universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Linked (Open) Data und semantische Interoperabilität in der Archäologie“

verfasst von / submitted by

Martina Simon, BA

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

Master of Arts (MA)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 801
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Urgeschichte und Historische Archäologie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Michael Doneus

Danksagung

Digitale Technologien ermöglichen es der archäologischen Forschung, Wissen weiter als bisher zu verbreiten, erhobene Daten effektiver zu analysieren und ungeahnte neue Verbindungen zu entdecken. Daher ist es besonders wichtig, diese Technologien ständig weiterzuentwickeln und die daraus gewonnenen Daten für die Zukunft zu erhalten, um neue Erkenntnisse gewinnen zu können. Im Zuge meiner Mitarbeit im Projekt „A Puzzle in 4D – Tell el-Daba. Digital Preservation and Reconstruction of an Egyptian Palace" wuchs mein Interesse an digitalen Methoden in der Archäologie und besonders an semantischen Technologien und so entstand die Idee zu dieser Arbeit.

Ohne Hilfe ist ein so umfassendes Projekt jedoch schwer umzusetzen. Mein besonderer Dank richtet sich an meinen Betreuer Univ.-Prof. Mag. Dr. Michael Doneus, der mich zu jeder Zeit unterstützt und mich zu neuen Ideen angeregt hat. Dr. Gerald Hiebel hat mich zu dieser Arbeit inspiriert und ist mir stets mit Rat beiseite gestanden. Dafür und für die große Hilfe bei der Konvertierung des Datenmodells möchte ich ihm besonders danken. Bei Ass.-Prof. Mag. Dr. Alois Stuppner bedanke ich mich herzlichst für die ermöglichte Verwendung der Daten zu urnenfelderzeitlichen Fundobjekten des Oberleiserbergs und die Bereitstellung der Funde für die Untersuchung. Für die vielen konstruktiven Gespräche und Diskussionen möchte ich mich bei meinen Kolleginnen und Kollegen bedanken. Ein großer Dank gilt auch meinen Freundinnen und Freunden, die mich stets moralisch und geistig unterstützt haben.

Mein größter Dank gilt jedoch meinen Eltern und meiner Familie, die mich in allen Lebenslagen immer unterstützt und mich das gesamte Studium hindurch begleitet haben. Ich möchte mich auch ganz besonders bei meinem Lebensgefährten Bernhard Müller bedanken. Er hat mich geduldig unterstützt und mich stetig motiviert, weiterzumachen. Ohne ihn wäre diese Arbeit nicht möglich gewesen.

Inhaltsverzeichnis

1 Einleitung

„This document gives a road map - a sequence for the incremental introduction of technology to take us, step by step, from the Web of today to a Web in which machine reasoning will be ubiquitous and devastatingly powerful."[1]

Sir Tim Berners-Lee, Erfinder des World Wide Web

Ein Netz von Daten, das als globale Datenbank fungieren kann, wird der Evolution menschlichen Wissens erheblichen Aufschwung verleihen. Diese Vision des Semantic Web stellt Tim Berners-Lee mit dem Dokument „Semantic Web Road map" vor und präsentiert, was mit digitalen Daten möglich ist.[2,3] Wie diese Technologie die Wege der archäologischen Forschung positiv beeinflussen kann, ist Thema dieser Masterarbeit.

1.1 Probleme mit digitalen archäologischen Daten

Digital Humanities ist ein großes Schlagwort unserer Tage. Anders als in vielen Disziplinen hat die Digitalisierung in der Archäologie bereits früh Fuß gefasst und ist allgegenwärtig in diesem Fach. Archäologische Ausgrabungen werden seit Jahren fast ausschließlich digital dokumentiert und die archäologische Prospektion arbeitet seit Jahrzehnten mit digitalen Big Data. Dadurch steht die Archäologie zunehmend vor größeren Herausforderungen des digitalen Datenmanagements.

Es gibt es eine Vielzahl an bereits abgeschlossener Ausgrabungs- und Forschungsdokumentationen. Diese wertvollen Daten nachhaltig zu archivieren, ist wichtig, um Datenverlust entgegenzuwirken und den Nutzen der Daten sicherzustellen. Dabei muss auf alte Dateiformate geachtet werden und darauf, dass die Daten wiederverwendbar und zugänglich gemacht werden können. Gleichzeitig erzeugen aktuelle Forschungs- und Ausgrabungstätigkeiten laufend neue digitale Daten, welche ebenfalls nachhaltig archiviert und verwaltet werden müssen. Es werden jedes Jahr unzählige archäologische Maßnahmen durchgeführt, die personelle Struktur in der Archäologie macht es jedoch unmöglich, die daraus resultierenden Daten auszuwerten und der Forschung zugänglich zu machen. Die meisten Daten werden dem Bundesdenkmalamt analog und in digitaler Form übergeben und sie verbleiben unberührt in Archiven. Das Bundesdenkmalamt gibt

[1] Berners-Lee 1998a, Introduction.
[2] Berners-Lee 1999, Machine-Understandable information: Semantic Web.
[3] Berners-Lee u. a. 2001, Evolution of Knowledge.

Richtlinien vor, in welcher Weise die Daten abgegeben werden müssen, dennoch sind diese heterogen, da unterschiedliche Standards und Vokabulare verwendet werden. Die Wiederverwendung für archäologische Forschungen dieser enormen Datenmengen wird dadurch erschwert. Dieser Umstand ist noch schwerwiegender zu sehen, da große Summen an Steuergeldern dafür aufgewendet werden, die Daten im Sinne des Denkmalschutzes zu erzeugen und zu retten. Die Möglichkeit der Wiederverwendung, Integrationsfähigkeit, Interoperabilität und nachhaltigen Nutzung sind notwendig für zukünftige archäologische Forschungsfragen. Es wäre ein großer Verlust, wenn diese Daten nicht genutzt würden.

Diskussionen über Open Data, das Wiederverwenden und Archivieren von archäologischen Daten werden zunehmend dringlicher. Doch um Daten sinnvoll wiederverwenden, archivieren und interoperabel mit anderen Daten verwenden zu können, gilt es einiges zu überwinden: archäologische Daten sind meist heterogen, wodurch oft ungenutzte Dateninseln mit uneinheitlichen Standards und Terminologien entstehen, zusätzlich erschweren veraltete Dateiformate und restriktive Zugangsmöglichkeiten das Archivieren und Wiederverwenden von Daten.

Die Methoden und Technologien des Semantic Web und der Linked (Open) Data bieten die Möglichkeit, Daten zu vernetzen, ihre Interoperabilität zu erhöhen und dadurch eine Wiederverwendung zu erleichtern. Archäologische Daten profitieren von diesen Möglichkeiten. Darüber hinaus geben semantische Technologien neue Perspektiven, um archäologische Daten aufzubereiten und zu nutzen. Bisher haben sich diese Technologien noch nicht durchgesetzt. Dies liegt zum einen an ihrer weitgehenden Unbekanntheit und zum anderen an ihrer Komplexität und Schwierigkeit in der Anwendung.

1.2 Ziel, Fragestellung

Semantische Technologien und Linked (Open) Data werden bereits seit einiger Zeit in der archäologischen Forschung diskutiert, entwickelt und angewandt (siehe auch Kapitel 7).[4] Die Publikationen sind jedoch meist technisch sehr komplex und sprechen lediglich eine Minderheit von Experten*innen in diesem Gebiet an. Daher ist es wenig verwunderlich, dass diese Technologien in der Archäologie wenig bekannt sind. Hinzu kommen nur einige wenige Beispiele, welche die Vorteile dieser Technologien darstellen können und zeigen, ob sich der Mehraufwand in der Aufbereitung der Daten lohnen würde.

[4] U. a. Hiebel u. a. 2021a; Doerr – Iorizzo 2008; Binding u. a. 2015; Isaksen 2011; Kintigh 2006; Felicetti u. a. 2015; Thiery u. a. 2019; Eichert 2021.

Diese Arbeit verfolgt das Ziel, die Methoden und Anwendungsbereiche von Linked (Open) Data und des Semantic Web in der Archäologie darzulegen. Dabei werden die potenziellen Vorteile und Hürden semantischer Technologien für archäologische Daten aufgezeigt und geprüft. Da es im Allgemeinen nicht die gängige Praxis für Archäolog*innen ist, Daten semantisch aufzubereiten, ist eine Zusammenarbeit mit IT-Spezialist*innen in diesem Gebiet zu empfehlen. Die Spezialist*innen wiederum sind dringend auf Archäolog*innen angewiesen. Denn diese können die archäologischen Daten sinnvoll modellieren und im Fach bekannte Standards und Vokabulare verwenden.

Im Zuge dieser Arbeit soll daher der Frage nachgegangen werden, wie Archäolog*innen ohne besondere IT-Kenntnisse archäologische Daten zu einer Ontologie modellieren, sie mit Standards und Vokabularen versehen können und in weiterer Folge in ein Linked (Open) Data-Format bringen können. Welche Vorteile bringen semantische Technologien für archäologische Daten? Welche Hindernisse sind für die Aufbereitung der Daten zu überwinden und ab welchem Zeitpunkt sind Experten in diesem Gebiet hinzuzuziehen?

Um dieses Ziel zu erreichen und die Fragen beantworten zu können, werden die Grundlagen der Technologien aufgezeigt, relevante Aspekte und Herausforderungen des Themenbereiches diskutiert und eine exemplarische Aufbereitung demonstriert. Es werden die notwendigen Schritte für die Anreicherung von archäologischen Daten mit Standards und Vokabularen, die anschließende Modellierung der Daten zu einer im kulturhistorischen Bereich verwendeten Ontologie und die Konvertierung der Daten in ein Format für Linked (Open) Data Schritt für Schritt aufgezeigt.

