DISSERTATION | DOCTORAL THESIS

Titel | Title

Making research FAIR

From principles to policy to practice

verfasst von | submitted by

Rasmus Kvaal Wardemann

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doktor der Philosophie (Dr.phil.)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 796 310 121
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Wissenschafts- und Technikforschung
Betreut von | Supervisor:	Univ.-Prof. Dr. Ulrike Felt Univ.-Prof. Mag. Dr. Barbara Prainsack

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Ulrike Felt and Barbara Prainsack, for their invaluable guidance, support, and mentorship throughout my doctoral journey. Their expertise and encouragement have been instrumental in my academic growth and the successful completion of this thesis.

I am also deeply indebted to my colleagues, Alexander Barth, Alice Vadrot, Anastasia Nesbitt, Andrea Schikowitz, Ange Pottin, Ariadne Avkıran, Artemis Papatheodorou, Bao-Chau Pham, Ben Skelding, Bettina Pospisil, Caitlin Harjes, Carsten Horn, Connor Hogan, Constantin Holmer, Dario Feliciangeli, Elaine Goldberg, Elisabeth Steindl, Esther Dessewffy, Florentine Frantz, Florian Bayer, Fredy Gàmez, Harry Wynands, Immanuel Bomze, Joseph Popper, Kai Strycker, Kamiel Mobach, Karin Neumann, Katharina, Paul, Katja Mayer, Katrin Hackl, Kaye Mathies, Laura Bomm, Lisa-Marie Ferent, Livia Regen, Lukas Schlögl, Maximilian Fochler, Meropi Tzanetakis, Michaela Scheriau, Natália Kancelová, Nikolaus Forgó, Nina Witjes, Noah Münster, Paul Dunshirn, Paul Trauttmansdorff, Philipp Kürten, Pouya Sepehr, Rafaela De Alcântara, Robin Rae, Roman Hansen, Ruth Müller, Sara Ramirez, Sarah Davies, Saskia Kaltenbrunner, Seliem El-Sayed, Stefanie Sterrer, Susanne Gahbauer, Theresa Henne, Ulrike Ritter, Vera Ulmer, and Virginia Vargolskaia, for their thoughtful feedback, constructive criticism, and support. Their insights and expertise have significantly enhanced the quality of this research.

I would like to thank the Research Platform Governance of Digital Practices and the Department of Science and Technology Studies at the University of Vienna for providing me with the necessary resources, facilities, and intellectual environment to conduct my research.

A large thank you also goes out to my informants, who supplied the basis on which this thesis is written.

Finally, I would like to express my heartfelt appreciation to my family and friends for their love, support, and patience throughout this challenging endeavor. Their belief in me has been a constant source of motivation.

Thank you.

Table of contents

Introduction

This thesis is about the production of knowledge, and ultimately about the elements and environments that govern the process. The thesis covers stories that pose the question whether there is a better way of doing it, and in the process interrogates what we consider "good" when it comes to knowledge production in the first place.

In the first instance, this thesis is about the FAIR principles, and the attempts at making knowledge[1] findable, accessible, interoperable and reusable. But as we follow FAIR's journey from principles to policy to practice, we get insight into broader topics: how we come to imagine better futures, how ideas about how the world should be take hold, and what happens when people try to build a different world, turning visions into practice.

But before we get into all of that, I'd like to set the scene and take you back to where this all started, from my perspective.

The story

The work on this thesis started for me four years before its completion, in October 2020. At that time, Europe, and the world as a whole, found itself in the middle of a pandemic it hoped was coming to an end. As politicians struggled to gain the public's trust and commitment needed to slow the spread of the pathogen, there was still one thing fueling their hope: science (see von der Leyen, 2020).

When COVID hit, one of the issues was that, unlike the regular spread of influenza, this new pandemic was exactly that: new. We were lacking knowledge on how it spread, how it circumvented our immune systems, how the disease developed in the body, and ultimately how it could kill us. To combat this lack of knowledge, in what was frequently called an "unprecedented" development, researchers started openly sharing their findings on COVID-related matters through preprint services like bioRxiv and medRxiv (Doudna, 2020; Gewin, 2020; Haseltine, 2021; Watson, 2022).

As the numbers of preprints skyrocketed, concerns about the quality of the knowledge within them started spreading (Lawrence et al, 2021; Watson, 2021). Throughout the years while the pandemic plagued the world, several studies had to be removed from pre-print services

[1] It is important to note that the form of knowledge addressed by the FAIR principles is a particular one centered on its digital embodiments – first and foremost "data", but also software and other artifacts (Wilkinson et al., 2016). We will get back to a discussion on the conception of knowledge later in this thesis.

because of allegations of poor science – infamously one recommending to treat the disease with the antiparasitic drug ivermectin (Watson, 2021). Part of the problem, some claimed, was the availability and quality of the research data (Gewin, 2020; Queralt-Rosinach et al., 2022; Basajja et al., 2022; Mitchell et al., 2022). If we wanted to combat COVID and later pandemics more effectively and efficiently, one researcher by the name of Barend Mons argued, we would have to improve our knowledge sharing practices and make our data FAIR, meaning findable, accessible, interoperable and reusable (Mons, 2021; 2022).

By following what was called “the FAIR principles” (Wilkinson et al., 2016), its proponents argued, dissemination of research findings could be made more efficient and effective, enabling machine analysis of data and ensuring that research would catch up with developments in information and communication technology, bringing science into “the e-science era” (FAIRPORT, 2014).

At the time the pandemic hit the world in late 2019, the FAIR principles had already been around for some five years. Formulated for the first time at a conference in Leiden in 2014 (FAIRPORT), and formally published in Nature Scientific Data in 2016 (Wilkinson et al.), the FAIR principles quickly gained significance as a core component of the European Commission’s Open Science strategy (EC, 2018).

So when I started the work on this thesis four years ago, in 2020, the FAIR principles had already gained a lot of attention. They had not only been endorsed by the European Union and made a mandatory requirement for research funding under the newly initiated Horizon Europe framework programme (EC, 2022), but also by the OECD (2018) and other governmental and non-governmental institutions and actors (OECD, 2020). This led research librarians Rosie Higman, Daniel Bangert and Sarah Jones (2019) to state that "Little in the research data field has gained such traction and universal acceptance as the FAIR data principles". But while the political momentum of FAIR impressed me, I wasn’t entirely sure exactly what it was.

On the surface, FAIR was a scientific/political initiative to improve research data management practices by computational means that resonated with claims made earlier by proponents of the Big Data paradigm, that the combination of ever growing amounts of digital data with computational methods of analysis would enable new and better forms of knowledge production[2]. While aligning with the basic idea of Big Data epistemologies, FAIR departed from this particular way of thinking in some distinct ways. First of all, Big Data was often concerned with the accumulation of more or less unstructured data stemming from

[2] See Anderson (2008) for praise, and Kitchin (2014) for critique.

users of digital platforms or the so-called internet of things. FAIR, on the other hand, concerned itself first and foremost with scientific data or health data from hospitals and other larger institutions that generated data for-purpose (Wilkinson et al., 2016).

Second, while growing amounts of data was seen as an undeniable good by proponents of Big Data, the people behind FAIR saw both potential and challenges in this development. More data, as they saw it, meant more knowledge, but it did not automatically mean more available and usable knowledge (Wilkinson et al., 2016). In the authors' experience, data could only be used for "knowledge discovery and innovation" effectively if it was curated properly – annotated with rich metadata, formatted according to established standards, and made readily available under clear circumstances (Wilkinson et al., 2016). If not, it would be rendered "reuseless" (Mons et al., 2017).

What the ideas behind Big Data and FAIR shared was that both considered data (in particular digital data) as inherently valuable for knowledge production. The main difference that set them apart was that proponents of FAIR claimed that unlocking or accessing the knowledge contained in data was a case of improving overall data practices, not only the tools and techniques used to analyze it. In other words, it was less about how to use the data, but more about how to share it in the first place.

This idea caught on within the European Commission who were trying to expand their Open Science strategy to not only include Open Access article publication, but also Open Data publishing through their new initiative: the European Open Science Cloud (or EOSC) (Burgelman, 2021). In an effort to realize the European Open Science Cloud and turn FAIR into reality, they commissioned three Expert Groups and funded a number of projects between 2015 and current day, 2024.

Between 2014 and 2024, several thousand people[3] worked on promoting FAIR – contributing to build an infrastructure for scholarly data reuse by suggesting citation practices, developing data formatting standards, updating data repository guidelines, making research policies, teaching FAIR practices, sharing research data and more. But as more and more people started engaging, it became clear that making research FAIR would be more difficult than first expected. What initially seemed relatively clear – making data FAIR means making it findable, accessible, interoperable and reusable – over time became muddy and uncertain. Already in 2017, just a year after the initial publication, the authors of the first paper started speaking out in conferences, keynotes and journal publications against what they saw as people misusing, misunderstanding and, essentially, undermining the aim of FAIR (Mons et

[3] Based on estimates by numbers of citations in Scopus (8442)

al., 2017; 2020; Mons, 2020; Jacobsen et al., 2020; GOFAIR, 2021). To make matters worse, not considering those who “misinterpreted” FAIR, it seemed like researchers’ uptake of FAIR practices was close to non-existent.

In spite of FAIR’s popularity in terms of research policies and paper citations, it remained a minimal niche in terms of scientific practice. Why weren’t people embracing FAIR wholeheartedly and adopting it as the creators intended? What was happening?

As the issues with uptake of FAIR became apparent, people who were promoting FAIR started working on diagnosing and solving the problem. Two common diagnoses were that either people didn’t understand it – which prompted the solution that they only needed clarification or to be convinced of the benefits of making their data FAIR (Mons et al., 2017; 2020; Mons, 2020; Jacobsen et al., 2020; GOFAIR, 2021; Higman et al., 2019; EC, 2018) – or that they didn’t have the available resources to make the transition – which prompted the solution that it should be made easier to become FAIR through external intervention[4] or technical innovation[5].

While there was a lot of engagement with FAIR, the overwhelming majority of it came from the computational life sciences, and outside of that it was mostly other data intensive disciplines[6]. As a result of this, most engagements with FAIR, even the most critical ones, did not question the tacit assertion that data sharing and reuse would be an unequivocal good. As critical scholarship from social sciences or humanities (SSH) was largely absent in terms of public engagement in conferences and journals – relegated to informal lamentacious conversations among colleagues – the underlying assumptions and potential unintended consequences remained unchallenged in public discourse.

“The gap”

That being said, FAIR had not escaped the radar of critical SSH scholars. In fact, there have been mentions of it in critical SSH scholarship, but in most of the cases FAIR had not been dealt with as a primary concern – only as a part of a larger whole: typically Open Science policies (see Leonelli et al., 2021) or research data reuse (see Thylstrup et al., 2022). As a result, critical scholarship from SSH typically glossed over the particularities and complexities of FAIR in favor of highlighting its commonalities with other phenomena. That being said, there were some notable exceptions, some scholars tending to FAIR not as an

[4] Examples include the assignment of data stewards or other support staff, different courses in FAIR practice, suggestions for more funding towards FAIR from institutional actors and transformation of incentive systems
[5] Examples include among other things FAIR measurement tools and Data Management Plan tools
[6] According to Web of Science citation statistics for Wilkinson et al. (2016)

expression of something else or part of a larger trend, but as an entity deserving attention and critical engagement in its own right.

The clearest example of this came from science scholar Louise Bezuidenhout in her article *Being Fair about the Design of FAIR Data Standards* (2020). While the article did go a long way in supporting FAIR in its “commitment to enhancing transparency and reproducibility in research” and as a “means by which to realise and improve open and data-driven science more generally”, Bezuidenhout articulated a concern for the lack of reflection in and around the FAIR movement. Aiming to ask questions rather than answer them, Bezuidenhout (2020) drew on scholarship in critical data studies (Kitchin and Lauriault, 2014) and sociology of standards (Busch, 2000) to suggest that an unreflexive implementation of FAIR standards for data management posed the risk of “epistemic injustice” (Fricker, 2007) towards under-resourced research environments (typically outside of the western hemisphere)[7].

At the same time, digital humanities scholar Mirjam van Reisen was making a similar argument (van Reisen et al., 2020a; 2020b) drawing on her experiences attempting to implement FAIR practices on the African continent through the GOFAIR[8] association. Also arguing for a sensitivity towards non-western environments, she went somewhat further than Bezuidenhout’s (2020) concern for the comparative lack of resources on the continent, adding an emphasis on the role of “African perspectives, objectives and epistemology” (van Reisen et al., 2020b).

From van Reisen’s perspective, FAIR could not only discriminate on the basis of a quantitative difference in resources, but also on the basis of qualitative differences in ways of producing knowledge – cautioning that a FAIR epistemology could threaten to suppress local epistemologies or be rejected on that basis (van Reisen et al., 2020b).

But in spite of these calls for critical engagement and reflexivity with the ethical and epistemic consequences of FAIR, and its large and increasing prevalence in research policies internationally, FAIR seems to have been overlooked in the field of Science and Technology Studies (STS)[9]. That being said, STS scholarship can still serve as a useful resource in helping us understand FAIR. Continuing this introductory chapter, I will present some avenues of STS research that will help inform the reading of this thesis.

---

[7] In spite of her clearly articulated call for reflexivity, Bezuidenhouts article received little to no attention among FAIR proponents, the paper gathering no more than 8 citations.

[8] An organization aiming to spread and develop FAIR. More about this in later chapters.

[9] Based on searching the keyword “Wilkinson et al., 2016” in 14 STS journals returning 10 articles that cite the FAIR paper with varying engagement with the concept

As it is one of the more common ways of understanding FAIR, I will begin with describing the work done on research governance, more specifically, the governance of research through policy. As Benoit Godin has demonstrated in his book *The Making of Science* (2009), the political steering of science and research has, since the postwar years, expanded, not only through the distribution of funds, but through the narrative framing of science, technology and social issues. Narratives, he argues, enable the governance of science through making things visible or invisible, by valuing certain things over others, defining roles and responsibilities, and producing concerns (Godin, 2009). In short, narrative frameworks are what enable policy to give science direction.

Though these narrative frameworks often take the form of disinterested and seemingly objective descriptions of matters of fact, scholars engaging empirically with the genealogies of these policy narratives show that this objectivity is little more than a performance (Mazur, 1987; Stone; 1997; Asdal, 2011). The aim of this performed objectivity, they argue, is to stabilize and close down controversies in order to enable the enactment of political power (Mazur, 1987; Stone; 1997; Asdal, 2011). Thus, the interrogation of science policy necessitates engagement with the seemingly objective and non-controversial through following policy narratives backwards to uncover their more apparent political origins. Scholars like Levidow and Neubauer (2014), Wyatt (2017) and Lahn (2020) suggest a particular sensitivity towards the practical work in shaping these policies and to what extent certain groups are allowed meaningful participation.

Further, Godin (2009) shows that policymakers have always struggled to understand the dynamics and relationships between science, innovation, and economic growth, and that in spite of numerous efforts spanning several decades, policymakers have yet to find a way of making science produce the outcomes they've desired. Still, they inevitably attempt through ever novel methods of governance to steer science towards serving what they dub "public interests" (Godin, 2009). In their attempts to gain control over science and exploit its perceived potential, policymakers have applied increasingly pervasive methods of accounting through measuring and quantification of scientific production of knowledge (Lapsley, 2008; Godin, 2009) – which in turn affects what kinds of knowledge that is produced (Müller and Rijke, 2017)[10].

In recent years, this quantification of knowledge production has reached a new frontier due to the increasing digitization of information in what is colloquially referred to as "data" (Kitchin, 2014; Rieder and Simon, 2016; Beer, 2019). I say "colloquially" because while the term "data" has been around for centuries, in recent years it has begun to be associated

[10] Speaking to the notion of "epistemic injustice" (Fricker, 2007) mentioned above

more and more intimately with the digital domain. Thus “data” in the way I refer to it here is to be understood as information stored in and expressed through a digital medium – bits and bytes, so to speak. According to the research of Rieder and Simon (2016), digitization of data has affected the quantification of knowledge production in two important ways. First, digitization of research data has made it more mobile and, thus, more easily available for measuring. Second, digitization of research data has also created new imaginations of its potential value as it becomes available for machine analysis. In other words, digital data is both easier to measure and perceived as more valuable, making it an attractive object for science policy (Rieder and Simon, 2016).

But these promises of digital data have been widely criticized and problematized from a strain of social sciences called Critical Data Studies (CDS). Scholars from this field often emphasize the prevalence of technological determinism and positivist epistemologies in the promotion of data-intensive methods (boyd and Crawford, 2012; Kitchin and Lauriault; 2014; Beer, 2019). This logic often frames data intensive methods as better or more efficient than traditional forms of knowledge production, which has caused CDS scholars to warn about the potential this “data imaginary” (Beer, 2019) has to suppress other forms of knowledge production, and thus also shape what we can know (boyd and Crawford, 2012; Kitchin and Lauriault; 2014; Rieder and Simon, 2016).

Critiques of this data imaginary often apply themselves to deconstructing the tacit assumptions that follow from it, of objectivity and efficiency. Since data often presents itself “as is”, critical data scholars argue, it often hides its origin, how it came to be “just data” (Gitelman, 2013). Empirical enquiries into the production of data therefore aim to show how data is never “raw” but is always a product of social and material practices that, once traced, undermine claims of objectivity and foregrounds the situated origins of data (Gitelman, 2013; Plantin, 2018; Crawford, 2021).

While efforts to share data between different researchers or to combine data from different sources tend to assume data has a rawness or objectivity, empirical accounts show that these efforts invariably run into challenges trying to overcome the difference between these data (Ribes and Bowker, 2009; Edwards et al., 2011, Borgman et al., 2014; Ribes, 2017). While successful implementations of data sharing and combination typically present themselves as seamless processes, scholars interrogating the production and maintenance of these efforts have found that data sharing and combination is always “seamful” (Inman and Ribes, 2018) and create “frictions” (Edwards, 2010; et al., 2011). This is to highlight that these efforts are not results of the natural and inherent qualities of data, but products of human labor, negotiations and symbolic forms of violence (Ribes, 2017; Poirier, 2019a).

What does all this mean for FAIR? STS scholar Christine Borgman put it succinctly in her book *Big Data, Little Data, No Data: Scholarship in the Networked World* (2015, p. 38):

> Policies of governments, funding agencies, journals, and institutions that are intended to improve the flow of scholarly communication often make simplifying assumptions about the ability to commodify and exchange information. While usually intended to promote equity across communities and disciplines, policies that fail to respect the substantial differences in theory, practice, and culture between fields are likely to be implemented poorly, be counterproductive, or be ignored by their constituents.

Taking FAIR seriously as a research policy that, by its own account, aims to enable and improve the sharing of research data, I follow the above strains of scholarship in two important ways. First, I follow their commitment to empirically informed critical scholarship, not readily accepting FAIR at face value, but interrogating its tacit and situated origins. Second, I follow their concerns for the consequences of policy and data sharing, intended as well as unintended. Put differently, I aim to interrogate and explore both how FAIR was shaped, and how FAIR shaped its surroundings as it traveled from principles, to policy, to practice.

The thesis

To achieve this, my thesis gives a more-or-less chronologically structured account of FAIR from the inception of the idea around the change of the millennium, to a more speculative account of its future. On the way there, I go through how it came to be picked up by the European Union as part of their science policy, the strategy employed by its proponents to develop and implement FAIR, what happened as actors attempted to translate FAIR from a set of prescriptions to actual practice, the negotiations around how to measure FAIR practices, the ontological tensions between the envisioned governmental structure of FAIR and scientific practice, and what the different visions of good science might mean for the future of FAIR.

Chapter 1 sets the scene for the rest of the thesis by giving a historical account of the genealogy of FAIR. By tracing back the ideas and developments that led up to the formulation of the FAIR principles in 2014, the chapter not only serves as a backdrop but also serves as a sensitizer for the prevalence of some of the ideas and attitudes that are promoted by certain proponents of FAIR, namely technological determinism, ontological singularism, and, as an embodiment of this, semantic web technology. Foregrounding FAIR's background, the chapter shows that while FAIR itself is often seen and presented as a novelty, it is a product of a long line of developments in the history of science and technology.

Chapter 2 continues looking backwards, but does so in order to explain how FAIR became taken up into the European Union’s science policy. It traces FAIR policy in the EU through a series of earlier policy initiatives – like Open Science, the Digital Agenda for Europe, and the Innovation Union – back to the launch of the European Research Area in 2000. The chapter notes that while the EU didn’t share the visions of the original proponents of FAIR, the flexibility of the concept allowed it to be presented as a response or solution to crises in the European Union.

Chapter 3 asks why – despite a seemingly high uptake – very few committed to practicing FAIR the way it was envisioned by the original proponents of FAIR. The chapter diagnoses this low “uptake” of FAIR as a product of exclusion of the subjects of the FAIR principles in the design and implementation process, and the sustained efforts to keep them out. Looking at the implementation strategy employed by the original proponents of FAIR, the chapter draws attention to a prevalence of determinism and reliance on deficit models in explaining the divergence between nominal and practical uptake.

Chapter 4 continues on the question of uptake of FAIR and asks why FAIR practices have become so different from the principles. This chapter turns its attention away from the principles and their authors and towards the people made responsible for turning FAIR into practice: a new type of academic professional called data stewards. Following the practice of these stewards, the chapter finds a widespread misalignment between the visions inscribed in FAIR and the environments in which academic practice takes place. Instead of rejecting FAIR on this basis, the stewards work as middle actors between principles and practice to negotiate a greater degree of alignment, modifying both FAIR and academic practice in the process.

Chapter 5 delves into the topic of evaluation and its ties to governance. Following three attempts at making FAIRness measurable it gives an account of the struggles to articulate and define what FAIRness should mean in practice. Looking at the different actors’ challenges in sourcing and performing authority, the chapter finds them struggling to escape a process I call the evaluator’s regress. As some of the actors realized the potential perversive effects of their metrics as tools of governance, they abandoned the goal of assessing FAIRness, settling on using limited metrics for assisting FAIRification, and thus circumventing the evaluator’s regress.

Chapter 6 explores the logics of the vision promoted by the original proponents of FAIR. It goes into the challenges inherent to semantic web technologies and the FAIR proponents attempts at solving or getting around them, revealing an underlying assumption about the ontology of scientific practice that prescribes what I call a “machine ontology”. Following the infrastructural aspirations inscribed into FAIR as they produce frictions with scientific practice, the chapter argues that the original vision of FAIR is necessarily incompatible with the ontological multiplicity of science – posing the question whether this vision of FAIR will ever be realized.

Chapter 7 picks up where chapter 6 left, and speculates about the future of FAIR. It does so by presenting four different types of visions for the future of FAIR: from the visions of its original authors, the visions of bureaucrats in the European Commission, the visions of the proponents of FAIR that found the a semantic web unrealistic, and the vision of the proponents of FAIR that found a semantic web potentially harmful. The chapter shows how these different future visions coexist in a state of tension with each other, and draws up ways in which this tension could resolve.

Through the thesis I go a long way in justifying the concerns of previous scholarly work on FAIR, research policy and data sharing. I find that there are clear cases of technological determinism in the beliefs and visions of the people promoting FAIR, building on an assumed linear trajectory of web technology going back to ideas espoused by “inventor of the web” Tim Berners-Lee – an idea that draws heavily on imaginations of data as inherently objective accounts of the real world (unless maliciously or incompetently made otherwise). I further find that policymakers attempt to employ FAIR as a tool for making science more innovative and accountable. Throughout, the thesis also demonstrates the seamfulness and frictions arising from trying to bring research data together in an infrastructure for the combination and sharing of research data.

Where my peers in critical scholarship on FAIR have mainly focused on its implications for (under-resourced) research environments in the global south, I demonstrate that these concerns are valid also within the European context – that “being FAIR” is not only a question of resources or a threat to radically different ways of knowing. The epistemic diversity within what some assume to be a relatively homogeneous environment – that being European science – turns out to be challenged by this new effort at governing science. In this way, my thesis is also an account of the complexity of scientific practices and the multitudes of factors governing it (not only formal science governance). FAIR, from principles to policy to practice is, thus, not a grand story about subjugation and resistance, but a humble story about tensions and the change it may produce.

Method and concepts

When I first started the work on this thesis back in October 2020 I had only a vague idea of what I would be working with. I had just finished my master's in Science and Technology Studies, exploring the digitalization of clerical work at a Norwegian insurance agency, and now I was about to start working on my doctoral project then titled "Knowledge-intensive human work in the age of digital transformation". While this at first seemed like a relatively clear framing to me, I quickly realized it still afforded substantial flexibility in terms of how I could interpret and operationalize it. That meant I still needed to find a case to study for my project.

I can't recall exactly how I stumbled upon it (probably just by googling "digital transformation science"), but soon two things caught my attention: High Performance Computing (HPC) and the European Open Science Cloud (EOSC). To explore if any of these two topics could be of interest, I reached out to a friend who was doing their PhD in nuclear organic biochemistry:

> R: Do you have any opinions on High Performance Computing and the European Open Science Cloud in your field?
>
> K: yes for the high performance computing and I've heard of the science cloud thing but it's super new and I don't think it's being used yet, at least to my knowledge.
>
> R: Ah ok. I'm thinking about writing my PhD about it
>
> K: Hmm. I would probably not do it on the cloud because as I said it is literally made this year and probably won't be practical for another few years at least. I think it's stupid, personally. Made by non scientists who don't understand how hard it is to communicate between fields.
>
> K: But the high performance computing is relevant and used in a lot of different fields.

In this case, my friend communicated two things that piqued my interest: 1) the EOSC practically didn't exist, and yet 2) it seemed very controversial. I decided to pursue the EOSC trail in spite of the warning not to, but still kept the HPC in mind. The next step was to find someone else who could serve as an informant and perhaps point me further along – to get the (snow)ball rolling, so to speak.

I contacted a senior Austrian bioinformatician and asked him about EOSC and HPC, and he insisted there was no reason for the bioinformatics community in Austria to become a part of the European Open Science Cloud – the existing infrastructures[11] were sufficient. On their own, his claims were interesting, but still a poor starting point to interrogate the EOSC. HPC,

[11] In particular the International Nucleotide Sequence Database Collaboration (INSDC) for data and Partnership for Advanced Computing in Europe (PRACE) for computing.

on the other hand, he could tell me, was something they used routinely. In other words, he more or less confirmed my friend's impression, albeit in a less explicitly EOSC-critical tone.

After this first exploratory interview I began a tumble to construct a research project. I tried to make sense of what I had learned by tentatively applying concepts like "knowledge systems" (Collins, 2010), "modes of ordering" (Law, 1994), and "sociotechnical imaginaries" (Jasanoff and Kim, 2015). I tried to find empirical cases to explore the depth and complexity of the issue, bringing me to look into phenological databases, data stewards, and meteorological citizen science. It was only after a few months that I felt like I had arrived at a good base for my research: The FAIR Principles (Wilkinson et al., 2016).

A sensible approach

Why is the above story methodologically interesting? Initially, for two reasons: 1) it describes the situated origins of this research project, and 2) it demonstrates the epistemic convictions of what the researcher (that being I) thinks of as "good research".

First off, I want to highlight the inherent uncertainty and unruliness of the first stage of a research project like this – where the investigator has a significant freedom to choose direction. In this uncertainty, what steers the search? Concretely, why did I not follow the recommendations to study High Performance Computing and instead pursue EOSC (which ultimately led me to FAIR)? Without the direction of a clearly pre-defined research project, I was left to choose for myself what to go after. While there are several strategies for selecting a research topic/subject/case, my research direction was inspired by what is sometimes called a sensible, affective, or bodied approach (see Star, 1990; Coenen; 1991; Pink, 2015; Mol, 2002; 2021; Land, 2023). The basic idea was to not let externalized logics guide the research, but, rather, use the researcher's organism to inform the design of the research. So instead of asking, "what is the most socially valuable?", "what is the research gap?", or "what will give the most citations, and increase my chance of further employment as a postdoc?", this approach encourages to listen to embodied sensibilities, instead asking, "what *feels* interesting, curious, surprising, funny, enraging, etc.?" (Star, 1990; Coenen; 1991; Pink, 2015; Mol, 2021; Land, 2023).

Through this approach, the idea goes, the researcher can open up new and unexpected avenues of knowledge. The caveat with the sensible approach is that "logics" are not deleted or removed, but rather made tacit. This demands a certain level of reflexivity to explicate. So what was it that sparked my interest for EOSC and later FAIR?

As mentioned, when coming into this project, I already had two years of STS under my skin, which sensitized me towards certain sociotechnical phenomena and dynamics. I already mentioned the classical STS sensibility towards controversy (see Jasanoff, 2019), which made my STS senses tingle at the first apparent mismatch between contemporary policy prescriptions and academic self-understanding: the visionary optimism in the European Commission's documents on the one hand and, on the other, the descriptions of EOSC as "stupid", "made by non-scientist who don't understand", and something irrelevant to good scholarly conduct. At the same time, nobody actually made a big fuss about it[12]. The controversy was not expressed as an actively unfolding conflict between opposing interests, it was articulated more as an irrelevance, as if to say "that's happening over there, but we're good over here". What was this? A non-controversial controversy? A controversial non-controversy? Whichever it was, it was curious, confusing, and maybe counterintuitive. It was worth a closer look.

In a sensible approach, the exploratory phase of a research project is in this way used to elicit an intuitive response to guide further research. Through this sensible guidance, it takes the project out of the exploratory phase and into a more focused approach – to pursue the path that resolves the feeling, so to speak.

This leads me to the second point, and why I choose to highlight that I *arrived* at the FAIR principles. First and foremost, this way of speaking of it is an attempt at illustrating a commitment to empirically oriented research as opposed to conceptually driven research. In Science and Technology Studies, this is expressed in the tradition of "following the actor".

While following FAIR around gave me something to anchor my research to, it would shortly become increasingly frustrating. No matter where I went, no matter who I asked, I could not get a clear answer as to what FAIR really was. Sometimes FAIR would be a policy object, other times a set of values. Sometimes it would be a technological infrastructure, other times just the vision of it. It became clear that it was far from clear what FAIR actually was.

What I was experiencing was not only a product of my field of research, but also my methodology. I was convinced that FAIR was one thing, an ontologically singular entity. Following this conviction, logically, my confusion had to be a product of misunderstanding, either on my end, my interlocutors' end, or both. This narrative of misunderstanding was also readily available in my material through frequent proclamations by the original proponents of FAIR that "clarified" what FAIR was and wasn't[13]. In my attempts to resolve my confusion, it was tempting to take these clarifications at face value and use them as the starting point for

[12] Unlike the "hot situations" described by Callon (1998)
[13] Notable examples being Mons et al. (2017) and Jacobsen et al. (2020)

my research project. This is, according to Francis Lee (2022), the risk of “following the actor” as a method, as it can lead us to “reifying our interlocutors’ matters of concern”. Accepting the explanation that FAIR seemed to defy definition because of the prevalence of misunderstanding, would not only lead me to accept the ontology of FAIR, but it could also lead me to adopt my interlocutors’ problem definition. Had I accepted this explanation, my research could have turned into an interrogation of why people don’t understand FAIR, why they refuse to accept it, or whether FAIR is good or bad. Ontological singularism, in this way, lends itself to an unwitting normative stance – taking sides in a conflict of matters of concern (Lee, 2022). Wanting to remain critical in my scholarship, I decided this was not the way to go. But this brought me back to the question: “what is FAIR?”, and introduced a new one: “how can I study something that seems to defy a stable definition?”

Studying things in the making

Luckily, I was not the first scholar to ask this question. In his book *Ontologies for Developing Things* from 2010, STS anthropologist Casper Bruun Jensen draws out some conceptual and methodological approaches to studying what he calls “developing things”. The problem, he argues, is that certain phenomena seem to resist ontological stability – they are “not yet” objects (Jensen, 2010). Jensen warns that identifying this “not yet”-ness can be tricky, as informants and materials often treat these phenomena as ontologically stable, as “existing” on their own accord, while they in practice are highly contested (Jensen, 2010). To cope with this conflict, Jensen suggests that the researcher should avoid the desire to uncover the “actual” meaning of the thing, and instead apply the principle of symmetry (see Bloor, 1976), treating each account of the thing, every ontological claim, as equal (Jensen, 2010).

So how does the researcher decide what to follow, if all claims should be treated as equally true or untrue and it’s unclear what the object of study even is? Drawing on Foucault (1990), Jensen suggests a “nominalist” approach, which supposes that all things are “in the first instance” words (Jensen, 2010). Studying developing things, one must therefore follow the word – in my case “FAIR”.

Methodologically, though, what are we studying when we follow a word? Do we study where it crops up, how frequently it is used in certain contexts, or its grammar and syntax? While these approaches can give us some insights, Jensen argues that we should focus our attention not on what the word *is*, but what it *does*. This approach is partly derivative of a “practical ontology” (Mol, 2002), an epistemic conviction that reality is a product of practices, not materials, and can thus only be judged or assessed by looking at processes. This methodology is by Mol (2002) and others (see Bueger and Gadinger, 2018) called "praxiography".

One of the challenges that I immediately encountered when researching this word, "FAIR", was that there didn't seem to be much practice around it[14]. I had seen the word so many times in policy documents, reports and journal publications, but whenever I went looking for researchers who were practicing FAIRness, so to speak, it never led anywhere. In 2020, when I started my research, it seemed like barely anyone had even heard of FAIR, and the ones who had, they didn't express any particular knowledge about what it entailed for them. Mostly, they would say that they went about their research in the same ways as before, but now their research proposals had the word "FAIR" in them. According to the researchers, "FAIR" barely did anything at all. To make matters worse, as I shadowed one of my informants, I noticed that the word "FAIR" – the very thing I was studying – never really came up in their practical research at all.

FAIR was, in other words, largely contained in the aforementioned documents. How would I study practices when my main corpus of materials was static text? The solution to this conundrum turned out to come about through a broadening of my perception of what constitutes practice, and an attentiveness not to what a document *is* but what a document *does*.

In her work on policy documents, STS scholar Kristin Asdal highlights that the view of documents as mere representations of a practical world "outside" is not only unproductive, but also naïve (2015; Asdal and Reinertsen, 2020). Documents do much more than just represent, they also act. Documents, Asdal argues, have a transformative capacity to work upon their environments (Asdal, 2015). This allows us to take a practice-oriented approach when we're studying documents, constituting an ethnographically and praxiographically informed methodology. Elements highlighted by Asdal (2015; Asdal and Reinertsen, 2020) include an attentiveness to the relationships between documents and how certain documents are being mobilized – not only in other documents, but also other practices. Developing a sense for the practices of documents, Asdal argues, is achieved through a close engagement with them over time, "getting to know them". Like Mol (2021), Asdal (2020) thus emphasizes the need for a process of sensitizing over time, a sustained interaction with the documents, in order to make the particular emerge from the otherwise seemingly trivial.

[14] Not to mention that "FAIR" is not readily searchable on the web, as search engines do not allow case-sensitive searches, and thus do not distinguish between "FAIR data" and "fair data" for instance.

Following Asdal's approach in my project, I read through and skimmed documents by the hundreds. In the first period, I attempted to identify the most important ones: the ones coming from what I considered significant actors (like the European Commission's Expert Groups or the authors of the FAIR principles) and the ones that were frequently referenced. Initially, little came across as self evidently interesting. My impression was that these documents mainly contained information about what FAIR was, what would/could/should be done to promote FAIR, or justifications as to why it was needed in the first place. As mentioned earlier, there were, to me, no clear expressions of controversy. It was only over time, after reading new documents and re-reading old ones, that I started picking up on the silent traces of controversy.

The earliest and most evident example of this was the troubled pairing of Open Science and FAIR. The conflict between these two terms, and the actors promoting them, was never explicitly expressed in any documents. The ones that did address the issue, did so in attempts to "clarify" the relationship between the two, not to oppose the combination/separation of them[15]. But what were these documents doing? By "clarifying", they were not only documenting what FAIR already was, they were practicing FAIR into being. Notably, they were doing so in different directions – they were practicing FAIR differently and adding tension to this "developing thing".

My gradually increasing awareness around this conflict and controversy over time made me notice muffled articulations of discord where I previously had seen none. Getting a sense for the Open Science/FAIR controversy, further sensitized me to other tensions between the European Commission and other proponents of FAIR – a recurring theme in this thesis. It became clear that what was articulated as stable and ontologically singular by the individual documents, taken together, actually was a developing and ontologically multiple FAIR.

Like its ontological underpinnings, praxiography's epistemology is multiple (Brives, 2013). Praxiography does not presuppose any particular theoretical commitment in studying developing things. Jensen (2010), Mol (2021) and others (see Bueger and Gardinger, 2018) emphasize the need to be empirically driven, to be agnostic about the relations between things and practices, not to assume that things work one way or another before having made any observation that it does.

[15] Mons (2017) and EC (2018) are good examples of this.

Sensitizing concepts

Being epistemically agnostic, however, does not mean to pretend that one doesn't believe that processes may follow certain patterns. So while praxiography rejects a theory-guided approach, it encourages the use of what Herbert Blumer (1954) called "sensitizing concepts" – the use of theory or concepts, not as frameworks to understand the world, but as sensemaking repertoires "that helps scholars to attune to the world" (Mol, 2010). Through this way of using concepts, the researcher does not apply an external logic to their object of study, but is guided by them to recognize patterns and seek out particular relationships empirically (Bueger, 2013; Bueger and Gadinger, 2018).

Approaching theory as sensitizing concepts is in this way similar to the process I described as a part of my initial exploratory phase of research. They are cognitive constructs that make the researcher sensitive to certain dynamics. In this way, the distinction is first and foremost a formal one between the tacit and the explicit application of sensemaking tools – consciously exploiting a bias that is present nonetheless (Latour, 1988). The difference, then, is that the use of "sensitizing concepts" invites active reflexivity from the researcher's side in what can be useful or otherwise support the research project.

Following the praxiographic principle of empirically driven research, the sensitizing concepts I've used in this research project varied for each chapter I wrote – from using the classical science studies concept of the experimenter's regress (Collins and Pinch, 1993) to explore the difficulties of measuring FAIRness; to the political anthropology concept of middle actors (Lipsky, 1969) to understand the role of data stewards; and the media studies concept of domestication (Silverstone and Hirsh, 1992) to explain how FAIR became multiple. While these concepts have been productive in analyzing specific cases, I could not have applied them across the entire thesis. With the aim of consistency, to keep a red thread and ensure that the thesis remained a cohesive whole, I've throughout relied on sensitizing concepts from the literature of infrastructure- and future studies.

Why infrastructure studies? When we think of infrastructures, we typically think of big material constellations – of roads, pipes, and wires that connect and allow the transportation of other things – like people, gas and electricity. In this sense, the FAIR I'm researching is not what you'd typically consider an infrastructure: it is a word. But social studies of infrastructure have challenged this simple definition of infrastructures as material constellations, complicating the matter and giving us new ways of conceptualizing what an infrastructure is, and, perhaps more importantly, what an infrastructure *does*.

The contemporary strain of infrastructure studies typically places its origins in the mid 90's with Geoffrey Bowker, Susan Leigh Star and Karen Ruhleder's work on Computer Supported Collaborative Work (Bowker, 1994; Star and Ruhleder, 1994; 1996; Bowker and Star, 1998; 1999). Their work marked a departure from earlier infrastructure studies by shifting focus from the things and people who constituted the infrastructures, to the consequences of infrastructures – how they reshape relationships between people and things (Bowker, 1994; Star and Ruhleder, 1996). According to this view, infrastructure is not a question of what, but *when* (Star and Ruhleder, 1996) – only when something acts as infrastructure does it become infrastructure. Star (1999) uses the example of stairs, how they act as infrastructure for the able bodied, but represent a barrier for people in wheelchairs. In other words, infrastructures are not only practical, but also relational (Star and Ruhleder, 1996).

By shifting the attention towards what infrastructures do, rather than what they are, Bowker (1994) argues, the scholar performs an "infrastructural inversion", bringing attention to some of the less visible components of infrastructure: the boring, mundane and trivial. This methodological approach prompts an understanding of infrastructure that is less concerned with the foregrounded elements of infrastructure – the roads, pipes and wires – and their intended outcomes – personal transportation, waste management and energy distribution.

A concrete example of how I applied infrastructural inversion in my research was in the early stages of my research, when I was looking at a particular infrastructural element of FAIR: the data repositories. Reading about and talking to people about ESOC and FAIR, I understood that data repositories were seen as an integral part of making research FAIR. When I started actually looking into it, reading about different repositories, talking to the people managing them, the people using them, it became clear that while some repositories implemented FAIR standards, they were not being used by researchers. A striking example was that while repositories contained large numbers of datasets, almost none of them would ever be downloaded – an obvious requirement for actual data sharing. In this case, looking at the structures of the repositories and the standards they implemented didn't make any sense, because the consequences of them (aside from the time and effort spent on implementation) were negligible.

Applying infrastructural inversion, makes the research less concerned with materiality, and more concerned with how infrastructures create the need for maintenance, create dependencies, discriminate for and against certain practices through the enforcing of standards (Star and Ruhleder, 1996).

Later works have combined infrastructure studies with a practical ontology approach to highlight the ontological consequences of infrastructures – how they are not only affecting a singular reality but reshaping and producing multiple realities (Larkin, 2013; Jensen and Morita, 2015; Carse and Kneas, 2019). The long-standing concern in infrastructure studies, with the tension between local practices and infrastructural drive towards globalized practices (Hughes, 1983; Star and Ruhleder, 1996; Hanseth, Monteiro and Hatling, 1996; Bowker and Star, 1998; 1999), then becomes a question not only of what practices are permitted by certain infrastructures, but what worlds they allow us to live in (Busch, 2000; Larkin, 2013; Jensen and Morita, 2015; Carse and Kneas, 2019; Poirier, 2019). Since the consequences of infrastructures are often unintended and resulting from their less visible components, Jensen and Morita (2015) warn that infrastructures often bring about "barely conceptible reconfigurations", or "silent transformations". As such, infrastructures are prime examples of what Annemarie Mol (1999; 2002) calls "ontological politics" – the reshaping of our realities through seemingly non-political means. This, Mol (1999; 2002), Larkin (2013), Jensen and Morita (2015) agree, calls for an attentiveness to the practical-ontological consequences of infrastructures *as* political interventions.

But this focus on the practical consequences of infrastructures only brings us back to an earlier issue. While FAIR as an infrastructure could have consequences for scientific practice – and, indeed, that was the whole point – they remained largely unrealized. While the aim of the people promoting FAIR was to make research products Findable, Accessible, Interoperable and Reusable, their main achievements in the period I studied FAIR was the production of documents, spread of awareness, and discussions on what FAIR really meant or should mean. Regardless of the prescriptions of how science *should* be done, very few would actually follow them. Truly, FAIR seemed to be an infrastructure mostly in terms of aspiration.

As Carse and Kneas (2019) would say, the FAIR infrastructure was "unbuilt and unfinished". While the state of the FAIR infrastructure surely was inconvenient for studying its consequences, Carse and Kneas (2019) make the argument that infrastructures are, in fact, never built or finished to completion – their ontologies never become total, never achieving what John Law calls "hideous purity" (1994). But by engaging with infrastructures, it becomes clear to us that they do strive towards a certain totality. FAIR "wants" to become a global infrastructure (Wilkinson et al., 2016; Wittenburg and Strawn, 2018). Carse and Kneas (2019) point to this as the necessary dual nature of infrastructures, as part material and practical, and part fictional and visionary.

Because of this divide between "real" and "fictional", the study of infrastructures, Carse and Kneas (2019) argue, always needs to give attention to not only the practical consequences of infrastructure, but also the controversies and frictions that arise from trying to realize the visions behind them. Other scholars have also highlighted how fictions or narratives are used to stabilize and diminish friction from infrastructuring (Star, 2002). This requires a sensitivity to the temporal aspects of infrastructure, in my case, particularly how the future comes to matter in shaping practices that contribute to realizing the infrastructure.

As Barbara Adam has noted in her seminal work on time and temporalities (2008), “as the realm the ‘not yet’, the future is not accessible to the senses”. Therefore, it could seem incompatible with praxiographic methodology which is so reliant on exactly “the senses”. But it only seems this way. While the future is typically understood as something that has yet to come into existence, the present is rich with enactments of the future, what scholars routinely refer to as “future-making practices” (see Halford and Southerton, 2023). As noted by Michael (2017), futures are everywhere, from political activity to everyday consumer goods. Michael contends that while we tend to focus on the dramatic stories of significant change – so called “big futures” – we should also be attentive towards the “little futures” as well (Michael, 2017). This means that a study of futures cannot solely focus on the explicit visions, expectations and imaginaries, but has to take seriously its mundane expressions as well (Michael, 2017). Since futures can be anywhere, a researcher cannot afford to look only where it would be most obvious.

In the case of FAIR, there are both obvious and less obvious futures at play. There are the clearly enunciated big futures of “revolutionary infrastructures” (Wittenburg and Strawn, 2018) and the tacit little ones expressed in different methods of measuring FAIRness[16]. Taking infrastructure seriously in my research project, thus, meant to remain sensitive to the consequences of these and other forms of future-making, no matter how mundane, and no matter how bold.

Sites of articulation

Choosing to approach FAIR praxiographically and through the lens of infrastructure studies led my research down many different paths. My first intuition was to look into the tensions between the original proponents of FAIR and the people in the European Commission, where the discussion largely centered around whether, and to what extent, FAIR and Open were compatible. Becoming sensitized to this muted conflict, I started noticing the different ways it became expressed or enunciated. Once I became aware of the means of articulating this controversy, I also started finding other controversies elsewhere. Everywhere the word “FAIR” came up revealed itself as a site of articulation where actors would make ontological

[16] See chapter 6 and 5 respectively for interrogations of these

claims either explicitly or tacitly. For instance, the desire to measure FAIRness was not only about being able to tell how FAIR a dataset was, it also challenged the notion of what FAIR was in ways that were non-obvious also to the proponents of measuring[17]. Taking the ontological politics of FAIR and its articulations seriously, led me to interrogate not only its consequences, but also its origins in an attempt to inform how the different articulations and enactments of FAIR gained their particular ontologies.

These sites of articulation were heterogeneous along several dimensions and scales, across space and time, in different arenas or domains, through various modalities. The multi-sited nature of my field presented the issue of selection, that there were more places that I could follow FAIR than was practically feasible[18]. Chapters on data management plans, librarians, and semantic vocabularies remain unwritten, sacrificed in favor of the seven chapters that did make the cut. The selection was a product of several considerations, from the more pragmatic concerns with practical feasibility, to the epistemic commitments towards generating a coherent and cohesive rendition of the topic.

Communicating praxiographic research

As a result of this, the chapters of this thesis are a mix of genealogies, contemporary histories, and speculation; based on documents studies, participant observation, and interviews; and using various theoretical concepts. In spite of their differences, or perhaps because of them, I've attempted to structure them somewhat like a narrative – one that progresses not only chronologically, but also logically from the development of FAIR itself, to the negotiations around its attempted implementations, and the possible future repercussions of the enactment of FAIR.

Since praxiographic research doesn't follow any strict procedures or recipes for how to study a given phenomenon, there are also no given ways to represent the findings (Bueger and Gadinger, 2018). Humphreys and Watson (2009) argue that since the guiding logic of inquiry is to follow the researcher's senses and sensibilities, the resulting knowledge will necessarily be a product of the researcher's experiences. Representations of this knowledge, then, should not attempt at making any claims on truth, but reflect its situated origins (Humpreys and Watson, 2009).

Still, the messiness of the praxiographic approach lends itself poorly to effective communication of research findings. Bueger and Gadinger (2018) therefore suggest to draw

[17] See chapter 5
[18] See Dujin, 2020 for a discussion on this issue

inspiration from documentarists, in the sense that they “construct narratives that include different places and voices and move between these in one story without creating a cacophony”. Crafting cohesive and engaging narratives in this way necessarily demands deletion of detail and the brushing over of certain detractions – much in the same way any other style or method of academic writing would. The difference then becomes by which logic the textual representation of knowledge is crafted. According to Bueger and Gadinger (2018): “Identifying what matters, and communicating it to a wide audience: that is the point of doing praxiography”. Whether this thesis will reach “a wide audience” is yet to be seen, but, dear reader, I hope it will reach you.

Materials

This thesis is an empirical engagement with the making of FAIR science from principles to policy to practice, and, as such, is based on a collection of different empirical materials. While I've given a broad sketch of what kinds of materials I've analyzed in this thesis, I will take this section to enunciate it and articulate it more clearly and systematically. I have categorized my empirical material according to three categories: documents, interviews and participatory/observational. To be clear, these categories are not pure, but overlap somewhat. More on that later.

Interviews

While the use of interviews is perhaps most evident in chapter 4 where I talked to data stewards about their experiences with and opinions on their practices, interviews have been part of the backbone for the entire thesis. From the "interview" I had with my friend about the European Open Science Cloud and High Performance Computing in the very first days of my project, to the interview I had with former Director General Robert-Jan Smits later on, interviews have informed my decisions on where to look, and sensitized me to what to notice. As such, interviews have been part of both my empirical material in a conventional sense – as the basis for my analytical writing – and my research process overall – as information for the decision making process of the project.

In the following list, I include all informants that have contributed to the final thesis in one way or another. Not all of them are quoted, and many of them are pseudonymized with a first name only. Only those who are identifiable by function of their role or are necessary to identify for other reasons will be named (this has been agreed to under explicit and informed consent).

Identifier	Role	Mode	Time of interview(s)
Greg	Bioinformatician	Video interview	2020.10
Lisa	Data steward	Video interview	2020.11
Paulina	Data steward	Video interview	2021.01
Tristan	Data depositor	Video interview	2021.02
Tamara	Biologist	Face to face interviews	2021.02-2023.11
Victor	Data steward	Video interview	2021.03
Andreas and Iben	Data steward	Video interview	2021.03

Tom	FAIRport conference participant	Video interview	2021.11
Niklas Blomberg	FAIRport conference participant and head of ELIXIR	Video interview	2021.12
Robert-Jan Smits	Former Director General of Science, Technology and Innovation	Video interview	2022.01
George Strawn	FAIRport conference participant	Video interview	2022.02
Mirjam van Reisen	GOFAIR member	Video interview	2022.02
Myles Axton	FAIRport conference participant and former journal editor	Video interview	2022.03
Fred	Biologist	Face to face and email interviews	2022.04-2023.01
Alex	Data steward	Video interviews	2022.12-2023.10
Pierre	Funder	Video interview	2023.02
Maria	Data steward	Face to face and email interviews	2023.02-2023.10

Table 1

It is perhaps not as striking as it used to be, but it might be noteworthy that most of these interviews were conducted via video call. The constraints of the Covid pandemic were quite evident. Video interviews meant that the flow of the conversation was difficult to maintain – the lack of eye contact and the slight asynchronicity made it a challenge reading the small cues informing when to speak, when to remain silent, when to dig deeper, and when to let something go. It also meant that an interview had a definitive start and a definitive end – no "do you want a coffee" with the ensuing conversation upon arrival, or "oh, by the way…" on the way out of the building. The strict delineation and formality of the situation made less room for the serendipitous and unexpected moments that characterize good ethnographic work – something I did get to experience during the empirical work that I did conduct in non-virtual space.

That being said, the pandemic also had its affordances. With people being used to video calling, it was relatively easy to set up and schedule interviews. My impression was also that the ability to record the interviews without the need for a secondary device not only made the sound quality better, but also made the informants more at ease – an impression informed by the recurring glances towards my recording device in the non-virtual interviews. Doing interviews in virtual space also meant I could reach more people than I would have been able to in non-virtual space – like with George Strawn in the US, or Niklas Blomberg in the UK, for that matter.

Still, you might notice some absences in my list of informants – notably Barend Mons. Originally, my intention was to interview more of the original proponents of FAIR, but after months of being turned down, ignored or even ghosted by the people I wanted to talk to, I decided to change strategy. Especially in the case of Barend Mons, there was an abundance of presentations, publications, interviews and other recordings available on the internet where people spoke more or less freely about their relationship to or positions on FAIR. While I wouldn't go as far as to consider this a perfect replacement for an interview, it did satisfy some of the informational needs and desires in the research process.

That being said, looking at the interview as a privileged source of empirical information can be misleading. While interviews can give a different insight into a person's experience, perspective and sensemaking of a given phenomenon than documents can, neither of them have more of a claim to "truth" or "reality" than the other. Documents and interviews are both products of context and are as such situational as well as relational. Seen this way, interviews offer the benefit over documents that the interviewer has more influence on the context – essentially: you can ask any questions you want (including follow-up questions). Still that does not mean the interviewer has full control. The interviewee has the option to say what they want, not say anything, and present themselves however they themselves please. This last point became particularly evident in one of the interviews, where the informant largely reproduced the same narratives and talking points that were present in the documents I'd studied.

Acknowledging that any articulation is a representation has some analytical consequences as well. First and foremost it necessitates treating statements as *claims*. This means that one can't say "the informant felt that things were going well", but rather that "the informant expressed feeling that things were going well". Same for "factual" statements like "the event happened in 1983", which should be expressed as "the informant recalled the event happening in 1983". Some times this epistemic commitment opens up for the interrogation of *why* certain claims are formulated (why would they say that?), and other times it merely prompts an inquiry into the accuracy of the claim (did it actually happen in 1983?). Throughout this thesis, this commitment is made clear through formulations that highlight the situational origin of claims. When claims are formulated otherwise, it is an expression of confidence from my side that the claims trangress their situational origin and are cross-checked with other sources. This, of course, is no less true in the case of documents.

Documents

Large parts of this thesis is based on an ethnographic engagement with documents (Asdal, 2015; Asdal and Reinertsen, 2020). While documents is a wide category that can include multimedial material, such as video recordings of conferences for example (see Pink et al., 2016; Shankar, Hakken and Østerlund, 2017), I will here use the term to refer to documentations in a strictly textual format. To be clear, this distinction is purely pragmatic and has no intention of making any ontological claims.

Furthermore, I only aim to list the most central documents for this thesis, as the true amount[19] is too great to practically or meaningfully present. What documents have been used in the specific chapters will hopefully be clear from the references and bibliography. The documents that are listed here, though, are the ones that have served as anchoring or reference points in the early stages and throughout my thesis. The documents listed underneath can be seen as the nodes in my document ethnography, the documents that are linked to other materials I've used. The list can, as such, also be read as a recommendation on where to start if one were to do a similar project to the one I have done.

Name	Author	Type	Year published
Which gene did you mean?	Barend Mons	Journal publication	2005
Science 2.0 (change will happen….)	Burgelman, Osimo and Bogdanowicz	Journal publication	2010
Ceci n'est pas un hamburger: modelling and representing the scholarly article	Pettifer et al.	Journal publication	2011
Jointly designing a DATA FAIRPORT	FAIRPORT conference participants	Conference report	2014
Realising the European Open Science Cloud: First report and recommendations of the Commission high level expert group on the European open science cloud	European Commission High Level Expert Group on the European Open Science Cloud	Report	2016
The FAIR Guiding Principles for scientific data management and stewardship	Wilkinson et al.	Journal publication	2016
Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud	Mons et al.	Journal publication	2017
Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data	European Commission Expert Group on FAIR Data	Report	2018
Common Patterns in Revolutionary Infrastructures and Data	Peter Wittenburg and George Strawn	Working paper	2018
FAIR Principles: Interpretations and Implementation Considerations	Jacobsen et al.	Journal publication	2020

[19] Somewhere in the hundreds, perhaps even the thousands, depending on how liberal you are in your understanding of documents (counting web-pages, for instance, would lead to quite a high number). I am sure this is a relatable case for anyone doing document studies in contemporary science.

Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story)	Jean-Claude Burgelman	Journal publication	2021

Table 2

The work on this thesis, as mentioned earlier, started with looking into the European Open Science Cloud, and, for this reason, the first documents I looked into were the publications from the European Commission. In this regard, the *Turning FAIR into reality* report (or "TFIR", as it is sometimes abbreviated) (EC, 2018) was an important point of entry. It is a rich document that contains a variety of different types of text: there is the foreword by then commissioner for Research, Science and Innovation, Carlos Moeads; the preface by chair of the expert group and director of CODATA, Simon Hodson; a wealth of problem definitions; suggested solutions; definitions of concepts, from FAIR to semantic technologies to metadata; and 135 links to other documents. It is a document that is explicitly political and normative, making both tacit and explicit ontological claims, sketching out a field of knowledge, and it's all presented with the authority of a group of Commission-mandated "experts". Thus, TFIR served as a great starting point for staking out a direction for further research. As I made my way through a growing corpus of documents, the report became less and less central to my analysis. So while it is not the empirical center of my analysis, it might be the methodological center of the thesis.

So what is the empirical center? With little doubt, it is *The FAIR Guiding Principles for scientific data management and stewardship* (Wilkinson et al., 2016), which has been referenced over 40 times throughout this thesis. The only parts of the thesis that do not mention this document are chapter 1 and 2, which deal with the time before the first publication of the FAIR principles. Unsurprisingly, it is also the document that is referenced the most by other documents that I've worked with during this project. But when it comes to what documents the Wilkinson et al. (2016) paper itself references, it is perhaps the absences that became the most evident to me. Most of the documents analyzed for the chapter on the history/genealogy of FAIR are not to be found in the bibliography of this publication – those had to be sourced from elsewhere, meticulously following traces from the publication history of Barend Mons, looking up connections mentioned in presentations or secondary documents, like the FAIRport report (FAIRPORT, 2014).

While the documents listed above are what I would call "consensus documents" – meaning relatively stable documents without visible internal disagreements or discussions – several of the documents I have been working with have been the opposite: very much in the making and with explicit tensions. The best example of this is perhaps the open draft of the TFIR report – a google document full of comments from people external to the expert group, voicing their concerns over different formulations and claims made in the draft. Documents like these were particularly helpful in identifying controversies, but also the product of these controversies by comparing the drafts to their final versions. The github repository of the RDA FAIR maturity indicator group was also a great documentation of the controversies and negotiations that would lead up to, but become absent in, the final reports and publications. In a similar vein, meeting documents from different groups gave rich insights into the processes behind different activities and publications. Although these were not necessarily meant for public consumption, it is important to acknowledge that they still don't reflect some honest and naked truth, but are contextual representations of events.

Documents, as such, can thus be considered a form of *space* where action and interaction takes place. In my thesis I've treated documents as *sites of articulation* – arenas that stimulate or force the explication and enunciation of the assumed, tacit and invisible. In this way, documents not only "reveal" things, but in effect also create or perform them. This notion of documents therefore allows a more practice oriented approach, such as the praxiographical approach I've based my research on.

Finally, I will have to note that much of the work I have done on documents these past four years would not have been possible without the Internet Data Archive. The amount of rotten links, 404's, unwanted redirects, and overwritten pages I have encountered has probably been in the hundreds. Without the *Wayback Machine* I would not have been able to write this thesis – which I realize is both ironic and fitting, as it means that much of the "data" I have used was only *barely* findable and accessible[20].

Participatory/Observational

From early on in the project, I was convinced that I wanted to do ethnographic research on FAIR. This was, unfortunately, made next to impossible due to the restrictions of the Covid pandemic. As mentioned, I could not meet up with people to do interviews, I could also not go to someone's lab and watch them work, and I could not go to conferences either – or, actually, could I? Yes and no. I could not physically transport myself to any conferences, but I could participate in conferences virtually – most conferences during this time were virtual anyway. Already after going to my first conference for this project – the EOSC symposium in October 2020 – it became evident that virtual conferences had both some drawbacks and some affordances that the non-virtual ones didn't have. The drawbacks were the

[20] Several times, documents have been neither findable nor accessible at all either. While this has been more than a slight nuisance, I might still consider myself lucky – at least some of them were saved.

aforementioned lack of serendipitous and unexpected moments, and the near impossibility of making any sort of connection with any of the other participants. While not making up for these drawbacks, there were two affordances that I came to appreciate greatly.

The first notable affordance of virtual conferences was that many of them had a chat function. This turned out to be a great source of discussion, often more interesting than the presentations that the participants were (supposed to be) chatting about. One thing that made the chats so rich was that they clearly elucidated some of the controversies. A recurring topic would for instance be the difference between Open and FAIR. The other affordance of virtual conferences was the fact that most of them were recorded for posterity, so anything that I might have missed or forgot, or that my notes only half-covered could be returned to later. But this is also where it gets complicated, because when the conference is documented so thoroughly, when do you call it a document and when is it observational material? In some cases, I was able to participate actively in the conferences, asking questions in the chat, so that made the distinction between participant and non-participant observation, but in the cases where I didn't, it was unclear to what extent there was any difference in being there "live", so to speak, and watching the conference afterwards. Because of this fuzzy boundary, I've decided to include here the conferences I did not attend live as well. This is, again, a pragmatic distinction.

Name	Type	Time
SWAT4LS Workshop - Semantic Web Applications and Tools for Life Sciences	Conference	2009
RDA plenary	Conference	2013, 2014, 2016, 2021
International FAIR convergence symposium	Conference	2020, 2022
FAIR festival	Conference	2021
CODATA Virtual SciDataCon	Conference	2021
Open Science FAIR	Conference	2021
FAIR Data Austria Online event	Conference	2021
EOSC Nordic General Assembly	Conference	2021
Horizon Europe Infor Days	Presentation	2021
EASA webinar: ISE policy paper on the transition to open science	Seminar	2021
Data sharing seminars for societies	Seminar series	2021-2022
Fred working	Shadowing	2022
FAIRsFAIR final event	Seminar	2022
DICE/EOSC pillar: FAIR data management gaps and solutions	Seminar	2022
FAIRPoints	Seminar series	2022

EOSC Nordic - value and limitations of FAIR evaluators	Conference	2022
FAIR data award	Evaluator board	2022
Research Data Management for the Life Sciences	Course	2023
Lorentz FAIR 10 year anniversary	Conference	2024

Table 3

Aside from the several conferences and seminars that I attended, I also conducted some more clear cut ethnographic work: shadowing Fred, participating on the evaluator board of a FAIR data award, and participating in a Research Data Management course for the Life Sciences. The shadowing of Fred was a quite modest endeavor with four sessions done over the course of four months, but with sustained contact over email for longer to see if anything of interest came up. For the most part, it didn't, as her relationship with FAIR was quite minimal (in spite of it being a mandatory requirement of her funding)[21]. While this encounter didn't provide much of an answer as to what FAIR practices were, it did lead to a lot of productive questions as to why FAIR practices seemed so absent in scientific practice.

The participation on the evaluator board for the FAIR data award was a more contained venture with a clear beginning, end, and goal. It consisted of a series of emails back and forth, receiving a link to several datasets and the instructions for how to assess them, and a meeting between the members of the board where we discussed how to rank and score the different datasets. The research data management course for the life sciences was something that I was invited to by one of the data stewards that I had been interviewing. They were planning to conduct the course for early stage researchers to give an introduction to data management (including FAIR). This was done over the course of two days with approximately 30 students, from master's students to post-docs. During the course, I took the same role as the life scientists, discussing and exploring data management together with them. Since everyone was informed about my research project and reason for being there, I was also sometimes referred to to answer questions like "what is FAIR?" or the likes. After the course, my informant and I spent some time reflecting upon the course and discussing the choices made in the design of the course (for instance talking about why the presence of FAIR was so modest). Aside from the informative nature of the encounter, it also demonstrated clearly the value of developing a connection with your informants – a task more easily afforded by non-virtual space.

[21] We did, however, have a look at her data management plan together. Although this was quite interesting, I did, as mentioned earlier, end up not writing a chapter on this due to practical constraints

Chapter 1: A history of FAIRness

I want to start the first chapter of this thesis with a simple question: what is FAIR?

It is a question I have asked myself, a question I have asked others, and one others have asked me. What I found early on was that despite the simplicity of the question, there was no simple answer. While this obviously made it harder to produce any knowledge on this topic in the first place[22], it also made it frustratingly difficult to talk about my research with others – because what do we talk about when we talk about FAIR? It often remains ambiguous. But ambiguity does not mean that something has no meaning. Ambiguity means that the meaning of something is uncertain or obscured, that it affords multiple interpretations.

Among the multiple interpretations of FAIR we'll encounter throughout this thesis, there is one that is invoked perhaps more often, and certainly with more persistence, than others: namely the "true" meaning of FAIR, what it is "really" about, the "original" vision. While I myself don't subscribe to this singular view of FAIR[23] – and recommend that you don't either – there are people who *do,* and we will encounter them several times over the following pages. The FAIR that these people subscribe to and invoke is FAIR in its earlier days, before it became an EU policy and before it became a part of the daily professional lives of researchers across countries and disciplines – FAIR before it became as multiple as it is today. Thus, while the time when FAIR seemed more singular in nature belongs to the past, it often haunts the present.

Therefore, I will devote this chapter to the early history of FAIR, not to uncover or define its true meaning, but to understand what FAIR once was, and to make sense of *why* it was that way. With that aim in mind, I am not attempting to write a *definitive* history, but a *productive* history, a history that helps us understand FAIR from principles to policy to practice. This story is told to put FAIR in a context, to show that it didn't emerge from nowhere, but that it was weaved together by different strands that, even though they have ceded into the background, remain a piece of the FAIR cloth. Histories are written because we believe that they matter, and this history of FAIR is no exception.

While other histories of FAIR tend to begin at the FAIRPORT conference in Leiden in 2014, I have instead chosen it as the terminus of this one. I do this for two reasons: first, because the period after 2014 will receive its fair share of attention later in this thesis, and, second, because the developments that preceded and led up to the formation of FAIR have shown to

[22] a challenge I've addressed in the method section of this thesis
[23] as would be possible to guess from the liberal use of scare quotes

persistently matter for its continued history even after they faded into its backcloth. With that in mind, we'll start this history in an office in Switzerland, with the invention of the World Wide Web.

Weaving the web

It is a well rehearsed story, the one of how the World Wide Web was conceived. A British computer scientist, Tim Berners-Lee, was working out how to share information across the large and heterogeneous research organization he was working in, the European Organization for Nuclear Research, better known as CERN (Berners-Lee, 1990; Berners-Lee and Fiscetti, 1999; CERN, n.d.). The way he tells it, he came up with the idea of applying hypertext as a way of connecting and maneuvering the different stores of information at CERN, replacing the closed off tree-structured systems with an interconnected web (or "mesh" as he first called it) (Berners-Lee, 1990).

