MASTERARBEIT | MASTER'S THESIS

Titel | Title

Assembling the Data Steward: Caring for data futures in academic research

verfasst von | submitted by

Anastasia Grace Nesbitt

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

Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 906
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Science-Technology-Society
Betreut von | Supervisor:	Univ.-Prof. Dr. Ulrike Felt

Abstract

In contemporary research contexts, data stewardship has emerged as an approach toward research data management (RDM) that centers openness and the FAIR Guiding Principles (Findability, Accessibility, Interoperability, and Reuse) of research data. This approach is currently being professionalized as a career in which practitioners support researchers through the planning, collection, analysis, archiving, publishing, and reuse of data. As data stewardship is a relatively recent framework and career path that is still actively debated by professional communities and research institutions, it offers a unique glimpse into how data's definitions, practices, material presence, and affective commitments are negotiated. This thesis explores data stewardship as it is practiced in the Austrian context as an infrastructure of care for the long-term futures of research data. Document analysis, qualitative interviews, and ethnographic engagement allow insights into the institutional framings of data stewardship, its entanglement in diverse data infrastructures and registers of value, and its daily practice by professionals. From this empirical work, I identify four modes of care that assemble the practice of data stewardship: the data guide, the data guardian, the data educator, and the data confidant. In these four modes, I highlight how the practices and affective commitments of data stewardship intersect with questions from STS, critical data studies, and information sciences around the visibility of data work, the impact of research funding structures, and the epistemic implications of data reuse. I bring together assemblage theory (De Landa, 2016) and matters of care (Puig de la Bellacasa, 2011; Murphy 2015) to argue that data stewardship is a form of care that contributes to the intertwined assemblage of data as promissory vision and material digital objects. I conclude by proposing that centering modes and infrastructures of care for data creation, archiving, and reuse will strengthen a long-durée approach to academic research data.

Abstract (DE)

In aktuellen Forschungsdiskursen hat sich Data Stewardship als ein Ansatz für das Management von Forschungsdaten (Research Data Management, RDM) herauskristallisiert. Dieser stellt Offenheit und die FAIR-Leitprinzipien (Auffindbarkeit (F für Findability), Zugänglichkeit (A für Accessibility), Interoperabilität (I für Interoperability) und Wiederverwendung (Reuse)) von Forschungsdaten in den Mittelpunkt. Es gibt bereits Bestrebungen, diesen Ansatz als Beruf zu professionalisieren. Data Stewards unterstützen Forscher*innen bei der Planung, Sammlung, Analyse, Archivierung, Veröffentlichung und Wiederverwendung von Daten. Da es sich bei Data Stewardship um einen neuen Beruf handelt, dessen Ausgestaltung noch aktiv in Forschungseinrichtungen diskutiert wird, bietet dieser die Möglichkeit, die Definitionen, Praktiken, die materielle Präsenz und die affektiven Verpflichtungen von Daten zu diskutieren. Die vorliegende Arbeit untersucht Data Stewardship, wie es im österreichischen Kontext als Care-Infrastruktur zur langfristigen Sicherung von Forschungsdaten praktiziert wird. Dokumentenanalyse, qualitative Interviews und ethnografische Erhebungen bilden die Grundlage für die Untersuchung der institutionellen Rahmung von Data Stewardship, ihrer Verflechtung mit verschiedenen Dateninfrastrukturen und Wertregistern sowie ihrer alltäglichen Praxis. Auf der Grundlage dieser empirischen Arbeit identifiziere ich vier Formen von Care im Verständnis der Befragten, die die Praxis von Data Stewardship ausmachen: den Data Guide, den Data Guardian, den Data Educator und den Data Confidant. Anhand dieser vier Kategorien zeige ich, wie sich die Praktiken und affektiven Verpflichtungen des Datenmanagements mit Fragen von STS, der kritischen Datenwissenschaft und der Informationswissenschaft überschneiden. Sie beschäftigen sich mit der Sichtbarkeit von Datenarbeit, den Auswirkungen von Forschungsförderungsstrukturen und den epistemischen Implikationen der Datenwiederverwendung. Ich bringe die Assemblage-Theorie (De Landa, 2016) und Matters of Care (Puig de la Bellacasa, 2011; Murphy, 2015) zusammen, um zu argumentieren, dass Data Stewardship eine Form von Care ist. Assemblagen von Daten werden sowohl als vielversprechende (Zukunfts-)Vision als auch als materielle digitale Objekte herausgearbeitet. Abschließend zeige ich, dass für den langfristigen Umgang mit akademischen Forschungsdaten in der Gestaltung von Infrastrukturen der Datenproduktion, -archivierung und -wiederverwendung Modi von Care eine Schlüsselrolle einnehmen müssen.

Acknowledgements

This thesis is part of the INNORES (Innovation Residues: Modes and Infrastructures of Caring for our Longue-durée Environmental Futures) project with funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation program (Grant Agreement 10105480) under the leadership of PI Univ.-Prof. Dr. Ulrike Felt.

The INNORES project not only afforded me the financial support to complete this thesis, but also gave me a brilliant group of colleagues: Carsten Horn, Kaye Mathies, Noah Münster, Sara Ortega, Ange Pottin, Livia Regen, Dominik Tkalcic, Matteo Vivi and Michaela Zuckerhut. Thank you for the many conversations about this thesis and for your good company. To Ulrike Felt, thank you for bringing this group together, for creating an inspiring framework with which to think, and for your supportive supervision of this thesis at each stage of its evolution.

To my interview partners, many thanks for the generous contribution of your time, experiences, and candor. This project would not have been possible without you.

My thanks to the members of the Department of Science and Technology Studies for making UniWien STS a wonderful place to work and study. A special thanks to Maximilian Fochler for his dedication to teaching STS.

To my fellow TA Office residents, Elaine Goldberg, Harry Wynands, Hanna Braun, Tereza Butková and Dario Feliciangeli: may the NIG elevator always work when you are running late and may your futures be as bright as the seminar room beamer. The TA Office also allowed me to get to know many students, professors, and guest-lecturers in the Science-Technology-Society master's program whose contributions in seminars and in casual conversation built my thinking about STS. Thank you!

Thanks to my family for always supporting my adventure to Austria and STS.

Finally, thank you to Jackie and Emily for making me feel at home in this world.

Table of Contents

1. Introduction

If you asked a researcher – in any academic discipline – what the left-behinds of research might be, I would wager they might not be able to tell you. What is residual, what gets forgotten, what gets used up – these are not the first questions we ask when we begin a research project, no matter the scale.

But what happens when we center these questions? What happens if we, for a moment, put the residues of research first in our inquiries?

In my experience, this switch in perspective opens a whole new world of people, practices, and things through which research moves. In an academic landscape characterized by the proliferation of data and its (their) logics, the reversal of the invariably forward-looking scripts of innovation points us to inquire after what happens to data when innovation is done, when grants are paid out and results are submitted to the preferred journals for publication. Where does the data that underpins so much of our epistemic processes and knowledge production go, and who takes care of it? What do they need to be kept alive, to be made available to others to reuse, interpolate, reform?

In the case of academic research, data are increasingly moved toward repositories, and cared for by data stewards. This might sound like a classical library dynamic, with all kinds of paper records, books, and documents entering a library to be cared for by a librarian. But there is something more to be parsed in the contemporary landscape of research data management, repositories, and data stewardship, even while this newer landscape intersects with the hallowed halls of the analog library, both in ancestry and even, at times, physical location.

From across university departments and units, data flows. It (they) proliferate(s) across projects, between administrators, and outside the bounds of precise quantification. Where there once was a library with concrete walls that limited the number of informational inhabitants living inside, there is now a digital interface upheld by dense material infrastructure that allows unknown volumes of digital objects to live on beyond the termination of the contexts that created them. It is in this landscape of proliferation that we find new technological solutions proposed for data's management, analysis, and storage, as well as a different kind of profession to deal with data.

Data stewardship, as a notion, emerged in the mid-2010s as proponents of the FAIR principles and Open Data practices and infrastructures further developed conceptual tools to make the data landscape of academic research a better place. With the rise of notions of Open science and Open Data, and their accompanying ethos of democracy, transparency, and equity, stewardship began to explicitly or tacitly accompany management as the guiding terminology for dealing with research data. We will explore exactly what data stewardship means to those who practice it in time. For now, we begin with the data steward as a rapidly professionalizing academic worker, whose position and practices help to assemble data as something(s) that can coherently move through the academic world and its vision of the future for data.

1.1 Defining data stewardship toward FAIR data

By now, we have established that this project concerns data stewardship and how it moves through spaces of academic research as a form of care for the futures of data. Before diving into literature in STS and related disciplines that contextualizes it, we need to ask one key question: what is data stewardship to those who theorize and do it? To answer this question, I bring in voices from across open science, research data management (RDM), and data stewardship to understand how data stewardship is framed as a notion and a practice.

We begin with data stewardship as a notion. Although there is no clear documentary trail of how and when data stewardship came to take precedence over, or at least coexist with, research data management, we do know that RDM predates data stewardship in terminologically expressing the project of dealing with research data (Mons, 2018, p. 24). If the wider academic uptake of the data management plan (DMP) in the early 2000s can be taken as a proxy for the emergence of RDM (Smale et al., 2018), then RDM predates data stewardship by roughly a decade. In fact, these two notions continue to share institutional and epistemic space, with long-standing RDM offices often overseeing the professional implementation of data stewardship. In conversation with those who work in this sphere, my interlocutors felt that data stewardship had emerged distinctly in the last decade to represent and act upon data in a longer-term temporality than the status quo encouraged. While data stewardship does have a specific documentary history connected with the FAIR principles, its uptake across institutions and literature is beyond the scope of this research. What is essential to know, however, is that there is significant crossover between these terms in both the literature and in institutional spaces. People who officially work in RDM offices might qualify themselves as data stewards. People working in data repositories might do the same tasks as a professional data steward in another institutional context, but not identify with the title of data steward. Despite this entanglement of roles and practices, we can still understand data stewardship as a notion undergoing specific definition in relation to FAIR and open science conversations.

First formulated in 2015 out of a workshop of information studies professionals linked with the biosciences, FAIR is defined as (F)indable, (A)ccessible, (I)nteroperable, and (R)eusable. Wilkinson et al. (2016) set out to elaborate these principles in the hopes of making the discovery and use of scientific data and other digital objects smoother for both humans and computers. Through the implementation of these principles laid out by Force11, GOFAIR, and the European Open Science Cloud, researchers and institutions aim to break down barriers to data's flow through diverse research contexts. Importantly, Wilkinson et al. (2016) argue that these principles are "domain-independent" and widely applicable (p. 4). However, the application of the FAIR principles has not been without its difficulties. For this reason, Barend Mons, preeminent figure in the FAIR data landscape and an original contributor to the formulation of the FAIR principles, proposed "data stewardship" as the ethos of action that would best foster FAIR data practices. His definition of data stewardship is as follows:

> "Data stewardship is now defined as:
> The process and attitude that makes one deal responsibly with one's own and other people's data throughout and after the initial scientific creation and discovery cycle" (Mons, 2018, p. 36).

Here, data stewardship is defined as a 'process' and an 'attitude' that enables whoever deals with data to deal with it well. This definition still leaves much to the imagination. Who exactly is dealing with data? And what is the cycle of scientific creation and discover? We also might be left pondering what kind of specific relational responsibility is set up here. Are there morals at stake when dealing with data? And if so, who is responsible for establishing them? In Mons' further elaboration, we get a clearer picture of what he means by data stewardship. This time, it is not an ethos but a professional practice that is 'not boring'.

> "These tasks are too complex and time-consuming to leave to researchers (I wrote a whole book on why it would be a grave mistake to train every future scientist to be a fully knowledgeable data steward). Few active researchers see data stewardship as their core business, especially because current incentive systems discourage data sharing and entrench an archaic and almost exclusively narrative-based culture of scholarly communication." (Mons, 2020, p. 491)

> "Therefore, data stewardship, here roughly defined as: treating data and the associated research objects with the utmost care, with the aim to make them reusable for discovery as long as they are valid, is not a boring, unavoidable task for dusty, conservative people, but a rapidly developing profession at the very heart of modern science." (Mons, 2018, p. 24)

To combine these three definitions by Mons into one, data stewardship is both a demeanor towards data and a professional practice that centers the accompaniment of research data through its creation, use, and reuse. Furthermore, it is something that should be laudable and not relegated away as merely an administrative task for the 'dustier' members of the academy. This definition is situated in the emergence of ideas around open science and FAIR data, making data stewardship – both as attitude and professional practice – embedded in a network of institutions, researchers, and associations with research data as its centripetal force.

1.2 Implementing data stewardship while under construction

However, despite this recent expansion in theorizing data stewardship and calls for its concrete implementation by institutions of diverse character, it is still a struggle to precisely define and carry out data stewardship across contexts. In library and information science (LIS) and informatics literature, there is an active debate covering divergent perspectives on the question of implementation. From quantitative literature studies arguing a lack of research on the implementation of data stewardship and RDM (Perrier et al., 2017) to studies of implementation in university curriculum against formal competence standards established by data science communities (Demchenko and Stoy, 2021), scholars in the LIS sphere have yet to establish what data stewardship means across contexts and how to teach or encourage it. These quite strict studies bely a discomfort that the intersection between data stewardship as a concept, notion, or attitude and data stewardship as a concrete profession with (as of yet unfixed) associated competencies has seemingly caused a lack of consensus and unified action. However, there are more qualitative propositions coming from LIS as well.

From bioinformatics, Wendelborn et al. (2023) have attempted to create a comprehensive definition of data stewardship that uses a stakeholder approach against what they characterize as a lack of cohesive understanding. They lay out their potential stakeholders and stakeholder rights, arguing that,

> "Data stewardship is the responsibility to optimise research data management and sharing to maximise the potential value of research data for the scientific community and the public and, at the same time, to respect the rights and legitimate interests of all other stakeholders concerned, such as data subjects, vulnerable groups or communities, data producing researchers, organisations in which the data stem from, research funders, scientific journals, and data repositories." (Wendelborn et al., 2023, p. 4)

While this definition is laudable for its focus on diverse stakeholders and its argument that this diversity should be central to the implementation of data stewardship, this approach does not specify an operational definition of data stewardship that might satisfy the two more skeptical previous papers. For an attempt at this more operational approach, we need to move backward, into the days before FAIR. Parsons et al. (2011), a group of data managers (in their self-description) on the International Polar Year project, emphasize the necessity for viewing data production through an ecosystems lens, with data managers deeply involved from the beginning of the research process. Making the caveat they are "practitioners, not theoreticians", these data managers argue the following toward what they see as a healthy data ecosystem:

> "It is very helpful for data managers to participate in that process of data creation because it helps engender trust and collaboration, but it also makes the boundary object, the data, more robust. We can readily see this with reference and community data because consensus building mechanisms such as science and technical working groups are necessarily included in their creation." (Parsons et al., 2011, p. 563)

So, in the terms of these pre-FAIR data managers – working before conversations shifted to take up data stewardship as the preferred terminology – data stewardship can be defined not only as acknowledging various stakeholders but as doing the work of building robust consensus-based data practices in a context when machine actionability for metadata and repository technologies were far less advanced. In their view, data stewardship is essential for the smooth flow of interdisciplinary data through research projects and spaces of reuse. In their fine-grained descriptions of different data dilemmas and solutions in the IPY project, these data managers suggest that there are indeed concrete practices and competencies that facilitate consolidating and archiving data across disciplines. In a far more recent consideration, Palmer and Cragin (2023) take up the legacy of LIS as a 'metascience' (see also Bates, 1999) as it relates to data stewardship. Data stewardship becomes a key to innovation as research moves toward 'convergence' (similar to complexity science) in which radically different approaches and data types are 'converged' in order to solve society's most pressing issues. In their view, "There is now a need for renewed attention to operationalizing and synergizing data-related metastructures for convergence purposes" (Palmer and Cragin, 2023, p. 123). And in order for data to be best prepared to be utilized in a convergence logic, it must be stewarded through metastructures that enable diverse data to find each other.

So, the FAIR principles kicked off the conversation about data stewardship and it is being defined in the context of a continued push toward science's promise of innovative solutions for wicked problems. But the process of translating these more abstract conceptions of data stewardship into practice requires a different kind of definition. TU Delft, well known for its innovation in LIS, and an inspiration for many in the Austrian data stewardship and RDM community, takes up many of the arguments we have thus seen in the LIS literature. In 2018, the TU Delft Library team published a report of its internal process of structuring data stewardship as a professional presence in the university. As the TU Delft team summarizes their experience with integrating data stewards across all faculties of the university, they use this definition of data stewardship that emphasizes disciplinary expertise:

> "Data stewards are disciplinary experts with knowledge of data management who are employed at faculties in order to advise researchers and faculty members on the various aspects of research data management" (Teperek et al., 2018, p. 142).

Here the TU Delft give an example that counters the somewhat concerned tone of the literature and curriculum studies previously mentioned. The TU Delft finds that disciplinary expertise – that is the previous experience of a data steward in a specific content area of data production – is the key to making data stewardship work across disciplines. Furthermore, the TU Delft embraces diversity in their RDM policies by allowing for distinct policy versions among faculties. In their attention to disciplinary specificity and diversity, the TU Delft approach links well with the

Parsons et al. (2011) and Palmer and Cragin (2023) studies and their cumulative argument toward a technologically structured and practice-oriented vision of data stewardship for innovative research.

1.3 Data stewardship defined in Austrian terms

Thus far, we have seen the process of defining data stewardship wind through a variety of scales and approaches. From its early mentions in the construction of the FAIR Guiding Principles to debates across sub-fields of LIS and related disciplines to its institutional elaboration at the TU Delft, data stewardship crisscrosses many academic and institutional communities. In Austria, the discussion over data stewardship has been primarily coordinated by the FAIR Data Austria project, which supported the implementation of the European Open Science Cloud in Austria from January 2020 to December 2022. It included six major research universities in Austria, as well as other partners. In the report "Data Stewardship – Austrian National Strategy and Alignment," the authors lay out the specific approaches and challenges to data stewardship that characterize the Austrian context. Importantly, they also provide their own definition of data stewardship and data stewards:

> "Data Stewardship encompasses a set of practices aimed at facilitating the transformation of researchers and RDM supporters towards FAIR (Findable, Accessible, Interoperable and Reusable) research data. Simultaneously, Data Stewardship acts as a bridge between stakeholders in research, infrastructure, and management, defining their roles and responsibilities in the transformation." (Bardel et al., 2023, p. 66)

> "Data Stewards are the experts in RDM responsible for implementing Data Stewardship in practice. These dedicated individuals empower researchers, considering the unique cultural context and available resources. To fulfil this role, Data Stewards must possess a range of professional, methodological, social, and personal competencies." (Bardel et al., 2023, p. 66)

Here the Austrian RDM community gives us a few key flavors in definition: its emphasis on 'bridging', the 'transformation of researchers' and the specification of data stewards as responsible for implementing everything that falls within the bounds of data stewardship 'in practice'. The data steward appears here as a coordinating and mediating figure, poised to enable a grand 'transformation' of what researchers can do or be. However, the "range of professional, methodological, social and personal competencies" that define an Austrian data steward remain enigmatic. How such a definition operates in practice remains to be seen.

In addition to providing these definitions, the report points to the three models of professional data stewardship that came out of the Fair Data Austria "Tasks and Profile of a Data Steward" workshop (Hasani-Mavriqi, 2020): generic (1 data steward per university), centralized (up to 3 data stewards per university in a central office), and decentralized (1 data steward per faculty or field). With these models established, the report then summarizes perspectives from across Austrian research universities concerning their choice and implementation of model. The report ends with a call for more support for data stewardship, echoing Mons' framing of data stewardship as essential scientific work in need of funding. The report argues there is still – despite this report itself defining data stewardship – a lack of consensus surrounding the "skills, roles, and responsibilities" of data stewards (Bardel et al., 2023, p. 65). In the summary report of the FAIR Data Austria project (Blumesberger, 2021), this lack of consensus is juxtaposed with the increasing demand for data science and programming skills among data stewards as,

> "Universities and other research institutions are fully aware that research is becoming increasingly data-intensive, interdisciplinary, and collaborative, and are therefore interested in retaining the value inherent in the vast amounts of research data" (Blumesberger et al., 2021, p. 39)

To summarize the situation in Austria based on these foundational documents, there is a movement to harness the value of proliferating data produced by Austrian research universities. In order to foster and protect this value creation, data stewardship must be funded and professionalized while respecting the diverse needs and structures of Austrian universities. The 'consensus building' pointed to by Parsons et al. (2011) becomes 'bridging', with Austrian data stewards responsible for the smooth travel of data through university communities and infrastructures. Data stewardship is still under definitional debate in Austria, with the outcomes of the FAIR Data Austria project providing a baseline for discussion. With its central term still under debate, the implementation of data stewardship is a complex process that relies on negotiation, tacit transfer of knowledge, and flexibility in order to progress. We will see this process further in the empirical chapters. For now, we can leave this brief history of the definition of data stewardship with the following working definition: data stewardship is both a notion that expands on RDM to include relational and communicative realms of data production and a professionalizing practice that centers the movement of data through research institutions toward FAIR reuse.

1.4 Assembling data stewardship in a thesis

This thesis concerns data stewardship as it is taken up in the realm of Austrian academic research. It is informed by the central questions of the INNORESS project, which ask how innovation societies deal with the residues of innovation in the contexts of nuclear waste, microplastics pollution, and data waste. Understanding data stewardship as a modality and infrastructure of care for the futures of research data allows me to place data stewardship in broader discussions of the material and temporal impacts of life with data. Therefore, I explore data stewardship as practice and a profession as it travels through conceptual expressions, university libraries, human-machine relations, funding offices, and more. This one of many possible tellings of the assembly of data stewardship contributes to conversations about a world datafied, and the people, practices, and things that populate that world. To begin, I present five bodies of literature that were essential to my understanding of data stewardship as I came to know it. I then move to the theoretical concepts that sensitized my engagement with data stewardship. This conceptual outline is followed by a description of the methods that guided the empirical work of this project. The empirical chapter begins with its guiding questions and moves through them via four modes of data stewardship. Finally, I discuss cross-cutting findings that wandered through each mode, and the broader conclusions that can be drawn from them. If you, as a reader, find anything in this project, I hope it is an expanded curiosity for the small moments, feelings, and things that make our data worlds what they are.

2. Literature Review

2.1 Introduction: Digitization, datafication, and critical data studies

With this glimpse into data stewardship behind us, it is time to introduce the bigger world in which it is assembled. A bit of a warning: this world is itself assembled from many institutions, interactions, and notions, and the many scholarly treatments of them. Here I will attempt to cut a clear path through the scholarship in critical data studies, STS, and information science that contextualizes data stewardship. As data stewardship is emerging (quite rapidly) as a notion with various definitions, sets of practices and values, and a rising academic profession, it is necessary to traverse the pockets of literature that offer deeper insight into these facets. While this study is firmly situated in STS, data stewardship has not yet received extensive treatment in STS venues. Nevertheless, STS offers essential analyses of data infrastructures and institutional dynamics going back decades. So, this literature review will weave together established strands of STS and related fields to best get an idea of how data stewardship emerged and is developing today.

To begin at the broadest point, this research is situated in the work with digitization and datafication coming from STS, critical data studies, geography, and other disciplines. These terms point toward the move to critically analyze the growth of digital devices and capabilities and the growth of data of various scales, uses, and characteristics across public and private science, the high-tech market, governance, education, and popular culture. Boyd and Crawford (2012) and Dalton and Thatcher (2014) succinctly summarize the central calls of a critical studies of data, big or small. Dalton and Thatcher (2014) call for us to be attendant to the historical specificity, non-neutrality, social embeddedness, constructed nature, ambivalent epistemic value, and possibility for non-hegemonic applications that characterize 'big data', or data regimes more broadly. Boyd and Crawford ask how 'we are our tools' (in Suchman's application of Levi Strauss) in the intellectual uptake of 'Big Data'. They question the definitional changes to knowledge, misleading association with objectivity, dubious utility of volume, risks of decontextualization, ethical vulnerabilities, and inequitable accessibility that Big Data carries into academic work. And indeed, the prompts offered by these scholars in the 2010s were already being taken up in various disciplines and have been significantly expanded in the decade since. While many detours in this mature scholarship could be taken, from its questioning of inbuilt algorithmic racism (Noble, 2018), to surveillance and the gig economy (Newlands, 2021), to gendered digital labor (Gregg and Andrijasevic, 2019), and much more, I will try to limit this first overview of critical data studies literature to what is directly applicable to this project (although I cannot resist recommending a deep dive into any and all of this literature).

For this purpose, I will move to Kitchin (2014; 2021; 2022) and Leonelli (2014; 2020) for their treatments of data as implicated in academic research. Their work allows for this project on data stewardship to exist in a narrower niche of critical data studies literature and, as the previous section might have hinted, this narrowing is a relieving necessity in the world of data. Kitchin's work focuses on what he terms 'Big Data' (defined in earlier 2013 work) as an epistemic shift, especially reacting to the assertion that Big Data are a 'fourth paradigm' in science. Responding to the apparent desire for a re-awakening of scientific positivism or 'new empiricism' through Big Data, Kitchin cautions that "to accompany such a transformation the philosophical underpinnings of data-driven science, with respect to its epistemic tenets, principles and methodology, need to be worked through and debated to provide a robust theoretical framework for the new paradigm" (2014). Kitchin's (2014; 2021; 2022) work does much to reveal how data, especially Big Data, become vulnerable to a kind of epistemic slippage in which all disciplines may come under its potentially positivist logic (as also pointed to by Boyd and Crawford, 2012). In sum, Kitchin suggests we turn a more trained eye at how data operates epistemically in realms beyond just data science.

Here, it might also be useful to distinguish between 'Big Data' and data ('Small Data'). While these terms are still debated, Kitchin (2014) offers a useful definitional summary. In his terms, 'Big Data' is huge in volume, high in velocity, diverse in variety, exhaustive in scope, fine-grained in resolution and uniquely indexical, relational in nature, flexible, extensionable, and scaleable. His definition of 'Small Data', or just 'data', would be a negative one that does not possess these characteristics. While the difference between 'Big' or 'Small Data' is not operationally critical to this project, this distinction is nevertheless important to understanding how data exists in university research contexts. Indeed, the movement to 'Big Data' has influenced epistemic environments beyond just the development of data science programs or the availability of unprecedented amounts of 'social' data from social media platforms. As Kitchin points out, this is perhaps a more tectonic epistemic shift, and one that requires care and political attention. From now on, this project subsumes is discussed and treated by data stewards as 'data' under that term, rather than distinguishing between 'Big' and 'Small'. I have three reasons for this choice: 1) the variety of data produced in Austrian research universities is beyond my capacity to categorize, even along Kitchin's clear terms and 2) 'data' is the term operationalized by universities in their vision documents and guidelines, subsuming 'Big data' into data, and 3) 'data' is the term that was used during my ethnographic work and interviews. Nevertheless, I see work analyzing 'Big data' as an epistemic movement as critically relevant to this project as it digs into the epistemic and organizational consequences of operating within a data-based framework. Whether data is 'Big' or 'small', there is a desire for more of it, and this is the context in which data stewardship moves.

At this juncture, I have a further terminological choice to justify that is crucially linked to Kitchin's identification of a broad epistemic shift related to data. As Kitchin (2021) makes clear, an overwhelming number of definitions, practices, things, and relations exist under the term 'data', in contrast to its classical definition. According to the Mayfield Handbook of Technical and Scientific Writing,

> "*Datum* is singular, meaning 'one piece of information' or 'one numerical result.' *Data* is the plural form of *datum* and should not be used as a singular noun. (Perelman et al., 1997).