Dies wird anhand einer Funddatenbank zu Fundobjekten gezeigt, welche am Oberleiserberg, einer bedeutenden Höhensiedlung im Niederösterreichischen Weinviertel geborgen wurden. Die einzelnen Schritte sollen einen Weg zur verbesserten Datenintegration darlegen. Die Daten zu den Funden des Oberleiserbergs stammen zum größten Teil von Funden aus der Urnenfelderzeit. Unter den Einträgen von insgesamt 381 Buntmetallobjekten befinden sich unter anderem Nadeln und Nadelfragmente, gegossene kleine Bronzeringe und ein großer Anteil an Werkstätten-Abfall. Die Funde wurden bereits von Alois Stuppner in einer unpublizierten Access-Datenbank mit 1.601 Datensätzen zu Metallfunden vom Oberleiserberg erfasst. Die Daten werden im Rahmen der Arbeit mit Hilfe eines einheitlichen Vokabulars und Standards für ein Semantic Web aufbereitet, damit in der Folge archäologische Suchen und Abfragen trotz der Heterogenität der vorliegenden Daten zu Ergebnissen führen, die einen bisher noch nicht dagewesenen archäologischen Mehrwert generieren.

1.3 Aufbau der Arbeit

Die Arbeit ist in zwei Abschnitte eingeteilt. Dabei werden in Teil 1 die wesentlichen Grundbegriffe und Verfahren des Semantic Web und Linked (Open) Data erklärt. Konkret befasst sich Teil 1 mit den Kapiteln 2 bis 6 mit den Grundlagen des Datenmanagements, des Semantic Web und Linked (Open) Data. In Kapitel 2 wird das Datenmanagement in der Archäologie diskutiert. Hier liegen die Schwerpunkte auf dem Forschungsdatenmanagement und dem Umgang mit archäologischen Daten. Das dritte Kapitel zeigt die grundlegenden Konzepte des Semantic Web und der Linked (Open) Data. Anschließend wird in Kapitel 4 auf die technischen Grundlagen für das Semantic Web und der Linked (Open) Data eingegangen. Die Bedeutung und Erläuterung von Ontologien in Hinblick auf semantische Technologien werden in Kapitel 5 erläutert. Dort wird der Fokus insbesondere auf die Ontologie CIDOC CRM und ihre Anwendung in der Archäologie gelegt. Am Schluss des ersten Teils werden in Kapitel 6 die Standards vorgestellt, welche für die Datenintegration wichtig sind.

Teil 2 der Arbeit mit den Kapiteln 7, 8 und 9 befasst sich mit der praktischen, archäologischen Anwendung von semantischen Technologien. Es werden Beispiele gezeigt, wie semantische Technologien in der Archäologie angewendet werden können. Zunächst wird in Kapitel 7 das Semantic Web mit der Archäologie in Verbindung gebracht und Beispiele zu Anwendungen von archäologischen Forschungsprojekten gezeigt. Die exemplarische, praktische Anwendung anhand von archäologischen Daten wird in Kapitel 8 vorgestellt. Darin werden zunächst die Hintergründe der Daten und die Datenbasis beschrieben. Im Anschluss wird die angewandte Methodik der Anreicherung der Daten mit Standards und Vokabularen, deren Modellierung in ein Datenmodell und die Konvertierung in ein Linked (Open) Data-Format Schritt für Schritt dokumentiert und führt zum Ergebnis der Case Study und Diskussion in Kapitel 9. Darin werden die Vorteile sowie die Grenzen semantischer Technologien für die Archäologie und die gewonnenen Erkenntnisse der Case Study diskutiert.

Diese Arbeit richtet sich an alle in der Archäologie Tätigen, welche vermehrt archäologische Daten erzeugen, verwalten, archivieren und verwenden. Die Technologien, die vorgestellt werden, sind nützlich, um Daten nachhaltig nutzen und wiederverwenden zu können. Die Arbeit soll die Aufmerksamkeit auf diesen Aspekt in der archäologischen Forschung, der Dokumentation der Ausgrabungen für das Bundesdenkmalamt und der Nachnutzung der Dokumentation erhöhen, sei dies im Rahmen von Forschungsprojekten, in Archiven, Instituten und Museen oder der täglichen Feldarbeit in Grabungsfirmen.

Teil 1 – Konzepte und Grundlagen semantischer Technologien

Integrationsfähigkeit, Interoperabilität und Wiederverwendungswert sind nur drei von vielen Eigenschaften, die für archäologische Daten vorteilhaft sind. Wie sie mithilfe von Datenmanagement, semantischen Technologien und Standards erreicht werden können und welche Konzepte dem Semantic Web und Linked (Open) Data zugrunde liegen, wird im ersten Teil der Arbeit analysiert.

Zuerst werden die Aspekte des Datenmanagements sowohl in Bezug auf Forschungsdaten im Allgemeinen als auch speziell in der Archäologie beschrieben. Im Anschluss werden das Semantic Web und Linked (Open) Data mit ihren fundamentalen Theorien und technischen Grundlagen im Detail erklärt. Eine wichtige Rolle spielen dabei auch Ontologien, wobei der Fokus auf die speziell für kunst- und kulturwissenschaftliche Bereiche ausgelegte Ontologie CIDOC CRM gelegt wird. Räumliche, zeitliche und terminologische Standards für die Archäologie werden im letzten Abschnitt ermittelt.

2 Datenmanagement in der Archäologie

Ein Überblick über den geläufigen Umgang mit Forschungsdaten und archäologischen Daten stellt die Ausgangslage dar. Um die Aspekte eines Semantic Webs oder Linked (Open) Data in der Archäologie zu beleuchten, ist es nützlich, zuerst einen Blick darauf zu werfen, wie mit digitalen Daten umgegangen wird. In diesem Kapitel werden verschiedene Faktoren betrachtet, die Einfluss auf das Datenmanagement in der Archäologie haben. Es ermittelt, wie mit offen zugänglichen Daten und dem Teilen der eigenen Forschungsdaten verfahren wird. Aber auch, ob und welche breit akzeptierten Standards existieren und wie es um die generelle Bereitschaft für neue Technologien im Feld der Archäologie bestellt ist. Der derzeitige Status-Quo wird betrachtet und diskutiert, welche Verbesserungen eventuell möglich oder nützlich sein können.

2.1 Forschungsdatenmanagement

Bevor die archäologischen Daten diskutiert werden, ist ein Überblick über das Datenmanagement in der österreichischen Forschungslandschaft sinnvoll. Die Rolle von digitalen Daten sowie deren Handhabung und Aufgaben wird zunehmend wichtig. Forschungsdatenmanagement ist essenziell für eine Forschung, die für Forschende und die Gesellschaft zugänglich, kompatibel mit weiteren

Forschungsergebnissen und fähig für Zusammenarbeit sein möchte. Diese Eigenschaften erhöhen auch die Nachvollziehbarkeit. Forschungsergebnisse sind besser überprüfbar und national und international sichtbarer.[5] Es gibt zum jetzigen Zeitpunkt keine einheitlichen Regeln für das Datenmanagement in der Wissenschaft. Doch sowohl von der internationalen als auch von der österreichischen Forschung wird nachhaltiges und transparentes Datenmanagement immer mehr beachtet. Forschungsförderer, Forschende und wissenschaftliche Einrichtungen erkennen die Notwendigkeit, einen Plan für die immer größer werdende Masse an kreierten Daten zu haben.[6]

Aber nicht nur das Management der Daten ist zu berücksichtigen, auch die Frage des öffentlichen Zugangs ist aktueller denn je geworden. Eine wichtige Rolle spielen dabei Forschungsförderer. Diese können Förderbedingungen aufstellen und setzen sich vermehrt für offen zugängliche Forschungsdaten ein. Die Europäische Kommission hat in den Richtlinien für Förderungen im Rahmen von „Horizon2020"[7] vorausgesetzt, dass Publikationen und Forschungsdaten öffentlichen Zugang erhalten müssen.[8] In Österreich sind bedeutsame Forschungsförderer beispielsweise die Europäische Kommission und der Österreichische Wissenschaftsfonds FWF (Fonds zur Förderung der wissenschaftlichen Forschung). Neben den Voraussetzungen zur Förderung bei der Europäischen Kommission hat auch der FWF in seinen Antragsrichtlinien als wesentlichen Punkt den Umgang mit den Forschungsdaten verankert. Seit 2019 sind Forscher*innen verpflichtet, einen offenen Zugang zu Forschungsdaten, welche durch die Finanzierung des FWF zustande kommen, herzustellen.[9]

Viele internationale Forschungseinrichtungen und Universitäten haben bereits Richtlinien für den offenen Zugang von Forschungsdaten und Datenmanagementpläne umgesetzt. Empfehlungen zum Umgang und dem offenen Zugang für Forschungsdaten werden beispielsweise von der League of European Research Universities gegeben.[10,11] In Großbritannien haben 80 Universitäten Richtlinien für die universitären Forschungsdaten veröffentlicht.[12] Der Mehrheit der österreichischen Forschungseinrichtungen ist bewusst, dass langfristige Konzepte und Richtlinien wichtig sind für die Archivierung und Verwendung ihrer Daten.[13] Die Universität Wien beispielsweise hat

[5] Bauer u. a. 2015, 6 f.
[6] Bauer u. a. 2015, 15.
[7] Siehe https://ec.europa.eu/programmes/horizon2020/en, abgerufen am 19.10.2021.
[8] European Commission 2017.
[9] Zur Open Access Policy für vom FWF geförderte Projekte siehe: https://www.fwf.ac.at/de/forschungsfoerderung/open-access-policy, abgerufen am 19.10.2021.
[10] Siehe: League of European Research Universities 2013, online unter https://www.leru.org/files/LERU-Roadmap-for-Research-Data-Full-paper.pdf, abgerufen am 19.10.2021.
[11] Siehe: League of European Research Universities 2018, online unter https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf, abgerufen am 19.10.2021.
[12] Horton – DCC 2016.
[13] Bauer u. a. 2015, 68.