Over time, Tim Berners-Lee's system became taken up among researchers at CERN, then at other institutions, before it gradually became available to the general public through the means of web browsers such as Netscape and Internet Explorer in 1995 (Hopgood, 2001). Towards the end of the millennium, people had started becoming used to sending and receiving information through the World Wide Web. Still in its early days, and far from gaining the status of ubiquity it has today, the future of the web was still unclear. Tim Berners-Lee, though, had a clear vision of how the web should be developed. In his book *Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor* (Berners-Lee and Fischetti, 1999), Tim Berners-Lee laid out his vision:

> I have a dream for the Web… and it has two parts. In the first part, the Web becomes a much more powerful means for collaboration between people. [...] In the second part of the dream collaborations extend to computers. Machines become capable of analyzing all the data on the Web – the content links, and transactions between people and computers. A Semantic Web, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy, and our daily lives will be handled by machines talking to machines, leaving humans to provide the inspiration and intuition.
>
> The intelligent agents people have touted for ages will finally materialize. This machine understandable Web will come about through the implementation of a series of technical advances and social agreements that are now beginning. Once the two-part dream is reached, the Web will be a place where the whim of a human being and the reasoning of a machine coexist in an ideal, powerful mixture (Berners-Lee and Fischetti, 1999, p. 157-158).

In his opinion, the web as it was (and largely still is[24]), had not reached its full potential. Merely connecting computers, allowing their users (human people) to communicate through sending documents back and forth was only the start. In the future, he argued, the computers would be able to communicate directly with each other, exchanging data, on what Tim Berners-Lee called "the Semantic Web" (Berners-Lee and Fischetti, 1999).

The main barrier between the conventional web, as *we* know it, and the Semantic Web of the future, was, according to Tim Berners-Lee, that people were not publishing their data in a machine-readable format and linking it to other existing data (Berners-Lee, 2006). As he put it: "The Semantic Web isn't just about putting data on the web. It is about making links". In order to guide people to contribute to this "web of data" (as opposed to the "web of hypertext"), Tim Berners-Lee developed four rules:

1. Use URIs[25] as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information.
4. Include links to other URIs. so that they can discover more things.

(Berners-Lee, 2006).

In later iterations, he would add an emphasis on the importance of open sharing, epitomized in a 2009 TED-talk titled "The Next Web", where he criticized government bodies and enterprises for "hugging" their data (as in not letting it go) (Berners-Lee, 2009). The way he expressed it, he – as a taxpayer – had already paid for the data, so it should be made available to him and other taxpayers freely and openly. With this, he encouraged anyone responsible for a dataset to "give us the unadulterated data", culminating in an attempt to make the audience chant "raw data now!" (Berners-Lee, 2009, 10m43s).

A year later, "in order to encourage people – especially government data owners – along the road to good linked data", Tim Berners-Lee developed a "star ranking system", published as an update to his blogpost with the four rules of linked data (Berners-Lee, 2010a). In order to achieve the different levels of linked *open* data stardom, Tim Berners-Lee suggested data owners should make their data:

[24] To be fair, the current day web does have applications of semantic web technology, like the open online knowledge graph Wikidata, the quick results sidebar in your Google search engine, or online advertising technologies that track user behavior on the web.
[25] Short for "Uniform Resource Identifier", a kind of social security number for things on the internet. An ORCID iD is one such example, pointing to a person. A DOI (digital object identifier) is another, pointing to more or less whatever. URI's are different than URL's as one is an identifier, and the other is a locator informing where a resource can be found. In other words, if a resource is moved, it retains its URI, but not its URL.

★ Available on the web (whatever format), but with an open licence

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV[26] instead of excel)

★★★★ All the above plus, Use open standards from W3C[27] (RDF and SPARQL[28]) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

(Berners-Lee, 2010a).

He later added that "Linked Data does not of course in general have to be open, [but…] under the star scheme, you get one (big!) star[29] if the information has been made public at all, even if it is a photo of a scan of a fax of a table – if it has an open licence" (Berners-Lee, 2010b).

Tim Berners-Lee suggested that if everyone would follow his recommendations, publishing 5-star linked data, the web would be well on its way towards becoming semantic, fulfilling his vision. But while he acknowledged the importance of the high-energy physics environment at CERN for developing the first iteration of the Web, he predicted that the future of the web would be driven forth by the life sciences (Bostrom, 2005). This was not without reason, because in parallel with Tim Berners-Lee's work promoting the next web, similar developments were underway in the life sciences – in fact, it had been for a couple decades already.

Linking genomic data

A couple decades earlier, in the mid-1980's, what is now known as one of the most prestigious big science projects in modern history, the Human Genome Project, took its first steps. At a conference in Alta, Utah, in 1984, a group of 19 biologists came together, not to initiate the significant undertaking of sequencing the entire human genome, but to ask whether it would be possible to develop methods to screen for mutations in people affected

[26] CSV is a file format that stands for "comma separated values", and is a basic and non proprietary format consisting of strings of values, separated by commas.
[27] W3C, The World Wide Web Consortium, an institution that is dedicated to developing and maintaining web technologies and standards, such as HTTP(S), HTML, RDF and more. They have no formal authority, but retain a significant influence on web design and development.
[28] RDF and SPARQL are, in a simplified sense, standards for data formatting (RDF) and data retrieval (SPARQL) that enable semantic encoding and use of data
[29] Naturally, much like the W3C, Tim Berners-Lee didn't have any formal authority and the star system wouldn't get you any rewards other than bragging rights that your data was 5-star for instance.

by the Hiroshima and Nagasaki bombings of 1945 (Cook-Deegan, 1989). In order for that to be possible, they argued, they would need to undertake a project so large, complex and expensive, that it was entirely unlikely that they would even get to even begin (Cook-Deegan, 1989).

In spite of the initial pessimism, the idea of creating a reference human genome stuck around. Thus, other meetings were organized throughout the second half of the 1980's (NIH, 2022). In these meetings, the aim of the initiative broadened beyond just screening for mutations in victims of nuclear war, to that of screening for genomic predictors of hereditary diseases (Sinsheimer, 1984). Recent developments in both computing and genome sequencing had instilled the researchers with more confidence that future technological progress would shift the idea of sequencing the human genome from the realm of absurdity to that of distant possibility. But with this new optimism they still had to answer the question: would the still significant investment of resources be worth it? (Sinsheimer, 1984; personal communication, November 19, 1984).

A series of meetings in Santa Fe, New Mexico seeked to address both the feasibility and value of such an undertaking. After all the discussions, the participants concluded that it *could* be done, but that there were still some issues to be solved – among them, the problem of computational methods (Sinsheimer, 1984; 1989; Genome Sequencing Workshop, 1986; DeLisi, 2008; NIH, 2022). Because even if they managed to produce large quantities of genome sequence data, there was no guarantee that anyone would be able to make use of it.

Without computational methods and resources, there was no feasible way to analyze the billions of nucleotides in the human genome. And while tools developed to analyze genomic data existed, people worried that "the computational methods available [...] at the time would yield relatively little information" (DeLisi, 2008). If they were to sequence the human genome, they therefore had to put more effort into developing new computational methods of analysis for life sciences. In this sense, the future of genomic research was put in the hands of bioinformatics.

As time progressed, the project, which only later became known as the Human Genome Project, started publishing sequence data and making it openly accessible, first on discs received in the mail, and later through the web (NCBI, 1998). It went from being an unlikely idea to a published draft human genome in 2000 (NIH, 2000) and a "finished human genome sequence" in 2004 (NIH, 2004). This rapidly increasing availability of genomic data gave researchers that were unable to produce their own data – but had the resources and knowhow to analyze it computationally – a sudden influx of material to work with. What had been a niche interdisciplinary field since the 1980's, found itself in a "deluge of data" (Williams, 1995) fueling what has been described as "the bioinformatics goldrush" (Howard, 2000).

Turning data into knowledge

With the increasing availability of genomic sequencing data, the envisioned potential of bioinformatics became even larger. But the field quickly encountered challenges turning this vision into reality. You'd be forgiven for finding it counterintuitive, but while the origin of bioinformatics was a "deluge" of data, this deluge was also cast as their biggest problem (The Economist, 2010). To make matters more confusing, one suggested fix for this problem was to publish even more data. As one leader article in the Economist from 2010 put it: "The best way to deal with these drawbacks of the data deluge is, paradoxically, to make more data available in the right way" (The Economist, 2010).

To make sense out of this conundrum, we need to get some clarity when it comes to the term "data". Because while most people have a somewhat intuitive grasp on what data is/are, the term remains fuzzy, unstable and contested (RIN and British Library, 2009; Leonelli, 2015; Offenhuber, 2017). "Data" does not defer to a stable ontological or epistemological category, but as a term is used to connote something of value in the process of knowledge production (RIN and British Library, 2009; Leonelli, 2015). The value of data, though, is not given or inherent, but relationally contingent (Borgman, 2017; Leonelli, 2015). For a bioinformatician studying genes, an instruction manual for a DNA sequencer might not be considered data, while for an STS scholar studying scientific practice it might. Likewise, humanities scholars often prefer to talk about their "materials" rather than their "data", even though they generally play the same role as elements of knowledge production[30]. This means that when people say "data" they might not be talking about the same thing (Borgman, 2017; Offenhuber, 2017).

In the case of the "deluge" in bioinformatics, the data they were talking about was mainly the information produced by automatic genome sequencers – strings of G's, and T's, and C's and A's (Williams, 1995; NIH, 2012). This string of letters did not become valuable for knowledge production on its own though. A researcher would need to know how it was produced, what it was supposed to point to, and whether they could trust the quality of the data, before they could use it for their research or anything else (NHGRI, 1996; Green, 1997; Green and Koonin, 1998).

[30] Humanities scholars' unwillingness to use the notion of data comes up again in chapter 4

So the challenges of bioinformatics, as they saw it themselves, was twofold: 1) the field of biology was producing more genomic sequences than it was collectively able to make use of, and 2) the sequences that were produced were not readily made use of beyond the site of their production (Bourne, 2005; Mons, 2005; Attwood, 2009; Bairoch, 2009). In other words, when the NIH or the European Bioinformatics Institute were talking about the deluge of data, they were referencing the increased quantity of *information* produced by more efficient sequencers – information they were struggling to turn into usable *data* (Williams, 1995; NIH, 2012). This perceived underutilization of information manifested as an anxiety that “data” was “lost”, “buried” or “hidden” (Attwood, 2009; Bairoch, 2009; Pettifer et al., 2011). The sheer volume of bits and bytes and strings of GTCA was too large for anyone to meaningfully be able to keep track of, and much less to make use of (Attwood et al, 2009; Sansom, 2010). Understood differently, what they were suffering from was therefore not an abundance of “data”, but an abundance of information that they couldn't make useful outside of its original context or situation.

While “immobile” data – meaning information that is data only within its original context – is nothing new or out of the ordinary (see Edwards et al., 2016; Leonelli and Tempini, 2020), the field of bioinformatics had created and nurtured an expectation that their data could, should and would be able to travel. As the mass of information produced by their new sequencers failed to live up to this expectation, this spurred a series of reflections on the nature of information, data, knowledge, wisdom, and how to turn one into the other. One early example came in the publication *Calling International Rescue: knowledge lost in literature and data landslide!:*

> Merely increasing the amounts of information we collect does not in itself bestow an increase in knowledge. For information to be usable, it must be stored and organized in ways that allow us to access it, to analyse it, to annotate it and to relate it to other information; only then can we begin to understand what it means; only with the acquisition of meaning do we acquire knowledge. The real problem is that we have failed to store and organize much of the rapidly accumulating information (whether in databases or documents) in rigorous, principled ways, so that finding what we want and understanding what's already known become exhausting, frustrating, stressful and increasingly costly experiences (Attwood et al., 2009).

In formulating their suggested solutions, bioinformaticians increasingly came to realize the necessity of distinguishing between different forms of what had previously been referred to generically as data. But their distinctions and descriptions seldom mapped perfectly onto each other. Still, they all dealt with variations of the same collection of terms – and the hierarchical relations between them – data as a basic form of *information*, that could be turned into *knowledge*, and ultimately (but not always) transform into *wisdom*[31]. The implications of this view was that lower grades or qualities of information could always be transformed into a higher form through some kind of applied effort. The challenge was figuring out how. One paper that defined data as “a collection of facts” suggested the application of interpretation as way of making the transformation into knowledge:

> For the purpose of this review, it is important to make a clear distinction between data and knowledge. The concept of data came into prominence relatively recently, mainly due to the widespread use of the information and communication technologies (ICT) and the advent of modern empirical technologies that outpour huge amounts of data. Data should not be confused with knowledge–the former is just a collection of facts that require interpretation in order to be converted into knowledge. Thus, knowledge is data plus an interpretation of its meaning (Antezana et al., 2010).

Another speculated in the development of certain tools as a way of *extracting* information and knowledge from data, and *producing* knowledge from data and information:

> Arguably, it is not the actual amount of data that is a problem as much as the difficulty of trawling those data – and the associated scientific literature – for the information and knowledge that we need. [...] Over the last few years, some publishers and technologists have been working with researchers to develop new tools to help make this process more efficient [...] All that remains is to see whether these enhancements in efficiency and ease of use, in the long term, aid the production of knowledge from data and information. Let alone wisdom… (Sansom, 2010).

No matter what solution they suggested, they generally agreed that the problem was that even though the amount of information was larger than ever, it was available in a qualitative state that restricted its usefulness. Thus, they aimed to find ways to transform the information in the field into usable data so they could produce useful knowledge.

Semantically enriching publications

The most immediately identified barrier towards the full exploitation of bioinformation was the issue that most researchers didn’t even publish their data in the native format. Instead, they would publish articles containing traces or descriptions of the data, creating the sentiment that researchers were “hid[ing it] in a often badly written text” (Bairoch, 2009). Researchers wanting to use this data would lament about how much work it would be to extract it, and

[31] This hierarchy of relations is often referred to as the DIKW pyramid, where the acronym stands for Data, Information, Knowledge and Wisdom (see for instance Rowley, (2016) for an introduction and overlook of the concept)

how it would make the data practically inaccessible (Mons, 2005; Attwood, 2009; Pettifer et al., 2011; Mons et al., 2011). The way they saw it, the academic publishing system was not keeping up with the developments of contemporary research practices.

In 2005, Philip Bourne – the Editor-in-Chief of PLoS Computational Biology and Co Director of the Protein Data Bank – floated the idea that the distinction between databases and journal articles not only would, but *should* collapse (Bourne, 2005). Pushing for the boundary between data-centric and knowledge-centric publishing to be dissolved, he suggested the solution to bioinformatics' problems lay in a future beyond the contemporary publishing system (Bourne, 2005). Because as data became more abundant, data also took a more central place in life science knowledge production (Borgman, 2008), ultimately making the discipline more reliant on what they perceived to be a purer form of information, as opposed to the cluttered arguments, narratives and subjective interpretations found in journal articles (Bourne, 2005; Mons, 2005; Mons et al., 2009; 2011; Attwood et al., 2009; Bairoch, 2009; Shotton, 2009; Pettifer et al.; 2011).

The question was then how to make data more accessible and exploitable for those who wished to use it. In the search for an answer, a number of scholars looked toward semantic web technologies (Clark et al., 2009; Attwood et al., 2009; Antezana et al., 2010; Sansom, 2010; Pettifer et al., 2011). The expectations towards this technology is illustrated by Shotton (2009):

> Recent developments in Web technology can be used for semantic enhancement of scholarly journal articles, by aiding publication of data and metadata and providing 'lively' interactive access to content. Such semantic enhancements are already being undertaken by leading STM publishers[32], and automated text processing will help these enhancements become affordable and routine. Publisher, editor, and author all have primary roles in that process; an incremental approach is needed. Publication of data and metadata to the Web make possible added-value 'ecosystem services'; semantic publishing will bring substantial benefits to scholarly communication.

The expected potential of semantic web technology served as an inspiration for a variety of different approaches. Shotton (2009), for example, suggested "semantic enhancement" of research papers, aiming to embed data directly into papers. As a demonstration of the feasibility and value of such an enhancement, Shotton et al. (2009) later made the effort to "semantically enrich" an existing article on disease management that could be read directly in a normal browser[33]. Another initiative, *The Semantic Biochemical Journal* and *Utopia*

[32] Scientific, Technical and Medical Publishers
[33] This article (Reis et al., 2008) can still be accessed, but some of the software needed to fully exploit the semantic enrichment has later been outdated

Documents (Attwood et al., 2009; 2010; Pettifer et al., 2011), tried out a similar approach, but instead of running the semantic enrichment of papers in the browser, they created a software that would “automatically” read the paper and annotate it for the reader, linking words to biological concepts, and making tables, graphs and other visualizations of data interactive directly in the reading software (Attwood et al., 2009; 2010).

Today, neither initiative remains active. Traces of Shotton’s approach can arguably be found in the way journals semantically enrich their papers today when they automatically highlight certain concepts or words linking them to, what is essentially, a keyword webpage. The Semantic Biochemical Journal and Utopia Documents are both gone with little trace, only available through the Internet Archive.

Publication post paper

But not everyone was seeking to dissolve the boundary between article and data publishing. A group of researchers under the moniker *Concept Web Alliance* (CWA, 2009), went against Bourne’s encouragement (2005), and wanted to keep the two separate, but shifting the priorities, making articles secondary to data (Mons and Velterop, 2009). Considering the value of data “lost” in article publication, they suggested that maybe the research paper wasn’t that important altogether anyways (Mons and Velterop, 2009). Instead of publishing their knowledge in article format, the Concept Web Alliance suggested, researchers should focus on publishing their data in a machine readable format rather than semantically enriching their human readable papers (Mons and Velterop, 2009).

In the paper presenting the Concept Web Alliance, Mons and Velterop (2009) contextualized their initiative as a way of exploiting the potential and opportunities brought forward by developments in information technology, and, more specifically, semantic web technologies:

> The dawn of the Semantic Web Era has brought a first wave of reduction of ambiguity in the Web structure as terms and other tokens are increasingly mapped to shared identifiers for the concepts they denote. Initiatives like Linked Open Data have gone a long way to connect Web resources at the ‘concept’ level, rather than at the term level, as done by Google and other word based systems. [...] Classical publication on paper, even when converted to electronic formats, has not even begun to seriously exploit the possibilities that Web Publishing, even in its current, still early stage of development, has opened up. Yet most available so-called electronic publications are mere analogues of the paper versions, and often only in PDF. Terms are rarely, if ever, mapped to unambiguous concepts and, together with the habitual repetition of factual statements in each consecutive paper for the sole purpose of human readability, analysing scientific information with computers can currently

> not be considered in any way close to its potential. [...] Computers can deal extremely efficiently with structured data. Unfortunately, people seem to dislike structured data entry, as evidenced by their reluctance to do it, and that is where the central problem of classical publishing arguably lies.

Their suggestion was to circumvent the academic paper in its entirety and rather publish what they called “nanopublications” (Mons and Velterop, 2009). While this was related to the idea of conventional data publication, it went one step further, drawing heavily on Tim Berners-Lee's concept of Linked Data. Instead of publishing data “as is” (or “raw” for that matter), they proposed that researchers publish the knowledge derived from data as “statements” (or nanopublications) (Mons and Velterop, 2009).

These nanopublications were imagined as something in between data and a traditional publication. They urged researchers to not publish a dataset showing a correlation between mosquito bites and malaria infection, not publish an article explaining how one causes the other, but rather to publish the insight as a *semantic statement*, often called “triples”. Triples are essentially statements structured in the format of object - relationship - object (Mons and Velterop, 2009)[34]. A simplified example could be “mosquito bites - transmit - malaria”. With this simple grammar, statements could be linked to others to build a network of statements allowing a computer program to combine entered statements into new ones. As an example, if we also had the triple “malaria - causes - death”, a computer could deduce “mosquito bites - cause - death”.

According to Mons and Velterop (2009), publishing only the essential statements of knowledge as nanopublications was the future of scholarly communication, as it would enable machines to automatically discover new knowledge faster than any human could. To reach this future though, they first had to overcome a problem related to the limits of machine reasoning.

The bioinformaticians described this problem as stemming from computers’ inability to resolve the ambiguity of data (Mons and Velterop, 2009; Pettifer et al., 2011). To illustrate the ambiguity issue, Pettifer et al. (2011) presented the phrase “fruit flies like a banana”. For a human, they argued, it is relatively simple to understand that the phrase refers to an insect enjoying a banana, not to a statement on the aerodynamics of fruit. For a computer, on the other hand, they claimed, the statement remains ambiguous because both interpretations are syntactically and logically sound, and the computer doesn’t have the ability to resolve it using context understanding or common sense. To make computers able to reason and understand, they argued, humans therefore would have to reduce the ambiguity of

[34] or “subject, predicate, object” by people in the discipline

statements and terms and make it clear what concepts they were denoting (Mons and Velterop, 2009; Pettifer et al., 2011). Pettifer et al. (2011) made the case that this was philosophical problem turned practical by “the advent of computers”:

> From the shadows on Plato’s cave wall, through Kant’s distinction between noumenon (‘the thing in itself’) and phenomenon (‘the thing as it appears to me’), to Magritte’s non-pipe and beyond, the subtle relationships between objects, concepts, and their representations have always been a topic of much philosophical and artistic interest. With the advent of computers – whose primary purpose is to manipulate concepts and their representations en masse – however, such questions take on an altogether more practical significance.

The way the Concept Web Alliance addressed this question, and aimed to solve the problem, was by formalizing the relationships between terms and concepts, making sure that one term denoted only one concept, through a document called a controlled vocabulary or ontology (CWA, 2009; Mons and Velterop, 2009). The idea is that if you can identify all relevant concepts and give them unique names (or identifiers) within an ontology, anyone formatting their data according to that ontology would be able to make their data interoperable and machine readable (CWA, 2009; Mons and Velterop, 2009; W3C, 2010; Pettifer et al., 2011).

A common method of making ontologies is to organize a delimited group of people to come together and map terms to concepts in a centralized fashion (Staab and Studer, 2009). But because of the scope of their envisioned semantic web, the people of the Concept Web Alliance suggested a distributed effort where individual researchers would contribute to building the ontology piece by piece (Mons and Velterop, 2009). By mapping the relationships between individual terms and concepts through nanopublication, they envisioned their concept web would grow incrementally, “anticipat[ing] more than 200 million terms in English alone” (Mons and Velterop, 2009).

Despite Mons and Velterop’s (2009) claim that “the efforts [to realize the concept web were] so small, and the benefits so great”, the idea struggled to take off. Researchers generally kept publishing papers in traditional journals, and nanopublications remained a niche. But the idea of an era of post-paper-publishing, sometimes called “eScience”, had caught on. A community of scholars, including personalities from earlier initiatives like Phil Bourne, David Shotton, Steve Pettifer, Theresa Attwood and Barend Mons, decided that coordinated action was necessary, and organized two conferences under the title “Beyond the PDF”, first in California in 2011, and two years later in Amsterdam (Bourne, 2010; Beyond the PDF, 2013; OpenPHACTS, 2013). With the persistent interest in the initiative, they decided to institutionalize the community under the name FORCE11, standing for “The Future of

Research Communications and e-Scholarship", and the number 11 referring to the year of the first conference (Groth, 2013).

Making principles for data publishing

Although many had deemed the combination of semantic web technology and biological data to be promising, it became clear that bioinformatics' efforts to semantically enrich papers and build ontologies would not bring about a semantic web for the sciences. In a 2010 paper, a group of computer and information scientists in England (Bechhofer et al., 2010) argued that "linked data is not enough for scientists", and that data needed to keep the same level of quality as people expected from other research publications in terms of "provenance, quality, credit, attribution, [and] methods" (Bechhofer et al., 2010). This approach to data publishing took its share of inspiration from the library sciences, that had been discussing the quality of data publishing for years already – although being more concerned with long term storage and retrievability of digital records than the interoperation of digital data for knowledge production purposes (see CCSDS, 2002; Research Libraries Group, 2002; Horwood et al., 2004; Nestor, 2006; RIN, 2008; Hunter, 2008; Data Seal of Approval, n.d.; NAS et al., 2009).

Perhaps the most central inspiration from the library sciences brought into the discussion on data publication for semantic knowledge production was the importance of contextual information for the published data, or so-called "metadata" (Bechhofer et al., 2010). If data were to be usable by others, the argument went, it had to be accompanied by additional information describing the data. Thus, the principal published component could not be the data itself, but a unit containing *and* describing the data (Bechhofer et al., 2010), in the library sciences, this was typically referred to as a "digital object"[35] (see Hunter, 2008).

In a 2009 blog post, one of the co-authors of the Bechhofer et al. (2010) paper, computer scientist David De Roure launched an idea to combine the digital object approach of the library sciences with the post-paper publication movement, stating that:

> I believe that the academic paper is now obsolescent as the fundamental sharable description of piece of research. In the future we will be sharing some other form of scholarly artefact, something which is digital and designed for reuse and to drop easily into the tooling of

[35] People familiar with academic publishing might recognize the "digital object" as the two first constituent letters of the acronym "DOI" (which stands for digital object identifier). The DOI system was originally an initiative by trade associations in the publishing industry beginning in 1997 (DOI, 2023). It built upon previous work on digital libraries commissioned by DARPA in the early 1990's, where Internet veteran Bob Kahn developed the Digital Object Architecture (Kahn and Wilensky, 2006; Internet Society, 2016). The "digital object" concept thus has clear origins in both library sciences and informatics.

> e-Research, and better suited to the emerging practices of data-centric researchers. These could be called Knowledge Objects or Publication Objects or whatever: I shall refer to them as Research Objects, because they capture research (De Roure, 2009c).

Throughout his work on digital knowledge sharing platform myExperiment, David had developed a list of characteristics a Research Object should have, all starting with the letter "R" (De Roure, 2009a; b; c). In the first iteration of David's list of R's, he argued that any Research Object, should it be of any value, had to be Repeatable, Replayable, Reproducible, Repurposable and Robust (De Roure, 2009a). Later iterations replaced Robust with Reliable and added Reusable, Referencable, Re-interpretable, Respectful, Respectable, Retrievable, Refreshable and Recoverable – culminating in a total of 13 principles starting with, not only R, but "Re"[36] (De Roure, 2009bc; 2010).

David's obsession with R's was mostly a personal pet project, published primarily on personal pages and in presentations, but he still managed to catch the attention of other colleagues, which culminated the aforementioned paper, titled *Why linked data is not enough for scientists* (Bechhofer et al., 2010; 2013). Aside from articulating the critique of previous initiatives, the paper also refined David's list of R's to a more reasonable 7: Reusable, Repurposable, Repeatable, Reproducible Replayable, Referenceable, and Respectful.

Meanwhile, the role of research data sharing had gained attention in the western policy sphere, with the US and the EU in particular taking the lead. Reports from both sides of the Atlantic had highlighted that the increasing volume of data in the sciences would necessitate some form of infrastructure development to make use of it (EC, 2010; NIH, 2012). The scientific community was soon to respond to these calls. One of the more notable contributions was the Royal Society's report, *Science as an open enterprise: open data for open science* (Royal Society, 2012). While mainly relying on ideas from the Open Science community[37], it did carry with it central elements from the post-paper publishing discourse, such as Semantic Web technology, the ordered hierarchy of knowledge, and the concept of Research Objects. Hence, the report urged to go further than just Open, and suggested researchers and institutions should pursue "intelligently open data", further refining David's R's into four principles: that data should be accessible, intelligible, assessable and usable (Royal Society, 2012, pp. 14-15).

A different group of data scientists, 75 in total, in parallel organized a meeting at the *International Conference on Research Infrastructures in Copenhagen* in 2012 to address the

---

[36] Hence the replacement of RObust. David later pointed out the rediculousness of the situation, in a blogpost titled "More Rs than Pirates" (De Roure, 2012).
[37] in part the Bermuda Principles that emerged from the Human Genome Project (NIH; 2022)

same call (DAITF, 2012a; b). The group, which then went under the name *Data Access and Interoperability Taskforce* (DAITF) had a round of presentations and discussion over the span of two days. In one of the presentations, computer scientist Larry Lannom put forth the idea that there was a barrier between the prospective users of data, and the data that they wanted to use (DAITF, 2012a). To overcome this barrier, he suggested, there was a need for “enabling technologies” to fulfill some fundamental functions in the data science workflow, namely: discovery, access, interpretation and reuse (DAIR)[38] (Ahalt et al., 2014). In order to achieve these functions, Larry envisioned a web of data as analogous to the conventional web, but instead of connecting web pages it would connect Digital Objects (Ahalt et al., 2014).

The infrastructure he imagined based itself on a common representation of the internet infrastructure we use today, the TCP/IP model. This model consists of four layers: the application layer, the transport layer, the internet layer, and the link layer. Other network models that competed with TCP/IP in the 1970’s and 1980’s would use different numbers of layers. IBM’s SNA network used only one, while TCP/IP’s biggest competitor, X.25, used three, but they all serve the same general purpose, to make sure that information packets get sent from one end of the network to the other. This is, simply put, why clicking a link in your browser sends a request for information through your router, to a server which sends the requested information back to you – instead of, for instance, requesting the information directly (peer to peer networking).

The success of TCP/IP over its competitors is often, by network- and computer scientists, attributed to the way it managed to coordinate the movement of information through the layers by forcing all the information to go through a single common layer, or a “spanning layer” which represented the Internet Protocol (IP) (Beck, 2019). The reason it was called the spanning layer, was because it was imagined to “span” across the entire width of the network, so to speak (see figure 1).

Layer 4 - Column A	Layer 4 - Column B	Layer 4 - Column C
Layer 3 - Column A	Layer 3 - Column B	Layer 3 - Column C
<- (Spanning) Layer 2 ->		
Layer 1 - Column A	Layer 1 - Column B	Layer 1 - Column C

Figure 1

[38] Some later references to this presentation have exchanged discovery with search, making the acronym “SAIR” (see Tochtermann, 2018 and RDA, 2018).

By spanning across the entire layer, IP allowed the movement of information across the vertical columns of the model without having to tailor special applications or protocols. At some point this spanning layer became imagined not as a wide layer going from one end to the other, but as a pinched together passage point, thus, some people replaced the visual representation of the stack model (figure 1) with that of an hourglass model (see figure 2) (Guenter, 1998).

This particular way of envisioning the structure of the web caught on and became a taken for granted truism propagated by people within the web-technology community. The hourglass model then took on the character of, not only a description of the web, but also as a prescriptive design principle for robust and adaptable networks, as seen in Larry Lannom's design (figure 2) (see also Beck, 2019).

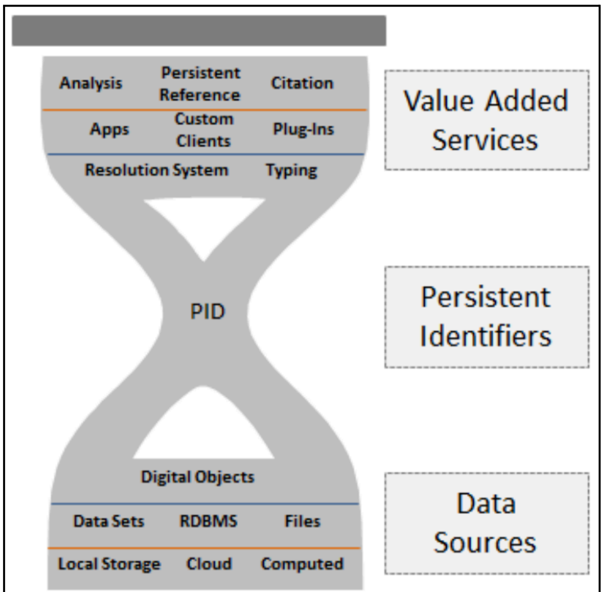

Figure 2 (Ahalt et al., 2014)

Larry sketched out his suggested structure as seen in figure 2, with PID, or Persistent Identifiers, serving as the narrow waist, or spanning layer, in the suggested network model. Placing PID at the center of the model, Larry's idea aligned itself with Tim Berners-Lee's emphasis on stable and unique identifiers as a central element. Combining a system of stable and unique identifiers with the Digital Object approach was imagined to solve the challenge of distinguishing between terms and concepts, and make it easier to link them together using the web infrastructure (Ahalt et al., 2014).

Larry's idea received considerable attention and recognition as an important concept within the task force. The ideas also propagated into other contexts as well, like the *Nairobi Data Sharing Principles* (CODATA et al., 2014), the *Joint Declaration on Data Citation Principles* (FORCE11, 2013) and the *G8 Principles for an Open Data Infrastructure* (Blatecky et al., 2013). The DAIR principles and the hourglass model with PID as the narrow waist, was ultimately described as instrumental in the development of the Research Data Alliance (RDA) in 2013 – one of the arguably most influential organizations when it comes to developments in research data sharing (Wittenburg, 2018; 2019; RDA, 2018; Tochtermann, 2019)[39]. In spite of the influence that Larry Lannom's ideas would gain, the DAIR principles

[39] and an organization that will reappear several times throughout this thesis

were ultimately overshadowed by a similar sounding, but ultimately different initiative – and the focus of this thesis – the FAIR principles.

Making research FAIR

While there is no formal, established or canonized history of the FAIR principles and how they came to be, there are some common stories, and they all place one man at the center of it: Barend Mons[40]. And the way Barend Mons tells the story himself, it all starts with his research into malaria on the African continent in the 1990's. One of the problems with malaria, as Barend Mons saw it (Mons, 1997; Mons et al., 1998) was that there was a large continent of many countries with one and the same problem – an endemic disease costing both lives and money. Meanwhile, the efforts to solve this problem remained fragmented. His suggested solution was to unify knowledge and efforts, partly through institutional efforts, but – more importantly to the story of FAIR – through translation of words from different languages into universal concepts, or "real-life entities" as he called it (Mons, 2002). In his own words:

> I realised that paludisme was the same concept as malaria, just another token to refer to it in human language. So, I became interested in thesauri, and I started to make a plea to refer in science to concepts with a unique identifier. And that's the end of my biology career almost because I was sucked into graph technology and later on even to conceptual modelling. […] Matching across languages is how it started (Mons, 2020b).

The turn of the millennium marked a shift for Barend Mons, who up until then had been publishing almost exclusively on malaria[41]. In 2000 he made his first publication within the field of informatics with the paper *Facilitating networks of information* (van Mulligen et al., 2000). Later, after becoming gripped by the idea of semantic web technology, he became one of the chief initiators of the Concept Web Alliance, making his appearance in this chapter (Mons and Velterop, 2009; CWA, 2009). While he remained largely in the background of this story in the years after, he spent that time developing his ideas of the Concept Web and Nanopublications with EU's Innovative Medicines Initiative (IMI)[42] and the pharmacological industry in 2011, resulting in the project "Open Pharmacological Concepts Triple Store" (or Open PHACTS, for short) (Mons, 2013). The project was promoted as an astounding success and proof of concept for Mons' semantic technology (Williams et al., 2012). As it was put in one of the publications from the project: "Open PHACTS clearly has

[40] One of my informants even referred to him as "the great brain that conceived of FAIR!"
[41] According to Barend Mons' google scholar profile
[42] Together with Stephen Pettifer from the Utopia Documents initiative

the potential to initiate a paradigm shift in the pharmacological space and far beyond" (Williams et al., 2012).

Encouraged by the success of Open PHACTS and finding himself professor of biosemantics at Leiden University in 2013, Barend Mons decided to invite some 30 colleagues to a conference at the Lorentz Center in Leiden to discuss the possibility of designing an infrastructure for the sharing and reuse of research data, an initiative they called "FAIRPORTS" (FAIRPORT, 2014). These FAIRPORTS echoed earlier initiatives in that they were imagined as environments prioritizing publishing of data as digital objects designed according to an hourglass model, with the final aim of establishing "a fully interoperable internet of data" through the application of semantic web technologies (FAIRPORT, 2014). What put the FAIRPORT initiative apart from its predecessors was the explicit intention of designing this infrastructure first and foremost for machines, making human reuse of data a secondary aim (FAIRPORT, 2014; Stehouwer and Wittenburg, 2015). Thus, the FAIRPORT initiative was explicitly designed to develop a web like the one Tim Berners-Lee had envisioned: "a place where the whim of a human being and the reasoning of a machine coexist in an ideal, powerful mixture" (Berners-Lee and Fischetti, 1999).

Although the groundwork of this initiative was laid, it would only become "FAIR" after a discussion between Barend Mons and one of the founding members of the RDA, and colleague of Larry Lannom, Peter Wittenburg. Barend Mons had realized the conceptual and nominal resemblance between DAIR and FAIRPORT (Ahalt et al., 2014), and soon after decided to scratch the "PORT" from his initiative. Some months later, the FAIR principles – now standing for findable, accessible, interoperable and reusable[43] – were published and promoted through the FORCE11 community (FORCE11, 2014) and presented at the 4th RDA plenary (Mons, 2014).

What next?

For the people who now supported the FAIR principles, the future was laid out before them. Drawing on futures past – from the Human Genome Project to Tim Berners-Lee's Semantic Web, from the eScience era and post-paper publishing, to object oriented hourglasses – the proponents of FAIR were committed to realizing their future and making research FAIR (FAIRPORT, 2014). But this is where the story ends for this chapter.

[43] as opposed to "fair airport". This acronym replacing the "discovery" in Larry Lannom's idea with its equivalent "findability", and the "interpretation" with "interoperability"

That being said, the story of FAIR will be continued throughout this thesis as we follow FAIR from principles to policy to practice. Throughout the following chapters we will look at several instances where the meaning of FAIR was transformed, challenged and contested. As FAIR started its journey in 2014, the aim behind it seemed relatively clear, but, as we will see, as time went on and it was taken up in new places and by new people, the nature and purpose of FAIR became less and less evident. In that regard, this chapter serves as a reminder of where FAIR came from and what its purpose once was.

I hope this will also reduce some of the ambiguity around the term “FAIR” and help make sense of some of the tensions and controversies we’ll encounter throughout this thesis – maybe even beyond it. While it may be difficult today to say what FAIR is, this chapter makes it clear that it wasn’t always like that. As we move on through the next chapters, I also encourage you to consider and keep in the back of your mind, how things would be if we instead had been building FAIRPORTS, making DAIR infrastructures, publishing Utopia Documents, or merely linking raw data. It could have been otherwise.

Chapter 2: Making EU FAIR in times of crisis

In November 2020, as the world was in the depths of the COVID pandemic and economic growth slowed down, researchers and policymakers were still organizing conferences and meeting at symposiums – although only from behind their computers. In spite of the financial downturn, the millions of deaths, and the inconvenience of wearing a face mask once in a while, at least one person was finding a silver lining in this global crisis: the researcher and ex-EU-policymaker Jean-Claude Burgelman.

In a pre-recorded message for the first plenary of the CODATA International FAIR Convergence Symposium, Jean-Claude Burgelman didn't talk about how bad things were, but how they could have been worse, and, importantly, how they would get better. He talked about what he saw as the main reason we were able to develop vaccines so fast: FAIR data. He talked about the power and potential of data science to change the world. But, he warned "we still need to push a lot to make sure this momentum is not lost [...] and finally – but that is an ex-policymaker speaking to you – *never waste a good crisis*" (Burgelman, 2020, 25:21-26:05).

Maybe some would find Jean-Claude Burgelman's suggestion that COVID would be something positive offensive[44], but then again, it is with politics as it is with sausages: if you want to keep your respect for them it's best not to know how they're made[45]. But respect for policy has never been a concern for Science and Technology Studies, a discipline more interested in critical scholarship and deconstruction. So in this chapter I will delve into the making of policy, how FAIR became EU research policy.

The reason I do this is because, in the context of the FAIR principles, policymaking has been of great importance. Without the support of the European Union – through making the principles a central component of their European Open Science Cloud – FAIR would likely not have held the position it does today. And while there are many explanations for why FAIR became taken up by the policymakers in the European Union (see *Data Intelligence* special issue on Open Science (Wittenburg and Strawn, 2021b), empirical interrogations of how this happened are largely lacking.

[44]At least that was the case when Obama's chief of staff Rahm Emmanuel was the (supposedly) first to say on record not to waste a crisis in 2008 in the aftermath of the financial crash (see Colton, 2020)
[45] This aphorism is often credited to Otto von Bismarck, but is likely apocryphal. For a investigation on the source, see Quote Investigator (2010)

What I aim to do with this chapter is then to contribute an explanation to how FAIR came to take its place in the European policy domain – an explanation that does not view the inherent qualities of FAIR as deterministic for its success (see Wittenburg and Strawn, 2018; 2021a; b; Burgelman, 2021), but an explanation that looks at the chain of events that created the environment where FAIR could become policy.

Listening to Jean-Claude Burgelman speak about not wasting a good crisis, thus tuned me into not only the potential a crisis could have on the future promotion of FAIR, but also what role crises could have had in the past to shape the political environment FAIR was promoted into. I couldn't help but think that COVID maybe wasn't the first crisis with significance to FAIR, and after doing my research, it became clear that it wasn't[46]. This led me to the question at the heart of this chapter: what was the role of crises in the introduction of the FAIR principles to the EU's research policy programme? While the focus on crises is sure to create blind spots and make me overlook other important contributions to FAIR's development, supplementing the desirable imaginaries presented in chapter 1 with the non-desirable imaginaries presented here, will contribute to a fuller understanding of the role of sociotechnical temporalities in shaping our societies.

As with the previous chapter, this story will end where most tellings of it begin, the launch of the European Open Science Cloud and the formal introduction of FAIR into EU research policy. Empirically, this chapter will engage with the policies that preceded and led up to this point, starting at the launch of the European Research Area in 2000 and following its journey as other policies were introduced in parallel or added on top of it.

Central to the theme of this chapter is the notion of crises, and the way they are mobilized to promote policy. Inspired by STS scholar Kristin Asdal (2015; 2020), this is not a story about the practices of individuals, but a story about documents and how they perform and configure our realities. Further, this is not a story about the FAIR principles or ideas behind them, but a story about EU research policy – and as such, it makes sense to start with an introduction into EU policymaking.

A Europe in crises

The EU has since its inception been a hub for controversy, always having to legitimize its existence – trying to justify the financial expenditures and relinquishment of political sovereignty of its member states to a supranational actor (Sternberg, 2013). A recurring narrative in this legitimation work has, since the early years of the union, been that without a

[46] As will become evident in this chapter

unified Europe, the individual nations would not be able to withstand the pressures of global developments (Sternberg, 2013). This kind of argumentation was used to justify the highly controversial Maastricht treaty of 1992, which itself, somewhat ironically, helped to *amplify* the already existing skepticism towards the EU as a legitimate actor (Sternberg, 2013, pp. 104).

Since then, one of the main struggles of the European Union has, according to Claudia Sternberg (2013), been one for legitimacy – a struggle she calls “a sisyphean aspiration” (pp. 225).

Uniting the European Research Area

Eight years after signing the Maastricht treaty in 1992, the European Council held a special meeting in Lisbon to formulate a strategy intended to increase Europe’s productivity and stop it from falling behind other international actors (European Council, 2000). As a part of the Portuguese presidency’s focus on an expansion of the knowledge-sector as a driver for innovation (Portugal 2000, 2000), the European Council signed into effect the so-called Lisbon Strategy, and with it launched the idea of creating a coordinated European Research Area, or ERA for short (European Council, 2000).

But, already at an early stage, the Lisbon Strategy struggled to gain traction and was seen as a faltering – if not an altogether failing – strategy (World Economic Forum, 2010). In spite of these struggles, the bureaucrats and policymakers in the European Commission would not let go of the planned European Research Area. In 2007, commissioner for science and research Janez Potočnik, attempted to reinvigorate the discussions around the ERA with an open consultation launched through a green paper[47] with the tagline “inventing our future together” (EC, 2007b). In the document, Janez repeated the need for a coordinated and unified European research area building on the idea of the European Single Market, promising that the ERA would bring European citizens a “fifth freedom”[48] securing the “free circulation of researchers, knowledge and technology” (Potočnik, 2007; EC, 2008a). As part of the initiative to realize this fifth freedom – in particular with regard to knowledge – the Commission promoted “open access to scientific publications and [...] raw data from publicly

[47] A green paper is a term used to denote a policy proposal document that is tentative and non-committing in nature, intended to spark discussion and debate around a given subject.
[48] on top of the already established notion of “the four freedoms” of the EU: the free movement of goods, capital, services, and people.

funded research" (EC, 2008a), but noted difficulties in realizing this vision due to industry backlash and legal barriers (Potočnik, 2007; EC, 2008a)[49].

In 2008 – under the Slovenian presidency of the European Council – the Competitiveness Council of the EU met to "revive the European Research Area" (Slovenian Presidency, 2008a). This revival was promoted as a driver of European competitiveness in order to catch up with US+Japan and to maintain or strengthen the advantage over China (Slovenian Presidency, 2008a; b). With the reinvigoration of the ERA came the forming of the European Research Area Board, tasked with advising how to achieve the vision of the common research area by 2030 (EC, 2009). In its first report, *Preparing Europe for a New Renaissance* (EC, 2009), the ERA board sketched out a vision for how Europe could exploit the opportunities of the growing "knowledge economy" and enter a "new renaissance". But they didn't want to claim what this new renaissance would look like. Instead, they delegated it to other actors to stake out the goal and way forward:

> We believe it an important contribution to our common future to stimulate thought and action on these two basic questions: Where do we want to go, and how do we get there? (EC, 2009, pp. 10).

They did, however, have some idea of where they wanted to go. The "new renaissance" as they envisioned it should – in addition to giving Europeans a "fifth freedom" – put Europe at the front of science and innovation, enabling it to solve "the grand challenges" of our time (EC, 2009). In their vision, a strengthened knowledge sector would produce better medicines, alleviate poverty, help deal with climate change, stimulate the arts, and strengthen democracy (EC, 2009). In order to reach this future and realize this renaissance, the European Research Area Board articulated 6 milestones that the ERA would have to reach within 2030. In accordance with these milestones, the ERA had to be 1) united; 2) driven by societal needs; 3) based on a shared responsibility between science, policy and society; 4) based on open innovation; 5) a deliverer of excellence; and 6) a cohesive entity.

In spite of the report's otherwise revolution oriented narrative, the milestones drew on an already established vision of the European Union – as a driver for peace, prosperity and unity in a continent threatened by fragmentation, stagnation and division. The difference was that, now, research was the activity that would deliver on this vision. How? By being an engine for innovation (EC, 2009).

[49] It is important to note that discussions around Open Access had been around for some time already in 2007. Outside of the EU policy context, the Open Access movement had started picking up steam in the early 2000's with the Berlin, Bethesda, and Budapest declarations. It was only in 2007 that it became part of EU policy with the EC communication on scientific information in the digital age: access, dissemination and preservation (EC, 2007a) and later with the Open Access pilot in the seventh framework programme (EC, 2008b).

As the Lisbon strategy came to its end in 2010, the focus on innovation had only become stronger (World Economic Forum, 2010). At this time, the EU was going through the aftermath of the 2008 financial crisis, and found itself at the height of the European Debt Crisis. The EU did not only have to deal with the ever present crisis of their legitimacy, but also the immediate one of the financial and monetary system the Union was built around.

The strategy that superseded Lisbon – the Europe 2020 strategy – thus needed to not only be a continuation of previous efforts, but also present solutions to the novel and urgent crisis. In the Europe 2020 strategy, the Union expressed a need to grow out of their current predicament and into “a sustainable future” (EC, 2010a). To this end, the Europe 2020 strategy introduced a strengthened emphasis on innovation as a driver for positive development (EC, 2010a). This renewed emphasis was illustrated by Janez Potočnik’s successor Máire Geoghegan-Quinn’s decision to add “innovation” to the job title, then becoming the commissioner for science, research *and* innovation. In addition, the European Research Area became part of a new “flagship initiative”: the Innovation Union (EC, 2010b). Launching the initiative, Máire Geoghegan-Quinn underlined the central role innovation would play in the upcoming years:

> “As we emerge from crisis in the teeth of fierce global competition, we face an innovation emergency. If we do not transform Europe into an Innovation Union, our economies will wither on the vine” (EC, 2010c).

A fear of missing out

Interestingly enough though, the flagship initiative that would become perhaps the most significant in the growth of FAIR in the EU was not the research-oriented Innovation Union, but the market-oriented Digital Agenda for Europe (EC, 2010a). For as the Innovation Union focused on quantitative change: the need for *more* innovation – the Digital Agenda for Europe had a different approach, centered around qualitative change: the need to transition from an analogue to a digital society (EC, 2010a). An important element in this way of thinking was the impression that while other economies (in particular the US) had managed to financially exploit the growing digital domain with businesses like “Google, eBay, Amazon and Facebook”, there was an impression in the commission that Europe had virtually no success in doing so (EC, 2010b). The European Commission further worried that European actors had become dependent on the infrastructures and platforms of other countries (EC, 2010b). They argued that this could not continue, reiterating the imperative from the Slovenian presidency, that the EU would have to catch up with those ahead (EC, 2010b). But, differently from the Slovenian presidency, the EC now emphasized that the EU would have to become less dependent on outside regions by building what they called “a digital single market” for Europe (EC, 2010b).

Still, in spite of its market-orientation, the Digital Agenda for Europe seemed to resonate more with actors in the science and research domain than what the research-oriented Innovation Union had done. Because – while the policymakers and bureaucrats in the EU had been concerned with the wellbeing of the Union – actors in science and research had been more concerned with changing trends and developments in the production of knowledge, specifically in terms of the role of data and computational methods (PARADE, 2009; EC, 2010d; Burgelman et al., 2010).

For some years already at that time, people in the scholarship domain – including the European Research Area Board – had been taking note of changes in the way research and knowledge production in general was being conducted (Science 2.0, 2006; Schneiderman, 2008; Waldrop, 2008; EC, 2009; Tacke, 2010). This perceived change was in particular tied to the spread and adoption of new information technologies, often dubbed “Web 2.0” (see Sterlig, 2005; Schneiderman, 2008; Nielsen, 2008; Waldrop, 2008; EC, 2009, Tacke, 2010). Researchers were increasingly putting their findings and data on the web, but not only through the “traditional” channels of journals, but also on blogs, as preprints, and in online databases, constituting a shift they called “Science 2.0” (Sterlig, 2005; Schneiderman, 2008; Nielsen, 2008; Waldrop, 2008; Tacke, 2010).

Commentators argued that this “Science 2.0” was challenging the established ways of doing and communicating science – and therefore, that approaches to research only concerned with production in the university and dissemination in journals was at risk of falling behind (Sterlig, 2005; Schneiderman, 2008; Nielsen, 2008; Waldrop, 2008; Tacke, 2010). The playing field, they would claim, was not the same as it was before – rendering continuation of old policy approaches inappropriate (PARADE, 2009; EC, 2010d; Burgelman et al., 2010). As part of the commentariat, researchers in the European Research Area Board (Burgelman et al., 2010) expressed a concern that as global competition to exploit the affordances of Science 2.0 increased, it would threaten “national system[s] of innovation”, making organization through “larger regional areas” necessary to ensure competitiveness (Burgelman et al., 2010). This was an explicit nod towards the European Research Area, promoting the initiative as *necessary* for the EU to keep up with other “regional areas” (Burgelman et al., 2010). In other words, the new crises could not be solved with the same approaches as the old – as Burgelman et al. (2010) argued:

> Innovation and research are on top of the policy agenda of developed as well as developing countries, especially in times of crisis and restructuring. [...]Yet governments struggle to design appropriate research policies for a fast–changing and unpredictable world.

"Innovation and research", as was the focus of the Innovation Union, the authors argued, would thus not be a legitimate solution to the oncoming crisis brought forward by Science 2.0. The changing landscape of science wasn't an expressed concern in the Innovation Union (EC, 2010b), and it wasn't in the Digital Agenda for Europe either (EC, 2010a) – but at least the Digital Agenda for Europe expressed the sentiment that there was radical social change on the horizon. Thus, it was through the Digital Agenda for Europe that the EC and the Science 2.0 discourse found a shared concern in the fear of missing out on the next developments in the digital domain. This fear that was perhaps most succinctly articulated in the report of the High Level Expert Group (HLEG) on Scientific Data: *Riding the Wave: How Europe can gain from the rising tide of data* (EC, 2010d):

> History can repeat itself. Today, it is the job of European policy leaders to prevent that happening – at least in the area of scientific data. The danger lies in the history of the Internet and World Wide Web. That technology, as outlined earlier, began growing with the scientific community – initially in the US, but soon in Europe, as well. But it was American companies that first spotted the commercial potential and launched user-friendly browsers and money-making services. Likewise, it was the US government that wrote the rules of the road for what was by then being called "the Information Superhighway." European politicians didn't really catch on until 1993, when the first EU R&D project on the Web was funded. Result: For most of the past 25 years, it has been American companies that have reaped most commercial benefit from the Web. Europe still has no real answer to the likes of Google, Linked-In, Facebook or Amazon; and while challengers are now rising, they come from Asia, not Europe.

The argument from both the European Research Area Board and the HLEG on Scientific Data was that the EU needed to break with contemporary innovation-oriented policies to deal with the new world of Science 2.0, because "The risks attached to this change should make us consider the appropriateness of current policy tools for this new framework." (Burgelman et al., 2010). In spite of firm recommendations to rethink EU science policy sooner rather than later, the political momentum was not sufficient to bring about a radical change as soon as the proponents of Science 2.0 would have it. It would take four more years to bring Science 2.0 onto the stage of EU policy again, in part helped by an emerging crisis in one of the EU member countries.

(Dutch) science in crisis

For some years already, before 2010, there had been stirrings in the world of science – studies that had been considered reliable, and results that had been considered true, started to be taken into question – because on closer inspection, they didn't seem to reproduce (Pashler and Harris, 2012; Fanelli, 2018; Romero, 2019). This led scientists, researchers and others in the academe to worry whether science could be trusted, not only for scientists themselves, but for democratic society at large. And while these stirrings had heated to a simmer on an international basis in 2010, it reached an almost rolling boil in the Netherlands when a combination of events came together to create a national science crisis[50].

Concerned with what they saw as science failing to live up to its promises and people's expectations towards it, four researchers organized a conference titled "Science in Transition" and wrote a position paper titled "Why Science Does Not Work as It Should And What To Do about It" (sic.) (Dijstelbloem et al., 2013). In the paper, these researchers sketched an outline of the changes science was and had been going through, drawing on contemporary societal debates around scientific quality, the developments of computational methods and data sharing, and increased competition in and market orientation of the university sector.

The authors of the position paper continued by diagnosing science as unable to deal with these changes, mirroring the concerns of earlier authors like Burgelman et al. (2010) and the HLEG on Scientific Data (EC, 2010d). But unlike their priors, the Dutch researchers added an emphasis on "trust[...,] reproducibility[...,] reliability and corruption" in science (Dijstelbloem et al., 2013). The way they saw it, science was losing legitimacy as an institution of democracy (Dijstelbloem et al., 2013). This, they argued, constituted a necessity for change, but when it came to suggestions as to what these changes should be, they were not so clear. Their position remained that they would rather start a discussion about how it could be done than dictate concrete solutions:

> [W]e have tried to outline the problems demanding resolution. We have not yet supplied a blueprint for a new and more sustainable university. But we hope to have given you sufficient material to start a fruitful discussion about the future (Dijstelbloem et al., 2013, pp. 31).

And in this sense, they had some success. In the following years, the Science in Transition group spurred, contributed and participated in a series of talks, debates and discussions in academic, political and public contexts (Dijstelbloem et al., 2014; Benedictus, 2015; Miedema, 2022). This academic debate about science in transition fed into an already

[50] This included two controversial vaccination programmes (HPV and swine flu) and the fraud of Diederik Stapel (Benedictus, 2015)

existing one in the political domain, where the then state secretary for education, culture and science, Sander Dekker, was pushing for Open Access to scientific publishing, mainly as a way of stimulating innovation and increasing return on investments in science and research (Dekker, 2013; 2014; Tweede Kamer der Staten-Generaal, 2013).

While these discussions seemed to engage a wide audience in the Netherlands, both the researchers and the policymakers made it clear that any changes to the academic system could not be limited to a national level. The science transition, its proponents emphasized, *had* to be international (Dijstelbloem et al., 2014; Benedictus, 2015; Dekker, 2013; 2014).

Mixing policies together

This, in 2014, is where we come back to Science 2.0. Because not long after the Science in Transition position paper was published and the debates in the Netherlands had picked up pace, the European Commission Directorates General (DG) for Research and Innovation and Communications networks, Content and Technology invited to a public consultation titled "Science 2.0: Science in transition" (EC, 2014a; b; c). As the title suggested, the consultation drew heavily on both the discussions on Science 2.0 *and* the Dutch Science in Transition initiative[51].

As the prior initiatives were mostly focused on starting a discussion about the academic system, the Directorates General wished to produce some more constructive outcomes. In the invitation to the consultation, they listed three main objectives:

> (1) to assess the degree of awareness amongst the stakeholders of the changing modus operandi;
> (2) to assess the perception of the opportunities and challenges and
> (3) to identify possible policy implications and actions to strengthen the competitiveness of the European science and research system by enabling it to take full advantage of the opportunities offered by Science 2.0. (EC, 2014a; b; c)

When the Directorates General published the outcome of the consultation, they went a long way in reproducing the original premise for it. Even though they found that only 25% of the respondents had an "awareness" of the transition to Science 2.0, they also found that 70% of respondents were in "agreement" that the transition was taking place (EC, 2015a, pp. 6). To explain this discrepancy, the report authors suggested that while many actors and institutions recognized a general shift in the "modus operandi" of science, they didn't agree fully on what this shift actually was (EC, 2015a).

[51] This is also highlighted in the background document for the public consultation (EC, 2014a, pp. 11)

A survey among the respondents to the consultation showed that what the different actors perceived to be “key drivers of Science 2.0” was neither unanimous nor in complete agreement with the DGs initial thesis. While 98% of the respondents agreed that “availability of digital technologies and increased capacities” was an important factor for this change, they disagreed whether factors such as “citizens acting as scientists” or “growing criticism of current peer-review system” were central to this change or not (EC, 2015a).

These results were further muddled by the fact that the respondents seemed to be talking about slightly different phenomena. As put in the validation of the public consultation: “The results of the consultation suggest that many stakeholders prefer using an alternative term to ‘Science 2.0’”[52] (Ec, 2015a, pp. 6). This problem led the DG to suggest a change of terms (EC, 2015a; Hoefler et al., 2015). Among a list of six terms, the DG landed on “Open Science” as the new moniker to be used going forward[53]. A central contribution of this process was thus the synthesis of different concerns and tying them to the ongoing Open Access initiative of the EU.

In this sense, the validation report presented both a continuity and a break. It performed a continuity by describing the changes as being in line with the original assumptions on Science 2.0 – even equating the terms Science 2.0 and Open Science. Yet, it represented a break in the obvious sense that it changed the name of the transition, but also less obviously in that it marked a shift in attention towards open publication and sharing of knowledge, not only in the form of articles but also data (EC, 2015a). Thus the outcome of the consultation on Science 2.0 ended up steering the EC towards other debates and developments tied to Open Scholarship and data sharing initiatives.

The same year as the public consultation on Science 2.0, the High Level Expert Group on Scientific Data was working on a follow-up report to their *Riding the Wave* Report (EC, 2010d) (now no longer as an HLEG, but under the RDA banner). The resulting report, *The Data Harvest* (RDA, 2014), echoed much of the sentiment of its predecessor and tied it into the ongoing debates on Science 2.0 – arguing that the growth in research data in particular would change not only research, but society on a wider scale, claiming that “this new data wave will matter eventually to every one of us, scientist or not” (RDA, 2014, pp. 5). This, as in the 2010-report, was paired with an argument that the EU stood at a risk of missing out:

[52] The reasons were not few. In addition to the respondents preferring other terms, the DGs also recognized the troubles tied to using a name that was already trademarked in the US. Further, they were afraid that the name Science 2.0 would imply that science was lagging behind industry, which was at that time imagined to go through a “fourth industrial revolution” earning it the similarly formatted name “Industry 4.0” (Hoefler et al., 2015).

[53] The other suggestions were: participatory science, science highway, better science, open research and open scholarship

“Europe’s leaders, including its new slate of European Commissioners and Parliamentarians, must act – or go down in history as the politicians who missed the Next Big Thing” (RDA, 2014, pp. 5).

What *The Data Harvest* (RDA, 2014) did differently than *Riding the Wave* (EC, 2010) though, was that it explicitly and forcefully promoted an economic imperative not only based on opportunity, but on cost. The authors expressed that they were content with how the Barroso Commission had taken the suggestions from *Riding the Wave* (EC, 2010d) and secured investments in data infrastructures across Europe (RDA, 2014, pp. 19). But, this was in the opinion of the authors not enough: the new Juncker Commission would have to follow up. In the words of the report:

> As a new Commission and Parliament begin work, we urge they pay heed. The seeds have been sown. Now is the time to plan the harvest (RDA, 2014, pp. 3).

Spotting a cloud on the horizon[54]

Now, both the Validation of the *Public Consultation on Science 2.0: Science in Transition* (EC, 2015a) and *The Data Harvest* (RDA, 2014) had made their cases clear: because of a digital transition, research – and in particular research data – was not being handled optimally, and this situation demanded action. While the calls for action in 2010 (Burgelman et al., 2010; EC, 2010d), had failed to generate any significant political action in the EU, these calls in 2014/2015 would become more successful.

On the 6th of May 2015, five months after publication of *The Data Harvest*, and two months after the *Validation of the public consultation*, the European Commission launched an idea they called “the European Open Science Cloud” (EOSC)[55] in a communication titled “A Digital Single Market Strategy for Europe”[56] (EC, 2015b). Three weeks later, the competitiveness council of the EU held a policy debate on “open and excellent science, as a follow-up to the Science 2.0 public consultation” finding wide support for the idea of an European Open Science agenda (Council of the EU, 2015b). In their conclusions on *open, data-intensive and networked research as a driver for faster and wider innovation*, the council

[54] This title is shamelessly stolen from Barend Mons, who chaired the first HLEG for the European Open Science Cloud. “A cloud on the Horizon” was his preferred title for the HLEG report, but it was stopped by Commissioner Carlos Moedas, in favor of the less ominous title “Realizing a European Open Science Cloud” (Mons, 2021).
[55] In that particular document it still didn’t carry the name yet, but was referred to as the “research open science cloud” (EC, 2015b).
[56] The name “Digital Single Market”, if you recall, stemmed from the Digital Agenda of 2010 (EC, 2010b), but back then only in lower case

> STRESSE[D] that open, data-driven and networked research can maximize Europe's digital potential through fostering faster and wider innovation while taking into account the legitimate interests of the stakeholders;

and

> WELCOME[D] the further development of a European Open Science Cloud that will enable sharing and re-use of research data across disciplines and borders, taking into account relevant legal, security and privacy aspects (Council of the EU, 2015a).

These policy intentions were not only tied to the recent Digital Single Market strategy and the public consultation on Science 2.0, but also on earlier commitments to the European Research Area, and the, now combined, Europe 2020 flagships *Digital Agenda for Europe* and *the Innovation Union* (Council of the EU, 2015a. pp. 2). In other words, this moment in European policymaking not only launched the beginnings of the European Open Science Cloud, but did so by drawing on and combining a long line of previous work in the Union. By amalgamating innovation with digital transition, the commission managed to create both a continuity of existing policies and a break with existing approaches – all packaged into the European Open Science Cloud.

But the EOSC was far from a solid commitment, nor was it a particularly fleshed out idea. According to Jean-Claude Burgelman (2021), the vague nature of the abstract "Cloud" combined with the estimated price of €10bn made it hard to generate further commitment to the EOSC (Burgelman, 2021). The way Jean-Claude Burgelman saw it, the concept had to become more concrete in order to justify the investment (Burgelman, 2021).

In order to bring the idea of the EOSC further, the new Commissioner for Research, Innovation and Science, Carlos Moedas, aimed to set up a High Level Expert Group on the European Open Science Cloud (EC, 2016c). This group needed experts – preferably someone who specialized in research data infrastructures and had ideas on how to realize this European Open Science Cloud. Somewhat coincidentally, at the same time, a group of scholars were working on promoting their ideas on what could be seen as a similar system: the internet of FAIR data and services – and through his previous engagements with the European Commission, the head of this group, the dutchman Barend Mons, happened to "have the ear" of the Director General for Science, Research and Innovation, Robert-Jan Smits[57]. Thus, in September 2015, Barend Mons was made chair of the commission's High Level Expert Group (EC, 2016a).

[57] Interview with Niklas Blomberg. See also the HLEG meeting documents (EC, 2016d)

It was not unimportant that the person in charge for the High Level Expert Group was a dutchman, because the following year of 2016, the Netherlands would take over the presidency of the Council of the European Union. On the background of the controversy around their national science system, they were, according to Burgelman (2021) still trying “to supranationalize some of the intense debates in the Netherlands" around their knowledge sector. This effort was also evident in a call for action prepared by the Dutch presidency pushing not only for Open Science, but also FAIR science (NL EU 2016, 2016). This call was soon after ratified, encouraging “Member States, the Commission and stakeholders to follow the FAIR principles in research programmes and funding mechanisms” (Council of the EU, 2016).

The Dutch call turned out to be a great support for the HLEG report, published five months later, with the title *Realising the European Open Science Cloud* (EC, 2016c), as it not only drew on the preceding narratives of federation, innovation, increased transparency, and a changing mode of knowledge production, but also on the *need* for FAIR data (EC, 2016c). The authors argued that if the coming wave of open research data (spurred on by the Horizon 2020 Open Data pilot) was not handled properly it would not create value, but could instead become a multi-billion-euro cost to the EU (EC, 2016c). Barend Mons underlined the urgency and gravity of the situation, stating:

> If we do not act, there might be a looming crisis on the horizon. (EC, 2016c. pp. 5).

After the introduction of FAIR to the EU, the principles only gathered momentum and cemented themselves as one of the core concepts of EU science policy. It became a central part of Commissioner Carlos Moedas science strategy, *Open Innovation, Open Science, Open to the World* (EC, 2016b); it gathered strong political and financial support from member countries, Germany and the Netherlands, who formed the GO FAIR organization (Germany and the Netherlands, 2017); it brought together over 40 institutions of science, committing to realize the European Open Science Cloud; it made its way into the budget of EU’s eighth framework programme, Horizon 2020 (Moedas, 2017); and in 2018, in Vienna, the European Open Science Cloud was officially launched as a policy commitment[58] (EOSC Vienna Declaration, 2018)[59].

---

[58] In the years after, and until now, it is still an open question whether the European Open Science Cloud is realized or not. It is accessible through a web-portal (www.EOSC-Portal.eu), but making research FAIR – its central component – remains an elusive task. This, we will get back to in the following chapters.

[59] Supposedly, the launch was hurried to 2018 because of the Cambridge Analytica scandal, and the Commission’s fear of losing control of European data (Burgelman, 2021)

A note on documents and futures

Before we get into understanding this process that led up to the launch of the EOSC, we will have to consider what we are looking at. The narrative above is in large part a weaving together of different threads – separate stories that in some ways are separate, yet in other ways intersect and come together, driving each other forward and – in this case – ending up at the introduction of the FAIR principles as EU science policy. There are, naturally and evidently, other ways to tell this story. One notable example is that of Jean-Claude Burgelman (2021) which has found its way into this text as well. In addition, there are the personal stories that my informants have shared with me through interviews, and the stories told in conferences and other places unavailable to the public and the prying eyes of a social science doctoral researcher. I want to make clear, yet again, that I am not attempting an official story – the explanation of how it *really* happened, but a story that opens up to a discussion and analysis on a particular topic: the use of crises and expectations as a basis for political action through the use of documents.