This definition might be familiar to you, but I ask that you temporarily suspend it for the purposes of this thesis. As datafication and digitization expand, 'data', as a term, carries more complex connotations than can be fit in this technical definition. Taking up the move by critical data studies and STS to complexify definitions of data, I use 'data' in both the singular and the plural.[1] In the singular, data *is* a notion, ideal, motivating force, or epistemic shift. Data *moves* through spaces of academic research in ways that go beyond epistemic or material boundaries. This use of 'data' might be closer to what critical data studies or STS scholars would term 'datafication'. However, the use of 'data' rather than 'datafication' is more faithful to the language and discourse of data stewardship and better displays the flow between and around these terms in practice. Data, as diverse digital objects supported by networked material infrastructure, also *move* through academic spaces as *their* construction and specificity demand certain forms of care. From this point forward, the singular and plural treatment of 'data' point to these distinct, yet intertwined, definitions.

Taking Kitchin's query into the epistemic implications of Big Data further, Leonelli (2013, 2014) asks if Big Data's quantity really constitutes an epistemic shift or if the cultural status and

[1] Further conceptual reasoning for this choice can be found in Ch. 3, Sensitizing Concepts.

methodological infrastructures built around it actually make more of a difference. Working with the travel of data through de-contextualization, re-contextualization, and (potentially) re-use, Leonelli (2014) argues that "what counts as data in the first place should be defined by the nature of their journeys" (p. 6). On the way to this argument, Leonelli (2014) exposes how critical the work of curation is to data becoming data through its selection, cleaning, and archiving. This curatorial process supports existing epistemic tendencies in the biological sciences but works with data in a way that fits into the expanded prominence that data holds across social arenas (Leonelli, 2013). Leonelli's work is critical to this project in that it takes up Big Data as being constituted by specific epistemic practices and communities, while simultaneously participating in broader scripts of datafication and digitization that nevertheless guide how data are produced and stored in scientific contexts. It is exactly this intersection of epistemic practice and cultural shift, in which data stewardship emerges away from disciplinary curation as a profession, that promises to shore up the dubious possibilities for data re-use. Here we will leave the foundational works of critical data studies and enter reuse as a complicated landscape of promise and possibility.

2.2 Data sharing and reuse

If data stewardship can be understood as, "optimising the management, preservation and sharing of research data in order to maximise its utility as a reusable resource" (Wendelborn et al., 2023, pp. 1-2), then we must take up reuse as a central logic that fuels data stewardship. Reuse has been examined in depth by scholars of informatics and information studies, and their arguments contextualize the emergence of data stewardship as a rapidly evolving profession. Wallis et al. (2013), in their study of the Center of Embedded Networked Sensing (CENS) in the U.S., ask the question "If we share data, will anyone use them?" as their main inquiry. They answer this question with a heavy degree of ambivalence, identifying that the current state of affairs, as of 2013, was quite abysmal when it came to data sharing. Despite there being specific conditions and motivations that might produce data sharing, data reuse still falters in the face of career demands around publishing and the difficulty of fitting one person's data into another's research. Ultimately, Wallis et al. (2013) argue that data sharing remains a culture of gifting to trusted colleagues or collaborators when the demand for reusable data in research is too low to justify making data reusable through rigorous metadata creation and archiving processes.

Christine L. Borgman, also an author of Wallis et al. (2013), has long explored the diverse dynamics that characterize data sharing and reuse, with her work crossing the variable conundrum of research data sharing (2012), legal implications of open data and gray data (2018), 'data life cycle' and reuse models (2019), and the university communities involved in sharing data (Borgman and Bourne, 2022). This last work provides a perspective on data that highlights its communal nature. From Borgman and Bourne (2022), I take their argument that "[b]road-brush policies, whether for data sharing or other scientific practices, rarely accommodate the full diversity of scientific practices in a domain" to be extremely useful when considering how data stewardship and management operate in practice (p. 5). They embrace the idea that managing and sharing data 'takes a village' and that the diversity of communities and activities that contribute to data practices must be taken seriously and made visible. In a further complication of reuse as a simple one-way process between scientists, Staunton et al. (2021) argue that sharing data does not equate to data equity. The reuse of data exists in webs of institutional hierarchy in which research institutions of diverse prestige and geographical location – especially taking geopolitical histories of violence into consideration – experience intense differentials in the resources needed for sharing and the possible benefits or drawbacks if data is made sharable. The 'village' that it takes to share data might have drastically different ideas about what that sharing can and should entail.

If reuse can be understood as a locally negotiated process that requires the collaboration between diverse actors and is often fraught by epistemic and political questions, one might wonder how this complex process is institutionally implemented through growing frameworks of open science and open Data. The phrase "as open as possible, as closed as necessary" crops up often in the conversation around data stewardship, but this adage by no means points to a unified expression of openness. Levin and Leonelli (2017) argue that openness, like reuse, is situated, relational, cultural, and negotiated among diverse scientific communities. Importantly, openness also contains judgements of value in scientific work. Researchers have complicated feelings and practices around making various parts of their work open or not, often based on what they perceive as being the most valuable elements of their research process. 'Dilemmas of openness' arise when researchers variably view attempts at formalizing openness as possible wastes of time, threats to their current or future work, or counter-productive. In order to resolve these 'dilemmas of openness', openness must be matched with institutional incentive structures in which academic funding and assessment might catch up to the widespread institutional calls for openness (Leonelli, 2015).

One further twist in the reuse story is the question if data, big or small, can be reused at all. From scholars in the qualitative sciences comes the critique that the proposition of data reuse is fundamentally incompatible with the epistemic foundations of qualitative disciplines (Parry and Mauthner, 2004; Mauthner and Parry, 2013; Feldman and Shaw, 2019). If one takes seriously the embedded nature of the researcher in social scientific research (Law, 2004), the collective nature of archival research, or any of the theorizing of standpoint theory (Haraway, 1988), then it becomes difficult to sustain an argument that qualitative data must be reusable in order to be valuable. The non-positivist assumptions that underpin much of contemporary social research are incongruent with the separation of evidence, in the form of data, from the conditions in which it was built. As we will soon see in the context of teaching research data management, this rather foundational epistemic difference in the conception of reuse rears its head. But, it is not only qualitative scholars and their epistemic proclivities that stick in the gears of reuse. Thylstrop (2022; Thylstrop et al., 2022) argues that the very nature of quantitative digital data sets used in algorithmic processes as unruly remains that defy removal from machine systems and as fundamentally entangled in their production contexts make the notion of reuse unstable. Similar arguments stand for reproducaibility, not only reuse (Leonelli et al., 2018). In discussions of reproducibility, Leonelly et al. (2018) argue that the quantitative sciences could learn much from the qualitative sciences in terms of reflexively documenting data practices and making transparent the diverse data work that goes into data creation. In sum, the practice of performatively separating and representing data away from the context of data production opens reuse to grave ethico-political interventions.

2.3 From information infrastructures to data infrastructures

Data sharing happens in a wider context of information and data infrastructures, with specific infrastructural dynamics contributing to the professionalization of data stewardship. Star and Ruhleder (1996) start this conversation about information infrastructures with a foundational argument: information infrastructures are mutable hybrids formed from practice and information technology that moves transparently through geographical and professional communities. Studying large scale information system for scientific research, Star and Ruhleder (1996) create a comical image of an infrastructure under development:

> “Trying to develop a large-scale information infrastructure in this climate is metaphorically like building the boat you’re on while designing the navigation system and being in a highly competitive boat race with a constantly shifting finish line” (p. 4).

Here, we are almost 30 years on and this metaphor still resonates. In the definitional debate around data stewardship, the RDM community is in just such a boat race. Technological tools and standards (like FAIR and CARE) around data sharing proliferate while still lagging behind the explosion of data science and computing power. So, it is essential that we attempt to clarify exactly what is being done as data stewards participate in information infrastructures. Bowker and Star (2000) went on to clarify information infrastructure further as they chart the history of the International Classification of Diseases (ICD) as an infrastructure. While arguing that the ICD itself did specific work – not only modulating pre-existing work practices – to establish what kinds of past information should have bearing on a living document, Bowker and Star make clear that information infrastructures do not appear out of nowhere:

> “Infrastructural routines [are] conceptual problems. No knowledge system exists in a vacuum, it must be rendered compatible with other systems. The tricky, behind the scenes work of ensuring backward and sideways compatibility is not only technical work, it challenges the very integrity of any unifying scheme. Such work, however, is itself often classed as ‘mere’ maintenance and deemed unworthy of public, historical pride of place” (Bowker and Star, 2000, p. 108).

This ‘backwards and sideways compatibility’ crops up repeatedly in the work that data stewards do in order to make data stewardship and management a coherent practice that researchers across disciplines can do. And indeed, this work does not receive recognition for what it does to piece together openness, software solutions, privacy, IT deparments, repositories, DMPs, storage hardware, and the many other components that assemble academic research data. This earlier work proposing information infrastructures as embedded in broader infrastructural networks but maintained by specific practices (often thankless) informs the lens I take when looking at data stewardship. Baker and Bowker’s (2007) proposal that data functions in an ecological manner further illuminates the non-linearity, intangibility, and bricolage that go into making data work in information infrastructures. Here, information management occupies an intermediary position that requires diverse kinds of knowledge in order to overcome the tensions – if not chaos – characteristic of information systems. In defining this work, Baker and Bowker argue we lack the language to fully describe the variety of labors that go into making information infrastructures work:

> "We just don't have good ways of talking about the work or measures of its success: the creative aspects of the work of the information manager get buried under its image as purely a service occupation" (Baker and Bowker, 2007, p. 139).

Edwards (2010) responds with a critical piece of language in the concept of data friction. Data friction describes the creation and management of data as effortful, time-intensive, and often contradictory. Edwards et al. (2011) work with metadata to concretize this idea of friction, as they finds the process of creating metadata to be fragmented, divergent, iterative, and deeply local. Edwards et al. (2011) complicate the common definition of metadata as 'data that describe data' that can exist in controlled vocabularies that travel through scientific disciplines to maximize findability. He paints a picture of metadata creation and management as fraught with inconsistencies, miscommunication, and a lack of resources invested in sustainable metadata schemes. When research teams attempt to implement a fixed metadata strategy (in this case without an information manager or data steward), both carrots and sticks are necessary to get the job done – as well as other less black-and-white methods of 'lubrication'. Importantly, Bates (2017) expands the notion of 'data friction' by arguing it is a space of politics. Not only do moments of data friction complicate the epistemic process of data sharing and reuse, but they "can be understood as an important constitutive force in the development of social relations" (Bates, 2017, p. 413). In this thesis, 'data friction' refers to this extended definition that encompasses not only the unruliness of data travel but also its political nature. In a further step on politics and metadata, Mayernick (2019) argues that the choice to differentiate metadata from data is revelatory of what constitutes evidence in science and where accountability for research processes and products can be located. In Mayernick's (2019) perspective, this is also a political distinction in that it requires deeply tacit and embedded knowledge of data production process to make descriptions of data meaningful. In this framing, the utility of metadata as a form of communication that can cross boundaries across disciplines, or even projects, is dubious.

Thus far, our discussion of infrastructures has taken us from information to data to metadata. On each step of the journey, we find that building infrastructures presents tensions and friction. The epistemic process of producing, describing, archiving, and communicating data is a project of immense proportion, with many opportunities for things to go awry. And things to indeed go awry. However, Murillo (2023) would argue that this is not the result of mere happenstance, but the consequence of power relations in-built in information infrastructures. Here, the stewardship of data is a justice issue. In Murillo's analysis, the openness of data by no means guarantees equity or transparency in how data are produced or distributed. The CARE Principles for Indigenous Data Governance (2019) have opened the conversation around provenance and sovereignty, calling for data from indigenous groups to be characterized by (C)ollective Benefit, (A)uthority to Control, (R)responsibility, and (E)thics.[2] But, principles alone do not an infrastructure make. Writing from the perspective of both an anthropologist and a data technologist working with local communities in a climate-oriented NSF project in Alaska, Murillo 'studies-while-caring' for this environmental data. But, during this work, "the 'infrastructual blues' strikes when the possibilities of mutual-aid are foreclosed, when collective data work is deemed 'superfluous or inefficient,' when data infrastructures only serve the machine-automated use of human beings for the purposes of data extraction" (Murillo, 2023, p. 12). Even when data and the software and repositories that facilitate data travel are 'open' from a technological and even ideological point of view, there still remains an uneven distribution of power in the joinery that scaffolds data production, description, archiving, and distribution together.

2.4 Projectification and assetization in the data university

The uneven power distribution pointed to by Murillo (2023) does not only exist in the context of working specifically with data from marginalized communities, but also can be found in the broader academic conditions in which data stewards find themselves embedded. From the early work by Hackett (1990) and Slaughter and Leslie (1997) on academic capitalism, we can understand the university – even the public one – as an institution un-free from the infiltration of market logics. Both research activities and research structures reflect the entrance of market-orientations that make the attraction of private investment through grants and endowments, industry partnerships, tuition fees, and short-term contracts familiar characteristics in the contemporary university. In Hackett's (2014) more recent estimation, this makes the university a place with dubious power as an independent 'moral force' in society. Fochler (2016) makes the essential jump from academic capitalism to epistemic capitalism, arguing that the influence of capital in contemporary academic life and structures can not only be seen in the strategic prioritization of financial value, but in the absorption of dynamics of accumulation into the cultural norms that shape university knowledge-making. In this comparative study, the impact of accumulation patterns on biological research is visible both in the private sector and in public

[2] For further work on Indigenous theorizing of data and data governance, see Carroll et al. (2019), Raine et al. (2019), and the International Indigenous Data Sovereignty Interest Group (IG) in the Research Data Alliance.

university research, with regimes of accumulation appearing stronger in the public research environments, even to the detriment of epistemic efforts:

> "Academic postdocs, however, felt compelled to subordinate all other forms of worth relevant to them to the overarching aim to be as productive as possible in accumulating career-relevant capital, with strong epistemic and biographical consequences." (Fochler, 2016, p. 943)

Across the phases of a research career, researchers are concerned with accumulating the right kinds of publications, contracts, and projects that will sustain them into the upcoming step in the temporal trajectory. The particular 'timescapes' that inform the 'regimes' of accumulation to researchers unconsciously or consciously respond are shortening. Dollinger (2020) draws on 'projectification' (Maylor et al., 2006; Midler, 1995) to argue that contemporary academic research is increasingly characterized by the atomization of processes and practices into discrete parts that can be best organized along short-term project timelines. Midler's (1995) organizational analysis of Renault's project structure in car production identified a key conflict between continuous participation in fixed-term projects and the fostering of "permanent structures and processes," such as career longevity (p. 14). Likewise, Dollinger (2020) highlights that projectification poses problems for teaching, curriculum design, the building and maintenance of academic networks, and the quality of research. In Dollinger's (2020) view, some of the most important elements of academic research, such as meaningful collaboration and mentorship, resist the process of atomization that project timelines demand. When atomized components of academic life and epistemic practices become linked with formal assessment measures, the elements pointed to by Dollinger (2020) might further fade in favor of behaviors, practices, and affects that better fit into quantified indicators, despite critiques of their 'chronopolitics' (Felt, 2017). As we will soon see, projectification's shortened timescapes impact how data move through the university, and how data problems arrive at the data steward's desk.

If individual researchers now exist in an environment in which university rankings and prestige, external funding, and fixed-term projects influence epistemic choices and structures, then it is important to know exactly how this happens. Fochler et al. (2016) investigate the regimes of valuation that PhD students in the life sciences draw on when they describe their current work and future prospects for an academic career. In a setting characterized by hyper-competition, the transition from the doctoral phase to the post-doc phase brings with it a narrower array of regimes of valuation, with collectivity and academic sociability losing ground to the pressures of individualized achievement towards tenure. This transition implies the increasing amount of time, effort, and emotional space that must be sacrificed in order to achieve security through competition for funding, permanent positions, and prestige. Sigl et al. (2020) further argue that the linear model of innovation still stands when it comes to conceiving of an academic life and trajectory, and formal metrics for excellence still maintain pride of place in determining upward academic mobility. As already referenced in the literature considering data sharing and reuse, contemporary academic reward structures currently compete with the integration of 'societal responsibility' (Sigl et al., 2020) and, I would argue, communal behavior around data. Literature further specifying assetization in research allows us to understand not only the constraints placed on the production and treatment of research data through narrowing regimes of valuation and responsibility, but also reveals how data itself can become a traveling asset.

Birch and Muniesa (2020) propose assetization as the process by which previously non-economic entities become converted into assets that can be the objects of investment and return. Buying and selling of assets still occurs in the process of capital accumulation, but what is more important in technoscientific capitalism, Birch and Muniesa (2020) argue, is that value can be extracted through control and ownership, which facilitate the manipulation of access. In the case of data, we can look to Berend Mons' (2020) framing of data stewardship as offering "excellent returns on investment" (p. 491) by stretching the euros and dollars invested in public research to their maximum potential in data reuse. While Mons does not explicitly connect this kind of return on investment with a profit logic, his portrayal of data stewardship as the key to shoring up data and investment losses quickly turns research data into an asset to be forever connected to its producers and reused in perpetuum. Reuse is not the same as direct financial revenue, but in the logical environment of FAIR and Open science, it is certainly a kind of currency. And various kinds of currency are needed to propel an academic career. Falkenberg and Fochler (2023) investigate assetization as it operates to valorize innovative technologies. The use of these innovative technologies buffers the evaluation of research and promises increased productivity key to career mobility in pressurized academic settings. Although specifically treating innovative technologies in life science research, Falkenberg and Fochler's (2023) following argument is also directly applicable to how data operates in assetized regimes of valuation around academic research:

> "For the university, this set of technologies takes the form of an asset that promises future revenues not only because it can allow research groups to successfully acquire third-

> party funding, but also because it raises the university's prestige as part of a broader digitalization agenda" (Falkenberg and Fochler, 2024, p. 11).

Not only innovative digital technologies, but data practices and data sets that travel across research contexts can become a revenue stream. This revenue might take a literal form in financial profit (for example, through industry partnerships) or figuratively as it contributes to individual researchers' and universities' prestige, which in turn attracts financial revenue. In these ways, assetization is an integral operation in contemporary academic research and illuminates some of the boundaries and motivations around academic data production.

From this literature establishing the university as an institution in which public good can no longer be understood as the sole stimulus for activity, we can draw three key conclusions that contextualize data stewardship. Firstly, the conditions under which academic research data is produced are often characterized by the perception of stress and pressure on behalf of the producing researcher. Second, we can understand the valuation of data sharing and reuse as conflictual. On the one hand, reuse is a moral impetus and an efficiency booster, according to FAIR and Open science discourse. On the other hand, it is a waste of precious time and potential prestige, both of which are in scarce supply in a hyper-competitive research environment. We will see how this conflict appears for data stewards across the empirical chapter. Finally, we can understand data as itself(themselves) a possible asset to be harnessed for direct financial gain or indirect financial gain via prestige. From academic research data contextualized by academic capitalism, we will move to the kind of specific practices of care and labor that produce research data.

2.5 Data work and data care

If neoliberal approaches to assetization characterize contemporary university operations and surround the work of data stewards, is that only factor at play when looking at data stewardship's process of professionalization? Indeed not. Rather, there are specific dynamics of work and care identifiable in many data-heavy contexts that serve to deepen our understanding of data stewardship before diving into how data stewards themselves understand and portray their work.

Starting off with a bang, Plantin (2021) argues along an autonomous Marxist line that data processing is a form of factory labor following Marx's notion of alienation. In the context of participant ethnography in a U.S. data archive, Plantin observes how the data processing environment does not allow for meaningful transferrable skill development or the encouragement of a deeper relationship with the data processors work with. Nevertheless, processors find ways to subvert the boundaries of their work environment through micro-resistant choices and a bricolage approach to data. They take time for personal matters on the job, or dive deep into a data set and explore it for interest's sake. In highlighting these micro-practices, Plantin (2021) argues for the coexistence of immaterial labor (such as data processing) and factory labor. Importantly, despite forms of micro-resistance, data processors hit a wall in that they will never become fully formed archivists with the cache and professional trajectory it brings. Gabrielsen (2023) builds on Leonelli's (2014, 2016) work on biocuration to argue along similar lines that female biocurators too hit professional walls due to the gendered nature of biocuration work. While necessary to fuel the turn toward bioinformatics and computational biology, the work of biocuration that provides essential data sets is not acknowledged as 'real' scientific work. In place of connotations of innovation and impact that characterize contemporary biology, service, detail-orientation, and communication – all gendered female in Gabrielsen's (2023) understanding – accompany biocuration. This has critical impacts on the career paths of female biologists. On the one hand, the gendering of biocuration creates a cycle that attracts female-socialized persons who might be able to fulfill these gendered expectations or might want the flexibility that the field offers in order to fulfill other biographical desires, like family formation. On the other, it limits career paths as a return to research after moving toward biocuration is an uphill battle. Gabrielsen (2023) sums up the gendered nature of data work succinctly, stating:

> "When technology is granted epistemic agency through envisioned machine actionability, the work that goes into facilitating the machines is further removed from scientific recognition. The effect is that biocuration is positioned as service work for machines and biocurators become parts of an invisible gendered infrastructure enabling computational biology" (p. 19).

Grugulis and Vincent (2009) lend a key perspective on the 'soft skills' that go into the 'invisible gendered infrastructure' proposed by Gabrielsen (2023) in their study of the privatization of public benefits casework in the U.K. They zoom in on what these 'soft skills' entail and what they do to labor relations. With the entrance of private contracting companies into public services, the authors identify a move toward soft skills follows a move toward target compliance rather than experience in management. Grugulis and Vincent (2009) argue that these soft skills disadvantage those in positions that place a high value on soft skills by entrenching gendered and racialized assumptions related to them, individualizing responsibility through a ‘customer relations’ approach that ignores relational dynamics, and preventing the development of technical skills that can travel more easily to other positions or organizations. This focus on ‘soft skills’ centered around communication and collaboration pops up repeatedly along the professionalizing trajectory of data stewardship, as we will soon see.

But it is not only ‘soft skills’ that are desired in the data-focused arenas. In the case of data science as a rapidly formalizing academic and professional discipline, Dorschel (2021) finds that data scientists are expected to be exceptionally hybrid beings that can reconcile conflicting data realities through their simultaneous identities as “generalists and specialists; technicians and communicators; data exploiters and data ethicists” (p. 2). Occupying this ‘in-between’ identity allows for data science to be optimally situated to ‘mine’ and predict new niches for capitalist expansion. Generalized competencies matched with deep domain knowledge make data scientists valuable for market decision-making, just as data stewards are similarly valuable for facilitating the coherent assembly of data for the university. But, the work needed to ‘mine’ data for its maximum value, in any domain, comes with complications. Green et al. (2023) collect findings from a remarkable body of literature concerning Danish health data (Wadmann and Hoeyer 2018; Holt 2020; Hillerdal and Svendsen 2022; Hoeyer 2023). The authors argue that when repurposing enters the scene as a key motivation for data collection, unplanned demands on data work and data workers ensue. In the Danish health system, the incorporation of private sector actors as sponsors for clinical trials in exchange for access to remarkably detailed and complete health datasets pushed data work away from clinical relevance and toward external parameters for repurposing in research. This dynamic created time and resource demands on both clinicians and patients that cause the authors to argue that data work is both political and ethical. The cost of smoothing out data friction (Edwards, 2010; Bates, 2017) through intensified data work is something that must be acknowledged and negotiated actively by “mov[ing] close to the actual work practices and articulat[ing] the dilemmas at hand” (Green et al., 2023, p. 129).

With these considerations of data work as bounded by demands for ‘soft skills’, as gendered, as relational, and as implicated in profit logics, we can now move to literature that explicitly treats data work through the lens of care. These studies (Pinel et al. 2020; Choroszewicz 2022) enable us to better understand how the boundaries of data work come to function and be contested in practice. Pinel et al. (2020) examine how caring for data creates value in the context of genomics research. In the Twinomics lab, everyone from PhD students to clinical nurses to P.I.’s cares for the lab's twin data, but in distinct ways that follow existing academic hierarchies. Importantly, it is the early career researchers' (both PhDs and postdocs) care through 'donating' their time to collecting twin data, connecting personally with twin study subjects, and developing knowledge of 'how the data like to be treated' that enables the Twinomics database to expand. Pinel et al. (2020) argue, "Such affective and attentive engagements with data, which are currently not valued and relegated to the domains of junior staff, are in fact value generators for laboratories like Twinomics, and should be recognized as such" (p. 193). Choroszewicz (2022) also proposes recognition for the emotional labor undertaken by data workers across three phases: the production, processing, and analysis of data. In the Finnish healthcare system, data workers invested a great deal of attention and skill into making data fit for repurpose, making their data practices legible for management, and making results interpretable for users. Much frustration, excitement, and exhaustion from the "Knowledge Team" went into the process of interpreting the language of healthcare into data products and systems that produced results for management and users. Choroszewicz (2022) proposes that we answer to "the work involved in seeking to cope with uncertainty and the messiness inherent in everyday tasks concerning data work and the construction of data technologies, specifically inducing anxiety, stress and frustration" by recognizing data as process rather than product (p. 9). Taken together, these studies provide a key baseline of understanding data work as data care in which emotional states and affective commitments come to constitute data world-building.

2.6 Conclusion

From the early calls for a critical data studies and attention to 'the boring things' in invisible information infrastructures to targeted studies of specific data practices in biology, healthcare, and environmental research, this review has attempted to pull pockets of scholarly work around data's definition, reuse, reproduction, conditions of production, and ethico-political implication together as a kind of map onto which we can begin to read data stewardship. As there is no distinct body of secondary literature taking up data stewardship as such, this is necessarily a patchwork map of my own creation, constructed from concepts and empirical cases that guided my understanding of data stewardship. These studies locate the human effort in making data practices and systems into infrastructures. That is, rendering them invisible. But, as this literature has shown, diverse, relational, and contested work goes into pushing data infrastructures to the backs of our minds. There are stakes in this process as well, with marginalized communities suffering from the loss of provenance over their data and various kinds of data workers laboring in gendered, racialized, or restrictive conditions. With an attention to localized data practices, what emerges from this literature is a picture of working with academic data as a process of negotiating friction across funding bodies, computers, researchers, software, repositories, community stakeholders, and data workers, and the many iterations of entities and materialities in between.

To cut a cross section of such an entanglement coalesced by data would produce an untold number of case studies, practices, and affects from which to glean many conclusions. However, as this literature also hints, such cross-sections might not allow us to come any closer to an operational definition of data itself. In fact, previous scholarship and my own work were confounded by just this: a definition of data. Not only does the literature around data practices and infrastructures grapple with the definitional question – if not explicitly, then implicitly – but the community around data stewardship does too. As the question of defining data amidst its localized and practiced enactments haunted my empirical work, I looked to assemblage and care for directional guidance. In the next chapter, I will work with these notions to better understand how the professionalizing process of data stewardship and the question of defining data co-produce each other.

3. Sensitizing concepts

3.1 Dealing with the multiplicity of data

There can be no exploration of what data stewardship is without an exploration of what data is. Those familiar with critical data studies or studies of data situated in STS will know that this is a heavy direction indeed, but an infinitely fascinating one. To attempt to define data might pull you across time, discipline, sub-field, mental state. Thankfully critical scholars of data have already paved the way for thinking through data even as it evades ontological grasp. In particular, I found the definitions by Kitchin and Lautiault (2018) and Leonelli (2015) to be the most relieving as I muddled my way through. Their definitions of data are as follows:

> "Data then are situated, contingent, relational, and framed and are used contextually to try and achieve certain aims and goals" (Kitchin and Lauriault, 2018, p. 6).