ein umfangreiches Forschungsdatenmanagement ausgearbeitet, das den FAIR Guiding Richtlinien folgt (siehe Kapitel 2.1.2).[14]

2.1.1 Umfrage zu österreichischen Forschungsdaten

Die Universität Wien hat 2015 im Rahmen des Projekts e-Infrastructures Austria eine Befragung zum praktischen Umgang mit Forschungsdaten in Österreich realisiert, um einen Überblick zu gewinnen. Es wurden alle öffentlich-rechtlichen Universitäten und drei außeruniversitäre Forschungseinrichtungen befragt. Ziel war es, die Gegebenheiten der österreichischen Wissenschaftswelt im Umgang mit Forschungsdaten festzustellen.[15] Dies ist ein geeignetes Instrument, um abzuschätzen, inwieweit Forschende bereit sind, Open Data, also eine offene Datenpolitik, anzuwenden und ob eigene oder fremde Daten wiederverwendet werden. Ebenso kann die Befragung die vorliegende Infrastruktur evaluieren und einordnen. Die Ergebnisse der e-Infrastructures Austria zeigen außerdem auf, welche Datentypen bevorzugt verwendet werden, wie die Archivierung vorgenommen wird und ob Forschende die rechtlichen Fragen zum Datenschutz beantworten können: [16]

- Der Großteil der *Datentypen* besteht laut der Umfrage aus unstrukturierten Textdateien, Tabellen und Grafiken. In den Geisteswissenschaften und der Medizin werden im Verhältnis mehr Datenbanken verwendet als in den restlichen Forschungsgebieten.
- Die *Datenarchivierung* geschieht mehrheitlich über dienstliche oder private Rechner, USB-Laufwerke und externe Festplatten und wird von mehr als zwei Drittel der Forschenden als uneinheitlich beschrieben. Neun von zehn Forschenden sind für die Archivierung selbst verantwortlich. Datenarchive und Repositorien werden nur von jedem siebten verwendet.
- Bei *rechtlichen Fragen* haben ein Fünftel der Befragten Unklarheiten mit dem Datenschutz beim Verwenden von fremden Daten. Ethisch oder rechtlich sensible Daten werden in etwa von jedem oder jeder siebenten Forscher*in behandelt.
- Bei der *Zugänglichkeit und Wiederverwendung* von Daten ist deutlich zu sehen, dass für viele Wissenschafter*innen die Nutzung von Fremddaten einen wesentlichen Teil ihrer Forschung einnimmt. Ein Viertel der Befragten verwendet nie Daten von Dritten. Den Zugriff auf die eigenen Daten geben die meisten nur eingeschränkt, mehr als die Hälfte sogar nur auf Anfrage. Nur jeder oder jede zehnte Forscher*in veröffentlicht seine oder

[14] Für das Forschungsdatenmanagement der Universität Wien siehe: https://datamanagement.univie.ac.at/forschungsdatenmanagement/, abgerufen am 19.10.2021.
[15] Bauer u. a. 2015, 15.
[16] Bauer u. a. 2015, 7 f.

> ihre Daten als Open Data. Zehn Prozent der Befragten überlassen ihre Daten Kolleg*innen auch nicht auf Anfrage für eine weitere Nutzung. Ein Drittel der Befragten macht es möglich, seine Daten wiederzuverwenden. Die Forschenden teilen ihre Daten meist über E-Mail und externe Datenträger, aber auch über Cloud- und Webseiten-Anwendungen. In der Medizin, den Sozial- und Geisteswissenschaften werden Daten seltener als bei anderen Forschungsrichtungen herausgegeben. Die Vorteile, die sich Forscher*innen erhoffen, wenn sie ihre Daten teilen, sind beispielsweise höhere Sichtbarkeit ihrer Forschung, neue Kooperationsmöglichkeiten, Anerkennung und Berücksichtigung im Forschungsgebiet. Nachteile seien ein erhöhter Zeit- und Kostenfaktor, Datenmissbrauch, rechtliche Unsicherheit, mögliche Datenverfälschung, nicht erwünschte Kommerzialisierung und die Erhöhung des Konkurrenzdrucks.

Viele der Befragten sind der Meinung, die Voraussetzung für ein besseres Datenmanagement sei eine institutionelle, nationale oder fachspezifische Infrastruktur in Form von Datenarchiven und Repositorien. Wünschenswert sei auch eigens dafür vorgesehenes Personal und feste Richtlinien, an die sie sich halten können. Die Bereitschaft von Forscher*innen für organisiertes Datenmanagement und eine offenere Datenpolitik ist demnach durchaus vorhanden, wenn dafür Unterstützung in Form von Infrastruktur und Richtlinien zur Verfügung gestellt wird.[17] Inwiefern diese Infrastrukturen in Form von Repositorien für Archäologiedaten bestehen, welche Projekte diese bereits verwenden und welche Anforderungen es für Projektförderungen gibt, wird in Kapitel 2.2 näher erläutert.

2.1.2 Die FAIR Richtlinien (FAIR Guiding Principles)

Das Angebot an Daten in der Wissenschaft wird laufend größer und somit schwerer zu überblicken. Forscher*innen wünschen sich Unterstützung bei der Bewältigung dieser Herausforderungen durch Bestimmungen.[18] 2016 stellten M. D. Wilkinson u. a. Richtlinien für Dateninhaber*innen vor, welche den Wiederverwendungswert ihrer Daten erhöhen möchten.[19] Durch die Vorschläge können sich Forscher*innen orientieren und erhalten Hilfestellung im Umgang mit ihren Daten. Diese Richtlinien werden die FAIR Richtlinien (FAIR Guiding Principles) genannt. Das Akronym steht dabei für „Findability", „Accessability", „Interoperability" und „Reusability"

[17] Bauer u. a. 2015, 7 f.
[18] Bauer u. a. 2015, 7 f.
[19] Wilkinson u. a. 2016, 1.

(siehe Abbildung 1).[20,21] Daten sollen demnach im Netz gefunden, abrufbar, kompatibel und wiederverwendbar sein. Die Richtlinien sollen dies für Mensch und Maschine in Zukunft erleichtern. Es wird empfohlen, maschinenlesbare Daten zu veröffentlichen, da Menschen mehr und mehr Unterstützung durch Computer benötigen, um große Datenmengen zu verarbeiten. Denn es kommen stetig mehr Daten in Umlauf, die Daten werden komplexer und sie werden immer schneller und einfacher erstellt.[22] Auch in der Archäologie werden laufend mehr digitale Daten produziert, was zu einem höheren Bedarf an Datenstrukturen und Datenaustausch führt.[23] M. D. Wilkinson u. a. betonen den Vorteil der Branchenunabhängigkeit der Richtlinien, da unterschiedliche Forschungsgebiete davon profitieren würden.[24]

Datenmanagement ist essenziell in der Wissenschaft. Immer mehr öffentliche Forschungsförderer verlangen Management- und Verwaltungspläne für die erstellten Daten. Projekte müssen offenlegen, wie sie die Daten erfassen, archivieren und verwalten und Langzeitarchivierungspläne vorzeigen. Dies soll dazu beitragen, wichtige Forschungsdaten zu erhalten. Des Weiteren sollen Forscher*innen neue Daten finden und wiederverwenden können. Wissenschafter*innen profitieren davon ebenfalls mehrfach: zum einen wird ihre eigene Forschung geteilt und sie werden zitiert und zum anderen finden sie selbst andere Daten und können diese in ihrer Forschung verwenden.[25] Die FAIR Richtlinien können helfen, Daten transparent aufzubereiten und nachhaltig zur Verfügung zu stellen.[26]

M. D. Wilkinson u. a. weisen jedoch auf die besondere Herausforderung bei Wissenschaften mit hohem Datenaufkommen hin. Es sei schwierig, die Neuentdeckung von Wissen zu optimieren. Die Forscher*innen und ihre maschinellen Hilfsmittel sollen unterstützt werden, damit sie wissenschaftliche Daten entdecken und darauf zugreifen können und um sie integrieren und analysieren zu können.[27]

Die Archäologie lässt sich in die Kategorie von Wissenschaften mit erhöhter Datenproduktion einordnen. Die FAIR Richtlinien werden in der jüngeren Zeit bereits von einigen Projekten angewandt. Ein Beispiel ist das SEADDA-Projekt (Saving European Archaeology from the Digital Dark Age), das einerseits auf die Notwendigkeit hinweist, digitale archäologische Daten langfristig

[20] Wilkinson u. a. 2016, 1.
[21] 2019 wurde ein Addendum veröffentlicht mit der aktuellen Online-Repräsentation der FAIR Guiding Principles: https://www.go-fair.org/fair-principles/, abgerufen am 19.10.2021, siehe Wilkinson u. a. 2019, 1.
[22] Wilkinson u. a. 2016, 1.
[23] Filzwieser – Eichert 2020, 1385.
[24] Wilkinson u. a. 2016, 4.
[25] Wilkinson u. a. 2016, 3.
[26] Hiebel u. a. 2021a, 267-269.
[27] Wilkinson u. a. 2016, 2.

zu sichern und deren Zugang sicherzustellen und andererseits den „Open Data Konsens“ in der Archäologie zu stärken. [28]

Aber auch andere archäologische Projekte wenden die FAIR Richtlinien für ihre Forschungsdaten bereits an. Das Thanados-Projekt[29] des Naturhistorischen Museums in Wien und der Österreichischen Akademie der Wissenschaften hat das Ziel, Daten von ca. 500 frühmittelalterlichen Friedhöfen aus Österreich und seinen Nachbarländern online und mit freiem Zugang zur Verfügung zu stellen. Das Datenmodell, welches zur Strukturierung der Daten verwendet wurde, ist das CIDOC CRM-Modell (siehe Kapitel 5.2).[30] Es ist speziell auf Daten archäologischer und kulturhistorischer Herkunft ausgerichtet. Die Daten wurden mithilfe eines Online-Datenbanksystems für archäologische, historische und geografische Daten, dem OpenAtlas[31], eingegeben, welches vom Projektteam in vorigen Projekten entwickelt wurde. OpenAtlas ist webbasiert, kann online verwendet werden und ist eine freie Software. Es wird bereits von anderen Projekten verwendet, welche ihre Daten mit dem CIDOC CRM-Modell modellieren möchten.[32] Die Mehrzahl an Projekten, die das CIDOC CRM verwenden, benutzen es, um bereits existierende Daten zu „remodellieren“. Die OpenAtlas Software hat das Datenmodell bereits im Hintergrund mitintegriert. So können neu eingegebene Daten automatisch zu diesem Modell modelliert werden und dies, ohne die Notwendigkeit das ontologische Modell kennen zu müssen.[33] Dies ermöglicht eine niederschwellige Anwendung semantischer Technologien und eine nachhaltige, faire Datenproduktion für neue Projekte.