Both the the methodological and the theoretical work of this chapter[60] is based on the concept of document ethnography: following relationships between texts and how they act upon each other and the world (Asdal, 2015; 2020). For, as Asdal (2015) puts it: "Paperwork does not simply describe an external reality 'out there': Documents also take part in working upon, modifying, and transforming that reality".

In my analysis of the documents in this story, I've paid particular attention to the justification and legitimation of political will and agency – how strategies and actions are argued for and linked into longer chains. This form of discoursive action places ideas into societal and political context, affecting the meaning they carry, thereby acting as a form of "modifying work" (Asdal, 2015). Now, why is this interesting?

Policies and ideas are not inherently meaningful on their own, but are interpretively flexible, affording different meanings: a research policy can be seen as both an economic matter, and a democratic matter, or something else entirely. This means that when policies are ascribed or given meaning, this is an expression of a particular version of reality (Asdal, 2015). This modifying work is not innocent, but can be applied as tactics used to achieve or avoid certain (un)desired outcomes (Hoff, 2023). In spite of performances of neutrality or objectivity, descriptions and ascriptions of policy concepts are still politics (Asdal, 2015; Hoff, 2023).

[60] A method assemblage (Law, 2004), if you will

Scholars of Science and Technology Studies have concerned themselves in particular with how this type of political will is exerted through mobilizing narratives about the future (see Oomen et al., 2021). Although the terminology used differs – whether it is expectations, imaginaries, futures or visions – the general essence of the argument remains the same: by telling stories about the future, actors can shape the present.

There are stories about good futures (Jasanoff and Kim, 2015), and there are stories about bad futures (Felt, 2015). There are stories about likely futures (Borup et al., 2006), and unlikely futures (Tutton, 2011; Gardner et al., 2015). And importantly, there are stories about how to avoid or realize those futures. By using these stories tactically, actors can enroll others into their strategies, and thereby give them power (Paltieli, 2021).

An important note is that these stories, or narratives, are never settled, and often exist in a state of contestation with alternative ones (Hoff, 2023). A central element of enacting these stories is therefore about establishing one as more likely than or desirable as the others, and thereby closing down discussions about the way forwards into the future (Levidow and Neubauer, 2014). In this aspect, documents can be central tools (Asdal, 2015; Sepehr and Felt, 2023). Studying them, therefore, can give a rich insight into both the use of these stories and their influence (Sepehr and Felt, 2023).

Now, I will not look at the stories presented in this chapter in an attempt to explain how they gained agency – instead, I will look at which narratives that were used in the establishment of FAIR as EU policy and how they have been mobilized to fulfill certain futures (now past). This will further allow us to gain insight into what narrative tactics that resonate with the political project(s) of the European Union and what this might mean for the future of FAIR.

With that in mind, let’s get to it.

Crisis as a driver

In the introduction I mentioned the work of Claudia Sternberg (2013), who posited that the European Union seems to be in a constant state of crisis. In her conclusion, she argues that this is a burden, comparable to that of the Greek tragedy of Sisyphus: that the EU has to eternally work an uphill struggle for legitimacy.

I argue the opposite: that the EU's constant state of crisis is a resource – a repertoire of political agency. To paraphrase the existential thinker Albert Camus, I suggest that we must imagine the EU happy.

In this chapter, we've seen that from the launch of the European Research Area to *Realising the European Open Science Cloud*, crises have been used as motivators for political action. We should note that while the solutions were sometimes strikingly similar, the crises that were used to justify expansion of the EU's research policy would often change. Sometimes it was about economic crisis, sometimes about democratic crisis, sometimes about political crisis – and often a mix of these. But I wish to draw attention to another distinguishing factor of the crises we find expressed in these policy documents: their temporality.

In the story leading up to the launch of the EOSC in 2018 there were three main mentions of crises: past, present and future ones. This is perhaps most clearly exemplified in the three following statements.

Past: Máire Geoghegan-Quinn's statement on the launch of the Innovation Union: "As we emerge from crisis in the teeth of fierce global competition, we face an innovation emergency. If we do not transform Europe into an Innovation Union, our economies will wither on the vine" (EC, 2010c).

Present: the European Research Area Board's third report, with the subtitle: "Innovating Europe out of the crisis", suggests turning their vision of the ERA into reality is "the only way out of our present crisis".

Future: the EOSC HLEG report, suggests that if their vision of the European Open Science Cloud isn't realized, "there might be a looming crisis on the horizon" (EC, 2016c. pp. 5).

In these cases, the crises are used to the same end: to legitimize, justify and generally motivate the promotion of research and innovation policy – and in a hurried manner. The structure of the arguments are also similar, in that they suggest that "*if* you don't follow this suggestion, something bad will happen (sooner rather than later)". What this "something bad" is supposed to be differs slightly in these narratives, but centers around the same thing: not recovering from the crisis, not getting out of the crisis, or encountering an entirely new one.

But these crises are not necessarily "real", in the sense that they are taken as evident and agreed-upon[61]. This is perhaps most intuitively true in the case of the future crises, but can also be applied to the past and the present.

While the financial crisis of 2008 and the ensuing Eurozone Crisis were generally undisputed, there were contentions around the existence of the crisis in science. In the position paper *Why Science Does Not Work as It Should And What To Do about It*

[61] In a social constructivist sense

(Dijstelbloem et al., 2013), the authors were careful and critical of using the term, stating: "Is the university in crisis? It is much too easy – and also not very productive – to use this much abused label for the current situation". A similar sentiment is found in Fanelli's *Is science really facing a reproducibility crisis, and do we need it to?* (2018), where he suggests that "The new 'science is in crisis' narrative is not only empirically unsupported, but also quite obviously counterproductive".

This caution towards the crisis narrative is telling. In the case of Dijstelbloem et al. (2013), the authors are critical of the term because they would rather inspire careful reflection than drastic change. Fanelli (2018) rejects "crisis" mainly because he sees it as a threat to the autonomy of the sciences, as the narrative could legitimize external intervention instead of his preferred internal inspiration – as if to say "science can deal with this challenge on its own". In other words, the crisis narrative can be seen as a potential justification for outside forces to drive radical change.

This is part of the power the crisis narrative has, that it acts as a door-opener for intervention and radical change – that it deems current practices insufficient and in need of replacement. In a crisis narrative "business as usual" can not be applied to get out of a new and unusual situation. This is evident in the way the Innovation Union initiative struggled to gain legitimacy as a policy response to "Science 2.0", while the Digital Agenda for Europe gained resonance through its focus on qualitative shifts in the "modus operandi" of societies and, in extension, knowledge production (EC, 2015a). Still, the fact that the concerns around Science 2.0 took so long to materialize as policy is testament to how a crisis can be stretched out over several years in spite of the initial claim of urgency to attend to the matter. Repeating the crisis narrative thus extends the urgency beyond its original frame of reference.

With that in mind, it is then notable that several "old" policies stuck around, even after the crises they addressed disappeared into the background. While the financial crisis of 2008 goes wholly unmentioned in documents of later date, the solutions suggested to combat it, like federation of national research efforts and increased innovation remained. Even after the threat of financial breakdown ceded, it remained urgent to federate the nation states' efforts, but for different reasons, be they European competitiveness, digital transformations, or concerns for the legitimacy of democratic institutions. This is exemplified by how even though the Innovation Union failed to sufficiently address the concerns within science, it still became a part of the EU's science policy strategy – but only by combining it with the Digital Agenda for Europe. In other words, as crises pass or fail to materialize, the reactions to, or anticipations of them remain, making them a form of, what Berg et al. (2020) call a sticky policy.

The persistence of policy solutions beyond their acute relevance thus suggest that crises are not the primary concerns of policymakers, but, rather, that crises are used as justifications for policy making. Not wasting a good crisis, as Burgelman suggested in the introduction of this chapter, is then less of a call for innovation to combat the crisis than a call for exploitation of the potential the crisis affords.

With this in mind, the image of the EU as a suffering Sisyphus in suit and tie might not serve as the best metaphor. For, while challenges to the EU's legitimacy might return for as long as the institution exists (Sternberg, 2013), the labor of defending it does not need to be undone like a boulder rolling down a hill. In the case we have looked at in this chapter, legitimation efforts do not dissipate as the challenges do. Instead, they accumulate and stick around. Crises, real or imagined, thus serve as repertoires for policymaking and political agency.

One strategy for the EU?

With the policy responses of these different crises coming together it becomes pressing to ask if these accumulating policies add up to a single unified whole, or only form a disparate, fragmented and directionless policy landscape?

In the documents presented in this chapter we see a number of different crises being utilized as justifications for policymaking that not only accumulate, but (for this chapter) culminate in the launch of the EOSC in 2018. The crises, or dystopian narratives, that were mobilized range from the EU becoming fragmented, left behind in global competition, losing out on (knowledge-)industry developments, and weakening its democracy. While many solutions were often recycled, each crisis brought with it a novel response: fragmentation was met with federation, competition with innovation, losing out with reorientation, and democratic erosion with transparency.

But while the responses were different, the EU policymakers managed to tuck all of these policies into the same package: the European Research Area. This, of course, did not happen out of some inherent relationship between the concepts – as demonstrated by the ERA's pivot from the Innovation Union to the Digital Agenda for Europe (and later, the Digital Single Market). Throughout the policy journey of the European Research Area, different policy objectives (including the ERA itself) were adapted in order to fit the relevant narratives.

The flexibility of these concepts were integral to allow this modifying work, and thus propel them into the future. To this point, the ERA has been presented as a way of uniting EU member-states, as a driver of innovation, and as a tool for saving costs and increasing return on investments. Similarly, the notion of Open (be it access, science or data) has demonstrably gone through a similar process where it was introduced as an initiative to ensure the free movement of knowledge in the ERA, and ended up as an imagined new paradigm of knowledge production. These policies can then be seen as a form of vessels able to carry new policies, functioning as more-or-less stable entities that policies can stick to.

But as these policies get modified to carry others and to align with contemporary narratives, they might come into conflict with each other. This is most clearly made evident by the addition of the FAIR principles to the EOSC. The emergence and development of the EOSC, even while drawing on the input of international scholars and concepts, was always clearly embedded in the European Union. As an effect, the EOSC was intended to address EU needs and concerns, in line with the preceding policies stuck onto the ERA. But with the introduction of Barend Mons and his FAIR principles shortly after, this EU-centeredness turned into a problem.

The conflict between EOSC and FAIR is perhaps best illustrated on page 8 in the EOSC HLEG report *Realising the European Open Science Cloud* (EC, 2016d), where Barend Mons states that, in spite of its name, the European Open Science Cloud can neither be just European, open, only about science, nor a cloud (Mons, 2021). This redefinition of the EOSC was not an attempt at undermining the concept per se, but rather at modifying it, making it fit better with his concept of the FAIR principles, and *the Internet of FAIR Data and Services* (EC, 2016d, pp. 7).

This modification effectively unsettled the EU's policy imperatives of a federated European Research Area on the EU's terms, and the idea of the EOSC as a source of competitive advantage for the EU. It remains to be seen whether these two policy concepts are flexible enough to carry each other or if one has to be jettisoned to keep the other intact. EU policymakers and proponents of FAIR contend that the two are compatible, but as later chapters of this thesis will demonstrate, this remains a ground for contestation.

Challenging flexibility

So what is the role of crises in the introduction of the FAIR principles to the EU's research policy programme?

It might be obvious, but the FAIR principles did not come to the EU out of nowhere, and it did not come into an empty space. In this chapter I've tried to show the processes and events that paved the road for FAIR's introduction as EU research policy. Through following these, we see that the road to EU policy is paved with stories of crisis – each tile laid down with a concern for something bad happening – each expense justified by the cost of inaction. These crises, real or imagined, have fueled the development of EU research policy, functioning as a repertoire for political agency. They have propelled policy concepts like the European Research Area, Open Science and the European Science Cloud into the EU policy sphere.

Even after the crises ceased to be relevant, the policies that were justified by them stuck around, accumulating and sticking to each other, creating compound policies. Still, using crises as policy justification is not without consequences. It necessitates modification of policy concepts in order to align them with the narrative(s) of the crises. This means that dependence on a crisis-narrative to push policy, thus, necessitates a certain flexibility of the policy introduced. As policies become tacked onto existing policies and clustered with others, they become weaved together into compound narratives that may reduce their flexibility. The modification of Science 2.0 into Open Science stands as a clear example of this – how the general notion of a changing science was transformed into the more narrow matter of openness.

As adapting a policy to a narrative also changes its meaning, and clustering reduces flexibility, policymakers run the risk of putting policies in tension with each other, threatening their continued coexistence. Most of the EU research policies presented in this chapter have been shown to be both flexible and sticky, meaning that they went through modifications to align with contemporary narratives, *and* remained a part of established policy after their original purpose ceased. This is not to be taken for granted though, and may not become the case for the FAIR principles, as they stand in tension with the European Open Science Cloud. Time will show whether FAIR is flexible enough to adapt to crises past, present and future, or if it's sticky enough to stay around regardless.

As we progress into the following chapters of this thesis we will see the FAIR policy challenged, its flexibility and stickiness tested, and its future speculated. When we do so, I encourage you to keep in mind, what crisis does FAIR respond to, if any?

Chapter 3: FAIRsplaining

When I stumbled across the FAIR principles in the autumn of 2020, it struck me that this must be something big. It stood at the center of what was then the newly launched European Open Science Cloud (EC, 2018), and the paper publishing the principles four years prior had already amassed thousands of citations[62]. In the words of Higman et al. (2019): "Little in the research data field [had] gained such traction and universal acceptance as the FAIR data principles".

In the virtual space – safe from the viral dangers of the physical world outside – I attended conferences and talks, read papers and comments. I didn't understand all of the technical jargon – about PID, RDF and knowledge graphs – but everything I read supported what I had already inferred to be true: the FAIR principles were the future of research.

As the pandemic withdrew, and the physical world with all its inhabitants once again became accessible, I started to notice cracks in the picture from before. People I talked to were less impressed by the FAIR principles than what my previous sources suggested they would be – most of them didn't even know what FAIR was. I tried to explain, but I couldn't seem to get the message across.

I wasn't alone though. According to several publications, many – actually most – researchers didn't understand FAIR (Digital Science, 2020; EC, 2020; David et al., 2020, DANS et al., 2022), to the extent that some saw the lack of understanding as one of the main barriers for the uptake and realization of the FAIR principles (EC, 2020). In order to overcome this barrier, a number of papers were published to clear up any confusion and explain FAIR to those who struggled to understand (see Mons et al., 2017; 2020; Jacobsen et al., 2020). Regardless, FAIR seemed to be remarkably resistant to implementation. Maybe the people promoting FAIR, like me, also struggled to get the message across?

The more people I talked to, and the more publications I read, the more I started to get the feeling that maybe people's capacity to understand wasn't really the problem – despite the proponents of FAIR insisting that it was. Following a praxiographical approach – where the ontological state of a phenomenon is determined by how it is practiced, rather than some intrinsic nature – the notion of "misunderstanding" FAIR makes little sense. So in this chapter, I do not aim to make a judgment on people's capacity to understand. Instead I interrogate the discursive strategy of labeling dissent as misunderstanding, and why the

[62] Clarivate data on Wilkinson et al. (2016)

people closest to the FAIR principles began to use it, and why they were so persistent in their position that everyone else was the problem.

Phases of development

In order to understand the particular strategy of the original proponents of FAIR, we will take a look at how they expressed it throughout their attempts at making science FAIR. But before we get to that, we will need to go back and look at where it started, at the conception of FAIR.

The conception of an idea

To answer this question, we will need to go back in time, to before anyone had heard of FAIR, to when the FAIR principles was called the FAIRport convention[63]. A handful of people from science and surrounding domains had gathered at the Lorentz Center in Leiden, Netherlands to design an initiative that would pave the way for what they saw as the coming revolution of data-intensive science (Lorentz Center, 2014; International Innovation, 2014)[64]. Among them were people “representing leading research infrastructures and policy institutes, publishers, semantic web specialists, innovators, computer scientists and experimental (e)Scientists” (FAIRPORT, 2014). What brought them together was their shared interest in the potential of combining computational methods and digital research data.

Throughout four days the conference participants had several workshops, plenaries and discussions addressing “the context, user needs, technological challenges, access & business models, funding requirements and governance of such a global initiative” (FAIRPORT, 2014) The outcome of these activities was “a remarkable consensus” where “all participants agreed that a global infrastructure [...] is essential for effective data driven research”. They argued that their FAIRports would realize this global infrastructure by copying the developmental strategy of The Internet. Just as the TCP/IP protocol had acted as the “narrow waist” of an hourglass, tying the users and service providers of The Internet together, FAIR would do the same for the users and providers of “a fully interoperable ‘Internet of Data’” (FAIRPORT, 2014).

Now, the next question was how to move this idea out from the conference center and to the rest of the world.

---

[63] According to Tom, “The FAIRPORT Convention” had to be scrapped in order to “[avoid] copyright issues about this name”. The reason being that it was already taken by a 60s British folk-rock band.
[64] For an more in-depth look into the FAIRport conference and the ideas behind FAIR, see chapter 1

On the final day of the conference, they held a meeting with representatives from "policy circles and funders", most notably the European Commission[65]. The FAIRport idea got a warm reception, with "verbal commitments" that were found "very encouraging" by the conference participants (FAIRPORT, 2014). Outside the walls of the conference center, they reasoned, were more people who were inclined to agree with their newly developed idea (FAIRPORT, 2014). Some of these external actors had either resources, influence or agency that was perceived as helpful to realize "the internet of data". The FAIRport group would "need to find a way to make all [of them] part of this initiative" (FAIRPORT, 2014). A subgroup of the participants were then tasked with reaching out to "all enablers and other stakeholders that were not present at the meeting" and get them on board.

An idea worth spreading

In the months after the initial conference, the participants submitted updates to each other, describing how "Niklas Blomberg and Barend Mons visited the European Commission with H2020 and DG Connect staff attending to present the Data FAIRport initiative"[66], or how "Bengt Persson[...] has also taken the initiative to set up an Interest Group in RDA"[67] (Data FAIRport, 2014a). In total, at least six additional organizations and institutions were contacted by the end of the reporting in September 2014, with more in the pipeline.[68] The report authors were clear in that these networking activities were strategically employed in order to scale the FAIRport principle (FAIRPORT, 2014). As they put it themselves: "the ultimate aim [of these activities is] to gain enough momentum and critical mass to convince the [Global Research Council] in 2015 that this community is the most solid embedding for a global policy" (FAIRPORT, 2014).

In the meantime, they continued to develop the idea of the FAIRport convention in different fora, including the RDA, ELIXIR and the NIH (the US National Institutes of Health) (Data FAIRport, 2014a; Wilkinson et al., 2016). But perhaps the most substantial work was done in the FORCE11 community (Wilkinson et al., 2016) – an interest organization advocating the transition from publishing papers to publishing data "through the effective use of information

[65] The full quote from the report lists the European Commission, IMI, NGI, BBSRC, the RDA, DANS, and the Dutch Ministry
[66] This was one of the meetings that ended up in the adoption of FAIR by the EU as we saw in the last chapter. H2020 stands for Horizon 2020, the eighth framework program of the European Union
[67] This was one of the meetings that got Barend Mons and Peter Wittenburg together, culminating in the reformulation of the FAIRport to just "FAIR", as we saw in the first chapter.
[68] The FAIRport participants confirmed to have been in touch with the EC, NIH, NSF, RDA, ELIXIR, Force11 (JDCP), EUDAT, DONA. Meanwhile, they were planning to contact the IMI, ORCiD, VIVO and other "enablers that we identified at the FAIRport meeting" (Data FAIRport, 2014c)

technology" (FORCE11, n.d.)[69]. This work resulted in the publication of an open draft of the first FAIR guiding principles on the FORCE11 website (FORCE11, 2014).

Later the same year, from the 22nd to the 24th of September, 2014, the RDA was organizing their fourth plenary in Amsterdam. During the first day, the Director General of Research and Innovation at the European Commission, Robert-Jan Smits, held an appeal to the scientific community, saying that:

> We are counting on the activities of the Research Data Alliance and other initiatives such as FAIRport. They have the task of coming up with bottom up, user-led solutions, to ensure that the data that is generated in Horizon2020 or other funding programs is FAIR: findable, accessible, interoperable and reusable (RDA, 2014a, 17:10 - 17:36).

When it was Barend Mons' turn to come to the stage and present the FAIRport initiative, he continued on Robert-Jan's plea, encouraging all participants to visit the first draft of the FAIR principles and leave their comments. But time was of the essence, because a week later, on the 1st of October, the principles would be closed for comments and opened for endorsements (Mons, 2014).

But while FORCE11 and the RDA were arguably "bottom-up" initiatives, they were arguably not demographically representative of science as a whole – as they both consisted mainly of western computer- and information scholars. As the groups of people participating in the formation of FAIR remained rather homogenous, FAIR also became characterized accordingly. After its initial presentation at the RDA plenary in September 2014 by Barend Mons, Christine Borgman, who held a keynote there, criticized the dutchman for applying his assumptions too widely: "Genomics found ways to work for itself, but it doesn't generalize very well across other sciences. [...]Genomics is the exception" (Borgman, 2014). After the keynote, her concerns were brushed away by Barend Mons: "I was clearly, yesterday, talking about data publishing [...]. I was not extending that to the humanities or anything, or review papers or whatever". Christine's intervention on stage was thus classified as irrelevant, and any off stage comments seemed to have left little impact on the principles: because a week later, when the FAIR principles were fixed in place, published on the FORCE11 webpage, traces of Christine's comments were hard to find.

According to the internal notes of the FAIRport initiative, the months after the publication were characterized less by development of the principles and more by efforts to spread the word (Data FAIRport, 2014a; b; c). Within the end of 2015, the proponents of FAIR had succeeded in getting the support of significant actors in the science policy world. In the EU,

[69] This is the FORCE11 we encountered in chapter 1

they had the support of the European Commission through their Horizon2020 funding programme and the European Open Science Cloud project. In the US, they had the support of the National Institutes of Health through their (later defunct) Big Data to Knowledge programme[70]. In spite of their national and regional success, they struggled to achieve their "ultimate aim" of convincing the Global Research Council to endorse their principles.

Still, failing to convince the Global Research Council to make FAIR a "global policy" in 2015, the involved actors kept the original principles largely the same. The following discussions were less concerned with opening up for changes to the principles, and more concerned with tuning the emphasis and focus of the initiative. Supposedly, one of the main discussions leading up to the official publication of FAIR in 2016 was whether a strong emphasis on semantic interoperability was necessary, or if it would be better to have a more technology agnostic approach to achieve reusability[71].

Anticipating concerns from outside of the group was not considered: they already saw themselves as representing "a diverse set of stakeholders [from] academia, industry, funding agencies, and scholarly publishers" (Wilkinson et al., 2016). Throughout the first publication in 2016, they repeatedly used language to argue for the general validity of the FAIR principles. Unlike other "domain-focused publications", FAIR was described as "domain-independent", and something that would come "to the benefit of the entire academic community" (Wilkinson et al., 2016). After this performance of inclusivity, the paper rounded off with an invitation to publishers and anyone producing data: "by working together towards shared, common goals, the valuable data produced by our community will gradually achieve the critical goals of FAIRness" (Wilkinson et al., 2016). Thus, the doors for participation seemed to be officially opened (again).

At the same time, the European Commission had given FAIR a central spot in their new research strategy, building their prestige project, the European Open Science Cloud around it. A High Level Expert Group (HLEG), chaired by Barend Mons, recommended that the EC funnel money into building prototypes of the necessary infrastructures of the Internet of Data, which they had renamed the Internet of FAIR Data and Services – or IFDS (EC, 2016). Additionally, they suggested that FAIR was not a part of EOSC, it was the other way around – EOSC should be "the EU contribution to an Internet of FAIR Data and Services" (EC, 2016, p. 7)

[70] The Big Data to Knowledge (BD2K), was an NIH initiative, spearheaded by FORCE11 founder and FAIR author Philip Bourne. Together with the NIH Commons, it was supposed to serve as a US Semantic Web of Research Data – an equivalent to the EU's Open Science Cloud. Philip left the NIH in 2017, and after its initial pilot run the BD2K was scrapped. (NIH, n.d.; Bourne, Bonazzi et al., 2015; NIH, 2017; Jagodnik et al., 2017)
[71] Interview Niklas Blomberg

After this, the EC put north of 250 million euros towards advancing the European Open Science Cloud and/or FAIR science in Europe (EC, 2020b). This gave the FAIR principles a considerable boost by spreading the word to every researcher who would be interested in research funding from the EU: make your research FAIR.

Reception and reaction

After the publication in 2016, the paper quickly became highly cited by academics from an array of different backgrounds and fields, although most of them still from the life sciences[72]. The reactions to the publication of the FAIR principles were mixed. One of the very first papers citing Wilkinson et al. (2016) was authored by two biologists critiquing FAIR for not focusing enough on “data liberation”, like Open Science did (Alquraishi and Sorger, 2016). But most of the incoming citations were decidedly supportive of the principles, some claiming they were such a good idea that they were already following them (Zerbino et al., 2017), or that they soon would (Buniello et al., 2018)! And with each citation, the community of people participating in the formation of the FAIR principles grew.

But the large amount of citations and support was not only welcomed by the authors. In particular two issues emerged from this: 1) the way the citing papers interpreted and expressed FAIR was different than what Wilkinson et al. (2016) desired, and 2) the overwhelming majority of citing papers still came from the computational- and life sciences, giving FAIR a reputation as a disciplinary niche concept.

Regarding the first issue, the Wilkinson et al. (2016) authors later expressed that those who cited the it were mistaking the FAIR concept as corollary to Open Science (Mons et al., 2017; 2020; Jacobsen et al., 2020), thought putting a DOI on a PDF made them FAIR (GO FAIR, 2021), and probably hadn’t even read the paper (Mons, 2020). Regarding the second issue, the authors expressed that those who didn’t cite the Wilkinson et al. (2016) paper, did so only because of “common misconceptions”, allegedly thinking “FAIR is limited to the EU and the life sciences – why should I care?” (Higman et al., 2019). In other words, both the lack of participation and the unwanted participation were deemed to stem from “misconceptions”. Jacobsen et al. (2020) put it this way:

> As support for the Principles has spread, so has the breadth of these interpretations. In observing this creeping spread of interpretation, several of the original authors felt it was now appropriate to revisit the Principles, **to clarify both what FAIRness is, and is not**[73].

[72] Data from clarivate
[73] Emphasis added

Over the years after 2016, several articles "explaining" what FAIR was really about became published. The focus of these papers were on how it supposedly would benefit everyone if they only would understand FAIR properly, making it clear that there were wrong ways and right ways of interpreting FAIR (Mons et al., 2017; 2020; Jacobsen et al., 2020).

The perceived need to make sure people were doing FAIR the "correct" way, led the proponents of FAIR to organize a gatekeeping authority through the GOFAIR network (Mons, 2021). As Barend Mons put it at a German material science conference:

> We [the GOFAIR community] will make sure people don't say 'We are FAIR because we are interoperable within our own silo'. [But] this little group can't develop this in all countries. You would need a team in Germany to make sure German communities are following FAIR (Mons, 2021).

A large part of the GOFAIR network's effort to make sure everyone had the same idea about FAIR, was through their so called "FAIR Convergence Symposiums" (GOFAIR, 2020a; 2022). There, a number of FAIR proponents and supporters got together to discuss the development of FAIR. Notably, despite the initial struggles of the FAIR proponents, the first keynote of the 2020 installment had a presentation by European Commission bureaucrat Jean-Claude Burgelman claiming that "[the Internet of FAIR Data and Services] will happen – for sure".

During the same symposium, Ana Persic from UNESCO took a more critical stance, saying that if FAIR were to succeed, it needed to be aware of the divide between the global South and North (CODATA, 2021, 59:15). As a response, Barend Mons was quick to remark that even though "the catchy acronym, FAIR" had "an ethical ring" to it, this was not something the principles were concerned with. In his words, "the connotations of FAIR [were] misleading" (CODATA, 2021, 1:15:00). Similar interactions happened throughout the symposium, where someone would voice their concerns, with one of the FAIR proponents shortly after rejecting their concerns as products of misinterpretation of the intentions behind FAIR.

This pattern of dismissive behavior was not just a quirk of personality, but grounded in the conceptual work behind the convergence symposiums themselves. In 2018, FAIR proponents Peter Wittenburg and George Strawn had published a paper formally introducing the concept of convergence to the FAIR movement. Drawing heavily on technology scholar Thomas Hughes seminal *Networks of Power* (1983), they argued that a period of divergence (or creolization, as they called it) was completely normal for "revolutionary infrastructures", but that convergence around a semantic web for science would happen. After all, they argued, that was the case for both electrical infrastructures and The Internet – so why not for the Internet of FAIR Data and Services then? (Wittenburg and Strawn, 2018). It just needed some time.

But the development sketched out in the convergence paper (Wittenburg and Strawn, 2018) didn't materialize in the way the authors imagined. Three years later, Wittenburg, Strawn, and a handful of other convergence supporters came forward to explain: "Some recent observations and developments [...] made clear that convergence enabling a vigorous data exploitation phase might be further away than was hoped" (Jeffery et al., 2021). The FAIR proponents were not giving up though. According to Wittenburg and Strawn (2021a), the convergence would happen, just not very soon. Everything was on track, it was just that "revolutions take time" (Wittenburg and Strawn, 2021a). What remained was to get some social and political issues out of the way (Wittenburg and Strawn, 2021a). As for the researchers that hadn't adopted FAIR yet, it was an issue of "persuading [them] to adopt best practices[, by] decreasing the cost[,] and increasing the benefit" (Jeffery et al., 2021). Thus, while the horizon was pushed a bit further away, the strategy remained largely the same.

The problem of change

But who are we to criticize the planning and execution of FAIRification? It's not often that change goes according to plan. In their seminal paper, Rittel and Webber describe planning as a "wicked problem", that is, a problem which is "terribly difficult, if not impossible" to resolve (1973, pp. 157). Planners, they claim, tend to think of change as a "tame problem" which can be solved through expert knowledge and analysis (Rittel and Webber, 1973). But on the contrary, Rittel and Webber argue, approaching wicked problems in this manner is likely to spur new ones (1973). Like the hydra of Greek mythology, for each problem solved, more seem to appear.

A host of other scholars have made similar points, that thinking of planning as something that should be done by a centralized (see: detached) body is setting yourself up for failure (see Friedman and Friedman, 1980; Mintzberg, 1994; Rip, 2010). The arguments against this top-down governance differ, but what they have in common is the basic notion that societies – from nations to organizations – are too complex for any actor or group of actors to sufficiently grasp, much less control. Yet, change happens all the time.

Locating agency

But if agency is not with governing bodies, where is it then? The answer to this question tends to vary, and is largely what sets the different theorists apart. Here we can put a sliding scale from hard determinism – where technology, ideas or markets act as unstoppable forces – to chaos theory, where any small stirring can emerge and become influential. With that being said, most scholars exist somewhere in the middle, agreeing that no one thing determines everything, but that some things influence the development of certain events more than others. Some popular concepts range from the simple overton window (Mackinac Center, 2010) to the more complicated multi-level perspective (Geels, 2002).

But the concepts I will be drawing on are more related to what John Law (1994) calls "a modest sociology". As opposed to much of the aforementioned literature on change, Law argues, a modest sociology "[...]will not distinguish, before it starts, between those that drive and those that are driven." (1994, pp. 13). This does not mean that change can come from anywhere, but that, analytically, we should not assume that it can't. And without those assumptions, we're only left with what we can know and observe, meaning that anyone aiming to learn anything about change, should do so through empirical study, not conceptual abstractions (Law, 2014, pp. 34).

Both the fields of Science Studies and Technology Studies have histories that are rich in inquiries into change within their respective fields of study. In technology studies, scholars like Thomas Hughes (1983), Trevor Pinch and Wiebe Bijker (1984) were trying to figure out how technologies were developed by their designers, but also how they would become implemented into or accepted by societies. In science studies, scholars like Latour and Woolgar (1979), Knorr-Cetina (1981), Brian Wynne (1991), David Bloor (1976), Steven Shapin and Simon Schaffer (1985) were asking similar questions, but about scientific knowledge (Bloor, 1999). As a common denominator, they found that the development and implementation of both scientific knowledge and technology was not determined by some intrinsic trajectory or force of its own. Neither did they depend singularly on the mind and labor of a lonely genius or entrepreneur. Instead, scientific, technological, and societal change more broadly, seemed to be products of a host of different factors that were neither purely social nor purely technical in nature (Pinch and Bijker, 1984). The challenge of locating the sources of action and agency famously led to debates within Science and Technology Studies on whether or not objects (scallops being one of the most known examples) should be considered actors as well (see Jones, 1996). Regardless, scholars on both sides of the debate agreed that change was a complex and sociotechnically contingent affair.

Locating experts

How does one then plan, govern and design change after acknowledging its sociotechnical complexity and contingency? With growing concerns that science and technology might not only produce positive outcomes, but also risks (Beck, 1992; Lash et al., 1995) or other unwanted outcomes (Latour, 1993; Irwin and Wynne, 1996), scholars started thinking about how to answer this question (Elzen et al., 1996).

The general consensus was that the general public needed to be included. This could be seen as both a normative and a pragmatic position. In terms of normativity, inclusion and participation seemed the most democratic approach, and in terms of pragmatism, inclusion and participation was expected to lower the risk that innovations in both science and technology would be rejected by the public and thereby failing.

But the methods of inclusion and participation were not always democratic and didn't always lower the risk of rejection. Several critical analyses of change processes showed that among people promoting change, be it in science or technology, there was a common assumption that the public/users would reject sociotechnical change, not because they had a good reason to, but because they did not understand it (Woolgar, 1990; Wynne, 1991). According to this understanding, also called the "deficit model" (Wynne, 1991), "experts" who developed the knowledge or technology knew best. Any lack of success would therefore be explained as caused by a deficit of knowledge among the "recipients" of the innovation. Strategies for dealing with this deficient public was thus based on informing what was considered "lay people" about the correct knowledge (Wynne, 1992; 1996) or disciplining them to develop correct behavior (Woolgar, 1990).

Scholars of science and technology studies would counter this deficit model, claiming that lay people didn't necessarily need any education or discipline. Quite the opposite, they were deemed experts in their local context, and sometimes would best know how to adapt these innovations accordingly (Epstein, 1995; Wynne, 1996; Sørensen, 1996; Aune, 1996; Laet and Mol, 2000). The distinction between "lay" and "expert" was thus deemed performative and situational.

Opening up

The redefinition of the public from a barrier to be overcome, to a resource to be utilized meant that instead of putting effort into convincing people to adopt change, they could and should be included in the process of planning and design (Elzen et al., 1996). It is important to note that the inclusion of "others" into the process of knowledge production, design or policymaking was not new. The novelty came in the form of scholarly and public attention to it – making it a trend, so to speak (Bjerknes and Bratteteig, 1995).

Scholars within the field of Science and Technology Studies followed up by conducting studies of how participatory processes would, could, and sometimes should, be done. Notable contributions from Science and Technology Studies were Latour's (1993) parliament of things, Callon's hybrid forums (1998), and Nowotny et al.'s (2001) agora. While they all insisted on the participation of heterogeneous actors, Latour (1993) and Callon (1998) extended this invitation to non-humans as well.

The inclusion of participants without a voice of their own naturally raised the question and concern of who is able and allowed to speak on behalf of them[74]. Callon et al.'s (2009) argument is, simply put, that specialists – those most intimately familiar with the relevant phenomena – should be invited to represent them in the hybrid forum. Importantly – drawing on Wynne (1996) – Callon argued that the position of specialist did not need to be filled by a scientist, but could be appropriated by what in other terms would be called a lay-person (Callon et al., 2009). This approach, though, would not delete the danger of misrepresentation. For, Callon argues, representation is a sort of paradox – while its aim is to give actors a voice, it can only do so by silencing them (Callon et al., 2009). Put differently: when someone speaks on another's behalf, the other is not allowed to speak.

Boxing it in

Callon (1998) argued that the selection of relevant concerns, actors and objects was important for the success and long term stability of any process of change. The framing, as he called it[75], of an issue was necessary to define what a process was about, but also to define who would participate in it. Conversely, framing would also define what a process was *not* about, and who were *not* allowed participation.

An inherent quality of framing is its inability to encompass all things in the world, and sometimes even the most immediately relevant ones. This, Callon argued, posed a problem for governance (1998). Anything out of frame, so called externalities[76], could at any point present themselves as overflow, forcing themselves into the process with varying severity, from manageable leaks to chaotic breaches (Callon, 1998). In practice, these overflows would mean that previously excluded actors[77] would enter the frame without the framer's deliberate will or intent.

[74] See Bloor (1999) and Latour (1999) for a debate on this topic.
[75] Drawing heavily on Erwing Goffman's frame analysis (1971)
[76] Here, Callon takes from economists market-analysis
[77] In Callonian terms, the term actor would include both humans and non-humans (see: actants).

In spite of growing attention and interest, the participatory turn in governance and decisionmaking didn't always lead to any substantive inclusion (Wynne, 2007; Stirling, 2008). In an attempt to explain the continued prevalence of "persistently deterministic notions", Andy Stirling (2008) made the distinction between two forms of governance participation: appraisal and commitment. *Appraisal* denoted any action that could contribute to informing governance, but without a guarantee of influence. Real power, Stirling (2008) argued, lies in *commitment* – any action that builds towards a particular future – intentional or otherwise.

The asymmetry between the two forms of action, raises questions about participation in decisionmaking processes: "Are they about informing, or actually forming, the commitments themselves?" (Stirling, 2008). Without commitment, participation may be little more than a "technology of legitimation" (Glasner, 2001) for pre-decided commitments. Governance processes that don't commit to appraisals, scholars argue, are therefore at best pointless, and at worst harmful (Stirling, 2008; Owen et al., 2012).

Making a frame and sticking with it

With this in mind, we'll go back to the case of this chapter, interrogate the actions of the FAIR proponents and sketch out the characteristics of their strategies. In order to do this, we will focus the analysis around the concepts of frame, externalities and maintenance.

Frame

Right off the bat, the FAIR principles introduce a striking tension between inclusion and exclusion. The initial FAIRport conference had only a handful of participants, composed entirely of people with ties to bioinformatics. Yet it was later presented as diverse (FAIRPORT, 2014; Wilkinson et al., 2016). The 2016 paper (Wilkinson et al.) was authored by a similarly composed group, and was similarly presented as "a diverse set of stakeholders" – in the second sentence of the paper, no less (Wilkinson et al., 2016).

From the beginning, the promotion of the FAIR principles have contained this duality, this oscillation between homogeneous and diverse, top-down and bottom-up, special and general. In this analytical chapter, we'll take a closer look at how this dynamic has been constructed by the proponents of the FAIR principles. We start by looking at the very initial framing of FAIRification.

In 2014, when Barend Mons invited 30 colleagues to the Lorentz conference center, the expressed intention was not to create a global omnidisciplinary research infrastructure, but rather "to discuss how they could create an environment that enables effective sharing and re-use of biological research data" (International Innovation, 2014, emphasis added). In order to do this, they defined a set of concerns, namely "the context, user needs, technological challenges, access & business models, funding requirements and governance of such a global initiative", and invited people who they considered to be able to cover these concerns. Looking at the list of participants (FAIRPORT, 2014), there's a reasonable coverage of these concerns, with people from the fields of life sciences and informatics, including publishing[78] and other research infrastructures, both private and public[79]. Several of them also had experience with similar initiatives[80] or work within the European policy sphere[81]. With this set of representatives, the conference managed to produce the "remarkable consensus" that the FAIRport concept was the future of life science research in "the e-science era" (FAIRPORT, 2014).

This initial definition of concerns and inclusion of representatives managed to establish a relatively clear frame for the envisioned change process, and importantly, a reasonable match between concerns and representatives (Callon, 1998). Putting it coarsely, they were a group of bio-informaticians that wanted to change bio-informatics.

But at some point something happened that misaligned the concerns and representatives. "I don't know when that was or whether it was a seminal moment or whether this was something that did grow.", one of the conference participants, Tom told me, but somewhere along the line they decided to expand their framing to encompass other disciplines as well. Here's Tom again, explaining why:

> What we notice is that a lot of the work that we are doing is applicable outside [of our field]. It's very dangerous to say that you know how things work in other fields, but you can at some point say this cannot be specific for the life sciences. There must be more generic things.

Using their own experience with research data, they gathered that other fields must have had similar concerns. The way they justified this decision in the report was that "the wildly variable character of life sciences data, more so than their sheer volume are a major challenge, putting Life Sciences in an 'exemplar' situation for eScience data needs" (FAIRPORT, 2014). In this way, the conference participants expanded the concerns of their initiative, while keeping the representatives the same. They could do this because they considered themselves capable of representing "all eScience approaches, also outside of the Life Sciences" (FAIRPORT, 2014).

---

[78] Traditional publishers like Elsevier and Springer Nature, but also smaller initiatives like ScieELO
[79] Notably ELIXIR, BBMRI, The Dutch Techcentre for Life Sciences and The Internet/W3C
[80] Like the Budapest Open Access Initiative, the Human Variome Project, WikiProteins
[81] Barend Mons, notably worked in the European Commission before his work on the FAIR principles

Because of their perceived ability to represent the concerns of other actors, their outreach to other actors, like the FORCE11 community or RDA, was not motivated by a wish to expand their representative capacities, but to expand their base of supporters. This motivation was explicitly made clear in the reports from the subsequent FAIRPORT meetings (Data FAIRport, 2014a; 2014b; 2014c).

As we've seen, the attempt at expanding the influence of FAIR later also included inviting others to contribute to the formation and formulation of the principles themselves:

> We call on all data producers and publishers to examine and implement these principles, and actively participate with the FAIR initiative by joining the Force11 working group. By working together towards shared, common goals, the valuable data produced by our community will gradually achieve the critical goals of FAIRness (Wilkinson et al., 2016).

But as we've seen, the appraisals that did come in were not necessarily welcomed by the original proponents, and therefore not turned into commitments.

Externalities

With the initial framing of the FAIR principles set, we can start to take account of the different externalities that presented themselves after FAIR went public. The very first came from Christine Borgman when she took stage on the 4th RDA plenary in Amsterdam in September 2014. There, she publicly voiced her concerns with the idea behind the FAIR principles, suggesting they were not as generalizable as Barend Mons might have thought, and that this in turn could produce problems later on (Borgman, 2014).

In the opening statement of her talk, Borgman expressed a desire to contribute to the formation of a system for data sharing, but with reservations and a warning:

> "This is the critical time to be building this new infrastructure. The challenge is that we need to build an infrastructure around the way that the scholars and the research community *really* work, not the way we would like them to work, not the way we think they ought to work, but the way they actually *do* work." (Borgman, 2014, 22:56-23:16).

Although this was a critique of Barend Mons' governance strategy – which in Christine's opinion based itself off of narrowly informed assumptions and ideals of researchers' practices – it could also be seen as an appraisal of and wish to (in)form (Striling, 2008) Barend Mons' governance strategy. Differently put, she claimed FAIR *could* be bad because it didn't consider "actual" practices, but it *could* be good if it did. In this sense, Christine was acting as a representative of a concern that she thought hadn't been considered by the FAIR proponents. As she put it: "Genomics found ways to work for itself, but [...]Genomics is the exception". Outside perspectives were needed.

Using Stirling's (2008) lens, we can take a look at a couple more instances of what we could call appraisals during the timespan of FAIR development. For the purpose of this chapter, we will look at the following: the European Commission, the papers citing Wilkinson et al. (2016), and the convergence papers (Wittenburg and Strawn, 2018; 2021a; Jeffery et al., 2021).

The European Commission's embrace of the FAIR principles as the central concept to their envisioned European Open Science Cloud (EOSC) had two important impacts. First, as discussed, it gave FAIR a considerable added momentum. Second, it also meant that the FAIR principles became strongly associated with the European Commission, their research strategy, and their Open Science Cloud[82]. In the commission, FAIR was seen as an extension to their Open Science strategy (Burgelman, 2021), a *way of seeing* that was adopted by a significant portion of the larger public. In other words, making the FAIR principles a part of the EC's Open Science strategy effectively pulled the Open Science concern into frame of the FAIR principles.

The concern for open sharing could also be seen among the scholarly papers citing FAIR – like Alquraishi and Sorger (2016) – but this group of people also introduced other (often more tacit) concerns. For instance, a paper describing the genomics database *Ensembl*, claimed that the database had been FAIR "long before these [principles] were formalized" (Zerbino et al., 2017). According to the explicit mentions in the paper, this meant making data "freely and programmatically available", defining "globally unique and persistent identifiers", following "international naming standards" and so on. But in addition, claiming to be FAIR for a long time already, didn't only imply that FAIR was what they were *saying*, but also what they were *doing*. The same went for publications tied to other databases like the Protein Data Bank (wwPDB, 2019), BBMRI (ter Horst, 2016), or the Gene Ontology (Gene Ontology Consortium, 2020), to mention a few. And in a similar vein to the EC's adoption of FAIR, these publications not only added momentum, but also created associations of what it meant to be FAIR – ever so slightly introducing themselves and their concerns into frame, causing externalities to start leaking into the frame put around FAIR.

[82] As made clear by Higman et al. (2019)

Maintenance

According to Callon (1998), when externalities force themselves into frame, they can cause uncontrolled spillovers, breaches or overflows. These, in turn, threaten to cause the governance of the given process to fail. Under the assumption that this is generally undesirable by the framing actors, they will generally set out to manage these externalities by expanding or maintaining the frame – opening up or closing down in Stirling's (2008) terms. In the case of the FAIR principles we see a clear tendency towards the second option, maintaining frame and closing down.

When Christine Borgman presented her assessment of FAIR at the 4th RDA plenary, the reaction from Barend Mons was not to use her comments to (in)form the shaping of FAIR. Rather, he suggested, she might have misunderstood what he was talking about. So instead of taking Christine's concerns about the principles' generalizability into frame, Barend Mons kept them out. Labeling her assessment as a product of misunderstanding, Barend Mons effectively reinforced the original framing of the principles, maintained his role as legitimate representative, and silenced the appraisal (Callon, 2009) – effectively closing down the framing of FAIR (Stirling, 2008).

In later instances we see a similar dynamic: that when actors submitted "unwanted" appraisals, the proponents of FAIR would choose to delegitimize them and keep them out of frame. This is evident in the case of the European Commission's involvement with FAIR, where the attempt from the EC to make FAIR a part of their Open Science strategy got flipped on its head, making the EC's strategy a part of the FAIR movement. Thus, the EC's tacit claim to FAIR as corollary to Open Science was resisted and rejected.

Later, the papers "explaining" what FAIR was supposed to mean (Mons et al., 2017; 2020; Jacobsen et al., 2020) acted as a form of maintenance work on the original framing of the principles. Externalities imposing themselves were defined as as "misconceptions" (Higman et al., 2019) or "interpretations" (Jacobsen et al., 2020), stemming from "confusion" (Mons et al., 2020), a lack of knowledge (EC, 2020b), lack of understanding (David et al., 2020; EC, 2022), or from not having read the principles at all (Mons, 2020).

Suggesting that FAIR should be otherwise, or that FAIR was not as universally applicable as its proponents claimed, was routinely responded to with a deficiency narrative. Rebuttals of deviations often framed themselves as *explanations* rather than arguments (Mons et al., 2020). If someone claimed to support FAIR, but in a way that was deemed wrong by the original authors, the *explanation* was that they either didn't understand it, or had some interest in undermining the FAIR movement (Mons, 2020). If someone didn't support FAIR at all, it could be explained by their lack of awareness or understanding (David et al., 2020; EC, 2020a; Khodiyar et al, 2021; EC, 2022). In other words, those who didn't subscribe to FAIR as proposed by its proponents, were presented as if suffering from some cognitive, perceptive or moral deficiency.

This explanatory framework left no room for legitimate participation other than support that was aligned with the original FAIR. In this way, only actors representing already framed concerns were allowed to be included. This gatekept and limited inclusion thus gave an impression of participatory governance, while in reality creating a quite high threshold for qualification by effectively labeling dissenting actors as a form of lay-people.

Even when FAIR was threatened with stagnation – an increase in supporters, but lack of practitioners – its proponents held fast on their strategy. Wittenburg and Strawn's (2018) paper on convergence can in this light be seen as an attempt to legitimize the strict framing around FAIR. Diagnosing the problem as a lack of convergence, explaining it as a natural part of the development of "revolutionary infrastructures", the lack of "qualified participation" was not seen as a sign that something needed to change, but rather, that more time was needed for the strategy to work.

This explanatory framework helped to bolster the existing framing of FAIR. Externalities challenging the integrity of the original framing were effectively kept out by defining them as products of deficiency and/or part of the natural developmental trajectory of "revolutionary infrastructures" (Wittenburg and Strawn, 2018). This notion of convergence was then later used by the GOFAIR association to gatekeep and legitimize the original framing of FAIR and justify the decision to hold on to it through the convergence symposiums and other activities (GOFAIR, 2020a; 2022; Mons, 2021).

Origins for a strategy of change

To summarize, there are three central elements to the governance strategy of the FAIR proponents:

1. a relatively narrow initial framing of concerns,
2. inviting others to participate in the framing, and
3. strict gatekeeping of participation through deficiency narratives

Looking at it this way, the internal contradiction becomes strikingly apparent. The sustained call for participation and refusal to accept diverging participation introduced a tension that was hard to resolve. Should they stick with their narrower frame and stay within bioinformatics, or allow participation in order to expand it to other fields? Throughout the development of FAIR, it seems like the concept's proponents chose to do a little bit of both: sticking with the narrow frame *and* expanding to other fields.

But we should resist the temptation to attribute this paradoxical decision to some form of stupidity, illiteracy or cognitive deficiency. Rather, we should ask: what could be the reason for this peculiar strategy?

When we look back at the inception of the FAIR principles, there's one step in their development that catches our attention: the transition from data intensive life sciences to virtually every scientific discipline, so that's where we'll start.

Big dreams in a small group

In order for a framing to be successful – in the sense that it accomplishes the change it sets out to make – there needs to be a certain alignment of the frame and what it sets out to represent (Callon, 1998). In the case of FAIR, the initial framing was, if not perfect, then at least decent. The group of actors that had convened in Leiden represented the field they wanted to change – making an initiative by bioinformaticians, for bioinformatics. Of course, this would not have been without its challenges either – as indicated by the deviance of the bioinformaticians citing the first paper (see Alquraishi and Sorger, 2016; Zerbino et al., 2017; Buniello et al., 2018) – but by expanding their goal, the FAIRport conveners made their task significantly more challenging – making an initiative by bioinformaticians, for bioinformatics *and* everyone else.

Now the question is why would they want to do such a thing? Why not just stay with the original framing of the initiative? Why bother to expand it beyond its initial boundaries?

Part of the answer can be found in their strategy to gain influence through leveraging the position of powerful actors – trying "to make all 'enablers' part of this initiative" (FAIRPORT, 2014, p. 4). When we look at the list of organizations and institutions the FAIRport members of the FAIRport convention reached out to, we mainly find actors with similar interests and concerns, like FORCE11, Elixir and the RDA – but we also find actors with a wider scope, most notably the European Commision and the Global Research Council (FAIRPORT, 2014, Data FAIRport, 2014a; 2014b).

Committing to these actors necessitated bringing their concerns into frame, and meant that FAIR *had to* expand from life sciences to everything else. The European Commission was looking for an extension to their Open Science strategy, a strategy that was already being applied across all domains of research in the EU (Burgelman, 2021). FAIR could therefore not be an add-on just for life sciences – a European Open Life Science Cloud, so to speak. FAIR needed to be transdisciplinary. In addition to this, the FAIRport participants' aim to get the endorsement of the Global Research Council meant the scale also could not remain European (or western for that matter), but needed to strive towards "global policy" (Data FAIRport, 2014b).

But the FAIR proponents' commitment to wider-framed actors does not explain their scope expansion entirely. After all, even as the group failed to convince the Global Research Council, and after their US embedding through the NIH's BD2K faltered, they still retained their global scope[83]. Additionally, there must have been something that made the FAIR proponents believe that this expansion was feasible. This brings us to the next part of the explanation.

Undeterred determination

One of the reasons for the FAIR proponents' persistence in promoting their principles was their strong belief in themselves as representatives for others, and in the inherent momentum of the concept they had developed.

One of the central beliefs for allowing the expansion of FAIR from life sciences to the rest of the world, was that the participants at the conference seemingly believed that their experiences within bioinformatics qualified them to represent the concerns of actors outside their fields, effectively silencing those they believed to represent (Callon et al., 2009). They conceded that things were different in other places, but that, in general, data was data, and life sciences was an exemplary case for data needs more broadly (FAIRPORT, 2014).

This belief in their ability to represent other actors' and groups' concerns was also combined with what we can call a technological determinism. As Stirling (2008) noted, in spite of the growing popularity of participatory governance strategies, and strong critique against it, determinism has remained prevalent in processes of governance. Subscribing to deterministic notions effectively deletes the agency of actors, because it posits that agency is something that belongs to some external factor – the inherent logic or force of "revolutionary infrastructures" (Wittenburg and Strawn, 2018), for example.

In the case of FAIR, the traces of determinism can be found in their belief that The Internet would evolve from the current web into the future "semantic web" (Mons, 2005; Wittenburg and Strawn, 2018). "The goal", as Barend Mons put it, was to "create a new internet, but now for machines" (Mons, 2019). In their eyes, the world was on a trajectory towards this future.

[83] In spite of the EC's wish to keep the EOSC tied to the EU and the "European Data Space"

Anyone disagreeing, were in the proponents' eyes simply unable to see the tracks that they were moving on.

This becomes most evident in Peter Wittenburg and George Strawn's paper, *Common patterns in Revolutionary Infrastructures and Data,* from 2018. And even in subsequent papers revisiting their claims, holding onto the idea that FAIR would prevail – it just needed some more time (Wittenburg and Strawn, 2021a; Jeffery et al., 2021; Strawn, 2021). Notably, they posited, the technological barriers were already overcome – what remained was social issues (Wittenburg and Strawn, 2021a; Jeffery et al., 2021; Strawn, 2021). In other words, the future was already here, people only needed to accept it. As Jean-Claude Burgelman put it during the first convergence symposium: "[the Internet of FAIR Data and Services] will happen – for sure".

In this way, the actors rejecting or diverging from FAIR were granted some agency, but only to slow progress down. On the train tracks leading to the semantic future, actors could put branches in the way, pull the brakes, maybe even blow up a bridge, so to speak, but according to the FAIR proponents' view, they could not alter the direction. Any appraisals from actors seeking to change the course of FAIR thereby had to be rejected.

So why this performance of welcoming inclusion and invitation to participation?

Closing the gate

It is peculiar, this performance that keeps repeating, that FAIR needs to be built bottom up through involvement from the communities (Mons, 2014; Wilkinson et al., 2016; EC, 2016; 2018; 2020a; 2022; Mons et al., 2020). Because as we've seen, when the "bottom" and representatives from "the communities" tried to get involved, their appraisals were rejected.

It can seem like the participatory language works as a form of "technology of legitimation" as Glasner (2001) describes it. But when the proponents of FAIR talked about the participatory nature of their governance process, they typically didn't pretend to be more inclusive than they actually were. When they claimed to represent a "diverse set of stakeholders" in the 2016 paper (Wilkinson et al.), they may have been exaggerating their diversity, but they hadn't invited an actually diverse group of stakeholders and rejected their appraisals. In fact, the FAIR proponents were quite open about their rejection of appraisals – repeatedly and quite publicly denying their concerns in articles and talks (Mons et al.,2017; 2020, Jacobsen, 2020; Mons, 2014; 2018; 2020; 2021). The legitimation of FAIR was not performed through some pretended inclusivity, but – as discussed above – through the proponents' presentation of themselves as adequate representatives in a change process with a predetermined outcome. Since FAIR was seen as self-evident, its proponents identified more as apostles or missionaries than salesmen or negotiators. The dissenting researchers that rejected or challenged FAIR were thereby labeled a lay public, deficient in the understanding of revolutionary infrastructures, and in need of information or education. Participation in the continued design and development of FAIR was therefore not identified as the necessary strategy to implement it.

What motivated the calls for bottom-up process and community involvement was not a need for legitimacy, it was, rather, the need to realize the determined trajectory of FAIR. With no users, there would be no one to provide FAIR data or services for. And with no FAIR data or services, there would be no one interested in using them. The calls for participation in FAIR were, in other words, not intended as invitations to develop FAIR, but invitations to supply FAIR with data, services and users in order to realize an already determined outcome. Contributions outside of this scope were therefore not considered interesting or even welcome.

As FAIR continues to spread into science policies, grant proposals and the lives and practices of ever more researchers, externalities are going to grow and exert more pressure on the framing around FAIR. As we continue this thesis, we will see different actors trying to form and inform the development of FAIR, making appraisals, and encountering resistance. The question is whether the frame around FAIR can withstand the pressure, or if externalities will overflow. Change will happen, but who will be the agents behind it, and what strategies will they employ?

Chapter 4: Stewarding data, implementing FAIR?

So far in this thesis, we've been looking backwards with a historian's gaze, ruffling through documents, rummaging archives and scouring recordings to figure out how this all started – how FAIR was conceived, turned into policy and introduced to the world of science. So far, we've seen how proponents of FAIR tried to grapple with complex situations across domains, trying to make a clear case for how science should be governed. But, so far, we've seen little of the people who the FAIR principles concerned perhaps the most – the people "on the ground", so to speak. But no story about research governance is complete without the governed researchers. So, four years ago, I set out to find out how researchers related to FAIR and to learn what FAIR looked like in practice.

Looking for research subjects, I quickly realized, would be harder than I had anticipated. While some of the researchers I reached out to had heard about the FAIR principles, or at the very least found them familiar-sounding, nobody could really tell me exactly what they were – much less explain how they themselves practiced FAIR research.

But one day, after years of searching[84], I lucked out. A soil microbiologist whom I'd talked to before, Tamara, reached out to me and said "You know what? We had a meeting about FAIR at our lab the other day!". Finally, I thought, I'd get some answers.

"So what did they say about FAIR?", I asked. Well, she said, first of all, it had something to do with proper research data management. Second, it was something that research funders would require. Third, the letters in the acronym stood for Findable, Accessible, Interoperable and Reusable (or was it Reproducible?). Generally, FAIR, she said, was about how the lab could ensure that their data was properly managed – storing it in the correct locations, labeling it so it could be understood easily, and sharing the code they used to analyze the data. Basically, the way she imagined it, FAIR was just another word for good research practice. So when I asked whether there had been any discussion around machine-readability, access licensing or semantic vocabularies, the puzzled look on her face was answer enough.

How come researchers' perception and ways of practicing FAIR differed so much from what its proponents had imagined? Where did the researchers get this impression from?

[84] admittedly, only two years, but still plural.

I started looking for answers, and after an almost embarrassing amount of time, it dawned on me that the researchers' knowledge about FAIR rarely ever came to them directly through the official channels of the European Commission or from other proponents of FAIR. Rather, it was mediated to them through a third party – a middle actor.

When it came to who these third parties were, there was naturally a substantial variation. A typical first meeting with FAIR for many of the researchers I talked to was through some general mention in another researcher's presentation, much along the lines of "in this project we have followed the FAIR principles…". While these cases did put the acronym into the audience's consciousness, the mentions of FAIR could be seen as more of a gesture than an actual introduction. The more thorough and formal introductions to FAIR were increasingly coming from a type of academic professional called a data steward. In my effort to explain the discrepancy between FAIR principles and practice, I therefore chose to continue my interrogation there: with the data stewards.

Data stewards were far from common back then, and chances are that when you're reading this, they still aren't – but that depends on your definition. We'll get back to this later – the definition of a data steward – but when I started looking into them back in the early 2020's, a data steward was inextricably linked to efforts at making research FAIR. But if a data steward's responsibility was to make research FAIR, then, judging by the responses from the researchers I talked to, the data stewards seemed to be doing a pretty poor job. That being said, I didn't assume they were just sitting on their butts all day[85], which led me to the question: if the data stewards weren't making research FAIR, then what were they doing?

Answering this question started, as with all good research, by answering another question first:

What is a data steward?

While the question seems straightforward enough, the answer – as often is – is far from it. That being said, I will try to make it as brief as possible in this chapter. I will therefore have you note that although the term data steward(ship) has been used prior to the introduction of the FAIR principles, I will focus on data stewardship in its relation(ship) to FAIR.

In the first official publication of the FAIR principles in *Nature Scientific Data* (Wilkinson et al., 2016) the notion of data stewardship was a central element, present clearly in the title of the paper: *The FAIR Guiding Principles for scientific <u>data management and stewardship</u>*. In spite

[85] Although, chances are that, like the rest of us, they did spend a lot of time sitting (on their butts) in front of a computer

of this, the paper engaged little with the data steward as a role, and mostly with data stewardship as a practice. In other words, stewardship in the paper was presented as something to be done, but it didn't say much about *who* would do it.

So what was data stewardship according to Wilkinson et al. (2016)? In the authors' definition, they started by differentiating stewardship from management, stating that

> data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data (Wilkinson et al., 2016).

Data management on the other hand, the authors implied, had only a concern for the immediate and short-term. Further, they underlined that machine-readability should become "a first-priority for good data stewardship", and how the 15 FAIR guiding principles could be used to inform this practice. According to the paper, a data steward's ultimate purpose would be "supporting knowledge discovery and innovation" (Wilkinson et al., 2016).

In the years after the first publication, the data steward transitioned from being someone *supported by* the FAIR principles, to being seen as a key role in *realizing* FAIR (see EC, 2016; EC, 2018). In most publications from that time on, the data steward was presented as some sort of link between the then current (unFAIR[86]) practices and future (FAIR) practices (EC, 2018; David et al., 2020). Throughout the years after 2016, the role of the data steward flipped from being a product of FAIR to being a producer of FAIR, to put it crudely.

In the publications after Wilkinson et al. (2016), as before, the data steward continued to be (implicitly) defined simply as someone who stewarded data – while data steward*ship* itself was defined as any practice that realized FAIR science. The empirical examples of data stewardship in these texts notably tended to point towards roles with other titles, such as (data) librarian, repository manager, or data curator – or towards the practices of institutions like data repositories and databases (Whyte et al., 2019; David et al., 2020; Palmer and Cragin, 2022). In these texts, the data steward in itself thus remained a mostly conceptual entity, a theoretical precondition for FAIR science, like a research data management higgs boson, or an elusive missing link between contemporary science and e-science.

But the elusive nature of the data steward did not stop it from becoming established as an important component of European research data policy (EC, 2022). In response to the policy push from the EU (EC, 2017) and national bodies (FWF, 2019; RCN, 2019; NWOn n.d.), institutions – like universities – gradually started opening positions for data stewards around the year 2020. Up until this period, studying data stewards and stewardship would have had

[86] Or "reuseless", as is the preferred term of Barend Mons (Outsell, 2023).

to be an analysis of expectations rather than practices. But because of developments in the early 2020's that we won't get into right now, data stewards started emerging in different academic institutions – if not plentiful, then at least to an extent that created an empirical basis for beginning to answer the question: "what is a data steward?".

So since the start of my research project in 2020, I've talked to several people who either called themselves data stewards, or held formal positions as such. In line with the literature on the subject, these data stewards inhabited different institutional spaces, roles and identities. Two of them were located in their university's library, both physically and organizationally. Four others were physically located at their respective faculties as data stewards, while having their position's funding coming from different pots in the university system. The two last stewards had their positions assigned to data repositories, but had their institutional ties to a university and were physically located at their university library[87]. Out of the eight, only four bore the formal title of data steward[88].

With this significant variation and multiplicity in mind, there's a need to establish some structure to what I will talk about in this chapter. Because, I believe there is something to be said about these people, and, specifically, their roles as data stewards. Throughout my research it became evident that these people shared some commonalities, not only in terms of their job descriptions, but also in their strategies for carrying out their jobs. That being said, their abilities to carry out these strategies were not the same. Therefore, I have structured this section in a way that is focused on the commonalities first and the differences second.

Governed by uncertainty

"It's not really a strongly defined job title".

I was talking to one of my informants, Maria. She was one of the people I talked to that most clearly identified as a data steward. Not only did she carry the title of one, but she had also gone through a training program for data stewards at her University. She was, in some sense, a poster child for data stewardship. Yet, with all this formality, she struggled to understand exactly what her job was.

When I asked her what being a data steward entailed, she produced the above quote, and said that her employer had told her: "you can always find out". Perhaps in an attempt to do so, she returned the question back to me: what is a data steward? I couldn't tell her much

[87] See appendix for overview

[88] I have chosen to categorize them as data stewards for three reasons: 1) because the work they do to a lesser or larger degree overlaps with the theoretical prescriptions, 2) because they partially self-identify as data stewards, and 3) because it enables me to capture the multiplicity of this role

else than that it seemed to be an open question with many answers. But I had talked to people like her before – people who bore the title, or in some way or other considered themselves data stewards. I told her that there were a lot of differences between the people I talked to, but that she was not alone in her doubt. In fact, most of the stewards I talked to had brought up a similar concern.

One steward, employed at faculty level, Alex, told me about the process of being hired to the position of data steward, and how their boss made it evident that they did not have a clear image of what a data steward would do, and that it would be up to Alex to figure it out on their own: “I basically fill this position with what I want”, they said.

Another steward, Andreas, told me that he was pretty sure the announcement for the job he applied for said “research data steward”, but other than that, he didn’t know exactly what that meant. Since he worked at the university library, he identified more with his colleagues who were “just” librarians. Other than that, he could only describe his position in this way: “my job is about open science and research data management at the university library”.

Other stewards I talked to also identified as a form of librarian in transition. The typical situation would be that their university had a new strategy around research data management, and the librarians would see the opportunity to work more within that field while more traditional library work became less and less attractive: “the librarian role it is changing, so we’re doing a lot of tasks today that weren’t relevant only ten years ago[...], but I think this is way more exciting than ordering books, to put it that way!”. Still, the role came with no clear demands or measures of success (nor failure). The stewards had to figure out their role on their own.

In one way, this lack of guidelines, demands and job description was seen as a desirable aspect of the job – giving the stewards the freedom to do whatever they wanted. Maria told me she would spend some of her days just slacking off in the office because there was nothing she could do and no one expecting her to do anything either.