> "Data thus consist of a specific way of expressing and presenting information, which is produced and/or incorporated in research practices so as to be available as a source of evidence, and whose behavior and scientific significance depend on the context in which it is used. In this view, data do not have truth-value in and of themselves, nor can they be seen as straightforward representations of given phenomena. Rather, data are essentially fungible objects, which are defined by their portability and their prospective usefulness as evidence" (Leonelli, 2015, p. 2).

Here, both Kitchin and Lauriault (2018) and Leonelli (2015) provide definitions of data that point to its multiplicity. Data cannot be defined by discipline, software, granularity, or quantification. Rather, these scholars argue that data must be detected in its relational enactment as 'a source of evidence' that is employed to 'achieve certain aims and goals'. Rather than a fixed category of information or bits of information, data become(s) data through the relations that produce it and through the negotiations data underpins. Leonelli (2015) goes so far as to argue that "There is no such thing as data in and of themselves, as what counts as data is always relative to a given inquiry…" (p. 10). With this argument, we are welcomed into the wonderful world of the multiplicity of data. Data are indeed discursively formed and contextualized. How we discuss and debate data takes place in different spatial and temporal scales. However, data are not merely discursive objects, but inhabit multiple material, cognitive, and representational forms. Data can be stored in a data center or a USB stick, or both. They can be PDFs, CSVs, docxs, MP3s, HTMLs, WAVs, to not even scratch the surface of the world of file types. Data can be analog too!

Paper archives still exist, as do their digital twins as archives are digitized. Social media platforms and streaming services have proliferated the uploading and downloading of untold amounts of data for personal use, which are then transformed into research data through scraping and analysis.

Are all of these same? Can we speak about them in the same breath?

In order to attempt this discussion, I must dive a bit deeper into the discussion of assemblage as a way through ontological discussions, beyond data. If data is defined as multiple, contextual, and relational evidentiary proofs, then can they be further defined ontologically?

3.2 Assembling data and data stewardship

So, we move forward with a groundwork of data as multiple, contingent, and relational. But, this admission is only the starting point. Data is (are) spoken about simultaneously as a monolithic concept/object and as infinite material pluralities that pour in and out of our personal devices, private sector profit endeavors, and academic research projects. Even while speaking of multiple data, these multiple 'datas' confound classification and stratification – especially in the story of data stewardship. Over the course of conversing with data stewards, being taught research data management, and interacting with the guiding documents of this world, it became clear to me that there are co-existing worlds in which data is(are) moving. Here I turn to Manuel de Landa (2016) to argue that data, data stewards, and data stewardship exist in assembled relations. In this section, I hope to clarify my understanding of assemblage and the relief it may bring to the overwhelmed soul tasked with thinking about data.

It may be useful to begin by discussing the ways in which the term assemblage has already been explicitly applied to data. Here is the first:

> "Kitchin defines a data assemblage as a complex socio-technical system that is composed of many apparatuses and elements that are thoroughly entwined and whose central concern is the production, management, analysis, and translation of data and derived information products for commercial, governmental, administrative, bureaucratic, or other purposes" (Kitchin and Lauriault, 2018, p. 8).

This definition of data assemblage as a socio-technical system points to the many arenas of social life that are fundamentally entangled with data practices and processes. From governance to art, Kitchin would argue that there is a ‘central concern’, a motivational thrust, towards the production and use of data not only as an evidentiary proof, but as a guiding source. In this definition, a data assemblage is ‘complex’ and ‘composed of many apparatuses.’ So, beyond being relational, contingent, mutable, and fungible, they are also composed of distinct ingredients. Here we enter very interesting territory. To my knowledge, Kitchin does not cite De Landa in his use of the term assemblage (Kitchin and Lauriault, 2018; Kitchin 2014). Nevertheless, we enter the territory of data as assembled from constituent parts.

In an opposite direction, Boyd (2022) applies the notion of assemblage in a very technical sense to “data cleaning, curation and other data practices involving data repositories” (p. 1342). Boyd takes the 2019 US Census dataset and argues that it expresses eight characteristics of an assemblage. Through this nimble analysis of one dataset, Boyd (2022) shows how assemblage notions applied to a seemingly single digital tabular data object in the publicly available US census data CSV file can reveal data as composite, historically specific, and scalar. However, it is not only the presence of these components but also the “autonomous, machine-like or algorithmic interaction of their components” that makes this data object an assemblage (p. 1342). While working with this specific assemblage of the 2019 US census dataset, Boyd points toward the significance of assemblage thinking for highlighting the various kinds of labor that assemble such a dataset and points to the need for further research that locates and describes particular data assemblages. However, in attempting to visually and textually model the “dataset’s bones, sinews and organs,” Boyd undercuts some of the best conclusions of the application of assemblage thinking to data. With only passing mentions of how data workers’ practices assemble the dataset and sweeping allusions to the use of the US census, De Landa’s key notion of assemblages as themselves sets of nested assemblages is lost. And this is an essential notion.

Thus far assemblage has been applied to data in a more Foucaultian sense by Kitchin (2018) and in a narrow De Landian sense by Boyd (2022). The former emphasizes the broad socio-technical, power-laden entanglement of data with other, seemingly distinct, social entities. The latter zooms in on a single dataset to illustrate its properties as an assemblage but misses the relational movements that put it in circulation. For the purposes of this project, the distance between these applications of assemblage to data must be closed. This can be done without exiting the corpus of assemblage theory to look for a conceptual bridging mechanism, but rather by highlighting existing conceptual tools. To this end, we will now move back to De Landa and his interpretation of Deleuze and Guattari’s (1980) assemblages.

In De Landa's terms, an assemblage is "a nested set of individual emergent wholes operating at different scales" (2016, p. 16). In his project to make the effusive theoretical language of Deleuze and Guattari more applicable and, in his estimation, to extend their line of thought, De Landa builds the anatomy of the assemblage as it can be identified moving through social life. De Landa takes up Deleuze and Guattari's arguments of distributed, componential ontology, their 'rhizomatic' thinking, to explore how distinct entities can be put together or taken apart. His is a flat ontology, with assemblage characterizing entities across scales of size and complexity, from individual military battles and linguistic dialects to governments and nation-states. While being flat, De Landa's ontology is nevertheless realist. He strongly asserts that,

> "Much of the academic left today has become prey to the double danger of politically targeting reified generalities (Power, Resistance, Capital, Labour) while at the same time abandoning realism" (De Landa, 2016, p. 48).

While I do not propose that this project holds tightly to a realist perspective, it nevertheless benefits from De Landa's assertion that 'reified generalities' result in stunted political thinking. And I propose that 'data' – as it is used in the singular – should be added to De Landa's list of generalities waiting for dissolution. 'Data' in its broadest use becomes an incoherent term with no boundaries and no chinks in its armor. It becomes invulnerable to attack. Here De Landa's assemblage thinking is useful for me to break down 'data' (singular), as a sweeping cultural or ideological shift, and multiple 'data' (plural), as the digital objects and epistemic inputs/outputs of academic research, into nested sets of assemblages that interact, but nevertheless maintain compositional distinction and historical specificity. From this point forward in this thesis, the singular or plural usage of the term 'data' reflects this understanding. This conceptual framework is crucial in a realm where seemingly infinite discursive propositions, materialities, and actions exist under one term. To situate this research, and not be lost in the data deluge, I rely on a few key concepts in assemblage thinking to carve out and distinguish how data (singular) as broad cultural ideal, data (plural) as diverse material digital objects, academic research data, data stewardship, data stewards, RDM, and researchers exist in parallel and inform each other. Here, I will take a moment to define these terms and their utility in my project:

1) Assemblage

Following De Landa's reading of Deleuze and Guatarri (1980), I understand assemblage as a way to speak about things that stick together as 'wholes' in conversation, policy, and formal education (among other arenas), but that nevertheless unfold into multiplicity on further inspection. One thing is never just one thing. Assemblages are historically contingent, compositionally heterogenous, exist in nested arrangements, and display emergent relational properties. In the case of data stewardship, multiple data assemblages feed into the practice of data stewardship so that this term carries meaning. Likewise, data stewardship, as a set of contingent, relational practices, can contribute to the 'sticking together' of 'Data' as an object upon which universities can act.

2) Coding/decoding

Coding/decoding refers to the "special expressive components in an assemblage in fixing the identity of a whole" (De Landa, 2016, p. 22). De Landa describes a constitution document that defines an organization as a prime example of a code. Genes and languages also serve as examples of codes in that they limit the form and inclusion/exclusion of components of a given assemblage. In social assemblages, codes might be otherwise described as the registers (Heuts and Mol, 2013) or regimes of valuation (Fochler et al., 2016) to which members of a group respond in their decision-making or actions. Repertoires or languages of valuation can exist in conflict with one another and are negotiated relationally and locally. That is to say, given a set of circumstances, an actor may impact the process of assemblage by drawing on one register of valuation over the other. In this way, codes are not simply deterministic cultural scripts that facilitate assemblage but exist through active negotiation. To make coding/decoding more concrete, here is an example. In the opening of the report on the first year of the European Open Science Cloud (EOSC), EU Commissioner Carlos Moedas wrote the following in his forwarding message:

> "Open science is indeed one of my priorities. It is a move towards better science, to get more value out of our investment in science and to make research more reproducible and transparent" (Directorate-General for Research and Innovation, 2016, p. 4).

In this example of an 'expressive component', Moedas emphasizes open science (to which accessibility and reusability of data are key) as a universal good, and as a way to 'get more value' out of the investment made in science. This science will produce data which can and should accessible via EOSC's networked projects. Such statements and documents can be considered codes that express registers of value which then influence the assembly of data stewardship as a practice taken up by research institutions. These *codes* also serve as professional visions by which a data steward might orient their individual practice of data stewardship.

3) Territorialization/deterritorialization

The process of territorialization and deterritorialization refers not only to spatial boundary-making, but the process through which an assemblage becomes more or less homogenous. Through territorialization, a ‘thing’ becomes more ‘thingy’. It feels more concrete. Through deterritorialization, that thing gets fuzzier, more difficult to name. This concept creates an attunement to how data, as concept and material object, relates to data stewardship. To illustrate with an anecdote from my empirical work, let’s say a researcher in geographical area studies uses PDF scans of an ancient text made available by another university’s library, annotates them, and uses these annotations to support a publication. Is this ‘their’ research data? Should it be archived in a disciplinary repository to be reused by other researchers? How can it be *re*used at all if it has already been used by generations of scholars? In this case, this particular assemblage of academic research data (ancient object, digital PDF representation, annotation software, library archive) has a *deterritorializing* influence on the assemblage of research data management (RDM) as a set of homogenous practices that can be applied to all kinds of data. In this case, RDM becomes fuzzier.

In sum, assemblage thinking allows me to navigate the many layers of meaning-making and material practices that create the landscape of data stewardship without giving way to an overwhelming, indistinguishable singularity. Data as encompassing digital objects, ideas, and overarching institutional movements (to name just three components) is indeed a vast landscape to enter. Conceptual tools that allow for the parsing of the distinct components (and there are distinct components to be found!) enable us to find spaces of intervention. Although De Landa’s reading of Deleuze and Guattari is but one interpretation of that body of thought, I nevertheless find the clear elaboration of the conceptual tools originally presented by Deleuze and Guattari helpful in making visible the heterogeneity and relationality of data as comprising of and constituting social arenas.[3] When this is visible, we can more actively move toward intervening in these arenas toward creating more livable data worlds.

3.3 Caring for data and data stewards

[3] De Landa’s interpretation is not free of critique. See Buchanon (2015) and Bell (2018) for this discussion. However, these critiques do not fundamentally undermine the utility of De Landa’s concepts for this project.

Finally, before we begin to assemble the data steward, I must introduce matters of care as the last guiding sensitivity of this project. By now well-known in STS literature, Puig de la Bellacasa's (2011, 2014) notion of care has traveled far and wide (Martin et al., 2015). Not only has it and its feminist predecessors traveled through many epistemic communities, but it has perhaps moved beyond definition. As Martin et al. (2015) introduce, "… there can be no singular vision of what care is or what it might become" (p. 634). On its face, this is a frustrating conceptual proposition. Although the "open field of potentialities" that encompasses a sensitivity to care is indeed vast, matters of care, nevertheless, provides a concrete, graspable pulse in this work. Its graspability, for me, relies on Puig de la Bellacasa's original formulation:

> "We can think on the difference between affirming: 'I am concerned' and 'I care'. The first denotes worry and thoughtfulness about an issue as well as the fact of belonging to those 'affected' by it; the second adds a strong sense of attachment and commitment to something. Moreover, the quality of care is more easily turned into a verb: *to care*. One can make oneself concerned, but 'to care' more strongly directs us to a notion of material doing" (Puig de la Bellacasa, 2011, p. 90).

Material doing, attachment, and commitment. These are elements of caring practice that I detect in the small corner of data stewardship that this project inhabits. But, I argue, that care can be detected at different scales. On the one hand, data stewards care for data as it refers to the material digital objects that both plague and animate researchers, and the relations and practices that constitute those material digital objects (no matter how difficult it may be to define them). As seen in the literature around data curation, the care that enables movement of material digital objects is still not widely recognized or rewarded. Here again, Puig de la Bellacasa calls for attention to care in this sphere:

> "Potentially, matters of care can be found in every context; exhibiting them is valuable especially when caring seems to be out of place, superfluous or simply absent" (Puig de la Bellacasa, 2011, p. 93).

As this project seeks to show, research data exist through caring relations. While an argument could be made around data stewardship as care work, in a more literal sense of reproductive labor, I move with Puig de la Bellacasa's proposition that material doing, attachment, and commitment enliven a discussion of care. She argues with a subtlety that avoids a patronizing valorization of care labor while still linking the necessary acknowledgement of material conditions of power with the diverse kinds of practices that uphold familiar assemblages. This is where the other hand of data comes in. In addition to the more literally caring for material digital objects, data stewards are situated in an academic landscape that sees data more broadly as means of value accumulation and a guiding force. This 'Data' in the uppercase can be thought of in terms of Kitchin's epistemic shift of 'Big Data' and its promissory cultural narratives. Data stewards take up care here too as they work with the relations around data that allow Data to be the influential organizing force that it is.

As most elaborated in the data confidant chapter of the empirical section of this thesis (but present throughout), data stewards care in a relational sense that is not visible through the framing documents and public representation of data stewardship. Caring for the affective entanglements that exist between researchers and their data, caring for the longevity of archival objects, caring for the material conditions under which researchers work do not find themselves advertised in the formal characteristics or responsibilities of data stewards. These kinds of care indeed seem "out of place, superfluous or simply absent" (Puig de la Bellacasa, 2011, p. 93). The critique cannot end by identifying the support-provision, advising, bridging, networking, and training proposed by institutional definitions of professional data stewardship as reproductive labor, though it might be. But in addition to this identification, Puig de la Bellacasa (2011) would argue we need to go further by exploring how critique can end not in "corrosion", but in creating new ways to live well (p. 91).

In the move between matters of concern and matters of care, Puig de la Bellacasa argues the necessity for the analysis of what is at stake in an issue to not stagnate with worry or association, but that deeper investments and their complicated consequences still be given their due attention. Accounting for the practices and relations that hold things together does not preclude highlighting how such a process of holding-together has differential results for its constituents. In fact, this accounting can all the more support an ethico-political stance in regard to such a holding-together. In order to accomplish this, a sensitivity to care must not fall into the trap of associating care with feminized, positive, 'good feelings' (Murphy, 2015). Taking up various forms of feminist care projects in the context of international aid, Murphy (2015) makes the necessary argument that discomfort, unease, and rage can be fundamental to care. Care as such requires its unsettling from entanglements with hegemonic histories. Murphy asks,

> "Emphasizing matters of care risks drawing politics into the microcosm of legible attachment and affect as its narrow domain of engagement. Is there such a thing as a scope of attachment, feeling, and responsibility that might make technoscience accountable to its world?" (2015, p. 732).

Here, Murphy questions the very role of centering matters of care in science studies as it threatens to build a weak position for ethico-political stances by narrowing away from "a world already violated" (2015, p. 732). This warning is key for my study. As my research questions and methods section show, I indeed dive into the narrow domain of affective engagement. My interlocutors are attached and committed to the material doing of making data stewardship work in contemporary university infrastructures. In the empirical chapter, I will attempt to describe exactly how this manifest. But, I find that my study avoids the depoliticizing potential of becoming sensitive to 'legible attachment and affect' by following my interlocutors' interpretation of the entanglement of their work with shifts in academic research contexts that cause suffering, and my own presence in these contexts. To answer Murphy's critique and shed further light on my own motivation in this project, I turn back to Puig de la Bellacasa and Susan Leigh Star.

> "Is there something embarrassing in exposing what we care for? Not only politically embarrassing, but also affectively? The closer we come to the worlds of science and technology, the more we confront something like what Leigh Star has called 'The Wall of Transcendental Shame'. The wall is particularly high 'when we try to speak of our technological lives in a philosophical manner which includes experience, suffering, or exclusion'. We feel it when 'we are silently shamed – either within academia or within the swamps of convention' (Star, 2007: 225)" (De La Bellacasa, 2011, p. 9).

This project, as it presents the confusions, frustrations, and precarities that manifest data stewardship – as well as its triumphs – attempts to scale this wall towards an ethico-political stance that explicitly names conditions in academic research that render suffering. This project sees affective matters and researchers' and data stewards' experiences as inseparable. In caring for data, data stewards care for much more. Their practices of care pry open a world of data that spins too quick and shifts too fast to allow easy categorical walls to be built, around data itself or around the practices of the people who work with it. Instead, their work demands that we care for the conditions under which data is made. Narrow in scope, this project nevertheless takes up an attention to unsettled care and puts into focus the practices that assemble data into both relational epistemological materials (Kitchin, 2014; Leonelli, 2015) and a broader cultural shift

(Beer, 2018; Zuboff, 2018). This focus on unsettled care allows for intervention – it allows the vague to become concrete, and for multiple ways that data could be otherwise to assemble on a plane of possibilities.

4. Methods

4.1 Introduction

By this point, we can understand data stewardship as a rapidly evolving notion and professional practice that exists within the confines of data sharing and reuse, Open science, and the relations of labor and care that propel data through academic spaces. The task then becomes to explicate exactly how to see data stewardship assembling toward the assembly of data. This chapter will shed light on the methodological approach and sensitivities that I worked with while working with data stewardship as a notion and practice. In this, I follow Star's (1999) call to 'study the unstudied' as following this call opens "a more ecological understanding of workplaces, materiality, and interaction, and underpin[s] a social justice agenda by valorizing previously neglected people and things" (p. 379). Data stewardship, despite its current expansion, does not enjoy a high degree of awareness among academic circles, let alone non-academic ones. It is a form of information infrastructure still under construction. It encompasses administrative, communicative, and technical tasks that many would class as the 'boring', 'unavoidable', or 'dusty' (Mons, 2018, p. 24). Furthermore, those who are practitioners of data stewardship continue to find their work unacknowledged on many fronts. Therefore, data stewardship benefits from Star's approach to 'studying the boring things'.

Returning to the notions of infrastructures that we reviewed previously, Star (1999) identifies key characteristics that make infrastructures a system of substrates that disappear into the background, until they are picked up in moments of conflict or disrepair. Three of these characteristics are operationally guiding for my methodological choices as I understand data stewardship as a form of infrastructure of care under assembly: the 'transparency' of infrastructures, their already 'installed base', and their 'learning as part of membership' (Star, 1999, p. 379–382). By transparency, Star means the ease of use that comes with a successful infrastructure. When transparent, the hybrid, mutable, and even messy maintenance of infrastructures disappears and engagement with the infrastructure becomes second nature. As data stewardship is still under assembly and therefore has not reached this level of transparency (for researchers, universities, or data stewards themselves), I found it useful to pay attention to where data stewardship *does* "have to be reinvented each time or assembled for each task" (Star, 1999, p. 381). These moments of assembly required specific methodological attention to be detected and discussed. Moreover, data stewardship is professionalizing from and in an existing network of library, IT, and research infrastructure; it is therefore negotiated and educated locally within the boundaries of that existing 'installed base'. Finally, data stewardship coalesces a particular membership around its practices which is nevertheless not formalized. Therefore, I became attentive to how existing systems and communities interacted with newer notions and practitioners of data stewardship. With these infrastructural characteristics in mind, I was more able to conceptualize the design and execution of this research as sensitive to the assembly of data stewardship as an infrastructure of care works in practice.

4.2 Exploratory documents and sampling

I came to understand data stewardship as an assembling infrastructure of care for data through a variety of materials and people. This journey began with me attempting to understand what data stewardship was, which inevitably turned out to be the crux of this project rather than an initial inquiry to answer. From initial key term searches combining "data stewardship," "Research Data Management" I parsed international and Austrian documents concerning the definition and practical expression of data stewardship, as visible in the previous section "Defining data stewardship". These documents both participated in and contextualized my further methodological choices, which included ethnographic participation in training workshops and semi-structured interviews with members of the informal Austrian RDM/data stewardship network. For example, documents published by the FAIR Data Austria project led me to interview interlocutors, who then pointed me back to specific moments in those documents, as well as other written materials from the broader institutional network around Open science. These documentary materials and snowball-style sampling of interview partners informed each other in a cyclical manner. In this sense, documents were active, material, and relational (Asdal and Reinersten, 2022) in how they traveled between my interlocutors and me. Through this network of documents and interview contacts, I also identified RDM workshops that allowed for ethnographic participation in the pedagogical environment around RDM and data stewardship. In this way, my materials and sampling became the guide for my methodological choices as opportunities for various kinds of methodological engagement arose. Documentary exploration, ethnographic participation, and semi-structured interviews all took place from September 2023 to February 2024.

4.3 Access, positionality, and ethical considerations

Before I explain my approach to maneuvering the empirical landscape of this project, it is necessary to discuss how this empirical material came to be through access and positionality. Thanks to the support of the INNORES project for this master thesis and other members of the Department of STS at the University of Vienna, my pathway to access was somewhat paved. I was, and currently am, part of an academic community that carries institutional legitimacy reaching out to members of other academic communities that also function with similar languages of legitimacy. This ease of access is a form of privilege, with this kind of inter-institutional research less likely to be possible with a non-institutional positionality. But, being an academic colleague, in the broadest sense of the term, did not only impact field access. It also influenced how I related to my interview partners and the learning spaces in which I ethnographically participated. This project cannot be deemed ‘insider' research in that I do not work with or formally practice data stewardship or RDM. However, as the empirical chapters attempt to show, my interview partners shared experiences and feelings that extended into the broader contexts of academic research, and I often connected with these experiences on some level. This enabled natural-feeling conversations and further lines of inquiry, but it also meant that I was not cleanly separate from my interlocutors or the statements they made. In fact, I often found myself feeling deeply connected to the many elements of working life in academic research that my interlocutors brought to my attention.

As John Law (2004) and many others have argued, this entangled, potentially ‘messy' nature of social scientific research is neither a disadvantage nor an invalidator. Rather, it can produce more genuine, committed projects. Nevertheless, it requires delicacy in its ethical implications. I attempted to be aware of these ethical implications in how I approached informed consent, with all of my interlocutors being highly (if not the most) qualified to comment on how this consent should look in a formal sense. In several interviews, we had active discussions about the options available in the consent form and how these might operate. My interview partners were very generous in lending their insights and personal desires to this process. Nevertheless, I am still commenting on a deeply important aspect of peoples' lives that occupies much of their day: their work. There is no way that this is a risk-free endeavor. Even with care taken for pseudonymization and communication about the project, there is still the potential that this work might cause emotional discomfort. The possibility of this is not insignificant to me, especially considering the affinity I developed for the work that my interlocutors do. I hope that what follows in this thesis grants these ethical concerns.

4.4 Ethnographic participation

Over the course of the winter semester 2023, I participated in four 2–3-hour-long workshops hosted by one large Austrian research university, but facilitated by diverse members of the Austrian RDM/data stewardship network. Each workshop was facilitated entirely digitally via video call and online learning platform that included discussion boards, real-time exercises, and learning materials. These workshops were explicitly introductory and/or geared toward specific populations of PhD candidates. As I did not formally qualify for these workshops, I gained access by personally contacting workshop facilitators based on their public registration as facilitators on workshop listings. From there, facilitators manually added me as a workshop participant. I did not receive any rejection from these spaces, even with my explicit self-framing as a student researcher. However, I had initially attempted to gain access to the University of Vienna's "Steward Certificate Course." This attempt failed due to the positioning of this course within the tuition structure of the Postgraduate Center. While this environment would have indeed been illuminating to the professionalization of data stewardship, the inability to access it alone pointed towards important distinctions between teaching RDM and training professional data stewardship. This example of a field access 'failure' brought me into a more nuanced understanding of what ethnography could mean.

Inspired by Wittgenstein's descriptive methods, Mark Whitaker (1996) argues that, "ethnography should be approached contingently, as a form of learning, rather than absolutely, as a form of representation" (Whitaker, 1996, p. 1). This perspective on ethnography, Whitaker argues, dissolves the tension inherent in much ethnographic theorizing that poses constructivist positionality (in which objectivity and representation are eschewed) against political engagement (in which representation can claim epistemic authority in order to take a normative position). Whitaker (1996) rather frames ethnography as an attempt at 'tries'. These are "public acts of attempted understanding that aim to promote rather than repress communication" (p. 5). In my ethnographic participation in learning spaces of RDM and data stewardship, this learning 'try' takes on a double meaning. I entered a learning space as a learner and entered the same space as an ethnographer learning contingently from an ethnographically sensitive 'try'. Entering this learning space, I attempted to adopt an open and explicit sensitivity to how the teaching and learning around RDM and data stewardship unfolded in context. In each workshop, I announced my dual positionality as student and researcher to be transparent with my fellow participants. Taking paper field notes during the workshops allowed me to mark impressions, striking moments, and inquiries of other participants as we all learned about data management at the university. From these field notes and the materials provided to workshop participants via the online learning platform, I compiled an ethnographic account that was analyzed alongside the interview transcripts. Rather than attempting to create a precise record of each workshop for representational comparison, thereby extricating myself from participating in the learning 'experiment' in Whitaker's (1996) terms, I attempted to maintain a loose sensitivity that allowed me to oscillate between the attentions of an engaged learner and an attentive ethnographer. The outcome allowed me to develop more of a sense for how data and data practices are learned and contested.

4.5 Semi-structured interviews

In parallel with ethnographic participation in RDM workshops, I also conducted 9 semi-structured interviews with professionals in various areas of the Austrian RDM and data stewardship sphere. As explained in the introduction, RDM and data stewardship share physical, conceptual and professional space. Therefore, my interlocutors were situated across university libraries, repositories, and dedicated RDM offices and held an array of titles that did not only include 'data steward'. I located possible interlocutors through previously mentioned documents and the public presence of RDM and data stewardship in Austrian universities. I additionally received recommendations for further interview partners in the process of interviewing. Interviews ranged in length from roughly 35 minutes to 1.5 hours and were conducted via video call or in presence. Informed consent forms with granular consent options for data publication and pseudonymization were provided prior to the interview and discussed with all interview partners. Due to the nature and size of the RDM and data stewardship community in Austria, the tailored nature of the interview protocols, and the request by the majority of participants for pseudonymization, I will not include interview protocols in this thesis. Interviews were recorded, stored via secure university cloud service, and transcribed by a transcription service before coding.