Das Projekt ORD4MiningArchaeo – „Open Research Data for Prehistoric Mining Archaeology“[34] des Forschungszentrums HiMAT (The History of Mining Activities in the Tyrol and Adjacent Areas - Impact on Environment & Human Societies) der Universität Innsbruck[35] setzt sich zum Ziel, Daten von Forschungsergebnissen des multinationalen DACH-Projektes (Prehistoric copper production in the eastern and central Alps)[36] der prähistorischen Bergbauforschung im Hinblick der FAIR Richtlinien aufzubereiten. Im Projekt wurden dafür ebenfalls Technologien des Semantic Web und das Schema des CIDOC CRM verwendet, um die Daten auffindbar, abrufbar, interoperabel und wiederverwendbar zu machen. Um Daten für andere Forscher*innen wiederverwendbar zu sein, müssen sie transparent sein. Dies ist ein zentraler Punkt der FAIR

[28] Siehe auch https://www.seadda.eu/?page_id=1274https://www.seadda.eu/?page_id=1274, abgerufen am 19.10.2021.
[29] Siehe auch https://thanados.net/about, abgerufen am 19.10.2021.
[30] Eichert 2021, 4.
[31] Siehe https://openatlas.eu/, abgerufen am 19.10.2021.
[32] Filzwieser – Eichert 2020, 1385 f.
[33] Eichert 2021, 6.
[34] Siehe https://www.uibk.ac.at/projects/ord4mining-archaeo/, abgerufen am 19.10.2021.
[35] Siehe https://www.uibk.ac.at/himat/, abgerufen am 19.10.2021.
[36] Siehe https://www.uibk.ac.at/himat/projekte/dach/index.html.de, abgerufen am 19.10.2021.

Richtlinien. Im ORD4MiningArchaeo-Projekt wird neben den Metadaten auch der Prozess der Datengewinnung dokumentiert und durch Verlinkungen der Daten wird der Wiederverwendungswert weiter erhöht.[37]

Im Folgenden werden die Richtlinien der FAIR Richtlinien im Detail betrachtet.[38] Die Verfasser*innen M. D. Wilkinson u. a. verwenden den Begriff „(meta)data", der in der Erläuterung mit (Meta-)Daten übernommen wird. Bei diesem Begriff soll das jeweilige Prinzip sowohl auf die Metadaten als auch auf die Daten angewandt werden.[39, 40]

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Abbildung 1 – Die FAIR Richtlinien (Quelle: Wilkinson u. a. 2016, 4).

Findable – Auffindbar

F1. „Die (Meta-)Daten brauchen global einzigartige, permanente Identifikatoren (Identifier)."[41] Dieses Prinzip wird als das wichtigste von allen angesehen. Denn ohne Identifikator sind die übrigen Richtlinien nur schwer umzusetzen. Die Verwendung von Identifikatoren soll Mehrdeutigkeiten und Unklarheiten verhindern und anderen Nutzern und Nutzerinnen helfen, die Daten eindeutig

[37] Hiebel u. a. 2021a, 276 f.
[38] Nach den Erläuterungen der Online-Repräsentation der FAIR Guiding Principles unter: https://www.go-fair.org/fair-principles/, abgerufen am 19.10.2021.
[39] Wilkinson u. a. 2016, 4 f.
[40] Siehe https://www.go-fair.org/fair-principles/, abgerufen am 19.10.2021.
[41] Siehe https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/, abgerufen am 19.10.2021.

zu lokalisieren.[42] Maschinen können Suchanfragen effizienter ausführen und die Integration mehrerer Datensätze wird erleichtert. Identifikatoren sollen bevorzugterweise als HTTP URIs (siehe Kapitel 4.1), die zu einer Webseite mit den entsprechenden Daten führen, angegeben werden.[43]

Identifikatoren sollen die folgenden zwei Eigenschaften besitzen: global einzigartig sein und dauerhaft zur Verfügung stehen. Als global einzigartig gilt der Identifikator, wenn er eine unverwechselbare Adresse hat und sich somit die von ihm gekennzeichneten Daten eindeutig lokalisieren lassen und in weiterer Folge von Nutzer*innen zitiert werden können. Permanent verfügbare Identifikatoren sollen nur Links verwenden, die aktiv sind. Dies kann beispielsweise mit Registrierungsservices[44] sichergestellt werden.[45] Als Beispiel eines permanenten Identifikators kann der DOI, der digitale Objektidentifikator (Digital Object Identifier), genannt werden. Dabei werden Objekten digitale Identifikatoren zugewiesen, welche sie identifizierbar machen.[46] Das digitale Langzeitarchiv Zenodo[47] vergibt für jedes Datenset, jede Publikation oder jede Software, welche gespeichert und abrufbar ist, eine DOI. Dem Datenset „Alpiner Ortsnamen Gazeteer“[48] wurde beispielsweise die DOI „10.5281/zenodo.3703076“ zugewiesen.

F2. „Die Daten werden mit umfangreichen Metadaten beschrieben.“[49] Metadaten sind laut wörtlicher Definition „Daten über Daten“.[50] Sie sollen so großzügig wie möglich beschrieben werden. Dies bedeutet, den Kontext, die Qualität, den Zustand und Charakteristika der Daten miteinzubinden. Die Beschreibung soll für eine breite Nutzergruppe hilfreich sein. Dabei sollte die angenommene Zielgruppe nicht vorweggenommen werden. Je ausführlicher die Gestaltung der Metadaten geschieht, umso besser stehen die Chancen für Daten, gefunden, für vertrauenswürdig befunden, wiederverwendet und zitiert zu werden. Als Beispiele für Metadaten können solche Daten genannt werden, wie sie teilweise von Maschinen automatisiert mitgeneriert werden. Bei digitalen Fotos werden zum Beispiel zahlreiche Hintergrundinformationen, wie der Zeitpunkt der Aufnahme, das verwendete Objektiv oder die GPS-Informationen mitaufgenommen. Aber auch Informationen zu Messinstrumenten und Aufnahmeprotokollen von Proben zählen als Metadaten.

---

[42] Deutz u. a. 2020, 53.
[43] McMurry u. a. 2017, 2.
[44] Einige Beispiele für Registrierungsservices sind unter https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/ aufgelistet, abgerufen am 19.10.2021.
[45] McMurry u. a. 2017, 1–3.
[46] ISO 2012, Introduction.
[47] Das Repositorium Zenodo wird vom CERN-Institut betreut und wurde im Zuge des Projekts OpenAIRE (siehe https://www.openaire.eu/about, abgerufen am 19.10.2021) entwickelt. Siehe für das Zenodo Repositorium: https://zenodo.org/, abgerufen am 19.10.2021.
[48] Der „Alpine Ortsnamen Gazeteer“ wurde im Zuge des Projekts Semantics for Mountaineering History (Österreichische Akademie der Wissenschaften) veröffentlicht (siehe unter http://www.semanticmountain.at/, abgerufen am 19.10.2021).
[49] Siehe https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/, abgerufen am 19.10.2021.
[50] Riley 2017, 1.

F3. „Metadaten beinhalten in eindeutiger Weise den Identifier für das Datenset selbst."[51] Die Metadaten und das Datenset selbst sind meist in getrennten Dateien enthalten. Deswegen ist es wichtig, die Verbindung eindeutig zu kennzeichnen und den global einzigartigen Identifikator, der das Datenset identifiziert, in den Metadaten explizit anzugeben. So können die Metadaten dem Datenset eindeutig zugeordnet werden.

F4. „(Meta-)Daten sind registriert oder indexiert in einer Ressource, die durchsuchbar ist."[52] Um die Auffindungsquote der Daten zu verbessern, können sie indexiert werden. Dies kann z.B. durch ein Repositorium oder einen Datenspeicher geschehen. Repositorien sind digitale Langzeitarchive, welche alle Arten von Forschungsresultaten, wie Publikationen, Daten, Software und Ähnliches, speichern und mit den ihnen zugeteilten Lizenzvereinbarungen veröffentlichen.[53] Passende Repositorien können beispielsweise bei OpenAIRE[54] oder re3data[55] gesucht werden. Bei re3data finden sich Repositorien wie Phaidra[56] der Universität Wien oder ARCHE[57] der Österreichischen Akademie der Wissenschaften. Der FAIR Data Point[58] wurde dafür entwickelt, um als Infrastruktur zur Veröffentlichung und Entdeckung von Metadaten zu fungieren.