But, as I’m sure would be evident to you, this freedom was also experienced as a burden. Aside from the mental anguish of having, what could at times seem, a meaningless job[89], the stewards had to essentially invent their own role. In this venture of self-definition it was clear that the supposed freedom was far from absolute. Like many other positions in academia, the data stewards experienced a form of precarity, their jobs or employing institutions being funded as time-limited projects with no guarantee of continued support. One data steward, Victor, who was formally employed at the library reflected on the fragility of his position: “If

[89] Or a “bullshit job”, as proposed by late anthropologist David Graeber (2013)

[the Ministry for Education Science and Research] say 'this is not working', I guess [the university library] could reassign me. But I'm not a librarian, so where would they put me?".

The looming threat of budget cuts and non-renewals of contracts meant that the stewards had to continuously justify their existence to some funding entity (sometimes several). But with no clear and established metrics or indicators, the stewards struggled to figure out exactly *how* they could make themselves seem valuable. This was illustrated by Maria who was telling me about a meeting she was preparing for. The vice rector – who had initiated the stewardship program – had asked her and the other stewards at the university to present their work. Maria was not entirely sure how it would go "because [the vice rector] is kind of nervous about [the meeting], because we are presenting to the whole rectorate and this is kind of his big project". In other words, she had a feeling that the meeting could affect the continued existence of the project, and her job with it. The challenge for the data stewards, Maria told me, was that "there are no real criteria" to assess them according to, and, thus, no guidelines for how to best present themselves as valuable.

So while it could seem like the data stewards were free to do their job as they pleased, at closer inspection it became evident that they were, rather, governed by uncertainty. In other words, they would have to *assume* what their funders/employers would find valuable, and devise some goals based on this that they could then later try to attain in order to justify their continued existence – well knowing that their assumptions might have been wrong.

To complicate this further, the data stewards I talked to also found it challenging to attain the goals they ended up setting. Because – unlike a policeman who has the authority to punish deviance from policy (or law, as it is called), or an employer who can reward goal attainment with a bonus or a pizza party – the data stewards had no formal authority to help them implement FAIR. With no institutionalized source of authority, the stewards had to find alternative ways of gaining agency. Talking to the data stewards, it became apparent that they had developed methods to change researchers' practices in spite of their apparent powerlessness. These methods, generally speaking, consisted of what I sketch out as two steps: sensing and negotiating. I will note that while these methods were iterative and entangled in nature, with the purpose of clarity, I have chosen to present them separately in this chapter.

Sensing

When the data stewards started their jobs, they had, as mentioned, not much to go on. They had an office, a salary and a vaguely defined job description. Naturally, they had their education and other qualifications that enabled them to land the job in the first place, but generally, the beginning of their employment, as with most other jobs, consisted of finding out exactly what they were supposed and able to do. Maria spent her first weeks of employment walking around the halls talking to people, presenting herself in meetings, explaining what she was supposed to do[90]. She noticed quickly, though, that whenever she would bring up FAIR, "people [would] seem confused".

The other stewards had similar experiences, noting that, aside from a small group of what one steward called "idealists", most researchers didn't know what FAIR was, and weren't that inclined to become FAIR either. Alex even had one experience of a researcher coming up to them after a talk, expressing that they found the notion alienating. Clearly, there was some form of misalignment between FAIR and the researchers.

This misalignment was not the same everywhere though. Iben and Andreas, who had responsibility across different faculties and disciplines, noted that there was a distinct difference in the responses depending on the discipline:

> Iben: A lot of our job is to make people aware that there's something called FAIR, and that there's something called good data management. And for some disciplines this is completely evident and they already have good practices. For example, here with us, there's a lot of biologists that are very good – and not only the system biologists [...]. While for others, it's enough that we just mention Open Data and they put their spikes out. [...] Maybe in particular the technical disciplines – we have a lot of technically heavy engineering disciplines – they were not so interested.
>
> Andreas: …and the humanities faculty wasn't interested at all. We managed to get a meeting there and then there were like – I don't know – like three researchers there and some admin.
>
> Iben: I think it's a challenge that FAIR emerged in a sort of technical natural science context, and it doesn't immediately make a lot of sense in other disciplines. [...] Some disciplines are saying they don't have data, they have sources…

This last observation, on the difference between the more and less data-intensive disciplines, was something that was pointed out by several of the stewards. Paulina expressed that she felt FAIR had neglected "the long tail of science" – a term used to describe less data-intensive disciplines. Victor – who worked at a social science data repository – said that FAIR didn't make as much sense for the qualitative social science as the quantitative, making it hard for them to comply (and deposit their data).

[90] At least what she *thought* she was supposed to do

This misalignment between FAIR ideals and scientific practice was not only a phenomenon outside of the biological sciences. Maria, who worked within the biological domain, said that the prescriptions of the FAIR principles were “not close to how people actually work”.

But it wasn’t only disciplinary aspects that affected the alignment of FAIR and research practice. Other environmental factors also affected the ability of researchers to take up FAIR practices. The most notable and oft mentioned misalignment was the sense that researchers in general didn’t have the available time to make their research FAIR. Iben put it this way:

> I think that’s the main challenge, because it’s not something you have to do, and how will you put aside time and resources to share your data when you have other things you have to do? You have to publish, you have to do this and that…

Victor had made the same observation, stating that “researchers need to take the time”, before adding: ”but I’m not sure if they’re rewarded for the time that they invest in this”. In other words, what the stewards had identified was not necessarily a lack of time, per se, but a reward system that didn’t enable FAIR practices among researchers. Maria sensed the same, suggesting that even those who were interested in doing FAIR were unable to: “People don’t have time and it takes very long. Putting readme-files in every folder is a lot of work… and there are no rewards for doing it.”

The incongruity between the prescriptions of FAIR and the practices of researchers were readily evident for the data stewards. In particular, this was centered around the focus on the role of machines[91], which several of the stewards thought was at the very least not that helpful for the researchers, here illustrated by Paulina:

> The thing about machine readability is really important to FAIR, but we mustn't forget that [data] should be usable by a human – so human readability and human reusability. [...] And that is often very under-communicated, I think.

Andreas took his critique a bit further. He had an impression that some proponents of FAIR thought that “if only we make all data FAIR, then big data and artificial intelligence can solve every problem, and all knowledge will be available”. In his opinion, this idea was little more than a pipe dream and a distraction from the more positive potential of FAIR. In other words, the version of FAIR promoted by its proponents was in his opinion not helping most researchers.

Alex had spent some time reflecting on why it had become this way, and concluded that “the chief ideologues weren’t thinking about human interoperability”. This resonated with Victor’s

[91] Under the monikers of machine readability, machine actionability, machine interoperability, and so on

understanding of FAIR, who confided that “to be honest, the I of FAIR – interoperability – is a mystery to me”. The product of this focus on machines and lack of concern for humans, Alex reasoned, was that FAIR would be rejected by the researchers it overlooked. In their words: “if FAIR can't do anything for anthropology, there's no reason to believe that anthropology will do anything for FAIR”.

What the above examples illustrate is that through their attempts at making research FAIR, the data stewards experienced and made note of points of friction – where ideals and policy came into conflict with the practices of the researchers they attempted to govern. Through a combination of their experiential knowledge from academia and interaction with academics, the data stewards were able to sense the conflicting forces in their environment. Misalignments between epistemes, resources and interests meant that implementing FAIR came with significant resistance. But with no formal power and their jobs on the line, the data stewards had to find other sources of agency.

Negotiating multiplicity

Without the option of making researchers do anything against their interest, the data stewards instead took to finding ways of aligning these seemingly contradictory elements. Their approach to this was to find flexibility where they could – in the policies they were set to implement, among the researchers they were supposed to change, and in their own roles as data stewards.

This strategy was perhaps most clearly formulated by Alex, who said that their approach to the job was to “interpret policies in a way that is faithful to the policy, but also in the most generous way to the researchers in the field”. This negotiation was in other words a balancing act. If their interpretation was too liberal, they were at risk of getting researchers in trouble with their funders, journals or repositories, or getting themselves in trouble with their employer. If it was too strict, it put them at risk of being rejected by the researchers, and thereby, again, getting themselves in trouble with their employer.

In balancing between policy and researcher, they used their own sense of what they felt was the right amount of FAIR. Because even if they disagreed with some of the prescriptions of FAIR, the data stewards generally found that the interpretive flexibility of the concept afforded agreeable or even productive outcomes. Like Andreas said: “I’m a bit skeptical towards FAIR myself[, but] FAIR can be so incredibly many things”. The data stewards would therefore try to find and pick out the concepts in FAIR that worked and made sense in their context.

Examples of these practices could be Maria choosing to interpret interoperability on a smaller scale than what the FAIR principles prescribed. She would, instead of trying to make data interoperable across disciplines, focus on making data interoperable between researchers on the same project. Concretely, this work would mainly center around file naming conventions, folder structures, data formatting and standardizing metadata within projects, labs or groups. By working at this level instead of a broader one, Maria could tailor the data practices according to the relevant group's needs instead of having them cater to some unknown (often hypothetical) other data user. Not only would this be less work for the researchers, but, she claimed, it would also be more rewarding for them.

Another example would be Iben's strategy where he would survey the researchers needs and suggest solutions to them that could be conceived as part of a FAIR practice:

> We always approach it more with 'What is useful for you and this project, and what could be exciting to work with?' – trying to sell it by saying 'if you use Jupyter notebook[92], which is a great tool, then it's going to be easier to document and easier to fulfill these formal demands'.

Cody used a similar approach, rejecting parts of the policies that generated what he deemed unnecessary work for the researchers. So instead of focusing on FAIR, he highlighted "*intelligent* data management, saving researchers from bureaucracy – simplifying their lives" (my emphasis). Through this approach, the data stewards managed to use resources intended for increasing FAIRness to improve the workflow or working conditions of researchers in a manner more aligned with their interests. This strategy built upon taking the needs of the researchers' as a basis, and then seeing how they could be achieved within the interpretative space of the relevant policy.

But convincing researchers that FAIR (or an interpretation of it) would make their research more productive, efficient or otherwise didn't always work. As Iben put it: "it varies how much people get in on that approach…". When that would fail, the data stewards could instead negotiate directly with the researchers by appealing to their values. Alex described this approach as "weaponizing [the researchers'] research integrity against them". Iben acknowledged that "it's not particularly fun with formal demands", so he tried to frame it in a different way for the researchers, appealing to their morals or sense of "good research", "so we talk a lot about reproducibility and transparency in research [...] and that by thinking about FAIR data you make your project better, and it can become better for the rest of the world *if* you share it".

[92] A software tool for writing and annotating code

But while the different data stewards employed similar strategies to each other's, their abilities to enact them could differ quite substantially according to their sociomaterial conditions. The variation among the stewards I interviewed allows us some insight into how this would play out.

Affording differently

The interpretive flexibility of the data steward concept and the vagueness of the data stewards' jobs meant that the role became expressed differently from one place to another according to personal and environmental idiosyncrasies. While this flexibility made the data stewards more adaptable to their local contexts, it also meant that they were granted different affordances in their jobs, affecting their ability to carry out the above mentioned strategies.

By comparing the experiences of the data stewards I followed, one factor in particular stood out as important across all cases – one of proximity. How deeply embedded the stewards were in their environments seemed to be one of the biggest factors affecting the stewards' ability to perform their sensing and negotiation. While I won't rule out the potential effect of others, the two most pronounced dimensions of proximity arising in my research were the epistemic and spatial dimensions. The following parts of this section will therefore be structured accordingly, starting with spatial proximity.

Spatial proximity

While it can be hard to determine exactly what difference spatially close versus spatially distant stewardship makes, Iben and Andreas' experiences with COVID-19 serve a lucid illustration of the effect spatial proximity had on their work. When I interviewed them, it was in the middle of the pandemic, and they were forced to conduct most of their work online. This meant that they had to do their courses and presentations online through a video conferencing tool. This gave them fertile ground for comparison between stewarding in person and remote stewarding. Andreas described the experience this way:

> When we could be out talking to people physically, then we could pretty quickly… you feel the mood, because you quickly get comments and questions that make you understand how it is.

In other words, being denied the option of interacting in physical space adversely affected the stewards ability of sensing their environments. Alex also shared that his approach of "weaponizing [the researcher's] research integrity against them" heavily relied on his ability to sit down with them in their offices and properly talk to them – sometimes for several hours.

Cody expressed that the ability to seek out and talk to researchers in person made their job easier, underlining that “talking to people gives us more agency”.

Now, talking to researchers didn’t necessarily demand spatial proximity, as evidenced by Andreas and Iben who were located in their university’s library. In their case, they emphasized how the quality of the interactions they had suffered from them being further from the researchers. For Victor, who talked to researchers exclusively in virtual space – through email and sometimes over the phone – this effect was perhaps even more detrimental to his work. He described one researcher who contacted the repository he worked at, wanting to deposit a large and seemingly valuable dataset. The issue was that Victor found substantial issues with the dataset that had to be solved before publishing. The repository sent feedback suggesting a series of changes to fix the dataset and make it publishable. “We reminded him twice or three times: ‘Can we help you? Do you need anything?’ over the course of one and a half years”, Victor recalled, but in the end “he never wrote back”. After that, there was nothing Victor could do, as the researcher might have chosen to publish their dataset in a different repository with more liberal policies instead.

But quality of interaction wasn’t the only aspect that was affected by the stewards’ spatial proximity. Also the *access to* and *frequency of* interaction diminished. For Andreas and Iben this was experienced by their need to be invited to departments if they wanted to talk to people – a frequent occurrence at some departments, but conspicuously rare at others (particularly the humanities faculty). It was therefore easier for them to implement policy at the biology department than the department of interdisciplinary cultural studies.

For Maria, distance was not a significant barrier. Since she had her office in the same building and same floor as the researchers she was working with, she could get in touch with people with more ease – but not entirely without challenges. “So, I’m trying to force myself on people”, Maria said with a slight chuckle, “talking with them in the corridor and asking them again and again and again if they need something”. But even though the researchers would reject her approaches some times, at other times they would not. Therefore, being able to incidentally or coincidentally encounter the researchers more frequently was a significant strength of her situation. In this sense, Maria’s ease of access to researchers – in the hallways, the lunch room or by the coffee machine – was invaluable.

In short, the closer the data stewards were located to the researchers they worked with, the better equipped they were for doing their jobs. Closeness enabled them to sense the researcher's needs and desires more effectively, which in turn also improved their basis for negotiation. The possibility of frequent *quality* interaction also made negotiations easier. But, as mentioned, the spatial proximity was not the only one that mattered.

Epistemic proximity

One of the central and most difficult aims of FAIR was to achieve machine interoperability of research data across and between different scientific disciplines[93]. Essentially, the idea was to bring data across the disciplinary divides in science, and to achieve this, the original proponents of FAIR suggested that data stewardship should be a “domain-independent” data scientist (Wilkinson et al., 2016). According to this logic, the data stewards should be able to see the general value of data that a domain researcher wouldn’t be able to (Mons, 2018). Because of this outsiders view, the argument went, the generalist steward would therefore be well suited to make the data interoperable across disciplines, and thus help make the data more FAIR (Mons, 2018).

In practice, most data stewards were not domain-independent computer scientists, but tended to have a background in one field or another. The stewards who acted as generalists – stewarding across different fields – were not computer scientists either. Several of them came straight from a PhD in a specific field, like biology or social sciences. The closest to domain independence the stewards would get were those with a background in information or library science. Regardless of speciality or generality, all data stewards had a clear disciplinary identity, considering themselves to “belong” to a certain field. Typically, they would say that they “came from”, for example, linguistics or neuroscience, or that they had their “background” in humanities or political science, even though they were stewarding across different disciplines.

In several cases, the stewards with responsibility outside of their home discipline expressed a lack of ability to help researchers as much as they would like to. One of the stewards, Iben, said that most of the recommendations he gave had to be of a simple nature because he didn’t know the specificities of the other researcher’s field. He would then give “[recommendations] that weren’t that smart [...]. Like it’s stupid to have files named ‘final_final_final3_final’, and so on. Just these simple things”. When Paulina struggled to help someone, she had the possibility to delegate more domain-specific questions to “subject specialists” in a network of different university libraries. Still, from time to time, it would happen that there was no available specialist for a specific field or dataset:

> As a main rule, it is the subject specialist that curates the datasets within their fields. I curate language datasets because linguistics is kind of my field. But at other, smaller universities, it might be others than subject specialists – maybe a special librarian – that does that kind of work, and they might not have that researcher background. And then it might not be that easy

[93] This belongs to a larger discussion we will get back to in chapter 6

> to decide if a methods section in a readme-file is sufficient [to make a dataset FAIR], so yeah… (Paulina)

For Paulina, the availability of domain specific support entirely rested on whether the university library had a subject specialist for the relevant dataset, and if the subject specialists at the university libraries were responsible for data stewardship at all. This meant that, depending on a researcher's discipline, they could end up falling between the cracks, only getting more generic advice on their data practices.

Some of the generalist stewards also experienced a lack of legitimacy towards the researchers they were responsible for. Iben and Andreas expressed that they would sometimes be met with a disinterest from researchers thinking that this data stewardship thing was irrelevant for them – questioning why "someone from the library" would teach them about data management.

On the other side of the spectrum, the domain specialized data stewards felt more capable when it came to guiding and counseling researchers looking for help. They were more confident in identifying best practices in the field, relying on their own experiences as researchers in the domain. This would, for example, express itself in the choice of repositories to deposit datasets in. Perhaps more importantly, it helped the stewards identify what forms of metadata would be necessary to make a researcher's data reusable by other researchers within the same discipline, department or lab.

The skill to identify what the researchers would find useful in their own projects was particularly pronounced among the epistemically close data stewards. Maria talked about how the researchers at her department would use her to set up data sharing practices within certain labs, groups or centers – a form of help the researchers supposedly appreciated so much they would call her into group meetings on a semi-regular basis. In this sense, her disciplinary background made her help more attractive to the researchers, which in turn enabled her to align the researchers' practices with FAIR ideals.

But the stewards' epistemic proximity not only helped them identify and promote productive practices: it also enabled them to filter out the not so helpful ones. In particular Maria was a big fan of what you could call a "pragmatic" approach, and would routinely recommend researchers to not bother doing anything more than the minimum requirements from funders if it didn't help them in their job. One example would be her recommendation not to spend any time making "raw data" FAIR, since biologists preferred using the processed CSV[94] files

[94] CSV stands for comma separated values, and is the simplest non-proprietary format, consisting only of a string of numbers (and text), separated by commas (as the name indicates)

to bigger files in clunky formats anyways[95]. As for the social sciences, Victor said that he sometimes would be sent qualitative datasets for publishing that he thought would have little value if they were properly anonymized. In those cases, he would recommend the researcher to just not publish the data instead of trying to make it FAIR and GDPR compliant.

Through their epistemic proximity, data stewards were thus able to sense or know what kinds of practices that would and wouldn't be valuable for the researchers, minimizing the workload, and thereby lowering the tension between policy and practice. The generalist data stewards, who were ultimately more epistemically distant, lacked a certain embeddedness in the relevant scientific field were on the other hand at risk of increasing this tension by giving the researchers less helpful advice, and potentially a bigger workload without much payoff. That is not to say the generalists were not helpful, but that they had a harder time becoming helpful and carrying out their job.

Implementing policy through middle actors

So what do data stewards do? We already have a decent picture of their concrete practices, so when I ask this question here, I do not mean it in the sense of what actions they perform, but what the outcome of those actions are – what their transformative capacity is. In other words, I'm asking not what data stewards do *in* their environments, but what data stewards do *to* their environments?

Both in theory and in practice, the role of the data steward is a form of middle actor. In theory, data stewards are the actors between the FAIR principles and research practitioners. In practice, this still rings true, but quickly becomes more complex as data stewards have to intermediate between researchers, managers, their institutions, policies and a plethora of other larger or lesser actors. For the purpose of this chapter, it is this function as middle actors I will focus on, as it reflects both a Weberian ideal form of data stewards and their empirically observed practices.

In social science, the role of middle actors has garnered a fair amount of attention over several decades – first in political sociology and later also in science and technology studies (STS). Though both fields interrogate the role of middle actors, they differ in their focus, and in what kinds of actors they look at. The most striking difference is that political scientists tend to analyze how middle actors affect the implementation of policy (see Lipsky, 1969;

[95] This recommendation was also motivated by her own opinion on the resource impact of storing the data. As she put it: "Most of the data is never used and never deleted, and it's there forever. CSV is small, so that's fine, but mass spectrometry is a waste of resources".

Coburn, 2001; Spillane et al., 2002), while STS scholars typically look at how intermediaries affect the policymaking itself (Jasanoff, 1998; Duncan et al., 2020; McEntee et al., 2023). Put in simpler terms, political scientists have been more interested in top-down processes, while STS scholars have been more interested in bottom-up processes (or sometimes middle-out (Parag and Janda, 2014; Ingeborgrud et al., 2023)).

In the case of the data steward, my initial interest was in how they shaped the enactment of FAIR policy at the practitioner level – so, to this end, I will rely more on the political science branch of middle-actor research. In this tradition, the work of Michael Lipsky is particularly notable. In his seminal paper *Toward a theory of street-level bureaucracy* (1969), he conducted a study of, what he called, "street-level bureaucrats": state representatives who were in regular contact with citizens. What he found was that these bureaucrats – who ranged from teachers to policemen – often struggled to fulfill their jobs according to the relevant laws or policies when they interacted with their subjects (pupils or welfare recipients, for example).

Lipsky (1969) contended that these bureaucrats – middle actors between state and citizen – struggled to implement laws and policies due to lack of resources and inadequate work environments. Whereas he argued that this was particularly bad in the US during a period of welfare crisis, he noted that this would, to a varying degree, be the case in most bureaucratic institutions (Lipsky, 1969, p. 4). But his most important insight, at least for this chapter, was in how these middle actors would overcome their challenges. He argued that the limitations in their positions would put them in a dilemma and force them to negotiate between the prescriptions of their institutions and the context they worked in. To do this, they would apply their creative capacities and non-conventional resources, essentially improvising in order to attain results that were as good as possible according to their own convictions (considering the limitations of the situation). The bureaucrats would, in other words, use their own discretion to both assess and prioritize actions and outcomes. The concept of "street-level bureaucracy", Lipsky argued (1969; 2010), could be used to explain parts of the incongruity between policy and its implementation – why policy and practice remained different from each other.

Several later studies, conducted on teachers in specific, pointed to the same transformative capacities of middle actors as they attempted to implement educational policies (Weatherly and Lipsky, 1997; Cohen and Ball, 1990; Tyack and Cuban, 1995; Coburn, 2001; Spillane et al., 2002). The authors behind these studies argued that one of the main reasons policies would fail was because they rarely took into consideration the multiplicity of factors that shaped actors' practices other than the policy itself. The argument went that because of policymakers' naïve assumptions, middle actors were forced to modify policies and negotiate with their environments in order to make them fit with each other – typically to the detriment of policy. Coburn (2001) thus suggested that any study of policy would need to also take into account middle actors, but also their environments.

One striking quality of the above cited literature is that, in dealing with middle actors, it frames them as somehow deficient. Which element that is considered deficient ranges from the environment of the middle actors, like their inadequate resources or work contexts (Lipsky, 1969); to the middle actors cognitive or moral capacities (Coburn 2001; Spillane et al., 2002). Thus the authors frame middle-actor policy implementation as a problem that should/could be solved with extra funding, training or other additive means.

But an alternative approach to the implementation problem had already been suggested by Majone and Wildavsky (1978), who observed that policies would often fail – not because of deficient middle actors – but because the policies were "subject to an infinite variety of contingencies" (p. 113). This multiplicity, the authors argued, would be almost impossible to take into consideration in the design of most policies. Trying to make a policy fit perfectly within its context would therefore be a waste of resources. Implementation efforts, Majone and Wildavsky (1978) argued, would "inevitably shatter, reform, or reject elements of original policy" as policy and practice collided with each other. First attempts at policymaking would, in other words, always fail to some degree, and thus, they argued, policymaking should be an iterative or "evolutionary" process. Through this approach they introduced a focal shift away from the deficiency narrative and linear models of other scholars, and towards a more flexible and pragmatic understanding of policymaking and social change – in their own words:

> Under this model, the 'success' or 'failure' of a social action enterprise will not be be so much determined by whether or not a predetermined policy was carried out to the letter, but by the determination of what has been accomplished in a particular social area as the result of a rational, evolutionary use of resources as a part of the dynamic affecting the target area (Majone and Wildavsky, 1978).

Their solution was thus to fix policy through continuous feedback processes. To this end, Majone and Wildavsky (1978) saw middle actors as instrumental as a form of sensors of incongruence – arguing they would be able to sense policy/practice misalignments in their experience with implementation challenges first hand – and, ideally, feed that information back to policymakers.

Doing away with failure

Judged by the goals formulated by proponents of FAIR – like achieving machine-readability, an Internet of FAIR data and Services, and interdisciplinary interoperability – data stewards have so far failed to realize FAIR. Using the scholarship on middle actors, diagnosing the reason for this failure becomes relatively straight-forward. One could, like Lipsky (1969), point out the lack of resources the stewards had, their precarious working conditions and lack of strongly defined job descriptions; like Coburn (2001), one could claim that the stewards lacked understanding of the FAIR principles; or like Spillane et al. (2002) remark their moral opposition against FAIR. Applying this success/failure dichotomy gives the story about the implementation of the FAIR principles a reasonably straight-forward conclusion, but not a very interesting, helpful or symmetrical one (Bloor, 1976). Because, looking at the empirical material presented in this chapter, we realize that the process of implementation is more complex, messy and multiple than presented in much of the conceptual work on middle actors. Furthermore, in all this mess, it is not entirely clear to what degree the implementation of FAIR was a failure at all.

In this chapter, I align with the observations and arguments of Majone and Wildavsky (1978), arguing that mess is not a sign of failure, but of potential. Mess, in its inevitability, can be a productive opportunity for improvement (Majone and Wildavsky, 1978). But Majone and Wildavsky (1978) still judged implementation of policy according to a failure/success dichotomy (even if the metrics for success were less clear). This leaves us with the question: failure/success according to whom? In a process where agency is distributed throughout a messy constellation of actors and environments, maintaining a centralized authority for assessment might not be productive – especially when the authority's legitimacy is questioned by its subjects.

According to the data stewards, the goals of policymakers were in opposition to the researchers' desire to not increase their workload – a position some of the data stewards adopted. The general perspective was that policymakers didn't understand how researchers' jobs were, that they didn't have time to do extra data work without getting any reward for it.

On the opposite side, we've seen that the original proponents of FAIR didn't consider researchers' position legitimate either – claiming they didn't understand how FAIR worked (see chapter 3). Proponents of FAIR wanted convergence, but implementation had only led to fragmentation and multiplicity. With convergence (or standardization) across situations and contexts as a stated goal, the situations described in this chapter would not achieve what these policymakers desired. For that to happen, there was too much flexibility in terms of the implementation tools: the principles, the policies and the data stewards' jobs were either vague, ambiguous or unenforceable. In combination with an environment that was in (partial) opposition to the implementation, this led to the data stewards interpreting and modifying policies to achieve only as much alignment between policy and practice as they could within their limited means.

That is not to say that a stricter, more concrete implementation strategy would pay off. The signs of resistance among the researchers in this chapter suggests that such an approach might have struggled to achieve the policymakers' stated goals. Assessing policy implementation according to success or failure thus leaves us at a stalemate – a situation where policymakers and policy subjects both fail to achieve their goals or fulfill their desires.

In assessing policy implementation, I instead suggest operating with a sliding scale going from alignment to misalignment. This approach rejects the judgment of policies according to attainment of their goals (which are likely to fail anyways), and replaces it with a judgment according to consensus between policymakers and policy subjects.

This approach may seem similar to the concept of convergence promoted by FAIR proponents. The alignment/misalignment approach that I will suggest, though, differs in an important way from the convergence idea. While alignment and convergence both highlight consensus, the GOFAIR association and authors like Wittenburg and Strawn (2018) assessed outcomes according to how close the convergence was to their stated and desired outcome of convergence in standards that would enable semantic interoperability. Convergence around any other outcome would, thus, still be considered failure.

Assessing policy implementation according to alignment/misalignment does away with the privileged position of policymakers to assess the quality of an implementation. Taking what I would call a more agnostic view – in terms of what is considered good or bad – instead acknowledges the interdependency of factors in a given process. This agnosticism also does away with deficiency narratives as explanatory models and instead sees the behavior of actors as efforts to construct stable, viable and productive constellations in complex environments. In other words, behavior that does not work towards stated policy goals are not to be seen as products of deficiency, but as efforts to align policy with unexpected or unaccounted for practices and environments. Applying this perspective, policy and practice is then not seen as an outcome of conflict between two opposing forces, but an outcome of iterative and distributed adaptation, negotiation and co-production.

Taking the agnostic view, we see how Maria, for example, negotiated with her environment in order to align policy and practice. Considering the researchers' lack of time and motivation when it came to data sharing, she practiced an interpretation of FAIR and the relevant policies to promote practices that were adapted to the researchers' needs. So instead of promoting interdisciplinary interoperability, she promoted intra-project interoperability, thereby demanding less time and giving a greater sense of reward for the researchers. We also see that the feasibility of this approach in part relied on material factors such as Maria's spatial proximity to the researchers – bumping into them in the hallway or the shared lunch room[96]. Further, sociocultural factors such as her epistemic proximity and ability to anticipate and empathize with researchers' needs and desires also played a role. The importance of the wider environmental factors become evident when we compare with data stewards in different contexts, like Andreas' or Victor's lacking contact with researchers affecting their ability to change their practice, or Paulina's limited ability to support researchers from other disciplines.

None of the data stewards transformed research practices in a way that aligned them with the ideals of FAIR, but they did contribute to changing research practices in some way or another. We see that the data stewards' ability to make productive change in their environments differed significantly according to how they were configured within them. Doing away with the failure/success dichotomy could shift the focus away from trying to attain an unlikely/impossible goal and towards supporting productive change that is adapted to its contexts: "good research", as Iben put it.

What would this support look like? Drawing on Majone and Wildavsky's (1978) strategy of "evolutionary"[97] implementation, we can see policymaking as an iterative learning process that draws on feedback from middle actors. To this end, data stewards could be seen as a useful source of information and knowledge. This approach would be supported by literature from the field of STS suggesting that mediating local knowledges into processes of policymaking can improve alignment (Jasanoff, 1998; Pielke, 2007; Ingeborgrud et al., 2023; McEntee et al., 2023). In previous studies this has largely focused on scientists (Jasanoff, 1998; Pielke, 2007) or on bureaucrats (Ingeborgrud et al., 2023; McEntee et al., 2023) as mediators. Through my research for this chapter I suggest this would apply to middle actors such as data stewards as well.

A barrier to evolutionary implementation is that it would require a certain amount of humility. For policymakers, taking middle actors seriously, supporting their work, would mean opening

[96] Even factors like researchers not home-officing and having the desire and ability to take lunch or grab a coffee would play a role here

[97] This notion of "evolution", it is important to note, is not attempting to imply a sort of linearly progressing improvement (as in stronger, faster, better), but a process that aims to create a stable configuration of environmental elements while allowing for adaption according to changes in said environment.

up to the idea that their policies won't be realized, and giving up on attaining their goals. As this thesis has shown, and will continue to show, the original proponents of FAIR had a hard time giving up on their vision. But as this chapter has shown, their persistence in pursuing their vision has not seemed to pay off. Middle actors will use the interpretive space they have to adapt policies regardless. Following an evolutionary implementation strategy means embracing adaption instead of treating it as a hurdle.

The value of the data stewards, as I have argued, stems from their ability to sense their environment and negotiate with(in) it to create alignment of the different elements. Without an evolutionary implementation strategy, this ability becomes only half-utilized, as middle actors become limited to negotiating policy mainly in terms of their own interpretation. Integrating data stewards in the process of shaping policy could contribute to making policies that better fit with local contexts and are more easily aligned with the practices and environments of their subjects[98].

The case for middle actors

At the very beginning of this chapter, I aimed to answer why the FAIR principles were so differently expressed by its proponents and its supposed practitioners. While I was quick to point towards data stewards – the actors tasked with implementing FAIR – as the source of this discrepancy, it was clear that this was at best only part of the answer.

A more detailed picture formed after a closer look at the practices and perceptions of a handful of data stewards in different countries, different positions and different disciplines. After interrogating the data stewards it became clear that the differences in FAIR policy and practice had multiple sources – an environment consisting of a multitude of interconnected, sometimes opposing, factors. The assumptions about science inscribed into FAIR would not always line up with the sociomaterial environment, putting itself at odds with researchers' desires and practices.

It became clear that, what initially seemed as a policy failure, could through a different lens be seen as a successful venture into aligning two at first incompatible interests: that of FAIR proponents and of research practitioners. By using their discretionary freedom afforded to them by weakly defined job descriptions and ambiguous policy definitions, the data stewards managed to navigate and negotiate the environments they were embedded in – and ultimately created more or less stable configurations of policy/practice relationships.

[98] This strategy would also align with the desires of some of the data stewards that I talked to, who expressed that they would like to participate more actively in shaping FAIR policy but felt they had little influence or legitimacy in the eyes of policymakers.

The process of alignment expressed itself differently according to the variations in the different environments. In particular epistemic and material factors seemed to make a significant difference – such as the data stewards proximity to researchers' disciplinary knowledge and practice, and proximity to researchers in a physical sense. The outcomes also varied according to available resources and epistemic compatibilities with policy. FAIR did therefore not only express itself differently in principle and in practice, but also varied significantly between different enactments.

While this variation in outcome – a product of adaption to local contexts – may have been seen as positive by the researchers affected by this policy change, it was a negative in the eyes of the FAIR proponents who seeked uniformity in implementation – or convergence, as they put it themselves. Judging by the insights gathered from this empirical study, such a convergence seems unlikely. An approach that includes middle actors and focuses on alignment, continuous development, and openness towards alternative outcomes could be more successful at making research FAIR – with the caveat that this FAIR would be otherwise.

Chapter 5: making a FAIR assessment

In February 2022, I was invited to participate in a panel that would give out an award for data FAIRness. Eight datasets were nominated for the prize, but only one could win. Although I had spent over a year already studying the FAIR principles, I felt uncertain about how to evaluate the FAIRness of each dataset. In preparation, each panelist was encouraged to make use of the FAIR principles and the FAIRplus indicators[99] when assessing the FAIRness of each dataset. Additionally, we were told we could use two automatic assessment tools – F-Uji or the FAIRmetrics evaluator – but it was made clear that these should not be used to score the datasets themselves, only the repository that the data was stored in. This was important, as the panel wanted to score the effort of the dataset depositors, not the infrastructure they were using, as this was deemed to be unfair.

Clicking through the links for the datasets, I felt a tad under-qualified for the task. In many cases, I didn't even know what I was looking at. When you're accessing a dataset, what you're often seeing is a webpage hosting that dataset. Sometimes, these look a bit like the webpage for a journal paper, just without the paper itself – instead there's a button that says "download". As the datasets I was evaluating were from the life science domain, there was little I could make out of them. A spreadsheet with numbers and codes made no sense to me, but I could tell if the data had a usage license for instance. So using the FAIR principles as a checklist, I went to work.

> "(Meta)data are assigned a globally unique and persistent identifier"? Check!
>
> "Data are described with rich metadata"? I guess. Half a point.
>
> "Metadata clearly and explicitly include the identifier of the data they describe"? Hmm… I don't think so. No points?

As I went on, the task felt less and less like qualified assessments, and more and more like guesswork. I had no way of putting two lines under my answers, confidently saying that “this is FAIR enough”.

In between my invitation to assess these datasets and the convention of the assessment panel, there was a seminar on FAIR evaluation, organized by the EOSC Nordic project (2022a; 2022b). And even though they had been working on the topic for several years – some of them being authors of the original FAIR paper – they seemed to have a problem

[99] A set of indicators to measure FAIRness developed in the ELIXIR network. More on this later

similar to mine. They had no sure way of saying one thing was FAIRer than the other. Erik Schultes – one of the original authors of FAIR – summarized it this way:

> A FAIR evaluation on a resource from two different evaluators can sometimes give wildly different outcomes, and this is problematic, to say the least, because **it can really undermine confidence in the FAIR principles** or in the approaches and where the state of the consensus is. So I would say there's a limitation or maybe a risk right now of divergence, either because everybody's creating new evaluators all the time, or because **we haven't really come to an agreement on these interpretations of what the FAIR principles are asking for and how we should actually test them** (EOSC Nordic, 2022a, 15:52) (emphases added).

On the one hand, this was a relief to me, because it turned out I wasn't the only one struggling with this. On the other hand, it made me question: why are people struggling so much to establish a way of evaluating FAIRness?

If you're going to measure something, it is obvious that you'd need something to measure it up against, be it bushels, feet or the highly standardized and maintained kilogram. If you're wondering if you are observing a kilogram, there are several ways of figuring it out. The first step would be to put it on a scale which will give you an answer. But how do you know that the scale is not wrong? If you have a consumer-grade scale, you could check it against a liter of water (if you have a way of measuring that) – which used to be *the* definition of a kilogram. The weight and volume of water is unfortunately not that precise, so you might want to use a scale officially certified by a legal body – like the one in your local grocery store. If you didn't trust that scale either, you could check it against a reference kilogram, kept by different national and regional measuring authorities. If you suspected someone had tampered with your reference, you could check that against *the* reference kilogram in Saint-Cloud, France (Sample, 2018). Could you then be certain that your scale was correct or not? Not really.

For a long time, this reference lump of metal encased in two glass bells in a vault in France was enough to settle the score once and for all. But increasing demands on precision from researchers across the globe, and the fact that *the kilo* seemed to be losing mass, led to calls for replacing it as a measure. This lump of metal could no longer be trusted to lay claims on reality, what a kilo *is* in its very nature. A new definition of the kilo was developed taking the Planck constant as its basis, and using that to calibrate a device called a Kibble balance. And by decree of the International Bureau of Weights and Measures in 2019, *the kilo* was dead. Long live the kilo. Thanks to this complicated process, any doubt that you're observing a kilo can (theoretically) be settled, according to the mantra of the International

Bureau of Weights and Measures, “for all times and all people” (or at least until demands for more precision arise) (Sample, 2018).

But that is the kilo. When we look at research data, we get a different story. Because, if you’re wondering whether a dataset is FAIR or not, you are out of luck. The FAIR principles have no exact definition (Wilkinson, Sansone et al., 2018), no proper reference (EOSC Nordic, 2022a, 57:45), and no ultimate authority (FAIR Data Maturity Model WG Case Statement, 2018) to settle the score. Disagreement is widespread, and there seems to be no way of settling a claim that something *is* FAIR.

In this chapter of my thesis, I explore why it has been so difficult to say whether something is FAIR or not. By looking at the niche field of FAIR evaluation between 2016 and 2022, this chapter will take us through the making of FAIR metrics and assessment tools, and how they were used later on. In order to make sense of all of this, I will draw on the field of valuation studies, and in particular the work on valuation constellations.

Making metrics

In the original publication of the FAIR principles in 2016 (Wilkinson et al.), the authors expressed it quite clearly, that they didn’t want to impose any standards on data producers.

> These high-level FAIR Guiding Principles precede implementation choices, and do not suggest any specific technology, standard, or implementation-solution; moreover, the Principles are not, themselves, a standard or a specification. (Wilkinson et al., 2016).

They wanted “the community” to figure this out themselves – meaning that FAIR should be developed from the ground up, basing itself on existing use cases (Wilkinson et al., 2016). There are therefore no metrics or standards of evaluation in the original publication of the FAIR principles.

But, the 2016 paper didn’t leave the reader without any guidance on how to assess if something was FAIR or not. While there was no given system or instruction on ranking or rating data or other research objects, the authors supplied their audience with two important elements for valuation: some general principles and a handful of examples. Still, this framework was not enough to avoid a series of diverging interpretations and implementations of FAIR over the following years. As long as FAIR remained only vaguely defined, it was practically impossible to authoritatively claim that something was *not* FAIR. As long as something was arguably findable, accessible, interoperable and reusable (the key word here is “arguable”), anyone could claim it to be FAIR.

Wanting to avoid illegitimate claims on FAIRness and “subjective [...]assessments”, a handful of the 2016 paper’s authors got together in what they named “the FAIRmetrics group” to establish some clear ways of measuring whether something was FAIR or not, so that “data and service providers, journals, funding agencies, and regulatory bodies [could] qualitatively or quantitatively evaluate such claims” (Wilkinson, Sansone et al., 2018). This need for clear metrics or criteria for evaluation of FAIRness was also emphasized by the Expert Group report *Turning FAIR into Reality* of the same year (EC, 2018). So from 2018 on, several initiatives to develop ways of measuring, evaluating and assessing FAIRness were launched[100].

Knowing that neither of these initiatives succeeded in establishing a FAIR measure, looking at their attempts to do so will give us some insight into why exactly it is so difficult to make a measure. In this part of the chapter I will therefore present the development of three of the most influential efforts at measuring FAIRness: the FAIRmetrics group, the RDA FAIR Data Maturity Model Working Group, and the FAIRsFAIR project’s F-Uji tool.

FAIRmetrics

The FAIRmetrics group made itself notable through, in particular, two of its qualities: it consisted of central figures in the authoring of the original FAIR principles, and it was the first to focus on having machine-readable metrics. There were earlier initiatives[101] that had developed FAIR evaluation methods based on self-assessment questionnaires, but according to first author of the FAIR principles, and head of the FAIRmetrics group, Mark Wilkinson, “[q]uestionnaire-based approaches do not evaluate the entire point of FAIRness” (2019) – the point being machine-readability. The FAIRmetrics group further claimed that these questionnaires would not be “objective” or scaleable enough to be of any value to the FAIR movement (Wilkinson, Sansone et al., 2018). For the FAIRmetrics group it was therefore important to make an automated FAIRness assessment tool (Wilkinson, Sansone et al., 2018).

Trying to make an automatic assessment tool came with the added challenge of having to translate the human-readable principles into machine-readable code. This meant that the group not only had to formalize what words such as “rich” or “detailed” meant, but also figure out how a machine could measure this. After three meetings over the timespan of half a year, the group had only managed to produce three metrics, and they “weren’t even entirely happy with those” (Wilkinson, 2019b, p. 8). Instead of trying to completely solve FAIR

[100] Perhaps too many, as we’ve seen some claim already
[101] see: FAIRdat, DANS FAIR Data Review and CSIRO 5-star Data Rating

metrics at their first attempt, the group aimed for “good enough”, and saved the harder problems for later. For example, in their first published F2 metric, they stated that: “This metric does not attempt to measure (or even define) ‘Richness’ - this will be defined in a future Metric” (Wilkinson, 2019a).

When the FAIRmetrics group finally had a full set of metrics, they set out to test them on real world cases. They invited “a variety of resources [mainly data repositories] to participate in a self-evaluation” (Wilkinson, Schultes et al., 2018). The purpose of the test was, purportedly, not to assess the FAIRness of the repositories, but “to evaluate the legitimacy, clarity, and utility of the metrics themselves”. According to Mark, some of the resource owners felt that the test was “unfair” (Wilkinson, 2019c, p. 20). The repositories claimed their practices were FAIR, and yet, the test told them they weren’t (Wilkinson, 2019c). This happened in part because a data standard would have to be registered as FAIR in order to be recognized as FAIR. Alternative, non-registered methods would therefore produce a low score – to the irritation of one of the groups that were tested:

> We did provide linkset for genes and diseases mentioned in this dataset, but we didn't follow the proper linked data principles. [...] Instead of using exiting [sic.] URIs[102] we defined our own datatype to represent these identifiers (in Wilkinson, Schultes, et al., 2018, p. 13).

The first takeaway for the FAIRmetrics group from this test was “that certain metrics (and in some cases, the FAIR Principle underlying it) were not always well-understood” (Wilkinson, Schultes et al., 2018). In other words, they thought the problem was not necessarily the test, but the testees. The idea was that if only the testees had a better understanding of the FAIR principles or how they were represented in computer code, the testees would not disagree with the evaluation[103].

But after a period of reconsideration, the FAIRmetrics group started thinking that maybe they themselves were in the wrong, and that it wasn’t the testees, but the test that needed to adapt. As Wilkinson later admitted: “We were insufficiently aware of the diversity of metadata-provision mechanisms in-use in the repository community; our evaluations were not ‘fair’” (Wilkinson, 2019b). This led to the group realizing they could not create a set of metrics that could be applied universally, in spite of their original ambitions (Wilkinson, Sansone et al., 2018; Wilkinson, 2019b). Rather, the FAIRmetrics group encouraged “communities” to make and register their own metrics (using the FAIRmetrics template) to measure FAIRness in a way that would be deemed relevant and legitimate by the users

[102] Uniform Resource Identifiers
[103] A way of thinking that should be recognized from chapter 3

(Wilkinson, Schultes et al., 2018; FAIRmetrics, N.D.). After four years, the number of communities following suit was exactly one:

> I expected communities to build their own metrics and tests that fulfilled the R1.3 principle. Now, is that ever going to happen, or is that just wishful thinking? Well this is happening now… So, within the rare disease community – largely because we have a lot of infiltration into that community and we're encouraging them to do this – they've started to build rare disease-specific FAIR metrics. They've got some of the Fs but also a couple in the R's, and specifically R1.3, where they're actually validating the metadata according to the model that we have within the European joint program on rare disease (Wilkinson, 2022b, 31:38).

At the date of this thesis' publication, the number of communities registered with tailored metrics remains one: the rare disease community (Fair Maturity Indicators List, n.d.). Whether other communities will follow suit is still an open question – because FAIRmetrics is not the only option for those who want to measure the FAIRness of their datasets or other resources. For anyone who would like to conduct a FAIR assessment, there are a number of alternatives to choose between, and none of them have authority as the de facto way of measuring FAIRness. So in 2018 a Research Data Alliance Working Group set out to solve this problem (FAIR Data Maturity Model WG Case Statement, 2018).

RDA FAIR Data Maturity Model

Almost exactly like the FAIRmetrics group before them, The RDA FAIR Data Maturity Model Working Group (henceforth "the RDA group") argued that their existence was needed due to the ambiguity of the FAIR principles and the resulting divergence in interpretation and implementation (FAIR Data Maturity Model WG Case Statement, 2018). While the FAIRmetrics group's approach was to solve divergence by making a set of metrics, the RDA group realized that this hadn't solved the problem because

> [a]s a result [of FAIR ambiguity], a number of incompatible methodologies to assess FAIRness have been developed already. [...] Due to the lack of a common set of core assessment criteria for FAIRness, researchers and organisations cannot evaluate the readiness and implementation level of their datasets vis-à-vis the FAIR data principles in a coherent way (FAIR Data Maturity Model WG Case Statement, 2018).

To solve the problem of divergence in interpretation, the RDA group came up with the idea to create another set of FAIR metrics, or "a common set of core assessment criteria for FAIRness" (FAIR Data Maturity Model WG Case Statement, 2018).

Unlike the FAIRmetrics group, the RDA group consisted mostly of people who weren't involved in the authoring of the FAIR principles, but consisted of a small number of what they called "representatives from organisations with experience in the area of FAIR data uptake, real-life implementation and compliance across borders/disciplines" (FAIR Data Maturity Model WG Case Statement, 2018). And, unlike the FAIRmetrics group before them, the RDA group chose not to base their metrics too strongly on the original FAIR principles. This choice received some resistance, but was defended and kept by the members, as illustrated by this interaction between Barend Mons and Francoise Genova at the group's first face to face meeting:

> Mons: Been lots of interest in FAIR - but also lots of misinterpretations. [The working group] could make focus on what FAIR really is, by sharing original, relevant papers in the GitHub and suggest that all read them first.
>
> Genova: FAIR has moved beyond conceptions of initial creators. (RDA FAIR Data Maturity Model Working Group, 2019b)[104].

Instead the RDA group aimed to build their metrics according to practices in the "scientific communities", wanting to "define a common meaning of the FAIR principles through a good description of the assessment criteria" (RDA FAIR Data Maturity Model Working Group, 2019c). Their approach for designing FAIR metrics was therefore based on reviewing pre-existing metrics and assessment initiatives, and building a common set of criteria based on those (RDA FAIR Data Maturity Model Working Group, 2019a).

After reviewing the existing assessment and metrics initiatives, the RDA group concluded that if they wanted to create a viable and applicable set of metrics, they would also need to consider the complexity of the material they aimed to assess. Notably – like FAIRmetrics before them – they found that data from different communities needed different forms of evaluation. This led them to conclude early on – unlike FAIRmetrics – that there was no way of creating an automatic assessment across disciplines at that point in time (RDA FAIR Data Maturity Model Working Group, 2019c). As group member Jonathan Petters put it:

> I would argue that a world where we can fully automate assessment of all FAIR metrics (especially for evaluating the reusability of data within a domain) is a world that does not exist yet (Petters, 2019).

While the opinion that disciplinary diversity posed a challenge was far from controversial, the departure from automatic assessment as a way to "solve" this challenge created a stir within

[104] These "quotes" are from notes taken at the meeting. Hence the awkward representation of the interaction.

the group. By some, this departure was understood to effectively be a rejection of one of the core tenets of the original FAIR paper – a rejection of machine-readability and machine-actionability (RDA FAIR Data Maturity Model Working Group, 2019c).

Following the first review of the community practices, the RDA group set out to develop their "core assessment criteria" (Herczog & Russell, 2019). A central part of this work happened through discussions on the RDA group's GitHub page (RDA FAIR Data Maturity Model Working Group, n.d.), where it became evident that the different views on community standards and machine-readability were challenging to overcome (RDA FAIR Data Maturity Model Working Group, 2019e; 2019c).

Disagreements between those who wanted to take a more pragmatic route and those who wanted to stay close to the original principles quickly became an issue for the RDA group. This was not only because the discussions were difficult to settle, but also because it took them a lot of time to do so, and – unlike the FAIRmetrics group – the RDA group had a deadline. At an early stage, the European Commission had announced its interest in the group's work, and indicated that it would adopt the group's assessment criteria *if* the RDA group managed to finish their recommendations within a timeframe of 18 months (RDA FAIR Data Maturity Model Working Group, 2019b). As a consequence of the pressure, the RDA group chose to put significant effort into coming to an agreement, and getting there fast. To hurry the process along, the European Commission hired a group of experts on consensus-building from PricewaterhouseCoopers, led by Makx Dekkers, to be in charge of the editorial team of the RDA Group (RDA FAIR Data Maturity Model Working Group, 2019b).

One of the ways the RDA group tried to resolve the disagreements between the pragmatists and the principlists was to include all the criteria for machine-readability in the metrics, but adding a ranking system according to how prioritized the metrics should be. The result was that all "core assessment criteria" would range from optional/useful at the lowest, recommended/important in the middle, and mandatory/essential at the highest level (RDA FAIR Data Maturity Model Working Group, 2019f). This allowed the group to recommend steps they didn't see as easily achievable yet (like machine-readability), and give a score according to the steps they did find achievable (like findability). One interaction on the group's GitHub forum serves to illustrate:

> keithjeffery commented on Aug 21, 2019
> [W]e need to find a way to include (as some kind of FAIR) the large amount of assets currently under DoI schemes. As I tried to indicate, the problem arises mainly for autonomic processing of the metadata and assets. (Jeffery, 2019a)

> makxdekkers commented on Aug 23, 2019
>
> @keithjeffery Based on your analysis, we may need to downgrade F-01D, F1-02D and F3-01M from mandatory to recommended; otherwise most of the (meta)data in currently common DOI-based publishing approaches would have to be declared not-FAIR.
>
> Making these indicators recommended would signal that it would be more FAIR to have persistent identifiers for metadata and for data, not in terms of pass/fail, but as a suggestion for future improvement. (Dekkers, 2019b)

The prioritization levels were, in other words, not first and foremost based on how essential the different principles were for achieving FAIRness, but rather on what a resource or dataset could be reasonably expected to achieve. This is most clear in the decision to make none of the indicators for interoperability classified as mandatory (RDA FAIR Data Maturity Model Working Group, 2020b), despite interoperability being considered a central goal of the FAIR principles (Wilkinson et al., 2016). Dekkers justified the decision by claiming it would be impossible to achieve for disciplines that didn't have standardized vocabularies or common practice of linking datasets, sparking a lengthy debate on the RDA group's GitHub (Dekkers, 2019a).

So despite the consolidation efforts of Dekkers, when the group finalized the first iteration of their assessment criteria, the different stakeholders were still not happy. After testing the criteria, it became clear that the problems regarding community standards and machine-readability were still unresolved. One tester responded: "Several indicators require compliance with community standards, but who defines them?", another: "FAIR principles are aspirational and ambitious, aiming at full machine understandability, but current practices are not well aligned at this point in time", and, altogether, the editorial team noted that testers were generally stricter with the priorities than those given by the group, in particular concerning community standards and machine-readability (RDA FAIR Data Maturity Model Working Group, 2020a).

Taking all the feedback into consideration and attempting to reach a final consensus on the RDA FAIR Data Maturity Model before the deadline, the editorial team finalized and got the maturity indicators officially recommended by the Research Data Alliance in June 2020 (RDA FAIR Data Maturity Model Working Group, 2020b). The final result ended up containing no less than 41 indicators for FAIRness, of which 20 were deemed essential, 14 important and 7 useful, as seen in figure 3.

Principle					
Priority	Findable	Accessible	Interoperable	Reusable	Total
Essential	7	8	0	5	20
Important	0	3	7	4	14
Useful	0	1	5	1	7
Total	**7**	**12**	**12**	**10**	**41**

Figure 3 (RDA FAIR Data Maturity Model Working Group, 2020b)

Still, after much objection, none of the indicators for interoperability got the highest priority, while all of the indicators for findability did. This was addressed by the authors in the final document, where they noted that the indicators for interoperability in particular posed a challenge – but one that could be solved in the future by a group maintaining the indicators. The authors underlined that there would be a need for work on a registry of standards that could be applied in assessment of FAIR entities (RDA FAIR Data Maturity Model Working Group, 2020b, p. 38). At the date of writing, the maintenance group has made no updates or changes to the maturity indicators. But even though the document stayed the same from 2020 on, the ways the indicators were put to use changed significantly – most notably through the FAIRsFAIR project and their F-Uji assessment tool.

FAIRsFAIR F-Uji

FAIRsFAIR was a Horizon2020 funded project that ran from 2019 to 2022, with the aim (and subtitle) of "Fostering Fair Data Practices in Europe" (sic[105]) (FAIRsFAIR, 2019). Their ambition was wide-ranging and pertained to different elements of FAIRness – including, but not limited to, FAIR assessment (Gaiarin, 2020). But FAIRsFAIR had a slightly different approach than the two other groups. Whereas the FAIRmetrics and RDA groups aimed to assess the FAIRness of data sets, FAIRsFAIR wanted to assess the FAIRness of the repositories containing the datasets. The FAIRsFAIR project figured that the best way to do this was to look at the FAIRness of the datasets contained in the repository. So in the end, their method for assessing repositories was going to be built upon the assessment of datasets after all.

The approach of measuring individual datasets was not only a practical answer to the question of how to assess data repositories, but also allowed the FAIRsFAIR project to draw upon the work of others in the field of FAIR assessment. As part of their work, several members of the FAIRsFAIR project had therefore been participating in the RDA group's workshops and meetings (RDA, 2020). So when the RDA group finished their indicators in

[105] The letters in "Fair" being uncapitalized

2020, FAIRsFAIR had already announced that they would be formally adopting them for use in their planned automatic assessment tool, F-Uji (L'Hours et al., 2020).

In the paper describing F-Uji, the authors contextualized their work, making the, by now, familiar claim that the FAIR principles were too vague to practically measure anything (Devaraju and Huber, 2021). And while they applauded the work of the RDA group in making some more concrete metrics, they insisted that the maturity indicators still were not enough to practically measure FAIRness. In FAIRsFAIR review of the RDA indicators, they claimed the indicators lacked "clarity and context" and that their current form could only support "human evaluation" (L'Hours et al., 2020). They later argued that this would make evaluation according to the RDA's metrics un-scalable. Considering the "vast volume of data curated at data repositories", they argued, presented "a strong need for automated data evaluations" (Devaraju and Huber, 2021). So to realize this need, FAIRsFAIR had to make their own metrics.

The way Devaraju and Huber (2021) describe it, the FAIRsFAIR project started out with reviewing previous work on FAIR metrics (including the RDA group's), comparing them, combining them, testing them, and finally assessing them according to the RDA maturity indicators. But even though the FAIRsFAIR project claimed their metrics were strongly influenced by the RDA group's work, they differed substantially in practice. The RDA group's 41 metrics were reduced by FAIRsFAIR to 13 at first, but later expanded to 17; the RDA group's prioritization levels were replaced with an equal weighting of each letter of the FAIR acronym; and the RDA group's claim that FAIR could not be assessed by machine was ignored (Devaraju et al., 2020).

By modifying the metrics for FAIRness in this way, the FAIRsFAIR project was able to develop a set of machine-actionable tests of FAIRness, which in turn allowed for the development of an automated computer programme for FAIR assessment (L'Hours et al., 2020). This computer programme, F-Uji, enabled anyone to test the FAIRness of any dataset with an identifier (like a web-address or DOI) through a simple web interface (f-uji.net) – much like the FAIRmetrics group's tool. But, in spite of combining FAIRmetrics' machine assessment approach, and the RDA group's metrics, the completion of the F-Uji tool didn't settle the disagreements on how to definitively measure FAIRness either.

After the development and release of the FAIRmetrics assessment tool, the RDA FAIR Data Maturity Indicators, and FAIRsFAIR's F-Uji tool, the landscape of FAIR assessment had changed considerably. Anyone who would want to evaluate the FAIRness of a dataset would no longer have to base themselves off of just the notoriously ambiguous FAIR principles, but could make use of more clearly defined metrics or automatic tools. But the problem of FAIR metrics was never the lack of assessment methods. The problem of ambiguity – as formulated by all three groups (Wilkinson, Sansone et al., 2018; FAIR Data Maturity Model WG Case Statement, 2018; Devaraju and Huber, 2021) – still remained. While each group had found a solution to the ambiguity problem, neither aligned with the others, leaving the question “how FAIR is this data?” with several answers, none of them conclusive. Wanting convergence around a definitive method of assessment, why couldn’t the community around FAIR settle the issue?

The theory of valuation

Before I go on to discuss this question, we will have to go through some theory on how assessment, evaluation and *valuation* works. This brings us to the social science field of valuation studies.

Valuation studies is a particular strain of social science, perhaps even amounting to being a field on its own – it even has a dedicated journal, titled *Valuation Studies*. In the inaugural editorial article, the authors, Claes-Fredrik Helgesson and Fabian Muniesa (2013) pointed out that valuation is something we do every day, but is still a process that often eludes our understanding. Weiss (2017), trying to answer what value is, conjures the words of Augustine of Hippo: “If no one asks me, I know what it is. If I wish to explain it to him who asks, I do not know”.

In one sense, we’re so used to things being valued in the context of market capitalism: the price of the butter you buy in the supermarket, the value assessment of your possessions for mortgaging, the continuous calculation and fluctuating rates for the electricity in your sockets. Monetary and quantitative valuations are everywhere, and have become so ubiquitous that we sometimes can’t think of any other way of putting a value on something (Mau, 2019). The ability of valuation (in particular of the quantitative sort) to coordinate societies through a performance of objectivity has fueled the spread of Neo-Liberal policies (most notably New Public Management), but also motivated their critique (Johnsen, 2005; Lapsley, 2008). Because valuation allows us to arrange the elements of our reality according to worth, it is strongly tied to order and hierarchy (Lamont, 2012; Rottenburg et al., 2015). In other words, there is power in placing value on something (Bourdieu, 1993; Bessy & Chauvin, 2013).

But valuation does not confine itself only to markets and bureaucracies, in everyday life we encounter forms of valuation that sometimes defy the rational. The sentimental value we place on a dear possession, the worth we ascribe to a friendship, the “invaluable” feeling of watching the leaves turn yellow and marking the passing of seasons – they are intrinsically intimate, situated, contingent. As established institutionalized valuations often take the spotlight in critical scholarship (see Lamont, 2012; Mau, 2019), it's easy to forget that values only gain their aura of universalism and objectivity through a deletion of multiplicity (Latour, 1993).

This institutionalization of values is made possible through the repeated interaction of human as well as non-human actors – what valuation scholars sometimes call "valuation practices" (Dussauge, Helgesson and Lee, 2015; Waibel et al., 2021). But how do these fragile and situated interactions develop into stable values? With a focus on temporally and spatially delimited interactions and situated practices, this question has been a challenge for the field of valuation studies (Waibel et al., 2021).

In their seminal work on justification, Boltanski and Thevenot (2006) argue that valuations become stabilized through the arrangement of objects in relation to subjects (p.142). Later scholars (Kornberger et al., 2017; Waibel et al., 2021) develop this idea in combination with the field of infrastructure studies (see Star and Ruhleder, 1999; Larkin, 2013), making the point that objects not only make values durable, but also allow them to transgress their spatiotemporal situation, making values "travel". Infrastructures are thus seen as the material context in which valuations take place (Waibel et al., 2021).

Valuation scholars have generally referred to objects and infrastructures to solve the problem of durability, focusing less on how those objects gained their valuatory properties, or how those infrastructures came to be in the first place (see Kornberger et al., 2017; Waibel et al., 2021). The wider field of infrastructure studies in STS has been turning away from materiality, and shifted their focus towards practices; not how infrastructures are, but how they are being done; rejecting the notion of infrastructures and embracing infrastructur*ing* (see Star and Ruhleder, 1996; Karasti et al., 2016; Carse and Kneas, 2019).

In this chapter I therefore aim to incorporate a more praxiographical approach (Karasti et al., 2017) into valuation studies through an empirical study of valuation in the context of the FAIR principles, going beyond the study of valuation in delimited situations (Waibel, 2021).

FAIR is a particularly good case for exploring the infrastructuring of valuations, as the concept of FAIR is still an unstable construct. Valuation in the context of FAIR, therefore, needs to be done from the ground up; meaning each step of the process is open for contestation, and therefore well suited for study (Jasanoff, 2019).

Analytically, I'll draw on Waibel, Peetz and Meiers (2016; 2021) valuation constellations, which not only serves as a theoretical concept, but also a heuristic "highlighting the interconnectedness of valuations" (p. 47). The basic idea behind valuation constellations is that valuation does not happen as an isolated occurrence between only a valuator and a valuee, but is always a part of an interconnected constellation beyond just two actors.

Waibel et al. (2021) expand the dyad of valuator and valuee into a triad, where the third party is "the audience". According to the authors "it is often the audience that stabilizes valuations as it forces the valuee to not only recognize the valuator's public judgment but to adapt to it" (p. 38). In their valuation triad, the audience is important because it exerts the social pressure to accept the valuators valuation as valid or legitimate. The inclusion of the audience is also used to explain why a valuation can be seen as valid in one context, and invalid in another. If the audience changes, so does the outcome of the valuation.

In addition to the valuation triad of valuator, valuee and audience, Waibel et al. (2021) argue that rules and infrastructure matters. Infrastructure is, in their conception, the material context of the valuation. Rules are defined as formal or informal expectations that actors in the valuation triad have towards themselves or each other – in other words, the social context (Waibel et al., 2021). As we will see in the following paragraphs, the social and material contexts might be harder to separate than Waibel et al. (2021) suggest.

A lack of authority

As the observant reader surely would have noticed, the problem of measuring FAIR did not stem from a lack of methods. Even before the FAIRmetrics initiative, people were measuring the FAIRness of their data according to the original FAIR principles (Wilkinson, Sansone et al., 2018). Measuring FAIRness has then, strictly speaking, been possible since the FAIR principles themselves were published. At least anyone could *argue* that their data or repository or practices or whathaveyou was FAIR.

In fact, every initiative to create methods for measuring FAIRness acknowledged that the problem wasn't a lack of method for measuring, but rather that the existing ones tended to produce deviating and undesirable results (Wilkinson, Sansone et al., 2018; FAIR Data Maturity Model WG Case Statement, 2018; Devaraju and Huber, 2021). The problem with using the FAIR principles as a basis for evaluation, they claimed, was that it was seen as too interpretatively flexible. F-Uji. FAIRmetrics, and the RDA group agreed:

> Some of the FAIR principles are rather vaguely described (Devaraju and Huber, 2021).
> This ambiguity has led to a wide range of interpretations (Wilkinson, Sansone et al., 2018).

> As a result, a number of incompatible methodologies to assess FAIRness have been developed already (Herczog et al., 2018).

So when Wilkinson, Sansone et al. (2018) said they would "pursue the goal of defining ways to measure FAIRness", what they were saying was that they were pursuing the goal of defining *the way* to measure FAIRness. As authors of the original FAIR principles, the group expressed that, for them, making a set of metrics was "a natural and timely step" (Wilkinson, Sansone et al., 2018). Their metrics could – in other words – be seen as a natural extension of the original FAIR principles (Wilkinson et al., 2016), and, thus, the legitimate interpretation of them.

But the FAIRmetrics group realized that the legitimacy of their metrics would have to be grounded in more than their relationship to the original publication. After creating their first metrics, they, therefore, tested them on existing resources they considered to be relatively FAIR. As they put it "the objective was not to evaluate their resource, but rather, to evaluate the *legitimacy*, clarity, and utility of the metrics themselves." (Wilkinson, Sansone et al., 2018, emphasis added). But even as the test failed to create an agreement between the evaluators and the evaluee, the FAIRmetrics group continued to insist on their legitimacy. The source of the disagreement, they assessed, was not the illegitimacy, unclarity or non-utility of the metrics, but the lack of understanding among the resource owners (Wilkinson, Schultes et al., 2018).

As the FAIRmetrics group programmatic approach failed to produce assessments perceived as legitimate by the "audience" (in Waibel et al.'s terms (2021)), the group came to the conclusion that they not only needed to assess FAIRness, but also needed "[g]overnance over the metrics, and the mechanisms for assessing them" (Wilkisnon, Sansone et al., 2018). This governance organ, would, according to them, be required to settle "valid disagreements" (Wilkinson, Sansone et al., 2018). As if answering this call, the RDA FAIR Data Maturity Model Working Group sprang into existence only a few months after Wilkinson, Sansone et al.'s publication (2018), and with much of the same justification: solving FAIR ambiguity and resulting divergence. The RDA can be considered one of the most influential institutions in the field of research data – much thanks to their structure being built upon representation of interest groups from industry, policy and academia (RDA, 2016). In this way, the RDA theoretically manages to collapse the distinction between valuator, valuee and audience – as all parties participate in the same organ. Ideally, this would produce a stable valuation, making the RDA group a clear candidate for taking the role of the governing body suggested by Wilkinson, Sansone et al. (2018).