Because of the diversity of career length and prior experience of RDM and data stewardship professionals, my interview protocol was a flexible document. As per Jensen and Laurie (2016), the semi-structured interview is ideal for working with such diversity by allowing for central themes to run through elastic and collaborative interactions between researcher and interview partner. Whitaker's (1996) notion of ethnography as learning 'tries' followed me into interviewing as well. Rather than attempting to construct a strict questionnaire that would produce comparable results, I was more interested in exploring how data stewardship was framed in practice – however variably – by those who work under its umbrella and what kinds of life-worlds coalesce around data. Data stewardship is not merely a set of abstract concepts for guiding research data or a professional practice; it motivates and structures ways of living and working with data. Centering this, my interview guides were tailored to the current positions and publicly available professional biographical information about each interlocutor. Each interview began with biographical questions that opened space to discuss the variety of professional and personal trajectories that lead to data stewardship. The interview protocol then moved to questions around the emplacement of data stewardship and data in the specific institutional and professional situation of the interlocutor. Towards the end of the conversations, I asked more abstract and theoretical questions around data's longevity and ontology. In this structure, I follow Galletta's (2013) proposed protocol for semi-structured interviews that initially opens biographically or narratively, creating an experiential backdrop and level of comfortability in the interaction, and moves progressively toward more targeted or theoretically grounded questions. This initial biographical phase was key to exposing the diversity of data stewardship as a professionalizing field, and was revelatory toward the distinct notions of data that exist in this environment.

4.6 Coding and analysis structure

To begin the analysis of the empirical material, I approached the ethnographic reflection and interview transcripts with a grounded coding approach (Charmaz, 2006). As I understand it, a grounded theory approach allows for the conceptual work and further sharpening of the research focus to emerge in a 'grounded' manner rather than deductively analyzing material with a priori theoretical framework in mind. I coded this textual material using with ATLAS.ti. In vivo codes captured specific metaphoric language and more action-based descriptive codes marked themes in the framing of data stewardship, the professionalization of data stewardship, the entities that participated in this process, and the definitional discussion of data (Rivas, 2012). In this process, Star's (1999) words once again lent guidance: "Only by describing both the production task and the hidden tasks of articulation, together and recursively, can we come up with a good analysis of why some systems work and others do not" (p. 287). My open coding attempted to note both the explicit processes, emplacements, and trajectories described by my interlocutors and to capture the implicit affects and quandaries that characterized these more concrete descriptions.

From open coding, I moved to clustering the in vivo and descriptive codes into broader themes. What became clear during this process was the variety of activities and affective or relational states that my interlocutors inhabited while working around data stewardship. I then began to create clusters around these roles that were diversely situated in institutions and were associated with particular affective or metaphoric language. Through this coding process, it became clear that, situated in the emergent status of data stewardship as a notion and practice and its local expression in the Austrian context, discussing data stewardship was a discussion of multiplicity. But what kind of multiplicity? The question of how to continue sorting the codes posed a problem. Workshop attendees wondered what data was, or if they had it. Historians refused that they possessed data to be archived. Data stewards wondered how and who to be through it all. These started to feel like ontological questions. Or were they epistemic questions? Like Whitaker (1996) who chose a third option away from representationalism and anti-representationalism in ethnography, Lynch (2013) offers a way to work with the potentially ontological without advocating for a particular, bounded ontology: staying empirically close to how ontologies are discussed in situ. My discomfort at fully declaring the work of data stewards as ontological work found echoes in Lynch's (2013) definition of an ontological slippage that can tip into adhering to a Theory of Everything (TOE):

> "Accordingly, if we assume that any empirical description reverts to a prior theory, and that ontology is the most comprehensive of theories, then any attempt to describe 'ontologies' in the world must revert to an ontology of the world – a comprehensive picture of the worlds constitution" (p. 453).

What came forward from my coding, with Lynch in my ear, is not a strict argument for what data stewardship or data stewards are ontologically, but rather an attempt to describe the complex ontological discussion of both academic research data and the people and infrastructures that surround it. To capture the less-than-explicit quality of the framings that my interlocutors provided, I organized the coming empirical reflections along four modes that professional data stewardship encompasses: the data guide, the data guardian, the data educator, and the data confidant. These modes are my own interpretive typology separate from the formal responsibilities and roles described by the literature and professionalizing efforts around data stewardship. With this analytical organization, I hope to 'ontographically' display, beyond the normative discussion of what data stewardship *should* be, the internal practices and definitional negotiations of data stewardship on the way to its assemblage.

5. Empirical findings

5.1 Introduction

With the methodological protocol of this project behind us, we can now begin to examine data stewardship alongside those who practice it in the Austrian context. After reviewing existing scholarship from STS and information studies, we move from an understanding of data sharing, infrastructures, and work as characterized by a negotiation of data friction. Ontologically, data will not sit still long enough for anyone to reach consensus on reuse or reproducibility. Furthermore, frictions in the journey are laden with ethico-political consequences. After establishing this baseline for the movement of data through academic communities and institutions, I explained how assemblage thinking helps me to relate data stewardship as enacted by data stewards with the assemblage of data as both a broad cultural notion and concrete, material objects. As presented while working through the methodological choices of this research, to attempt to locate data stewardship is to look under many institutional rocks to see what diverse populations might be bustling there under a seemingly uniform surface. Semi-structured interviews, document work, and ethnographic participation provided the venues through which data stewardship as an institutional idea, a personal work practice, a profession, and an ideal could be explored. In what follows, I will attempt what I understand to be the process of assembling data stewardship in practice following this guiding question:

How is data stewardship assembled as an infrastructure of care for the long-term futures of data?

It is possible to work with the assembly of data stewardship in different ways. Data stewardship occupies and coalesces various spaces, places, definitions, sets of practices, and sets of values. But it is not enough to say that data stewardship exists in a multiple fashion. Rather, it is situated and constituted by the specific conditions under which it grows and in the diverse arenas it occupies. Therefore, the following chapters will be organized along the different modes of data stewardship, as I found them. These modes are ***the data guide, the data guardian, the data educator, and the data confidant***. We will spend the next chapters exploring the practices and affects that make up these roles. While they may exist in a distinguishable fashion, the modes that data stewards step into do not adhere to strict distinction in practice. Individual preference, institutional positioning, and fluctuating data interact in ways that have consequence, that leave detectable marks, and that make inflexible category-building a doomed project. Nevertheless, defining these modes allows for the assemblage nature of data stewardship as multiply constituted to become visible, and therefore actionable.

This research – as situated within broader questions of the left-behinds of innovation in academic settings – sees it as useful to move through the roles of data stewardship, beginning with its institutionalization in **the data guide**. The data guide was constructed in attempting to answer this question: *How do data stewards maneuver and participate in (re)building data infrastructures?* This question queries how data stewardship, as enacted by the Austrian data stewardship community, moves through academic institutions and their pre-existing data infrastructures. Understanding infrastructures as mutable hybrids (Star and Ruhleder, 1996; Star, 1999), this mode of data stewardship highlights what existing institutional systems and boundaries mean to the assembly of data stewardship. We then move to the mode in which the data steward has the strongest interaction with the process of ontologically defining data – **the data guardian**. This mode emerges from asking *how data stewards understand and engage in defining data*. Here we see data stewards interact with the notion of 'good' data on the way to preserving it for eventual reuse. Then we explore the duality produced by teaching and training research data management (RDM) and data stewardship in **the data educator**. Data stewards not only steward data but also educate along the vision of potential data futures in line with good RDM practices. They are also the recipients of formalized educational programs that prepare prospective stewards for a professional career. To this mode, I ask *how data stewards undertake to transfer knowledge and practices about data*. Finally, the affective commitments that penetrate through data stewardship's enfolded constituencies become lively in practice in **the data confidant**. Here I ask, *how do data stewards engage with their own affective states and those of others that accompany data?*

With our guiding questions and structure laid out, we can now move to the empirical exploration that produced the data guide, the data guardian, the data educator, and the data confidant.

5.2 The data guide

As Mons (2018) defines it, data stewardship is "the process and attitude that makes one deal responsibly with one's own and other people's data throughout and after the initial creation and discovery cycle" (p. 36). And while he, as one of the originating contributors to the FAIR principles and a major name in the realm of research data management and stewardship, emphasizes the necessity for the professional uptake of this definition of data stewardship, he leaves open exactly *how* data stewardship will exist in institutions. Precisely where data stewardship sits is institutionally situated and relationally determined, with an array of constellations represented across large Austrian research universities. In this chapter, we will explore how data stewards become parts of institutions and how they participate in infrastructures as "hybrid creations of work practice and information medium" (Bowker and Star 2000; Star and Ruhleder, 1996). To integrate into an institution, to participate in infrastructure is – in the case of data stewards – to exist in a multiple form. Data stewards both become parts of institutions and guide data and researchers through those same institutions, occupying a liminal space that hovers between institutional members and external institutional analysts. In order to best guide data and researchers through the maze of the contemporary research university, data stewards must become well versed in the general topography, the specific cultures of centers or institutes, and the technical capacities of the university.

In training data stewards (we will meet the data educator soon), the following role definition becomes key to inducting prospective stewards in to this emerging professional community:

> *... data stewards open channels of communication between various stakeholders. The idea is very much... that of a bridge of connecting researchers, students, with research infrastructure providers, with other services available within an institution, but also outside of the institution.* 8:21 ¶ 23 in DW_U2_230211.docx

Here, the data steward takes on the responsibility of communicating between communities inside and outside the institution, with data as the unifying force of this network. From funding bodies to student groups to administrative units, data stewards move between communities within the university to best facilitate good data management. But this has not always been the case. In the past, RDM was marginal. As one interlocutor recalls,

> *[...] fifteen years ago, we were an island in this library*. 7:15 ¶ 63 in DW_P1_231004.docx

The broad trend of digitization and the exponential increase in the production of academic research data have constituted a critical shift in those fifteen years. While a year-to-year genealogy of this shift is beyond the scope of the questions at hand, it is worth noting the velocity at which research data management and data stewardship evolved within university structures. In the case of the one Austrian university, research data management moved from an office within the library primarily associated with institutional repository management to a rapidly expanding office that included multiple professional data stewards embedded across faculties. The speed at which this transition continues to take place both reflects and creates the conditions of multiplicity in which data stewards work.

> *Functionally day-to-day, again, I'm a lot like an air traffic controller who connects people, who answers a lot of questions and then connects people to further resources pretty frequently and really tries to save them* [the] *time and frustration of trying to navigate this infrastructure by themselves alone and try automatically find them the most direct route to something or to have a conversation on their behalf, serves them well*. 6:44 ¶ 47 in DW_U4_230611.docx

The 'air traffic controllers' of the university, data stewards leverage institutional literacy to benefit researchers and their data, not without a decent dose of exhaustion in coping with this demand. Especially considering that prior to training as a data steward, through a course or some other professional structure, data stewards generally have no experience with RDM or data stewardship.

> *...there is no, 'This is RDM course if you start your PhD' or a refresher, 'this is what RDM is of course' if you start a postdoc contract.'* 1:90 ¶ 47 in DW_A1_230928.docx

So, data stewards enter their institutional homes with relatively new knowledge of RDM and perhaps very little knowledge of the topography of their institution. This knowledge must be gained through self-teaching, experience, and knowledge-sharing among data stewards (if there are multiple at one university). But, the relationship between data stewards and their institutions does not end there. In the embedded model of data stewardship (Bardel et al., 2023), stewards must integrate into their disciplinary contexts in order to best guide data through the institution.

5.2.1 Becoming a guide: institutional embedding

As data stewards described their process of undertaking data stewardship as a professional career, they told tales of getting to know the people, places, and materials that constituted their institutional situation. This entailed an process of outreach in the hopes of raising awareness that data stewards exist and are available to work with researchers on an array of data-related questions.

> *So, this was one of my first things I did. I tried to talk to all the group leaders and find out what are their problems, and* [it] *came up that they had or they have quite similar problems as I encountered during my time as a researcher*. 4:6 ¶ 11 in DW_U1_233110.docx

During this 'getting to know you' process, data stewards must accomplish several things at once. They explain RDM and its broader implementation at the university; they argue their specific relevance and utility to the department or faculty in which they are embedded; and they emphasize their own disciplinary expertise in connection with their new professional environment. Embedding is a process of relation-building across professional registers. There are more subtle conditions at play beyond university policies that state, "Researchers should be supported in processing (i.e., collecting, storing, changing, using, spreading, deleting, etc.) research data in the best possible manner" (University of Vienna, 2021, p. 1). Disciplines, research groups, and even individual researchers differ in their understandings of data and how it should be treated (Borgman, 2022; Green et al., 2023; Feldman and Shaw, 2019). And they differ in how they tacitly form communities with distinct values and conceptions of research practice (Traweek, 1988) that are expressed through data. Data stewards must sense this out to be effective.

> *I think it also depends on the field. I mean, it also depends on the chair, obviously, because I think there's this personal element there.* 1:91 ¶ 50 in DW_A1_230928.docx

This process can become further complicated when we take into account the remaining blurriness that characterizes the boundaries between professional data stewardship and the research contexts in which data stewards are embedded. While analyses of data stewardship might validate that developing technical solutions for RDM is "not a task for the researchers, but Data Stewards can act as their representatives" (Bardel et al., 2023, p. 79), this dynamic is far more complicated in practice.

> *I think a lot of them still don't really know what I can do and where the boundaries of my position, how far I can help them in deep in their project. I think that's also something we are still trying to figure out*. 4:17 ¶ 35 in DW_U1_233110.docx

> *It's on the one side offering services and on the other side trying to not being completely immersed into one research group (crosstalk) because if you offer people maybe to help them, then the demands will get bigger and bigger of what you maybe can do for them because they're all happy if they have an additional person in their group and do some stuff*. 4:27 ¶ 55 in DW_U1_233110.docx

Here, a data steward describes the porous nature of the border between researcher and data steward during the process of institutional embedding. In this case, a piece of code for file organization was the border-blurring object in question. Building tools for the organization or analysis of data, in this steward's eyes, is too much. Even in the case that a data steward has the technical skills necessary to build and implement a solution for a certain data problem, this is not necessarily desirable in combination with the many other things data stewards do.

> *It also depends on the level they work at, the institution, because working on the level of the faculty you're probably not going to be able to—you're not going to have the time to dedicate to actually writing a script for something for someone. Even if you might be able to do it, you probably won't have the time because you need to go through twenty data management plans and provide feedback on them, you need to do training courses for your PhD students and blah-blah-blah. You have a whole bunch of things that you're doing, and this is very low on the list on your agenda.* 8:62 ¶ 87 in DW_U2_230211.docx

Other data stewards described their processes of boundary building as they became embedded. Their boundaries centered not on building technical solutions, but on taking up tasks that researchers themselves should do in the framework of RDM and data stewardship that is currently under construction across Austrian universities.

> *Of course, some say, "Why don't you write my data management plan?" But it wasn't many people. But I had a longer discussion with a researcher who expected me to write it and I said, "When you take the time to explain me your project in the detail that I can write your data management plan, you can write it yourself."* 5:50 ¶ 91 in DW_U3_230311.docx

> *I never touch anyone's data. They have to do all of this themselves, but I'm here to sit with them while they do it and answer their questions and all that stuff.* 6:43 ¶ 43 in DW_U4_230611.docx

These data stewards find ways to reject requests that they deem to be outside the reasonable bounds of data stewardship work, as it is conceived of in their university contexts, and in the broader data stewardship community. For the sake of their own time limitations and workload, data stewards must not only advocate for data stewardship and advertise their services, but also set firm boundaries when requests for those services extend too far. There is, after all, a professional category to be built here. It would not do to accept every number of requests for any type of service in the process of developing data stewardship as a legitimate, respected, professional practice.

> *It's more having a common understanding of this area of responsibilities and tasks that data stewards typically take on, especially because many of these tasks have been taken on in the past by people in various areas, both researchers and research support staff, working in different parts of the university or other research institutions.* 8:18 ¶ 23 in DW_U2_230211.docx

So, there is, admittedly, a bit of confusion in building the boundaries of the data steward in that data stewardship professionally practiced encompasses tasks that have been absorbed from other areas of the university or from researchers themselves. As Star (1999) reminds us, an infrastructure "wrestles with the inertia of the installed base and inherits strengths and limitations from that base" (p. 382). Considering data stewardship as an infrastructure under construction, out-of-bounds requests from researchers become more understandable. But even if a data steward has the skills and time capacity to work with researchers in an in-depth way, there are other barriers to be considered, like funding.

> *I think I could easily also support them* [members of another faculty] *and it really makes sense content-wise or discipline-wise to have them with me. But as the faculties pay half of my salary, my faculty pays half of my salary, of course they don't want me to work for another center.* 5:59 ¶ 107 in DW_U3_230311.docx

Here, the harsh boundaries come not from the data stewards, but from the structures of institutional funding and their concomitant distribution of responsibility. Even if this data steward feels a disciplinary affinity for another department or faculty, the demands of scarce temporal and financial resources mean that funding boundaries must be kept clean and coherent. There should be no blurring of boundaries here, at least from an institutional point of view. But there are other desires at play beyond what makes the most institutional 'sense'.

> *We are currently trying to figure out how deep I can go into projects because I also find it interesting for me to do a bit of research on the side*. 4:22 ¶ 39 in DW_U1_233110.docx

This data steward, in the future, might want the boundaries to flex – perhaps even contrary to the current effort to institutionally legitimize data stewardship through boundary-making. This data steward sees a future in which they can not only participate in tasks related to data management in their institute, but also contribute as a researcher 'on the side'. The disciplinary experience that makes embedded data stewards desirable in the first place – for their understanding of specific data norms and cultures – also makes drawing strict boundaries difficult in the face of intellectual curiosity and camaraderie. It is difficult to predict if this flexibility will be feasible considering the volume of tasks that occupy a data stewards' day. Furthermore, data work is still structurally siphoned away from 'real' science and its associated publication and prestige pathways (Gabrielsen, 2023; Pinel et al., 2020; Plantin, 2021). While we can hope for a time when resource limitations and pre-established academic career trajectories do not limit the freedom of scholars/stewards, they are currently busy guiding data through the winding halls of the university.

5.2.2 Guiding data: navigating technical infrastructures

From the perspective of a data steward, those halls can become a maze for researchers, departments, and data.

> *[…] the infrastructure here is massive. It's absolutely massive and sometimes departments that run services and stuff, they're aware of a problem or an inconvenience and they don't have the capacity to advertise a solution to the whole world and to run it for everybody. But if you go and ask them, they'd be like, "Yeah, I think we got to [a] work around for that. We could make this work for this person better and remove that headache." Even if I think I know what is publicly available, sometimes I'm surprised*. 6:31 ¶ 27 in DW_U4_230611.docx

The technical services that universities can provide are not necessarily common knowledge. Though many infrastructures exist to analyze, store, preserve, publish, or graphically represent data, the scale at which large research universities operate makes this infrastructure anything but transparent. In Star's (1999) terms for defining an information infrastructure, "Infrastructure is transparent to use, in the sense that it does not have to be reinvented each time or assembled for each task, but invisibly supports those tasks" (p. 381). When data stewards and researchers interpret the data infrastructure of the university, this transparency is not a given.

> *It was also in this case they had an external server because the external person had access, it would always also have been possible with the university infrastructure. They even paid out of their own pocket. They paid the server privately. It was maybe not that much, but I thought 'it can't be'. It should not be necessary to pay for infrastructure out of your own pocket because you can't find a solution in the institution. 5:36 ¶ 59 in DW_U3_230311.docx*

This data steward's incredulity when discovering that an institute had paid from its own funds for a server solution that could be run over the university's existing infrastructure belies its opacity. It also reveals the investment that data stewards make in the process of embedding. They must argue their own disciplinary experience and the relevance of RDM, while discovering (or detecting) the unadvertised tinkering that goes on in the background of academic research. If a research group or institute has not managed to tinker their way into a solution, data stewards guide them to it.

> *He contacted me and said, "What should we do? Maybe we are not at the university maybe for months in between* [fieldwork trips]*." So, I arranged a meeting with the computer centre with the right department, with the IT support for research department, and they had a solution. So, my job was identifying the problem, not staying on the surface – like the help desk did and it's perfectly okay that they did – and connecting him with the right department in the computer center because it showed that we had an institutional solution in place, but he did not know about it.* 5:27 ¶ 55 in DW_U3_230311.docx

The 'air traffic controller' allusion again becomes relevant. But this time, with a deeper investment in the technical needs of researchers and their data. Data stewards have the mandate to go further and 'not stay on the surface'. Educated into communication and institutional maneuvering, they are primed to accomplish coordination efforts (Grugulis and

Vincent, 2009; Chorowzewicz, 2022). Data stewards' disciplinary experience and personal commitment to the intellectual projects of the researchers they support uniquely situate their guidance through university data infrastructures. This is a critical dynamic as data practices and data volume expand.

> *People so frequently at this point begin to climb outside of the storage units of the cloud service that you get here the university [...] as either they're organizing a whole lab or they need more storage than that. I get a lot of questions about what should I do, what's it going to cost me, that sort of stuff, how easy is it to set it up on my computer*. 6:23 ¶ 19 in DW_U4_230611.docx

Data is also becoming more complex. The file formats that researchers produce are not always the simplest to preserve in the long term. Data stewards come in to guide complex data into a form that can be technically maintained by the university. After all, "In practice ... the infinite possible ontologies of objects are limited by the pragmatics of data collection and by the inescapable inertia of categories already in use" (Bowker and Star, 2000, p. 117).

> *So, you need not only the content but you need to preserve the functionality. This is the difficulty. We have a pilot now running where a young software developer designed and is currently developing a tool where* [...] *it's like containers, docking containers, where the databases run, so there's no security risk* [...]. 5:20 ¶ 43 in DW_U3_230311.docx

Data stewards might not know how to construct a technical solution to a problem, but they must at least know the parameters of that problem well enough to find someone who can build a solution. As a data steward, technical ingenuity and a deep knowledge of the specific needs of research communities produce tailored solutions to complex data problems, working toward the kind of transparency that makes infrastructures infrastructures (Star, 1999). This makes the mode of the data guide critical to data journeys through the university and beyond. Data stewards actively engage with building infrastructures based on their prior experience across research contexts.

> *Especially coming from a researcher position, you're maybe thinking from — you want more stability and you're maybe not happy with the system as it is – and you're thinking what are the ways I could maybe do something that helps adjust this and help fix this system as it is now?* 8:54 ¶ 71 in DW_U2_230211.docx

As data stewards guide data through institutional infrastructures, they leverage capacities for communication and affinity across academic communities in the project of helping data live on. Tasked with 'bridging' between offices, services, departments, and faculties, they think systemically about the journey of data through the university and the affective and resources consequences this journey has for researchers (Grugulis and Vincent, 2009; Chorowzewicz, 2022).

5.2.3 Guiding resources: a question of funding

After following data stewards as they become institutionally embedded and guide data and researchers through the technical infrastructure of the university, one might guess that this work is highly valued in academic research. Unfortunately, this has yet to fully become the case. In Mons's (2018) framing, "as these efforts are not properly awarded and supported, the experts who can deal with data are undervalued, do struggle with job security, and frequently disappear to better-paid jobs, leaving academic science crippled and stuck with half-baked solutions" (p. 21). To make a seemingly basic point clear, data produced by the university exists in the thick web of funding strands that defines contemporary academic practice (Mirowski, 2011; Borgman and Bourne, 2022). However, in the context of research data management and data stewardship, this web takes on a specific configuration and contributes to their bounding and definition.

> *In Austria we do use the term like data steward and data structure a fair amount also because it's very present on the policymaker level. The ministry uses that term a lot. So, obviously it's stands to reason that we use it as well when we adapt to their wording because that way we can be able to understand each other better when we're speaking to them about funding these positions, which is the major issue in this area*. 8:17 ¶ 19 in DW_U2_230211.docx

In an academic context that incorporates not only intellectual and educational mandates, but also demands that researchers seek and secure their own sources of funding, data practices and the notions that guide them are deeply implicated in the financial dynamics of the university.

Digitization writ large is a stated goal and product of the university, and large Austrian universities specifically as evidenced by University of Vienna's (2020) "Digitization Strategy," TU Wien's ".digital" office, and the University of Graz's (2024) digitization section of the "Development Plan

2025-2030." Narratives of catching up to and contributing to the already-present digital future – at this point a ubiquitous narrative across global educational institutions – characterize the digitization schemes at large Austrian universities. The digital is essential. To lag behind the digital is to lag behind 'society' and 'innovation', as mutable as those terms may be. Nevertheless, the question of funding for the operation of digitization projects does not follow these stated goals in an analogous fashion. One respondent involved in institutional repository management described the precarity of the funding underpinning repository operation.

> *It's not so good because many things—you cannot do anything and cannot do some developments without money. It's not possible*. 7:18 ¶ 95 in DW_P1_231004.docx

To run an institutional repository, a key infrastructural node in the web of RDM at the university, requires not only researcher-facing staff but technical staff and computing equipment. In order to meet the demand by university researchers for up-to-date functionalities and the kind of user-friendly interface that early-career researchers have grown accustomed to from non-institutional repositories like Zenodo, institutional repository teams need to invest in purchasing or building hardware and software solutions and tailoring those to meet the specific needs of the institution. In cases where collaboration across institutions, even international ones, is desirable to enable cross-institutional functionalities and shared resources, funding is still sparse. One such project involving large Austrian university bodies received only around half of its proposed budget from the responsible ministry. Especially in the arena of institutional repository management, the capacity of RDM and technical teams to build up the kinds of services that researchers need or want in order to store data long-term with the university depends directly on the kind of funding granted to such an endeavor. When RDM offices exist at only a fraction of the size of a full academic department or a human resources office, the volume and type of services they can provide will necessarily be limited. Taken in combination with the lofty goals for digitization and data as a means through which universities might gain a return on their investments in research and a revenue stream (Teperek, 2018; Falkenberg and Fochler, 2022), this seeming mismatch of funding for RDM with the proposed significance of digitization for the university might suggest a deeper mismatch in the values implicit in academic data futures.

However, a shortfall of funding is not the only question linking broader financial trends in academic research with RDM and data stewardship. Indeed, many respondents pointed to the data management plan (DMP) as a kind of obligatory passage point (Callon, 1984) that must be reckoned with in order to secure third-party funding for research. The data management plan can take various forms, depending on the discipline and institutional situation of the data-producer. But, most data management plans demand a few key pieces of information, like data production practices (on which kind of servers will research data be stored during a project, under whose ownership data is produced, what privacy and security measures will be implemented during production), long-term storage plans (repository type, length of storage) and reuse measures (licensing, assignment of a persistent identifier, publication in a data journal). Major funders, like the FWF (Austrian Science Fund, n.d.) have specific demands depending on the disciplinary content of the data produced with funding. At the EU level, the DMP is mandatory as a deliverable within the first six months of the Horizon Europe-funded projects (Horizon Europe, 2024). Researchers and the data stewards that assist them in the creation of a DMP must comply with these standards in order to win research grants. Researchers often approach the DMP with either trepidation or panic, many unfamiliar with this object of scientific practice that has only come into regular use in the past few years. According to one data steward,

> They don't understand—they get this template from a funding agency that's like, "answer these questions," and they're like, "Yeah, that would be really great, but I don't know what this question is asking me. I don't know what concept this relates to." 6:1 ¶ 19 in DW_U4_230611.docx

As data and the practices that create and accompany them are multiple, it is not self-evident how the output issuing from the DMP guidelines should look. As clear as guidelines might seem from the perspective of the funding body, the process of interpreting those guidelines across diverse types, volumes, and temporalities of data proves tedious at best and impossible at worst. In a data landscape in which just a few years can produce monumental changes in popular software, file formats, and institutional visions, the promissory temporality of the DMP for funding becomes dubious.