Accessible – Zugänglich

A1. „(Meta-)Daten können anhand ihres Identifikators durch ein standardisiertes Kommunikationsprotokoll abgerufen werden."[59] Bei diesem Prinzip geht es um die Art und Weise, wie Nutzer*innen die (Meta-)Daten bekommen. Dies soll ohne Hürden und den Bedarf an proprietären oder speziellen Werkzeugen möglich sein. Die am häufigsten verwendeten Kommunikationsprotokolle hierfür sind HTTP (Hypertext Transfer Protocol) oder FTP (File Transfer Protocol).

> *A1.1. „Das Protokoll ist frei zugänglich, kostenlos verwendbar und universell umsetzbar."*[60] Um die Zugänglichkeit zu den Daten zu erhöhen, sollte das Protokoll kostenlos, frei zugänglich und weltweit einsetzbar sein. Die einzigen Voraussetzungen, die man als Nutzer*in

[51] Siehe https://www.go-fair.org/fair-principles/f3-metadata-clearly-explicitly-include-identifier-data-describe/, abgerufen am 19.10.2021.
[52] Siehe https://www.go-fair.org/fair-principles/f4-metadata-registered-indexed-searchable-resource/, abgerufen am 19.10.2021.
[53] Deutz u. a. 2020, 52.
[54] Um Repositorien zu suchen siehe https://explore.openaire.eu/participate/deposit/learn-how, abgerufen am 19.10.2021.
[55] Die Registrierungsseite für Repositorien „Registry of Research Data Repositories" findet sich unter https://www.re3data.org/, abgerufen am 19.10.2021.
[56] Siehe unter https://www.re3data.org/repository/r3d100010472, abgerufen am 19.10.2021.
[57] Siehe unter https://www.re3data.org/repository/r3d100012523, abgerufen am 19.10.2021.
[58] Siehe auch: https://github.com/FAIRDataTeam/FAIRDataPoint-Spec/ für das FAIR Data Point Repository, oder https://www.fairdatapoint.org/ für nähere Informationen, abgerufen am 19.10.2021.
[59] Siehe https://www.go-fair.org/fair-principles/ metadata-retrievable-identifier-standardised-communication-protocol/, abgerufen am 19.10.2021.
[60] Siehe https://www.go-fair.org/fair-principles/a1-1-protocol-open-free-universally-implementable/, abgerufen am 19.10.2021.

mitbringen muss, sollten ein Computer und Internetzugang sein. Diese Kriterien erfüllen beispielsweise HTTP und FTP.

A1.2. „Das Protokoll ermöglicht bei Bedarf Authentifizierung und Autorisierung."[61] FAIRe Daten sind nicht zwangsläufig zur Gänze freie und offen zugängliche Daten. Auch private oder sensible Daten können den FAIR-Prinzipien gerecht werden. Die Konditionen für einen Zugang müssen allerdings klar ersichtlich sein für Datennutzer*innen und im Idealfall von Maschinen interpretiert werden können. Benutzerkonten oder Anfragen für Repositorien können eine effiziente Möglichkeit darstellen, individuelle Zugriffsrechte zuzuweisen bzw. zu verwalten.

A2. „Die Metadaten sollen auch zugänglich sein, wenn die Daten selbst nicht mehr verfügbar sind."[62] Datensets online verfügbar zu halten und zu betreuen, kann Kosten verursachen. Daher können Daten in manchen Fällen nicht weiter verfügbar sein und verschwinden. Dabei führen an anderen Stellen angegebene Links ins Leere und bedeuten einen erheblichen Zeitverlust für Nutzer*innen. Es ist einfacher und weniger kostspielig, lediglich die Metadaten dauerhaft zur Verfügung zu stellen. Dieses Prinzip besagt, die Metadaten sollen auch zur Verfügung stehen, wenn die Originaldaten es nicht mehr sind. Es steht in Verbindung mit dem Punkt F4, der Registrierung und Indexierung der Metadaten. Der Vorteil darin liegt in der Möglichkeit, von der Existenz der Daten zu erfahren, auch wenn sie nicht mehr direkt verfügbar sind. Dadurch können Institutionen oder Forscher*innen dennoch kontaktiert werden.

Interoperable – Kompatibel

I1. „(Meta-)Daten verwenden eine formale, zugängliche, gemeinsame und allgemein anwendbare Sprache für die Wissensrepräsentation."[63] Die Daten sollen sowohl menschen- als auch maschinenlesbar sein. Um Daten zu generieren, welche auffindbar und interoperabel sind, seien zwei Punkte wesentlich: Erstens sollen kontrollierte Vokabulare und Thesauri verwendet werden, deren Definitionen über zugängliche Identifikatoren eingesehen werden können (siehe Kapitel 6). Zweitens soll ein Datenmodell mit einer Ontologie, welches die (Meta-)Daten strukturieren und beschreiben kann, benutzt werden (siehe Kapitel 5). Beispiele hierfür sind das RDF (Resource Description

[61] Siehe https://www.go-fair.org/fair-principles/a1-2-protocol-allows-authentication-authorisation-required/, abgerufen am 19.10.2021.
[62] Siehe https://www.go-fair.org/fair-principles/a2-metadata-accessible-even-data-no-longer-available/, abgerufen am 19.10.2021.
[63] Siehe https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/, abgerufen am 19.10.2021.

Framework, siehe Kapitel 4.3), um Daten zu strukturieren und die Ontologie Dublin Core (siehe Kapitel 5.1).

I2. „(Meta-)Daten verwenden Vokabulare, die den FAIR Richtlinien entsprechen."[64] Die Vokabulare, die Datensets beschreiben, müssen dokumentiert und für jeden im Netz mithilfe von weltweit einzigartigen Identifikatoren auffindbar sein. Dadurch können die Definitionen der Begriffe von jedem eingesehen werden und Missinterpretationen können verhindert werden.

I3. „(Meta-)Daten beinhalten qualifizierte Verweise zu anderen (Meta-)Daten."[65] Qualifizierte Verweise sind Querverweise, die die Absicht hinter den Verbindungen der (Meta-)Daten erklären sollen. Es sollen so viele bedeutungsvolle Links wie möglich kreiert werden, um den Inhalt des Datensets in einen Kontext zu stellen. Gibt es beispielsweise Datensets, die die Daten vervollständigen oder auf die verwiesen werden soll im wissenschaftlichen Kontext, so sollen diese mit identifizierbaren Links verbunden werden.

Reusable – Wiederverwendbar

R1. „Die (Meta-)Daten werden umfassend mit einer Vielzahl an genauen und relevanten Attributen beschrieben."[66] Um Daten effizienter wiederverwenden zu können, ist es hilfreich, die (Meta-)Daten so genau wie möglich zu beschreiben. Den potenziellen Datennutzer*innen soll die Entscheidung erleichtert werden, ob die Daten für den entsprechenden Zweck nützlich und verwendbar sind. Die Metadaten sollten demnach mehr Informationen enthalten, als nötig sind, um die Daten nur zu finden. Der Kontext, in dem die Daten entstanden sind, ist wichtig. Es sollen so viele Informationen über die Daten wie möglich dargestellt werden. Denn die Dateninhaber*innen können im Vorfeld nicht wissen, wer ihre Daten wiederverwenden möchte.

Beispiele, um die Daten genau zu beschreiben, wären, genau anzugeben, welche Methoden und Geräte zur Datengewinnung verwendet wurden und welche Bedeutung jede Spalte in einer Datenbank hat. Die Daten sollen verständlich und mit frei zugänglichen Programmen lesbar sein. Fachspezifische Abkürzungen sollen unbedingt vermieden werden. Die Daten sollen im Kern beschrieben werden, um den generellen Grund hinter der Datenerfassung zu erläutern. Darüber hinaus soll auf potenzielle Schwächen oder Grenzen der Daten hingewiesen werden. Das Datum der Herstellung, die Laborbedingungen/Ausgrabungsumstände, Verantwortliche und Details zur

[64] Siehe https://www.go-fair.org/fair-principles/i2-metadata-use-vocabularies-follow-fair-principles/, abgerufen am 19.10.2021.
[65] Siehe https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/, abgerufen am 19.10.2021.
[66] Siehe https://www.go-fair.org/fair-principles/
r1-metadata-richly-described-plurality-accurate-relevant-attributes/, abgerufen am 19.10.2021.

Datenerfassung wie Namen der Hardware und Software sollen angegeben werden. Den Zustand der Daten zu kennen ist ebenfalls wichtig. Denn die Information, ob es sich um Rohdaten oder um bereits verarbeitete Daten handelt, ist nützlich für den weiteren Gebrauch. Alle Namen bzw. Vokabel sollen in einer Felderläuterung erklärt werden. Versionierungen der Daten sollen ebenfalls gut dokumentiert und angegeben werden. Diese Angaben erleichtern eine Wiederverwendung der Daten und fördern die Transparenz.

R1.1. „(Meta-)Daten werden mit eindeutiger und zugänglicher Benutzer-Lizenz veröffentlicht."[67] Die Daten können nur wiederverwendet werden, wenn die Lizenz zur Nutzung der Daten definiert wird. Gebräuchliche Lizenzen sind zum Beispiel Creative Commons[68]. Die Konditionen zur Nutzung der Daten sollen in menschen- und maschinenlesbarer Form angegeben werden.

R1.2. „(Meta-)Daten werden mit genauer Herkunftsangabe veröffentlicht."[69] Nutzer*innen sollen wissen, woher die Daten stammen, wie die Verfasser*innen zitiert werden möchten und unter welchen Bedingungen die Daten weiterverwendet werden können. Genau definierte Quellen- und Rechtsangaben erhöhen die Bereitschaft von Nutzer*innen, die Daten wiederzuverwenden. Zur besseren Darstellung der Herkunft sollte auch eine Beschreibung des Workflows und möglicherweise der Prozessierung vorhanden sein. Hinweise, ob die Daten zuvor schon veröffentlicht wurden oder andere Autor*innen Daten hinzugefügt haben bzw. die Daten von anderen hinzugefügt wurden sind ebenfalls hilfreich.