Perhaps disappointingly for the group of the original FAIR authors, it became clear quite early that the RDA group would not govern metrics according to the original FAIR principles. Following the RDA guiding principles of consensus and harmonization (RDA, 2016), the RDA group would rather base the metrics on practical interpretations of FAIR than the "conceptions of [the] initial creators" (RDA FAIR Data Maturity Model Working Group, 2019b). Authority, in other words, should according to the RDA group not have stemmed from the authors of FAIR, but from what they called "the communities". One of the main tasks of the RDA group therefore became to figure out what the communities wanted, or, rather, to establish a *consensus* on what the communities wanted (RDA FAIR Data Maturity Model Working Group, 2018).

But like the programmatic approach of the FAIRmetrics group, the RDA group's attention to consensus also failed to gain the approval of the audience. Where FAIRmetrics was too rigid to cater to the different needs and wants of various groups, the RDA metrics ultimately became too loose and flexible to fulfill the desires of specific groups. While the RDA metrics were flexible enough to be used across a range of contexts, not everyone agreed they were actually measuring FAIRness, or that they were practically applicable (RDA FAIR Data Maturity Model Working Group, 2020a). In other words, the RDA group managed to establish stable valuations within communities, but not across them. The solution to this was attempted by the FAIRsFAIR project through their F-Uji tool, combining the programmatic approach and closeness to the original FAIR principles of the FAIRmetrics group, with the community-based authority of the RDA. Still, the F-Uji tool's authority kept being questioned by people who thought it was either too lenient, too strict or not precise enough (EOSC Nordic, 2022a).

To explain why people struggled so much to establish a consensus on how to measure FAIRness, we can look back to the controversies around measuring a quite different entity: gravitational waves. In their seminal book on the production of scientific facts, *The Golem* (1993), authors Harry Collins and Trevor Pinch tell, among other stories, about Professor Joseph Weber's attempts at convincing the scientific community that he was the first to have measured gravitational waves. Not entirely unlike FAIR, up until 2015, gravitational waves was mainly an idea – anticipated, but not yet proved to exist. When Joseph Weber claimed he had been able to measure the waves with a sensor of his own design in the 1960's, physicists across the world were not entirely sure what to think. Had Weber actually measured gravitational waves, or was it simply noise produced by his sensor? Because nobody had seen gravitational waves or knew how they were supposed to behave, there was no way of knowing. People started building sensors of their own, but could not reproduce Weber's results. Why? Was it because their sensors were poorly made, or was it because Weber was wrong and never actually measured gravitational waves in the first place?

> To find this out we must build a good gravity wave detector and have a look. But we won't know if we have built a good detector until we have tried it and obtained the correct outcome. But we don' t know what the correct outcome is until... and so on ad infinitum (Collins and Pinch, 1993, p. 98).

This is a cycle of unknowns, called “the experimenters regress” by Collins and Pinch (1993). The only way it gets broken is when a “correct outcome” can be established with authority. Most of the time, the criteria for success is “known” beforehand, and the authority stems from “facts”, thereby avoiding this regress. In the cases where criteria for success are not known, the quality of the experiment must be defined otherwise, “independent of the output of the experiment itself” (Callon and Pinch, 1993, p 98).

In the case of Joseph Weber, the case was closed not on the basis of the existence of gravitational waves or the quality of the sensor, but rather when a colleague, Richard Garwin, mobilized his authority in the field to discredit Joseph Weber and his work. Thus, gravitational waves remained in the realm of theory – that was, until three independent sensors were able to produce converging measurements of two black holes colliding in 2015 (LSC, 2016). Because mainstream theory had predicted that the collision of two black holes would produce a burst of gravitational waves, the “correct” result was “known” beforehand. And because an array of three independent sensors were able to produce the same result, they validated each other as “good detectors”. Thereby, the researchers were able to escape the experimenter’s regress, and assess with authority that gravitational waves were real and measurable – they were a scientific fact (LSC, 2016)[106].

The production of knowledge and establishment of it as fact has been described within “the ontological turn” of social research as equal to a production of realities (Mol, 2002; Law, 2004; Woolgar and Lezaun, 2013). If production of fact is production of reality, then an authority to establish a fact could in effect be understood as an *ontological authority*. The experimenter’s regress is in this view not only a barrier to the production of scientific fact, but also a barrier to an actor or assemblage’s achievement of ontological authority. In the following paragraphs I will argue that we can see a similar dynamic in the case of FAIR, and that the lack of ontological authority is central to the challenges of establishing a FAIR evaluator.

[106] This claim has not been left entirely unchallenged though. In 2018, researchers at a Danish institute raised questions around the interpretation of the data from the sensor array. Had the researchers conjured a signal from just noise? The controversy around gravitational waves lives on (Brooks, 2018).

The evaluator's regress

In order to make this argument, let's have a look at one concrete situation of valuation – in particular the case of the FAIRmetrics group testing their metrics for the first time. They had just made an evaluator based on their interpretation of the FAIR principles, but they needed to test it to see if it worked – if they'd built a "good detector", so to speak. All the existing evaluators were producing different, or "diverging", scores, as they put it (Wilkinson, Sansone et al., 2018). This in itself was taken as a clear sign the scores weren't trustworthy, and couldn't be used as comparison for the group's own evaluator. Instead, the group ran an evaluation of a dataset, and asked the dataset's owner to discuss the "legitimacy, clarity, and utility of the metrics themselves" (Wilkinson, Sansone et al., 2018). In lack of legitimate or trustworthy evaluators, the group used the assessments of the dataset owners as comparison for their own evaluator.

Using Waibel et al.'s (2021) valuation constellation, we ascribe the roles of valuator, valuee and audience to the different actors in this valuation situation. We set the FAIRmetrics evaluator in place of the valuator and the dataset in the place of the valuee – but who was the audience? There are (at least) two answers to this. One answer is that the audience was both the dataset owners and the FAIRmetrics group, but as they weren't aligned in their assessment, categorizing them as a singular actor doesn't produce a stable constellation. An alternative answer is that there is not one, but at least two valuation situations: 1) where the audience is the data owners, and 2) where the FAIRmetrics group is the audience. In this case, we also have to add a third evaluation situation, where 3) the FAIRmetrics group is evaluating the dataset owners' valuation. This produces the following model (figure 4):

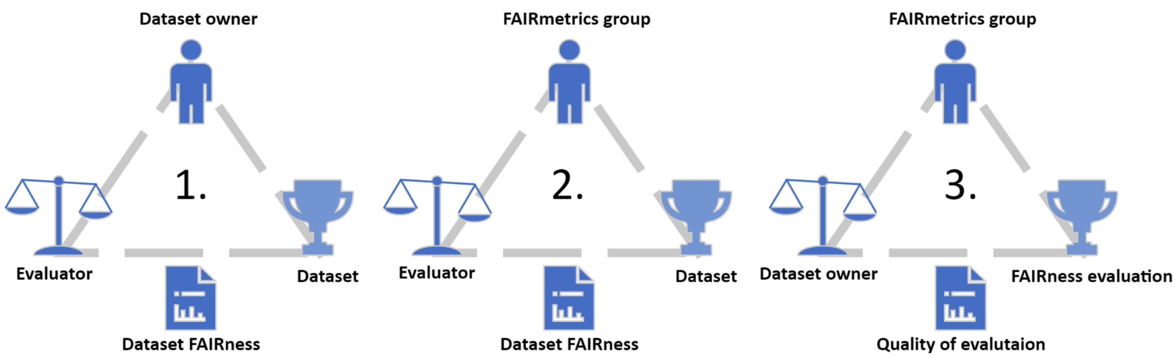

Figure 4

But there's more! "This process made it clear that certain metrics (and in some cases, the FAIR Principle underlying it) were not always well-understood." (Wilkinson, Sansone et al., 2018). This quote hints at the presence of an important element not included in the evaluation triads above – namely, the FAIR principles themselves.

To avoid branching off in endless valuation triads, I suggest a perhaps more helpful model based on what I'd call situational valuation *assemblages* (figure 5). Analyzing the same case in this way, we see that in the first situation (1.), the FAIRmetrics group, the data owner, the evaluator and the dataset produced a FAIRness score for the dataset. In the second situation (2.) – implied by Wilkinson, Sansone et al. (2018) – the actors performed a second valuation, producing two separate valuations of the dataset based on the actor's interpretation of the FAIR principles themselves. The FAIRness score of the dataset based on the evaluator (1.) and the interpretive FAIRness of the dataset based on the FAIR principles (2.) were then compared by the actors in a third situation (3.) in order to assess whether the first evaluation actually produced a result in accordance with the FAIR principles.

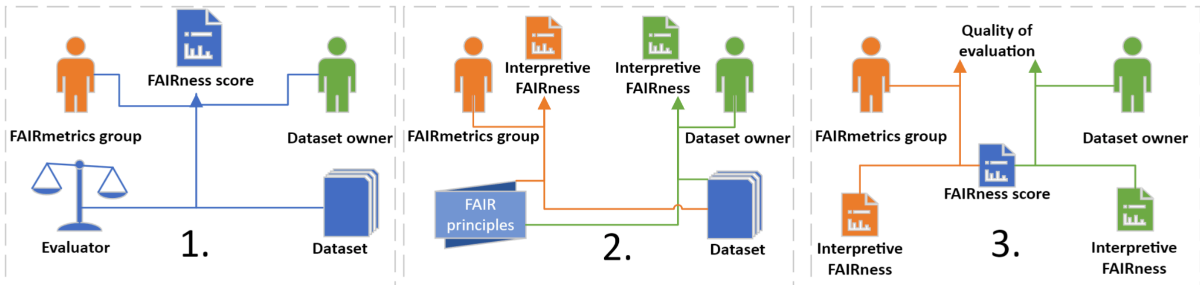

Figure 5

When we look at it this way, we see that the score in situation 1. was not an object of contestation. The actors agreed that this score was produced by this particular valuation assemblage. They both wanted to validate the legitimacy of this valuation though. This led to situation 2. where the evaluator was replaced by the FAIR principles, and the alignment between the actors was severed. What this means, is that they each had separate interpretations of the FAIR principles. Because they didn't agree on what the FAIR principles meant (either explicitly or tacitly), they produced two different FAIRness "scores" for the dataset. Both actors had then produced two valuations of the dataset each: one according to the evaluator, and one according to the FAIR principles. But remember, the aim was not to evaluate the dataset, but the evaluator itself (Wilkinson, Sansone et al., 2018).

In order to measure the quality of the evaluator, they did as the astrophysicists in Collins and Pinch's story (1993). In situation 3, they checked whether the result produced with the evaluator was the same as the result produced via their interpretation of the FAIR principles. Did the evaluator produce the expected outcome? The dataset owners came to the conclusion that it didn't – the quality of the evaluation was low. The FAIRmetrics group – having based the evaluator on their interpretation of the FAIR principles (which they themselves co-authored) – disagreed. To explain this discrepancy, they claimed the data owner did not understand the FAIR principles, or the evaluator for that matter. The FAIRmetrics group performed this process several times, and it kept producing discrepancies. Finally, they had to conclude "our evaluations were not 'fair'" (Wilkinson, 2019b). Why not?

The answer to why the FAIRmetrics evaluator kept failing at establishing ontological authority can be put briefly: nobody knew anything. Nobody knew how good the evaluator was. They didn't know because nobody knew what result it should have produced. Nobody knew what result it should produce because nobody knew how to produce an authoritative result. Nobody knew how to produce an authoritative result because they didn't have a good evaluator… and so on ad infinitum. In other words, they were going through what we can call an *evaluator's regress*.

When I say "nobody knew" here, it is important to note that what I actually mean is that nobody *agreed* what was correct, right or – in the last instance – true. This process just shows how FAIRmetrics – in this particular case – struggled to accumulate ontological authority by leveraging the process of evaluation on its own. We see that they realized this shortcoming, and that they tried to circumvent it, suggesting there was a need for a governing body that could make authoritative judgements to settle what they call "valid disagreements" (Wilkinson, Sansone et al., 2018)[107].

They never suggested who or what this governing entity should be, but it became clear that being the originators of the FAIR principles themselves did not grant enough authority. First of all, the FAIRmetrcis group, consisting of FAIR originators, evidently did not possess the authority to lay any claims on the ontological state of FAIRness. Barend Mons, the FAIR author par excellence, was also shot down by a member of the RDA metrics group when he suggested that any FAIR metrics should remain a close interpretation of the original FAIR principles (RDA FAIR Data Maturity Model Working Group, 2019b). If the FAIR principles didn't have enough authority, what would?

The question of where legitimate authority could or should come from, makes sense of the RDA group's approach of grounding their metrics in the existing community practices. Basing their metrics on already accepted practices and expanding them had the obvious benefit that the basis for evaluation would be "known entities". The practices were "known" to *be* FAIR by the people enacting them. If they could make a metric that would be able to measure those practices and produce a positive result, the metric would be valid. Already

[107] Presumably, they would also settle what made a disagreement "valid".

knowing the right answer was a clear benefit. So why did the RDA metrics still fail to accumulate the necessary ontological authority?

It turned out that the RDA group's strength was also their weakness. By aligning a wide array of practices into one set of metrics, it turned out they produced a compromise too big for some. Particularly the departure from machine assessment and interoperability as a core criteria became a camel too big to swallow for those who subscribed to a close interpretation of the original FAIR principles. According to them, whatever score the RDA metrics produced, it did not measure *true FAIRness*. Essentially, this was the same challenge that the FAIRmetrics group encountered, boiling down to differences in interpretations of FAIR. Only with the RDA, the roles were flipped in the sense that the data practitioners made the metrics, and the principlists rejected them. In the end, the result was the same: evaluator's regress, failing to achieve ontological authority.

The value and limitations of metrics

If you have any experience with measuring stuff, you might have noticed that ontological authority is far from necessary in most situations. When we're reading recipes for cooking, we encounter pinches of this and heaps of that all the time. We don't need to know exactly how big a chunk of butter is supposed to be to make a delicious stew. But if we're trying to make an eclair, we can't trust that my handful of flour is the same quantity as the recipe author's handful of flour. That is why recipes in need of precision usually provide the measurements in a metric scale[108]. Making eclairs, we might therefore prefer to delegate the measuring to our trusty kitchen scale.

The kitchen scale is for most people the most precise way to measure the mass of anything. In the public sphere, the deceivingly simple act of weighing something can quickly become more complex. If you buy anything by weight – like fruit, veggies or nuts – you have to trust that the grocer is not cheating you (on purpose or unknowingly). Therefore, scales used in legal transactions (like of money and apples), must be certified by an official organ, like an institute of weights and measures[109]. Still, only the stingiest among us would care about the grocer's scale's precision to the milligram. But for physicists, scales that are not precise to the milligram or even microgram can be a disaster. Knowing that *the kilo* itself had been losing approximately 50 micrograms over the years since it was enshrined in its three bell jars, was enough for physicists to call for the redefinition of the kilo that we finally got in 2019 (Georgia Institute of Technology, 2007; Sample, 2018).

[108] With the imperial units of measure as exceptions, naturally
[109] Look for a certification sticker on your grocer's scales the next time you're out shopping

As we can note from the example of the humble kilogram: the precision we demand of a measuring device depends on the context. More precisely, the authority of the measurement has to match with the practical ontology of the situation. When we measure a kilo of flour on our kitchen scale, we shouldn't be fooled to believe that the flour is an *actual* kilo in the sense of a universal kilo. It is probably off by some milligrams or even entire grams, but we don't care. For our intents and purposes, that heap of flour *is* a kilo. If we bring the same amount of flour to our grocer to trade, it doesn't matter what our kitchen scale said – in that situation, the grocer's scale's word is law (quite literally). If the scale says 990 grams, in that situation the flour *is* 990 grams regardless of how many grams it was at home on the kitchen counter. At the grocer's, the kitchen scale is virtually useless. In the same way, the grocer's scale is useless in a high-precision physics experiment. What we draw from this is that the ontological authority of a measurement depends on the demands of the situational ontology.

In the context of FAIR evaluators, we can recognize a similar dynamic. The people who had participated in making metrics for FAIR had realized that their evaluators and metrics couldn't actually measure FAIRness with a universal ontological authority. In this sense, it was equivalent to the kitchen scale in the above illustration. This led them to reject the notion that an object's FAIRness could be settled across all situations (or "for all times and all people", if you will) (Wilkinson, Sansone et al., 2018; Bahim et al., 2020; EOSC Nordic, 2022a). Did that mean that the metrics and evaluators were useless though? No. They had to be useful for some purpose, but for what? This question had not been addressed for a long period of time, as it seems everyone just assumed that they were making metrics for the same purpose: to measure FAIRness to the end of increasing FAIRness. This was, of course, at best, an imprecise assumption.

The tacit imagined purposes of FAIR evaluators became explicit perhaps most clearly when it came to the topic of FAIR scoring.

Running an assessment through the FAIRmetrics tool would essentially just provide a list of the indicators tested against, and show which tests were passed and which were failed. A FAIRmetrics score could therefore be represented as 10/22, but would not tell you the general FAIRness of the dataset – that judgment was reserved for whomever/whatever was running the test.

The RDA group did not want to leave the judgment of FAIRness to an external evaluator, and put a sizable effort into designing a scoring system for their metrics. The first suggestion from the editorial group was a tiered system of five levels, where the FAIRness level would be assigned according to the proportion of mandatory, recommended and optional criteria achieved:

> Level 0 : The resource did not comply will all the mandatory indicators
>
> Level 1 The resource did comply with all the mandatory indicators, and less than half of the recommended indicators
>
> Level 2 The resource did comply with all the mandatory indicators and at least half of the recommended indicators
>
> Level 3 The resource did comply with all the mandatory and recommended indicators, and less than half of the optional indicators
>
> Level 4 The resource did comply with all the mandatory and recommended indicators and at least half of the optional indicators
>
> Level 5 The resource did comply with all the mandatory, recommended and optional indicators
>
> (Bahim, 2019)

This immediately caused a stir. Because the RDA metrics were based on *achievable* FAIRness and not *ideal* FAIRness, Rob Hooft – one of the original authors of FAIR – argued that such a scoring mechanism would not provide useful feedback on how to improve the FAIRness of the evaluated dataset (Hooft, 2019a). Instead, he suggested that each letter of the FAIR acronym should get a "3-number tuple, indicating the percentage of mandatory, recommended, and optional indicators met, maximized to 99.99.99." (Hooft, 2019a). In this way, the feedback from the evaluator would be more granular, but as a drawback, it would also make it less clear what the overall FAIRness of the dataset would be. This comment sparked a long conversation about how the RDA score should look like, and why. Some thought it was overly artificial to separate the F, A, I and R (Jeffery, 2019b; Khalsa, 2019), others were afraid that the scoring could be interpreted as a pass/fail scoring system, as data with a non-perfect could be seen as "not FAIR" (Dekkers, 2019c). Four months into the discussion (after a workshop meeting), Mark Wilkinson (of the FAIRmetrics group) pitched in, saying that

> In my opinion, an overall FAIRness score is not a useful measure, for many of the reasons mentioned above, but in addition…
>
> In the paper on the FAIR Evaluator, we suggested to approach FAIRness testing from the perspective of contracts between a data provider, and a data consumer - promises that both humans and machine users can rely on as they attempt increasingly complex data transactions with a provider. From this perspective, a "FAIR score" is a totally meaningless artefact, since it tells me nothing about the behaviors of that provider that I can code my agent to expect (Wilkinson, 2020a, underscore added).

Makx Dekkers promptly responded:

> @markwilkinson
> Yes, in the perspective you depict, I agree that a score doesn't tell you much.
> However, there are potential users of the indicators, for example funding agencies, who may want to verify that the data produced by projects they fund complies with a certain level of FAIRness. Such a "FAIRness score" should be accompanied by the more detailed observations on the indicators, so an evaluator can see which indicators have been satisfied and which have not (Dekkers, 2020a).

From the discussion above, it became clear that the two actors, Wilkinson and Dekkers, had different expectations as to what situations the metrics would be used in. Dekkers' suggestion that the indicators would be used by funders in a situation of evaluation was not welcomed warmly by the members of the group. From the beginning of the discussions it had been claimed by Christophe Bahim from the editor's team that "The aim of this evaluation is NOT to pose a judgement but instead objectively score a resource to identify ameliorations" (Bahim, 2019). This was, as noted on several occasions, the consensus in the RDA group – that the purpose of the evaluators and metrics were to improve FAIRness (Dekkers, 2020b; L'Hours et al., 2020). When the notion of funders using the RDA metrics to judge the FAIRness of datasets was put forth, it triggered a less than enthusiastic response among several of the group members. Among them, Ruth Duerr stated:

> @makxdekkers Uff...! The example you just gave leads me to think about the current practices of publication metrics which are indeed used in just such ways, much to science's detriment in my opinion. Do you really want to create such an environment?
> (1👍 markwilkinson) (Duerr, 2020)

Mark Wilkinson added: "Indeed.... in fact, I think we should be DISCOURAGING the idea of 'FAIR Scores'!" (Wilkinson, 2020b, emphasis in original). Drawing on the debates around indicators and metrification elsewhere in academia[110], some of the people working on FAIR evaluators started worrying they could become part of the problem. Experiencing how the employment of metrics in governance could produce what they considered perverse ontologies, they were reluctant to allow their metrics to be used by funders. In an attempt to respond to this concern, Dekkers said that the group would "abandon the idea of an overall score" (Dekkers, 2020c) – thus making the metrics virtually useless for research funders like his employer, the European Commission.

This reluctance towards employing metrics in the context of governance also became a topic of discussion during a conference on the "value and limitations of assessment tools" organized by the EOSC Nordic project (2022). In their keynotes, Mark Wilkinson and Robert

[110] See the DORA declaration (n.d.) and Leiden Manifesto (2015)

Huber (from F-Uji) claimed that their tools were not good enough to be used to assess the true FAIRness of datasets. The evaluators, according to Mark, did “not work as a scoring system”, the reason being that “there’s a lack of agreement in interpretation”.

In spite of this general agreement on the limitations of evaluators, the conference participants still argued they had value (in line with the title of the conference). While people were uncomfortable with the evaluators making definitive and universal claims on FAIRness, they were open to them making more relative and situated claims.

What’s the use of useless metrics?

There were two important experiences that had demonstrated this relative value of FAIR evaluators. The first one was a series of tests conducted by the EOSC Nordic, and another by the ELIXIR FAIRplus project.

As a part of their project to support the realization of the European Open Science Cloud in the Nordic, the EOSC Nordic project wanted to figure out the FAIRness of the data repositories in their region. After compiling a list of repositories that could be tied to their geographical region, they started running tests on the datasets in the repositories using the FAIRmetrics tool, which produced an average score of 30% FAIRness[111] (EOSC Nordic, 2022b). Generally speaking, the project did not consider 30% a particularly high score. Neither did they consider the score itself particularly useful either. And upon repeating the test six months later, they found the score had changed to 35%, indicating an increase in the general FAIRness of the repositories. This change surprised the EOSC Nordic team, not because of the change itself, but since they “knew that some repositories had significantly increased the FAIRness of their metadata during this period” they had expected the increase to be *more significant* (EOSC Nordic, 2022b). This discrepancy between scoring and “knowing” made them look into what could have gone wrong, and, after some inquiry, they learned that the FAIRmetrics evaluator itself had been changed over the same timespan.

While the instability of FAIRmetrics proved to be an ongoing problem for the EOSC Nordic, the F-Uji developer team reached out and asked if the EOSC Nordic wanted to test their new evaluator. While moving over to the F-Uji tool produced consistently lower FAIRness scores for the tested repositories, being able to measure them with a fixed and stable evaluator proved useful in establishing that there had been a positive development in the FAIRness of the nordic repositories over time (EOSC Nordic, 2022b; Nordling et al., 2022).

[111] This percentage score was generated by translating the n/N score produced by the tool (Jaunsen et al., 2020)

In the other case, the ELIXIR FAIRplus project had set out with a goal to increase the reusability of their data holdings (ELIXIR, 2020). This started out with a collaboration with the RDA group to test out the usability of their metrics. The FAIRplus project also quickly realized that the metrics couldn't give "objective" scores to the datasets. Still, by using the metrics as a guide for their own local assessment, they could identify which of the criteria the relevant datasets met, and to which degree. Instead of using the insight from this process to establish the quality of the datasets, they used it as a starting point to identify what they should work on to improve its quality according to their own judgment (RDA FAIR Data Maturity Model Working Group, 2019g).

Taking the feedback from the evaluation, the FAIRplus project would then try to increase the amount of assessment criteria met, and then measure the dataset again after the attempted improvement to see if their work had the desired impact. What the tests showed was that the datasets would increase their FAIRness score, and thereby, the FAIRplus project believed, their reusability (RDA FAIR Data Maturity Model Working Group, 2019g). For research infrastructures like ELIXIR, this was considered useful, as data reuse was one of the key performance indicators they were measured according to by their funders in the EU (ESFRI, 2019). The FAIRness of the datasets, in this situation, was therefore not the goal itself, but the means to achieve that of reuse, indicating a belief that a FAIRer dataset would be reused more.

The two cases of EOSC Nordic and ELIXIR FAIRplus showed that while the FAIR metrics couldn't be used to establish authoritative measures across situations, they could be used within them. Even if "objective" FAIRness couldn't be measured, the FAIR metrics could be used to measure a comparative change in FAIRness. This "solved" the evaluators regress since the process didn't necessitate a known value, only an expected change of value. In both cases both the resource owners and the human evaluators expected the scores to increase after improving the dataset or repository, and as long as the predicted discrepancy between pre and post improvement matched the discrepancy in scoring, the evaluation method was considered legitimate and authoritative. If the prediction and result did not match, as happened with EOSC Nordic's use of FAIRmetrics, the evaluator would not be considered legitimate. With reference to figure 6, the human evaluators and the resource owners were able to create stable evaluations, only because they were able to agree on the predicted discrepancy between the original and modified resources' FAIRness.

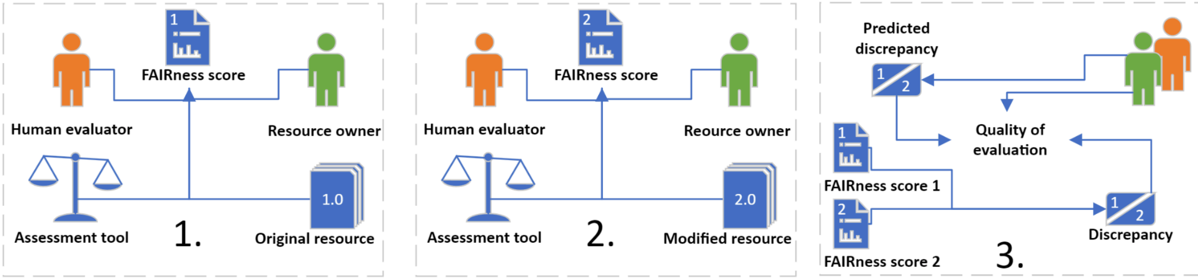

Figure 6

To begin with (1.), the resource owner, the human evaluator, and the assessment tool created a FAIRness score for the original resource. After establishing the first score, the resource owner improved the resource, and the actors came together (2.) to create a new FAIRness score, now with the modified resource. Next (3.), the human evaluator and the resource owner came together to compare the difference between the two FAIRness scores with the difference they expected. As long as the expected and measured differences were somewhat in agreement, the evaluation was considered legitimate.

The usefulness of the FAIR evaluators in this situation can be compared to a kitchen scale you know is uncalibrated. It won't be useful to measure out a kilo of sugar, but will be useful if you want to check whether your kids have eaten any of it. In other words, the evaluators didn't have authority to say "this is (not) FAIR", but they did have the authority to say "this is more/less FAIR than it used to be".

Realizing this limited, but real, usefulness of FAIR evaluators led Mark Wilkinson to suggest that the focus of FAIR evaluation should "Evolve away from 'FAIR Assessment' to 'FAIR Assistance'" (Wilkinson, 2022a). This notion gained a significant popularity during the EOSC Nordic seminar on the value and limitations of assessment tools, leading the seminar to conclude that the evaluators' value was in assistance, and that their limitations were their inability to assess FAIRness (EOSC Nordic, 2022b).

Since an evaluator's usefulness is directly linked to its authority, claiming that the usefulness of evaluators was limited was, in effect, also an effort at *making* them useless in any other context. While this limited authority had been accepted and appropriated by the academics of the FAIR evaluation community, those representing the policy sphere (more specifically, the EC) had a harder time accepting it. The RDA metrics that were supposed to be adopted by the EC at finalization, became downgraded to "a necessary first step towards building a sound methodology to assess the FAIRness of publicly funded research outputs" (Bahim et al., 2020).

Successive reports from the EC and EC-funded projects repeated that existing metrics had a limited usefulness for funders (Directorate-General for Research and Innovation et al., 2021; Mustajoki et al., 2021; Lacagnina et al., 2022). To make evaluators useful for funders, the EC's three most central elements on their wishlist were convergence between metrics, acceptance in the scientific community, and scalability through automatic assessment (Directorate-General for Research and Innovation et al., 2021; Mustajoki et al., 2021; Lacagnina et al., 2022)[112]. This sustained desire to develop metrics for governance purposes worried some of the people who had been working on FAIR metrics, leading to a publication of their concerns in an open discussion paper stating:

> The community is genuinely concerned that mechanisms to evaluate FAIRness can be misused and misinterpreted, especially when these become a decision-making instrument in funding scenarios (Wilkinson et al., 2023).

So, while the makers of the FAIR metrics had seemed to have settled on assistance rather than assessment, the threat of perversion still loomed.

FAIR enough?

This chapter started out with a short story about my struggles with assessing FAIRness, and the question that arose from it: why is it so hard to measure how FAIR something is? Looking into the history of FAIR evaluation, we not only find an answer to this question, but also insight into the dynamics of measuring in general. What was it that really happened between 2016 and 2022?

In 2016 the field of FAIR evaluation was fragmented. A small amount of evaluation tools – mostly in the form of self-assessment questionnaires – were circulating within certain contexts. They might have been seen as legitimate authorities on FAIRness within their contexts, but actors in the FAIR community realized these assessment methods would not produce results that would hold up as legitimate across different settings (Wilkinson, Sansone et al., 2018; EC, 2018). In the terminology of valuation studies, they were not able to make these valuations travel across contexts.

The main reason these valuations could not travel was because they didn't produce the same results as each other. In practice, this lack of alignment meant that the valuations were

[112] The method the EC uses to ensure researchers' compliance with the FAIR principles is through a human estimation of FAIRness according to a mandatory data management plan (Bahim et al., 2020). What is measured then is not FAIRness itself, but adherence to a promise of actions and behaviors listed in a document.

in contestation with each other – and none of them were able to achieve a level of authority needed to settle this contestation.

So in around 2018, a handful of actors tried to solve this stalemate by mobilizing different forms of authority: the FAIRmetrics group, representing the original FAIR authors (Wilkinson, Sansone et al., 2018) – and the RDA FAIR Data Maturity Model Working Group (Herczog et al., 2018), representing the wider scientific community. They were both relatively successful at making their valuations travel, but neither could achieve the level of authority needed to reach their goal: to settle the question of how to measure FAIR across contexts. The later initiative of the FAIRsFAIR project tried to solve this by combining the two approaches – representing both the FAIR principles *and* the scientific community – but to no avail.

The reason these actors failed to achieve the necessary authority stemmed from a problem I have called the evaluator's regress. Neither could "prove" that their methods would produce a valuation according to "true FAIRness" because there existed no agreement on what that would look like. Their valuations could be validated if their outcomes matched with a known outcome, but no such known outcome existed. As the community described it, the FAIR principles were too ambiguous to give any correct answer (Wilkinson, Sansone et al., 2018; Herczog et al., 2018; Devaraju and Huber, 2021), and there existed no calibration dataset (EOSC Nordic, 2022a). Since nobody could overcome this evaluator's regress, nobody could claim to measure true FAIRness – they couldn't achieve ontological authority.

Being aware of the potential of valuations to create order – sometimes perverse – the creators of these metrics and evaluators became reluctant to allow journals and funders to enforce these measures. Therefore, they undermined the legitimacy of their own and other's valuations – saying that the measurement of FAIR was not possible, at least not yet (EOSC Nordic, 2022a)!

This did not mean that the metrics were useless though. Experience from tests of the different metrics, showed that people appreciated their ability to guide improvement of digital resources like data (FAIRplus, n.d.) and data repositories (EOSC Nordic, 2022b). This created a shift in the community of FAIR evaluation "away from 'FAIR Assessment' to 'FAIR Assistance'" (Wilkinson, 2022a). The European Commission, wanting to "assess" the FAIRness of the research they funded, had to accept the community's lack of faith, and postponed the employment of FAIR metrics until they would be able to produce an assessment that would be considered legitimate and authoritative across contexts. Whether this will happen is still an open question, but there are some barriers that become clear to us: the tension between FAIR in principle and FAIR in practice, and the evaluators regress.

The first thing that becomes evident to us is that the ontological state of FAIR is in tension. FAIR *is* the principles described in the original publication from 2016, and FAIR *is* the things people do when they try and believe they are FAIR. In other words, FAIR is both principles *and* practice. The problem arises when these two states of FAIR fail to align. We see this happen over and over again during the design and testing of the different FAIR evaluators, perhaps most strikingly with the FAIRmetrics group and the RDA group. Each group started out with the assumption that either principles *or* practices should be the basis for FAIR evaluation, and both met resistance when the other was introduced. The FAIRmetrics assemblage failed at the introduction of practices. The RDA assemblage failed at the introduction of principles.

The F-Uji tool is an interesting case, in that it tried to make both principles and practices the basis for their tool, but failed at the encounter of either. The principlists rejected it on the basis of interpretation of the original publication, and the practicologists rejected it on the basis of applicability to existing practices. The ontological ambivalence of FAIR thus remained unresolved.

This ambivalence both contributed to, and remained unresolved because of, the second barrier: the evaluator's regress. Whether they were predicting contestation or experiencing it, the makers of the different evaluators attempted to gain authority. A strategy they employed to gain authority was to demonstrate the legitimacy of their metrics and evaluators through a series of evaluation trials that ultimately failed to make their evaluations travel. In their attempts to improve the metrics, the makers found themselves grasping at straws, because how could they possibly prove that their method produced the most true results?

As Collins and Pinch (1993) note, proving anything at all can be a tricky task. The most important element to effectively settle any claim of proof is the presence of known entities – like knowing the correct outcome of a test, which can be obtained by knowing the quality of a test or testee. As we've seen over and over again, FAIR had no such known entities. As Robert Huber of the F-Uji tool said during the EOSC Nordic conference (2022a), the problem was a lack of standard samples and calibration procedures. Without these elements, none of them could prove their ability to discern "objective" FAIRness from situational FAIRness. With no known entities, there will likely be no measure of FAIRness for all times nor all people. If only someone would find a way to define FAIRness according to the Planck constant.

Chapter 6: infrastructuring FAIR

In September 2014, some hundred members of the newly formed Research Data Alliance were meeting for their fourth plenary meeting in the Dutch capital of Amsterdam (Meleco, 2014) for what was presented in the opening welcome as “the biggest data event ever on Dutch soil” (RDA, 2014a, 1:32). In the first session, after the welcome, a keynote from Director General for research in the European Commission, Robert-Jan Smits and a video message from Vice President of the European Commission, Neelie Kroes, it was time for the head figure of the FAIRport initiative, Barend Mons, to come to the stage (RDA, 2014c). While on his way up to the stage, the chair of the session, friend and colleague of Barend Mons, Ingrid Dillo, quipped: “knowing Barend, I’m sure we’re in for a very provocative talk” (RDA, 2014a, 33:33).

In his keynote, titled “Bringing data to broadway”, Barend Mons laid out his ideas about the current and future state of science. Currently, he proclaimed, we were producing more knowledge than ever, but we were unable to make good use of it. The way he presented it, science was stuck in the pre-modern era of knowledge sharing and needed to move on. The next step would be “the eScience era” (RDA, 2014a) where machines would take a more central role and enable new forms of knowledge production. Because of this shift from human to machine, Barend Mons envisioned that “In the future we will not have supplementary data for papers, but supplementary papers for data” (RDA, 2014a).

In this future, where data would become “first class citizens” (RDA, 2014a), he continued, machines would be able to analyze and combine datasets too big and complex for humans to deal with, producing previously inaccessible knowledge. One of the main barriers standing between this future and contemporary science, he claimed, was that research data was not machine readable, because it was not Findable, Accessible, Interoperable and Reusable – or FAIR. To solve this, he envisioned the creation of an infrastructure for research data, so called “FAIRports”, and encouraged everyone in the audience to support the initiative (RDA, 2014a).

Perhaps as an attempt to make his talk less provocative than Ingrid Dillo had predicted, Barend Mons was sure to add that the infrastructure he suggested would only be a minimal intervention:

> We have a very light set of principles, and as long as people keep these principles: whatever technology they use, we couldn't care less (RDA, 2014a, 1:02:00).

And as for the viability of this light set of principles, he assured the audience that this had been done successfully before:

> This is analogous to the internet: TCP/IP for data - the very tiny thing in the middle that everybody has to adhere to." (RDA, 2014a, 1:02:15).

Either his words were successful in soothing the audience, or the talk was not perceived as that provocative to begin with, because the only comment it garnered from Barend Mons' colleagues was from one Jeffrey who felt like correcting one of Barend Mons' earlier claims that DOIs were charged per piece. Supposedly, charging per DOI was only a crossref practice (RDA, 2014a, 1:15:40). No further comments.

Only the next day did it become clear that there was at least one person who did find Barend Mons' talk somewhat provoking: data scholar Christine Borgman. In her keynote the following day, she made sure to address Barend Mons' presentation and his vision for the future of science. She agreed with him that it was a "critical time to be building this new infrastructure" for research data, but added a warning:

> The challenge is that we need to build an infrastructure around the way that the scholars and the research community really work, not the way we would like them to work, not the way we think they ought to work, but the way they actually do work (RDA, 2014b, 22:58).

In this comment, she was drawing on the field of Science and Technology Studies (STS), and a particular branch of it called infrastructure studies. In STS, infrastructures have long been objects of interest. Although the explicit focus on infrastructure as a concept only really took off in the mid-to-late 90's with Susan Leigh Star, Karen Ruhleder and Geoffrey Bowker (Star and Ruhleder, 1994; 1996; Bowker and Star, 1998), early STS was arguably already starting to notice the pervasive power of infrastructures, with Thomas Hughes study of Large Technical Systems being a notable example (1983).

What scholars of infrastructure studies noticed, and Christine Borgman alluded to, was that infrastructures are far more than purely technical enablers of action and interaction – they are also expressions and enactments of politics and power (Winner, 1980; Star, 1990; Star and Ruhleder, 1996; Bowker and Star, 1998; Woolgar, 1999; Larkin, 2013). Put in other terms, infrastructures do not only ease the movement, communication or interaction within the world, but create entirely new worlds (Larkin, 2013; Jensen and Morita, 2015).

Trying to overcome the barriers between the present world of local practices and the future world of infrastructure, actors therefore inevitably encounter conflict, opposition and resistance (Star, 1990; Ribes and Bowker, 2009; Ribes, 2017; Timmermans and Epstein, 2010; Edwards, 2010; Edwards et al., 2011). In this chapter of my thesis I interrogate the infrastructural visions of the FAIR movement and their strategy to overcome the hurdles between them and a FAIR future as an infrastructure in the making.

Making and doing infrastructure

As stated in the introduction, the study of infrastructures has for long been a popular pastime in science and technology studies. But whereas earlier works mainly considered infrastructures an empirical category, of, for instance, roads, bridges and electrical networks (Winner, 1980, Hughes, 1983), I will focus on the later works that developed infrastructure as a conceptual category beyond its material and nominal grounding.

The emergence of infrastructure as a concept largely grew out of the need to define what infrastructures are at a more basic level – the essence of infrastructure if you will (Star and Ruhleder, 1994; 1996). The earlier infrastructure studies took this category more or less for granted – treating something as an infrastructure if it already was named as such by others. It was with the increasing prevalence of the less tangible “information infrastructures”, or more specifically “computer supported cooperative work”, that the field of infrastructure studies first started coming into its own as a study of phenomena *as infrastructure*, instead of infrastructure *as phenomena* (Star and Ruhleder, 1994).

In their seminal work on infrastructures, Susan Leigh Star and Karen Ruhleder (1994; 1996) took a marked distance from the study of infrastructure as merely an assemblage of material artifacts, rather suggesting that “infrastructure is fundamentally and always a relation, never a thing” (Star and Ruhleder, 1994). This did not mean they denied the material characteristics of infrastructure, but that they wanted to highlight the connections and configurations necessary to turn materials into infrastructure – in their words, infrastructure “is product and process” (Star and Ruhleder, 1996).

But what makes an infrastructure then? According to Star and Ruhleder (1996), while the embodiments and enactments of infrastructure vary, “an infrastructure [is something which] occurs when the tension between local and global is resolved“ (Star and Ruhleder, 1996). Noticeably, their definition is not one of *what* an infrastructure is, but rather *when* an infrastructure is (or, in their words, when it “occurs”). This shift from defining what an infrastructure is to what an infrastructure does – sometimes called an “infrastructural inversion” (Bowker, 1994) – gives us an understanding of infrastructure as temporally situated, processual and continuous. Infrastructural inversion then draws attention to, not only the making and establishment, but also maintenance and repair (and eventual demolition) of infrastructure (Carse and Kneas, 2019).

Applying this understanding of infrastructure, Ashley Carse and David Kneas (2019) point out that infrastructures rarely occur in full and finished states. Instead, they suggest, unbuilt and unfinished infrastructures are the norm (Carse and Kneas, 2019). Since subjects of infrastructure tend to resist infrastructuring efforts, Carse and Kneas (2019) argue, infrastructures are generally, at best, only partially able to resolve the tension between local and global, and thus remain ever unfinished. Finished infrastructures, the argument follows, only truly exist in fiction (Carse and Kneas, 2019). Infrastructure, Carse and Kneas argue (2019), is thus in a state of perpetual becoming, and, as such, we can only talk about it as infrastructur*ing*.

While this position can come off as unhelpful for those who wish to study infrastructures, Carse and Kneas (2019) argue for its positive methodological implications. Seeing infrastructures as something that is indefinitely in the making allows for, in particular, two fruitful entranceways for interrogation. First, as infrastructuring tries to resolve the tension between local and global, it produces conflict and controversy, thus revealing the implicit, tacit and otherwise invisible (Carse and Kneas, 2019). In this sense, Carse and Kneas make the same argument as controversy studies (see Jasanoff, 2019) and the study of developing things (Jensen, 2010). Second, seeing infrastructures as "becoming" draws attention to the fiction of the finished infrastructure (Carse and Kneas, 2019). Asking: "becoming what?", enables us to interrogate what kinds of arrangements and orders infrastructuring aims to produce (Carse and Kneas, 2019) – using the vocabulary of Madeleine Akrich (1997), looking at infrastructuring gives us the ability to read the "script" of the finished infrastructure.

Taking these two objects of study together – the practical outcomes of infrastructuring and the aspirations of infrastructure – further elucidates the tension between local and global, and the process of resolving it. This tension is particularly salient in the aspiration towards interoperability, which, in its most basic sense, means the ability something has to operate across (inter) two or more contexts (Ribes, 2017). In this sense, it is a core quality of infrastructure's ability to transcend local contexts and establish global practices (Star and Ruhleder, 1996).

The study of interoperability as a topic of interest in its own right, though, has only more recently come to the fore in infrastructure studies, spurred on by the increased pervasiveness of information infrastructures, computational modes of knowledge production and, in particular, the challenges surrounding the sharing and reuse of digital data in research settings (Ribes, 2017)[113].

[113] As discussed in chapter one, this development is not all that new, representing a continuation of efforts and discussions that began in the mid 80's.

As initiatives for data interoperability have proliferated, scholars of science and technology studies have found that efforts to achieve it are often plagued with what Paul Edwards (2010) terms "friction" (Ribes and Bowker, 2009; Edwards, 2010; Edwards et al., 2011; Poirier, 2019a; Ribes et al., 2019; Aula, 2019) – a process which "consumes energy and produces turbulence and heat – that is, conflicts, disagreements, and inexact, unruly processes" (Edwards et al., 2011). This friction does not necessarily entail a failure to interoperate, but, rather, an added layer of exertion, due to "the stark contrast between the goals [...]and the existing institutional realities" of these projects (Aula, 2019).

Overcoming these contrasts between goals and institutional realities will, according to the literature, necessarily be a balancing act between giving flexibility to the individual actor on one end and enforcing strict standards on the other (Hanseth et al., 1996; Timmermans and Epstein, 2010; Edwards et al., 2011; Poirier, 2019). Through flexibility, the infrastructuring deals with incongruencies at a local level, moving the task of interoperation away from standards towards individual actors. Through successful standardizing, on the other hand, infrastructures avoid incongruencies in the first place by deleting variance (Hanset et al., 1996; Timmermans and Epstein, 2010; Edwards et al., 2011; Poirier, 2019).

While standardizing action and interaction is often promoted as an effort to make knowledge production more effective or efficient, scholars have pointed out that standardizing, itself, can be a source of friction, dismay and resistance (Star, 1990; Bowker and Star, 1999; Borgman et al., 2014). The issue, they point out, is that standardizing necessarily deletes multiplicity and diversity in favor of common denominators (Karasti et al., 2016). On the other hand, too much flexibility leads to useless standards and impotent infrastructures (Timmermans and Epstein, 2010). With this in mind, designers and makers of infrastructure have to make the decision between strict enough standards to make a difference, and enough flexibility to actually work.

In the field of machine reasoning (sometimes under the moniker of "AI"), this dilemma between flexibility and strictness has been a persistent challenge – a gordian knot so to speak (Poirier, 2017; 2019a; 2019b). The way STS scholar Lindsay Poirier puts it, "designers have no choice but to consider trade-offs when it comes to deciding upon the complexity/restrictiveness, flexibility/hardness, scruffiness/neatness of their standards and frameworks" (Poirier, 2019a). This has not kept people from trying to work around the issue (Poirier, 2019a). In particular, proponents of Semantic Web technology – the technology at the center of FAIR[114] – have argued for the possibility of liberation from these dilemmas

[114] The relationship between the two is discussed further in chapter 1 of this thesis.

(Poirier, 2019a). The proponents of FAIR seem to follow the same strategy, as we will see in the next section of this chapter, dealing with the infrastructural fiction of FAIR.

FAIR fiction

The infrastructural ambitions of FAIR are readily apparent in its first iteration, the FAIRport initiative, where Barend Mons and colleagues expressed a desire and need to build “an airport for data” (FAIRport, 2014). While this particular infrastructural metaphor and imaginary largely disappeared shortly thereafter, the seeds of its replacement was already sown from the beginning – as they put it in the report of the FAIRport conference: “the FAIRPORT final aim is a fully interoperable ‘Internet of Data’.” (FAIRport, 2014). Just as the internet became an infrastructure for linking documents[115], what would later be known as the Internet of FAIR Data and Services (EC, 2016) was imagined to become an infrastructure for linking data (and services, mind you) – thus realizing Tim Berners-Lee’s vision for The Internet[116].

Inspired by a popular design strategy behind The Internet, they early proclaimed that “The DATA FAIRPORT is not the about the development of yet more standards” (FAIRport, 2014, sic) – a position that was repeated in the first publication of FAIR in 2016 (Wilkinson et al.) and in the follow-up paper from 2017, emphasizing that “FAIR is not a standard” (Mons et al., 2017). The choice to not promote the making of more standards was claimed to both enable maximum flexibility and to avoid “top-down” design choices (FAIRport, 2014; Wilkinson et al., 2016; Mons et al., 2017).

But even though FAIR was not presented as a standard itself, nor relying on creation of more standards, its designers realized FAIR would not work without them. Therefore, the proponents of FAIR encouraged researchers to adopt existing standards, or, if none existed, develop their own. The intended result of this approach would be that the Internet of FAIR Data and Services would be an infrastructure of multiple standards, giving the users added flexibility, but also ownership over their own standards.

By leaving the choice of standards to the scientific communities, the people behind FAIR aimed to circumvent the problem of applying their own epistemic frameworks to other fields of study, hoping to avoid or minimize resistance (Mons et al., 2017; Jacobsen et al., 2020). The way they imagined it, research communities knew best how to design standards for their own needs, and would therefore be able to design and implement the most appropriate

[115] Although not generally or intuitively understood as such, web-pages are little more than rich documents linked together on the internet and read by advanced machines
[116] More on this in chapter 1

standards for themselves. The only thing they asked was that these standards would follow their list of 15 principles for good data stewardship (Wilkinson et al., 2016).

When the proponents of FAIR were talking about standards, what they had in mind was what computational sciences often call "controlled vocabularies" or "ontologies" (FAIRport, 2104; Wilkinson et al., 2016). These standards are, essentially, a list of terms and the relationships between them (Staab and Studer, 2009; W3C, 2010). The use of these are meant to delete ambiguity by avoiding that the same things are designated under different names, or that different things are designated under the same name, so "the machine knows what we mean", as Barend Mons likes to say (Mons et al., 2020) – which in turn is supposed to enable interoperability (W3C, 2010). The example Barend Mons would use to describe it was the use of the words "malaria" and "paludisme" to refer to the same disease in medical research (FAIR-DI, 2020). By deleting the ambiguity between signifiers and concept – so that different actors "speak the same language" – the use of ontologies was intended and imagined to enable automatic machine analysis of data (FAIRport, 2014; Wilkinson et al., 2016).

As different scientific communities would start designing and implementing these ontologies, the people behind FAIR imagined, computer scientists could make translators to connect across these various standards: what one community would call "paludisme" could be mapped onto what another one called "malaria", for instance. The two denominators would then point to the same concept, and this concept could then be given a "persistent identifier" that a computer could use to combine and analyze all available knowledge about it – regardless of what scientific community it came from[117].

The benefit of this approach, proponents of FAIR claimed, was that it could be applied to "[achieve] the transition from the current closed and silo-based approaches to research towards more open and networked scholarship" (Mons et al., 2017), without having to demolish the silos and build from scratch. Since a scientific community, in theory, wouldn't have to adapt its practices to another, but rather just establish and follow best practices within itself, the intervention would supposedly be minimal (Wilkinson et al., 2016). In this way FAIR was imagined to enable global practice while also preserving local ones. As they put it in the Wilkinson et al. (2016, emphases added) paper:

> [T]hrough the definition of, and widespread support for, a *minimal* set of community-agreed guiding principles and practices, *all* stakeholders could more easily discover, access, appropriately integrate and re-use, and adequately cite, the vast quantities of information being generated by contemporary data-intensive science.

[117] For a more in-depth discussion about this idea, look to chapter 1

This idea was explicitly inspired by a design strategy popularized by the Internet development community: the hourglass approach (Guenther, 1998; Beck, 2019)(Figure 7)[118]. According to the logic of the hourglass approach, an infrastructure could gain flexibility by defining "minimal but strict conventions" (Wilkinson et al., 2016). In the case of FAIR (and the FAIRport), this was supposed to

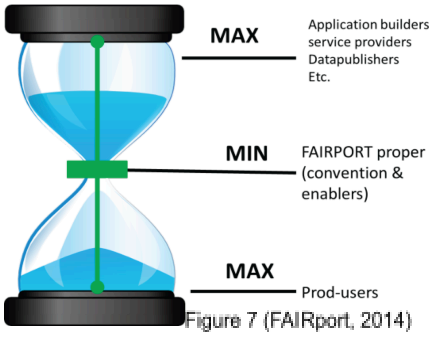

Figure 7 (FAIRport, 2014)

> enable a wide variety of services and applications needed to realize computer-friendly as well as human-friendly data interoperability, stewardship and compliance against data and metadata standards, policies and practices (FAIRport, 2014).

For the people behind FAIR, the image of the hourglass was mainly intended to illustrate the desire for a minimal set of "conventions" at the "narrow waist" (FAIRport, 2014). The wider ends of the hourglass – thus not considered as interesting for that cause – thereby only became vaguely described as "wider" or "maximal" endpoints, containing a larger diversity of elements. Looking at the illustration that was originally used for the hourglass design approach (Guenther, 1998; Beck, 2019), we get a better picture of what the image doesn't show (Figure 8).

Layer 4	Layer 4	Layer 4
Layer 3	Layer 3	Layer 3
Layer 2		
Layer 1	Layer 1	Layer 1

Figure 8

The "stack" model, as it is called (Beck, 2019)(see figure 8), more clearly visualizes different "layers". So instead of freely associated elements drawn together by a narrow passage point, we see clearly delineated entities adding up and filling the entire width, connected instead by a wide "spanning layer" (Guenther, 1998; Beck, 2019). These entities are often, by the proponents of FAIR, referred to as "domains" (as in *scientific* domains[119]) (see Wilkinson et al., 2016)[120]. Applying this language to the above model and identifying the remaining layers, we get something that looks more like figure 9.

---

[118] This is discussed more in-depth in chapter 1

[119] The notion of "domains" is commonly used in computer science to refer to epistemic spaces differentiated by matters of concern or ways of knowing (see Ribes, 2019)

[120] The formulation of principle R1.3 – which was one of the core contentions in the making of FAIR metrics – is a good example: "(meta)data should meet *domain*-relevant community standards".

	Domain 1	Domain 2	Domain 3
User	D1 Researcher	D2 Researcher	D3 Researcher
Service	D1 Service	D2 Service	D2 Service
Spanning layer	Domain independent FAIR		
Data	D1 Data	D2 Data	D3 Data

Figure 9

Compare this with a model illustrating knowledge silos (figure 10), and the allure of the FAIR hourglass approach becomes more apparent.

	Silo 1	Silo 2	Silo 3
User	S1 Researcher	S2 Researcher	S3 Researcher
Service	S1 Service	S2 Service	S3 Service
Data	S1 Data	S2 Data	S3 Data

Figure 10

In the silo model, all domains are separated by a boundary, only allowing vertical interaction between researchers, services and data in the same silo. In the FAIR model (figure 9), "minimal standards" (FAIRport, 2014; Wilkinson et al., 2016) allow data from different domains to move up, so to speak, through the minimal standards and to the services used by researchers in different domains – or, alternatively, for services to move from different domains down to the data level[121]. This, the proponents of FAIR suggest, would be made possible by inserting the spanning "domain-independent" layer between services and data (FAIRport, 2014; Wilkinson et al., 2016). This layer is not only supposed to allow interoperation, but also becomes an "obligatory passage point" (Callon, 1984) for those who use the infrastructure.

While the idea of a spanning layer was far from new – stemming from early internet engineering – the proponents of FAIR argued that this had not been practically attempted in science on the scale which they suggested, as illustrated in the following quote from Wilkinson et al. (2016):

> While there have been a number of recent, often domain-focused publications advocating for specific improvements in practices relating to data management and archival, FAIR differs in

[121] These notions of data mobility versus software mobility can be seen in Barend Mons' concept of data visiting, exemplified by the "personal health train" project (World Duchenne, 2019)

> that it describes concise, **domain-independent, high-level principles** that can be applied to a wide range of scholarly outputs (Wilkinson et al., 2016, emphasis added).

I've chosen to highlight the two hyphenated concepts in the above quote (domain-independent, and high-level principles) because they illustrate an important insight about the FAIR approach: it shows two essential components intended to overcome domain boundaries and linking across silos. First, as mentioned, the FAIR approach would not attempt at linking separate domains through applying the standards of an existing domain – like suggesting that sociology and environmental studies should use biology standards. Instead, a standard that is *independent* of domains would be applied, thus circumventing both epistemic mismatch and the ensuing discontent from having one domain adapt to the standards of another. Second, the domain-independence would be achieved by "moving up" to "higher-level principles". These two approaches are then interlinked – independence is achieved precisely through a higher level of abstraction, producing a sort of common denominator standard.

If domains have common denominators – the FAIR approach suggests – their differences could be resolved by translating up a layer to this denominator, and then translating down again into the desired domain. Using the example of malaria again, this could be expressed as "malaria" translating up to the persistent identifier "NCBI:txid5833"[122], which could translate down again to "paludisme".

This approach to achieving interoperability is what scholar David Ribes calls "the logic of domains" (Ribes, 2018; Ribes et al., 2019). According to Ribes,

> a characteristic feature of the logic of domains is positing (and then engineering) an intermediating object across discrete domains. Domains are thus not placed directly in contact, and their boundaries are not dissolved. Rather, difference is translated while leaving domains relatively unchanged.

Concretely, this "intermediating object" proposed by proponents of FAIR – what has been referred to as the "narrow waist of the hourglass" or "domain-independent layer" – is a concept called the "FAIR Digital Object" (or FAIRDO/FDO) (GOFAIR, 2021; Santos et al., 2022; FAIRDO, n.d.)[123]. This FAIR Digital Object is imagined to serve as the highest abstraction of a concept, intended to act as a common denominator for all representations of itself: presented as "A technical essence of a 'thing' in cyberspace" (FAIRDO, n.d.).

[122] The identifier used for malaria parasite *Plasmodium falciparum* in the National Center for Biotechnology Information (Schoch et al., 2020)

[123] You might remember the "digital object" approach from chapter 1.

While this “technical essence of a thing” was imagined to enable the transgression of domain boundaries, it was not intended for human use directly (FAIRport, 2014; Wilkinson et al., 2016). Rather, the FAIR Digital Object was intended to enable interdisciplinary automated machine analysis of research data (FAIRDO, n.d.) – aspiring to an “optimal state – where machines fully ‘understand’ and can autonomously and correctly operate-on a digital object (Wilkinson et al., 2016). By turning scientific knowledge into epistemically compatible, abstractable and identifiable units of data, the authors of FAIR envisioned to make this knowledge available for machines and finally fully enter the era of eScience (FAIRport, 2014; Wilkinson et al., 2016).

That is at least how FAIR was imagined to work according to its proponents. According to scholars of Science and Technology Studies, though, the assumptions behind this approach are counter to the way science is practiced and knowledge is produced – “the way [researchers] actually do work”, as Christine Borgman (RDA, 2014b, 23:20) put it (Ribes, 2017; 2018; Poirier, 2019a). What’s more, the inherent logic of it, they suggest, might be more controversial than its proponents presume (Ribes, 2017; 2018; Poirier, 2019a). Thus, as the fictional infrastructure of FAIR encounters scientific practice, frictions occur.

FAIR friction

It is perhaps apt that this section, following the section on fiction, deals with the concept of reality[124] – because one of the main sources of friction when it comes to FAIR is its conception of reality, or what we will refer to as “ontology”.

In philosophy of metaphysics, an “ontology” is a particular way of conceptualizing reality and what it consists of. Typically, this is distinguished from epistemology, which concerns itself with the question of how we can gain knowledge about reality. Within several disciplines over the last couple decades, STS including, this distinction has been drawn into contestation, in what has been dubbed “the ontological turn” (Woolgar and Lezaun, 2013). According to proponents of the ontological turn, the distinction between ontology and epistemology is meaningless, as realities can not be conceptualized as something “out there” independent from experience – realities must be simultaneously seen as both the products and sources of experience (Law, 2004). Concerned less with what reality *is* and more with what it *does*, and considering that ways of knowing affects ways of doing, the ontological turn therefore collapses epistemology and ontology into each other (Law, 2004; Barad, 2006; Woolgar and Lezaun, 2013).

[124] or realit*ies,* if you will

One of the driving actors behind the ontological turn in Science and Technology Studies, Annemarie Mol, broadly differentiates four different ontological paradigms: singularism, perspectivalism, constructivism and multiplism[125] (1999). The first distinguishing factor between these paradigms is their relationship to truth, where singularism is the only one to insist on *one truth*, while the others allow for a certain plurality of truths. Of the other ontologies, perspectivalism explains plurality of truths as a product of the different perspectives actors can have on the same object. In that way, it allows for different interpretations and experiences, but sticks with the idea that these truths stem from a common reality. Constructivism, Mol argues, contends that reality is continuously produced and reproduced with the potential for multiple different outcomes, but ultimately producing a common one. Multiplism, she claims, rejects the notion of a common or shared reality – what John Law calls "a one-world world" (Law, 2015) – and replaces it with the notion of multiple ontologies. These multiple ontologies, Mol suggests, are simultaneously coexisting realities that come into being through *enactment* (Mol, 1999; 2002). *The* world is not made – *worlds* are being practiced.

In her 1999 publication, Mol uses the example of anemia to illustrate her argument for multiple ontologies. Doing an anthropological study of anemia in different medical spaces, she discovered that although different people in different contexts talked about anemia as one and the same thing, it was enacted quite differently (Mol, 1999). In one place she demonstrated how anemia in the doctor's office was a matter of redness of the patient's eyelids. In the lab, on the other hand, it was a matter of statistical deviance of hemoglobin levels. To make matters more complex, in the medical literature, it was a matter of oxygen transportation to the organs (Mol, 1999).

While one might argue that this difference in enactment of anemia can be explained by perspectivalism – that the observation and enactment is different, but that the object remains the same – Mol counters this by showing how these different instances of anemia cannot be pointing to the same phenomenon. She insists that the anemia actually differs in significant ways – that it's not the same, but multiple. This multiplicity of anemia – or any other phenomenon – she argues, does not mean that one is more right or wrong than the other: instead, they are enactments of a reality that is, in practice, multiple itself (Mol, 1999; 2002).

So what does the multiple ontology imply? In the first instance, Mol argues (1999; 2002), a multiple ontology means that "there are options" – echoing the STS maxim that "it could be

[125] It is noteworthy that Mol does not use these exact terms in her writing. I employ these as close approximations to ease the reading experience.

otherwise" (Woolgar and Lezaun, 2013)[126] – that there isn't only *one* world to choose from. Second, she continues, ontologies are not neutral – rather, they are what Latour (1993) might have called "politics by other means". These two notions led her to develop the concept of ontological politics – the idea that ontologies have political consequences and, thus, are (and should be) arenas for political debate and action (Mol, 1999). The notion of ontological politics therefore encourages critical interrogation with ontologies – asking what they admit, prohibit, encourage and marginalize (Mol, 1999; 2002; Law, 2004; 2015).

Taking FAIR at face value, its ontological framework seems to be one of the strengths of the concept, that it allows for the existence of multiple ontologies through its hourglass approach. After all, every domain is encouraged to use their own ontologies while machines will translate across them through a separate high-level ontology. But, this is a misunderstanding of how the term "ontology" is used in these two contexts. For while ontologies in computer science can be translated and resolved into each other, the ontologies that STS scholars talk about cannot.

To illustrate the different meanings of ontology, I want to revisit the stack model from before, but with some of the superfluous elements stripped away. Having done that, the ontology that FAIR presumes looks something like figure 11.

Ontology 1	Ontology 2	Ontology 3
Ontology 1	Ontology 2	Ontology 3
FAIR Digital Object Ontology		
Ontology 1	Ontology 2	Ontology 3

Figure 11

The first quality of this model I want to draw attention to is the boundaries – the black lines drawn between the different domains or ontologies. Whereas Mol denies a hierarchical ordering of ontologies – one that implies a higher plane that all ontologies are derived from or can be abstracted to (Mol, 1999) – the horizontal line in the model, separating FAIR from community standards that is promoted by Mons suggests the opposite (FAIRport, 2014; Wilkinson et al., 2016). While ontologies in the lower levels of the model are not explicitly claimed to *stem* from the higher-level ontology, they are presented as being able to resolve into it – or even to be *reduced* into this plane of "essence" (FAIRDO, n.d.).

[126] This is, to be fair, a more modern version of the original "it could have been otherwise" – a notion stemming from the historian parts of STS. Mol ties this sentiment to the constructivist ontology – one which relegates plurality to the past – and distinguishes this from multiple ontologies by not merely implying it could *have been* otherwise, but that it could be otherwise here and now (1999).

In this embodiment, FAIR suggests a *singular* ontology, one that is perspectivalist: an ontology where each domain has an equal claim on truth, but only as long as this aligns or resolves into the “higher” ontology. Thus, while Mons’ ontology may seem multiple at first glance, it performs – in Mol’s terms – *plural*. I’ll call this particular form of ontology “machine ontology”.

The ontological freedom and multiplicity promised by this machine ontology, thus, is only a relative freedom of *computational* ontologies (or controlled vocabularies). For what happens when knowledge in one domain deviates from another? An example that could serve to answer this question comes in the form of a Q&A with Barend Mons during the *Conference on a FAIR Data Infrastructure for Material Genomics* in 2020, where one of the people in the audience, James Warren, asked:

> Is it also possible to detect bad data to avoid sloppy science to fund malicious mischief or worse so? The question is: how do you build trust in the FAIR model? (FAIR-DI, 2020, 10:20).

The issue of trust in data was founded in a concern that if anyone would “upload” data into the model[127] that was somehow not good, it would mix bad data into the good and poison the well, so to speak. Since the data would be read by computers, and the knowledge mediated as a compound, James worried it would be difficult to notice the “bad data” in the mix. Although Barend Mons recognized the concern, he reassured that this would not be an issue:

> Yeah… Very important question. I've even been invited to Brussels by [EU commissioner Frans] Timmermans to see whether we could detect fake news, but of course… You have a knowledge model of all established knowledge [...] …of course if there is a new assertion coming into that system and it causes a lot of tension in the model there is an alert, but **this could be a lie or a scientific breakthrough,** because scientific breakthroughs do not fit in the common knowledge as well. Of course they should not be completely out there, but tension in the graph is a sign something is happening (FAIR-DI, 2020, 10:40).

In this quote (and in other places (see CODATA, 2021, 1:10:40)), Barend Mons makes clear that “the model” allows for diverging claims, but that these “assertions" will create tensions that indicate that either the claim is false, or that the pre-existing model is false. In other words, one claim, or set of claims, would have to cede for the other one – there would be no room for the coexistence of two opposing truths. Making different domains subject to a machine ontology would, in other words, not in fact allow for multiple ontologies, but pit them against each other in a claim for essential truth – an ontological survival of the fittest spurred on by semantic incompatibility.

[127] “The model” here, likely refers to the Semantic Web of interlinked data

One might interject that this is how scientific knowledge works irrespective of the machine ontology, but in scientific practice, deviating ontological states of objects and practices is actually rather commonplace (Ribes, 2018; Ribes et al., 2019). Furthermore, different ontological enactments in scientific practice do not prohibit interaction and cooperation between or across ontologies (Star and Griesemer, 1989; Mol; 1999; 2002) – which takes us from the horizontal to the vertical lines in figure 11 – the ones delineating different ontologies.

While Mol (1999; 2002) suggests that ontologies can exist next to, but separate from each other, she underlines that this is far from the rule, but, rather, the exception. Ontologies, she suggests, go into, around, beside, under, over and on top of each other, their boundaries more porous than solid (Mol, 2002).

Instead of a messy, overlapping topology of ontologies – as suggested by Mol and others (1999; 2002; Barad, 2004; Law, 2004; 2015; Woolgar and Lezaun, 2013) – the horizontal and vertical separations of the machine ontology ultimately represents science as fractal and cumulative. This way of representing science is what David Ribes calls the logic of domains, meaning a belief that individual domains can be added up together to form science as a whole (Ribes et al., 2019). Imagined this way, we can picture science as a cake divided into different slices named Natural Science, Social Science, Humanities (figure 12).