> *I think it's also a bit difficult that FWF wanted that because they should know that the university can't really promise that. Of course, they're [the university] hosting, yes. But keeping it running? Well, you have to do it for the funding because they say we won't fund a digital project if it's offline after three years. I completely understand that, but who can promise that in fact*? 5:47 ¶ 83 in DW_U3_230311.docx

Cases like these drive researchers toward data stewards to work out a technical solution on the way to securing funding. Here, data stewards become a portal through which researchers must pass on their way to securing funding. The DMP becomes an object that not only lays out the guidelines by which a funding body could potentially audit or review a research group's data practices, but is also one element in the constellation of proofs that researchers must provide in order to gain necessary legitimacy for their research and continue funded research activities (Fochler et al., 2016; Sigl et al., 2020; Falkenberg and Fochler, 2023). In this way, DMPs nudge researchers toward data stewardship, which in turn tenuously drives funding toward data stewardship.

Despite this seeming drive from funding agency to researcher to data steward, there remain questions about the long-term financial security of data stewardship. While projects and communities exist to develop data stewardship and communicate its value, like FAIR Data Austria and the Research Data Alliance, there nevertheless remains, at this moment, a shortfall in resources for the institutional grounding of data stewardship. In one large Austrian university, data stewards worked on the basis of a pilot program. Institutionalized data stewardship is still in the testing phase to see whether it - as an idea or a practice - will bring the university closer to its digital goals. This means data stewards work on a limited contract. Once again, data stewardship and institutional funding exist in a state of tension.

> *I can answer their little questions that are easy and that we can build good enough relationships with each other that when we enter into a place where we have to have a more challenging conversation or we have to deal with something that's more complicated, we have a mutual relationship, we have a mutual respect, we have a mutual understanding of one another, and a mutual trust where we can have that conversation productively. Again, that's my hope for the future and what I hope this is long term*. 6:2 ¶ 87 in DW_U4_230611.docx

Here, this data steward expresses the desire to pursue meaningful embedding at their department, fostering and nurturing relationships that can move forward in time. Such long-standing, trustful relationships, this respondent argues, are what will facilitate data practices that give research data a long life. Without this longevity, the work of the data stewards and the data produced in researchers' projects stand to suffer. It might be an easy move to quote the adage 'put your money where your mouth is' in critique of the contemporary neoliberal university here. And this quote might reflect an honorable spirit. But it would not serve to help untangle the fiscal and organizational web in which data stewards find themselves working. And it is the specificities of this web that offer opportunities for shifts toward more livable futures for researchers, university administrative staff, and data stewards alike.

5.3 The data guardian

As data stewards build an institutional topography of the university and use it to guide data producers and data through data infrastructures, the overarching goal is to get data into its permanent resting place in its final form. But how do they participate in deciding what data will be stored? If data stewardship is "the process by which institutions steer this transformation [of data practices] along the whole research data lifecycle" (University of Vienna, n.d.), then the data guardian is the role within this process that most contributes to an ontological discussion of data. The data guardian answers and influences the question "What *is* 'good' data?" on the way to storing it. By preventing loss during the production of data, preserving research data once it is deemed desirable, and encouraging the deletion of data that does not meet the standards for long-term storage, data stewards intervene in the process of defining what makes data 'good' and, therefore, what makes data data.

5.3.1 Preventing the loss of 'good' data

In a world where more data is better (Kitchin, 2021; 2022), the avoidable loss of data is a tragedy. And in an academic research context in which researchers' careers, reputations, and livelihoods – not to mention passion or hours spent – are staked on the volume and quality of the research they publish, data loss is no small threat. Unfortunately, or fortunately, data loss does not only come from an inexplicable disappearance of data or the crash of some far-off cloud service's servers. More often it comes from far less exciting human error. In their capacity as educators, data stewards take it upon themselves to plead with researchers to forestall data loss caused by less-than-ideal RDM practices.

From folder systems that become indecipherable labyrinths to storing essential data on only a singular vulnerable device, research data storage comes in all shapes and nerve-wracking forms. Data stewards – both in individual interactions with researchers and in workshops – attempt to guide researchers away from practices that will most easily result in their research data being lost. As one data steward narrativized,

> *I said, "Please, have this rule[s] for file naming and so on in place before you start because when five people work at the same folders, it's really bad if they don't keep to rules." If it's*

> *your own data, okay, but if you work together with others, please think about that now*. 5:61 ¶ 119 in DW_U3_230311.docx

In one RDM workshop, data stewards prompted attendees to react via emoji to the prompt, “How would you react if your computer was stolen?” The array of reactions, including my own, justify the data stewards’ stress on this point. In their capacity as the guardians of the university’s research data, they must do what they can to beat back our worst data nightmares. But a compromised device isn’t the only source of data loss. There are gaps in institutional policy that make consistent data practices difficult to foster.

> *If you didn’t know anything, then you always have your peers but that is such informal, not codified, not institutionalised, not documented, never written down. So much knowledge again lost. 1:96 ¶ 58 in DW_A1_230928.docx*

While colleagues in research institutes certainly have the capacity and motivation to guide newcomers through localized data practices, data stewards argue that this is a fragile safety net. Researchers sense this, and often search for something more - if they find a venue, like an RDM workshop in which to inquire.

> *They want to know how they should keep their lab notebook and stuff like that, or how they should organize their files and their research group, so that they don’t lose things when people come and go from the lab*. 6:19 ¶ 15 in DW_U4_230611.docx

Researchers, especially early-career ones, know that high turnover is now a characteristic of contemporary university life. With the entrance of New Public Management notions into university operation, projectification and the proliferation of short-term contracts form the basis of contemporary academic knowledge production (Midler 1995; Packendorff and Lindgren 2014; Burke and Manathunga, 2023). Nevertheless, this quote shows that researchers – despite experiencing insecure conditions – want to improve the state of data practice. Whether motivated by the daily operations of a research group running smoothly or by a loftier belief in the value of the intellectual work they do, researchers are cognizant of the potential for data loss linked to the turnover of academic staff, and want to find solutions to prevent this from happening. Here, data stewards come in to provide advice and solutions.

> *I was here to help with all of that, laying out some kind concrete examples of things I could help them figure out [....] some things that I suspected would maybe be things that bothered them, but they never really thought that they could come up with a solution for, like onboarding and offboarding plans for when people arrived and left their research groups or departments. You lose a lot of data and information that way. So, we could come up with a plan, we could fix that*. 6:38 ¶ 39 in DW_U4_230611.docx

Here, a data steward gives an example of some of the 'concrete things' that they could suggest for a senior researcher or project leader. The data steward – with previous disciplinary experience that might clue them in to the possible outcomes of high academic turnover – offers the possibility that data loss can be avoided through conscious, ordered, verbalized data management. In this case, the data steward as data guardian stands between the precarious conditions of the university and the loss of data. Whether an asset or a return on investment (Fochler et al., 2016; Mons, 2020; European Commissions, Directorate General for Research and Innovation, 2016), 'good' data must survive on to fulfill its mandate as a reusable resource. The data steward alone cannot do much about the standard length of contract in a given institution or the temporality of a project, but data stewards can and do intervene in the knotty conditions of the neoliberal university via data practices when the loss of data is at stake.

5.3.2 Preserving 'good' research data

Beyond a preventative approach to instilling good RDM values into the young and not-so-young researchers of their universities, data stewards are also the purveyors of knowledge relating to the long-term future of academic research data. If the final resting place of data is not the obscure depths of a researcher's PC, then it ideally should be in a data repository where it can be stored for at least 10 years, easily accessed, and referenced by potential data re-users. In the defense of data, the repository is a fortress in a storm. But it is not necessarily common knowledge which repository would be most fit for purpose, or how to use it once it is found. Once again, data stewards enter to fill in gaps in missing knowledge, pointing researchers in the best direction possible. The ability to be this kind of signpost depends on the institutional situation of the data steward, as referenced previously. If a university is working with a more generalist model, data stewards might not have the discipline-specific knowledge necessary to point researchers in the right direction. However, as was the case with one large Austrian university, embedded stewards were able to provide very targeted advice to researchers concerning the appropriate repository to approach with their data. In workshops, participants,

> *...want to know what storage facilities for their data are available at the University of Vienna. They want to know what a repository is and how they choose a good one.* 6:18 ¶ 15 in DW_U4_230611.docx

In RDM workshops, facilitators introduced re3data.org as a port of call in the search for an appropriate repository. They also emphasized that general repositories (such as Zenodo), which have few requirements for the admission of data and are not discipline-specific, should be used only as a last resort if more fitting repositories are not found. In some cases, discipline specific repositories are not found, especially in the humanities, where the notions of 'data' and 'source' or 'repository' and 'archive' have not yet reached the clarity of distinguishability (Feldman and Shaw, 2019). One attendee in a very specific sub-field of language studies struggled as – in their account – no one in their department knew about a repository that would be appropriate for depositing the already worked-over texts that constituted their research data. The student further wished to know how the data their specific project produced would be handled if such a project were based on already-archived and public copies of ancient texts. Unfortunately, these were not the kind of questions that the RDM workshop format could handle (see the data educator). However, the facilitator made clear that issues such as these could be handled on a more detailed basis by working with a data steward one-on-one.

And as a repository respondent made clear, all kinds of data can be desirable for preservation.

> *But so far, I think if the data producer thinks their work is of value for others and is willing to do the work to make sure it's in a shape or form that we can work with, at least personally, I wouldn't say no, we're not taking in that data because even if something is very niche and strange, we can still have it.* 1:119 ¶ 94 in DW_A1_230928.docx

Here, the 'data producer' is the primary determiner of the value of their data. There is some value to be gained by having 'niche and strange' data in a repository. Those working with research data from both the production side (data stewards) and the preservation side (repository staff) argue that the determination of 'good' data belongs with the academic community, specifically with the researchers who produce said data. In their defense of 'good' data, data stewards embrace the idea that valuable data is determined by those who value it. This is not to say that anything goes. Data stewards encourage researchers to use the DMP to reflect on the quality of the data they produce *before* the initiation of the production process.

> *Well, you write about how your output is preserved. For example, who is responsible for the data, which licenses you would apply or which licenses are on the material you use. There's also some quality criteria for repositories. You'd also write about that*. 5:53 ¶ 99 in DW_U3_230311.docx

Data stewards hope that through an RDM consciousness, researchers will have more success in archiving their research data in a repository, thereby expressing the values of open data and open science, more broadly.

> *But I also think that from a more, not political, but strategic [...] standpoint, I think as soon as you're going into the Open science perspective and 'sharing data is good and should be done and it gives us work' and all that, it's really hard to justify not taking certain types of data or certain datasets. I feel as soon as you go into the 'we're open, we're sharing, we're archiving,' there's this consequence that comes with it that you need to take everything if it's in a shape that is actually publishable and reusable and well done. We're not talking about fake data or badly documented data, obviously. I think that is actually nice. I like the thought that there's very niche research being done. If some professor in a small office after all decides to archive what they've been working on for the past ten years on their own, that they can also do that*. 1:122 ¶ 90 in DW_A1_230928.docx

Here, there is a reference to the time, effort, and idiosyncrasy of individual researchers that imbue their data with value. By explicitly pointing to researchers as the source of value creation for research data, data stewards value the valuers. In the case of an epigenetic database centered on the genetic profiles of twins, Pinel et al. (2020) argue that the personal relationship between PhD-level researchers and the twins that donate their time to be measured, recount medical histories, and give genetic samples becomes "inscribed in the data and in the researchers and contributes to the value of the dataset in that it creates ways of knowing" (p. 189). Data stewards or repository staff, as then tasked with preparing data to be archived long-term, recognize this inscription – if not actively then tacitly. Expressions of value in many forms – emotionality, explanations of datasets, stress about the unpreparedness of datasets – from researchers travel through data stewards on the way to becoming archived. And this expression is taken up as partially constituting the value of the data to be archived. In the face of timescales of academic research that demand quick, efficient and atomized units (Fochler, 2016; Dollinger, 2020), paints a picture of research that hearkens back to academic days gone by when a researcher could perhaps work with one set of data for ten years. Whether or not this was ever a reality is less important than the idea that research data could be produced in conditions that not only prioritize efficiency, utility, and reusability, but also acknowledge research data as precious parts of researchers' lifeworlds. Even more significant is that this respondent sees data stewardship as playing a role in creating those conditions.

In the terms of Pinel et al. (2020), "It is through this relational labor that the data exist as valued objects that can be used to make new claims about the world" (p. 191). Data stewards and repository managers, by working directly with researchers to prepare data for preservation (and possibly publication), come to understand this relationality on a granular level and consider when it comes time to determine if data is 'good' enough to preserve. And they hope that it is. But, at the end of the day, data must either be stored or not. The affective investment in research data must be negotiated with existing codes (De Landa, 2016) that make data desirable and coherently assembled. This negotiation produces 'good' data that is worth being entered into a repository for long-term storage and is worth being preserved in a protected format that can potentially be accessed and interpolated by others in their own research. But, this process can also function in reverse. Data stewards guard the preservation of 'good' data against data that is not 'good' enough.

5.3.3 Deleting for the sake of 'good' data

Data stewards, as they guide research data toward long-term preservation, are tasked with its negative definition through deletion. To illustrate how central the question of deletion is, a data steward provided me with this anecdote:

> *...some of the projects built up databases and then the people left their university and went to another university and the databases were here [...] What to do? You cannot destroy it because maybe they're very important for somebody, but you cannot reach the people [...] nobody will ask about this data anymore because people are spread around the world, not thinking about their data [from] twenty years ago. Now we have data policies, research data policies in the university which say you should archive and you should make it open and*

> *transparent. Maybe the next policy says you have to destroy all the data.* 7:43 ¶ 275 in DW_P1_231004.docx

With all of this data occupying precious material infrastructure, data stewards and repository staff have the onerous job of figuring out who owns it and what to do with it. In parallel to encouraging researchers to be mindful of preventing data loss, data stewards attempt to make researchers more comfortable with the active decision to delete data that is not destined for long-term preservation. Determining what data can be deleted is, like creating and archiving data (Choroszewicz, 2022), a relational and value-laden process. Through deliberation with and support of researchers through this process, data stewards guard against data becoming something unusable by being archived in an undesirable state ('fake' or 'badly documented'). This is, as ever, a process of affective work.

> *So, I try and see where they're at and what the anxieties are and what possibly likely very real constraints they're feeling professionally in relation to that and say, "Could we at least maybe say if I haven't published this data in x number of years, then I will put it in this archive and make it open," or something like that. Do we feel comfortable saying that at least and navigating it that way?* 6:72 ¶ 127 in DW_U4_230611.docx

Here, a data steward describes the first step in the journey toward deletion: querying how data that is lying around, unused and unarchived, can be published and made open for the possibility of reuse. The data steward must acknowledge that the looming shadow of publication is a very real stressor for researchers and that data producers might hold onto data in the hopes of it supporting subsequent publications (Sigl et al., 2020). Only when this concern has been cleared can the question of preservation or deletion begin. And when the answer is not preservation but deletion, the process of guiding data towards its final destination can be even more daunting for researchers, especially if RDM has not been a topic of discussion. Data stewards try to work with researchers to avoid uncomfortable situations at the end of a project, when decision-making is more limited and heightened emotions can kick in.

> *It's always the same thing, just thinking through these things beforehand and trying to figure out plan this out and not just float along and by the end of the project you're like, "I have gigabytes and gigabytes of data and I have to sort through it and find out what actually needs to be kept and what doesn't." Ideally, it doesn't come to that, although we know that it does*

> *happen and when it does happen, you can reach out to the data steward*. 8:74 ¶ 99 in DW_U2_230211.docx

This data steward recounts a familiar dilemma in the RDM universe: a researcher waits until the end of a project to approach a data steward for help. A researcher then admits they have 'gigabytes and gigabytes' of data that need to be parsed through, often under time-constrained conditions. Data stewards hope that, through broader exposure to RDM and targeted trainings, this panicked situation can be avoided in the future. But unfortunately, that ideal is not the current status quo. In reference to the laissez-faire attitude that can be taken by institutes, one data steward describes the incoherence in expectations for research data management:

> *On the one hand, they don't really care about the people having most of the data on their personal computers. But on the other hand, if it's on a backup infrastructure and published, then they are keeping it there forever. So, there's no real logic consistency between these ideas here and no explanation why that is the case*. 4:74 ¶ 207 in DW_U1_233110.docx

> *There, we see this problem of the data does not even get deleted when it's already preserved somewhere*. 4:73 ¶ 199 in DW_U1_233110.docx

In this case, the data steward expresses frustration that data is 'kept around forever', doubled on institute-owned or personal devices, when viable solutions for long-term preservation are already in use. In this case, it will require the data steward to engage with their prescribed role as a communicator, relying on a relationship of trust built through the effort of integrating into an institutional setting (see the data guide and data confidant) to nudge toward establishing norms that have lasting power. This infrastructuring work is not for the faint of heart. As Bowker and Starr (2000) argue, "Infrastructure does more than make work easier, faster, or, more efficient; it changes the very nature of what is understood by work" (p. 108). If, as also argued by Bowker and Starr (2000), infrastructural routines are conceptual problems, then the conceptual problems of data infrastructures often land in the laps of data stewards. They must then marshal their skills, experience, and sensitivity to implement more sustainable practices. In the case of deletion, this process has consequences not only for the interpersonal or affective dimension of a data stewards' integration into their working community, but also for the data they work with. One data steward describes the series of questions that stewards move through with students as they attempt to solve the conceptual problem of deletion:

> *I think [...] it's more that we remind students that their data can be deleted also. I mean, not if it underpins a publication or there's instances where you do not delete data, but letting them know that: Does this data underpin a publication? We save it. Is this data highly unique? We will probably save it. Did it cost a lot of money to generate, or time? Probably save it. Is it a complete usable data set that would be interesting to someone else? We'll probably save it. But also recognising that people do generate bits and bobs that were maybe part of a project that didn't get finished or you have some questions about the quality controls of the data, you're a little uncomfortable. It's okay to let things that go. I think for our students, they're more surprised by the concept that you can make the decision in good ethics and in good conscience to let stuff go as well*. 6:68 ¶ 115 in DW_U4_230611.docx

The task becomes exposing students and researchers to the idea that data, in 'good conscience', can be deleted, can be purposefully lost. If the list of parameters given by this data steward are not met by a specific data set and cannot be made to fulfill them through cleaning data sets or improving metadata, then the logical choice in the paradigm set up by the RDM community is to delete the data – to 'let go'. But this is not at all a comfortable proposition at first. One data steward even recalled that,

> *[...] my perception of that changed over the last year. When I started, my impression was we have to keep everything for as long as possible*. 4:68 ¶ 195 in DW_U1_233110.docx

This data steward, with previous experience as a researcher, came to professional data stewardship with notions of research data management carried over from their time in academic research and the tacit knowledge of data passed through that environment. Only during training as a professional data steward did the 'save everything' mentality shift to an awareness of deletion as a legitimate and even desirable choice. If "Universities and other research institutions are fully aware that research is becoming increasingly data-intensive ... and are therefore interested in retaining the value inherent in the vast amounts of research data," then it follows that the assumption guiding tacit data practices not under professional RDM influence would lean toward 'saving everything' (Blumesberger et al., 2021. P. 9). There is a revenue stream to be lost by deleting useful data, and it is in universities' best interest to work against this. Nevertheless, data stewards are aware that 'saving everything' is neither desirable nor possible, and argue for deleting data that does not hold up to questioning for quality and utility.

> *They relax a little bit when you remind them of all of that, that storing data isn't a zero-cost endeavor and there can be ecological costs, there can be running costs. Actually, again, the ethical choice and the conscious choice can be to let some things go and that I think seems to make them clicks a little bit, "Oh, it is all right, I'm not being bad by deleting this." Of course, we always tell them if something has legal or ethical implications that trumps everything else that we're talking about, you can't break the law and you never want to hurt somebody or something if it's vulnerable*. 6:69 ¶ 119 in DW_U4_230611.docx

Here, the argument linking unnecessary preservation with the material impacts of long-term data storage is used to support conscious deletion of research data. In contrast to the digital archive as dumpster, as elaborated by Hogan (2015), data stewards are aware of the connection between the storage of digital data and the material infrastructure that enables it and use this connection to ease researchers' minds when the time comes to let go of data that is not 'good' enough. Although researchers are the ultimate determinants of data's value, in the eyes of data stewards, there are nevertheless basic restraints that make data not 'good' enough to be stored long-term. In this world, researchers must come to the point when differentiations can be made between versions of data sets, critical and superfluous data, and when these differentiations can or should result in data deletion.

> *So, maybe once we have more of this cultural shift towards more responsible research practices, it's also going to become easier to give up on those old files because you're like, "Well, I know all my data is safely kept here. I can always access them, I can always reuse them. It's all legal and ethical the way I've done this work and I don't have to worry about it and I don't have to worry about my computer crashing or something*." 8:75 ¶ 103 in DW_U2_230211.docx

In this ideal world, RDM has become part of the basic enculturation of scientific researchers. Gone are the days of workshops and RDM as a subsidiary practice. In this vision, RDM is fully integrated into scientific practice, and researchers can be confident in the form in which their data live on in the long term. This is a vision free from worry, free from the regret that comes with data decisions wrongly made. Here, data is fit for reuse and perpetually accessible. This is the vision that enables deletion. If there are external storage options – like academic repositories – that enable the long-term preservation and publication of research data, then researchers can make assured decisions to delete that which does not meet the barriers for entry into a repository. Data stewards serve as the guide toward this vision of deletion as a peaceful step in the path toward 'good' data.

But, archiving data toward reuse is not a simple nor linear story (Borgman, 2012; Borgman, 2019; Edwards et al., 2011; Staunton et al. 2021; Green et al., 2023; Leonelli, 2018). It is fraught with friction, haunted by academic hierarchies, and epistemically debated. Despite data stewards' appreciation for novel and idiosyncratic data, the propulsive goal of secure archiving of 'good' data for eventual reuse does not grapple with the fundamental questions inherent to reuse. Rather, the intervention of data stewards toward preserving 'good' data and deleting 'bad' data reflects the personal commitments of data stewards to individual researchers and data objects in negotiation with the institutional and cultural demand for open, reusable data. This negotiation remains open as data confound strict standards for archiving and reuse, being technically incompatible or epistemically committed in ways that preclude reuse. As data stewardship holds an institutional position that is simultaneously unfixed and professionalizing, it struggles to fully incorporate deeper ontoepistemic questions of data. Nevertheless, it moves forward through more and more academic communities through its training and teaching in universities.

5.4 The data educator

As we have established, data stewardship as a concept and a practice is still under a process of active definition. Various professional and/or academic groups, such as FAIR Data Austria (Hasani-Mavriqi et al., 2022) and community-based organizations, such as the "Professionalising Data Stewardship" interest group of the Research Data Alliance (Research Data Alliance, n.d.) have pursued the cause of assembling data stewardship as concrete and coherent through workshops, reports, and forums. These venues for thinking about and discussing data stewardship are populated by people invested in the professional and ideological futures of data stewardship. While this formal attempt at meaning-making for data stewardship is ongoing, the practice of data stewardship by data stewards on the ground must move forward without a unified, international understanding of its contents and boundaries. In their daily work, data stewards reference the publicly available material that they can to determine a course of action and work together to create a professional environment that functions for them and the academic staff they service. One of the realms in which this bricolage approach can be located most clearly is in the mode of the data steward as data educator. This role has a twofold expression: 1) the data steward takes on the role of an educator of RDM norms and best practices in their interactions with students and researchers and 2) data stewards self- and community-educate around becoming a data steward. This section explores how data stewards educate, and towards what outcomes.

My first exposure to data stewardship was through the idea of teaching it. Through an interest in teaching and training across a variety of data practices, data stewardship first came to my attention because it is explicitly being taught, unlike many other data practices that remain tacit and more difficult to see. As mentioned in the methods section of this thesis, I engaged with the first strand of education around RDM through ethnographically-inflected attendance at RDM workshops held by the professional development office of a large Austrian research university. Unfortunately, access to professional data steward training courses was not possible. However, engagement with the promotion materials and online presence of data steward training courses and semi-structured interviews with designers and facilitators of a certificate course lent insight into the training of professional data stewards. The following sections will explore both of these realms of teaching and learning around data stewardship and RDM.

5.4.1 Teaching Research Data Management

If data stewardship is the solution to the lamentable waste of researchers' time and resources that Mons (2018) points to in his schema for data stewardship, FAIR, and open science, then it only makes sense that universities are taking it up as an explicit learning outcome. Even institutions carrying the most tacitly prestigious reputations frame research data management as an area for improvement and training. A brief survey of the RDM landscape of research universities reveals that data stewardship and RDM as guiding concepts or attitudes have not yet reached widespread awareness across university populations. Large Austrian universities are no exception to this status quo. To increase the exposure to and utilization of RDM concepts and skills, Austrian universities offer a variety of resources to researchers and students, including visual and textual guides, coursework, and workshops. These materials come from RDM offices or libraries, where research data management and/or data stewardship variably coexist as intersecting notions and practices that are contextual and non-standardized.

At one large Austrian research university, the RDM office was a separate entity, associated with, but not within the library's infrastructure. Training workshops in RDM were offered through the professional development portal of the university for staff and were conducted by facilitators with a variety of titles, including data stewards, repository data managers, and IT specialists. Workshops were primarily accessible to employed staff of the university and explicitly marketed toward PhD researchers. Some trainings were directly titled for and open only to populations of PhD students in given faculties. Nevertheless, there was enough flexibility and personal communication possible in the registration process that I was able to attend five such workshop sessions. All workshops were held via videocall and online learning platform. They started with a round of introductions from the participants, describing their projects and situation in the university. The workshops then began in earnest by asking “What are data?” or “What are your research data?” via workshop presentation slides. In social science and humanities workshops, it was clearly anticipated, with a dose of humor, that not all attendees present might consider that they have research data at all. However, it seemed that all attendees, across disciplines, did indeed consider that they produced or worked with research data.

5.4.1.1 The research ‘data life cycle’

Once it was established that every attendee did indeed bring experience with research data, the workshops were roughly structured around the ‘data life cycle’, a conceptual model ubiquitous across the RDM/data stewardship landscape. Here it is useful to pause and describe the life cycle in our attempt to grasp education around RDM. As it provides structure to workshops themselves and the sense of RDM that attendees take away from them, it is critical to notice how it operates as a register of value for data, a code for what makes data well-kept.

Despite its ubiquity, the ‘data life cycle’[4] varies across disciplines and applications. For instance, the Harvard Data Science Review defines the ‘data life cycle’ as a linear progression: “generation → collection → processing → storage → management → analysis → visualization → interpretation” (Wing, 2019). Accordingly, “*Data science is the study of extracting value from data*. ‘Value’ is subject to the interpretation by the end user and ‘extracting’ represents the work done in all phases of the *data life cycle*” [italics in original] (Wing, 2019, p. 2). This kind of model is explicitly oriented toward value production in data science. In contrast, The University of Vienna’s RDM office offers “planning → data collection → preparation/analysis → archiving publication → reuse” as its interpretation of the research ‘data life cycle’ (University of Vienna, n.d.). The Vienna

[4] The ‘data life cycle’ is also seen spelled as a formal noun, ‘Data Life Cycle’. This is more often used when such the model is framed as proprietary. I continue to write ‘data life cycle’ in single quotes to distinguish this term as coming from within RDM and data stewardship literature and in lowercase to refer to the general non-proprietary model.

University of Economics and Business suggests “Plan and Fund → Collect and Analyze → Preserve and Store → Publish and Share → Access and Reuse” as the ideal model (Vienna University of Economics and Business, 2024). The University of Graz, in a slightly different flavor, guides us through “Plan & Design → Collect & Capture → Collaborate & Analyse → Manage, Store & Preserve → Share & Publish → Discover, Reuse & Cite” as their version of the ‘data life cycle’ (University of Graz, n.d.).