R1.3. „(Meta-)Daten entsprechen den Standards der jeweiligen Domäne."[70] Daten sind einfacher wiederzuverwenden, wenn in der Domäne existierende Standards eingehalten werden (für Standards in der Archäologie siehe Kapitel 2.2.1). Hier sind beispielsweise Dateiformate, Dokumentation, Ontologien oder Vokabulare gemeint. Sollten „Best Practices" bzw. allgemein anerkannte Vorgehensweisen bestehen, wie die Daten archiviert oder angeboten werden sollen, ist es ratsam, diese zu berücksichtigen.

Die FAIR Richtlinien ermöglichen Wissenschafter*innen ein durchdachtes Datenmanagement. Eine wichtige Rolle spielen dabei auch Repositorien, in denen Daten gespeichert werden können. Sie sind wichtig, um die Daten auffindbar und identifizierbar zu machen. Es existieren bereits

[67] Siehe https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/, abgerufen am 19.10.2021.
[68] Siehe auch https://creativecommons.org/about/cclicenses/ zu Creative Commons Lizenzen, abgerufen 19.10.2021.
[69] Siehe https://www.go-fair.org/fair-principles/r1-2-metadata-associated-detailed-provenance/, abgerufen am 19.10.2021.
[70] Siehe https://www.go-fair.org/fair-principles/r1-3-metadata-meet-domain-relevant-community-standards/, abgerufen am 19.10.2021.

viele Datenrepositorien, eine Liste von Beispielen findet sich auf der „How to FAIR“ - Webseite[71]. Neben Repositorien können Forscher*innen auf dieser Webseite auch Hilfestellungen zu Datenmanagement-Plänen, Identifikatoren und Lizenzen finden, die den Einstieg in die Thematik erleichtern sollen.[72] Die Möglichkeiten zur Langzeitarchivierung speziell für archäologische Forschungsdaten werden in Kapitel 2.2.3.3 näher beschrieben.

D. B. Deutz u. a. fassen auf ihrer den FAIR Richtlinien gewidmeten Homepage die sechs wichtigsten und fundamentalsten Praktiken zusammen, um FAIRe Daten zu erhalten: die Daten müssen dokumentiert sein, es müssen die richtigen Dateiformate ausgewählt werden, die Metadaten müssen ausreichend Auskunft über die Daten geben, der Zugang zu den Daten muss geregelt sein, die Daten müssen mit einer passenden Lizenz versehen werden und es müssen permanente Identifikatoren für die Daten vergeben werden.[73] Von diesem Startpunkt aus können Wissenschafter*innen gemeinsam ein globales Netz an FAIRen Daten entstehen lassen, die auch von Maschinen gelesen werden können. Für das künftige Wissenschafts-Datenmanagement können diese Richtlinien von großem Nutzen sein.

2.2 Archäologische Daten

Archäologische Daten werden durch verschiedene Methoden generiert, dabei sind jedoch oft die Standards der Dokumentation und die verwendeten Terminologien uneinheitlich. Dadurch sind die Daten meist sehr heterogen.[74] Die Digitalisierung schreitet schnell voran und die Archäologie produziert eine Vielzahl an Daten. Daher sind Fragen des Datenmanagements und Standards für die Dokumentation und Metadaten sowie der Terminologien aktueller denn je.

Diese Arbeit geht vor allem von den Anforderungen archäologischer Forschungsdaten aus. Ein Großteil archäologischer Daten stammt jedoch auch aus dem Denkmalschutz und den unzähligen archäologischen Maßnahmen, welche aufgrund von aktuellen Bauarbeiten durchgeführt werden (Rettungsgrabungen). Diese Daten weisen andere Herausforderungen auf. Die Themen dieser Arbeit, wie Interoperabilität, Wiederverwendung, Langzeitarchivierung und Linked (Open) Data können und sollen jedoch auch auf diese Daten Auswirkungen zeigen. Um Daten sinnvoll weiter nutzen zu können, sollten sich auch Fragen zum Austausch und dem Zusammenwirken der Daten

[71] Datenrepositorien und Archive sind unter folgender Adresse aufgelistet: https://www.howtofair.dk/links-additional-reading/#data-repositories-and-archives, abgerufen am 19.10.2021.
[72] Siehe hierfür https://www.go-fair.org/resources/rdm-starter-kit/, abgerufen am 19.10.2021.
[73] Siehe Deutz u. a. 2020, auf der Homepage unter folgendem Link: https://www.howtofair.dk/what-is-fair/, abgerufen am 19.10.2021.
[74] Hagmann 2018, 55.

gestellt werden. Daten interoperabel und integrationsfähig zu produzieren, leistet hier einen wertvollen Beitrag. Viele Wissenschafter*innen sehen weiters noch ungeahnte Chancen für eine transparente Forschung durch Open Access und Open Data, um Daten wiederverwenden zu können. Doch ebenso wichtig ist, Daten in Repositorien nachhaltig zu sichern, um ein „digitales dunkles Zeitalter" zu verhindern. Wie diese Themen in der Archäologie aktuell diskutiert werden und in weiterer Folge gelöst werden könnten, wird in diesem Kapitel beleuchtet.

2.2.1 Datenmanagement und Standards

2.2.1.1 Definition von Datenmanagement

Datenmanagement ist ein wichtiger Bestandteil der Forschungsarbeit. Beobachtungen werden dabei als Forschungsdaten definiert, organisiert, dokumentiert, gesammelt, verarbeitet und archiviert. Die Ausführung beeinflusst die Qualität der resultierenden Daten und deren Wiederverwendungswert. Das Datenmanagement betrifft den kompletten Lebenszyklus der Daten, den sogenannten „Data Life Cycle". Dieser Zyklus wird in drei Phasen unterteilt (siehe Abbildung 2): die Ursprungsphase, in der die Daten gesammelt werden, die aktive Phase, in der sich weitere Daten ansammeln und verändern und die inaktive Phase, in der keine neuen Daten hinzukommen und die Daten bereits unverändert bleiben. Die FAIR Richtlinien (siehe Kapitel 2.1.2) kommen dabei in jeder Phase zum Einsatz.[75]

M. Zozus beschreibt die Mindestanforderungen an das Datenmanagement mit zwei Komponenten: Datenqualität und Rückverfolgbarkeit. Hierbei ist die Qualität notwendig, um die Forschungsfragen beantworten zu können. Die Reproduzierbarkeit ist wichtig, um nachvollziehen zu können, wie aus den Rohdaten die verarbeiteten Daten entstanden sind. Dies entscheidet auch, ob die Daten konsistent sind.[76]

Abbildung 2 – Phasen des Datenmanagements (Grafik: M. Simon nach Zozus 2017, 3).

[75] Zozus 2017, 1–3.
[76] Zozus 2017, 9.

2.2.1.2 Datenmanagement und Standards in der Archäologie

Das Kernziel der Archäologie ist, zu versuchen, mit Hilfe menschlicher Hinterlassenschaften vergangene Gesellschaften und die dahinterstehenden Menschen zu erforschen. Für diesen sehr weitreichenden und komplexen Anspruch werden unterschiedlichste Theorien und Methoden verwendet und es wird mit vielen anderen Disziplinen zusammengearbeitet. Das macht die Archäologie zu einem diversen, breitgefächerten und interdisziplinären Feld. Durch diese Heterogenität und Interdisziplinarität würden Standards durchaus helfen, Informationen effektiver austauschen zu können. Doch gerade diese Eigenschaften erschweren es Schematisierungen, breit anerkannte Standards oder ein weitläufig anwendbares Datenmanagement zu etablieren und alle Bereiche zufrieden zu stellen. Hinzu kommt die Tatsache, dass auch die zugrundeliegenden Daten dem Forschungsgegenstand entsprechend ebenso heterogenen und zudem hochgradig fragmentiert überliefert sind.[77]

Ein weiterer Punkt, der es der Archäologie erschwert, Standards festzusetzen, ist die Situation der archäologischen Archive. Diese werden mehrheitlich als unabhängige, geschlossene Einheiten betrachtet. Sie waren ein Neben- bzw. Endprodukt individueller Projekte und Ausgrabungen und wurden oft von Einzelpersonen oder einem kleinen Team betreut. In solch kleinen Rahmen war es problemlos, eigene Aufnahme- und Dokumentationssysteme zu verwenden. In größeren Projekten oder Forschungsverbünden, in denen unterschiedliche Ausgrabungs-, Prospektions-, und Analyseergebnisse miteinander kombiniert zu neuen Erkenntnissen führen sollten, waren Integrations-Strategien von Vorteil. Dies geschah zum Beispiel durch offline bearbeitete Datenbanken.[78] Auch heute noch wird diese Praxis in archäologischen Einrichtungen, wie Museen oder Archiven, ausgeübt, vor allem, wenn es an finanziellen Förderungen oder wissenschaftlichen Anreizen fehlt.

Das Feld der universitären und institutionellen archäologischen Forschung stellt nur einen kleinen Teil der Daten produzierenden Archäolog*innen dar. Grabungsfirmen arbeiten in Österreich stetig an zufällig georteten Fundstellen und geben die Dokumentation gemäß den Richtlinien des Bundesdenkmalamtes weiter. Museen, Galerien, Archive und Bibliotheken sind wiederum für die Kuration des Materials zuständig und haben jeweils ihre eigenen Methoden zur Datenaufnahme.[79] Alle Produzenten von Daten verfolgen dabei andere Ziele mit unterschiedlichen finanziellen Möglichkeiten.

[77] Isaksen 2011, 7.
[78] Richards – Hardman 2008, 168.
[79] Isaksen 2011, 8.