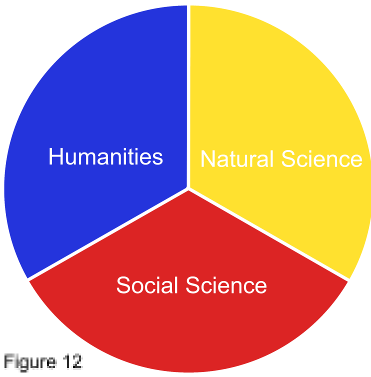

Figure 12

According to the logic of domains, these slices could then be divided further into smaller slices named Biology, Engineering, and Meteorology; Sociology, Economics, and Anthropology; and History, Linguistics, and Musicology for example (figure 13). This slicing into smaller parts could go on and on until the smallest epistemic unit of science. Importantly, the sum of these slices would always make up the same total cake of science, nothing more and nothing less and everything with its designated place.

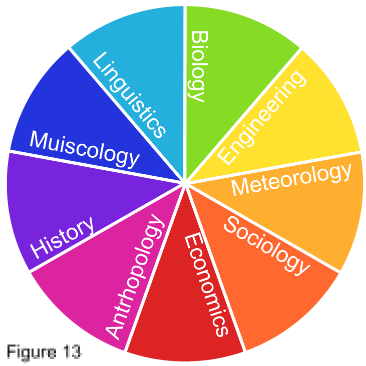

Figure 13

For STS scholars this is intuitively misconstrued for two reasons. First, because the discipline of STS itself often defies disciplinary categorization. Second because the concept of boundaries in itself has been repeatedly contested in seminal STS literature (Gieryn, 1983; Star, 1989; Knorr-Cetina, 1999; Galison; 1999; Epstein, 2007).

In STS scholarship, science is not seen as a cake – it is not seen as a clearly defined and delineated entity with strict rules and distinguishable contours. In fact, some of the foundational work in STS established that "science" in itself cannot be clearly distinguished and kept separate from "society" – that the laboratory is not isolated from its surroundings, but clearly embedded within them (Gieryn, 1983; Latour, 1993). Similarly, and as an extension of this line of thinking, Karin Knorr-Cetina argued that the disciplines "within" science also lack any clear boundaries (1999). While two laboratories might be categorized as belonging to the same discipline or scientific community, she argues, they are far from the same, and will be differentiated by a host of "un-scientific" factors traditionally seen as belonging to social, political, economic and interpersonal domains. Consequentially, she refuses the idea of disciplines and replaces it with the notion of "epistemic cultures" (1999).

Epistemic cultures, as opposed to the notions of disciplines or communities, highlight the intricate relationship between ways of knowing and producing knowledge, with scientific as well as extra-scientific factors. As these factors become accounted for, the seemingly neat order of science and its disciplines crumbles to an unorderly mess. Instead of having clear boundaries, epistemic cultures, which on the surface seem different, might be found overlapping with each other in a number of ways – for example, individual researchers may be moving across epistemic boundaries (Davies, 2020) or the same methods may be used across different disciplines (Law, 2004).

This last point about the methodic overlap between disciplines is sometimes used by proponents of FAIR as an argument to show how it is possible to overcome disciplinary boundaries: that medical scholars use interviews in the same way as social scientists do[128], or that the humanities have embraced computational methods (sometimes referred to as "digital humanities") (EROSS, 2020). In their way of presenting it, this serves as proof that transepistemic knowledge production, like the one intended to be enabled by FAIR, is not only entirely possible, but already happening.

While embracing the weak boundaries implied by epistemic cultures as support for their infrastructural ambitions, they overlook the other side of the coin, which is that these transepistemic ventures are not signs of traversed boundaries, but a symptom of the lack of definitive boundaries in the first place (Knorr-Cetina, 1982). In other words, interdisciplinarity – like an economist using satellite data to monitor economic growth, or a neurologist using hydrodynamic models to understand blood flow in the brain – is not a sign that scientists are dismantling or reaching over silo walls, but an indication that there might not have been walls there to begin with (Law, 2004).

[128] From interview with coauthor of Wilkinson et al. (2016), Tom (see table1)

In addition to overestimating the differences between disciplines, the proponents of FAIR also underestimate the differences within. A discussion at an EOSC Nordic webinar on data sharing serves as an example: During a presentation on community metadata standards, a message appeared in the chat, from Mari Kleemola, an employee of the Finnish Social Science Data Archive. She was voicing skepticism towards the efficacy of community standards.

> From Mari Kleemola to Everyone 12:01 PM
> Even when a widely-accepted community metadata standards exists, like DDI for social sciences, the real-life experience has been that we have tended to create our own DDI dialects!

Angus Whyte, a Research Data Specialist with the Digital Curation Centre at the University of Edinburgh countered:

> From Angus Whyte to Everyone 12:03 PM
> @Mari @Joakim isn't this the benefit of metadata application profiles, i.e. to mix and match existing standards?

Before we move on, I'll take some time to explain what Angus Whyte is talking about. The idea behind "metadata application profiles" builds on the same idea as FAIR more broadly – that differences between domains can be overcome by referring to a higher level of standards. In the case that Angus Whyte referred to, this principle would be applied also within domains with more granularity, cutting the epistemic cake into smaller pieces, if you will. Following this method, one could go as far as to make a standard for every individual project, and resolve it into a slightly higher level standard (let's say a community standard). One concrete example of this is the CLARIN[129] Component Metadata Infrastructure (CMDI), illustrated in figure 14. In their model, the "CCR definitions" act as the spanning layer, or narrow waist of the hourglass, "to offer the advantages of shared conventions, common semantics and predefined structures while maintaining the flexibility required to serve the needs of CLARIN's core communities" (CLARIN, n.d.-b). In other words, Angus had suggested that the use of such an approach – based on the FAIR principles

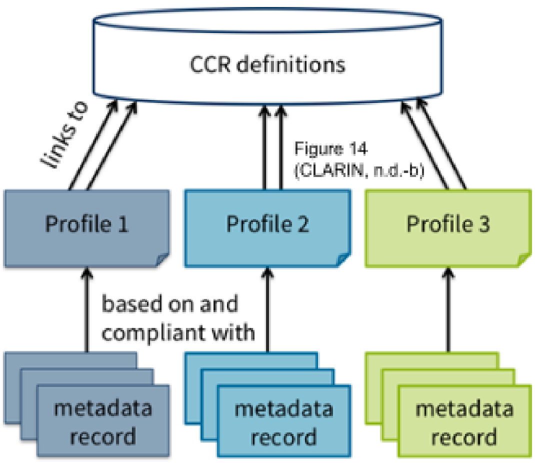

Figure 14
(CLARIN, n.d.-b)

[129] CLARIN stands for Common Language Resources and Technology Infrastructure [sic], and is part of the European Research Infrastructure Consortium (ERIC) (CLARIN, n.d.-a)

(CLARIN n.d.-a) – would solve the issue of intracommunity variation, or “dialects” as Mari called them.

Another participant of the webinar, Philipp Conzett, data steward at Dataverse Norway, quickly replied to Angus’ comment:

> From Philipp Conzett (UiT) to Everyone 12:04 PM
> @Angus: Combining standards into profiles also has its challenges, cf., e.g., the proliferation of CLARIN CMDI profiles...

The challenge Philipp pointed to was not uncommon, nor unheard of. Susanna-Assunta Sansone, one of the original authors of the FAIR principles and Director at the Oxford e-Research Centre, had encountered it before in her work:

> From Susanna-A Sansone to Everyone 12:05 PM
> @Philipp - There is another issue with the use of existing standards (checklists, ontologies, models etc): it is relatively easy to create a metadata template that complies to one checklist/standard; but experiment are complex, e.g. Proteomics-based investigations of neurodegenerative disease. These will require use of several checklists, ontologies and models. If of interest we highlight this issue here https://doi.org/10.5281/zenodo.5596465

In her experience, the epistemic diversity within communities would quickly create frictions with the standards, exploiting the flexibility of the approach to nearly “tip the standard into uselessness”, to use Timmermans and Epstein’s (2010) words (Batista et al., 2021). In the report she linked to in her comment, the authors put it this way:

> The wealth of domain-specific metadata standards is proof that communities operate within their domains of interest and expertise, and these activities should be encouraged. However, the vast majority of community-focused activities creates fragmentation that leads to a lack of interoperability amongst the resulting standards (Batista et al., 2021).

That being said, the authors didn’t consider these challenges to be impossible to overcome. In their concluding remarks, they noted:

> The need for coordination among community standards and their interoperability are not trivial issues. There are social and technical challenges. Nevertheless, it is pivotal to foster convergence, modularity and interoperability among the community-standards if we are to realise the vision of FAIR data. (Batista et al., 2021).

The sentiment that “there are social and technical challenges” to be overcome was widespread among the proponents of FAIR, who kept working to realize the vision of the Internet of FAIR Data and Services – to overcome the friction between local and global in order to make their fiction come true. This sentiment is succinctly summarized by the GOFAIR organization (GOFAIR, 2020):

> We have learned that, classically, domains operate in silos and that even within domains multiple standards, vocabularies, languages, and approaches will continue to emerge. This is not only a nuisance and a lack of coordination and discipline, it is also an intrinsic part of the creative process that should be supported in order to further our knowledge and drive innovation. This means that 'mapping tables', 'libraries to choose from', 'community standards registries', etc. will continue to be crucial elements of the IFDS[130] support infrastructure.

In the following section I will sketch out the potential consequences of this attitude, and the danger that might be tied to it – should it succeed.

FAIR in the making

So far in this chapter I've attended to the infrastructural ambitions of FAIR, inquiring into the tensions they create between global and local (Star and Ruhleder, 1996), between fiction and practice (Carse and Kneas, 2019). In my attempt to pin down what the fiction of the FAIR infrastructure is, I've drawn attention to its own ontological underpinnings, its "logic of domains" (Ribes, 2018; Ribes et al., 2019): the hourglass model. Following the notion that ontologies are not purely descriptive, but gain a prescriptive character through their enactments (Mol, 1999; 2002), I'll argue that the infrastructural dilemma between strict and flexible (Star and Ruhleder, 1994; 1996; Bowker and Star; 1998; Poirier, 2017; 2019a; 2019b) is not only a question of epistemic efficacy and efficiency, but a question of ontological politics.

In the fiction of the FAIR infrastructure, ontological multiplicity is not seen as something to be conserved, but "a nuisance" that creates frictions (GOFAIR, 2020). Multiplicity is cast as a problem that needs to be overcome in order to realize a future where scientific domains can make use of data and services from other, separate domains – a future where science is interoperable. This idea, that data and knowledge not only can be, but also should be interoperated across separate domains/disciplines/communities of science has its roots in the Semantic Web approach that serves as the basis for the FAIR infrastructure (Ribes and Bowker, 2009; Guns, 2013; Ribes, 2017; Poirier, 2017; 2019a; 2019b; Ribes et al., 2019).

While the history of Semantic Web technology has shown a persistent dilemma between flexibility on one side and strict standards on the other (Poirier, 2017), the proponents of FAIR have promoted their approach as a way of working around it. Through the combination

[130] Internet of FAIR Data and Services

of the hourglass design approach and digital objects, they've aimed to achieve both the wide uptake associated with flexibility, and the high degree of interoperability and automation associated with standards (Poirier, 2017).

Still, practical applications of the FAIR approach have invariably failed to deliver on the promise of flexibility and efficacy. Thus, the Internet of FAIR Data and Services remains a fiction as epistemologies resist the efforts to make them widely interoperable. Projects attempting to create higher level ontologies, or translating across ontologies produce considerable frictions, and struggle to achieve a high enough level of goal attainment to make this fiction reality (Star and Ruhleder, 1996; Carlile, 2004; Ribes and Bowker, 2009; van Damme et al., 2022; Guiliano and Estill, 2023; Luschi et al., 2023). While continuing to produce the same frictions as previous semantic web efforts, finding that "the worldviews and concepts diverse people use to describe their world often don't fit so 'neatly' into tighter systems of representation" (Poirier, 2019), the efforts of FAIR proponents have persisted. In other words, as the practices of researchers resist the FAIR fiction, the fiction of FAIR simultaneously resists the frictions this tension produces. Thus, the infrastructuring continues, maintaining the tension between local knowledge practices and a fictitious global infrastructure.

But while the FAIR infrastructure struggles to manifest, and the literature is full of unrealized infrastructures of interoperability (RDA, 2014b), there is no inherent impossibility in it happening down the line. After all, both the past and the present are full of infrastructures that have managed to overcome incompatibilities with their subjects/users through a combination of negotiation, discipline and/or violence (Larkin, 2013; Karasti et al., 2016; Poirier, 2019a; Bogart, 2020).

While remaining in this unfinished state – still in the making – FAIR still affords options: will it realize its machine ontology and delete the multiple ontologies of scientific practice, will it fail to overcome the tension between local and global, leaving the world multiple and as it was, or will it appropriate the nature of its subjects and become multiple itself? I will leave this question hanging in the air for now, but will come back to it in the next chapter of this thesis.

Chapter 7: FAIR futures

In this thesis, we've spent a lot of time looking backwards to the near and immediate past – to the genealogy, history and more current development of FAIR. We've largely gravitated around the time of these principles' inception in 2014, moving backwards and forwards some years according to the needs of the story being told. But as we have kept our gaze fixed in the rearview mirror, the past has been staring back at us, towards the future. Whether it's the future of research communications and e-scholarship (FORCE11, n.d.), the future of the European Union (EC, 2007b), or the future of revolutionary infrastructures (Wittenburg and Strawn, 2018): the future has been a recurring topic in the development of FAIR.

While the future often entered my research through statements about how it would or could look like, it frequently manifested through questions, and one in particular: what is the future of FAIR? Oftentimes, I would be the one asking – myself or others – but not seldom would it be my informants or my colleagues posing the question. It was typically not a question that instilled confidence in the one who was asked, and, thus, it typically only produced careful, tentative answers. Because, when I started doing my research in 2020, FAIR had encountered some challenges. Even though many had subscribed to FAIR, and were promoting it in their own work, the original proponents noticed that far from everyone was pulling in the same direction (Mons et al., 2017)[131]. In an attempt to solve this problem, the original proponents tried "clarifying" what FAIR was and wasn't – and, importantly, what the goals behind the concept were intended to be (Mons et al., 2017; 2020; Jacobsen et al., 2020)[132]. But these efforts had only limited success, as diverging enactments of FAIR continued.

Observing the difficulty of aligning all FAIR proponents around a common understanding, two of the original proponents attempted to explain why this divergence was happening in the first place, and further, what it would mean for the future of FAIR. RDA founding member Peter Wittenburg and FAIRporter George Strawn started looking to literature on processes of technological innovation and change, and used this to write the article *Common Patterns in Revolutionary Infrastructures and Data* (2018). In it they claimed that the challenges FAIR was encountering – with multiple interpretations pulling in different directions – were completely natural for innovations such as FAIR, and dubbed this multiplication process "creolization"[133]. Over time, though, Wittenburg and Strawn (2018) argued, the different

[131] See also previous chapters
[132] See chapter 3 for more in-depth discussion on the matter
[133] Not to be confused with the way Peter Gallison (1997) uses the term "creolization", which is to describe a process that brings different actors together in hybrid constellations through development

actors would converge around one common interpretation that would lead to the original goal of FAIR finally becoming fulfilled: a semantic web of data.

Since the publication of the Wittenburg and Strawn (2018) paper, much effort has been put into achieving this "convergence", including the organization of "convergence symposia" (GOFAIR, 2020; 2022). Yet, little suggests that this convergence has been reached or that it is reaching a close. In their follow-up paper, *Revolutions take time*, three years after the first paper, Wittenburg and Strawn (2021a) acknowledged the failure to create convergence, but reassured that it was still going to happen, it would just take some more time – the future staying ever on the horizon.

But is this convergence first and foremost a question of time? Will this future eventually realize itself, or is there reason to believe that we have a different future in store? In this chapter of my thesis I will explore a collection of alternative futures represented and enacted by other proponents of FAIR (or "opponents" of FAIR, from the perspective of some), and how the original proponents' reluctance towards accommodating alternative futures may very well kept them from realizing their own preferred future.

Studying the future

Before we get started, we have to consider what "the future" actually is. In scientific discourse, as elsewhere, the future is a contentious topic. For natural sciences, the future is often something to be studied in a predictive manner – of what will happen with the universe as it keeps expanding, of what kind of weather we will have tomorrow, or even in more immediate causal terms, of what will happen after you pour water into an acid. In the social sciences, nostradamic efforts are generally more frowned upon, something to keep your hands off of (unless you're an economist) (Adam, 2023). In spite of this aversion (or, perhaps, because of it), the future has gained significant attention within parts of social sciences over the past couple decades, even spawning a dedicated journal called *Futures*.

In the branch of social sciences dubbed "future studies", scholars distinguish between different forms of futures. Barbara Adam (2008) emphasizes that futures have different modalities that can be expressed by the future's relationship to the present. A "future present" is what most people, including the examples mentioned above, refer to when they talk about "the future" – what will be experienced as the present after a certain time has passed, like the weather tomorrow. Conversely, and central to this chapter, is the "present

---

of creole languages – which in many respects is the opposite of what Wittenburg and Strawn are trying to describe.

future": presently expressed visions (Lösch, 2006), expectations (Borup et al., 2006), and imaginaries (Jasanoff and Kim, 2015) of how the future *could* look like – what we today say that the weather might be tomorrow (Adam, 2023).

Concerned with the agency and effect of present futures, scholars have developed some methods for studying them. As a foundation, they often emphasize how the future is not a distinct object to be observed, but, rather, an emergent effect of the present. The future, thus, can be found in various sociotechnical assemblages that consist of practices, enunciations and materials (Jasanoff and Kim, 2015; Michael, 2017; Schneider and Lösch, 2019; Halford and Southerton, 2023). In other words, we can look for "the future" in everyday objects, such as wheeled luggage (Michael, 2017), in policy documents, such as national nuclear strategies (Jasanoff and Kim, 2015), and in the fictional utopias of artists, academics and activists (Greenwood, 2023), to name a few examples. But, importantly, these do not necessarily enact the same future – which is why we rather talk about futures in a multiple sense.

But futures are not only a collateral of assemblages (Felt, 2013), they are necessary for their stability and continued existence. In this way, futures contribute to the present's extension into the future – its "telos" (Appadurai, 2012; Felt, 2016). This simple but profound notion – that the present is imbued with meaning by the future – can be found outside of future studies as well. It can be found in psychoanalyst Victor Frankl's work on concentration camp captives, where he observed that robbing an individual of their future, also robbed them of meaning and will to live on (Frankl, 1946). It can be found in sociologist Anthony Giddens notion of "ontological security", that only the sense that the present will extend into the future enables individuals to meaningfully extend themselves into this future (Giddens, 1991). In other words, without a future to orient themselves towards, actors become unable to act with purpose and motivation.

The notion that a futureless present is a present without purpose and motivation remains true not only for individual actors, but also for collectives. In the same way that futures motivate the wills of individuals, futures also coordinate wills *between* individuals (Jasanoff and Kim, 2015; Andersson, 2017). This ability is perhaps most evident in the case of strongly scripted futures – like organizational strategies, political roadmaps or vacation plans – that allow a group of people to act in concert (Andersson, 2017). But fuzzy, ambiguous and vague futures can have a similar effect (Rist, 2007; Bensaude Vincent, 2014; Bos et al., 2014; Ribeiro, Smith and Millar, 2017).

Whereas defined futures are used to enroll actors into a more or less *stable collectives* (Jasanoff and Kim, 2015), the collectives around fuzzy futures are per definition unstable

(Cornwall, 2007; Ribeiro, Smith and Millar, 2017). Often, these futures are expressed and circulated in brief form, such as phrases or even singular words: examples from the literature including "development" (Cornwall, 2007; Rist, 2007), "public engagement in science" (Bensaude Vincent, 2014), and "responsible research and innovation (RRI)" (Ribeiro, Smith and Millar, 2017). As these "buzzwords" only offer fuzzy, ill-defined[134] and highly interpretively flexible suggestions of the future (Cornwall, 2007; Ribeiro, Smith and Millar, 2017), their multiple meanings afford what the strongly scripted futures do not: the attraction of actors with differing ideas, values and goals. Through their flexibility, they gain the "ability to unify [...] a wide range of [...] ethical and policy issues" (Ribeiro, Smith and Millar, 2017). One of the defining characteristics of these "buzzwords" is thus that they organize actors in what Bernadette Bensaude Vincent (2014) calls "unstable collectives", where actors weave in and out of alliances and alignments with each other. This ability to unify and enable coordination across difference has led some scholars (Randles et al., 2014; Ribeiro, Smith and Millar, 2017) to liken buzzwords to boundary objects, in the sense that they allow for the coexistence of "heterogeneity and cooperation" between actors (Star and Griesemer, 1989)[135].

Throughout the chapters of this thesis, we've seen FAIR bringing together a variety of groups and actors with different goals for and visions of the future – be they tacit or explicit. This has led to several discussions and debates around what the goal of FAIR should be, but also on what FAIR is. In this chapter, I aim to enunciate and articulate some of these differences, what futures these enactments express or contain, and to what extent FAIR is able to coordinate these actors.

We will begin by articulating the original future of FAIR, as envisioned by Barend Mons and his colleagues.

A data-centric future

For the original proponents of FAIR, the most basic element of their vision was the interoperability of heterogeneous data – a semantic web[136]. But their semantic web wasn't promoted as a standalone future technology, it was presented as a part of a more complete future. To sketch this future out, I will look at the futures presented alongside FAIR by its original proponents.

[134] Borrowing the language of Star (1989)
[135] It is important to note that "boundary object" is not a claim on the essential character of an object, but rather an emergent effect of its relations (Star, 2010). As such, buzzwords are not boundary objects, but perform the characteristics of boundary objects.
[136] as discussed in chapters 1 and 6

When Barend Mons talked about the future he would sometimes talk about the future he imagined in the past. The story goes that in the 90's he was working as a biologist in Africa and studying malaria (FAIR-DI, 2020). Part of the challenge back then, he said, was that even though the same disease[137] was endemic across large parts of the continent, researchers and medical professionals were using different terms to describe it – ultimately hampering cooperation (FAIR-DI, 2020). According to Mons, this could be solved – and he came up with an idea that became the first ideas behind FAIR:

> I realized that paludisme was the same concept as malaria, just another token to refer to it in human language. So, I became interested in thesauri, and I started to make a plea to refer in science to concepts with a unique identifier. And that's the end of my biology career almost because I was sucked into graph technology and later on even to conceptual modeling. […] Matching across languages is how it started (Mons, 2020).

According to this story, the experience made Barend Mons realize that "graph technology" and "conceptual modeling" – a strain of computational methods – were the future of knowledge production. Throughout the mid-to-late 00's, Barend Mons expressed this view, that science was transitioning from a narrative-centered to a data-centered paradigm that he called "the e-Science era" (Mons, 2005; Mons and Velterop, 2009; Mons et al., 2011). This future was simultaneously promoted by several other scholars – in particular in the field of bioinformatics – through the Beyond the PDF movement (later dubbed "The Future of Research Communications and e-Scholarship", or FORCE11). One of these visions were explicated and published in the FORCE11 manifesto:

> We see a future in which scientific information and scholarly communication more generally become part of a global, universal and explicit network of knowledge; where every claim, hypothesis, argument—every significant element of the discourse—can be explicitly represented, along with supporting data, software, workflows, multimedia, external commentary, and information about provenance. [...] This vision moves away from the Gutenberg paper-centric model of the scholarly literature, towards a more distributed network-centric model; it is a model far better suited for making knowledge-level claims and supporting digital services, including more effective tracking and interrogation of what is known, not known, or contested (Bourne, Clark et al., 2015).

A product of this envisioned future – as summed up by Barend Mons – was that "in the future we will not have supplementary data for papers, but supplementary papers for data"

[137] The notion of "the same disease" stems from Barend's account. Following Annemarie Mol's understanding of disease, I would argue that it, more likely than not, was not "the same disease", but the disease multiple (Mol, 2002).

(RDA, 2014a). In other words, their envisioned future of science would be more centered on publishing data than on publishing articles.

While scholars like Jasanoff and Kim (2015) emphasize how futures gain their agency from drawing up a vision of what is attainable and good, Barend Mons instead conjured an *inevitable* future that contained the potential for both good and bad. This can be seen in the opening sentence of the first publication of the FAIR principles in Nature Data Science, where the authors claim that "[t]here is *an urgent need* to improve the infrastructure supporting the reuse of scholarly data" (Wilkinson et al., 2016, emphasis added). Why action was urgent is never explicitly stated in the Wilkinson et al. (2016) paper, and neither is the time of arrival for this future specified – but reacting to it is presented as decidedly urgent. This is not to imply that action could have hindered this future from arriving (when?). Instead, the future is presented as if it will happen regardless, following a trajectorist understanding of development (Appadurai, 2012). Thus, action is a matter of preparing for this future to avoid falling behind when it finally arrives at some point down the road – presumably sooner than late. This is perhaps most clearly exemplified by a statement from Barend Mons in 2022 as a comment on the then recently released UNESCO recommendations on Open Science:

> We have entered a disruptive decade, in which data is growing so exponentially that incremental changes in the current system of scholarly communication are doomed to fail (Mons, 2022).

With the backdrop of this inevitable future where science would be unable to deal with unprecedented amounts of data, a pivot towards a new "system of scholarly communication" was presented as the reasonable solution. This particular expectation towards the future, that data would grow exponentially (measured in volumes of bytes, we must assume), was not first introduced by Barend Mons or his colleagues, but had been (and continues to be) spread in popular culture and media. One notable example is IBM's widely circulated claim that "90% of the data in the world today has been created in the last two years" (IBM, 2012; Wonder, 2017)[138], which nurtured an expectation that the growth would continue in a similar fashion as the famous "Moore's Law".

In both the case of IBM's 90% claim and Moore's Law, the future is presented as a logical or statistical extrapolation of the past: because growth has increased according to an exponential pattern in the past, proponents claim, it will continue to do so in the future, ultimately making it predictable (Mackenzie and Wajcman, 1999). This logic has the obvious

[138] A claim that continues to be circulated well beyond its publication in 2012 (see Marr, 2018; Duarte, 2023). Furthermore the original source of the claim can no longer be found on the internet or in the Internet Archive.

flaw that these developments – both in growth of data production and processor speed – are contingent on a host of factors beyond what the simple statistical models include. Still, the future of Moore's Law remained in line with the present for a long time – in part because the expectations it created spurred its fulfillment by companies like Intel and AMD (Schaller, 1997).

But while Moore's Law has been applied in the sustained efforts to fulfill itself, the expectation of exponentially growing data has been mostly focused on how to react appropriately to its repercussions. This particular data future has been presented as one that has the capacity to overwhelm society in similar fashion to a natural disaster – a sentiment found in the terms "data tsunami" (Jaunsen et al., 2021), "data deluge" (The Economist, 2010; Apathy and Brown, 2023) and "data flood" (Borgman, 2014).

What Barend Mons and his colleagues envisioned to meet this oncoming disaster was the construction and implementation of a semantic web, where machines would be able to read digital data and reason on it with minimal human intervention (Berners-Lee and Fischetti, 1999). Barend Mons' and colleagues' suggestion was to take this unrealized vision and make it real within the scientific domain, "a fully interoperable 'Internet of Data'" (FAIRPORT, 2014)", later dubbed "the internet of FAIR data and services" (EC, 2016).

What Barend Mons and colleagues did, was, thus, effectively to use the expectations of data growth to promote the necessity of their vision – a data-centric science relying on semantic web technology. Thus, the semantic web was made "inevitable" only through presenting it as the logical solution to an undesirable expected future – as such, it was vision disguised as expectation.

An Open future

The data future presented by Barend Mons and his colleagues in their promotion of the FAIR principles found resonance with the European Commission who had been expecting a similar data crisis since at least 2010 (see EC, 2010 and Burgelman et al., 2010). The bureaucrats at the European Commission had been worried about the challenges of increasing amounts of data, more or less in parallel with the proponents of FAIR (Burgelman et al., 2010)[139]. Thus, these two groups shared the same expectation, creating an entrypoint for FAIR into the EU's data policy. But, it would become evident, they didn't share the same vision.

[139] More on this in chapter 2

The differences between Barend Mons' and the EU's vision manifested quite early on, as exemplified by the introduction to the first High Level Expert Group on the European Open Science Cloud's report, where Mons stated that the European Open Science Cloud could neither be European, entirely open, only about science, nor be a cloud (EC, 2016). The then commissioner of Research, Innovation and Science, Carlos Moedas, made it clear that the European Open Science Cloud was first and foremost "a vision for the future of open science, *in the context of the Digital Single Market*" (EC, 2016, p. 4, emphasis added), effectively stating that FAIR and EOSC represented a different vision in the European context than it did with Barend Mons and colleagues. But why did they disagree?

When FAIR was introduced to the bureaucrats of the European Commission in 2014, the proponents of FAIR and the bureaucrats of the Commission both agreed that production and dissemination of scientific knowledge was changing, in large part due to developments within information and communication technologies (Burgelman et al., 2010; EC, 2010; FAIRport, 2014). While they both expected the production of data to increase, they had different expectations of what problems it would produce. While the proponents of FAIR expected that scientific data would become "reuseless" (Mons et al., 2017), the bureaucrats in the Commission were more concerned with an expectation that this "new paradigm" could harm the EU's position in global competition (Burgelman, 2010). As the expectations differed, so did the envisioned solutions.

The future circulating in the corridors of the Berlaymont building in Brussels was less concerned with how to optimize scientific data for automatic machine analysis on a global scale, and more concerned with ensuring the competitiveness of the EU and the "European Research Area" (Burgelman et al., 2010)[140]. Part of this vision, according to Director General Robert-Jan Smits[141], was to make sure that the return on the EU's billion euro investment in European science was as high as possible. This was envisioned to be achieved by making as much as possible of the data produced within the EU to be made available to actors within the EU – like researchers, but also companies and other private actors (EC, 2016). Making data openly available, was, according to Robert-Jan Smits, motivated by a "responsibility to the taxpayer[142] which spends the money", which meant "that you really make sure that you are indeed using that money in the most efficient way [...] as public authority, you have to harvest your investments"[143].

[140] see also chapter 2
[141] Interview with Robert-Jan Smits
[142] A view also espoused by Neelie Kroes (EC, 2010)
[143] Interview with Robert-Jan Smits

The role of research data is important here because, historically, measuring the return on investment in science has been a difficult task (Godin, 2009). While investments have been easily quantified in dollars, euros or any other currency, the return – or output – has been more elusive and harder to quantify. Not for a lack of trying though. The attempts at quantifying scientific output have been many, from measuring numbers of publications, citations and patents produced, to the development of indexes like the Impact Factor, the H-index and a number of university rankings. Further, there are examples of attempts to establish correlations between science investment and other more general societal phenomena like "innovation" (Godin, 2009). While these indicators of scientific output – or return on investment – have not escaped scrutiny or enjoyed much legitimacy within the scientific domains they were made to measure[144], up until this day they have remained in use as measures of "return on investment" in science among policymakers and institutional actors (Godin, 2009; Rieder and Simon, 2016).

The increasing prevalence and attention towards data as an output of scientific conduct, nurtured the expectation that data could become a regulatory object enabling a quantification of knowledge production (Rieder and Simon, 2016). This is illustrated by the way Robert-Jan Smits talked about the efforts to regulate and "harvest" the data output of European scientists, expressing a regret that the Commission didn't "pay more attention" to scientific data earlier. He said, "Of course it was not possible because the digital technology was not there", and "that before this whole [Open Data] policy came into place, something like 90 or 85% of the data just disappeared". Thus, with the vision of a data-centric science came also the vision of a more accountable science.

The transition to data-centric science (big data science, e-science, science 2.0, open science, and so forth) thus spurred a new and quantified view of knowledge production. With this new awareness came also the notion of waste, that all data that wasn't saved and made Open effectively "just disappeared" when it could have been put to use. This imagined loss, combined with the notion that "data is the new oil"[145], and the EU's "responsibility to the taxpayer" to maximize return on investment, made it seem irresponsible to not do something about this waste (EC and PwC, 2018), to "let those [data] just sit on local systems or just go into the wastebasket"[146]. In the words of Robert-Jan Smits, the investments of the past needed to be harvested in the future – and for the EU, FAIR would enable that harvest (EC, 2018).

[144] See DORA (n.d.) as an example
[145] Interview Robert-Jan Smits
[146] Interview Robert-Jan Smits

So although the bureaucrats in the Commission and the original proponents of FAIR shared some similar expectations, their visions ended up being significantly different. The future that the bureaucrats feared was not so much one in which data became unusable for machine analysis, but one in which data became lost and unavailable for exploitation in the EU. An increased production of data in the EU would not pose a problem in itself, the way they saw it – instead it was perceived as an asset, albeit an underutilized, not-yet-harvested one. And since the problem differed, so did the solution. Where the proponents of FAIR imagined an infrastructure for standardizing data quality, the bureaucrats of the Commission imagined a "Cloud" where European data would be stored and shared "openly", hindering it from disappearing into "local systems" or "the wastebasket", as Robert-Jan Smits put it. For this, they didn't need (or want) a semantic web, so why did they appropriate the FAIR principles in their science policy then?

According to Niklas Blomberg, who participated in some of the first meetings between the FAIRport initiative and the bureaucrats in the Commission, there were two main reasons:

> Robert-Jan Smits saw the potential in the acronym and the momentum. Because he wanted to drive Open Science, he was happy to hang on to the coattails and actually give the occasional gusts that held the whole thing[147].

In spite of the FAIRport initiative having little to do with the Open Science movement, the acronym and its components were flexible enough to be interpreted in a way that was congruent with the idea of Open Science – at least by the bureaucrats in the Commission. The way Niklas Blomberg put it: "The beauty of the acronym at that level is that you go to a science policymaker and they completely understand". But the original proponents of FAIR would not agree that the science policymakers understood FAIR – in fact, they would claim that policymakers had *mis*understood FAIR (see Mons et al., 2017; 2020; Jacobsen et al., 2020). What Blomberg then points to, was not an agreement or an understanding, but a *sense* of agreement and understanding. FAIR was, for the policymakers, easy to make sense of as a part of their own vision. In other words, FAIR afforded a vision of Open Science being fair, findable, accessible, interoperable and reusable. The semantic web technologies and other details of FAIR, as envisioned by its original proponents, was evidently less important for the bureaucrats in the Commission and could ultimately be disregarded.

The original FAIR proponents, on the other hand, were not willing to disregard any visionary differences. The original proponents of FAIR were early to note that their acronym was attractive to actors with different visions and agendas than themselves, and that it was being

[147] interview Niklas Blomberg

appropriated. In a 2017 paper titled *Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud* (Mons et al., 2017), the authors proposed "to clarify both what FAIRness is, and is not". Among the several "clarifications" in the paper were two that applied particularly to the differences with the EOSC: human- versus machine readability, and Open versus FAIR (Mons et al., 2017). Notably, the authors did not completely reject neither human readability nor Open as part of FAIR, but presented them as secondary concerns that *could* align with the goals of FAIR, even if they weren't parts of it originally.

This conflict continued into the writing of the 2018 Expert Group report titled *Turning FAIR into reality* (EC, 2018). According to the expert group's meeting documents, the assignment given to the group from the Commission was to "Move FAIR from concept to operation reality that serves open science" (FAIR EG, 2017) – in other words, to make FAIR work for the Commission's vision of Open Science. The attempt to fold FAIR into Open did not come without some form of resistance. In an open working document (FAIR EG, 2019), original proponents of FAIR – like Barend Mons, Michel Dumontier and Susanna Assunta Sansone – voiced their concerns that the report would neglect the aim of machine readability[148] and that it conflated the aims of FAIR and the aims of Open (FAIR EG, 2019).

In the draft, under the heading "Definitions of FAIR", the report authors stated that "FAIR is not limited to its four constituent elements: it must also comprise appropriate openness". This spurred a comment from Barend Mons, saying: "Level of openness is a matter of (EOSC) policy, not of the FAIR principles" (FAIR EG, 2019, p. 63). As a result of this, the formulation that FAIR should entail openness was then later taken out of the report, replaced with the statement that "The concepts of FAIR and Open data should not be conflated" (EC, 2018, p. 21). Rather, the final report suggested that as data would become more FAIR and/or more Open, the concepts would overlap, as suggested in figure 15.

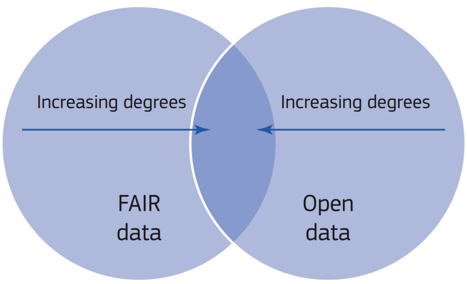

Figure 15
(EC, 2018)

The product of these discussions was thus that in the context of the European Open Science Cloud and the EU's science funding frameworks *Horizon 2020* and *Horizon Europe*, FAIR and Open were presented as separate but complementary entities that could converge towards a shared optimum. Effectively this implied that working towards a FAIR future meant working towards an Open future, and vice versa – just with an even more distant horizon.

[148] Barend Mons stating that "the principles were first and foremost formulated for machines" (FAIR EG, 2019)

But while the proponents of EU research policy and the original proponents of FAIR had found a shared vision, there were still those who didn't subscribe to it.

A reusability future

Both the proponents of FAIR and the bureaucrats in the Commission based their visions of FAIR on established expectations to coordinate the actions within and between each group, but those outside of the core FAIR group and the European Commission were not all readily accepting these visions of Semantic Webs and Open Science. According to head of the ELIXIR infrastructure, Niklas Blomberg, this was an issue that arose soon after the FAIRport meeting of 2014 had concluded. When the prospective authors came together to write the first formal publication of the FAIR principles, a divide between them became apparent. According to Niklas Blomberg "[the Wilkinson et al. (2016) paper] was a very difficult paper to write. The reason why it was difficult is because there was a huge tension [...] between the I and the R – between the interoperability and the reusability."

The tension between Interoperability and Reusability was in broad terms a conflict between what the ultimate vision of FAIR should be, whether it should be an implementation of the Semantic Web, or if it should be a set of guidelines to increase the reuse of data, regardless of technology. Niklas Blomberg believed it should be the latter, and explained his conflict with Barend Mons this way:

> Barend and I had some massive conflicts around it, which was content conflicts. They come down to the practicality of the whole thing. [...] He genuinely believed – and I think still believes – in that vision of a semantically interoperable world where everything can be abstracted up to a foundational format. [...] I think it's a great idea, but actually my mind works differently. I lead a large organization. The reason why I lead a large organization is because I'm operationally focused and I enjoy seeing things being put to work and solve real problems. Theoretical perfection is nice but it's not half as nice as something that works and gives results. [...] I don't know if this was genuinely or just the way it appeared, but a lot of people – me included – had the impression that Barend tried to then move from the principle to dictating the technology choice around semantic web technologies and the particular flavor of semantic web technologies as the only way of being FAIR[149].

The way Niklas Blomberg presented the two parties – the I and the R "camp" – the defining difference was between visionaries and the pragmatists: "it's a difference between the search for a theoretical optimum and a group that says, 'But hang on a second, we only

[149] Interview Niklas Blomberg

have a million euros'"[150]. The pragmatists didn't believe FAIR was useless, but that the grand future of the visionaries was unreasonable and unrealistic. Another member of the FAIRport initiative, and co-author on the first publication, Myles Axton, shared this view, albeit from a different position than Niklas Blomberg. While Niklas Blomberg grounded his view on FAIR in his position as the head of a research infrastructure, Myles Axton did so in his experience as an editor of a scientific journal. In his opinion, the vision of people like Barend Mons was partially based on a misconception of the role of publishing in academia:

> What a publisher gives is not a stamp of quality. It's the freedom to move on and to exchange the complexity of your project for the simplicity of the report of the project. [...] So, in a way, the potential for FAIR data interferes in some fields with the process of using publication to forget a project because the data may never have had any value to anybody to begin with, and the data may not be even what it's supposed to be. If you don't have data that's good enough quality to generate some of the conclusions that you generated in that paper, what is gained by putting headers, data types, storage, backup, and connectors onto that data? It was there for one purpose, and now is it really fit for another purpose?[151]

According to Myles Axton's experience, the scholarly communication that the proponents of FAIR were hoping to revolutionize by making data a "first class citizen" (RDA, 2014a), didn't work the way they thought. Most researchers, Myles Axton claimed, wouldn't be interested in making their data FAIR. Rather, they would generate their data to produce a paper and then prefer to forget about it. That being said, Myles Axton was not against FAIR – he believed it had value and

> that for some types of research – where there is a community that needs standard identifiers and where the resources are big and which are used over multiple years and in fact are stable like genomes and geophysical data and chemical libraries, even medical entities and medical records – we will need to have interface language improvement between human and machine. That will include licenses, that will include semantics, and will include tools. But ultimately, it's going to be the quality of the tools demanded by the user group that determines whether the particular infrastructure for FAIR principles will be useful or onerous – time-consuming.[152]

In other words, Myles Axton rejected a grand vision of FAIR: believing that it had use cases, but that those were limited to a set of contexts – that not all scientific data would need to be made interoperable in a Semantic Web, and that sometimes it wouldn't make much within the epistemic living spaces of researchers.

[150] Interview Niklas Blomberg
[151] Interview Myles Axton
[152] Interview Myles Axton

A similar sentiment was presented by Christine Borgman at the 4th RDA symposium in 2014, where she agreed with Barend Mons that something needed to be done to the way scientific data was handled, but that

> the challenge is that we need to build an infrastructure around the way that the scholars and the research community really work, not the way we would like them to work, not the way we think they ought to work, but the way they actually do work. (RDA, 2014b, 23:20).

Looking at the way researchers appropriated FAIR after its introduction, it becomes clear that there was something to Blomberg, Axton and Borgman's objections. As noted in chapter 3 of this thesis, Barend Mons and other proponents of FAIR spent a significant amount of time attempting to gatekeep FAIR from people with the "wrong interpretation" – seeing the thousands of citations on their 2016 paper as a sign that people were only citing out of convenience and that they hadn't even read the paper (Mons, 2020). Many, if not most, of the people citing the 2016 paper were not citing it because they had implemented (or even aimed to implement) semantic web technology in their data practices, but rather to state that they viewed generic findability, accessibility, interoperability and reusability[153] as a good to aspire to.

The prevalence of the idea that FAIR could stand for nothing more than the words constituting the acronym[154] has also been a recurring finding in chapters 3, 4, 5 and 6 of this thesis. Researchers, data stewards and others who worked to promote FAIRness often did so using their own interpretations of what it should mean to be FAIR, adapting it according to their own needs.

After asking Niklas Blomberg what he thought about the number of citations on the Wilkinson et al. (2016) paper, he expressed that he didn't share Barend Mons' concern. The way he saw it, the people citing Wilkinson et al. (2016) could be seen as

> a research community coming together to define how to publish that data for reuse [...] by communities. If you look at the citation of that original FAIR paper, it's thousands of papers. There is a very common pattern from a cursory scan of titles which is "community A wants to make the community data more reusable for reference within the community because we think we can by pooling the data, learn something of the whole body of experiments" [...] and they are coming together to define good data management practice and develop that. When they do so, they write a paper around that good data management practice, and then typically they cite the FAIR paper. [...] But I think that's a pattern I see from these citations. [...] Clearly, it's struck a chord somewhere[155].

[153] As opposed to the stricter definitions of these qualities by the FAIR principles
[154] As opposed to the vision of a Semantic Web infrastructure for research data
[155] Interview Niklas Blomberg

Notably, what Niklas Blomberg noted about the people citing the FAIR paper was that they were not doing it necessarily because they were following the principles, but because they were trying to "define good data management practice" within their field, and then cite the FAIR paper to signify that that's what they were trying to do. What this citing practice points to, is that for many researchers, FAIR seemed to be synonymous with "good data management" – similarly to how data stewards were using FAIR[156], or similarly to how people involved with making FAIR metrics defined FAIR[157].

A common denominator between these groups of actors was that – unlike the original proponents and EU bureaucrats – they seldom invoked any sense of an expectation or a vision of a science that was radically different from the one that they're operating within at the moment. To a larger extent they were placing FAIR within the context of their present – what challenges they were encountering, what needs and desires they had then and there. The people who were more "operationally focused", as Niklas Blomberg put it, did not have the affordance of grander visions with a more distant horizon – they were not visionaries like the original proponents of FAIR. Niklas Blomberg characterized the difference between Barend Mons and the people trying to implement FAIR as an inherently conflictual one, saying that "people who are running resources invariably rub very hard against him, because nothing is impossible for the man who doesn't need to do it himself". In other words, those with concern for a more immediate future did not afford a substantially different future like the one the EC, Barend Mons and colleagues were peddling. As one researcher I spoke to put it: "I like the idea of Open Science, but, on the other hand, science is on a dead end on how many responsibilities are on a person"[158]. Misalignments with this researcher's epistemic living space – in particular the lack of free time in their work – meant that visions of a better future remained a low priority when it came to their scientific practice.

Thus, a FAIR future for the pragmatists was not the distant and visionary semantic web for research data, but one that was more aligned with their present constraints. A FAIR future for the pragmatists, then, largely became a future not too dissimilar from their current situation, but with gradual or limited improvements in data management[159]. So even though the futures of the visionaries and the pragmatists didn't align, what ensued was not a rejection of FAIR, but rather an assimilation of it: not "unFAIR", but FAIR on other terms. Before we get back to this point, I will introduce another group that promoted FAIR with their own vision of its future.

---

[156] See chapter 4
[157] See chapter 5
[158] interview Fred
[159] As seen in chapter 4

An ethical and equitable future

When FAIR started gaining momentum after its initial launch and uptake by the European Commission, some communities started becoming concerned about what a future FAIR science could mean not only for them, but for others as well, and in particular for science as a larger whole.

One of the early examples of this can be found in a Wellcome trust report that was authored with the aim to improve data sharing during global health crises (Pisani et al., 2018). The report was based on the experience with the Ebola epidemic of 2014, where data sharing wasn't happening at the scale that the report authors claimed was necessary to deal with the disease properly (Pisani et al., 2018). Part of the cause for this sub-optimal level of data sharing was in the aftermath of the epidemic claimed to stem from poor ethical and equitability procedures, overlooking the differences in epistemologies and resources between the west and the rest (Pisani et al., 2018). The authors of the Wellcome report thus reasoned that the future of data sharing should be more than just FAIR: it needed to include ethics and equitability, to be "FREE-FAIRER" (Pisani et al., 2018)[160].

A similar sentiment was later presented by a different group of academics belonging to the Global Indigenous Data Alliance, who launched their "CARE principles" during the Research Data Alliance's 12th plenary in Botswana in late 2018 (GIDA, n.a.). The principles, which stood for "Collective benefit", "Authority to control", "Responsibility" and "Ethics" came as a reaction to the lack of concern for indigenous data in the FAIR principles, suggesting researchers and institutions should "be FAIR and CARE" (Russo Carroll et al., 2020; 2021).

This type of expansion of the FAIR principles by adding new letters to the acronym or adding companion acronyms was not unheard of after the launch of FAIR. In fact, these additions were so prevalent that it was suggested in the discussions around the 2018 High Level Expert Group report *Turning FAIR into reality* that added letters and acronyms was a sign that FAIR *needed* to be expanded (FAIR EG, 2019). Ultimately this notion was scrapped after a, by now familiar, protest from Barend Mons that "personal 'extensions' to the FAIR principles are in many cases because they did not read or understand the FAIR principles properly" (Mons, 2018)[161]. To make the case clear, Barend Mons expressed that additions like "ethics, equitability issues and timeliness of sharing are expressly NOT covered by the

[160] The acronym in full stands for Findable, Rapidly available, Ethical, Equitable, Forever, Accessible, Interoperable, Reliable, Economically viable, and Reusable
[161] In the final report it says: "the term FAIR is widely-used and effective so should not be extended with additional letters" (EC, 2018).

principles and are (again) policy choices." (Mons, 2018). The disregard for ethics and equitability was, in other words, not entirely unintentional.

All of these cases were very much in line with the findings presented in chapter 3, that Barend Mons and other original proponents of FAIR were trying to minimize the "creolization"[162] of their concept by gatekeeping and separating FAIR from, for example, "policy issues". While this had relative success – in terms of avoiding additions to the acronym in the majority of settings – it did not rid the acronym of its interpretive flexibility, or its buzzwordiness (Bensaude Vincent, 2014).

A particularly thorny quality of the acronym was that it was imbued with a moral connotation through its reference to fairness. On the one hand, as FAIRport member Tom put it, "there are people who say that it's nasty to be mentioned as unfair. But I think that is part of the success that people don't want to be unfair"[163]. On the other, according to Barend Mons, "One downside of the catchy acronym 'FAIR', is, of course, it has an ethical ring" (Mons, 2020; 1:15:28). Choosing "FAIR" as an acronym thus introduced a tension, or a double-communication even, as it implied a certain morality that the creators of the acronym later actively tried to take distance from.

But this lack of concern for ethics and equitability in the original vision of FAIR was not only seen as an opportunity to introduce it (through initiatives like FREE-FAIRER and CARE), but, by some, also as a potential for harm. One of these scholars was Mirjam van Reisen, who at the time of the original publication of FAIR in 2016 had just become chair of *Computing for Society* at the Leiden Institute for Advanced Computer Science[164]. When she first heard about FAIR, it piqued her interest in the possibilities it could produce, but at the same time, she was skeptical, saying that

> I was also thinking if the European Union is going to make this a condition for science, who can fulfill those conditions and what is going to be the semantic dominance of the application in terms of what, as science, we will discover or will be able to discover?[165]

In other words, Mirjam had a vision of a FAIR future that could be exclusionary towards certain demographics and epistemologies. But instead of rejecting or opposing it, she chose to become part of the FAIR movement and advocated for its implementation through the GOFAIR network championed by Barend Mons. Through her position in GOFAIR she worked as head of the "implementation network" for Africa – a role that included building

[162] Referring to a multiplication of interpretations and implementations. See chapter 3 and Wittenburg and Strawn (2018)
[163] Interview with Tom
[164] Interview with Mirjam van Reisen
[165] Interview with Mirjam van Reisen

FAIR infrastructures for sharing of data in the health sector in countries like Uganda, Kenya and Ethiopia (GOFAIR, 2019). But why would she work to promote something she was so concerned about?

The way Mirjam van Reisen explained it, her involvement was motivated by an idea that it would be better to change FAIR from the inside at an early stage, than to criticize it from the outside when it was too late. She explained it this way:

> [FAIR] has really the potential to be very revolutionary. I think that is the challenge, because it has a revolutionary potential. It is at this point where that direction is being formed. That's why I was so keen to start. Actually, my intention was to make sure that with my African colleagues, we wouldn't walk behind, and then we had the advantage of having no legacy and whatever: suddenly we end up being in front. This is very confusing but it's actually important because it is within the initiation of the architecture where you set what it will be[166].

Mirjam van Reisen envisioned that FAIR would have a large impact on the future, but was under the opinion that the trajectory of this future could be influenced by an intervention at an early enough stage. In her view, then, the negative outcomes of a FAIR future were only collateral that could be avoided by infusing a different vision of the future, one where FAIR didn't cause harm, but instead protected researchers. When Mirjam van Reisen described the aims of the implementation network she was heading, she therefore emphasized the importance of data ownership.

Similarly to the Wellcome trust report with the FREE-FAIRER principles, Mirjam drew on the experiences African scholars and medical professionals had had with the extraction of data from the continent by western scholars and institutions[167]. Seeing the negative effects of data sharing across the west/rest boundary, she envisioned a way to make data sharing work for the benefit of non-western scholars. If data sharing across the west/rest boundary were to have any future, it would, in her opinion. have to be based on mutually beneficial relationships. This, she believed, could be achieved through the access control that was already part of the FAIR vision.

At the same time, she was wary of the negative relationships a FAIR future could entail. This tension between the potential for harm and potential for good acted as a double motivation for Mirjam. Her engagement with the FAIR movement was motivated both by a wish to avoid an undesirable future, and a wish to attain a desirable one:

[166] Interview with Mirjam van Reisen
[167] Interview with Mirjam van Reisen

> The issues that came out again and again, why the African colleagues wanted to participate and why we made so much pace was the localisation, the ownership, and the responsibility, and the co-creation. That is possible in FAIR if that is given that space. But the danger is in overselling it as something that is representative of everything, which is such a completely non-agentic, deterministic, and crazy idea. That is why people also don't want to do it. We need to find the humility in it, but the thing is that for my African researchers, it also really is a possibility to reclaim. The thing that is in essence the problem is also potentially the strength[168].

What Mirjam was saying was that the motivation of her African colleagues for appropriating FAIR, was a desire to reclaim control over the data they produced, and to ensure that it would not be extracted by others without consent or compensation. The future Mirjam envisioned for her African colleagues, and the future they envisioned for themselves, was one in which they were empowered to combat the type of exploitation they had experienced historically[169] by highlighting the A in FAIR[170]. Rejecting the notion of Openness, or free and unrestricted sharing of research outputs – as promoted by the EC's interpretation of FAIR – they embraced the difference between Open and FAIR, namely that research outputs should be made accessible, but under the data owner's control[171].

Even with the embrace of FAIR, the future vision represented by Mirjam van Reisen was a far cry from Barend Mons'. While they shared the perceived importance of access control, they had stark disagreements on what FAIR should and could be in its fullest extent. In van Reisen's opinion, the underlying ontological politics of FAIR were incompatible with the continued existence of African ontologies[172]. One example she used was her experience working with one of her PhD students, Raymond Mubaya (2020), who had been studying the phenomenon of the Njuzu – a mermaid-like spiritual being – in Zimbabwe. In western countries, she said, the Njuzu would be reduced to imaginaries and beliefs, but for people in Zimbabwe, the Njuzu were real[173]. Drawing from this experience, van Reisen claimed, she got the idea that "we must also have similar problems probably in Europe. We just haven't identified them".

This idea that a FAIR epistemology/ontology could oppress local ones was not uniquely held by Mirjam, but was shared by other colleagues in the humanities and part of social sciences.

---

168 Interview with Mirjam van Reisen
169 The ebola epidemic and COVID pandemic being two recent examples (Maxmen, 2021)
170 Interview with Mirjam van Reisen
171 A notion central to the CARE principles as well. Notably, the proponents of CARE didn't find access control represented to a satisfactory degree in the FAIR principles, like Mirjam van Reisen did.
172 Interview Mirjam van Reisen
173 A view echoing Law's argument in *What's wrong with a one-world world?* (2015). Also, see chapter 6 of this thesis for a similar discussion.

As FAIR became increasingly present as part of European science policy (also other funders than the EU), more scholars got the impression that demands for FAIR science would not disappear at first occasion – and they also started to become concerned with what would happen should this “life science hobby” (Mons et al., 2017) become a governing force in less data-intensive epistemologies (see Tóth-Czifra, 2019; ALLEA, 2020).

The concerns voiced by scholars in the (qualitative) humanities and social sciences were mainly centered around the potential threat FAIR posed to “epistemic diversity”, in particular epistemes that didn’t share the positivist inclinations of Barend Mons and his colleagues (Jones et al., 2018; Juros, 2022). Jones et al. (2018), understanding qualitative social sciences and humanities to be characterized by their “constructivist epistemology”, offered this explanation to the reaction:

> A purely positivist approach to research should have no epistemological problem sharing data. [...] In contrast, constructivist epistemology is likely in pure form to reject the notion that qualitative data could be re-used in appropriate fashion, since the data themselves are generated through the relational process of an individual researcher engaging with research subjects.

This fear that a FAIR future would be a future where social sciences and humanities became marginalized was not entirely new, but drawing on existing anxieties within the communities, as seen in Emma Uprichards concerns for social sciences in the face of the 2010’s hype around “Big Data” (Uprichard, 2013).

Scholars like Bezuidenhout et al. (2017), Poirier et al. (2019) and van Reisen et al. (2020), took a broader approach, arguing that the threat to epistemic diversity went along a range of other dimensions than just the disciplinary, like economic, geographical, cultural. These scholars were envisioning a dystopian future where FAIR, in its logical extension, marginalized epistemic cultures that didn’t align with positivism or data-intensive methods – at worst, deleting epistemic diversity or committing what Caroll et al. (2020) had dubbed "epistemicide".

While the authors left little reason to believe that they anticipated this dystopia to actually materialize, they mobilized the vision as a justification for their involvement in supporting the FAIR principles – taking the stance that they could change it from within (Bezuidenhout et al., 2017; Tóth-Czifra et al., 2019; Poirier et al., 2019; ALLEA, 2020; van Reisen et al., 2020). Policymakers and institutions, they argued, should attempt to curb the negative potential of FAIR by interpreting it differently than the original proponents suggested. Further, they argued, policymakers should avoid applying the FAIR principles in the same way to different disciplines, geographies, cultures, and so on. For example, Poirier et al. (2019) suggested that in the context of anthropology, the R in FAIR should stand for Re-Interpretable instead of Reusable; van Reisen et al. (2020) suggested that African communities should be encouraged to select which elements of FAIR they wanted to implement themselves; and Bezuidenhout et al. (2019) suggested a shift from production and sharing of FAIR data to a focus on the utilization of FAIR data as a tool to minimize “the digital divide” between the west and the rest. All of these approaches, they claimed, didn’t need a rejection of FAIR, but were possible just by interpreting FAIR differently.

Understanding change as domestication

The processes I’ve described above, of different actors and groups joining the FAIR movement, have been described as an “uptake” of FAIR (Mons et al., 2017). This way of describing it presents it as if people were becoming followers of the original principles. But, as should have become clear from this text, appropriating the word “FAIR” didn’t always mean that those actors agreed with Barend Mons. Keeping that in mind, what’s been happening over the years since 2014 seems like less of an “uptake” and more towards what Wittenburg and Strawn (2018) called “creolization”. That being said, I would urge caution to buy into that idea too readily. When Wittenburg and Strawn (2018) talked about creolization, they imagined groups of actors that worked on different implementations of the same idea – aligned in ideals and goals, yet misaligned in terms of practice until further notice. The processes I’ve described, on the other hand, show groups of people misaligned in both their practices *and* ideals and goals. Yet they remained aligned around a term, or a buzzword: FAIR.

I therefore suggest that we look at this process not as uptake, and not a process of creolization, but a process of domestication.

Now, domestication as a theoretical concept was originally developed in the 1980’s to study particular media technologies – like the TV and the computer – as an alternative to more linear and deterministic models of technological uptake (like the diffusion model) (Berker, 2023). The basic idea was that as users brought technologies into their homes (domus), they would not readily accept them at face value, or follow the user scripts (see Akrich, 1997) of the developers. Instead, users would negotiate with and adapt the technologies, using them in ways different than intended – like putting a curtain in front of their TV’s or using the text-based SMS technology to send “pictures” (or emoticon[174]). Part of the logical extension

[174] See “:-)” as an example

of this idea, was that if the technology wasn't interpretively flexible enough to allow domestication, or to become "tamed", the user could end up rejecting it altogether (Berker, 2023).

The concept of domestication resonated with the growing field of Science and Technology Studies (STS), who were trying to break down the dichotomy between "user" and "technology" (Berker, 2023). As they started applying this concept, STS scholars ended up "domesticating" it themselves, applying it to cases outside of media technologies, and increasingly using it to explain how users related to infrastructures, thus marrying the concept to the work of Star and Ruhleder (see 1996) (Hartmann, 2020; Berker, 2023). As the concept was appropriated by STS scholars, they applied "domestication" to analyze processes of design and implementation beyond media technologies, including energy efficient housing (Korsnes et al., 2018), milking robots (Egseth, 2018), energy monitoring devices (Hargreaves et al., 2010), and analytical concepts (Hartmann, 2020).

What these studies suggested was that the understanding of a domus should not be limited to homes in the conventional sense – like the houses in early domestication studies (see Silverstone and Hirsh, 1992). Instead it should be seen as a collected and connected network of cognitive, symbolic, practical and material components (Sørensen, 1996) – a form of sociotechnical habitus (or rather "habitat" in this context) (Weber, 2005)[175]. This understanding of home or domus has significant overlaps with the concept of "epistemic living space" as suggested by Ulrike Felt (2009) to study academic contexts. In this concept, "living spaces" are not only seen as spatial environments, but "habitations" that contain elements along epistemic, spatio-material, temporal, symbolic and social dimensions (Felt, 2009; Felt and Fochler, 2011). As this chapter deals with academics, I will use the term "epistemic living space" henceforth to refer to homes/domi when applicable.

Acknowledging the complex interconnectivity of what constitutes a domus, the concern for fit between the different components becomes foregrounded. Domestication then becomes less of a process of addition or inclusion, and more a process of negotiation between components to make them fit or align with each other. In other words, when something is introduced into a domain, it either triggers a process of adaptation and change among the components of the domain, or it triggers a rejection (Berker, 2023). In this sense, domestication theory has a lot in common with the epistemic underpinnings of infrastructure studies.

---

[175] It is tempting to use the term "domain" here, but I will refrain from it due to the established use of the term to describe scientific domains and domains of knowledge (a practice I have critiqued in the previous chapter).

As a research approach, domestication studies differ from infrastructure studies in a central manner: where the latter focuses on the infrastructure at hand or the infrastructurer, domestication studies focus on the infrastructuree (Berker, 2023). Domestication is seen as a process not driven by the designers of infrastructure, but the users of it (Berker, 2023). That being said, domestication is not a unidirectional process. After an actor has domesticated a technology or infrastructure, information of the domestication *can* be used by the designer(s) to feed into the continued design of the technology or infrastructure, leading to a continuous process of design – resulting in a form of co-production that blurs the lines between user and designer (Weber, 2005).

Domesticating FAIR

How does this matter to FAIR and how does the future play a role in it?

When we look at the different groups presented in this chapter, it becomes clear that while all of them subscribe to the idea that a FAIR future is both desirable and attainable, the understanding or interpretation of what that means is quite different. What happened was that these actors managed to agree that a FAIR future could be good, but disagreed on what it should look like in order to be good. What kept them together was the organization around the acronym “FAIR”. In the words of Bensaude Vincent (2014) they established and maintained an unstable collective. Collective because they all promoted a future under the same name: a “FAIR” future. Unstable because while their future was nominally the same, it was conceptually different, making their interactions characterized by controversy and disagreement. As I will argue, both the instability and the collective are products of domestication.

In chapters 3 and 6 of this thesis, we’ve looked at FAIR as a process of sociotechnical change and infrastructuring, but mainly from the perspective of the drivers of change and designers of infrastructure. The general finding from those chapters was that Barend Mons and other proponents of FAIR generally struggled to account for multiplicity and to create alignment with other contexts than the one FAIR originated in. As a result, *implementation* of FAIR has been met with a series of difficulties prohibiting the realization of Barend Mons' FAIR future.

But in spite of difficulties and misalignments, the implementation of FAIR had not been met with significant amounts of outright rejection. Actors with distinctly different, and sometimes mutually exclusive, visions of the future rallied around the concept, supporting it through co-authorship on papers, political promotion, participation in conferences and organizational work. Maintaining different visions of a FAIR future, these actors simultaneously worked against and with each other.

An important factor in enabling this was the interpretive flexibility of the acronym FAIR, often taken to mean findable, accessible, interoperable, and reusable, and not including the actual FAIR principles of which there are a total of 15 (Wilkinson et al., 2016), nor the philosophical/epistemological/ontological underpinnings of the concept[176]. This allowed policymakers, scientists and everyone between and beyond to subscribe to a more general and generous interpretation of what it would mean to be FAIR. Rarely ever is principle A1 (as an example), that "(meta)data are retrievable by their identifier using a standardized communications protocol" (Wilkinson et al., 2016) mentioned in papers that cite FAIR or reports that promote it. Similarly, few contemporary proponents of FAIR envisioned a semantic web of data as the end goal of the FAIR movement.

In practice, FAIR seldom embodied the characteristics of a strict and elaborated concept[177], and increasingly resembled something of a buzzword (Bensaude Vincent, 2014) containing a multitude of meanings. The principles had, according to one of the original authors[178], "no disagreeable thing" in them, and only consisted of positive attributes that were easy to subscribe to and align with. After all "nobody wants to be unfair"[179], and nobody would oppose findability, accessibility, interoperability, nor reusability. But evidently this didn't translate into people following the original intention of FAIR. As attention was drawn to the precise meaning and the future implications of these terms, the discord between these diverse proponents became apparent: their futures did not align.

While this misalignment, and, essentially, disagreement, had the potential to hinder the formation of a wider movement around FAIR, actors like Robert-Jan Smits, Niklas Blomberg and Mirjam van Reisen still found a way to organize themselves around the concept – not by abandoning their visions, but by interpreting FAIR in a way that allowed it to be domesticated into their epistemic living space. The result was that the FAIR movement then consisted of a group of actors with different – and often contradictory – ideas about how the future could and should be.

Considering the importance of futures for identity and agency at individual as well as collective level, misaligned futures can pose significant challenges for social and political movements such as the one surrounding FAIR (Rist, 2007). In spite of this, the disparate

[176] See chapter 1 and 6 for a discussion on this topic
[177] Even though it most certainly is
[178] Interview Tom
[179] Interview Tom

groups promoting FAIR managed to maintain a level of unity, largely avoiding splinter groups or new schisms of FAIR[180].

Still, many of the people subscribing to and promoting FAIR often made a point of distinguishing themselves from the group that originated the FAIR principles. As an extension of this, they made clear their own understandings or opinions on what FAIR should mean within the interpretational limits of the acronym: for the EU bureaucrats, “findability” and “accessibility” meant that research data should be shared with “the taxpayer”, not that data should be indexed and assigned a DOI and a license; for the pragmatists, “interoperability” meant for data to be interoperable within a project or specific field of study, not interoperable across disciplines on a semantic level; for the ethics and equitability oriented, “accessibility” meant the option to apply access control in terms of social justice, not to apply access control in terms of privacy or corporate IP. These distinctions were not contained to rhetorics, but carried over into practice as well, something that became particularly tangible in the discussions around making FAIRness metrics seen in chapter 5, and the practices of data stewards seen in chapter 4.

The domestication by different groups and actors meant that the number of actors supporting FAIR grew – but as it grew it also increased the tension within the FAIR movement. As discussed in chapter 3, the reaction from Barend Mons and the original proponents of FAIR was to reiterate their original intentions behind the acronym and the concept to avoid “creolization” (Wittenburg and Strawn, 2018) and “diverging interpretations” (Mons et al., 2017). The unwillingness to cede on the core components of the principles meant that the process of domestication of FAIR didn’t take on a symmetric or mutual nature between those who originally designed them, and the ones who were intended to use them. The original proponents remained reluctant to accommodate the domesticated versions of FAIR, insisting on their own version as the correct one.