There are at least semiotic, if not semantic, differences between these definitions, as well as among those offered by other research institutions around the globe. An in-depth analysis of the linguistic and functional differences between these definitions alone would indeed be intriguing but is unfortunately beyond the scope of this project.[5] But, what is key to note is that these definitions exist in a relatively similar way, even with the subtle variances between them. They are all available via the various RDM/data stewardship sections of university websites and are framed as educational resources which students and researchers can (and should) reference. Data stewards then point to these definitions in workshops and trainings as the standards by which good RDM is done. From these capacities, we can consider the ‘data life cycle’ as it is framed in the academic research context as a code by which RDM is assembled in collaboration with data stewards. Through this model, a certain concept of data travel to students of RDM that streamlines and simplifies friction. ‘data life cycle’ models attempt to display and transmit data’s construction in a more holistic way beyond the common understanding of data waiting to be ‘picked’ from epistemic air. They might even be expanded to highlight the skills, time and effort required to bring them into existence (Green et al., 2023; Pinel et al., 2020). However, as they currently function in the Austrian context, data life cycles segment the process of working with data, and data themselves, into discrete elements that do not represent the blurriness or temporal length between them. These blurry zones or timespans are where data frictions occur, and where those who work to resolve them come in. Whether it would be possible or desirable to represent the complex and networked affects and communities that collaborate to produce data in a simplified model is debatable. However, current ‘data life cycle’ models stop short of incorporating notions of data and the varieties of data practice that can travel across disciplines and contexts, as we will soon see. With this understanding of the data life cycle, we can now shift back into the RDM trainings that it guides.

[5] See Borgman (2019) for a discussion of various ‘data life cycle’ across disciplines. Borgman’s analysis, however, stays at the level of academic discipline and does not attend to possible institutional elements of the structure and content of ‘data life cycle’ models.

5.4.1.2 RDM workshops in the Austrian context

From the point of the introduction of 'data life cycle' visualizations, this notion guided the remainder of the training courses. In explaining this notion, trainers emphasized that good research data management "is a key part of good scientific practice." With customizations for discipline and specific interests of attendees, the 'data life cycle' is the conceptual model through which university research or support staff become familiar with data management as a whole. Policy, problems, specific skills, emotional reactions, all found their place hung along the specific iteration of the 'data life cycle' provided by the facilitators of RDM workshops. While the ubiquity of the 'data life cycle' as a heuristic device in teaching RDM might suggest that there is a consensus that this model reflects some kind of data reality. But data stewards and repository staff, both of whom facilitate RDM workshops in the Austrian context, offer a far more nuanced understanding of this model. For example, one data steward illustrated the flexibility of the 'data life cycle' in pedagogical practice:

> *When we do the [...] course, actually... the first activity we do is that we have placards with the names of all the different steps in the research data life cycle. We asked them to organize them in a way that makes sense to them, but we also give them a blank card where they can add steps to the research data cycle if they want. Again, it's totally flexible in how they want to use it*. 6:77 ¶ 111 in DW_U4_230611.docx

Here, the 'data life cycle' becomes less linear, less defined by unidirectional arrows. Students can, and are encouraged to, intervene in the modeling of data that frames RDM, as they might or should practice it. They are invited to integrate their discipline-specific experience of data work into the discussion of how best to manage data as an entity separate from their research projects.

Another data steward expressed the following:

> *It's maybe the life cycle—Well, somehow it shows that it starts again, this is good. But also I think the concept is not [good, implied] —because it's not always— it's still linear, it's one after the other, and then you start again. But in fact, it's not like that, it's like this [multi-directional hand motion]. You touch all the topics somewhere and it's simplified, of course, and it's better than to have this model. But I think it's also a bit problematic maybe.*
> 5:43 ¶ 71 in DW_U3_230311.docx

Here the feeling is that the ‘data life cycle’ is valuable as a teaching tool because it makes clear that reuse underpins the cycle – “*it starts again, this is good*.” However, this response also points to the incommensurability between the model and the lived data practices of researchers. On this level, there is the need to balance the necessity of simplification for the purpose of knowledge communication – teaching – and the tension that exists between this heuristic simplification and the world of practice.

In the context of RDM workshops in large Austrian research universities, the ‘data life cycle’ guided the structure and content of learning research data management. The connection between the production and care for data and the loftier ethos of good science was the thrust and logical underpinning for the subsequent units of the workshops. From the planning phase (including the DMP and file naming systems) to the reuse phase (working with repositories, metadata, and licensing) attendees were exposed to an array of discrete and connected practices that could embody ‘good scientific practice’ across the Data Life Cycle. And in the question-and-answer sessions that generally followed the forward-facing teaching units, attendees found this model sufficient enough for it to elicit no explicit question or critique. However, attendees nevertheless struggled to conceptualize their work and the data they worked with into this model. One student working with a specific software as the object of their research project questioned such an object would be archived long term. Is a version of a software data? If so, which part? At what time? Where and how can this be stored? Such questions were beyond the scope of a general introductory workshop in RDM, but facilitators did their best to point the attendee in the right direction, perhaps towards a software repository. However, these workshops were not the venue in which to debate the relational and shifting status of data as evidence or epistemic object, no matter if those deeper debates impacted the pragmatic questions of when, where, and how to store data.

5.4.1.3 The burden of learning RDM

In general, attendees interacted with RDM workshops as a venue in which to clarify the concerns and conundrums that had already arisen in their research projects. The Q + A served as a sounding board for insecurities and often too-specific queries to facilitators. From the national-legal boundaries and dark histories of archeological data collection to the seemingly dubious benefit of data journals to the early-career publication progress, attendees brought in concerns with data that ran the gamut of academic life. By and large PhD or early-career researchers, again because of explicit targeting of that academic population, these workshop attendees were interested and collaborative during group exercises. They have a reason – increasing demand by funding institutions for the explication of RDM practices – to be interested in gaining competencies in RDM, and seemed ready and willing to take on what facilitators offered in the workshops.

However, the alternate facet to this motivation is the anxiousness of an academic career in many ways permeated by precarity. Attendees mentioned feeling the workshops necessary because their superiors or PIs were unfamiliar with RDM practices in general, a dynamic summed up by a respondent associated with a data repository:

> *I think it's a generational thing as well. More senior scholars might not be aware of them, thus not disseminating that knowledge or the lack of knowledge towards their students.* 1:121 ¶ 23 in DW_A1_230928.docx

This 'generational thing' and its accompanying anxieties touch broader trends in academic life that suggest a general shift toward precarity and insecurity. If a senior scholar finds themselves tasked with elaborating their project's RDM practices in a DMP, but they have no clue what research data management is, it might seem entirely unclear by whom, when, and how this gap should be filled. In their capacity as institutional guides, data stewards indeed step in here, as often as they can, to advise and guide researchers across generations toward consistent RDM practices. If a young researcher has the good fortune to become exposed to RDM early in their career, they might be able to avoid stressful situations in which the language of RDM and its implementation in mandatory documentation must be achieved in crunched time spans or under pressure. However, this does not seem to be the case. As one repository respondent with previous research experience explained,

> *I mainly learned how to analyze data, but I'd never heard of the term research data management at all....In general, I learned a lot of this research data management stuff 'by doing' through the work here, through the work in the projects*. 2:37 ¶ 19 in DW_A2_231016.docx

Another respondent painted a clear picture of the conditions in which early-career researchers might be expected to learn and execute RDM:

During that time, obviously I learned about a lot about data management by doing it and a lot by doing it badly. I ran into that problem or what I felt like was doing a PhD was hitting a moving target a lot and that target wasn't only moving, but there were so many of them. 1:48 ¶ 11 in DW_A1_230928.docx

Active PhDs attending RDM workshops and professionals in the sphere of RDM alike expressed that learning RDM was no small feat. Even the realization that RDM exists at all and that it can and/or should be practiced in an academic context is not obvious. As PhDs are currently starting their careers in an already-dataified (or datafying) context, it follows that they be the members of a project or department that would invest time and energy into RDM. However this might benefit their future careers, it is still an additional task that will consume precious energy and time, both of which are already strained.

The situation may not be entirely dire after all. According to one respondent, younger PIs reaching senior positions might have a refreshing position on RDM and its incorporation into their research design:

There's now one young group leader who had just started as a group leader and she has a quite small team....She also asked me about the data management plan, how to set that up, and we discussed also that maybe the data management plan in her case then could be also a document for her team to know how to handle data within their team. [That] *it's not an administrative document that exists and she's familiar with it but nobody else. But that they really use this as a leading document. Let's see how that works out in the future*. 4:79 ¶ 63 in DW_U1_233110.docx

It seems that the burden of the 'generational difference' might become less present in the future, a future in which RDM would ideally be taught at a much earlier phase of the research trajectory, even in the bachelor.

Also, the university has other priorities and that's totally fair and valid. (...) Honestly, what would also be a very good way to go - as to something that we talked about now before - is to develop those skills as part of all the curricula for all degrees, so that once you know it from the beginning of your degree, you know how to do this stuff properly. 8:31 ¶ 47 in DW_U2_230211.docx

Pushing exposure to RDM earlier in the university career, potentially as early as possible, would allow for some relief from future stressors that currently fall to early-career researchers. Furthermore, disciplinary differences across practices could perhaps be handled simultaneously as students learn the foundations of their disciplines and fields.

5.4.1.4 The future of RDM education

In the current approach of targeting young researchers for training, the scheme for RDM education understands RDM as part of 'good scientific practice' that should be learned alongside discipline-specific methods and the grander scientific method. The 'data life cycle'provides a useful heuristic through which to expose RDM learners to the variety of tasks, practices, and ideas that can be subsumed within RDM. However, respondents who teach RDM workshops gave a complex understanding of the 'data life cycle'that exists beyond strict linear phases. Understanding that disciplinary differences, technological barriers, and other unexpected complexities confound this notion, data stewards – in their education approach – make space for this code to be flexible. If it becomes too rigid, then students of RDM might become frustrated or lose interest, and the threat of disassembly looms. To that end, RD workshop facilitators did their best to keep it loose and answer questions to the best of their knowledge. Until RDM is more commonplace in research life, the 'data life cycle'will continue to be an unfamiliar, if obnoxious, teaching tool for RDM education. But, my interlocutors described a world in which it might not be necessary. To take the liberty of summing up an idealized future for training RDM on behalf of those who really do it, RDM would be best understood and taken up before it is needed from an institutional or professional perspective so that learners can come to know what data means for them and how best to treat it within their disciplinary specialty. However, this would require a whole lot more data stewards to do a whole lot more training. This need brings us to our next inquiry: how to train the trainers.

5.4.2 Training the trainer: becoming a data steward

Thus far I have elaborated how RDM and data stewardship are taught – in what formats, with what concepts, and to what reception from students. But perhaps we have put the cart before the horse. Before RDM can be taught to students and researchers, there must be someone to do the teaching. If a data steward is the one to do this teaching, as is common in the Austrian context, then we must try to understand how a data steward comes to be a data steward. While this 'becoming' varies across contexts, the following will focus on the University of Vienna's "Data Steward Certificate Course," which provides professional training in data stewardship to prospective data stewards across Austrian research institutions (Bardel, 2023).

5.4.2.1 The Data Steward Certificate Course: the beginning

The "Data Steward Certificate Course" began in 2021 with its first cohort of students enrolling in 2022 and finishing the two-term program worth 15 ECTS in the spring of 2023 (Online Info Event slides). The certificate program grounds its data stewardship education across three pillars: "competence acquisition," "peer-to-peer learning," and "community." Across these pillars hang the learning modules of the course, which cover content from RDM and Open science, the FAIR principles, and project management. With instruction from IT to data science to training and requirements engineering, students cover much ground in the academic year. The course begins in presence in Vienna but is primarily conducted remotely using well-known e-learning tools. In terms of target audience, the training course is marketed towards researchers, PhD students, and research support staff ("such as library, IT..."). The course requires a university entrance qualification, apprenticeship and work experience, or a Master's or diploma degree, in addition to a B2 level of English language proficiency. In reporting on the first cohort of students, the program organizers emphasize the national diversity of students, with representation from ten countries. The 2022/23 cohort also reportedly contained disciplinary representation across the academic spectrum, from arts to natural sciences.

This training course is the first of its kind in the Austrian context. According to a member of the first cohort,

> *"There was such excitement in the room. On the first day you could tell everyone, "We're going to be the very first data stewards, how amazing. This is the new frontier." So, that was really cool. It was a lot of excitement and energy."* 8:47 ¶ 67 in DW_U2_230211.docx

With its emphasis on peer-to-peer learning and community building, the course offered participants a venue through which to explore data stewardship and all that this realm could contain for them. Through diverse modules, a small cohort of colleagues (roughly 25), and group work, participants could come to know data stewardship as a professional possibility. In the past, this specific focus on data stewardship as a set of skills relating to the long-term care of data, let alone a distinct career path, seems to have been missing from the German-speaking context:

> *But back when I was doing the degree [library and information science], if I remember correctly, we didn't even use the term 'data stewardship' at all, which doesn't necessarily mean we weren't talking about data stewardship. There is no real consensus in the international community as to what is research data management, what is data stewardship*. 8:56 ¶ 19 in DW_U2_230211.docx

If, as this respondent argues, there really is no consensus in the international community concerning the distinction between research data management and data stewardship, or what could be contained within either term, then how to begin teaching data stewardship? How does an institution, a community, or a person train towards a position that escapes definition? Well, in the Austrian context, the development of the training course and the definition of the role of data steward went hand in hand:

> *Working together with five other universities in Austria, we co-developed this curriculum, and this was a very open process where we were brainstorming what are the topics that data stewards need to know about, what are the skills they need, what are the things that would be helpful to them but maybe we don't need to train.* 8:77 ¶ 51 in DW_U2_230211.docx

With demand for the skills and presence of data stewards already there, the University of Vienna was reportedly 'jumping at the chance' to develop and run the certificate course. Nevertheless, the uncemented nature of data stewardship – as a term and a practice – put the training course in the position of defining data stewardship *through* training it. From the selection of modules for inclusion to the relative weight that different modules carried, the course was a bricolage of available resources, structures, and locally debated decisions. Representatives from major universities across Austria came together in a consortium to facilitate the exchange of knowledge around FAIR, Open science, and data management. One stated goal of this consortium project was to produce a plan for the implementation of these data policies and values in the form of training and support services. Here, the University of Vienna's certificate program can serve as one concrete output from this goal. While available materials and examples from other institutions, such as TU Delft, TH Köln, the U Bremen Research Alliance, and State Initiative for

RDM in North-Rhein Westphalia (Bardel et al. 2023), the design and execution of curriculum at the University of Vienna were necessarily situated within its institutional boundaries and local context. As a diverse institution encompassing disciplines across the arts, humanities, social sciences, and natural sciences (however those loose categorizations can be delineated), the university's institutional identity and history impact how data stewardship can be assembled.

5.4.2.2 Defining data stewardship in professional practice

Situated within a multidisciplinary institution, the training program gives more of a diverse definition to data stewardship than what might be found at a technical university oriented exclusively toward STEM/MINT subjects. The training program references both the broader LIBER (Ligue des Bibliothèques Européenes de Recherche – Association of European Research Libraries, 2020) call for embedded data stewards that connect researchers to various resources and services across the university and aligns with FAIR Data Austria's third model of decentralized data stewardship in which data stewards are situated within their disciplinary fields of expertise (FAIR Data Austria, "Workshop: Tasks and Profile of a Data Steward, 2020). Thus, the University of Vienna's certificate course kicks off with the understanding of data stewardship as deeply entwined with diverse discipline-specific knowledge. Working with this embedded model means that data stewards – in an ideal world – would work from a pre-existing understanding of the needs, idiosyncrasies, and common data practices of their field.

In the context of a massive multi-discipline university such as the University of Vienna, this approach makes sense. However, a training course that aims to produce professional data stewards across disciplines cannot possibly focus on the discipline-specific knowledge that might be necessary for a deep embedding in a given institute or department. Rather, the certificate course focuses on those skills and competencies that can unify data stewardship across disciplines. These are: "Networking and collaborating (inter)nationally," "Providing advice and support to researchers," "Conducting needs-assessments and requirements engineering," "Bridging the gap between researchers and support staff," and "Designing and delivering training" (Kalová, 2023). Thus, what the certificate program can teach is not the fine-grained knowledge of particular data types and data work that uphold fields and sub-fields, but the 'softer' skills, the organizational sensibilities that can make data infrastructurally legible at a broader scale (Grugulis and Vincent, 2009; Gabrielsen, 2023). The multiplicity of data in academic research means that data stewardship can be linked to no narrow definitions of working with data. Rather, it must interact with data multiplicity through flexibility, and by building an institutional territory through which data and data makers can move.

In the certificate course, emerging data stewards can expect to become familiar with the established lines of discussion within RDM, basic licensing, and intellectual property law, Open science, data security, and more. From a 'softer' point of view, they will embody the personal and social characteristics that make for a good "translator": patience, empathy, and curiosity, to name just a few. However, once finished with the course and ideally embedded into a department or institute in which they have field-relevant experience, the process of learning to be a data steward does not end:

> *[...] there's also a lot of independent self-study time of reading websites from the university, reading things from outside the university, teaching yourself about a lot of research data management concepts*. 6:28 ¶ 27 in DW_U4_230611.docx

As this reflection makes clear, even employed data stewards must continue to grapple with data stewardship as a mutable, multiple practice. As the status quo stands, data stewardship is a profession that is still under active definition and assembly. As data stewards proliferate, the specific institutional constellations in which they work will continue to place demands on how data stewards interpret and express data stewardship. Nevertheless, data stewards themselves, at this moment, exercise a relatively high degree of agency in the formation of this multiple practice. Working with the structural boundaries of the embedded model of data stewardship, the prior disciplinary experience of data stewards, and their close working relationship with researchers allows for data stewards to practice as they see fit. In training data stewards, this prior experience is highly valued and incorporated into notions of practice.

5.4.2.3 Producing professional data stewards

As prospective data stewards work their way through the training course, their prior experience is both required and deeply appreciated:

> *They get to keep in touch while they're all working towards the same goal and at the same time, they're not being treated as children or as 18-year-old undergraduate students, but they're treated with respect where we show them, 'We appreciate your expertise, you're*

> *not coming here to us as students, you're coming to us as experts on your field and we want you to be able to infuse your experience into this program and to be able to share and have free discussions with your colleagues and learn in terms of peer-to-peer learning a lot.'* 8:51 ¶ 67 in DW_U2_230211.docx

Here the data steward emerges as a separate category of academic professional that is not just an 'undergraduate', assumed to be in the position of intaking knowledge in a passive way, but a fully-fledged 'expert' in the contrasting position of actively contributing discipline-specific knowledge to the course, their cohort, and eventually to the institution in which they will practice data stewardship. This quote belies a heavy dose of boundary work (Gieryn, 1983) in which the data steward certificate program fights for the epistemic legitimacy of data stewards by asserting their pre-existing expertise. These are not intellectual children, but intellectual adults with a requisite set of experiences that coalesce into expertise. Even as data stewardship exists under the 'meta-science' of LIS (Palmer and Cragin, 2023) and often as a mediatory intervention, the data steward certificate course seeks to form data stewards as legitimate members of an academic epistemic community. Here, data stewardship's unique claim to epistemic authority comes through the blending of already-achieved – and identifiable – disciplinary expertise with the 'softer' skills that the certificate course provides.

This essential prior expertise serves to assist the professional data steward when becoming an embedded member of a disciplinary institution:

> *I have the perception that people don't really work with the data stewards as closely if they're not embedded at their faculties or centres. It's also because they have this very, how to say it, strong identity of the centre and belonging together. For them, it's very important that I'm not one of these library people, but that I'm one of them.* 4:52 ¶ 143 in DW_U1_233110.docx

This data steward needs to be 'one of them' to function in their institutional context. Creating a sense of trust (see the data confidant) relies on the data steward being an intellectual 'adult', a member of an existing epistemic community. Only then can the data steward speak with authority on disciplinary matters of data. Even if an embedded data steward has experience that does not exactly match the research carried out by a given institution, the principle still follows.

I have wet lab experience, but then because of my background I have a diversity of other disciplines that I've touched on. So, it set me up really well to be interested in what a lot of people are doing across discipline divides. 6:5 ¶ 3 in DW_U4_230611.docx

Despite this data steward's diverse interests, they were embedded in a natural science institute. Here the divide between the 'natural' or 'hard' sciences and the 'social' or 'soft' sciences carries weight. Even if a data steward's experience diverges from their eventual disciplinary home (experience in cultural studies v. embedding in history, for example), expertise in either path of this divergent academic road matters when the proverbial rubber hits the road in working with the needs and habitus of different academic communities.

I think the biggest challenge is to keep up with all the developments in the research field you're part of or that you're there for and also all the developments in research data management. I think that's maybe for data stewards is one of the biggest challenges and those establishing them as part of all the services that we offer. 4:76 ¶ 247 in DW_U1_233110.docx

In the process of this research, a data steward that jumped the natural/social divide was nowhere to be found. However, this is not merely a matter of quantitative/qualitative methods and the data they produce. After all, the social sciences and digital humanities heavily rely on quantitative, statistical methods, and data visualization to produce their final intellectual products. As one respondent emphasized,

But I do think that the social sciences are especially hard because the variety of data generation processes is so large. 1:86 ¶ 43 in DW_A1_230928.docx

This respondent, though not breaking the natural/social divide, needed to get accustomed to new data formats in order to best work with social scientists.

That is something that is exclusively used by sociologists for some reason. I have never worked with SPSS in my own education, (...) let's say it that way. But we still use that. 1:105 ¶ 70 in DW_A1_230928.docx

Here the complexity and diversity of social science data demands that data stewards undertake that 'self-learning' that participants in the data steward certificate course pointed to. This is not only a necessity in the natural sciences, in which disciplines, sub-disciplines, fields, and sub-fields might develop their own data processing software and work with disciplinary data types. While perhaps not on the same scale, the social sciences and humanities also produce diverse data. Nevertheless, data stewardship – even as an in-between, or supra intervention – does not manage to bridge the natural/social divide. Epistemic authority remains bifurcated as data stewards integrate into the daily processes of researchers and attempt to gain their trust.

5.4.3 The dual educator

In this chapter, we have explored the two modes in which data stewards inhabit an educational role: in teaching RDM to the wider academic population and in training prospective professional data stewards. Perhaps it is obvious that a data steward would take up the role of educator while situated in an educational institution. Perhaps it is obvious that data stewards should train their own 'kind'. But the story might be more particular than that. When the data steward accesses an educational mode, education is a game of bricolage (Varenne and Koyama, 2011). If data stewardship is a moving target, then educating toward it necessarily requires tools with different scopes, ranges, and levels of precision. The choices made and materials forged through teaching and training data stewardship exist in the tension between the agentic decision-making of data stewards and the constraints of institutional norms and resources, not to mention the intervention of learners.

Varenne and Koyoma (2011) define education this way: "the ensemble of the concrete, local activities that people conduct in real-time when they are in 'communities of intelligence' as a polity figuring out their exact present conditions, and what to do next" (p. 11). Data stewards and the RDM network grant that they are most certainly attempting to figure out what exactly data stewardship is. Definitions posted on university websites might seem self-explanatory enough to justify its institutional presence. But, it is data stewards – through the process of educating – that do the partial pulling together of disparate things (notions of data, confusions, disciplinary divergences, institutional norms, community-building, frustrations) and make data stewardship coherent enough to communicate. While this work is necessary for the institutional implementation of openness, FAIR, and reuse, stewards do not find teaching RDM to be a simple task. Relaying the practices and principles of data stewardship seems too general for some learners, to specific for others, and too late for even early-career researchers to implement rapidly. Despite these doubts, it is clear that the work of making data stewardship is indeed being done. In Varenne and Koyama's words, "Everyone produces culture out of their ignorance and with the stuff they find around them" (2011, p. 12). This is not to argue that data stewards are ignorant or that their process of teaching is defined by some lack of information. On the contrary, what we are seeing is the assemblage of a notion and a practice in real-time, in process. As institutional complexity and precarity loom, "their work produces culture by deeply inscribing instructions about what to do or not do to accomplish a task that no one ever had to accomplish *heretofore*" (Varenne and Koyama, 2011, p. 9). As data proliferates and demands our care, data stewards take on the role of attempting to territorialize the instructions by which we might care for our data.

5.5 The data confidant

In this final section, we will explore how data stewards articulate and work with the complex affects that accompany data through academic research. This mode, the data confidant, can be seen as cumulative, or weaving through the modes we have previously seen. Rather than a set of concrete practices of technical problem-solving or pedagogical approach, the data confidant represents a quality of engagement that centers the sensory and emotional elements that assemble data. Data stewards draw on their personal histories and diverse affective registers to maneuver and facilitate data movement. Following Murphy's (2015) notion of care as making "technoscience accountable to its world" (p. 732), I understand care in the context of data stewardship as an acknowledgment of the rich affective and sensory worlds that bring data into being, in whatever assembled form it may take. In this chapter, we will explore this kind of care through the mode of the data conduit. The mode of the data conduit works with the affective, emotional, and sensory components that assemble research data. These components are not always 'positive' and reflect a data world 'already violated' (Murphy, 2015). As researchers make attempts to care for data within the frameworks of RDM and data stewardship, data friction (Edwards 2010; Bates, 2017) inevitably arises in multiform ways, bringing engendering multiform affective and emotional states. As the data confidant, data stewards work with the accounts of these affective states to resolve data friction and begin the process of building data worlds differently.

5.5.1 From a 'beam of pressure' to a relationship of trust

In our conversations, my interlocutors shared their experiences maneuvering the varying affective states of the researchers with whom they work. These states are framed as being linked to the stressful, pressurized conditions of work in the contemporary university. To best work with these affective conditions, data stewards seek to establish a trustful relationship with researchers, as alluded to in previous chapters. They do this by emphasizing that their role is not a penal one with an aim to subject researchers to assessments of their data practice.

> *I always stressed that I'm not the data management police, and I'm not here to enforce a policy or something*. 5:67 ¶ 59 in DW_U3_230311.docx

Not being the 'data management police' is an essential negative definition. Without it, researchers might feel yet another source of pressure coming from the RDM demands of funding bodies and university in the form of a DMP, or other RDM practice.

> *So, they're like, "I don't know, do I have to write another research proposal?" [the answer is] No, this is what they're asking you [in the DMP], here's how we answer these kinds of questions in a productive way. And peppered in that is also getting them to shift their thinking about it because, again, this stressed out researcher brain is like, "Oh, God, it's another bureaucratic thing I have to do." I talk with them about ways that we can shift that a little bit to make this a document that actually serves them…* 6:22 ¶ 19 in DW_U4_230611.docx

In this quote, the data steward frames themselves not only as an 'air traffic controller' that guides the researcher and their data through the data infrastructures of the university but as someone who relieves stress by reframing RDM as something that can 'serve' the researcher, instead of stressing them. This is a subtle process of understanding the current placement of the researcher in their career, their research process, or their specific project.

> *There is understanding how academic publishing works, for example, and understanding how research groups tend to operate internally and understanding where some of the tension points in data collection are, where your data is either useful or useless […].* 6:11 ¶ 7 in DW_U4_230611.docx

When the data produced by a researcher can be deemed ‘useful’ or ‘useless’, we are indeed dealing with a pressurized environment. Early career researchers especially feel publication pressure as they attempt to progress through an academic career (Sigl et al., 2020). Given the importance currently granted to producing reusable and FAIR data, coming to the end of a project with too much ‘useless’ data would indeed be a sensitive point for a researcher of any career stage. Another data steward framed the imposition of RDM standards and practices on the research process as ‘yet another beam of pressure’.