Ein weiterer Punkt, welcher die Etablierung von Standards betrifft, ist ein gewisses Misstrauen einiger Archäolog*innen gegenüber Daten-Standards. Dies könnte auf vorangegangene Versuche, archäologische Arbeitsweisen zu standardisieren und auf die dadurch befürchtete Kontrolle und Beschränkungen von archäologischen Interpretationen und Freiheiten zurückzuführen sein.[80] Doch mit der weiteren Entwicklung von Online-Ressourcen und der Digitalisierung wurden einige Vorteile sichtbarer. Etwa die Notwendigkeit bei datenübergreifenden Suchen und Analysen.[81,82] Darüber hinaus sind archäologische Fachbereiche, welche eng mit Naturwissenschaften, wie etwa der Archäometallurgie oder der Bioarchäologie, oder mit Formalwissenschaften, wie Statistik, zusammenarbeiten stärker mit standardisierter Dokumentation konfrontiert.

Daher wäre es eine Möglichkeit, die Archäologie unter anderen Wissenschaften und deren Bereitschaft für Datenmanagement und Standards zu untersuchen, indem Standards mit „Naturwissenschaftlichkeit" verknüpft wird. Wenn die Verwendung von Standards der Messfaktor für „Naturwissenschaftlichkeit" wäre, würde sich nach J. D. Richards die Archäologie in etwa in der Mitte befinden bzw. sich sogar eher in Richtung der Naturwissenschaften bewegen, da sie oft sehr eng zusammenarbeiten. Er lokalisiert die Archäologie im Hinblick auf die Anwendung und den daraus bestehenden Nutzen von Standards zwischen den Natur- und Geisteswissenschaften.[83] L. Isaksen sieht bei den Naturwissenschaften eine Mischung aus Abhängigkeit und Nutzen aus den vielschichtigen Datenmanagementsystemen und Standards. Die Geisteswissenschaft hingegen ist nicht abhängig davon und kann auch ohne solche komplexen Systeme funktionieren. Er sieht die Archäologie wie J. D. Richards ebenfalls zwischen diesen beiden Feldern angesiedelt. Dabei ist es für L. Isaksen unklar, ob und wann komplexes Datenmanagement unerlässlich sein wird. Bis 2011 sah er noch keinen Druck auf die Archäologie ausgeübt, breitgefächerte Standards oder schwierig anzuwendende Technologien einzuführen.[84] J. D. Richards sieht die Archäologie jedenfalls offener eingestellt gegenüber Standards als vergleichbare Geisteswissenschaften, wie etwa die Geschichtswissenschaft. In der Archäologie nimmt die Aufmerksamkeit, Daten wiederverwendbar und interoperabel zu generieren, merklich zu.[85] Dabei gibt es allerdings Variationen innerhalb des Feldes. J. D. Richards merkt an, es wäre beispielsweise einfacher, mehrere Datenbanken mit Tierknochen zu integrieren, als komplexere Keramik- oder Stein-Datenbanken.[86]

[80] Baines – Brophy 2006, 210.
[81] Richards 2009, 31.
[82] Richards – Hardman 2008, 168.
[83] Richards 2009, 27.
[84] Isaksen 2011, 7.
[85] Richards – Hardman 2008, 170.
[86] Richards 2009, 31.

Doch es gibt bereits eine Vielzahl an Standards, die sich durchaus früh, wenn auch oft nur regional oder national, durchgesetzt haben. Die Methode der Typologie könnte als einer der ersten Standards gelten. Sie wurde 1903 von O. Montelius veröffentlicht[87,88]. Darüber hinaus werden beispielsweise einheitliche Koordinatensysteme zur Vermessung von archäologischen Maßnahmen verwendet. Für alle archäologischen Maßnahmen gelten in Österreich die „Richtlinien für archäologische Maßnahmen" des Bundesdenkmalamtes[89]. Die Standards und Richtlinien des Bundesdenkmalamts sind wohl die wichtigsten, da sie verbindlich sind. Darin werden die Standards zur Dokumentation, Vermessung, Behandlung von Funden bis hin zu Dateiformaten und der abzugebenden Ordnerstruktur erläutert.

Die stärkere Technologisierung und Nutzung von Computern lässt Datensammlungen immer schneller anwachsen. Durch archäologische Messdaten von Ausgrabungen oder die Möglichkeiten der naturwissenschaftlichen Beprobung und Auswertung sammeln sich viele Daten an. Hinzu kommen die zahlreichen Fund- und Befunddatenbanken. Darum wird in der Archäologie die Wichtigkeit von Datenmanagement und anwendbaren Standards deutlicher, wenn die Menge an Daten bewältiget und nachhaltig genutzt werden möchte.

2.2.1.3 Digitale Standards in der Archäologie

International werden von der Internationalen Organisation für Normung (ISO)[90] Normen definiert. Diese werden oft aus vorhergehenden Standards, welche von Konsortien entwickelt werden, erarbeitet. Ein Beispiel eines Konsortiums ist das World Wide Web Consortium (W3C)[91], welches Standards wie HTML oder XML veröffentlicht hat[92]. Digitale Standards betreffen die Art und Weise, wie Daten dokumentiert und beschrieben werden ebenso wie die verwendeten Hilfsmittel. Im Folgenden werden die Standards näher beschrieben, welche für die Archäologie besonders relevant sind. Sie können grob in die folgenden drei Kategorien eingeteilt werden (siehe Abbildung 3):[93]

1. Technische Standards – Hard- und Software
2. Inhaltliche Standards – System in der Datenaufnahme

[87] Eggert 2005, 181–186.
[88] Montelius 1903.
[89] Bundesdenkmalamt 2018.
[90] Siehe unter https://www.iso.org/standards.html, abgerufen am 19.10.2021.
[91] Siehe unter https://www.w3.org/Consortium/, abgerufen am 19.10.2021.
[92] Standards des W3C sind unter https://www.w3.org/standards/ einzusehen, abgerufen am 19.10.2021.
[93] Richards 2009, 28.

3. Metadaten Standards – Dokumentation der Daten

Abbildung 3 - Digitale Standards in der Archäologie (Grafik: M. Simon nach Richards 2009, 28).

Unter den technischen Standards können vor allem Dateiformate, Kommunikations- und Computerstandards verstanden werden. Die Computerstandards, die sich in der Archäologie etabliert haben, wurden großteils vom Markt bestimmt. Microsoft und Apple sind die gängigen Betriebssysteme, ebenso wie Microsoft Office den Bereich der Textverarbeitung, Tabellen und Datenbanken prägt. Andere Software wie AutoCAD und ArcGIS oder QGIS sind in der Archäologie ebenso etabliert. Wichtig sind jedoch auch die Dateiformate u. a. von Tabellen, Datenbanken, Text- und Bilddateien, die produziert werden. Diese sollen mit anderen Programmen kompatibel und auch nachhaltig hinsichtlich der weiteren Archivierung sein.[94] Durch die immer öfter eingesetzten Methoden der dreidimensionalen Visualisierung von Befunden und Funden wurden 2012 die Prinzipien von Sevilla[95] veröffentlicht. Sie bilden die Standards für 3D-Visualisierungen in der Archäologie und wurden aus der Londoner Charta[96] für digitale Visualisierungen historischer Objekte aus dem Bereich des Kulturerbes heraus entwickelt.[97]

An dieser Stelle sei auf die Empfehlungen des IANUS – Forschungsdatenzentrums[98] hingewiesen. Dieses Forschungszentrum für Archäologie und Altertumswissenschaften behandelt Fragen des Datenmanagements und bietet außerdem ein Langzeitarchiv für digitale Daten. Es stellt Vorschläge für den Umgang mit digitalen Daten in den Altertumswissenschaften vor und gibt IT-Empfehlungen für Datenmanagement, Dateiformate, Forschungsmethoden und Projektphasen, die sicherstellen sollen, Daten nachhaltig nutzen zu können.

Für Anleitungen, Beschreibungen und Beispiele für Standards, die in der Archäologie Tätigen helfen sollen, einen Überblick zu gewinnen und diese Standards verwenden zu können, gibt es

[94] Richards 2009, 28.
[95] ICOMOS 2017.
[96] Denard 2009.
[97] Printz 2014, 18–22.
[98] Siehe https://www.ianus-fdz.de/it-empfehlungen/, abgerufen am 19.10.2021.

neben den IANUS-Richtlinien beispielsweise weiters das FISH-Forum und das CIfA-Institut. Das „FISH - Forum on Information Standards in Heritage“[99] ist ein Forum, welches Links und Informationen bereitstellt für die „Best Practice“, also das empfohlene Vorgehen, bei der Dokumentation von Kulturerbe. Auf der Webseite findet sich ebenfalls eine Vielzahl an frei verwendbaren Thesauri[100] in englischer Sprache. „CIfA - Chartered Institute for Archaeologists”[101] ist ein britischer, staatlich anerkannter Berufsverband, der Standards und Leitfäden entwickelt. Es ist auf britische Archäolog*innen zugeschnitten, doch es existieren auch Arbeitsgruppen in Deutschland und Australien.[102]

Inhaltliche Standards, wie die Dokumentation und Beschreibung von Ausgrabungen oder Funden, Vokabulare, Typologien und Terminologien, werden bereits seit langer Zeit diskutiert. Wenn alle Ausgrabungen auf die gleiche Weise dokumentiert werden würden, würde dies die Kontrolle der Qualität und Vergleiche erleichtern. Auf der anderen Seite nimmt jede Art der Dokumentation Bezug auf eine konkrete Fragestellung. Deshalb ist es nicht möglich, objektiv und allumfassend zu dokumentieren. Auch das Bundesdenkmalamt verweist in seinen Richtlinien auf die Möglichkeit von inhaltlichen Unterschieden: „Inhaltliche Abweichungen von den gegenständlichen »Richtlinien für archäologische Maßnahmen« können aufgrund besonderer Rahmenbedingungen, besonderer Befundsituationen oder besonderer Projektziele sinnvoll sein oder auch von äußeren Umständen erzwungen werden.“[103] Inhaltliche Standards und Vokabelkontrolle sind jedoch nützlich, um beispielsweise die Muster-Erkennung zu ermöglichen. Diese ist wichtig für Fundklassifikationen, Typologien und Objektgruppenbenennungen.[104]