Unlike the classic cases of domestication, where the intellectual property rights over technological artifacts are more clear and simple to enforce, FAIR is not a trademarked or patented product that anyone can claim the right to control. Thus, Barend Mons can’t (even though he certainly has tried) make people not use the concept of FAIR to promote their own futures. Because of this, the unwillingness to accommodate the domesticated versions of FAIR could have the potential to lead to a takeover from actors who no longer see the original proponents to be the most fitting stewards of FAIR – further blurring the lines between “designers” and “users”.

[180] In spite of what I’ve sketched out in this chapter, there are no groups within FAIR that identify themselves as the Open group, the pragmatist group or the ethical group.

As the work aimed at "turning FAIR into reality" (EC, 2018) progressed – in particular around the development of metrics – it became increasingly clear that the lack of clear ownership and governance of FAIR was an issue with real repercussions (EC, 2021). As discussed in chapter 5, a core issue of the struggle to define measures of FAIRness was the lack of authority to settle disagreements on what FAIR was or should have been. As FAIR got domesticated by an increasing number of actors, and the meanings of FAIR multiplied, the question of authority became more pressing.

Some scholars working on the European Open Science Cloud's FAIR working group made note of this and decided to create a poll for the EOSC Consultation Day, asking the participants whether there was a need to govern FAIR, and if so, then who should be in charge of their maintenance (EC, 2021, p. 17). While the sample of respondents was small and not necessarily representative[181], they found that 60% wanted FAIR to be governed, while 16% expressed that FAIR should not change and thus didn't need governance. Further, only 7% thought that the "original authors" should be in charge of maintenance, while 77% expressed that "an international group representing all research communities should be in charge" (EC, 2021, p. 18). That being said, the report authors noted that there was no desire "to question Findability, Accessibility, Interoperability and Reuse as essential requirements and aims", but instead, "to assess the 15 guiding principles" (EC, 2021). In other words, they wanted to keep the buzzwordy acronym intact, but opened the possibility of negotiating the specific contents suggested by Wilkinson et al. (2016) and essentially attempting to settle their meaning into something adapted to the epistemic livingspaces of the users of FAIR.

As I've shown in chapters 3, 5 and 6, both the original proponents of FAIR and the original concept of FAIR demonstrated a significant inflexibility when it came to accommodating, negotiating or accepting alternative enactments of FAIR. What maintains the unstable collective around FAIR seems thus to be only the flexibility of the term itself: its buzzwordiness. The question then becomes how long this buzz can be sustained. Rist (2007) suggests there is no inherent limit to the duration of a buzzword, and Ribeiro et al. (2017) underlines that it is precisely the ability to unify across difference that characterizes a buzzword. That being said, there are signs that FAIR might be unable to sustain its internal tension indefinitely. Aside from the incompatibilities of the different futures of FAIR, as we've seen, several actors have already started defining one particular group's future out of the collective, namely the original proponents' future. In the concluding remarks of this chapter I will discuss what that might mean for the future of FAIR.

[181] 44 respondents in a breakout room about FAIR metrics at an EOSC event

Making the future

Earlier in this chapter I mentioned that the social sciences typically avoid predicting the future – and I will not readily fall into temptation myself either. Still, I will dare to speculate in some possible future presents of the FAIR movement based on the empirical and analytical work I have done over the span of my research project[182]. In my defense, this will not be the first time someone has used STS theory to predict the future, and not even the first time in this very chapter that someone has used STS theory to predict the future. The first instance of this is the Wittenburg and Strawn publication *Common patterns in revolutionary infrastructures and data* (2018), where they applied insights from Thomas Parke Hughes seminal STS book *Networks of power: electrification on western society, 1880-1930* (1983) to argue that over time FAIR would – not unlike the electrical infrastructure of 1880-1930 – come to a point of convergence where the best solutions would attract all actors to embrace a common set of standards (Wittenburg and Strawn, 2018; 2021).

Aside from missing the point of Hughes' work (1983)[183], the main issue with this description and prediction of FAIR's development is that it is based on what my empirical work suggests is a wrong assumption – specifically the assumption that the difference between actors' subscriptions to FAIR is a question of method and not of aspiration. Wittenburg and Strawn can be excused for assuming that everyone involved is working towards FAIR, because that is largely what everyone involved were saying they were. The problem is, they did not account for the possibility that although people were organized around the same acronym, they were not organized around the same vision, not sharing their expectations, not committed to the same future.

As noted earlier, futures play an essential role in coordinating actors and action (Jasanoff and Kim, 2015), and because of their entanglement with and embeddedness within the sociotechnical networks of groups and actors, they are both theoretically and demonstrably resistant to change (Lösch et al., 2019). This is what we've seen in the empirical material of this chapter, that even though actors were enrolled into promoting FAIR, they were not appropriating the intended practices, nor promoting the original vision. You can bring a horse to water, but you can't make it imagine a future drinking it.

[182] If not as a Nostradamus, maybe as a Cassandra

[183] Hughes shied away from deterministic notions in favor of a more contingent view on the development of infrastructure. A central element of his story about the development of electrical infrastructure was that "there was – and probably is – no one best way of supplying electricity", and that "despite the momentum of systems, [...] contingencies push systems in new directions." (Hughes, 1983, p. 16-17).

The result of the appropriation of the term “FAIR” combined with the durability of epistemic living spaces was a reversal of the imagined roles of FAIR implementation. Where the intention was to have enrolled actors promote FAIR (in terms of Barend Mons' vision), what we saw happen instead was a case of FAIR (in terms of the buzzword) being used to promote other actors’ futures. While the EU bureaucrats promoted FAIR, they simultaneously used FAIR to promote their own agenda of Open Science[184]. While pragmatists like Niklas Blomberg did promote FAIR, he used it to promote the reuse of the ELIXIR data holdings. While ethics and equitability oriented scholars like Mirjam van Reisen did promote FAIR, she used it to promote more equitable international relations around research data. These promotions seem mutual only until you realize that the FAIR that these actors promoted was not the same FAIR as the one Barend Mons and his colleagues imagined.

The visions of FAIR promoted by the different proponents were not the same as the original, but domesticated versions, tailored to fit the actors’ epistemic living spaces and the futures they afforded or contained. As new actors were enrolled into the FAIR movement, they continued to birth new and alternative versions of FAIR. Aside from creating a conceptual sprawl (or creolization, if you’d like), this also introduced visions that contradicted those of the original FAIR. As these newly enrolled actors started enacting their multiple FAIR futures, they invariably created tensions within the collective organized around FAIR – making both the collective and the word unstable. In spite of efforts to achieve “convergence”[185], difference persisted. The visions held by the various actors and groups proved durable in encounter with standardization pressures, and as a result, FAIR remained an “ill-structured” (Star, 1989) buzzword, and the collective remained unstable (Bensaude Vincent, 2014).

But enough about things that have happened in the past, what will happen with FAIR in the future? On a general level, we can imagine three possible trajectories: 1) FAIR sustains its buzz, 2) the buzz fades out, or 3) FAIR stabilizes (losing its buzz). To explore the future of FAIR, we will delve a bit deeper into these trajectories.

1) FAIR sustains its buzz
In this scenario, the actors stick with their incompatible futures never reaching convergence, instead using the buzz around FAIR to work towards their own goals. This future would hinder the development of the original FAIR future – an Internet of FAIR Data and Services – but could still achieve other goals, like the FAIR movement demonstrably has throughout the infrastructuring process the past eight years. While, as noted by Rist (2007), a buzz can be

[184] More on this in chapter 2
[185] Notably through the GOFAIR association and several convergence symposia

sustained indefinitely, there are some ways FAIR could lose its buzz. This could make it unable to coordinate and hold together its unstable collective – which brings us to the next scenario.

2) FAIR's buzz fades out

If FAIR becomes unable to generate the buzz it has before, it may lose its ability to sustain an unstable collective of actors. While this wouldn't necessarily mean the end of the concept of FAIR in itself, it would mean a dissolution of the collective around it, resulting in the agential loss of FAIR as the people promoting it decreases. One way this could happen is through a gradual loss of momentum (see Ribeiro, Smith and Millar, 2017). Judging by Niklas Blomberg's opinion, this might be the most likely scenario, as he claimed this was already happening back in 2021:

> I think we've reached peak FAIR. [...] It's been a buzzword that's been incredibly valuable for funding acquisition. I think that value is rapidly diminishing. That's what I mean with peak FAIR.

If the value of FAIR is experienced as diminishing to its proponents (not only in terms of funding acquisition), then they may no longer see the benefit of promoting it, as the relationship becomes less mutually rewarding. Alternatively, the tensions within the movement could reach a boiling point, leading to a more intentional expulsion of members of the collective – leading us to the next scenario.

3) FAIR stabilizes

Throughout the history of FAIR, there have been few moments of stability. Perhaps the closest it got to one was during the FAIRport conference in 2014, but since then, it has been unable to avoid controversies and tensions[186]. As I've noted earlier, this has been in large part due to the inability of its original proponents to adapt to the contingencies of the contexts into which the concept has been introduced.

The original FAIR concept is characteristically inflexible, and has thus been unable to accommodate the domesticated versions of FAIR. Domestications of FAIR, on the other hand, do not, to the same extent, share the original concept's need for standardization. An approach to achieving a stable FAIR, could thus be to expel the original FAIR proponents from the collective. Although this may seem counterintuitive, we have already seen traces of this approach, with actors taking clear stances against the original FAIR vision. This includes Mirjam van Reisen characterizing the infrastructural ambitions of FAIR as "impossible" or Niklas Blomberg "always push[ing] back on [the] theoretical aspect of the acronym". If this

[186] As covered in the chapters of this thesis.

tendency continues, spreads and intensifies, the end result might be that Barend Mons and his colleagues become excluded, relegated to the periphery, losing their influence on the FAIR movement, having to find elsewhere to realize their future. Rid of this wicked tension, FAIR *could* stabilize around a more flexible Open, pragmatic or ethical and equitable understanding – maybe even one encompassing all three.

Concluding remarks

A question I get often – and one that readers of this thesis likely have faced themselves – is "what is your thesis actually about?". Be it a colleague too focused on the details to remember the bigger picture, an uncle who's never had it explained, or a friend who's had it explained several times but needs a refresher – the question is hard to answer. In most social settings, there's no time (nor patience, for that matter) to give a detailed account of four years of research. Therefore, the story needs to be condensed into a bite-sized snack that satiates the listener's appetite, like an academic elevator pitch with nothing else to sell than an explanation.

So what is my thesis about? Throughout my project, and depending on my audience, I've given a host of different answers: "it's a thesis on european research policy", "it's a thesis on data sharing", "it's a thesis on AI"[187], or, perhaps most commonly, "it's complicated". Taking a bit more time and space to give an explanation, I wrote in the introduction that this thesis is about the production of knowledge and the environments that govern it, about tensions, and about the messy entanglements of envisioned change and continuity of practices. In this sense, this thesis has been both a story *about* FAIR and a collection of stories told by *looking at* FAIR. Each and every chapter, and the thesis as a whole, speaks to more than just FAIR. The insights and knowledge produced and disseminated throughout this collection of texts shouldn't be mistaken as relevant only for those interested in making data FAIR, or, conversely, those interested in conserving contemporary scientific practices in spite of FAIR.

Because, although we've learned about FAIR, FAIR has also helped us learn about things beyond itself. In an attempt to articulate these insights, I will spend some time and space in this final chapter to recap and condense the previous ones, drawing these individual stories together to point to some overarching themes of the thesis, and finally presenting some suggested takeaways from this project.

Recap

While the chapters in this thesis interrogate a broad range of themes, they are also parts of a story about FAIR that – reflecting the central theme of temporalities in this thesis – have been structured along a loosely chronological narrative: from the inception of FAIR in the first chapter, to the future of FAIR in the last.

[187] For when the listener's interest is more important than accuracy

In the first chapter we explored a genealogy of the FAIR principles, starting with the dream of The Semantic Web by proclaimed inventor of The World Wide Web, Tim Berners-Lee. Kicking off from that point, I argued that the growth of sequencing data resulting from the Human Genome Project, and the following realization among bioinformaticians that data is not inherently reusable, set the scene for the combination of Semantic Web technology and the life sciences through the growing field of bioinformatics. What followed was a period where different semantic solutions to bioinformatics' data challenges were suggested and tried out – most of them to little fruition. But as time went on, the different suggestions on how to solve this challenge started to become more and more similar[188], highlighting the need for data to be both accessible and understandable if they were to be (re)used by other people or – importantly – machines. The chapter ended by positioning FAIR as a sort of culmination of these developments. While the chapter set up the question of how FAIR came to gain its later significance, that was left for the following chapters to answer.

The second chapter gave a partial but significant explanation as to why FAIR became so widespread, through following its uptake by the European Union. Partial, because it only addressed the political environment FAIR became embedded within. Significant, because the account showed that FAIR was not adopted due to its Semantic Web elements (demonstrated to be a core component in chapter one), but in large part due to a series of crises in the EU relating to its legitimacy as a supranational governing body. Notably, these crises only partially overlapped with the concerns of the people behind FAIR, as they were more rooted in concerns about federal cohesion, economic development through innovation, and the legitimacy and value of science as an institution. Coincidentally, around the same time as the FAIR principles were being articulated, and right before FAIR's country of origin, the Netherlands, took presidency over the European Council, these bureaucrats' concerns were synthesized to launch the idea of the European Open Science Cloud. The second chapter introduced the notion that crises – as collectively held imaginations about the past, present, and future – were essential for the adoption of FAIR, but that these imaginations might not be able to sustain FAIR over time. The second chapter was also a demonstration of the interpretive flexibility of FAIR, and the role this had in its uptake and promotion.

In the third chapter I interrogated the strategies employed by the originators of FAIR in their development and attempts to spread their concept beyond the context in which the idea was conceived. Despite the successful wide spread of the word "FAIR" (also outside of Europe), the reception of the concept "FAIR" was characterized by a significant amount of both tacit and explicit critiques. The explicit critiques were largely directed towards the lack of moral or

[188] or "converge" if you'd like

ethical considerations in the principles[189], while the tacit critiques took the form of people pledging support without following the prescriptions of the principles' original authors – an appropriation enabled by FAIR's interpretive flexibility. The chapter went on to describe how the original proponents of FAIR reacted to this "backlash". In spite of using a language of inclusivity, they would continuously label different opinions as a product of misunderstanding – attempting to minimize the interpretive flexibility that enabled its wide adoption in the first place. I characterized this strategy as a form of gatekeeping driven by technological determinism, deficiency narratives, and the proponents' belief in their ability to represent the needs and desires of other researchers.

Chapter four attempted to explain *why* researchers were not doing FAIR the way the original proponents desired by following the people made responsible of helping researchers become FAIR: the data stewards. Following a handful of stewards, I attributed the discrepancy between FAIR ideals and FAIR practice, in part, to the stewards' translational work as "middle actors". Based on the previously made assertion that FAIR was designed according to the assumptions of a contextually delimited group of people, I showed how the original proponents' rejection of external input was compensated for by the data stewards. Finding themselves in positions with limited agency and authority, the data stewards ended up having to adapt and negotiate principles, policy and practice, exploiting the interpretive flexibility of the concept to achieve a more or less stable alignment between the three. While literature on middle actors tends to see the adaption of policy as a failure, I suggested instead that adaption could be seen as a step in making policies implemented more productively.

In chapter five we followed a number of actors attempting to measure FAIRness. Underway we realized that while all of them claimed to strive for objectivity, they all had different reasons for wanting to measure FAIRness, and, thus, also came up with substantively different approaches to measure it. As these actors tried to establish one definitive way of measuring FAIRness, they were forced to reflect and articulate what FAIR meant for them, and what they thought it should be – making visible the conflicts and tensions that had remained tacit. As these conflicts and tensions arose, the chapter showed how various actors tried closing them down by drawing on different sources of authority, from the original inventors of FAIR to the scientific community more generally[190]. As this appeal to authority failed, it became clear that the main barrier to establishing a measurement of FAIRness was not a technical issue, but a lack of authority to claim and settle the ontological status of FAIR – to delete its multiplicity. Unable to establish a single measurement of FAIRness, the actors

[189] That the FAIR principles didn't address openness or equity concerns
[190] Represented by the Research Data Alliance (RDA)

pivoted from attempting to assess FAIRness, to using the “imperfect” measures to guide and assist the improvement of their datasets according to their own understandings of FAIRness – maintaining multiplicity.

Chapter six took a turn for the more abstract by not following the FAIR history or practices, but interrogating the infrastructural ambitions of the people behind it. Instead of looking at the current infrastructural components of FAIR, such as data repositories, standards, policies, etc., the chapter dove into the future FAIR infrastructure, or the Internet of FAIR Data and Services, as imagined by the original proponents of FAIR. Taking the ambitions of making a semantic web for scientific data seriously, I scrutinized the logics behind it, in particular the epistemic and ontological assumptions the proponents of FAIR built on. Drawing on previous conceptual and empirical works within STS, I made the argument that the core components of FAIR were misaligned with the practical ontologies of knowledge production in the scientific domain. Further, I showed how this misalignment between imagined infrastructure and practiced practice not only belonged to the abstract and hypothetical, but that its consequences manifested as practical challenges to building a FAIR infrastructure as well. While actors aiming to realize the original vision of FAIR identified these issues as practical hurdles to overcome, I suggested that it is, rather, an ethical, epistemological and ontological consideration – forcing the choice between violence or failure.

The seventh chapter was dedicated to studying the futures of FAIR and how different expectations and visions of the future played, and could play, a role in its development. In the chapter, I made the assertion that the imagination of the future is essential for the coordination of practices in the present. Identifying a number of different, and sometimes mutually exclusive, visions of the future among the supporters of FAIR, I argued that FAIR will struggle to reach the level of standardization necessary to fulfill the infrastructural ambitions of its original proponents. I traced this multiplicity of futures back to the interpretive flexibility of “FAIR” as a buzzword, which allowed actors to adapt FAIR to their own needs and desires. The chapter concluded by suggesting three future scenarios: that FAIR becomes only a vague and ambiguous notion of data quality, that FAIR becomes a marginalized and niche concept, or that it stabilizes as a more inclusive concept than today. Notably, in neither scenario does FAIR lead to the creation of an Internet of FAIR Data and Services, as imagined by its original proponents.

This FAIR story

While each and every chapter stands on its own, telling a fairly contained story, there is also insight to be gained from seeing them all together. What do they have in common, and is there something more to be said from bringing them together? Separately, they are stories about EU politics, measuring, and middle-actors – together they are the story about FAIR from principles to policy to practice. But what is that story?

As I've argued earlier in the thesis, the original proponents of FAIR tended to imagine the story of FAIR quite deterministically, from information technology developing along its inherent path, towards semantic web technology gradually being implemented and accepted by researchers, policymakers and other scientific institutions (see Wittenburg and Strawn, 2018; 2021). Quite evidently that story has been too simple.

While it might not be particularly controversial to say this as an STS scholar, the development of FAIR was not deterministic, but contingent – belonging to neither a realm of necessity, nor of impossibility. This was particularly salient through the development of the idea and the circumstances that led up to FAIR becoming appropriated by the European Commission. FAIR was not the only initiative to suggest changes in scientific practice around data sharing and management, and social and temporal circumstances seem to have been decisive in turning it into policy – and, thus, spurring its popularity. If Barend Mons was not Dutch, had not worked in the European Commission, or did not propose the idea of FAIR in the same time as the EC proposed the European Open Science Cloud, things could have been otherwise. If not for these contingencies, EU's science policy could have promoted "intelligently open data" (Royal Society, 2012), the 13 R's for good data management (De Roure, 2012), or continued promoting Open Science the way they had done before (EC, 2015a).

A non-negligible component for FAIR's spread and adoption was also the acronym's appeal and considerable interpretive flexibility. So while many, if not most, actors outside of the group of original proponents didn't subscribe to the technological prescriptions of FAIR, they were ready to promote findability, accessibility, interoperability and reusability as qualities to aspire to in scientific data management. Perhaps ironically, this has also been one of the most persistent and profound challenges to the realization of the original FAIR proponents' visions. The flexibility of FAIR made it adaptable to other contexts and practices, a characteristic that other actors made sure to exploit. This led to FAIR being interpreted and enacted in a multitude of different ways, which, in turn, undermined the infrastructural ambitions of the original FAIR proponents which depended on practices being standardized.

The multiple ontologies of scientific practice in stark contrast with the singular machine ontology of the infrastructural ambitions of the FAIR principles meant that the continued development and spread of FAIR was plagued with tensions: tension between machine-readability and human readability, tension between interoperability and reusability, tension between standards and adaptability, tension between openness and controlled access, tension between pragmatic limitations and visionary expectations, tension, tension, tension, and no resolution in sight.

Yet, FAIR has shown itself well suited to sustain this tension over time. The opportunity to imbue the word with multiple meanings led to a diverse crowd becoming involved in negotiations to define what a FAIR should and could be – and in extension, what science itself should and could be. While this negotiation has not transformed FAIR in such a way as to generate a stable consensus among the relevant actors – like researchers and policymakers – I would be reluctant to rule out a future in which this happens. FAIR thus retains its instability, but also an aspirational nature – while most agree that science is not yet FAIR, many think it still can be in the future. What this future will look like, and when it will arrive remains uncertain.

Making it otherwise?

When I do get the occasion to answer the question "what is your thesis actually about?", like I have had here, I sometimes get a follow-up question: "what should have been done differently?" or "how should we proceed?". Both are difficult questions to answer, not only because I am careful in claiming too much authority when it comes to dictating the "right" or "wrong" course of action, but also because the questions – straightforward as they seem – conceal tacit valuations of what is deemed desirable and not.

What should have been done, or should be done, quite obviously depends entirely on what the desired outcome is, and, classically, desire is not something that belongs to the realm of STS, a discipline that has traditionally been more concerned with critique[191] and "otherwising"[192] (as I myself have done)[193]. Yet, I will go beyond the boundaries of this STS tradition, venture out of this institutional comfort zone, and suggest some key takeaways in response to those who have asked, and anyone who in the future might. To do that, I will first have to articulate what remained tacit in the aforementioned questions, the "should" of it all,

---

[191] See Latour (2004) *Why has critique run out of steam?* for a discussion on this

[192] A concept lent from Jane Calvert in her talk "On not being a bioethicist: normativity, otherwising and synthetic biology"

[193] It has to be noted that there are scholars attempting to shift away from merely critiquing and otherwising and towards "making and doing" (see Lachney and Foster, 2020)

the desires of those asking. Instead of covering all possible positions and values, I will reduce it to three ideal types: the original FAIR proponents, the policymakers, and the researchers as subjects of policy.

Taking the position of the original proponents of FAIR, I will assume the establishment of a semantic web infrastructure for science as the desired outcome. In the interviews I did with the people who attended the FAIRport conference in 2014, several of them mentioned that if they would have done something differently, they would have been more patient and moved slower in terms of promoting the FAIR principles. This stemmed from a sentiment that as FAIR became picked up by the EU it wasn't "ready" yet – that when FAIR became policy, it still wasn't fully developed. While I do understand where that perspective came from, it is simultaneously evident that the support from the EU was essential in the popular spread of the principles, and that the coinciding timing of FAIR, the EOSC, and the Dutch presidency in the European Council was a type of opportunity that would not soon present itself again. Postponing the work on turning FAIR into policy would therefore be likely to limit its impact in a way that would have harmed the prospect of a successful establishment of a semantic web for science.

The idea that FAIR could have benefited from "more time" also neglects the fact that there has been ample time to develop the concept of FAIR further – and that FAIR actually has been through a significant degree of change after its uptake in EU policy. The somewhat vague and flexible nature of FAIR was also one of the reasons it became adopted so widely. Therefore, any attempt at limiting the interpretive flexibility of FAIR by taking more time to elaborate and stabilize the concept, would have potentially diminished its uptake by other actors. At the same time, the level of standardization needed to infrastructure science across disciplines, the way it was imagined by FAIR's original proponents, would necessitate deletion of multiplicity on a scale they would be unlikely to have capacity for. So to sum up, this thesis does not support the attainability of a future semantic web for science, barring radical and unforeseen changes.

In the case of policymakers it is a bit more difficult to reduce their desires into a simple one. One obvious reason is that policymakers are individuals with different values, beliefs and desires. Additionally, policymakers – as representatives of different institutions, organizations and interests – can enact diverging or opposing goals. As the focus of this thesis has been on the European Union and European Commission, I will take them as the basis for this position. But as demonstrated in chapter 2, the interests, goals and desires of the EC are also internally differentiated. I will therefore not speak of the attainment of a specific goal, but rather of the attempt at making FAIR a successful policy in terms of spread, nominal uptake and resulting change in practices.

As such, the EC had notable success when it came to making FAIR an element of other institutions policies, particularly among other science funding institutions. The question is whether FAIR policies have led to actual change of scientific practice. In that regard, the EC seems to have had some success – although more limited than what the original proponents of FAIR had imagined, and perhaps differently than what the policymakers had envisioned. To make the case, most chapters of this thesis display some sort of transformation of science or scientific practice, from the local improvement of data practices through stewardship, to the added political momentum of certain epistemic futures. What is notable though is that these are changes that played along the interests of its subjects. This is most clearly elucidated in the case of making metrics for FAIRness, which highlighted the difficulty of pushing for policy change against the interests of policy subjects.

FAIR has, in this regard, not lived up to the expectation in the European Commission that it could become a tool to increase the accountability and governability of science. Still, FAIR policies might have contributed to developing more productive research practices. In spite of the resistance from the EU's policy subjects against a logic of accounting[194], or against research data sharing for interoperability[195], FAIR policies still contributed to improving data practices – exemplified by FAIR metrics for assistance, or data stewardship for local needs. To what extent this could speak to the EC's desire to increase their return on investment in science is less certain, but it wouldn't be too radical to suggest the situation at least affords such an interpretation.

Making FAIR a successful science policy, thus, necessitates to a lesser extent a stricter enforcing of policy goals, and to a larger extent a reorientation in what policy success means. That is to say, while FAIR has largely failed at establishing a semantic web for science[196], we can imagine it successful in other terms. FAIR has been successful in starting debates on data management, in improving local data practices, and in creating new ways of imagining a better future for science. Taking this perspective, what policymakers should do and should have done is perhaps to not only *allow and accept* a shifting target, but to *encourage and nourish* it in order to utilize the willingness to change and adapt among their subjects.

[194] Seen quite explicitly in the process of making FAIR metrics
[195] As seen in the case of the data stewards' practices
[196] Or the European Open Science Cloud, for that matter

This brings me to the researchers. Reducing researchers to one group with one set of interests and desires is not only reductive, but also unproductive – not only when the actors in this thesis do it, but also if I were to do it as well. Throughout this thesis I've made the point that epistemic practices are not only multiple, but impossible to divide into stable categories. I will therefore refrain from doing so, and instead point out a common denominator between the researchers in this thesis: their stated interests have generally not been initially aligned with the interests of the policymakers. Still, in spite of having critical positions on FAIR, some of the subjects chose to embrace this concept and adopt it in a way that could align with their interests or convictions: be it the data stewards who used it to promote "good research" within their institutions, repository workers using the opportunity to gain funding for their data management projects, or scientific communities using the political momentum to forward their own interests – sometimes in discord with the stated goals of the original formulation of the FAIR principles themselves. For researchers as policy subjects, the suggestion is thus to not treat policy as a threat to epistemic integrity, but as an opportunity for epistemic development and a source of political momentum.

Beyond FAIR

But what can we take away from this, beyond the case of FAIR? I stated in the introduction of this thesis, that it is not merely a story about FAIR, it is also a story about the production of knowledge and the elements and environments that govern that process – it is about the relationship between practices and principles. Principles are – essentially – idealized prescriptions of practice, how someone thinks something should be done (see Mol and Berg, 1994). As such, principles are a source of judgment, valuation and ordering – making claims on what is the good or right way of doing things, and conversely what is wrong and bad. Principles are tools of discipline. But are they good tools?

Annemarie Mol and Marc Berg (1994) explored this question by looking at principles and practices of medicine, and found that – perhaps counterintuitively – while principles perform and idealize a singular ontology, they also enable and legitimize difference and multiplicity. Their argument is that the distinction between principle and practice functions as an admission that principles do not reflect how things are being practiced, thus "setting principles and practice apart facilitates the co-existence of inconsistent logics" (Mol and Berg, 1994, p. 260). As such, principles may be good tools, but not of discipline.

In this thesis, I've found the same to be true for the FAIR principles – they are poor tools of discipline, but productive nonetheless. The particular quality of FAIR – as opposed to the principles of medicine Mol and Berg analyzed – is that the FAIR principles were not "foundational" principles (Mol and Berg, 1994), but aspirational principles. The FAIR principles were not intending to uphold and maintain existing practices, but intending to encourage and nurture new ones.

FAIR is neither the first nor (likely) the last attempt at governing science through aspirational principles, with notable examples being Responsible Research and Innovation (or RRI) (see Macnaghten and Stilgoe, 2012), or Open Science (see Fecher and Friesike, 2013). One of the central characteristics of aspirational principles like these is that they will remain aspirational. Exactly what practices they aspire to is never settled, and exactly when they should manifest is never claimed. Thus, this aspiration can go on indefinitely: we will never know when we have arrived at a FAIR science, nor when we have failed to become FAIR. As such, aspirational principles insulate themselves from critique and collapse, always able to counter that we can't give up now, we'll be there soon, just not yet, in perpetuity.

A defining feature of these aspirational principles, then, is a constant nature of becoming or a not-yetness. This unsettled nature leaves an openness that sustains interpretive flexibility and multiplicity. As such, these aspirational principles are not well suited for achieving the standardization that the infrastructuring of epistemic practices necessitates. While this perpetual openness does not produce the desired outcomes, it does stimulate negotiations among subjects of governance. Thus, the use of aspirational principles would fit better with a more open-ended participatory governance strategy that invites subjects to contribute with their aspirations.

As opposed to governance approaches that build on the ideas of formal or centralized spaces for negotiation – like an agora or parliament – the negotiations with principles happen in a distributed fashion. Instead of taking place at sites of governance, these negotiations are expressed through practices that become related to the principles at various and different sites of articulation. While articulations can take place at sites organized for the purpose of negotiation, like at conferences, they often take place elsewhere: at any sites of practice that necessitate coordination, such as in scientific project work. As such, articulations are the collateral of coordinated practice. Since these articulations seldom take place at sites of explicit negotiation, they are not as prone to manifest as outright controversy. As these sites of articulation are located at sites of practice, the outcome of these negotiations are adapted to local contexts in which they take place – like that of a biosemantics conference, a group for development of indicators, or in the stewardship of everyday scientific practice.

It is only through their relationship to principles that these otherwise distributed and multiple articulations "appear to be a solid whole. Not a heterogeneous assemblage of bits and pieces with frictions between them" (Mol and Berg, 1994). The appearance of harmony between strictly opposing or exclusive practices is thus an inherent product of these principles as such. Infrastructuring through principles thus seems to be self-contradictory, as it simultaneously aims to achieve standardized practices and nurtures multiplicity.

In other words, while concerns for the discriminatory and violent potential of FAIR that characterizes prior scholarship are understandable, warnings against "epistemic injustice" (Bezuidenhout, 2020) or "epistemicide" (Caroll et al., 2020) voiced by critics of FAIR are likely overblown. Because while there's reason to take aspirations for standardization seriously, the multiplicity of epistemic practice is unlikely to succumb under the efforts at turning principles into policy into practice.

At least not yet.

Bibliography

Adam, B. (2008). *OF TIMESCAPES, FUTURESCAPES AND TIMEPRINTS*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=57c02886201ab5de160be7461638b9c39cb66781

Adam, B. (2023). *Futures Imperfect: A Reflection on Challenges - Barbara Adam, 2023*. Sociology. https://journals-sagepub-com.uaccess.univie.ac.at/doi/full/10.1177/00380385221113478

Ahalt, S., Berg-Cross, G., Heinzel, J. B., Broeder, D., Chen, Y., Isaac, A., Kahn, B., Lannom, L., Lautenschlager, M., Moore, R., Michelini, A., Pfeiffenberger, H., Ritz, R., Schwardmann, U., van de Sompel, H., van Uytvanck, D., Vieglas, D., Wittenburg, P., ZHU, Y., & Stehouwer, H. (2014). *RDA Data Foundation and Terminology DFT 1: Overview*. https://zenodo.org/record/2574407/files/DFT1%20-%20overview.pdf?download=1

Ahalt, S., Broeder, D., Heinzel, S., Kahn, B., Lannom, L., Lautenschlager, M., Moore, R., Rajasekar, A., Reetz, J., Schwardmann, U., Stotzka, R., Weigel, T., & Wittenburg, P. (2014). *RDA Initiative Data Foundation and Terminology Basic Concept Note Major Version 7*. https://www.datafairport.org/images/documents/rda_annotated.pdf

AlQuraishi, M., & Sorger, P. K. (2016). Reproducibility will only come with data liberation. *Science Translational Medicine*, *8*(339). https://doi.org/10.1126/scitranslmed.aaf0968

Andersson, J. (2017). Governing Futures. In D. S. King & P. Le Galès (Eds.), *Reconfiguring European States in Crisis*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198793373.003.0015

Antezana, E., Kuiper, M., & Mironov, V. (2009). *Biological knowledge management: the emerging role of the Semantic Web technologies*. *10*(4), 392–407. https://doi.org/10.1093/bib/bbp024

Asdal, K. (2015). What is the issue? The transformative capacity of documents. *Distinktion: Journal of Social Theory*, *16*(1), 74–90. https://doi.org/10.1080/1600910x.2015.1022194

Asdal, K., & Reinertsen, H. (2020). *Hvordan gjøre dokumentanalyse : en praksisorientert metode*. Cappelen Damm Akademisk.

Attwood, T. K., Kell, D. B., McDermott, P., Marsh, J. D., Pettifer, S., & Thorne, D. (2009). Calling International Rescue: knowledge lost in literature and data landslide! *Biochemical Journal*, *424*(3), 317–333. https://doi.org/10.1042/bj20091474

Aula, V. (2019). *Institutions, infrastructures, and data friction – Reforming secondary use of health data in Finland - Ville Aula, 2019*. Big Data & Society. https://journals.sagepub.com/doi/full/10.1177/2053951719875980

Aune, M. (1996). The computer in everyday life: Patterns of domestication of a new technology. In M. Lie & K. Sørensen (Eds.), *Making technology our own? Domesticating technology into everyday life* (pp. 94–124). Scandinavian University Press.

Bahim, C. (2019, September 23). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34

Bahim, C., Casorrán-Amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., Russell, K., & Stall, S. (2020). The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. *Data Science Journal*, *19*. https://doi.org/10.5334/dsj-2020-041

Bairoch, A. M. (2009). *The future of annotation/biocuration*. https://doi.org/10.1038/npre.2009.3092.1

Baker, K. S., & Yarmey, L. (2009). Data Stewardship: Environmental Data Curation and a Web-of-Repositories. *International Journal of Digital Curation*, *4*(2), 12–27. https://doi.org/10.2218/ijdc.v4i2.90

Barad, K. (2006). *Meeting the Universe Halfway*. Duke University Press. https://doi.org/10.2307/j.ctv12101zq

Basajja, M., Suchanek, M., Getu Tadele Taye, Samson Yohannes Amare, Mutwalibi Nambobi, Sakinat Folorunso, Ruduan Plug, Oladipo, F., & Mirjam van Reisen. (2022). Proof of Concept and Horizons on Deployment of FAIR Data Points in the COVID-19 Pandemic. *Data Intelligence*, *4*(4), 917–937. https://doi.org/10.1162/dint_a_00179

Batista, D., Gonzalez-Beltran, A., Sansone, S.-A., & Philippe Rocca-Serra. (2021). Machine actionable metadata models. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.5596465

Bechhofer, S., Ainsworth, J., Bhagat, J., Buchan, I., Couch, P., Cruickshank, D., David De Roure, Delderfield, M., Dunlop, I., Gamble, M., Goble, C., Michaelides, D. T., Paolo Missier, Owen, S., Newman, D., & Sufi, S. (2010). *Why Linked Data is Not Enough for Scientists*. https://doi.org/10.1109/escience.2010.21

Bechhofer, S., Buchan, I., David De Roure, Paolo Missier, Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Michaelides, D. T., Owen, S., Newman, D., Sufi, S., & Goble, C. (2013). *Why linked data is not enough for scientists*. *29*(2), 599–611. https://doi.org/10.1016/j.future.2011.08.004

Beck, M. (2019). On the hourglass model. *Communications of the ACM*, *62*(7), 48–57. https://doi.org/10.1145/3274770

Beck, U. (1992). *Risk society: towards a new modernity*. Sage Publications.

Beer, D. (2018). *The Data Gaze*. SAGE.

Benedictus, R. (2015). *What is Science in Transition?* https://scienceintransition.nl/app/uploads/2015/07/Science-in-Transition-summary-milestones-and-context.pdf

Berg, A., Lukkarinen, J., & Kimmo Ollikka. (2020). “Sticky” Policies—Three Country Cases on Long-Term Commitment and Rooting of RE Policy Goals. *Energies*, *13*(6), 1351–1351. https://doi.org/10.3390/en13061351

Berker, T. (2023). Domestication as user-led infrastructuring. In *The Routledge Handbook of Media and Technology Domestication* (pp. 29–41). Routledge. https://doi.org/10.4324/9781003265931-5

Berners-Lee, T. (1990, May). *Information Management: A Proposal*. W3.org. https://www.w3.org/History/1989/proposal.html

Berners-Lee, T. (2006, November 15). *Linked Data - Design Issues*. Web.archive.org. https://web.archive.org/web/20061115043657/https://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T. (2009, March 13). *The next web*. Www.ted.com. https://www.ted.com/talks/tim_berners_lee_the_next_web?subtitle=en

Berners-Lee, T. (2010). *Linked Data - Design Issues*. Web.archive.org. https://web.archive.org/web/20101224234432/https://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., & Fischetti, M. (1999). *Weaving the Web : the past, present and future of the World Wide Web by its inventor*. Orion Business.

Bessy, C., & Chauvin, P.-M. (2013). The Power of Market Intermediaries: From Information to Valuation Processes. *Valuation Studies*, *1*(1), 83–117. https://doi.org/10.3384/vs.2001-5992.131183

Beyond the PDF. (2013). *10*(2), 91–91. https://doi.org/10.1038/nmeth.2363

Bezuidenhout, L. (2020, September 29). *Being Fair about the Design of FAIR Data Standards | Digital Government: Research and Practice*. Digital Government: Research and Practice. https://dl-acm-org.uaccess.univie.ac.at/doi/10.1145/3399632

Bjerknes, G., & Bratteteig, T. (1995). User participation and democracy: a discussion of Scandinavian research on systems development. *Scandinavian Journal of Information Systems*, *7*(1), 73–98.

Blatecky, A., Bicarregui, J. C., & Morais Pires, C. (Eds.). (2013, December 12). *White Paper: 5 Principles for an Open Data Infrastructure*. G8+O6 DATA INFRASTRUCTURES WG MEETING LONDON, 12 DECEMBER 2013 . https://epubs.stfc.ac.uk/work/12236702

Bloor, D. (1976). *Knowledge and Social Imagery*. University of Chicago Press.

Bloor, D. (1999). Anti-Latour. *Studies in History and Philosophy of Science*, *30*(1), 81–112. https://doi.org/10.1016/s0039-3681(98)00038-7

Bogart, D. (2020). Infrastructure and institutions: Lessons from history. *Regional Science and Urban Economics*, 103626. https://doi.org/10.1016/j.regsciurbeco.2020.103626

Boltanski, L., & ThévenotL. (2006). *On justification : economies of worth*. Princeton University Press.

Borgman, C. (2014, September 23). *Data, Data, Everywhere, Nor Any Drop to Drink*. https://collegerama.tudelft.nl/Mediasite/Play/55eb502f04aa4124ac94b75584c62abc1d?catalog=3984a02f-bf33-4c70-a080-94a04d3e8112

Borgman, C. L. (2008). *Data, disciplines, and scholarly publishing*. *21*(1), 29–38. https://doi.org/10.1087/095315108x254476

Borgman, C. L. (2015). *Big data, little data, no data : scholarship in the networked world*. The Mit Press.

Borgman, C. L. (2017, November 17). *If Data Sharing is the Answer, What is the Question?* Escholarship.org. https://escholarship.org/uc/item/20w9g1k2

Borgman, C. L., Darch, P. T., Sands, A. E., Wallis, J. C., & Traweek, S. (2014). The ups and downs of knowledge infrastructures in science: Implications for data management. *IEEE/ACM Joint Conference on Digital Libraries*. https://doi.org/10.1109/jcdl.2014.6970177

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology Analysis & Strategic Management*, *18*(3-4), 285–298. https://doi.org/10.1080/09537320600777002

Bostrom, J. (2005, May 18). *Berners-Lee seeks killer app for Semantic Web: Internet News from The Industry Standard*. Archive.org. https://web.archive.org/web/20050525054841/http://www.thestandard.com/internetnews/001301.php

Bourdieu, P. (1993). *The Field of cultural production : essays on art and literature*. Columbia University Press.

Bourne, P. (2010). *One Scientist's Wish List for STM Publishers*. https://www.stm-assoc.org/2010_04_27_Spring_Conference_Bourne_Scientist_Wish_List.pdf

Bourne, P. E. (2005). *Will a Biological Database Be Different from a Biological Journal? 1*(3), e34–e34. https://doi.org/10.1371/journal.pcbi.0010034

Bourne, P. E., Bonazzi, V., Dunn, M., Green, E. D., Guyer, M., Komatsoulis, G., Larkin, J., & Russell, B. (2015). The NIH Big Data to Knowledge (BD2K) initiative. *Journal of the American Medical Informatics Association*, *22*(6), 1114–1114. https://doi.org/10.1093/jamia/ocv136

Bourne, P. E., Clark, T., Dale, R., de Waard, A., Herman, I., Hovy, E., & Shotton, D. (2015, May 12). *FORCE11 Manifesto - FORCE11*. Force11.org. https://force11.org/info/force11-manifesto/

Bowker, G. (1994). Information Mythology and Infrastructure. In L. Bud-Frierman (Ed.), *Information Acumen: The Understanding and Use of Knowledge in Modern Business* (pp. 231–247). Routledge.

Bowker, G. C., & Leigh Star, S. (2000). *Sorting things out : classification and its consequences*. Mit Press.

Bowker, G., & Star, S. (1998). *Building Information Infrastructures for Social Worlds -The Role of Classifications and Standards*. https://www.ics.uci.edu/~wscacchi/GameLab/Recommended%20Readings/info-infra-social-worlds.pdf

boyd, danah, & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Brives, C. (2013). *Identifying ontologies in a clinical trial - Charlotte Brives, 2013*. Social Studies of Science. https://journals-sagepub-com.uaccess.univie.ac.at/doi/full/10.1177/0306312712472406

Brooks, M. (2015, October 31). *Exclusive: Grave doubts over LIGO's discovery of gravitational waves*. New Scientist. https://www.newscientist.com/article/mg24032022-600-exclusive-grave-doubts-over-ligos-discovery-of-gravitational-waves/

Bueger, C. (2013). Pathways to practice: praxiography and international politics. *European Political Science Review*, *6*(3), 383–406. https://doi.org/10.1017/s1755773913000167

Bueger, C., & Gadinger, F. (2018). Doing Praxiography: Research Strategies, Methods and Techniques. *Springer EBooks*, 131–161. https://doi.org/10.1007/978-3-319-73350-0_6

Buniello, A., MacArthur, J. A., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., & Flicek, P. (2018). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(D1), D1005–D1012. https://doi.org/10.1093/nar/gky1120

Burgelman, J.-C. (2020, November 22). *201127 | 11 | Plenary 1 | Pre-Recording | International Open Science in 2020 | J.-C. Burgelman*. Vimeo. https://vimeo.com/482288812

Burgelman, J.-C. (2021). Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story). *Data Intelligence*, *3*(1), 5–19. https://doi.org/10.1162/dint_a_00069

Burgelman, J.-C. (2023, September 21). *201127 | 11 | Plenary 1 | Pre-Recording | International Open Science in 2020 | J.-C. Burgelman*. Vimeo. https://vimeo.com/482288812

Burgelman, J.-C., Osimo, D., & Bogdanowicz, M. (2010, July 5). *View of Science 2.0 (change will happen….) | First Monday*. Firstmonday.org. https://firstmonday.org/ojs/index.php/fm/article/view/2961/2573

Busch, L. (2000). The moral economy of grades and standards. *Journal of Rural Studies*, *16*(3), 273–283. https://doi.org/10.1016/s0743-0167(99)00061-3

Callon, M. (1984). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *The Sociological Review*, *32*(1), 196–233. https://doi.org/10.1111/j.1467-954x.1984.tb00113.x

Callon, M. (1998). An Essay on Framing and Overflowing: Economic Externalities Revisited by Sociology. *The Sociological Review*, *46*(1_suppl), 244–269. https://doi.org/10.1111/j.1467-954x.1998.tb03477.x

Callon, M., Lascoumes, P., & Barthe, Y. (2009). *Acting in an uncertain world : an essay on technical democracy*. Mit Press.

Carlile, P. R. (2004). Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries. *Organization Science*, *15*(5), 555–568. https://doi.org/10.1287/orsc.1040.0094

Carse, A., & Kneas, D. (2019). Unbuilt and Unfinished: The Temporalities of Infrastructure. *Environment and Society*, *10*(1), 9–28. https://doi.org/10.3167/ares.2019.100102

CCSDS. (2002). *Reference Model for an Open Archival Information System (OAIS)*. https://public.ccsds.org/Pubs/650x0b1s.pdf

CERN. (n.d.). *The birth of the Web | CERN*. Home.cern. Retrieved August 12, 2024, from https://home.cern/science/computing/birth-web

CLARIN. (n.d.-a). *CLARIN in a Nutshell | CLARIN ERIC*. Clarin.eu. Retrieved January 30, 2024, from https://www.clarin.eu/content/clarin-nutshell

CLARIN. (n.d.-b). *CMDI: Component Metadata Infrastructure | CLARIN ERIC*. Clarin.eu. Retrieved January 30, 2024, from https://www.clarin.eu/content/cmdi-component-metadata-infrastructure

Clark, T., Luciano, J. S., Marshall, M. S., & Prud'hommeaux, E. (2009). Semantic Web Applications in Scientific Discourse 2009. *CUER Workshop Proceedings*, *523*.

Coburn, C. E. (2001). *Collective Sensemaking about Reading: How Teachers Mediate Reading Policy in Their Professional Communities - Cynthia E. Coburn, 2001*. Educational Evaluation and Policy Analysis. https://journals-sagepub-com.uaccess.univie.ac.at/doi/abs/10.3102/01623737023002145

CODATA. (2020). Codata.org. https://conference.codata.org/FAIRconvergence2020/

CODATA. (2021a). *20211018_11_Opening Plenary.mp4*. Vimeo. https://vimeo.com/636516349

CODATA. (2021b, January 11). *201127 | 11 | Plenary 1 | International Open Science in 2020*. Vimeo. https://vimeo.com/499267657

CODATA. (2022, November 10). *FAIR Convergence Symposium 2022, Leiden, Netherlands - CODATA, The Committee on Data for Science and Technology*. CODATA, the Committee on Data for Science and Technology. https://codata.org/events/conferences/fair-convergence-symposium-2022/

CODATA, WFEO, MCIK, UNESCO, JKUAT, ICSU-WDS, GEO, ICA, ISPRS, ISDE, & RDA. (2014, July 27). *Principles for the Sharing of Data from Public Funding in Developing Countries (The Nairobi Data Sharing Principles)*. Jkuat.ac.ke. https://jkuat.ac.ke/wp-content/uploads/2014/08/Nairobi-Data-Guidelines-20142.doc

Coenen, H. (1991). Wandering through the Caves: Phenomenological Field Research in a Social World of Dementia. In D. R. Maines (Ed.), *Social organization and social process: essays in honor of Anselm Strauss* (pp. 315–332). De Gruyter.

Collins, H. (2010). *Tacit and explicit knowledge*. University of Chicago Press.

Collins, H., & Pinch, T. (1993). *The Golem: What Everyone Should Know About Science*. Cambridge University Press.

Colton, E. (2020, March 24). *Rahm Emanuel reprises “never let a crisis go to waste” catchphrase amid coronavirus pandemic - Washington Examiner*. Washington Examiner - Political News and Conservative Analysis about Congress, the President, and the Federal Government. https://www.washingtonexaminer.com/news/2635701/rahm-emanuel-reprises-never-let-a-crisis-go-to-waste-catchphrase-amid-coronavirus-pandemic/

Cook-Deegan, R. (1989, October). *The Alta Summit, December 1984*. Web.ornl.gov. https://web.ornl.gov/sci/techresources/Human_Genome/project/alta.shtml

Cornwall, A. (2007, August). *Buzzwords and Fuzzwords: Deconstructing Development Discourse on JSTOR*. Univie.ac.at. https://www-jstor-org.uaccess.univie.ac.at/stable/25548244

Council of the EU. (2015a). *Council conclusions on open, data-intensive and networked research as a driver for faster and wider innovation*. https://data.consilium.europa.eu/doc/document/ST-9360-2015-INIT/en/pdf

Council of the EU. (2015b). *OUTCOME OF THE COUNCIL MEETING: Competitiveness (Internal Market, Industry, Research and Space)*. https://www.consilium.europa.eu/media/23444/st09385en15.pdf

Council of the EU. (2016). *The transition towards an Open Science system - Council conclusions*. https://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf

Crawford, K. (2021). *Atlas of AI*. Yale University Press.

CWA. (2009, March 16). *Groups**. Wordpress.com; WordPress.com. https://conceptweblog.wordpress.com/groups/

DAITF. (2012a, February 14). *Pre ICRI DAITF Preparation Workshop, Copenhagen, 20/21 March | Research Data Alliance Support Forum*. Archive.org. https://web.archive.org/web/20130112235645/http://www.daitf.org:80/?p=48#1

DAITF. (2012b, June 5). *DAITF preparation forum | Data Access and Interoperability Taskforce*. Archive.org. https://web.archive.org/web/20120831101138/http://www.daitf.org/

DANS, DCC, Directorate-General for Research and Innovation (European Commission), EFIS, & Visionary Analytics. (2022). European Research Data Landscape: final report. In *Publications Office of the European Union*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/03b5562d-6a35-11ed-b14f-01aa75ed71a1/language-en

Data FAIRport. (2014a, April 4). *Work Packages Update 1*. Www.datafairport.org. https://www.datafairport.org/17/index.html

Data FAIRport. (2014b, June 1). *Data Fairport*. Www.datafairport.org. https://www.datafairport.org/component/content/article/8_news/18_work_packages_update_01_06_2014/index.html

Data FAIRport. (2014c, September 7). *WORK PACKAGES UPDATE 07/09/2014*. Www.datafairport.org. https://www.datafairport.org/latest/8_news/19_work_packages_update_07_09_2014/index.html

David, R., Laurent Dollé, Specht, A., Stryeck, S., yahia, M., Thomsen, M., Cambon-Thomsen, A., & Mabile, L. (2020). Implementing FAIR in data sharing: who are the actors and what are their responsibilities? *HAL (Le Centre Pour La Communication Scientifique Directe)*. https://doi.org/10.5281/zenodo.3743946

David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Dollé, L., Cambon-Thomsen, A., & SHARC. (2020, March 30). *Implementing FAIR in data sharing: who are the actors and what are their responsibilities?* Zenodo. https://zenodo.org/record/3743946

Davies, S. (2019). Epistemic Living Spaces, International Mobility, and Local Variation in Scientific Practice. *Minerva*, *58*(1), 97–114. https://doi.org/10.1007/s11024-019-09387-0

Davis, R. (2018). A short story on length. *Nature Physics*, *14*(8), 868–868. https://doi.org/10.1038/s41567-018-0243-4

de Laet, M., & Mol, A. (2000). The Zimbabwe Bush Pump. *Social Studies of Science*, *30*(2), 225–263. https://doi.org/10.1177/030631200030002002

De Roure, D. (2009a, July 28). *My Experiment*. Videolectures.net. http://videolectures.net/eswc09_roure_myexp/

De Roure, D. (2009b, August 1). *Research Objects - myExperiment*. Southampton.ac.uk. https://web-archive.southampton.ac.uk/wiki.myexperiment.org/wiki/Research_Objects.html

De Roure, D. (2009c, August 10). *Replacing the Paper: the six Rs of the e-Research Record*. Archive.org. https://web.archive.org/web/20150813214446/http://blog.openwetware.org/eresearch/?p=56

De Roure, D. (2010, November 27). *Replacing the Paper: The Twelve Rs of the e-Research Record - eResearch Blog | Nature Publishing Group*. Archive.org. https://web.archive.org/web/20120630003824/http://blogs.nature.com/eresearch/2010/11/27/replacing-the-paper-the-twelve-rs-of-the-e-research-record

De Roure, D. (2012, June 10). *More Rs than Pirates*. EResearch. https://web.archive.org/web/20130530035909/http:/www.scilogs.com/eresearch/more-rs-than-pirates/

Dekker, S. (2013). BRIEF VAN DE STAATSSECRETARIS VAN ONDERWIJS, CULTUUR EN WETENSCHAP. *Officielebekendmakingen.nl*. https://doi.org/kst-31288-354

Dekker, S. (2014, January 21). *Open Access to publications*. Archive.org. https://web.archive.org/web/20150906232540/https:/www.government.nl/documents/parliamentary-documents/2014/01/21/open-access-to-publications

Dekkers, M. (2019a, August 12). *Indicators prioritisation for Interoperability · Issue #32 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/32#issuecomment-520338836

Dekkers, M. (2019b, August 23). *Indicators prioritisation for Findability · Issue #30 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/30#issuecomment-524201257

Dekkers, M. (2019c, October 18). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-543584047

Dekkers, M. (2020a, February 13). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-585747982

Dekkers, M. (2020b, February 20). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub.

https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-589107446

Dekkers, M. (2020c, February 20). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-589107446

DeLisi, C. (2008). Santa Fe 1986: Human genome baby-steps. *Nature*, *455*(7215), 876–877. https://doi.org/10.1038/455876a

Devaraju, A., & Huber, R. (2021). An automated solution for measuring the progress toward FAIR research data. *Patterns*, *2*(11), 100370. https://doi.org/10.1016/j.patter.2021.100370

Devaraju, A., Huber, R., Mokrane, M., Herterich, P., Cepinskas, L., de Vries, J., L'Hours, H., Davidson, J., & White, A. (2020, October 12). *FAIRsFAIR Data Object Assessment Metrics*. Zenodo. https://zenodo.org/record/6461229

Digital Science, Hahnel, M., McIntosh, L. D., Hyndman, A., Baynes, G., Crosas, M., Nosek, B., Shearer, K., van Selm, M., Goodey, G., & Nature Research. (2020, December 1). *The State of Open Data 2020*. Digitalscience.figshare.com. https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_2020/13227875#:~:text=The%20report%20outlines%20the%20impact

Dijstelbloem, H., Huisman, F., Miedema, F., & Mijnhardt, W. (2013). Why Science Does Not Work as It Should And What To Do about It. In *science in transition*. https://www.scienceintransition.nl/app/uploads/2013/10/Science-in-Transition-Position-Paper-final.pdf

Dijstelbloem, H., Huisman, F., Miedema, F., & Mijnhardt, W. (2014). *Debate, progress and recommendations*. https://scienceintransition.nl/app/uploads/2014/07/Science-in-Transition-Status-Report-June-2014.pdf

Directorate-General for Research and Innovation, EOSC Executive Board, Aronsen, J. M., Beyan, O., Harrower, N., Holl, A., Hooft, R. W. W., Principe, P., Slavec, A., Jones, S., & Genova, F. (2021). Recommendations on FAIR metrics for EOSC. In *Publications Office of the European Union*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/ced147c9-53c0-11eb-b59f-01aa75ed71a1

DORA. (n.d.). *Read the declaration – DORA*. Sfdora.org. https://sfdora.org/read/

Doudna, J. (2020, June 5). *Jennifer Doudna on how covid-19 is spurring science to accelerate*. The Economist; The Economist. https://www.economist.com/by-invitation/2020/06/05/jennifer-doudna-on-how-covid-19-is-spurring-science-to-accelerate

DSA. (n.d.). *About | Data Seal of Approval*. Web.archive.org. Retrieved July 31, 2023, from https://web.archive.org/web/20131014195745/http://datasealofapproval.org/en/information/about/

Duarte, F. (2023, March 16). *Amount of Data Created Daily (2024)*. Exploding Topics; Exploding Topics. https://explodingtopics.com/blog/data-generated-per-day

Duerr, R. (2020, February 13). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-585943564

Duncan, R., Robson-Williams, M., & Edwards, S. (2020). A close examination of the role and needed expertise of brokers in bridging and building science policy boundaries in environmental decision making. *Palgrave Communications*, *6*(1). https://doi.org/10.1057/s41599-020-0448-x

Dussauge, I., Helgesson, C.-F., & Lee, F. (2015). *Value practices in the life sciences and medicine*. Oxford University Press.

EC. (2007a). *Communication on scientific information in the digital age: access, dissemination and preservation*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52007DC0056

EC. (2007b, April 4). *The European Research Area: New perspectives - Green Paper*. Web.archive.org. https://web.archive.org/web/20091121054533/http:/ec.europa.eu/research/era/pdf/era-greenpaper_en.pdf

EC. (2008a). *European Research Area - Specific ERA Initiatives*. Archive.org. https://web.archive.org/web/20090207180615/http:/ec.europa.eu/research/era/2020_era_vision_en.html

EC. (2008b). *Open access pilot in FP7*. https://op.europa.eu/en/publication-detail/-/publication/6719e951-d9b1-449f-a1b7-5271c8502b4d/language-en

EC. (2009). *Preparing Europe for a New Renaissance A Strategic View of the European Research Area*. https://publications.europa.eu/resource/cellar/89724e55-9579-4bd8-80cf-9a21ba645cbc.0001.02/DOC_1

EC. (2010a, May 19). *EUR-Lex - 52010DC0245 - EN - EUR-Lex*. Europa.eu. https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A52010DC0245#footnoteref35

EC. (2010b). *EUROPE 2020 A strategy for smart, sustainable and inclusive growth*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52010DC2020

EC. (2010c, October 6). *Press corner*. European Commission - European Commission. https://ec.europa.eu/commission/presscorner/detail/en/IP_10_1288

EC. (2010d). Riding the Wave: How Europe can gain from the rising tide of scientific data. In *https://ec.europa.eu/eurostat/cros/system/files/riding%20the%20wave.pdf*.

EC. (2014a). *Backgound document: Public consultation on "Science 2.0": Science in transition*. Web.archive.org. https://web.archive.org/web/20151013215726/https://scienceintransition.files.wordpress.com/2014/10/rtd_-public-consultation-science-2-0-final.pdf

EC. (2014b). *Consultation on "Science 2.0": Science in Transition - European Commission*. Archive.org. https://web.archive.org/web/20140707054209/http://ec.europa.eu/research/consultations/science-2.0/consultation_en.htm

EC. (2014c). *"Science 2.0": Science in transition*. "Science 2.0": Science in Transition; "Science 2.0": Science in transition. https://web.archive.org/web/20150227152029/http:/scienceintransition.eu:80/

EC. (2015a, February). *Validation of the results of the public consultation on Science 2.0: Science in Transition* . Web.archive.org. https://web.archive.org/web/20150501003941/http://ec.europa.eu/research/consultations/science-2.0/science_2_0_final_report.pdf#view=fit&pagemode=none

EC. (2015b, May 6). *A Digital Single Market Strategy for Europe*. Europa.eu. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52015DC0192

EC. (2016a). *European Open Science Cloud | Open Science - Research & Innovation - European Commission*. Archive.org. https://web.archive.org/web/20161223163407/http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud-hleg

EC. (2016b). Realising the European open science cloud: first report and recommendations of the Commission high level expert group on the European open science cloud. In *Publications Office of the European Union*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/2ec2eced-9ac5-11e6-868c-01aa75ed71a1

EC. (2016c). *Register of Commission expert groups and other similar entities*. Archive.org. https://web.archive.org/web/20170318003313/http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3353

EC. (2016d, May 17). *Open innovation, open science, open to the world - Publications Office of the EU*. Publications Office of the EU. https://op.europa.eu/en/publication-detail/-/publication/3213b335-1cbc-11e6-ba9a-01aa75ed71a1

EC. (2016e, October 10). *Realising the European open science cloud : first report and recommendations of the Commission high level expert group on the European open science cloud.* Europa.eu. https://op.europa.eu/en/publication-detail/-/publication/2ec2eced-9ac5-11e6-868c-01aa75ed71a1

EC. (2017, October 26). *EOSC Declaration*. Web.archive.org. https://web.archive.org/web/20171026151855/https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf#view=fit&pagemode=none

EC. (2018). *Turning FAIR into reality*. https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_0.pdf

EC. (2020a). *Horizon 2020 Work Programme 2018-2020*. https://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/main/h2020-wp1820-infrastructures_en.pdf

EC. (2020b). Six Recommendations for implementation of FAIR practice by the FAIR in practice task force of the European open science cloud FAIR working group. In *Publications Office of the European Union*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/4630fa57-1348-11eb-9a54-01aa75ed71a1

EC. (2021). *Recommendations on FAIR metrics for EOSC*. https://data.europa.eu/doi/10.2777/70791

EC. (2022a). *Horizon Europe (HORIZON) Programme Guide*. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

EC. (2022b). Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC). In *Publications Office of the European Union*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/f9b12d1d-74ea-11ec-9136-01aa75ed71a1

EC, & PwC. (2018). *Cost-benefit analysis for FAIR research data - Publications Office of the EU*. Publications Office of the EU. https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en

Edwards, P. N. (2010). *A vast machine : computer models, climate data, and the politics of global warming*. Mit Press.

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, *41*(5), 667–690. https://doi.org/10.1177/0306312711413314

Egseth, K. A. (2018). Bonden og melkerobote. Brukeres domestisering av en ny teknologi. In *RURALIS*. https://ruralis.no/wp-content/uploads/2019/08/r12_18-bonden-og-melkeroboten--brukeres-domestisering-av-en-ny-teknologi-kine-ariela-egseth-med-forsida.pdf

ELIXIR. (2020, July 31). *Data FAIRification with FAIRplus*. ELIXIR. https://elixir-europe.org/news/data-fairification-fairplus

Elzen, B., Enserink, B., & Smit, W. A. (1996). Socio-Technical Networks: How a Technology Studies Approach May Help to Solve Problems Related to Technical Change. *Social Studies of Science*, *26*(1), 95–141. https://doi.org/10.1177/030631296026001006

EOSC Nordic. (2022a, February 8). *FAIRification Step 5 webinar value and limitations of fair assessment tools*. Www.youtube.com. https://www.youtube.com/watch?v=CqUikcVX4Ig

EOSC Nordic. (2022b, May 9). *FAIR Incentives stakeholder workshop 9.5.2022*. Www.youtube.com. https://www.youtube.com/watch?v=nsFdYW8JLBA

EOSC Nordic. (2022c, August 29). *FAIR is possible! Increasing FAIR maturity of the Nordic and Baltic research data repositories*. EOSC-Nordic. https://eosc-nordic.eu/fair-is-possible-increasing-fair-maturity-of-the-nordic-and-baltic-research-data-repositories/

EOSC Portal. (n.d.). Eosc-Portal.eu. Retrieved September 21, 2023, from https://eosc-portal.eu/

EOSC Vienna Declaration. (2018, November 23). Eosc-Launch.eu. https://eosc-launch.eu/declaration/

Epstein, S. (1995). The Construction of Lay Expertise: AIDS Activism and the Forging of Credibility in the Reform of Clinical Trials. *Science, Technology, & Human Values*, *20*(4), 408–437. https://doi.org/10.1177/016224399502000402

Epstein, S. (2007). *Inclusion*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226213118.001.0001

EROSS. (2020, October 4). *Barend Mons: Q&A Session*. Www.youtube.com. https://www.youtube.com/watch?v=sqwWL2BC3JY

ESFRI. (2019, December). *WORKING GROUP REPORT Monitoring of Research Infrastructures Performance*. https://www.esfri.eu/sites/default/files/ESFRI_WG_Monitoring_Report.pdf

European Council. (2000). *PRESIDENCY CONCLUSIONS- LISBON EUROPEAN COUNCIL - 23 AND 24 MARCH 2000*. Www.consilium.europa.eu. https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/00100-r1.en0.htm

FAIR EG. (2017). *Minutes of the expert groups: Meeting of the FAIR Data Expert Group 27 March 2017*.

FAIR EG. (2019). *FAIR-Data-EG-INTERIM-Report-CONSULTATION*. Google Docs. https://docs.google.com/document/d/1FcnBOPUm-9gW3gjF18FDoPNUvIn-ojduLeLe0Y8yEH8/edit

Fair Maturity Indicators List. (n.d.). Fairdata.services. Retrieved October 3, 2022, from https://fairdata.services:7171/FAIR_Evaluator/metrics

FAIR-DI. (2020). Barend Mons: Discussion [YouTube Video]. In *YouTube*. https://www.youtube.com/watch?v=Cg5nUS2DreY

FAIRdat. (n.d.). *SurveyMonkey Powered Online Survey*. Www.surveymonkey.com. Retrieved October 31, 2022, from https://www.surveymonkey.com/r/fairdat

FAIRDO. (n.d.). *FAIR Digital Objects Forum |*. Fairdo.org. Retrieved January 22, 2024, from https://fairdo.org/

FAIRmetrics. (n.d.). *The FAIR Maturity Evaluation Service*. Fairsharing.github.io. Retrieved September 29, 2022, from https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#

FAIRplus. (n.d.). *FAIRplus FAIRification Results*. Fairplus.github.io. Retrieved November 13, 2022, from https://fairplus.github.io/fairification-results/

FAIRport. (2014). *FAIRPORT*. Docplayer.net. http://docplayer.net/149054256-Jointly-designing-a-data-fairport.html

FAIRsFAIR. (2019, February 13). *The Project*. FAIRsFAIR. https://www.fairsfair.eu/the-project

FAIRsharing. (n.d.). *FAIRsharing identifier schema register*. Fairsharing.org. Retrieved October 3, 2022, from https://fairsharing.org/search?fairsharingRegistry=Standard&recordType=identifier_schema&page=1

Fanelli, D. (2018). Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2628–2631. https://doi.org/10.1073/pnas.1708272114

Fecher, B., & Friesike, S. (2013). Open Science: One Term, Five Schools of Thought. *Opening Science*, 17–47. https://doi.org/10.1007/978-3-319-00026-8_2

Felt, U. (2013, January 25). Kollaterale Zukünfte: Zu den (An)Ordnungen von Morgen. *Zukunftsexpertise. Zur Generierung, Legitimierung, Verwendung Und Anerkennung von Zukunftswissen*.