> *So, there’s this multitude of requirements that you should meet and I think it’s really hard for good research data management practices to not be yet another beam of pressure pointing at you.* 1:68 ¶ 23 in DW_A1_230928.docx

Yet another data steward locates the pressure caused by RDM and data stewardship in the lack of incentive structure offered by research institutions to take up these normative standards for data practice (Leonelli, 2015).

> *Currently, we’re having no incentive structure. It’s more goodwill and you’re getting paid by tax money. So, it’s more an ethical obligation to do it or [it is] just enforced by the publishers. Then there are no incentives. It’s just their [the researchers] duty to do it.* 4:48 ¶ 115 in DW_U1_233110.docx

How do data stewards dissipate this ‘beam of pressure’ that carries no incentive in their interactions with researchers? By reframing RDM policy – and themselves as arbiters of it – in terms of help instead of harm.

> *But I’m here to help them and [...] our policy also says it codifies support [...] even in the first few paragraphs. [respondent emphasizes this] Because I know that everything that comes from a central place or from above is met with huge distress*. 5:33 ¶ 59 in DW_U3_230311.docx

In the face of a possible perception of RDM policy as centralized and top-down, data stewards position themselves to defuse the potential threat of increased ‘bureaucratic’ labor causing RDM and data stewardship to transform, in researchers’ perceptions, from a way to make data secure and reusable into a burden. For the data confidant, the goal becomes to care, perhaps in a traditional feminist reproductive sense (Gabrielsen, 2023), for researchers and their data. If we are to understand universities as data-producing institutions with the explicit goal of fostering and encouraging the production of data, then data stewards, in the mode of data confidants, indeed do the "productive doings that support livable relationalities" (de la Bellacasa, 2011, p. 93). While data stewards employed at universities might not frame their work as primarily enabling the data orientation of contemporary universities, their work is nevertheless critical to how data, in all its political, ideological, and material scales, assembles in university spaces. Theirs is a territorializing work. When the specific demands of data become unruly, frustrating, and disruptive, when the ideas and sensations around data become "connected as in a *delirium*" (de Landa, 2016, p. 27), instead of remaining within the boundaries of RDM policy and disciplinary norms, data stewards intervene to bring affective data states down to earth.

5.5.2 From data shame to data confidence

But, data stewards would ultimately like to reach a place where this territorializing work is less necessary. This would be a world in which RDM and data stewardship exist earlier in the academic career and research process, thus demanding less care to repair affective fractures.

> *I can answer their little questions that are easy and that we can build good enough relationships with each other that when we enter into a place where we have to have a more challenging conversation or we have to deal with something that's more complicated, we have a mutual relationship, we have a mutual respect, we have a mutual understanding of one another, and a mutual trust where we can have that conversation productively. Again, that's my hope for the future and why I hope this is long term.* 6:2 ¶ 87 in DW_U4_230611.docx

Data stewards care for data not only by dissipating punctuations of stress to the degree they can but also by digging into deeper affective relations with data that stand to endanger it and the researcher. These are feelings of shame, embarrassment, and insecurity around data practice.

> *I think too—I always like to tell, especially the PhD students, but everybody, that the research data management team is a judgment-free zone. So, whatever your problem is with your data or whatever, you can come and talk to us about it and we're going to do the*

> *best we can to help you. But we can't help you unless you tell us what's going on.* 6:51 ¶ 51 in DW_U4_230611.docx

Even in the context of training professional data stewards, this movement away from shame applies. In training prospective data stewards with backgrounds in academic research and administration, it is necessary to dispel any lingering notions of shame in service of caring for data in the most robust way possible. This applies to working with a dearth of technical skills related to data stewardship, and the possibility of gaining those skills in a judgement free environment.

> *You need to get in there and try things out a little bit yourself in a safe environment where nobody's really judging you for it and where it's not your end goal. It's not to be a developer yourself.* 8:61 ¶ 83 in DW_U2_230211.docx

When the immediate stresses of working academically with data become internalized, become a set of deeper affects that then express in reactions to problems related to data, data stewards describe building – or aiming to build – a working environment in which these affective states are not necessary for functioning with data. Researchers can approach data stewards to ask questions that might elicit feelings of shame around a lack of knowledge, a mistake with research data, or a misunderstanding of what the expectations from RDM policies and DMPs could possibly be. In these interactions, data stewards marshal their own recollections of their experiences in research to interpret and act upon the affective relations present in their interactions with researchers. This prior experience is a key factor in the legibility of affective exchanges around research data practices and relates to the value of the embedded model of data stewardship.

> *I think it's good to have this previous idea of how these problems feel and how they are. I think that's definitely a benefit of instead of going into with a very theoretical research data management background and trying to solve these problems* 4:8 ¶ 15 in DW_U1_233110.docx

Knowing how problems 'feel' and 'how they are' is not part of the explicit curriculum of the data stewardship certificate program or the definitional literature around data stewardship. This is something entirely different from the communicational bridging or 'soft skills' that are laid out as formal characteristics of data stewards. Nevertheless, this tacit knowledge is critical when dealing with the affective relations in which data participate. Data are not only products and objects of research practice, but carry personal meaning in a myriad of ways, from intense attachment to abject fear. Without prior research experience, data stewards might be unable to 'see' into this realm of data and counsel researchers through it. When data stewards reach for other modes, like the data guardian or educator, they carry their tacit understanding of academic research into a vision for the future of academic data work that is more caring for data and researchers, one in which worry dissipates in favor of confidence.

> *So, maybe once we have more of this cultural shift towards more responsible research practices, it's also going to become easier to give up on those old files because you're like, "Well, I know all my data is safely kept here. I can always access them; I can always reuse them. It's all legal and ethical the way I've done this work and I don't have to worry about it and I don't have to worry about my computer crashing or something."* 8:75 ¶ 103 in DW_U2_230211.docx

In the meantime, before this more secure future has materialized through the implementation and further infrastructuring of RDM and data stewardship, there is an affective in-between that can ameliorate friction-ful relations with data: recognition and support.

> *My hope is that reaction to stressed-out people and taxed people will be a ripple effect. I know that's very grandiose and I don't expect to suddenly change research environment, but there have been a lot of times in my professional life [...] where you wish someone would have recognized things were a lot right now [then] and you had really wished that someone had been like, "Actually, okay, I can't do all this for you, but I see this one thing and I can handle that one to make it a little bit better and a little bit easier."* 6:76 ¶ 143 in DW_U4_230611.docx

This interlocutor's response is revelatory as it balances acknowledging the entrenchment of contemporary research conditions while daring to hope for something else, perhaps an environment in which stress and the taxation of personal resources do not characterize data relations. This response weaves together personal history with future hopes as it summarizes what the data confidant can do: recognize the affective relations at play in data practice through trustful dialogue and identify concrete spaces of intervention that might bring relief. The same interlocutor pondered how this mode could be interpreted and valued on a broader scale, still with the hope that the personal importance of this mode translates to other participants in the practice and process of data stewardship.

> *I'm very curious to see what the value of those interactions with people is. I hope it's good and it's something that I guess is personally important to me. [...] We'll see how it goes, but hopefully it's a good thing*. 6:76 ¶ 143 in DW_U4_230611.docx

5.5.3 A caring confidant

In the role of the data confidant, data stewardship professionals care for data and researchers in a way that is both foundational and cumulative. By attempting to ameliorate the 'beam of pressure' locked onto researchers in contemporary academic conditions, data stewards see a way forward in the RDM world. The incorporation of RDM practices into research practice will become the norm, but it won't be characterized by data 'policing'. Rather, this will be an environment in which researchers can exist in some version of harmony with their data. DMPs will be shared guiding documents instead of a dreaded bureaucratic assignment on the way to a project grant. Moments of confusion and shame can be avoided altogether when RDM practices become second nature. Whether this data future can or should come to fruition stands to be seen.

Till then, data stewardship professionals can gather the strands of their prior disciplinary experience together to offer researchers a space in which embarrassment, doubt, and stress can – at least partially – dissipate in relation to data. Caring for researchers and their data performs a crucial role: it feeds back into the enactment of the other four modes that I propose in this project. In this sense, it is both foundational to and a product of these other modes. Without a baseline of trust built through experience and dialogue, problems around data storage and deletion might never be detected. They might languish in the web of daily stressors that researchers balance in their quotidian academic life. In wielding disciplinary experience, data stewardship professionals can best argue their relevance in disciplinary-specific institutional settings. Care might be what tips data stewardship professionals off that a mode-shift to the guide, guardian or educator might be necessary in a given moment working with data.

In these ways, the data confidant fits with Puig de la Bellacasa's notion of care as,

> "material and affective tasks related to communication, the production of sociability, and capacity of affect 'without which our lives do not work out,' the complexity of which makes them difficult to value, to reduce to a schedule, or to enclose in fixed tasks that 'start here and end there' (López Gil, 2007)" (2011, p. 93).

Indeed, the formal responsibilities or characterizations of data stewards explicitly point to a communicative or 'bridging' function. But what can be attributed to this communicative process is beyond what is framed in data stewardship's guiding policies or vision statements. Rather, this world is more characterized by the frictions and fractures that Edwards (2010), Bates (2017), and Murphy (2015) draw our attention to. In that complex affective states around data stand to disrupt data's proper planning, collection, analysis, and archiving, these states stand to decode the values that make data a coherent notion in the first place. On its face, care in the form of data stewardship might seem positive, restorative, even working towards harmony. But I see this work as constitutive, rather than communicative. By bridging between researchers, RDM policies and practices, and institutional resources – by caring – data stewardship professionals do constructive work of making data a legible notion despite its frictions. When the diverse materializations of data vibrate to deterritorialize visions of smooth flow toward reuse, this care comes in to pull the pieces back together.

6. Discussion

At this juncture, we have seen the data steward in many different lights, assembled from institutional situation, individual practice, community ideals, and more. The modes of the data guide, data guardian, data educator, and data confidant serve to display the assembled nature of data stewardship as a professionalizing practice, itself an assemblage that contributes to the multiple assembly of data. This topological approach makes clear the moments when data stewards interact with the coding (De Landa, 2016) of data from the RDM community, universities, and broader cultural narratives. It also helps to make visible the territorializing labor that data stewardship does to make research data itself a coherent assemblage. Running through these modes, however, there are some key cross-cutting elements that deserve attention and draw data stewardship into deeper waters of data.

6.1 Explaining data stewardship through metaphor

Across the modes presented thus far in the empirical portion of this project, my interlocutors illustrated their work using specific language. In reading the empirical chapter, the following terms might have stuck out as unique in the responses of my interview partners: ‘beam of pressure’, ‘grey zone of responsibility’, ‘air traffic controller’, ‘data management police’, ‘auxiliary brain’, and ‘the new frontier’. These framing phrases each bring a distinct dimension to the discussion of RDM and data stewardship. Here we will pause and examine the metaphoric framings of data stewardship as they operate to contextualize this emerging profession as a practice fraught with friction (Edwards, 2010; Bates, 2017). These metaphors point to the “special expressive components in an assemblage in fixing the identity of a whole” (De Landa, 2016, p. 22).

To begin, we enter a ‘new frontier’. This description appeared in a conversation about the training of professional data stewards in which newly qualified stewards celebrated their completion of the certificate program and their membership in the very first cohort of graduates. The professional world into which they would enter would be ‘new’ in that professional data stewardship as such, with certification from a tertiary educational institution, had not existed before. With the expansion of digital data and their material and epistemic infrastructures, there is now demand for specific competencies that care for research data. The language of ‘newness’ expresses the perception of a broader epistemic shift and the possibilities for contribution to this most recent version of science. It conjures an image of excitement, innovation, and novelty.

These notions are not in response to a job or set of tasks around keeping data clean and usable, but rather to a landscape of data work and knowing infrastructure that is perceived as fundamentally different.

Now, ‘frontier’ is a much more loaded term to handle. In the context of the treatment of Indigenous data by an NSF environmental science program, Murillo (2023) argues that data collection framed as a ‘new frontier’ is “grounded in a colonialist attitude and trope” which disregards Indigenous rights to data provenance and oversight (p. 3). In the context of this interview about data stewardship in Austria, such a direct implication of colonial legacy is unlikely. But, the relative comfort with which loaded terms like ‘frontier’ generally exist in conversation with data gives reason for pause. Whether or not my interlocutor understood this ‘frontier’ with its connotation of ‘empty’ space for the taking, this term nevertheless implies a landscape characterized by its uninhabited and unclaimed nature (however colonialist and mythical that characterization is). In the strictest definition of professional data stewardship, this is indeed an open zone. Furthermore, this zone is inextricably linked with the grand epistemic shift that data thinking has brought (Kitchin, 2014). However, this metaphor does not represent the array of people, machines, and data infrastructures that already exist around academic research data. The opposite of an empty space, this landscape is buzzing with data curators, all kinds of computers and software, and the researchers and administrators who did (and often still do) the tasks now taken up by professional data stewards (Gabrielsen, 2023; Pinel et al., 2020). Rather than a ‘new frontier’, what we can see is an already diverse world characterized by complex relations, discourses, and practices. The next two metaphors shed light on how this ‘new frontier’ is not a clean space primed for settling, but rather a realm already embroiled in friction.

Two other metaphors attempt to capture living and working in contemporary academic research conditions that shift away from the connotations of the ‘new frontier’: the ‘beam of pressure’ and the ‘grey zone of responsibility.’ Both metaphors portray working in the contemporary data university as, at least partially, working through data friction, albeit along different lines. The ‘beam of pressure’ refers to the coalescence of various streams of responsibilities and career mobility thresholds in academic life (Sigl et al., 2020; Fochler et al., 2016; Falkenberg and Fochler, 2023), for both researchers and data stewards. Eliciting sensations of intensity, with light and heat concentrated on a singular point, the ‘beam of pressure’ brings the perceived proliferation of tasks, responsibilities, and career considerations into the sensory realm. When additional demands around data stewardship arise, this just might push this beam of pressure into a dangerous affective zone. In contrast to this white-hot 'beam', we have the 'grey zone of responsibility'. Contained within this notion, there is an entirely different sensory evocation of a misty no-mans-land between the formation of a profile of formal responsibilities of data stewards, the kinds of requests that come in from researchers in the daily work practice of data stewards, and data stewards' own interests in research activities. There is something to be said for the contrast between these two illustrations. On the one hand, responsibilities, and career thresholds create a precise source of stress and discomfort, both for data stewards and for the researchers with whom they work. On the other hand, the unruliness of data and the refusal of intellectual interest to fully quiet down in favor of the service orientation of professional data stewardship create a zone that is characterized by its lack of boundaries. These exist simultaneously in the practice of data stewardship: attempted enclosure and persistent friction.

Next, we have a collection of alternative descriptors for data stewards: the 'air traffic controller', the 'data management police', and the 'auxiliary brain'. In these descriptors, data stewards draw parallels with other kinds of professions or entities and their perception of data stewardship. The 'air traffic controller' references an essential but strained logistical coordinator. From historical and contemporary examples of air traffic controller strikes, we understand air traffic control to be a linchpin in the flight industry that can enable or halt air travel. Without data stewards, research data also has a far trickier journey to its proposed final destination in a data repository. Data stewards not only guide data traffic but also guide researcher traffic through institutional data infrastructures and policies and across registers of value. However, on this journey, data stewards must avoid becoming the 'data management police' by investing in trustful relations with departments and individual researchers. Through this investment, they avoid the connotation of surveillance, shame, and enforcement that comes with pushing researchers into compliance with data policies without a buffer of trust. From Green et al. (2023) and Feldman and Shaw (2019), we know that implementing policies without the consultation or prioritization of data practitioners creates significant ethical consequences.

Finally, the description of a data steward as an 'auxiliary brain' points to the ways in which data stewards are seen as extensions of researchers, as an external support system. In being 'auxiliary', we can understand data stewardship as subsidiary or supplemental to what researchers do, yet still contributing in the sense of relieving researchers from tasks that could drain their 'brain' power away from 'real' research, much like biocurators do for biologists (Gabrielsen, 2023). Despite being essential, as the 'air traffic controller', caring for data does not hold the same status as producing research with that data. An air traffic controller is still not a pilot. These professional and anatomical metaphors of data stewardship express how data stewards understand their profession in broader cultural milieus of social ordering in academic research. In their perception, data stewardship is coded along existing value registers in which administrative or supportive roles in science exist in a separate identity sphere. Therefore, when the boundaries around the responsibilities and practices assumed to define data stewardship flout this coding, friction ensues with its accompanying affective relations.

6.2 Assembling data care in the data university

From the unique language used to describe data stewardship, we have thus far seen how data stewards conceive of their work with data in ways that stand in contrast to much of the formalizing literature around data stewardship (see Introduction). In practice, metaphoric comparisons to other professions are needed to explain the activities that encompass stewarding data through contemporary academic conditions. In providing these alternative languages and models, data stewardship professionals highlight the material doing, attachment, and commitment that constitute caring for data (Puig de la Bellacasa, 2011). Moving from a definition of care not solely as reproductive or affective labor, but as the manifold practices that hold things together, that assemble things, we can see the offerings of alternative languages and models as necessary adhesive components that allow for data stewards to acknowledge material conditions while still stewarding data along the lines of institutional and broader cultural expectations.

Here, Murphy's (2015) proposition that care often takes the form of less-than-palatable affective states or interventions is relevant. If we take research data as existing in an academic 'world already violated' (Murphy, 2015, p. 732), then the descriptions of 'beams of pressure', shame, 'grey zones', insecurity, 'data management police', and the open acknowledgment of these affective and sensory realms constitute a form of care for data that envisions a data future otherwise. Across every mode, from the data guide to the data confidant, projectification and assetization reared their heads as barriers to RDM practices by researchers and data stewardship by professionals (Falkenberg and Fochler, 2023; Fochler et al., 2016; Slaughter and Leslie, 1997). They spiraled through data's loss when short-term contracts terminated without exit plans for data, through associations with shame and confusion when researchers realized they had not sufficiently considered how to best organize collection and storage, and through stewardship by professionals who needed to constantly argue their relevance on the way to institutional stability.

In a discursive landscape obsessed with innovation and forward progression, it is a powerful intervention to acknowledge the long shadow that these notions leave in their wake. In the age of projectification and assetization, data stewards have deep knowledge of the specific conditions that impact data practices in academic research and have targeted suggestions for how to remedy the most potent faults of precarity as they impact data practices. RDM could be incorporated into undergraduate curricula to relieve the steep learning curve later in the academic career. Data journals could carry more prestige in the publication game. Investments into the technological capacity of repositories could encourage data reuse. These suggestions, and more, could be implemented in the short term to shore up the leaks in research data production as it stands now. But, these suggestions could also contribute to a more seismic shift in research towards stability in place of precarity, towards quality in place of objectivity (Funtowicz and Ravetz, 1994). If we care for how science is done, then understanding the development of professional data stewardship as both a symptom of and a solution to an academy plagued by neoliberal influence might be the starting place for re-imagining a future with data characterized by care.

6.3 Modeling the data care spiral

To re-imagine data futures is not a far-off proposition from the institutional, experiential, and caring vantage point of professional data stewardship; data stewards are already doing it. In Ch. 3, we saw how the 'data life cycle' is ubiquitous as a heuristic tool in the process of educating RDM. It is a visual and discursive simplification that purports to allow responsibility and agency to emerge in data practices. But, as my interlocutors made clear, it has dubious value when applied to concrete cases. The fixed endpoint of the 'data life cycle' in 'reuse' falls flat, as does the phased progression toward it (Borgman, 2018 and 2019; Leonelli et al., 2015 and 2018). My fellow attendees in RDM workshops made clear that their understanding of their own research material did not follow this model. What they considered the basis of their research either did not fit into the presented definition of data or did not submit to a bounded idea of reuse as the value-determiner of research data (Feldman and Shaw, 2019).

But, in the context of a conversation around the value of data reuse for research funders, one interview respondent suggested a new model: the data life spiral. Data's life is,

> *not just a cycle but also [...] a spiral. So, you can also start again and again, again, with this one dataset [...]* 7:44 ¶ 207 in DW_P1_231004.docx

This notion might sound similar to the already elaborated 'data life cycle'. After all, it seems to just emphasize multiple cycles of reuse in its conception of data life. However, based on our conversation and my overarching impression of my interlocutors' understanding of data journeys, there is more to be said for this spiral than just repeated reuse. The shape of the spiral is reveals an alternate framework for thinking with data. Rather than a flat, 2D cycle, we are offered a 3D spiral that moves out into time and space, creating more room for appreciating the non-atomizable elements that assemble data. It makes sense that an idea such as the 'data life spiral' would come out of the data stewardship community. After all, this community spends its working days negotiating the minute practices, grand narratives, and affective states that assemble data. They understand that not only the possibility for publication and reuse, but affinity, stress, and disciplinary specificity contribute to the creation of data.

The data stewardship community, while in certain registers espousing more simplistic models and heuristics for the sake of institutional legitimacy and legibility, is deeply aware of these confounding factors in their tacit framing of data stewardship and data. When workshop attendees pointed out the limited utility of the cycle for their data, workshop facilitators conceded their points. When discussing the 'data life cycle' in the context of their previous research experiences, my interlocutors expressed their misgivings about its representational or instructive power. Educating with the "data life cycle'model' was a process of bricolage that attempted to negotiate the friction inherent in data work (Varenne and Koyama, 2011; Edwards et al., 2011; Green et al., 2023). In these moments, it became clear that cramming data practices, diverse disciplinary epistemic components, and digital file formats (not to mention information or knowledge) into the term 'data' and then prescribing it a necessary path through life ceases to hold together – it disassembles.

As Pinel et al. (2020) argue, the value of data is not only determined by its capacity for reuse but also by the relations that create it. Data journeys are complex and non-linear (Leonelli, 2015; 2020). While the 'data life cycle' terminus in reuse might be important in conversations with funders to emphasize a worthwhile investment in research, the story is more complicated than proving distinct instances of reuse as the validation of value. Where the 'data life cycle' moves towards a prescriptive endpoint in reuse, the 'data life spiral' is more open to the variety of relational, institutional, affective, and historical ties that assemble specific data objects and that might complicate or preclude the process of reuse. Perhaps instead of adhering to the strict phases of the 'data life cycle' as what assembles 'good' data and data management, the 'data life spiral' would permit the recognition of the skilled, time-intensive, effortful labor that goes into creating and maintaining data. The data life spiral might enable us to envision alternate data futures in the face of current constraints.

If the 'data life cycle' fails as an effective heuristic or adequate model, and, in practice, RDM professionals move toward more complicated versions or away from it altogether, perhaps we should follow suit in our mundane lives with data. In the proposal of the 'data life spiral', there is an inherent acknowledgment of the complications, the unknowns, or untraceables, in working with data. In its 3D shape and extended temporality, the 'data life spiral' allows for longer historical ties to data to be maintained, following the long-standing proposition by Indigenous scholars of data (Murillo, 2023). In the 'data life spiral', the practices, affects, and things that coalesce to become research data exist closer in space, perhaps even overlapping, spiraling together. This might be a more frustrating way to conceive of epistemic work in the contemporary academy, and its consequences for policy or curriculum are less obvious. If the value in academic research data currently lies in its good management and subsequent reuse of a stable digital object, then registers of valuation would have to change in order to encompass the variety of lifeworlds that spiral around data (Baker and Bowker, 2007; Edwards, 2010; Edwards et al., 2011; Bates, 2017; Mayernick, 2019). This would necessarily require rather drastic shifts in how academic research communities negotiate value and legitimacy. Where data cleaning and curation are treated as less epistemically valuable than the research these practices produce, such efforts could gain formal recognition (Gabrielsen, 2023). There is much room for tinkering with how data spiral through academic work when the pace and orientation of contemporary university cultures can come under examination. In the process of such an examination, we might create more breathing room for the diversity of practices around data that assemble to produce academic work.

7. Conclusion

At the outset of this thesis, I asked how our understanding of contemporary research might change if we centered the left-behinds of academic research in the form of data. This focus led us not to repository management or archival practices as such, but into the present-focused realm of data stewardship as it seeks to make research data open and FAIR towards future reuse. As data and the promissory narratives around them proliferate, data must be cared for in specific ways to meet the expectation that it offers solutions to our most wicked societal problems (Palmer and Craigin, 2023). With the emergence of openness and FAIR data as proposed means through which data can live extended lives, I examined data stewardship as it is practiced in the context of Austrian research universities as one site where care for long-term data futures can be located. By asking how data stewardship is assembled to care for data, I focused on the construction of data stewardship from a pool of modes taken up by data stewards and RDM professionals that combine distinct practices and affective orientations with the ultimate goal of making research data archivable and reusable. Conceiving of data stewardship through assemblage (De Landa, 2016) further led me to understand data as an umbrella term that covers plural locally and relationally wrought assemblages. Data as a cultural movement, evidentiary proof, material entity, and more cannot be named in advance. Data must be located in the process of assembly (Leonelli, 2015). As we have seen, this process causes complications for data stewards and researchers when policies, tacit notions of openness, and institutional resources do not allow for data's unruliness. These moments of friction (Edwards, 2010; Edwards et al., 2011) formed the basis of my empirical conclusions. Through document work, ethnographic participation, and semi-structured interviews, I understood negotiating data friction as the kind of complicated care (Puig de la Bellacasa, 2011; Murphy, 2015) that enables data to be assembled. This kind of care demands multimodality in practice.

In the **data guide**, we saw data stewards maneuver the existing data infrastructures of the university. Data stewards not only guide data and researchers through the university but must also argue their relevance and assert themselves as legitimately participating in and as data infrastructure (Star 1999; Bowker and Star, 2000; Baker and Bowker, 2007. Data infrastructures create friction (Edwards, 2010; Edwards et al. 2011) on the journey of data through the university as technical capacities and the distribution of responsibilities for data care remain opaque. Funding for researchers and data stewardship and RDM activities is not always secure in the short-term, projectified timescales that often characterize contemporary university life (Hackett, 2014; Dollinger, 2020), adding a layer of resource scarcity to this journey. Through this friction, we see that the codes that determine research data's assemblage are often disjointed with the practices and institutional mazes that it must pass through. On the one hand, research data is touted as a prime output of the university and understood as a source of value. On the other hand, the care for data after projects terminate does not always receive the financial support needed to maintain and promote value-creation through reuse. As data guides, data stewards attempt to close this distance between the ideals for research data and its assembly on-the-ground in institutional data infrastructures.

The **data guardian** allowed us to see how data's definition is negotiated in practice, with the determination of 'good' and 'bad' data deciding their eventual archival future. Preventing data loss through lack of skill or awareness, preserving 'good' data for reuse, and deleting data that is not 'good' enough all make up the practices of the data guardian. Despite the seemingly harsh process of deciding the quality of data, data stewards argued that researchers should be the ultimate arbiters of data value for archiving, even displaying an affinity for niche or idiosyncratic data. In this way, data stewards acknowledge that 'goodness' and reusability are not the only determinants of data value. Affective investment and relational labor are as well (Choroszewicz, 2022; Green et al., 2023; Pinel et al., 2020). In practice, however, the stated purpose of data archiving and long-term storage as enabling reuse takes pride of place. This means that more complicated affective and relational entanglements with data must eventually take a backseat to the formal requirements of archiving and preservation in a reusable form (Green et al., 2023; Plantin, 2021). Nevertheless, the disciplinary diversity of research data (Feldman and Shaw, 2019) resists and confounds logics of reuse in its incompatibility with fixed notions of ownership, authorship, and epistemic labor.

The **data educator** gave us a glimpse into the twofold way that framings and practices of RDM and data stewardship are transferred to potential practitioners. Early-career academic populations learn the basics of RDM through the logic of the 'data life cycle' (Borgman, 2019), which attempts to describe and normatively prescribe a linear trajectory for research data before, during, and after the research process. However, teaching RDM becomes complicated when learners' conceptions of data or the data themselves do not fit within the model of the 'data life cycle'. Workshop facilitators negotiate this friction as it arises with an ad hoc, bricolage approach to education (Varenne and Koyama, 2011). In training professional data stewards, a different kind of friction must be balanced: generalist characteristics vs. specialist competencies. Professional training programs emphasize ‘soft skills’ in combination with disciplinary experience as the ideal recipe to produce a data steward (Gabrielsen, 2023; Grugulis and Vincent, 2009). This emphasis on soft skills exists in conflict with calls for more technical data science skills from RDM literature on curriculum and hiring patterns. In the end, the balance between soft skills and disciplinary experience is struck through oscillation in the work of data stewards.

Finally, the **data confidant** takes up affective engagements with data most explicitly. With the speed of research ticking up and high turnover characterizing academic projects, data stewards work with varying affective conditions as they take up the other three modes. Caring for data in this mode depends on prior disciplinary experience, toolkits of emotional regulation, and an awareness of the complex relationship between data and the researchers who make and use them (Choroszewicz, 2022; Gabrielsen, 2023; Pinel et al., 2020). This mode demands a specific sensitivity and vision that enables data stewards to ‘read between the lines’ of the problems that researchers bring them, excavating shame and distress in an effort to maintain a working relationship of trust. Without this mode, it is debatable if the other three would be operational. This affective sensitivity to the conditions that surround research data practices runs through data stewards’ capacity to negotiate frictions of institutional integration and guidance, to parse the diversity of data in determinations of quality, and to transfer knowledge about data stewardship and RDM to an uninitiated audience.

This modal map of data stewardship attempts to show that the assemblage of data is an arduous process that requires much of those who actively engage in it. Existing literature has explored the gendered and invisibilized dimensions of data curation, the frictions inherent in the construction and operation of data infrastructures, and the skillful effort required for data to travel toward reuse. This project fills a gap in the literature by probing data stewardship as a modality of care for academic research data that encompasses these existing explorations and extends them to understand how data and data stewardship exist in a relation of mutual assemblage through care. Previous scholarship makes clear that various forms of labor, affective investments, and relations go into building data infrastructures and cultivating reuse. However, there remains sparse work that examines how these elements are called upon to assemble data as encompassing both visionary notions and material digital objects. I argue that care, as enacted through the practices and positionalities of data stewards, is the essential adhesive that enables this assemblage. Codes, in De Landa’s (2016) sense, that declare ‘data as valuable’, ‘data as solution to complex societal issues’, and ‘data as the gold standard of evidentiary proof’, require territorialization to become legible and motivating to wider audiences. When moments of data friction stand to loosen the cohesion of data assemblages, interventions of care through affective engagement or definitional work stick things back together again, enabling data to move.

There is one final twist to this story about data stewardship. As you might recall from following the data guide through the maze of funding in the Austrian research landscape, one group of data stewards was actively negotiating the standing of data stewardship at their university while on fixed-term contracts under a pilot program. This status placed these data stewards in somewhat similar conditions to the research projects they described, in which researchers experienced career insecurity and needed to accumulate assets in order to move forward in on academic trajectory. By the end of the empirical engagement of this thesis, the pilot program was finished, and this group of data stewards were given permanent contracts within the university. After getting to know these interlocutors over the course of months, I was relieved to hear that they had gained some level of institutional recognition and security. But on an analytical level, I wondered how much this final incorporation of professional data stewardship into the university could change my thinking about its positionality and assemblage. Was data stewardship, along with its acknowledgment of data friction and affective commitments as practiced by Austrian data stewards, now a formally recognized part of institutional data production? Could this move by the university be seen as an embrace of data's complexity in addition to its valuable status?

Based on the multi-modal practice of data stewardship exhibited by my interlocutors and the dominant promissory and accumulative notions of data so easily found across academic research settings, this institutional choice does not constitute a genuine movement toward a new conception of data at the university. Rather, it incorporates data stewardship into existing data infrastructures insofar as it shores them up in the context of open and FAIR data discourses. If infrastructures, "encode the dreams of individuals and societies and are the vehicles whereby those fantasies are transmitted and made emotionally real" (Larkin, 2013, p. 333), then data stewardship, in its localized Austrian form, is incorporated into a fantasy of data in which universities gain institutional legitimacy and prestige by participating in openness and producing data assets. The acceptance of data stewardship into the institutional fold does not necessarily signal a radical shift in how data infrastructures, labor, openness, or reuse are understood. Rather, these material, practical, and technical components of data assemblages can exist in a substrate below the "forms of desire and fantasy" of infrastructures "that can sometimes be wholly autonomous from their technical function" (Larkin, 2013, p. 329). It is possible that data stewardship, for all its embedded diversity and complexity in practice, could be incorporated purely on the basis of its utility to a higher data fantasy.

Larkin (2013) argues that, “In the case of infrastructures, the poetic mode means that form is loosened from technical function” (p. 335). Even if this is the case and data stewardship poetically contributes toward the assemblage of data as a powerful notion of academic objectivity and prestige more than it technically enacts a localized form of care for data’s constitutive conditions, data stewardship still offers us a different way to think and work with data. Beyond the formal goals of data stewardship and the prescribed characteristics of professional data stewards, we can look to data stewardship *in care-ful practice* to see the beautifully unruly and deeply interdependent ways that researchers, materialities, information, evidence, and data assemble. Data stewards, with their intimate knowledge of data journeys through our epistemic projects, know better than most the possible points of intervention in the assemblage of data. If we would like to live differently with data, to know the constitutions and implications of data more fully, and to inquire into the gravity of data’s material expansion, then perhaps data stewards are the first people we should ask about making data worlds more livable.

8. References

Asamer, E.-M., Stryeck, S., & Kalová, T. (2020). *FAIR Office Austria Workshop Summary Report*. https://doi.org/10.25365/phaidra.235

Asdal, K., & Reinersten, H. (2022). *Doing document analysis: A practice-oriented method*. SAGE Publications Ltd.

Austrian Science Fund (FWF). N.d. *Research Data Management*. https://www.fwf.ac.at/en/about-us/what-we-do/open-science/research-data-management

Baker, K. S., & Bowker, G. C. (2007). Information ecology: Open system environment for data, memories, and knowing. *Journal of Intelligent Information Systems*, *29*(1), 127–144. https://doi.org/10.1007/s10844-006-0035-7

Baker, K. S., & Karasti, H. (2018). Data care and its politics: Designing for local collective data management as a neglected thing. *Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1*, 1–12. https://doi.org/10.1145/3210586.3210587

Bardel, A., Hasani-Mavriqi, I., Jean-Quartier, C., Schaffer, P., Macher, T., Kalová, T., Feichtinger, M., Bargmann, M., Sanchez Solis, B., Stork, C., Miksa, T., & Piroi, F. (2023). Data Stewardship – Austrian National Strategy and Alignment. *Zeitschrift Für Hochschulentwicklung*, *18*(Sonderheft Forschung), 65–88. https://doi.org/10.21240/zfhe/SH-F/05

Bates, J. (2018). The politics of data friction. *Journal of Documentation*, 74(2), 412–429. https://doi.org/10.1108/JD-05-2017-0080

Beer, D. (2019). *The data gaze: Capitalism, power and perception*. SAGE Publications Ltd.

Birch, K., & Muniesa, F. (Eds.). (2020). *Assetization: Turning things into assets in technoscientific capitalism*. The MIT Press.

Blumesberger, S., Gänsdorfer, N., Ganguly, R., Gergely, E., Gruber, A., Hasani-Mavriqi, I., Kalová, T., Ladurner, C., Macher, T., Miksa, T., Sánchez Solís, B., Schranzhofer, H., Stork, C., Stryeck, S., & Thöricht, H. (2021). FAIR Data Austria—Aligning the Implementation of FAIR Tools and Services. *Mitteilungen Der Vereinigung Österreichischer Bibliothekarinnen Und Bibliothekare*, *74*(2), 102–120. https://doi.org/10.31263/voebm.v74i2.6379

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078. https://doi.org/10.1002/asi.22634

Borgman, C. L. (2018). Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier. *Berkeley Technology Law Journal*, *33*(2), 365–412. https://www.jstor.org/stable/26533142

Borgman, C. L. (2019). The Lives and After Lives of Data. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.9a36bdb6

Borgman, C. L., & Bourne, P. E. (2022). Why It Takes a Village to Manage and Share Data. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.42eec111

Bowker, G. C., & Star, S. L. (2000). *Sorting Things Out: Classification and Its Consequences*. The MIT Press. https://doi.org/10.7551/mitpress/6352.001.0001

Boyd, C. (2022). Data as assemblage. *Journal of Documentation*, *78*(6), 1338–1352. https://doi.org/10.1108/JD-08-2021-0159

Boyd, D., & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Buchanan, I. (2015). Assemblage Theory and Its Discontents. *Deleuze Studies*, *9*(3), 382–392. https://www.jstor.org/stable/45331821

Burke, P. J., & Manathunga, C. (2023). *The Timescapes of Teaching in Higher Education* (1st ed.). Routledge. https://doi.org/10.4324/9781003380290

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_suppl), 196–233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

Carroll, S. R., Rodriguez-Lonebear, D., & Martinez, A. (2019). Indigenous Data Governance: Strategies from United States Native Nations. Data Science Journal, 18(1), 31. https://doi.org/10.5334/dsj-2019-031

Charmaz, K. (2006). *Constructing grounded theory*. Sage Publications.

Choroszewicz, M. (2022). Emotional labour in the collaborative data practices of repurposing healthcare data and building data technologies. *Big Data & Society*, *9*(1), 205395172210984. https://doi.org/10.1177/20539517221098413

Dalton, C. M., & Thatcher, J. E. (2014, May 12). *What Does A Critical Data Studies Look Like, And Why Do We Care?* Society and Space. https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care

Datum/data. (1997). In Perelman, L.C., Paradis, J. & Barrett, E. (Eds.), *The Mayfield Handbook of Technical and Scientific Writing*. McGraw Hill. https://www.mit.edu/course/21/21.guide/data.htm#:~:text=Datum%20is%20singular%2C%20meaning%20%22one,used%20as%20a%20singular%20noun.

De Landa, M. (2016). *Assemblage theory*. Edinburgh University Press.

Demchenko, Y., & Stoy, L. (2021). Research Data Management and Data Stewardship Competencies in University Curriculum. *2021 IEEE Global Engineering Education Conference (EDUCON)*, 1717–1726. https://doi.org/10.1109/EDUCON46332.2021.9453956

Denis, J., & Pontille, D. (2023). Cultivating Attention to Fragility: The Sensible Encounters of Maintenance. In *Repair, Remediation and Resurgence in Social and Environmental Conflict*. Bristol University Press.

Dollinger, M. (2020). The projectification of the university: Consequences and alternatives. *Teaching in Higher Education*, 25(6), 669–682. https://doi.org/10.1080/13562517.2020.1722631

Dorschel, R. (2021). Discovering needs for digital capitalism: The hybrid profession of data science. *Big Data & Society*, *8*(2), 205395172110407. https://doi.org/10.1177/20539517211040760

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

European Commission. Directorate General for Research and Innovation. (2016). *Realising the European open science cloud: First report and recommendations of the Commission high level expert group on the European open science cloud.* Publications Office. https://data.europa.eu/doi/10.2777/940154

Falkenberg, R., & Fochler, M. (2024). Innovation in Technology Instead of Thinking? Assetization and Its Epistemic Consequences in Academia. *Science, Technology, & Human Values*, *49*(1), 105–130. https://doi.org/10.1177/01622439221140003

Feldman, S., & Shaw, L. (2019). The Epistemological and Ethical Challenges of Archiving and Sharing Qualitative Data. *American Behavioral Scientist*, *63*(6), 699–721. https://doi.org/10.1177/0002764218796084

Felt, U. (2017). Under the Shadow of Time: Where Indicators and Academic Values Meet. *Engaging Science, Technology, and Society*, 3, 53–63. https://doi.org/10.17351/ests2017.109

Fochler, M. (2016). Variants of Epistemic Capitalism: Knowledge Production and the Accumulation of Worth in Commercial Biotechnology and the Academic Life Sciences. *Science, Technology, & Human Values*, 41(5), 922–948. https://doi.org/10.1177/0162243916652224

Fochler, M., Felt, U., & Müller, R. (2016). Unsustainable Growth, Hyper-Competition, and Worth in Life Science Research: Narrowing Evaluative Repertoires in Doctoral and Postdoctoral Scientists' Work and Lives. *Minerva*, *54*(2), 175–200. https://doi.org/10.1007/s11024-016-9292-y

Gabrielsen, A. M. (2023). Gendering data care: Curators, care, and computers in data-centric biology. *Science as Culture*, 1–25. https://doi.org/10.1080/09505431.2023.2260830

Galletta, A. (2020). *Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication*. New York University Press. https://doi.org/10.18574/nyu/9780814732939.001.0001

Green, S., Hillersdal, L., Holt, J., Hoeyer, K., & Wadmann, S. (2023). The practical ethics of repurposing health data: How to acknowledge invisible data work and the need for prioritization. *Medicine, Health Care and Philosophy*, *26*(1), 119–132. https://doi.org/10.1007/s11019-022-10128-6

Gregg, M., & Andrijasevic, R. (2019). Virtually Absent: The Gendered Histories and Economies of Digital Labour. *Feminist Review*, 123(1), 1–7. https://doi.org/10.1177/0141778919878929

Gregory, K., Groth, P., Scharnhorst, A., & Wyatt, S. (2023). The Mysterious User of Research Data: Knitting Together Science and Technology Studies with Information and Computer Science. In K. Bijsterveld & A. Swinnen (Eds.), *Interdisciplinarity in the Scholarly Life Cycle* (pp. 191–211). Springer International Publishing. https://doi.org/10.1007/978-3-031-11108-2_11

Grugulis, I., & Vincent, S. (2009). Whose skill is it anyway?: 'Soft' skills and polarization. *Work, Employment and Society*, *23*(4), 597–615. https://doi.org/10.1177/0950017009344862

Hackett, E. J. (2014). Academic Capitalism. *Science, Technology, & Human Values*, *39*(5), 635–638. https://doi.org/10.1177/0162243914540219

Hasani-Mavriqi, I. (2020, November 6). Workshop: Tasks and Profile of a Data Steward. *Cluster Forschungsdaten*. https://forschungsdaten.at/en/workshop-tasks-and-profile-of-a-data-steward/

Hasani-Mavriqi, I., Knopper, S., & Jetten, M. (2021). *Professionalising data stewardship in the Netherlands: Competencies, training and education: Dutch roadmap towards national implementation of FAIR data stewardship*. https://phaidra.univie.ac.at/detail/o:1188966

Hasani-Mavriqi, I., Reichmann, S., Gruber, A., Jean-Quartier, C., Schranzhofer, H., & Rey Mazon, M. (2022). *Data Stewardship in the Making: What Austrian Universities Look For?* https://forschungsdaten.at/en/report-data-stewardship-in-the-making-2/

Heuts, F., & Mol, A. (2013). What Is a Good Tomato? A Case of Valuing in Practice. *Valuation Studies*, 1(2), 125–146. https://doi.org/10.3384/vs.2001-5992.1312125

Hillersdal, Line, and Mette N. Svendsen. 2022. Cancer Currencies: Making and Marketing Resources in a First-in-Human Drug Trial in Denmark. In *Precision Oncology and Cancer Biomarkers. Issues at Stake and Matters of Concern*, eds. Bremer, A., and Strand, R. pp. 45–60. Cham: Springer.

Hoeyer, Klaus. 2023. Data Paradoxes. *The Politics of Intensified Data Sourcing in Contemporary Healthcare*. MIT Press.

Holmwood, J., & Marcuello Servós, C. (2019). Challenges to Public Universities: Digitalisation, Commodification and Precarity. *Social Epistemology*, *33*(4), 309–320. https://doi.org/10.1080/02691728.2019.1638986

Holt, Jette. 2020. At gøre ting med tal: En nexusanalytisk afdækning af overvågningssystemet HAIBAS diskursive aflæsning. Aalborg Universitetsforlag. Aalborg Universitet. Det Humanistiske Fakultet. Ph.D.-Serien. Available online: https://vbn.aau.dk/ws/portalfiles/portal/404967144/PHD_JH_E_pdf.pdf

Horizon Europe. 2024. *HE Programme Guide: V4.1*. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

Jensen, C. B., Morita, A., & Osaka University. (2015). Infrastructures as Ontological Experiments. *Engaging Science, Technology, and Society*, *1*, 81–87. https://doi.org/10.17351/ests2015.007

Jensen, E. A., & Laurie, A. C. (2016). Doing real research: A practical guide to social research. SAGE.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 205395171452848. https://doi.org/10.1177/2053951714528481

Kitchin, R., & Lauriault, T. P. (2018). Toward Critical Data Studies. In *Thinking Big Data in Geography: New Regimes, New Research*. University of Nebraska Press.

Kitchin, R. (2021). Data lives: How data are made and shape our world. Bristol University Press.

Kitchin, R. (2022). The data revolution: A critical analysis of big data, open data & data infrastructures (Second edition). Sage Publications.

Lagoze, C. (2014). Big Data, data integrity, and the fracturing of the control zone. *Big Data & Society*, *1*(2), 205395171455828. https://doi.org/10.1177/2053951714558281

Law, J. (2004). *After method: Mess in social science research*. Routledge.

Leonelli, S. (2018). Rethinking Reproducibility as a Criterion for Research Quality. In L. Fiorito, S. Scheall, & C. E. Suprinyak (Eds.), *Research in the History of Economic Thought and Methodology* (Vol. 36, pp. 129–146). Emerald Publishing Limited. https://doi.org/10.1108/S0743-41542018000036B009

Leonelli, S., Spichtinger, D., & Prainsack, B. (2015). Sticks *and* carrots: Encouraging open science at its source. *Geo: Geography and Environment*, *2*(1), 12–16. https://doi.org/10.1002/geo2.2

Leonelli, S., & Tempini, N. (Eds.). (2020). *Data Journeys in the Sciences*. Springer International Publishing. https://doi.org/10.1007/978-3-030-37177-7

Levin, N., & Leonelli, S. (2017). How Does One "Open" Science? Questions of Value in Biological Research. *Science, Technology, & Human Values*, *42*(2), 280–305. https://doi.org/10.1177/0162243916672071

Lucivero, F. (2020). Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives. *Science and Engineering Ethics*, *26*(2), 1009–1030. https://doi.org/10.1007/s11948-019-00171-7

Lynch, M. (2013). Ontography: Investigating the production of things, deflating ontology. *Social Studies of Science*, *43*(3), Article 3. https://doi.org/10.1177/0306312713475925

Mayernik, M. S. (2019). Metadata accounts: Achieving data and evidence in scientific research. *Social Studies of Science*, *49*(5), 732–757. https://doi.org/10.1177/0306312719863494

Maylor, H., Brady, T., Cooke-Davies, T., & Hodgson, D. (2006). From projectification to programmification. *International Journal of Project Management*, 24(8), 663–674. https://doi.org/10.1016/j.ijproman.2006.09.014

Midler, C. (1995). "Projectification" of the firm: The renault case. *Scandinavian Journal of Management*, *11*(4), 363–375. https://doi.org/10.1016/0956-5221(95)00035-T

Mirowski, P. (2011). *Science-mart: Privatizing American science*. Harvard University Press.

Møller, N. H., Bossen, C., Pine, K. H., Nielsen, T. R., & Neff, G. (2020). Who does the work of data? *Interactions*, *27*(3), 52–55. https://doi.org/10.1145/3386389

Mons, B. (2018). *Data stewardship for open science: Implementing FAIR Principles*. CRC Press.

Mons, B. (2020). Invest 5% of research funds in ensuring data are reusable. *Nature*, *578*, 491.

Murillo, L. F. R. (2023). How to avoid the "infrastructural blues"? Studying-while-caring for data stewardship. *Annals of Anthropological Practice*, napa.12208. https://doi.org/10.1111/napa.12208

Newlands, G. (2021). Lifting the curtain: Strategic visibility of human labour in AI-as-a-Service. Big Data & Society, 8(1), 205395172110160. https://doi.org/10.1177/20539517211016026

Noble, S. U. (2020). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press. https://doi.org/10.18574/nyu/9781479833641.001.0001

Packendorff, J., & Lindgren, M. (2014). Projectification and its consequences: Narrow and broad conceptualisations. *South African Journal of Economic and Management Sciences*, *17*(1), 7–21. https://doi.org/10.4102/sajems.v17i1.807

Palmer, C. L., & Cragin, M. H. (2023). Curating for Convergence: Data Stewardship for Interdisciplinary Inquiry. *Library Trends*, *71*(1), 113–131. https://doi.org/10.1353/lib.2023.0007

Parsons, M. A., Godøy, Ø., LeDrew, E., De Bruin, T. F., Danis, B., Tomlinson, S., & Carlson, D. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science*, *37*(6), 555–569. https://doi.org/10.1177/0165551511412705

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the Reuse of Scientific Data. *Data Science Journal*, *16*, 8. https://doi.org/10.5334/dsj-2017-008

Perrier, L., Blondal, E., Ayala, A. P., Dearborn, D., Kenny, T., Lightfoot, D., Reka, R., Thuna, M., Trimble, L., & MacDonald, H. (2017). Research data management in academic institutions: A scoping review. *PLOS ONE*, *12*(5), e0178261. https://doi.org/10.1371/journal.pone.0178261

Pickren, G. (2018). 'The global assemblage of digital flow': Critical data studies and the infrastructures of computing. *Progress in Human Geography*, *42*(2), 225–243. https://doi.org/10.1177/0309132516673241

Pinel, C., Prainsack, B., & McKevitt, C. (2020a). Caring for data: Value creation in a data-intensive research laboratory. *Social Studies of Science*, *50*(2), 175–197. https://doi.org/10.1177/0306312720906567

Pinel, C., Prainsack, B., & McKevitt, C. (2020b). Caring for data: Value creation in a data-intensive research laboratory. *Social Studies of Science*, *50*(2), 175–197. https://doi.org/10.1177/0306312720906567

Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data: Conceptualising data in an emerging world. *Big Data & Society*, *5*, 205395171775322. https://doi.org/10.1177/2053951717753228

Plantin, J.-C. (2021). The data archive as factory: Alienation and resistance of data processors. *Big Data & Society*, *8*(1), 205395172110075. https://doi.org/10.1177/20539517211007510

Rainie, S., Kukutai, T., Walter, M., Figueroa-Rodriguez, O., Walker, J., & Axelsson, P. (2019). Issues in Open Data: Indigenous data sovereignty. In *The State of Open Data: Histories and Horizons*. Zenodo. https://doi.org/10.5281/ZENODO.2677801

Research Data Alliance. (n.d.). *Professionalising Data Stewardship Interest Group*. https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig/

Rivas, C. (2012). Coding and analysing qualitative data. In C. Seale (Ed.), *Researching society and culture* (Fourth edition, pp. 366–392). Sage.

Sigl, L., Felt, U., & Fochler, M. (2020). "I am Primarily Paid for Publishing…": The Narrative Framing of Societal Responsibilities in Academic Life Science Research. *Science and Engineering Ethics*, *26*(3), 1569–1593. https://doi.org/10.1007/s11948-020-00191-8

Slaughter, S., & Leslie, L. L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Johns Hopkins University Press.

Slaughter, S., & Leslie, L. L. (2001). Expanding and Elaborating the Concept of Academic Capitalism. *Organization*, *8*(2), 154–161. https://doi.org/10.1177/1350508401082003

Smale, N., Unsworth, K., Denyer, G., & Barr, D. (2018). The History, Advocacy and Efficacy of Data Management Plans. *bioRxiv* preprint. https://doi.org/10.1101/443499

Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist*, *43*(3), 377–391.

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

Staunton, C., Barragán, C. A., Canali, S., Ho, C., Leonelli, S., Mayernik, M., Prainsack, B., & Wonkham, A. (2021). Open science, data sharing and solidarity: Who benefits? *History and Philosophy of the Life Sciences*, *43*(4), 115. https://doi.org/10.1007/s40656-021-00468-6

Task Force Data stewardship, curriculum and career paths. (n.d.). *Task Force charter Data stewardship curricula and career paths*. European Open Science Cloud Association. https://www.eosc.eu/sites/default/files/2021-12/eosca_tfdatastewardshipcurriculaandcareerpaths_draftcharter_20210614.pdf

Technical University of Vienna. (n.d.) .digital office. https://www.tuwien.at/en/tu-wien/organisation/central-divisions/digital-office

Teperek, M., Cruz, M. J., Verbakel, E., Böhmer, J., & Dunning, A. (2018). Data Stewardship addressing disciplinary data management needs. *International Journal of Digital Curation*, *13*(1), 141–149. https://doi.org/10.2218/ijdc.v13i1.604

Thylstrup, N. B. (2022). The ethics and politics of data sets in the age of machine learning: Deleting traces and encountering remains. *Media, Culture & Society*, *44*(4), 655–671. https://doi.org/10.1177/01634437211060226

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

Traweek, S. (1992). *Beamtimes and lifetimes: The world of high energy physicists* (1. ed). Harvard Univ. Press.

Varenne, H., & Koyama, J. (2011). Education, Cultural Production, and Figuring Out What to Do Next. In B. A. U. Levinson & M. Pollock (Eds.), *A Companion to the Anthropology of Education* (1st ed., pp. 50–64). Wiley. https://doi.org/10.1002/9781444396713.ch4

Vienna University of Economics and Business. (2024). *Research Data Management*. https://www.wu.ac.at/en/library/services/employees/research-support/research-data-management

University of Graz. (n.d.). *Research Data Management*. https://forschungsdatenmanagement.uni-graz.at/en/

University of Graz. (2023). Digitalization. In *Development Plan 2025-2030*. https://static.uni-graz.at/fileadmin/_files/_administrative_sites/_strategie-und-qualitaet/Entwicklungsplan_2025-2030.pdf

University of Vienna. (n.d.). *The research data life cycle*. https://rdm.univie.ac.at/what-is-research-data-management/the-research-data-lifecycle/

University of Vienna. (2020). Digitalisation strategy of the University of Vienna. https://digital.univie.ac.at/fileadmin/user_upload/p_digital/Dokumente/Digitalisierungsstrategie_ENG.pdf

University of Vienna. (2021) *Research Data Management Policy at the University of Vienna*. https://rdm.univie.ac.at/fileadmin/user_upload/p_forschungsdatenmanagement/Dokumente/RDM_Policy_UNIVIE_v1_en.pdf

Wadmann, S., & Hoeyer, K. (2018). Dangers of the digital fit: Rethinking seamlessness and social sustainability in data-intensive healthcare. *Big Data & Society*, 5(1), 205395171775296. https://doi.org/10.1177/2053951717752964

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE*, *8*(7), e67332. https://doi.org/10.1371/journal.pone.0067332

Wendelborn, C., Anger, M., & Schickhardt, C. (2023). What is data stewardship? Towards a comprehensive understanding. *Journal of Biomedical Informatics*, *140*, 104337. https://doi.org/10.1016/j.jbi.2023.104337

Whitaker, M. P. (1996). Ethnography as Learning: A Wittgensteinian Approach to Writing Ethnographic Accounts. *Anthropological Quarterly*, *69*(1), 1–13. https://www.jstor.org/stable/3317135

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., … Mons, B. (2016).

The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18
Wing, J. M. (2019). The Data Life Cycle. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.e26845b4
Winthereik, B. R. (2023). Data as Relation: Ontological Trouble in the Data-Driven Public Administration. *Computer Supported Cooperative Work (CSCW)*. https://doi.org/10.1007/s10606-023-09480-9
Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power* (First edition). PublicAffairs.