Felt, U. (2015). Keeping Technologies Out: Sociotechnical Imaginaries and the Formation of Austria's Technopolitical Identity. In S. Jasanoff & S.-H. Kim (Eds.), *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power* (pp. 103–125). The University of Chicago Press, Chicago.

Felt, U. (2016). Of Timescapes and Knowledgescapes. In P. Scott, J. Gallacher, & G. Parry (Eds.), *New Languages and Landscapes of Higher Education* (pp. 129–148). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198787082.003.0006

Felt, U., & Fochler, M. (2011). Re-ordering Epistemic Living Spaces: On the Tacit Governance Effects of the Public Communication of Science. *Sociology of the Sciences/Sociology of the Sciences Yearbook*, 133–154. https://doi.org/10.1007/978-94-007-2085-5_7

FORCE11. (n.d.). *Info – FORCE11*. Force11.org. Retrieved May 14, 2023, from https://force11.org/info/

FORCE11. (2013, October 30). *Joint Declaration of Data Citation Principles – FINAL – FORCE11*. Force11.org. https://force11.org/group/joint-declaration-of-data-citation-principles-final/

FORCE11. (2014a, September 3). *The FAIR data Guiding Principles - FOR COMMENT*. FORCE11. https://web.archive.org/web/20150815084745/https://www.force11.org/group/fairgroup/fairprinciples

FORCE11. (2014b, September 10). *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0*. FORCE11. https://web.archive.org/web/20151013011258/https:/www.force11.org/node/6062

Foucault, M. (1990). *The History of sexuality. Volume I: an Introduction*. Vintage Books.

Frankl, V. (1946). *Man's Search for Meaning*. Pocket.

Fricker, M. (2007). *Epistemic Injustice*. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Friedman, M., & Friedman, R. (1980). *Free to choose : a personal statement*. Penguin Books.

FWF. (2019). *Research Data Management*. Austrian Science Fund (FWF). https://www.fwf.ac.at/en/about-us/what-we-do/open-science/research-data-management

Gaiarin, S. P. (2020, December 17). *FAIR Assessement and Certification in the EOSC region*. Zenodo. https://zenodo.org/record/4486280

Galison, P. (1999). Trading Zone: Coordinating Action and Belief. In *Science Studies Reader* (pp. 137–160). Routledge.

Gardner, J., Samuel, G., & Williams, C. (2015). Sociology of Low Expectations. *Science, Technology, & Human Values*, *40*(6), 998–1021. https://doi.org/10.1177/0162243915585579

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy*, *31*(8-9), 1257–1274. https://doi.org/10.1016/s0048-7333(02)00062-8

Gene Ontology Consortium. (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Genome Sequencing Workshop. (1986). Https://Www.genome.gov/. https://www.genome.gov/sites/default/files/media/files/2021-02/1986_Santa_Fe.pdf

Georgia Institute of Technology. (2007, September 21). *Official Kilogram Losing Mass: Scientists Propose Redefining It As A Precise Number Of Carbon Atoms*. ScienceDaily. https://www.sciencedaily.com/releases/2007/09/070921110735.htm

Germany, & the Netherlands. (2017). Joint Position Paper on the European Open Science Cloud. In *https://www.dtls.nl/wp-content/uploads/2017/05/DE-NL-Joint-Paper-FINAL.pdf*.

Gewin, V. (2020). Six tips for data sharing in the age of the coronavirus. *Nature*. https://doi.org/10.1038/d41586-020-01516-0

GIDA. (n.d.). *CARE principles*. Global Indigenous Data Alliance. Retrieved April 25, 2024, from https://www.gida-global.org/care

Giddens, A. (1991). *Modernity and Self-Identity : Self and Society in the Late Modern Age.* Wiley.

Gieryn, T. F. (1983). Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review*, *48*(6), 781–795. https://doi.org/10.2307/2095325

Gitelman, L. (Ed.). (2013). *"Raw Data" Is an Oxymoron*. The MIT Press. https://doi.org/10.7551/mitpress/9302.001.0001

Glasner, P. (2001). Rights or rituals? Why juries can do more harm than good. *PLA Notes*, *40*. https://www.iied.org/sites/default/files/pdfs/migrate/G01290.pdf

Godin, B. (2009). *The making of science, technology and innovation policy: conceptual frameworks as narratives, 1945-2005*. Centre Urbanisation Culture Société Institut national de la recherche scientifique. https://www.csiic.ca/PDF/TheMakingOfScience.pdf

GOFAIR. (2017, November 17). *European Open Science Cloud Declaration online - GO FAIR*. GO FAIR. https://www.go-fair.org/2017/11/17/european-open-science-cloud-declaration-online/

GOFAIR. (2019, March 15). *GO-FAIR IN-Africa*. https://www.go-fair.org/wp-content/uploads/2019/08/Activity-Plan-with-the-Manifesto-of-the-GO-FAIR-in-AFRICA-Final-1-August-2019.pdf

GOFAIR. (2020a). *International FAIR Convergence Symposium*. GO FAIR. https://www.go-fair.org/events/international-fair-convergence-symposium/

GOFAIR. (2020b, July 2). *The Internet of FAIR Data & Services - GO FAIR*. GO FAIR. https://www.go-fair.org/resources/internet-fair-data-services/

GOFAIR. (2021a, June 16). *FAIR Digital Object Framework - GO FAIR*. GO FAIR. https://www.go-fair.org/today/fair-digital-framework/

GOFAIR. (2021b, July 1). *FAIR Festival 2021: Opening Plenary "FAIR minds open science."* Www.youtube.com. https://www.youtube.com/watch?v=aV1rIC1DDIk&t=1874s

GOFAIR. (2022). *2nd FAIR Convergence Symposium*. GO FAIR. https://www.go-fair.org/events/2nd-fair-convergence-symposium/

Graeber, D. (2013). *On the Phenomenon of Bullshit Jobs*. Atlasofplaces.com. https://www.atlasofplaces.com/essays/on-the-phenomenon-of-bullshit-jobs/

Green, P. (1997). *Comments on JASON report*. Archive.org. https://web.archive.org/web/20090906123638/http://bozeman.genome.washington.edu:80/ScienceLetter/JasonComments.html

Green, P., & Koonin, S. E. (1998). Human Genome Project:Data Quality. *Science*, *279*(5354), 1113–1117. https://doi.org/10.1126/science.279.5354.1113f

Greenwood, M. (2023). *Real Utopia as a Method? Utopian-Sociological Paths from Jameson's Universal Army to a Postcapitalist Post Office - Martin Greenwood, 2023*. Sociology. https://journals-sagepub-com.uaccess.univie.ac.at/doi/full/10.1177/00380385221133205

Groth, P. (2013). *Call for Organizers: 2014 FORCE Conference (formally Beyond the PDF)*. Google.com; Google Groups. https://groups.google.com/g/beyond-the-pdf/c/p-j-Dev-3Q4

Guenter, M. (1998, November 30). *Re: hourglass model*. Ietf.org. https://mailarchive.ietf.org/arch/msg/ietf/GN2d_mO5hXVbqksh-YmDbMqMj5I/

Guiliano, J., & Estill, L. (2023). What gets categorized counts: controlled vocabularies, digital affordances, and the international digital humanities conference. *Digital Scholarship in the Humanities*, *38*(3). https://doi.org/10.1093/llc/fqac091

Guns, R. (2013). Tracing the origins of the semantic web. *Journal of the American Society for Information Science and Technology*, *64*(10), 2173–2181. https://doi.org/10.1002/asi.22907

Hanseth, O., Monteiro, E., & Hatling, M. (1996). Developing Information Infrastructure: The Tension between Standardization and Flexibility. *Science, Technology, & Human Values*, *21*(4), 407–426. https://www.jstor.org/stable/690088

Hargreaves, T., Nye, M., & Burgess, J. (2010). Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors. *Energy Policy*, *38*(10), 6111–6119. https://doi.org/10.1016/j.enpol.2010.05.068

Hartmann, M. (2020). (The domestication of) Nordic domestication? *Nordic Journal of Media Studies*, *2*(1), 47–57. https://doi.org/10.2478/njms-2020-0005

Hartmann, M. (2023). The Routledge Handbook of Media and Technology Domestication. In *Routledge eBooks*. Informa. https://doi.org/10.4324/9781003265931

Haseltine, W. A. (2021, May 25). *How COVID Changed Science*. Scientific American; Scientific American. https://www.scientificamerican.com/article/how-covid-changed-science/

Healey, P. (2011). The universal and the contingent: Some reflections on the transnational flow of planning ideas and practices. *Planning Theory*, *11*(2), 188–207. https://doi.org/10.1177/1473095211419333

Helgesson, C.-F., & Muniesa, F. (2013). For What It's Worth: An Introduction to Valuation Studies. *Valuation Studies*, *1*(1), 1–10. https://doi.org/10.3384/vs.2001-5992.13111

Herczog, E., Peristeras, V., & Russel, K. (2018, September 23). *FAIR Data Maturity Model WG Case Statement*. RDA. https://www.rd-alliance.org/group/fair-data-maturity-model-wg/case-statement/fair-data-maturity-model-wg-case-statement

Herczog, E., & Russell, K. (2019). *Workshop report FAIR data maturity model Working Group Objectives Useful links RDA FAIR data maturity model WG RDA FAIR data maturity model Case Statement Workshop #3 material (Presentation, chat logs and report) RDA FAIR data maturity model GitHub RDA FAIR data maturity model Collaborative document*. https://www.rd-alliance.org/system/files/documents/20190618_FAIR_WG_Workshop_3_report_0.pdf

Higman, R., Bangert, D., & Jones, S. (2019). Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights the UKSG Journal*, *32*. https://doi.org/10.1629/uksg.468

Hoefler, A. M., Scholz, W., & Tochtermann, K. (2015, September 16). *Science 2.0 and Open Science: Similar but Still Different | ZBW MediaTalk*. ZBW MediaTalk. https://www.zbw-mediatalk.eu/2015/09/science-2-0-and-open-science-similar-but-still-different/

Hoff, J.-L. (2023). Unavoidable futures? How governments articulate sociotechnical imaginaries of AI and healthcare services. *Futures*, *148*, 103131–103131. https://doi.org/10.1016/j.futures.2023.103131

Hooft, R. (2019a, October 3). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-537798528

Hooft, R. (2019b, October 18). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-543761652

Hopgood, B. (2001). *History of the Web*. W3.org. https://www.w3.org/2012/08/history-of-the-web/origins.htm#c7p1

Horwood, L., Sullivan, S., Young, E., & Garner, J. (2004). OAI compliant institutional repositories and the role of library staff. *Library Management*, *25*(4/5), 170–176. https://doi.org/10.1108/01435120410533756

Howard, K. (2000). The Bioinformatics Gold Rush. *Scientific American*, *283*(1), 58–63. https://www.jstor.org/stable/pdf/26058789.pdf?refreqid=excelsior%3A657d5598c680e2fce8c4c437d9fba3bb&ab_segments=&origin=&initiator=&acceptTC=1

Hughes, T. P. (1983). *Networks of power : electrification in Western society, 1880-1930*. The Johns Hopkins University Press, Copyr.

Humphreys, M., & Watson, T. (2009). Ethnographic Practices: From “Writing-up Ethnographic Research” to “Writing Ethnography.” In S. Ybema, D. Yanow, & H. Wels (Eds.), *Organizational Ethnography. Studying the Complexities of Everyday Life* (pp. 40–55). Sage.

Hunter, J. (2008). Scientific Publication Packages – A Selective Approach to the Communication and Archival of Scientific Output. *International Journal of Digital Curation*, *1*(1), 33–52. https://doi.org/10.2218/ijdc.v1i1.4

IBM. (2012, March 19). *90% of the data in the world today has been created in the last two years.* Twitter. https://twitter.com/IBM/status/181711401335795712

Ingeborgrud, L., & Suboticki, I. (2023). *Planners as middle actors in facilitating for city cycling*. Mobilities. https://www-tandfonline-com.uaccess.univie.ac.at/doi/full/10.1080/17450101.2023.2186799

Inman, S., & Ribes, D. (2018). Data Streams, Data Seams: Toward a seamful representationof data interoperability. *Proceedings of DRS*. https://doi.org/10.21606/drs.2018.359

International Innovation. (2014). Data Fairport: enabling global exchange of research data. *International Innovation*, 98–101.

Irwin, A., & Wynne, B. (1996). *Misunderstanding science? The public reconstruction of science and technology*. Cambridge Cambridge University Press.

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., & Kaliyaperumal, R. (2020). FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence*, *2*(1-2), 10–29. https://doi.org/10.1162/dint_r_00024

Jagodnik, K. M., Koplev, S., Jenkins, S. L., Ohno-Machado, L., Paten, B., Schurer, S. C., Dumontier, M., Verborgh, R., Bui, A., Ping, P., McKenna, N. J., Madduri, R., Pillai, A., & Ma'ayan, A. (2017). Developing a framework for digital objects in the Big Data to Knowledge (BD2K) commons: Report from the Commons Framework Pilots workshop. *Journal of Biomedical Informatics*, *71*, 49–57. https://doi.org/10.1016/j.jbi.2017.05.006

Jasanoff, S. (1998). *The Fifth Branch*. Harvard University Press.

Jasanoff, S. (2019). Controversy Studies. *The Blackwell Encyclopedia of Sociology*, 1–5. https://doi.org/10.1002/9781405165518.wbeosc130.pub2

Jasanoff, S., & Kim, S.-H. (2015). *Dreamscapes of modernity : sociotechnical imaginaries and the fabrication of power*. The University Of Chicago Press.

Jaunsen, A. O., Alaterä, Tuomas J, Buckland, P. I., Holmstrand, Katrine F, Lahti, K., Essen, P., & Pfeil, B. (2021). *Nordic FAIR Data Collaboration Opportunities*. DIVA. https://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A1622519&dswid=4157

Jaunsen, A. O., Kleemola, M., Alaterä, T. J., Lehvaslaiho, H., Hasan, A., Nordling, J., & Assinen, P. (2020). D4.1 An assessment of FAIR-uptake among regional digital repositories. *Zenodo*. https://zenodo.org/record/4045402#.YwxAhHZBw2w

Jeffery, K. (2019a, August 21). *Indicators prioritisation for Findability · Issue #30 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/30#issuecomment-523423755

Jeffery, K. (2019b, October 15). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-542086350

Jeffery, K., Wittenburg, P., Lannom, L., Strawn, G., Biniossek, C., Betz, D., & Blanchi, C. (2021). Not Ready for Convergence in Data Infrastructures. *Data Intelligence*, *3*(1), 116–135. https://doi.org/10.1162/dint_a_00084

Jensen, C. B. (2010). *Ontologies for Developing Things*. BRILL.

Jensen, C. B., & Morita, A. (2015). Infrastructures as Ontological Experiments. *Engaging Science, Technology, and Society*, *1*, 81. https://doi.org/10.17351/ests2015.21

Johnsen, Å. (2005). What Does 25 Years of Experience Tell Us About the State of Performance Measurement in Public Policy and Management? *Public Money & Management*, *25*(1), 9–17.

Jones, K., Alexander, S., Bennett, N., Bishop, L., Budden, A., Cox, M., Crosas, M., Game, E., Geary, J., Hahn, C., Hardy, D., Johnson, J., Karcher, S., LaFevor, M., Motzer, N., Pinto da Silva, P., Pittman, J., Randell, H., Silva, J., & Smith, J. (2018). *Qualitative data sharing and re-use for socio-environmental systems research: A synthesis of opportunities, challenges, resources and approaches*. https://doi.org/10.13016/M2WH2DG59

Jones, M. P. (1996). Posthuman Agency: Between Theoretical Traditions. *Sociological Theory*, *14*(3), 290. https://doi.org/10.2307/3045390

Juros, T. V. (2022). Challenges of qualitative data sharing in social sciences. *European Science Editing*, *48*. https://doi.org/10.3897/ese.2022.e77781

Karasti, H., & Blomberg, J. (2017). Studying Infrastructuring Ethnographically. *Computer Supported Cooperative Work (CSCW)/Computer Supported Cooperative Work*, *27*(2), 233–265. https://doi.org/10.1007/s10606-017-9296-7

Karasti, H., Millerand, F., Hine, C. M., & Bowker, G. C. (2016). Knowledge infrastructures: Part I. *Science & Technology Studies*, *29*(1), 2–12. https://doi.org/10.23987/sts.55406

Khalsa, S. J. S. (2019, October 16). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-542610424

Khodiyar, V., Laine, H., O'Brien, D., Rodriguez-Esteban, R., TÃ¼rkyilmaz-van der Velden, Y., Baynes, G., Brack, M., Cambon-Thomsen, A., Carr, D., Carrus, E., Chisholm, L., Civai, C., Cruz, M., Grant, R., Newbold, E., & Nunes, F. L. S. (2021). Research Data: The Future of FAIR White paper. *Figshare.com*. https://doi.org/10.6084/m9.figshare.14393552.v1

Kitchin, R. (2014). *The Data Revolution*. SAGE.

Kitchin, R., & Lauriault, T. (2014, July 30). *Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work*. Ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112

Knorr-Cetina, K. (1981). *The Manufacture of Knowledge*. Pergamon.

Knorr-Cetina, K. (1982). Scientific Communities or Transepistemic Arenas of Research? A Critique of Quasi-Economic Models of Science. *Social Studies of Science*, *12*(1), 101–130. https://doi.org/10.1177/030631282012001005

Knorr-Cetina, K. (1999). *Epistemic cultures: how the sciences make knowledge*. Harvard University Press, , Cop.

Kornberger, M., Pflueger, D., & Mouritsen, J. (2017). Evaluative infrastructures: Accounting for platform organization. *Accounting, Organizations and Society*, *60*, 79–95. https://doi.org/10.1016/j.aos.2017.05.002

L'Hours, H., Devaraju, A., Stein, I. von, & Mokrane, M. (2020). FAIRsFAIR Comments Response on RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specification and guidelines. *Zenodo*. https://zenodo.org/record/3827109

Lacagnina, C., David, R., Nikiforova, A., Kuusniemi, M. E., Cappiello, C., Biehlmaier, O., Wright, L., Schubert, C., Bertino, A., & Thiemann, H. (2022, December 1). *charter draft - data quality v1*. Google Docs. https://docs.google.com/document/d/1J6M4oy8PIxkSLcivVNK4HzW4qKFaOIgnvlDEPIrD3Ag/edit#heading=h.gwxzh37zo2eb

Lachney, M., & Foster, E. K. (2020). Historicizing making and doing: Seymour Papert, Sherry Turkle, and epistemological foundations of the maker movement. History and Technology, 36(1), 54–82. https://doi.org/10.1080/07341512.2020.1759302

Lahn, B. (2020). *Changing climate change: The carbon budget and the modifying-work of the IPCC - Bård Lahn, 2021*. Social Studies of Science. https://journals-sagepub-com.uaccess.univie.ac.at/doi/10.1177/0306312720941933

Lamont, M. (2012). Toward a Comparative Sociology of Valuation and Evaluation. *Annual Review of Sociology*, *38*(1), 201–221. https://doi.org/10.1146/annurev-soc-070308-120022

Land, N. (2023). *Bodying Postqualitative Research*. Routledge. https://doi.org/10.4324/9781003353676

Lapsley, I. (2008). THE NPM AGENDA: BACK TO THE FUTURE. *Financial Accountability & Management*, *24*(1), 77–96. https://doi.org/10.1111/j.1468-0408.2008.00444.x

Larkin, B. (2013). The Politics and Poetics of Infrastructure. *Annual Review of Anthropology*, *42*(1), 327–343. https://doi.org/10.1146/annurev-anthro-092412-155522

Lash, S. M., Szersynski, B., & Wynne, B. (1996). *Risk, Environment and Modernity: Towards a New Ecology*. Sage Publications.

Latour, B. (1988). *Science in Action* . Harvard.edu. https://www.hup.harvard.edu/catalog.php?isbn=9780674792913

Latour, B. (1993a). *The pasteurization of France*. Harvard University Press.

Latour, B. (1993b). *We have never been modern*. Harvard University Press.

Latour, B. (1999). For David Bloor… and Beyond: A Reply to David Bloor's "Anti-Latour." *Studies in History and Philosophy of Science*, *30*(1), 113–129.

Latour, B. (2004). Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry*, *30*(2), 225–248.

Latour, B., & Woolgar, S. (1979). *Laboratory life : the social construction of scientific facts*. Princeton University Press.

Law, J. (1994). *Organizing modernity*. Blackwell.

Law, J. (2004). *After Method: Mess in Social Science Research*. Routledge.

Law, J. (2015). *What's wrong with a one-world world?* Distinktion: Scandinavian Journal of Social Theory. https://www.tandfonline.com/doi/full/10.1080/1600910X.2015.1020066

Lawrence, J. M., Gideon Meyerowitz-Katz, James, Nicholas, & Sheldrick, K. A. (2021). The lesson of ivermectin: meta-analyses based on summary data alone are inherently unreliable. *Nature Medicine*, *27*(11), 1853–1854. https://doi.org/10.1038/s41591-021-01535-y

Leonelli, S. (2015). *What Counts as Scientific Data? A Relational Framework*. *82*(5), 810–821. https://doi.org/10.1086/684083

Leonelli, S., Lovell, R., Wheeler, B. W., Fleming, L., & Williams, H. (2021). From FAIR data to fair data use: Methodological data fairness in health-related social media research. *Big Data & Society*, *8*(1), 205395172110103. https://doi.org/10.1177/20539517211010310

Leonelli, S., & Tempini, N. (2020). *Data Journeys in the Sciences*. Springer Nature.

Levidow, L., & Neubauer, C. (2014). EU Research Agendas: Embedding What Future? *Science as Culture*, *23*(3), 397–412. https://doi.org/10.1080/09505431.2014.926149

Lipsky, M. (1969). *Toward a theory of street-level bureaucracy*. Institute For Research On Poverty, University Of Wisconsin.

Lipsky, M. (2010). *Street-level bureaucracy : dilemmas of the individual in public services*. Russell Sage Foundation. https://muse.jhu.edu/chapter/504582

Lorentz center. (2014). *Jointly designing a Data FAIRPORT*. Lorentzcenter.nl. https://www.lorentzcenter.nl/jointly-designing-a-data-fairport.html

Lösch, A. (2006). Anticipating the futures of nanotechnology: Visionary images as means of communication. *Technology Analysis & Strategic Management*, *18*(3-4), 393–409. https://doi.org/10.1080/09537320600777168

Lösch, A., Heil, R., & Schneider, C. (2019). Visionary Practices Shaping Power Constellations. *Technikzukünfte, Wissenschaft Und Gesellschaft*, 67–88. https://doi.org/10.1007/978-3-658-27155-8_4

LSC. (2016, February 11). *GW150914 - The First Direct Detection of Gravitational Waves*. Www.ligo.org. https://www.ligo.org/detections/GW150914.php

Luschi, A., Petraccone, C., Fico, G., Pecchia, L., & Iadanza, E. (2023). Semantic Ontologies for Complex Healthcare Structures: A Scoping Review. *IEEE Access*, *11*, 19228–19246. https://doi.org/10.1109/access.2023.3248969

Mackenzie, D. A., & Wajcman, J. (1999). *The social shaping of technology*. Open University Press.

Mackinac Center. (2010). *The Overton Window*. Mackinac Center. https://www.mackinac.org/OvertonWindow

Majone, G., & Wildavsky, A. (1978). Implementation as evolution. *Policy Studies Review Annual*, *2*.

Marius Korsnes, Berker, T., & Woods, R. (2018). Domestication, acceptance and zero emission ambitions: Insights from a mixed method, experimental research design in a Norwegian Living Lab. *Energy Research & Social Science*, *39*, 226–233. https://doi.org/10.1016/j.erss.2017.11.004

Marr, B. (2023, September 12). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. *Forbes*. https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/

Mau, S. (2019). *The metric society : on the quantification of the social*. Polity.

Maxmen, A. (2021). Why some researchers oppose unrestricted sharing of coronavirus genome data. *Nature*, *593*(7858), 176–177. https://doi.org/10.1038/d41586-021-01194-6

Mazur, A. (1987). Scientific disputes over policy. *Cambridge University Press EBooks*, 265–282. https://doi.org/10.1017/cbo9780511628719.011

McEntee, M., Fabien Medvecky, MacBride-Stewart, S., Macknight, V., & Martin, M. (2023). Park Rangers and Science-Public Expertise: Science as Care in Biosecurity for Kauri Trees in Aotearoa/New Zealand. *Minerva*, *61*(1), 117–140. https://doi.org/10.1007/s11024-022-09482-9

Meleco, Y. (2014, November). *Conference Report: Research Data Alliance (RDA) 4th Plenary Meeting*. Www.dlib.org. http://www.dlib.org/dlib/november14/meleco/11meleco.html

Meyer, M., & Kearnes, M. (2013). Introduction to special section: Intermediaries between science, policy and the market. *Science and Public Policy*, *40*(4), 423–429. https://doi.org/10.1093/scipol/sct051

Michael, M. (2017). *Enacting Big Futures, Little Futures: Toward an ecology of futures - Mike Michael, 2017*. The Sociological Review. https://journals.sagepub.com/doi/full/10.1111/1467-954X.12444?journalCode=sora

Miedema, F. (2022). Open Science: the Very Idea. *SpringerLink*. https://doi.org/10.1007-978-94-024-2115-6

Mintzberg, H. (1994). *Rise and fall of strategic planning.* Free Press.

Mitchell, S., Lahiff, A., Cummings, N., Hollocombe, J., Boskamp, B., Field, R., Reddyhoff, D., Zarebski, K., Wilson, A., Viola, B., Burke, M., Archibald, B., Bessell, P., Blackwell, R., Boden, L., Brett, A., Brett, S., Dundas, R., Enright, J., & Gonzalez-Beltran, A. (2022). *FAIR Data Pipeline: provenance-driven data management for traceable scientific workflows*. https://arxiv.org/pdf/2110.07117

Moedas, C. (2017, June 12). *EOSC Summit: The European Open Science Cloud – The New Republic of Letters - European Commission*. European Commission. https://web.archive.org/web/20180321122850/https://ec.europa.eu/commission/commissioners/2014-2019/moedas/announcements/eosc-summit-european-open-science-cloud-new-republic-letters_en

Mol, A. (1999). *Ontological Politics. A Word and Some Questions*. *47*(1_suppl), 74–89. https://doi.org/10.1111/j.1467-954x.1999.tb03483.x

Mol, A. (2002). *The body multiple : ontology in medical practice*. Duke University Press. https://www.degruyter.com/document/doi/10.1515/9780822384151/html

Mol, A. (2010). Actor-Network Theory: sensitive terms and enduring tensions. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie. Sonderheft*, *50*. https://hdl.handle.net/11245/1.330874

Mol, A. (2021). Eating in Theory. In *Duke University Press eBooks*. Duke University Press. https://doi.org/10.1515/9781478012924

Mol, A., & Berg, M. (1994). Principles and practices of medicine. *Culture Medicine and Psychiatry*, *18*(2), 247–265. https://doi.org/10.1007/bf01379451

Mons, B. (2005). Which gene did you mean? *BMC Bioinformatics*, *6*(1). https://doi.org/10.1186/1471-2105-6-142

Mons, B. (2013, February 25). *Kennis is als liefde... men wordt van het delen niet minder*. https://scholarlypublications.universiteitleiden.nl/access/item%3A2958959/download

Mons, B. (2014). *Bringing data to broadway*. https://collegerama.tudelft.nl/Mediasite/Play/0844aefac5bb49ca9032069c6edc668f1d?catalog=3984a02f-bf33-4c70-a080-94a04d3e8112

Mons, B. (2018a, August 3). *Comment on FAIR-Data-EG-INTERIM-Report-CONSULTATION*. Google Docs. https://docs.google.com/document/d/1FcnBOPUm-9gW3gjF18FDoPNUvIn-ojduLeLe0Y8yEH8/edit?disco=AAAACCReETs

Mons, B. (2018b). *Data Stewardship for Open Science*. CRC Press. https://www.routledge.com/Data-Stewardship-for-Open-Science-Implementing-FAIR-Principles/Mons/p/book/9781032095707

Mons, B. (2018c, December 6). *How and why make your research data FAIR?* Www.youtube.com. https://www.youtube.com/watch?v=CsrWenP3Dfo

Mons, B. (2019, December 9). *The Guest - Barend Mons*. Www.youtube.com. https://www.youtube.com/watch?v=IaCuKW_SMec

Mons, B. (2020a, May 9). *201127 | 11 | Plenary 1 | International Open Science in 2020*. Vimeo. https://vimeo.com/499267657

Mons, B. (2020b, June 3). *The FAIR Guiding Principles in Times of Crisis*. Www.youtube.com. https://www.youtube.com/watch?v=lRciVnP9WIY&list=PLrRaxjvn6FDUXLEulK_krDTDiwZpGK1FW&index=6

Mons, B. (2020c, October 4). *Barend Mons: Q&A Session*. Www.youtube.com. https://www.youtube.com/watch?v=sqwWL2BC3JY

Mons, B. (2021a). Comments to Jean-Claude Burgelman's article *Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story)* —“EOSC is a bigger ME” and the Dunning Kruger effect. *Data Intelligence*, *3*(1), 32–39. https://doi.org/10.1162/dint_a_00074

Mons, B. (2021b). FAIR-DI - FAIRmat Colloquium with Barend Mons on How to materialise FAIR [YouTube Video]. In *YouTube*. https://www.youtube.com/watch?v=N4CBAqKQmQs

Mons, B. (2021c). How to materialise FAIR [YouTube Video]. In *YouTube*. https://www.youtube.com/watch?v=N4CBAqKQmQs

Mons, B. (2022, January 10). *Unesco Recommendation on Open Science – Thoughts by Barend Mons*. Frontiers Policy Labs. https://policylabs.frontiersin.org/content/commentary-unesco-barend-mons

Mons, B., Herman van Haagen, Chichester, C., Peter, Johan, Gert-Jan van Ommen, Erik, Bharat Bhushan Singh, Hooft, R., Roos, M., Hammond, J., Kiesel, B., Giardine, B., Velterop, J., Groth, P., & Schultes, E. (2011). *The value of data*. *43*(4), 281–283. https://doi.org/10.1038/ng0411-281

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, *37*(1), 49–56. https://doi.org/10.3233/isu-170824

Mons, B., Schultes, E., Liu, F., & Jacobsen, A. (2020). The FAIR Principles: First Generation Implementation Choices and Challenges. *Data Intelligence*, 1–9. https://doi.org/10.1162/dint_e_00023

Mons, B., & Velterop, J. (2009). *Nano-Publication in the e-science era*. https://www.semanticscholar.org/paper/Nano-Publication-in-the-e-science-era-Mons-Velterop/195c7924d58e729c1c66ace76a0bab06a856afd5

Mubaya, R. (2020). *[-In]tangible Heritage, Humans and the Environment: An Ethnographic Account of the Conservation of Chingoma Falls in South-eastern Zimbabwe*. https://pure.uvt.nl/ws/portalfiles/portal/32599931/Mubaya_in_tangible_06_03_2020.pdf

Mustajoki, H., Pölönen, J., Gregory, K., Ivanović, D., Brasse, V., Kesäniemi, J., Koivisto, E., & Pylvänäinen, E. (2021). Making FAIReR assessments possible. Final report of EOSC Co-Creation projects: "European overview of career merit systems'' and “Vision for research data in research careers.” *Zenodo*. https://zenodo.org/record/4701375#.Y20dX-zMI-Q

National Academy Of Sciences, National Academy Of Engineering, & Institute Of Medicine. (2009). *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. National Academies Press.

NCBI. (1998). *NCBI News NCBI News NCBI News Speed and Sensitivity: BLAST Version 2.0*. https://www.ncbi.nlm.nih.gov/Web/Newsltr/feb98.pdf

Needham, E. K., & Lanzer, E. L. (1993). The Puget sound environmental atlas update: Context and GIS database integration issues. *Computers, Environment and Urban Systems*, *17*(5), 409–424. https://doi.org/10.1016/0198-9715(93)90037-6

Nestor. (2006). Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. *Hu-Berlin.de*. http://edoc.hu-berlin.de/18452/2175

NHGRI. (1996, April 15). *NHGRI Workshop on DNA Sequence Validation*. Archive.org. https://web.archive.org/web/19980204143013/http://www.nhgri.nih.gov/HGP/Reports/dna_sequence_workshop.html

Nielsen, M. (2008, July 17). *The Future of Science – Michael Nielsen*. Michaelnielsen.org. https://michaelnielsen.org/blog/the-future-of-science-2/

NIH. (n.d.). *Big Data to Knowledge | NIH Common Fund*. Commonfund.nih.gov. https://commonfund.nih.gov/bd2k

NIH. (2000). *International Human Genome Sequencing Consortium Announces \"Working Draft\" of Human Genome*. Genome.gov. https://www.genome.gov/10001457/2000-release-working-draft-of-human-genome-sequence

NIH. (2004). *International Human Genome Sequencing Consortium- Describes Finished Human Genome Sequence*. Genome.gov. https://www.genome.gov/12513430/2004-release-ihgsc-describes-finished-human-sequence

NIH. (2012). *National Institutes of Health Data and Informatics Working Group Draft Report to The Advisory Committee to the Director*. https://acd.od.nih.gov/documents/reports/DataandInformaticsWorkingGroupReport.pdf

NIH. (2017, May 18). *Data Commons*. Commonfund.nih.gov. https://commonfund.nih.gov/commons

NIH. (2022, July 5). *Human Genome Project Timeline*. Genome.gov. https://www.genome.gov/human-genome-project/timeline

NL EU 2016. (2016). *Amsterdam Call for Action on Open Science*. https://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf

Nordling, J., Mihai, H., Meerman, B., Alaterä, T., Kleemola, M., & Livenson, I. (2022). D4.3 Report on Nordic and Baltic repositories and their uptake of FAIR. *Zenodo*. https://zenodo.org/record/6880904#.YwxAqnZBw2w

Nowotny, H., Scott, P., & Gibbons, M. (2001). *Re-thinking science : knowledge and the public in an age of uncertainty*. Polity.

NWO. (n.d.). *Research data management | NWO*. NWO. Retrieved November 29, 2023, from https://www.nwo.nl/en/research-data-management

OECD. (2018). OECD Science, Technology and Innovation Outlook 2018. In *OECD Science, Technology and Innovation Outlook*. OECD. https://doi.org/10.1787/sti_in_outlook-2018-en

OECD. (2020). *The Digitalisation of Science, Technology and Innovation*. OECD. https://doi.org/10.1787/b9e4a2c0-en

Offenhuber, D. (2017). *Waste Is Information*. MIT Press.

Okhowat, A. (2020, April 17). *Sharing tools and data globally will help us beat COVID-19*. World Economic Forum. https://www.weforum.org/agenda/2020/04/sharing-data-to-beat-covid-19/

Oomen, J., Hoffman, J., & Hajer, M. A. (2021). Techniques of futuring: On how imagined futures become socially performative. *European Journal of Social Theory*, 136843102098882. https://doi.org/10.1177/1368431020988826

OpenPHACTS. (2013). *Beyond the PDF 2 Conference – Open PHACTS Foundation*. Openphactsfoundation.org. http://www.openphactsfoundation.org/beyond-the-pdf-2-conference/

Outsell. (2023). Outsell – A Discussion with Barend Mons: The Future of Data—FAIR and its Implementation [YouTube Video]. In *YouTube*. https://www.youtube.com/watch?v=NkpAXEze-hw

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy*, *39*(6), 751–760. https://doi.org/10.1093/scipol/scs093

Paltieli, G. (2021). The political imaginary of National AI Strategies. *AI & SOCIETY*. https://doi.org/10.1007/s00146-021-01258-1

Parag, Y., & Janda, K. B. (2014). More than filler: Middle actors and socio-technical change in the energy system from the “middle-out.” *Energy Research & Social Science*, *3*, 102–112. https://doi.org/10.1016/j.erss.2014.07.011

Pashler, H., & Harris, C. R. (2012). Is the Replicability Crisis Overblown? Three Arguments Examined. *Perspectives on Psychological Science*, *7*(6), 531–536. https://doi.org/10.1177/1745691612463401

Petters, J. (2019, April 3). *What would be the format of the assessment? · Issue #9 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/9#issuecomment-479564686

Pettifer, S., McDermott, P., Marsh, J., Thorne, D. J., Villéger, A., & Attwood, T. K. (2011). *Ceci n’est pas un hamburger: modelling and representing the scholarly article*. *24*(3), 207–220. https://doi.org/10.1087/20110309

Pfotenhauer, S., Juhl, J., & Aarden, E. (2019). Challenging the “deficit model” of innovation: Framing policy issues under the innovation imperative. *Research Policy*, *48*(4), 895–904. https://doi.org/10.1016/j.respol.2018.10.015

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, *14*(3), 399–441. https://www.jstor.org/stable/285355

Pink, S. (2015). *Doing Sensory Ethnography*. SAGE.

Pink, S., Horst, H. A., Postill, J., Hjorth, L., Lewis, T., & Tacchi, J. (2016). *Digital ethnography : principles and practice*. Sage Publications.

Pisani, E., Amrita Ghataure, & Merson, L. (2018). *Data sharing in public health emergencies: A study of current policies, practices and infrastructure supporting the sharing of data to prevent and respond to epidemic and pandemic threats*. https://doi.org/10.6084/m9.figshare.5897608.v1

Plantin, J.-C. (2018). Data Cleaners for Pristine Datasets: Visibility and Invisibility of Data Processors in Social Science. *Science, Technology, & Human Values*, *44*(1), 52–73. https://doi.org/10.1177/0162243918781268

Poirier, L. (2017). A Turn for the Scruffy: An Ethnographic Study of Semantic Web Architecture. *Proceedings of the 2017 ACM on Web Science Conference*. https://doi.org/10.1145/3091478.3091505

Poirier, L. (2019a). Classification as Catachresis: Double Binds of Representing Difference with Semiotic Infrastructure. *Canadian Journal of Communication*, *44*(3). https://doi.org/10.22230/cjc.2019v44n3a3455

Poirier, L. (2019b). Making the Web Meaningful: A History of Web Semantics. In *The SAGE Handbook of Web History* (pp. 256–269). SAGE Publications Ltd. https://doi.org/10.4135/9781526470546

Portugal 2000. (2000, June 19). *Portugal 2000*. Web.archive.org. https://web.archive.org/web/20000619102140/http://www.portugal.ue-2000.pt/uk/frame.htm

Potočnik, J. (2007, April 4). *European Commission - Janez Potočnik: Theme of the week*. Archive.org. https://web.archive.org/web/20070703111300/http:/blogs.ec.europa.eu/blog_potocnik/page/potocnik?entry=giving_europe_a_fifth_freedom

Quote Investigator. (2010, July 8). *Laws are Like Sausages. Better Not to See Them Being Made – Quote Investigator®*. Quoteinvestigator.com. https://quoteinvestigator.com/2010/07/08/laws-sausages/

Randles, S., Dorbeck-Jung, B., Lindner, R., & Rip, A. (2014). Where to Next for Responsible Innovation? In *Innovation and responsibility: Engaging with new and emerging technologies* (pp. 19–35).

RCN. (2019, April 15). *Sharing research data*. Forskningsradet. https://www.forskningsradet.no/en/research-policy-strategy/open-science/research-data/

RDA. (2014a). *Plenary 1*. Tudelft.nl. https://collegerama.tudelft.nl/Mediasite/Play/0844aefac5bb49ca9032069c6edc668f1d?catalog=3984a02f-bf33-4c70-a080-94a04d3e8112

RDA. (2014b). *Plenary 2*. Tudelft.nl. https://collegerama.tudelft.nl/Mediasite/Play/55eb502f04aa4124ac94b75584c62abc1d?catalog=3984a02f-bf33-4c70-a080-94a04d3e8112

RDA. (2014c, February 6). *Plenary4-Programme*. RDA. https://www.rd-alliance.org/plenary-meetings/fourth-plenary/plenary4-programme.html

RDA. (2014d). *The Data Harvest: How sharing research data can yield knowledge, jobs and growth An RDA Europe Report*. https://rd-alliance.org/sites/default/files/attachment/The%20Data%20Harvest%20Final.pdf

RDA. (2016, March 22). *Get involved*. RDA. https://www.rd-alliance.org/get-involved.html

RDA. (2018). *RDA Europe Planning and Activity report D2.4*. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5d351c6b3&appId=PPGMS

RDA. (2020, September 22). *Accelerating the implementation of FAIR assessment in practice*. RDA. https://www.rd-alliance.org/accelerating-implementation-fair-assessment-practice

RDA FAIR Data Maturity Model Working Group. (n.d.). *FAIR-data-maturity-model-WG*. GitHub. Retrieved November 15, 2022, from https://github.com/rd-alliance/FAIR-data-maturity-model-WG

RDA FAIR Data Maturity Model Working Group. (2018). *Case statement*. Rd-Alliance. https://web.archive.org/web/20230602210817/https://www.rd-alliance.org/sites/default/files/case_statement/Case%20Statement-FAIR%20Data%20Maturity%20Model-v1.7.pdf

RDA FAIR Data Maturity Model Working Group. (2019a, February 22). *FAIR Data Maturity Model Online meeting #1*. Rd-Alliance.org. https://www.rd-alliance.org/system/files/documents/20190221_FAIR_WG_slides_v0.14.pdf

RDA FAIR Data Maturity Model Working Group. (2019b, March 6). *P13 Breakout Session - WG FAIR Data Maturity Model: First Face-to-face Meeting*. Google Docs. https://docs.google.com/document/d/1oRhn0eiY9TQ7X-E0rDmYEig5domC_CqyVbh8q7FEkUU/edit#

RDA FAIR Data Maturity Model Working Group. (2019c, March 6). *What would be the essential criteria to be FAIR? · Issue #6 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/6

RDA FAIR Data Maturity Model Working Group. (2019d, April 3). *Workshop report 13th RDA Plenary - Workshop #2*. Rd-Alliance.org. https://www.rd-alliance.org/system/files/documents/20190403_FAIR_WG_Workshop_2_report_0.pdf

RDA FAIR Data Maturity Model Working Group. (2019e, August 9). *Indicators prioritisation for Reusability · Issue #33 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/33

RDA FAIR Data Maturity Model Working Group. (2019f, September 23). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34

RDA FAIR Data Maturity Model Working Group. (2019g, December 4). *FAIR Data Maturity Model Workshop #6*. Https://Www.rd-Alliance.org/System/Files/Documents/20191204_FAIR_WG_slides_v0.04.Pdf. https://www.rd-alliance.org/system/files/documents/20191204_FAIR_WG_slides_v0.04.pdf

RDA FAIR Data Maturity Model Working Group. (2020a, February 13). *FAIR Data Maturity Model Workshop #7*. https://www.rd-alliance.org/system/files/documents/20200213_FAIR_WG_slides_v0.05_0.pdf

RDA FAIR Data Maturity Model Working Group. (2020b). FAIR Data Maturity Model. Specification and Guidelines. *Zenodo*. https://zenodo.org/record/3909563#.YGRNnq8za70

Reichmann, S. (2022). *Data sharing for global, unspecified reuse*. https://doi.org/10.31235/osf.io/tbex2

Reis, R. B., Ribeiro, G. S., Dias, R., Santana, F. M., Mohr, S., Astrid, Queiroz, A., Santos, A., Rogger Ravines, Wagner, Marilia Sá Carvalho, Reis, M. G., & Ko, A. I. (2008). *Impact of Environment and Social Gradient on Leptospira Infection in Urban Slums*. *2*(4), e228–e228. https://doi.org/10.1371/journal.pntd.0000228

Research information Network (RIN). (2008). *Stewardship of digital research data: a framework of principles and guidelines* . http://shapingthefuture.pbworks.com/f/Stewardship+of+digital+research+data+by+RIN.pdf

Research Libraries Group (RLG). (2002). *Trusted Digital Repositories: Attributes and Responsibilities An RLG-OCLC Report RLG Mountain View, CA*. https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf

Ribes, D. (2017). *Notes on the Concept of Data Interoperability | Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM Conferences. https://dl.acm.org/doi/10.1145/2998181.2998344

Ribes, D. (2018). STS, Meet Data Science, Once Again. *Science, Technology, & Human Values*, *44*(3), 514–539. https://doi.org/10.1177/0162243918798899

Ribes, D., & Bowker, G. C. (2009). Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization*, *19*(4), 199–217. https://doi.org/10.1016/j.infoandorg.2009.04.001

Ribes, D., Hoffman, A. S., Slota, S. C., & Bowker, G. C. (2019). The logic of domains. *Social Studies of Science*, *49*(3), 281–309. https://doi.org/10.1177/0306312719849709

Rieder, G., & Simon, J. (2016). Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data. *Big Data & Society*, *3*(1), 205395171664939. https://doi.org/10.1177/2053951716649398

RIN, & British Library. (2009). Patterns of information use and exchange: case studies of researchers in the life sciences. In *rin.ac.uk*. http://rin.ac.uk/system/files/attachments/Patterns_information_use-REPORT_Nov09.pdf

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169. https://doi.org/10.1007/bf01405730

Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, *14*(11). https://doi.org/10.1111/phc3.12633

Rottenburg, R., Merry, S. E., Park, S.-J., & Mugler, J. (2015). *The World of Indicators*. Cambridge University Press.

Rowley, J. (2016). *The wisdom hierarchy: representations of the DIKW hierarchy - Jennifer Rowley, 2007*. Journal of Information Science. https://journals-sagepub-com.uaccess.univie.ac.at/doi/abs/10.1177/0165551506070706

Royal Society. (2012). *Science as an open enterprise*. https://royalsociety.org/-/media/policy/projects/sape/2012-06-20-saoe.pdf

Russo Carroll, S., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, *19*(1). https://datascience.codata.org/articles/10.5334/dsj-2020-043

Russo Carroll, S., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, *8*(1). https://doi.org/10.1038/s41597-021-00892-0

Sample, I. (2018, November 9). *In the balance: scientists vote on first change to kilogram in a century*. The Guardian. https://www.theguardian.com/science/2018/nov/09/in-the-balance-scientists-vote-on-first-change-to-kilogram-in-century

Sansom, C. (2010). The Semantic Biochemical Journal: Navigating the data wilderness. *The Biochemist*, *32*(3), 37–38. https://doi.org/10.1042/bio03203037

Santos, L. O. B. da S., Guizzardi, G., & Sales, T. P. (2022). *FAIR Digital Object Framework Documentation*. Fairdigitalobjectframework.org. https://fairdigitalobjectframework.org/

Schaller, R. R. (1997). Moore's law: past, present and future. *IEEE Spectrum*, *34*(6), 52–59. https://doi.org/10.1109/6.591665

Schneider, C., & Lösch, A. (2019). Visions in assemblages: Future-making and governance in FabLabs. *Futures*, *109*, 203–212. https://doi.org/10.1016/j.futures.2018.08.003

Schneiderman, B. (2008, March 7). *Science 2.0*. Science. https://www.science.org/doi/10.1126/science.1153539

Schoch, C. L., Ciufo, S., Mikhail Domrachev, Hotton, C. L., Kannan, S., Rogneda Khovanskaya, Detlef Leipe, Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Vladimir Soussov, Sullivan, J. P., Sun, L., Turner, S., & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database*, *2020*. https://doi.org/10.1093/database/baaa062

Sepehr, P., & Felt, U. (2023). Urban Imaginaries as Tacit Governing Devices: The Case of Smart City Vienna. *Science, Technology, & Human Values*, 016224392311785-016224392311785. https://doi.org/10.1177/01622439231178597

Shankar, K., Hakken, D., & Østerlund, C. (2017). Rethinking documents. In U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr (Eds.), *The handbook of science and technology studies* (pp. 59–86). The MIT Press.

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump : Hobbes, Boyle, and the experimental life ; including a translation of Thomas Hobbes, Dialogus physicus de natura aeris, by Simon Schaf*. Princeton University Press.

Shotton, D. M., Portwin, K., Klyne, G., & Miles, A. (2009). *Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article*. *5*(4), e1000361–e1000361. https://doi.org/10.1371/journal.pcbi.1000361

Silverstone, R., & Hirsch, E. (1992). *Consuming technologies : media and information in domestic spaces*. Routledge.

Simonsen, J., Karasti, H., & Morten Hertzum. (2019). Infrastructuring and Participatory Design: Exploring Infrastructural Inversion as Analytic, Empirical and Generative. *Computer*

Supported Cooperative Work (CSCW)/Computer Supported Cooperative Work, *29*(1-2), 115–151. https://doi.org/10.1007/s10606-019-09365-w

Sinsheimer, R. L. (1984). Human Genome Institute: A position paper. *BMJ*, *288*(6420), 856–856. https://doi.org/10.1136/bmj.288.6420.856

Sinsheimer, R. L. (1984, November 19). [Letter to David Pierpont Gardner].

Sinsheimer, R. L. (1989, November). *Historical Sketch: The Santa Cruz Workshop | Genomics Institute*. Genomics.ucsc.edu. https://genomics.ucsc.edu/2018/09/21/historical-sketch-the-santa-cruz-workshop/

Slovenian Presidency. (2008a, April 15). *EU2008.si - Launch of "Ljubljana Process" to revive the European Research Area*. Eu2008.Si. http://www.eu2008.si/en/News_and_Documents/Press_Releases/April/0415MVZT_COMPET.html

Slovenian Presidency. (2008b, April 15). *Informal Meeting of Ministers for Competitiveness (Research)* . http://www.eu2008.si/en/News_and_Documents/download_docs/April/0414_COMPET/070Summary_Research.pdf

Sørensen, K. (1996). *Learning technology, constructing culture: Sociotechnical change as social learning*.

Spillane, J. P. (2002). *Policy Implementation and Cognition: Reframing and Refocusing Implementation Research - James P Spillane, Brian J Reiser, Todd Reimer, 2002*. Review of Educational Research. https://journals-sagepub-com.uaccess.univie.ac.at/doi/abs/10.3102/00346543072003387

Staab, S., & Studer, R. (2009). *Handbook on Ontologies*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-92673-3

Star, S. L. (1989). The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. *Distributed Artificial Intelligence*, 37–54. https://doi.org/10.1016/b978-1-55860-092-8.50006-x

Star, S. L. (1990). Power, Technology and the Phenomenology of Conventions: On being Allergic to Onions. *The Sociological Review*, *38*(1_suppl), 26–56. https://doi.org/10.1111/j.1467-954x.1990.tb03347.x

Star, S. L. (2002). Infrastructure and ethnographic practice: working on the fringes. *Scandinavian Journal of Information Systems*, *14*(2), 107–122.

Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420. https://doi.org/10.1177/030631289019003001

Star, S. L., & Ruhleder, K. (1994). *Steps towards an ecology of infrastructure | Proceedings of the 1994 ACM conference on Computer supported cooperative work*. ACM Conferences. https://dl.acm.org/doi/10.1145/192844.193021

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Stehouwer, H., & Wittenburg, P. (2015). *Principles for Data Sharing and Re-use: are they all the same?* https://b2share.eudat.eu/records/8913da0e5c934eb8b924e7b33bb4ec7f

Sterling, B. (2005). Ivory Tower. *Nature*, *434*(7034), 806–806. https://doi.org/10.1038/434806a

Sternberg, C. S. (2013). *The struggle for EU legitimacy Public contestation, 1950-2005*. Palgrave Macmillan.

Stirling, A. (2008). “Opening Up” and “Closing Down.” *Science, Technology, & Human Values*, *33*(2), 262–294. https://doi.org/10.1177/0162243907311265

Stone, D. A. (1997). *Policy paradox : the art of political decision making*. W.W. Norton.

Strawn, G. (2021). Open Science and the Hype Cycle. *Data Intelligence*, 1–7. https://doi.org/10.1162/dint_a_00081

Styhre, A. (2007). *The Innovative Bureaucracy*. Routledge.

Tacke, O. (2010). Open Science 2.0: How Research and Education Can Benefit from Open Innovation and Web 2.0. *Advances in Intelligent and Soft Computing*, 37–48. https://doi.org/10.1007/978-3-642-14481-3_4

ter Horst, R., Jaeger, M., Smeekens, S. P., Oosting, M., Swertz, M. A., Li, Y., Kumar, V., Diavatopoulos, D. A., Jansen, A. F. M., Lemmers, H., Toenhake-Dijkstra, H., van Herwaarden, A. E., Janssen, M., van der Molen, R. G., Joosten, I., Sweep, F. C. G. J., Smit, J. W., Netea-Maier, R. T., Koenders, M. M. J. F., & Xavier, R. J. (2016). Host and Environmental Factors Influencing Individual Human Cytokine Responses. *Cell*, *167*(4), 1111-1124.e13. https://doi.org/10.1016/j.cell.2016.10.018

The Economist. (2010, February 25). *The data deluge*. The Economist; The Economist. https://www.economist.com/leaders/2010/02/25/the-data-deluge

Thylstrup, N. B., Hansen, K. B., Flyverbom, M., & Amoore, L. (2022). Politics of data reuse in machine learning systems: Theorizing reuse entanglements. *Big Data & Society*, *9*(2), 205395172211397. https://doi.org/10.1177/20539517221139785

Timmermans, S., & Epstein, S. (2010). A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. *Annual Review of Sociology*, *36*(1), 69–89. https://doi.org/10.1146/annurev.soc.012809.102629

Tochtermann, K. (2019). *FAIR DATA MANAGEMENT*. https://www.medizininformatik-initiative.de/sites/default/files/2020-07/ZBW_MII_Jahresversammlung_Tochtermann.pdf

Tóth-Czifra, E. (2019). The risk of losing thick description: Data management challenges Arts and Humanities face in the evolving FAIR data ecosystem. *Archives-Ouvertes.fr*. https://halshs.archives-ouvertes.fr/halshs-02115505

Tsing, A. L. (2005). *Friction*. Princeton University Press. https://doi.org/10.1515/9781400830596

Tutton, R. (2016). *Promising pessimism: Reading the futures to be avoided in biotech - Richard Tutton, 2011*. Social Studies of Science. https://journals-sagepub-com.uaccess.univie.ac.at/doi/abs/10.1177/0306312710397398

Tweede Kamer der Staten-Generaal. (2013). Kamerstuk 29338, nr. 120 | Overheid.nl > Officiële bekendmakingen. *Officielebekendmakingen.nl*. https://doi.org/kst-29338-120

Uprichard, E. (2013, October). *Focus: Big Data, Little Questions?* Discover Society. https://archive.discoversociety.org/2013/10/01/focus-big-data-little-questions/

van Damme, P., Fernández-Breis, J. T., Benis, N., Miñarro-Gimenez, J. A., F. de Keizer, N., & Cornet, R. (2022). Performance assessment of ontology matching systems for FAIR data. *Journal of Biomedical Semantics*, *13*(1). https://doi.org/10.1186/s13326-022-00273-5

Van Duijn, S. (2020). Everywhere and nowhere at once: the challenges of following in multi-sited ethnography. *Journal of Organizational Ethnography*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/joe-12-2019-0045 van Mulligen, E. M., Diwersy, M., Schmidt, M., Buurman, H., & Mons, B. (2000). Facilitating networks of information. *Proceedings. AMIA Symposium*, 868–872. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243846/

von der Leyen, U. (2020, September 16). *State of the Union 2020*. State of the Union. https://state-of-the-union.ec.europa.eu/state-union-2020_en

W3C. (2010). *Ontologies - W3C*. Archive.org. https://web.archive.org/web/20110809030351/http://www.w3.org:80/standards/semanticweb/ontology.html

Waibel, D., Peetz, T., & Meier, F. (2021). Valuation Constellations. *Valuation Studies*, *8*(1), 33–66. https://doi.org/10.3384/vs.2001-5992.2021.8.1.33-66

Waldrop, M. M. (2008). Science 2.0. *Scientific American*, *298*(5), 68–73. https://doi.org/10.1038/scientificamerican0508-68

Watson, C. (2022). Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. *Nature Medicine*, *28*(1), 2–5. https://doi.org/10.1038/s41591-021-01654-6

Weber, H. (2005). *On "DOMESTICATION" or: Who is domesticating what or whom?* https://www.lancaster.ac.uk/fass/projects/dnc/wkshpsjan06/papers/12-13th/weber.pdf

Weiss, M. G. (2017). Valuography. *BioSocieties*, *12*(3), 461–465. https://doi.org/10.1057/s41292-017-0060-5

Whyte, A., Leenarts, E., De Vries, J., Huigen, F., Kuehn, E., Sipos, G., Kalaitzi, V., Dijk, E., Jones, S., & Ashley, K. (2019). *D7.5: Strategy for Sustainable Development of Skills and Capabilities*. https://eoscpilot.eu/sites/default/files/eoscpilot-d7.5-v1.1.pdf

Wilkinson, M. (2019a, February 22). *TITLE: FAIR Metric FM-F2*. GitHub. https://github.com/FAIRMetrics/Metrics/blob/master/MaturityIndicators/Gen1/FM_F2.md

Wilkinson, M. (2019b, March 15). *Examining FAIR "Maturity."* https://docs.google.com/presentation/d/1FzpOjx_jNZWHTpj7hfau4Y1p-W0zUApmd_cUMqpRYSw/edit#slide=id.p

Wilkinson, M. (2019c, March 28). *FAIR Maturity Indicators & Evaluation*. https://indico.neic.no/event/56/sessions/93/attachments/42/66/4.2._Day_4_Late_PM_FAIR_Maturity__Evaluation_.pdf

Wilkinson, M. (2020, February 13). *Indicators for FAIRness | Scoring · Issue #34 · rd-alliance/FAIR-data-maturity-model-WG*. GitHub. https://github.com/rd-alliance/FAIR-data-maturity-model-WG/issues/34#issuecomment-585689034

Wilkinson, M. (2022a, February 16). *FAIR Evaluation, at EOSC Nordic: FAIRification Step 5 webinar value and limitations of fair assessment tools*. https://www.youtube.com/watch?v=CqUikcVX4Ig

Wilkinson, M. (2022b). *FAIR Evaluation How to test tricky things and How testing guides improvements in FAIRness*. https://www.eosc-nordic.eu/content/uploads/2022/02/Mark-Wilkinson-EOSC-Nordic-FAIR-Assessment-.pdf

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Gonzalez-Beltran, A. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1). https://doi.org/10.1038/sdata.2016.18

Wilkinson, M. D., Sansone, S.-A., Méndez, E., David, R., Dennis, R., Hecker, D., Kleemola, M., Lacagnina, C., Nikiforova, A., & Castro, L. J. (2023). Community-driven governance of FAIRness assessment: an open issue, an open discussion. *Open Research Europe*, *2*, 146–146. https://doi.org/10.12688/openreseurope.15364.2

Wilkinson, M. D., Sansone, S.-A., Schultes, E., Doorn, P., Bonino da Silva Santos, L. O., & Dumontier, M. (2018). A design framework and exemplar metrics for FAIRness. *Scientific Data*, *5*(1). https://doi.org/10.1038/sdata.2018.118

Wilkinson, M., Schultes, E., Bonino, L. O., Sansone, S.-A., Doorn, P., & Dumontier, M. (2018, July 4). *FAIRMetrics/Metrics: FAIR Metrics, Evaluation results, and initial release of automated evaluator code*. Zenodo. https://zenodo.org/record/1305060

Williams, N. (1995). Europe opens institute to deal with gene data deluge. *Science*, *269*(5224), 630–630. https://doi.org/10.1126/science.7624788

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*(1), 121–136. https://www.jstor.org/stable/20024652

Wittenburg, P. (2018, December 12). *View of Keynote 3. FAIR - one pillar towards Convergence. Dr. Peter Wittenburg*. Library.wur.nl. https://library.wur.nl/ojs/index.php/FAIRdata2018/article/view/16308/15766

Wittenburg, P. (2019). *From Persistent Identifiers to Digital Objects to Make Data Science More Efficient*. *1*(1), 6–21. https://doi.org/10.1162/dint_a_00004

Wittenburg, P., & Strawn, G. (2018, February). *Common Patterns in Revolutionary Infrastructures and Data*. https://www.rd-alliance.org/sites/default/files/Common_Patterns_in_Revolutionising_Infrastructures-final.pdf

Wittenburg, P., & Strawn, G. (2021a). Revolutions Take Time. *Information*, *12*(11), 472. https://doi.org/10.3390/info12110472

Wittenburg, P., & Strawn, G. (Eds.). (2021b). *Data Intelligence*, *3*(1). https://direct.mit.edu/dint/issue/3/1

Wonder. (2017). *Please prove the claim that 90% of the world's data has been generated in the past two years, including hard statistics to back it up.* Askwonder.com. https://askwonder.com/research/please-prove-claim-90-world-s-data-generated-past-two-years-including-hard-utp5zijln

Woolgar, S. (1990). Configuring the User: The Case of Usability Trials. *The Sociological Review*, *38*(1_suppl), 58–99. https://doi.org/10.1111/j.1467-954x.1990.tb03349.x

Woolgar, S. (1999). *Do Artefacts Have Ambivalence: Moses' Bridges, Winner's Bridges and other Urban Legends in S&TS - Steve Woolgar, Geoff Cooper, 1999*. Social Studies of Science. https://journals-sagepub-com.uaccess.univie.ac.at/doi/abs/10.1177/030631299029003005

Woolgar, S., & Lezaun, J. (2013). The wrong bin bag: A turn to ontology in science and technology studies? *Social Studies of Science*, *43*(3), 321–340. https://doi.org/10.1177/0306312713488820

World Duchenne. (2019, March 21). *Highlights on Data sharing for Duchenne meeting • World Duchenne*. World Duchenne. https://www.worldduchenne.org/news/highlights-on-data-sharing-for-duchenne-meeting/

World Economic Forum. (2010). *The Lisbon Review 2010*. https://wbc-rti.info/object/document/7341/attach/TheLisbonReview2010.pdf

wwPDB. (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, *47*(Database issue), D520–D528. https://doi.org/10.1093/nar/gky949

Wyatt, S. (2017). Making Policies for Open Data. *Science, Technology, & Human Values*, *42*(2), 320–324. https://doi.org/10.1177/0162243916687645

Wynne, B. (1991). Knowledges in Context. *Science, Technology, & Human Values*, *16*(1), 111–121. https://doi.org/10.1177/016224399101600108

Wynne, B. (1992). Misunderstood misunderstanding: social identities and public uptake of science. *Public Understanding of Science*, *1*(3), 281–304. https://doi.org/10.1088/0963-6625/1/3/004

Wynne, B. (1996). May the sheep safely graze? A reflexive view of the expert-lay knowledge divide. In S. Lash, B. Szerszynski, & B. Wynne (Eds.), *Risk, Environment and Modernity Towards a New Ecology*. Sage Publications.

Wynne, B. (2007). Dazzled by the Mirage of Influence? *Science, Technology, & Human Values*, *32*(4), 491–503. https://doi.org/10.1177/0162243907301086

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., To, J. K., & Laird, M. R. (2017). Ensembl 2018. *Nucleic Acids Research*, *46*(D1), D754–D761. https://doi.org/10.1093/nar/gkx1098

Abstract

Since its formal launch in 2016, the concept of FAIR has become a central element in European research policies – making it mandatory for publicly funded research projects to share their data in a findable, accessible, interoperable, and reusable manner. While these prescriptions seem simple enough, academic practice has not followed suit. This thesis aims to answer why that is, and what happens as FAIR moves from principles to policy to practice.

Through a praxiographical approach to studying things in the making, this thesis follows FAIR, from its genealogy as a concept and a policy, through its enactments in practice, to its envisioned futures. The empirical basis for the study consists of a rich assemblage of documents, interviews, observation and participation.

Throughout, FAIR has demonstrated a persistent flexibility which, on the one hand, allowed it to be nominally appropriated by a wide array of different actors, but, on the other hand, disallowed it from achieving its infrastructural aspirations. The thesis shows how this simultaneously high nominal appropriation and low conceptual appropriation of FAIR produced multiplicity and tensions that could be observed at different sites of articulation – from conferences to courses to everyday research practices.

While multiplicity was routinely pitched as a barrier to overcome and a problem to be solved, this thesis makes the case that articulations of multiplicity should rather be approached as an opportunity to adapt principles and policy in ways that are better aligned with practical epistemic realities.

Abstract - German

Seit seiner offiziellen Einführung im Jahr 2016 ist das FAIR-Konzept zu einem zentralen Element der europäischen Forschungspolitik geworden. Es macht es für öffentlich finanzierte Forschungsprojekte zur Pflicht, ihre Daten auf auffindbare, zugängliche, interoperable und wiederverwendbare Weise zu teilen. Obwohl diese Vorschriften recht einfach erscheinen, ist die akademische Praxis diesem Beispiel nicht gefolgt. Diese Arbeit soll beantworten, warum das so ist und was passiert, wenn FAIR von Prinzipien über Richtlinien zur Praxis übergeht.

Durch einen praxiografischen Ansatz zur Untersuchung von Dingen im Entstehen verfolgt diese Arbeit FAIR von seiner Genealogie als Konzept und Richtlinie über seine Umsetzung in der Praxis bis hin zu seinen geplanten Zukunftsaussichten. Die empirische Grundlage der Studie besteht aus einer reichen Assemblage von Dokumenten, Interviews, Beobachtungen und Beteiligungen.

FAIR zeigte durchweg eine anhaltende Flexibilität, die es einerseits ermöglichte, es nominell von einer Vielzahl unterschiedlicher Akteure anzueignen, es andererseits jedoch daran hinderte, seine infrastrukturellen Ziele zu erreichen. Die Arbeit zeigt, wie diese gleichzeitig hohe nominale und geringe konzeptionelle Aneignung von FAIR Vielfalt und Spannungen erzeugte, die an verschiedenen Orten der Artikulation beobachtet werden konnten – von Konferenzen über Kurse bis hin zur alltäglichen Forschungspraxis.

Während Vielfalt routinemäßig als zu überwindende Barriere und zu lösendes Problem dargestellt wurde, argumentiert diese Arbeit, dass die Artikulation von Vielfalt eher als Gelegenheit betrachtet werden sollte, Prinzipien und Richtlinien auf eine Weise anzupassen, die besser mit praktischen epistemischen Realitäten übereinstimmt.

List of tables and figures

Tables

Figures

Appendix

Pseudonym	Title	Location	Domain	Institution
Maria	Data Steward	In faculty	Life science	University Faculty
Edward	Data Steward	In faculty	Life science	University Faculty
Alex	Data Steward	In faculty	Social science / Humanities	University Faculty
Cody	Data Manager	In faculty	Social science / Humanities	University Faculty
Iben	Data Steward	Off location	General	University Library
Andreas	Librarian	Off location	General	University Library
Paulina	Data Curator	Off location	General	Repository
Victor	Data Manager	Off location	Social science	Repository