Wenn Fragen an archäologische Daten gestellt werden, können sie in den meisten Fällen mit „Wann“, „Wo“ und „Was“ beantwortet werden. In diesen Kategorien schlagen A. Baines und K. Brophy eine Standardisierung vor, um Daten interoperabel zu kreieren. Das „Wann“ ist besonders wichtig, wenn Suchanfragen erfolgreich sein sollen. Die Terminologien gänzlich zu vereinheitlichen ist allein wegen der regionalen Unterschiede der Zeitperioden und deren Benennungen nicht möglich. Besser wäre eine örtlich getrennte Definition einer Zeitperiode. Die „Wo“-Frage ist einfacher zu lösen. Durch Längen- und Breitengrade sind Lokalisierungen einfach zu beschreiben

[99] Siehe http://www.heritage-standards.org.uk/, abgerufen am 19.10.2021.
[100] Siehe http://www.heritage-standards.org.uk/fish-vocabularies/, abgerufen am 19.10.2021.
[101] Siehe https://www.archaeologists.net/codes/cifa, abgerufen am 19.10.2021.
[102] Wait 2017, 126.
[103] Bundesdenkmalamt 2018, 2 f.
[104] Richards 2009, 28–30.

und zu definieren. Eines der wichtigsten Konsortien für Geodaten ist das Open Geospatial Consortium (OCG)[105].[106] Es entwickelt Standards und arbeitet mit der ISO/TC211 – Geographic information/Geomatics[107] zusammen, welche die Standards überprüft und als internationale Norm übernimmt.

Die „Was"-Frage ist am schwierigsten zu standardisieren. Es gibt viele verschiedene rivalisierende Thesauri, die unterschiedliche Artefaktgruppen in unterschiedlichen Ländern und Zeiten abdecken.[108] In Kapitel 6.4 wird genauer auf diese Aspekte eingegangen und Lösungsvorschläge vorgestellt. A. Baines und K. Brophy betonen, Thesauri seien wichtig, um Daten besser durchsuchen zu können.[109] Bei archäologischen Daten handelt es sich nicht um „Wahrheiten" über Funde und Befunde. Daten stehen miteinander im Kontext und Forschende im Bereich der Archäologie versuchen materielle Objekte und Befunde zu beschreiben und den Kontext herzustellen.[110] J. D. Richards meint, zentralisierte Thesauri wären unmöglich einzuführen in der Archäologie, da das Feld viel zu breit gefächert und jedes Projekt zu speziell dafür ist. Es würde wohl nie zu einer Einigung kommen. Stattdessen gibt es einen breiteren Konsens für dezentrale Datensysteme. Diese können an die jeweiligen Bedürfnisse angepasst werden und im Anschluss können von den Dateninhaber*innen Maßnahmen vorgenommen werden, die Daten interoperabel machen.[111]

Um Standards auf europäischem Level einheitlicher zu gestalten, müssen weitere Faktoren bedacht werden. Hier spielen die unterschiedlichen Sprachen und archäologischen Konzepte eine Rolle.[112] Einen international verwendbaren archäologischen Thesaurus zu entwickeln, würde Jahrzehnte an Arbeit bedeuten.[113] Diesem Thema widmen sich bereits einige Projekte, wie zum Beispiel ARIADNE Plus[114] oder SEADDA[115] (Näheres hierzu wird in Kapitel 7.1 erläutert).

Die dritte Gruppe der Standards, die Metadaten Standards, sind wichtig, um Daten interoperabel, integrierbar, wiederverwendbar und auffindbar zu machen. Um schnelle, akkurate und vollständige Suchergebnisse zu Metadaten zu bekommen, ist die Lösung ihre Standardisierung. Diese Standards sind flexibel anwendbar und schränken den Herausgeber nicht ein in der Gestaltung der Daten. Sie geben Schemata vor, um die „Daten über die Daten" vergleichbar und zuordenbar

[105] Siehe https://www.ogc.org/, abgerufen am 19.10.2021.
[106] Printz 2014, 25.
[107] Siehe unter https://www.iso.org/committee/54904.html oder https://committee.iso.org/home/tc211, abgerufen am 19.10.2021.
[108] Richards – Hardman 2008, 170–172.
[109] Baines – Brophy 2006, 210.
[110] Baines – Brophy 2006, 211.
[111] Richards 2009, 30.
[112] Oberländer-Târnoveanu 2005, Kapitel 3.
[113] Richards 2009, 32.
[114] Siehe https://ariadne-infrastructure.eu/, abgerufen am 19.10.2021.
[115] Siehe https://www.seadda.eu/, abgerufen am 19.10.2021.

zu machen und den Inhalt der Daten für den oder die Nutzer*in schnell ersichtlich zu machen.[116] Zwei international etablierte ISO Standards[117] für Metadaten sind beispielsweise die Ontologien Dublin Core[118] und CIDOC-CRM[119] (siehe zu beiden Ontologien Kapitel 5). Dublin Core ist ein Set von 15 Metadaten-Kernelementen, die eine domänenübergreifende Nutzung zur Beschreibung von Ressourcen erlauben. Es wird von der Dublin Core Metadata Initiative betreut.[120] CIDOC CRM wurde speziell für den Kulturbereich entwickelt, um den Informationsaustausch zu erleichtern.[121,122]

Diese drei Ebenen von Standards können sicherstellen, dass die Daten stimmig sind, die Dokumentation eindeutig festgehalten wurde und der Prozess der Datenerstellung dokumentiert wurde. Nur wenn diese Kriterien erfüllt sind, so S. Stead, sind die entstandenen Daten auch geeignet, um wiederverwendet zu werden.[123]

2.2.1.4 Vorteile digitaler archäologischer Standards

Um herauszufinden, wie stark die Motivation für Archäologen und Archäologinnen ist, in Datenmanagementsysteme zu investieren und sich an bereichsspezifische Standards zu halten oder komplexe Technologien anzuwenden, sollte man einen Schritt zurückgehen und sich ansehen, worin der Nutzen dabei bestehen würde. Dieser liegt hinsichtlich der Standards darin, Daten interoperabel zu machen, heterogene Datensets analysieren zu können und darin, Daten im Internet zu finden und verwenden zu können.[124]

Mit der Etablierung des Internets und der Online-Publikation und -Dissemination von Daten und Ergebnissen wurde das Potential, Daten anderer Personen wiederzuverwenden, erkannt und gefördert. Die Voraussetzung für den Austausch von Daten sind gemeinsam verwendete Standards.[125] J. D. Richards und C. S. Hardman sehen viele neue Möglichkeiten für archäologische Daten, wenn sie standardisiert und interoperabel generiert werden. Unterschiedliche Projekte und Forscher*innen können Daten austauschen und nutzen und dabei neue unbekannte Ressourcen entdecken, die für sie von großem Interesse sind und ansonsten unentdeckt geblieben wären.[126]

[116] Wise – Miller 1997, Kapitel 3.
[117] Siehe https://www.iso.org/standards.html, abgerufen am 19.10.2021.
[118] Siehe zu Dublin Core Metadata Initiative http://dublincore.org/, abgerufen am 19.10.2021.
[119] Siehe http://www.cidoc-crm.org/, abgerufen am 19.10.2021.
[120] ISO 2017b.
[121] ISO 2014.
[122] Richards 2009, 30.
[123] Stead 2007, 23 f.
[124] Isaksen 2011, 8.
[125] Richards – Hardman 2008, 168.
[126] Richards – Hardman 2008, 168–170.

Ebenso vorteilhaft stellen sich Standards bei Suchanfragen im Internet heraus. Man könnte meinen, mit der Google-Suche finden sich alle Daten und Dokumente im Internet und strukturierte Daten wären überflüssig. Google verwendet unstrukturierte Freitextdaten und gibt die relevantesten Ergebnisse heraus. Diese Suchmethode wird auch die „type and hope"-Philosophie genannt.[127] Dabei kann es geschehen, nachdem unzählige Suchergebnisse händisch durchgesehen wurden, dass sich die Mehrheit als unbrauchbar herausstellt. Dies verschlingt viel Zeit und Ressourcen. Durch standardisierte Ontologien und Vokabulare, wie sie auch für das Semantic Web verwendet werden, kann die Recherche erheblich erleichtert werden. J. D. Richards betont, Standards und Ontologien, auf die sich international geeinigt wurde, sind eine Voraussetzung für das Semantic Web. So kann die Bedeutung und die Beziehung der Daten richtig zugeordnet und Suchanfragen weitaus effizienter gemacht werden.[128]

Die Vorteile von Datenmanagement und Standards in der Archäologie sind demnach sehr vielfältig: Daten können besser genutzt und wiederverwendet werden. Weiters sind interoperable und konsistente Daten besser geeignet, um große Datenmengen zu bewältigen. Darüber hinaus können Forscher*innen völlig neue Datenquellen durch die Online-Publikation der Daten entdecken, welche ansonsten eher nicht gefunden worden wären. Die Motivation zur Anwendung ist jedenfalls gegeben und wird vielfach schon in die Tat umgesetzt, wenngleich sie noch mit Skepsis betrachtet wird und die Vorteile noch klarer ersichtlich gemacht werden sollten.

2.2.2 Interoperabilität und Integration

Einen Schritt weiter gehen die Überlegungen im Semantic Web, welche die Interoperabilität und Nutzbarkeit der Daten in den Fokus rücken. Mit Daten in einem strukturierten Format und semantischen Verbindungen untereinander können Daten interoperabel werden. J. D. Richards und C. S. Hardman teilen eine Vision der Zukunft, in der Forscher*innen alle integrierten Ausgrabungsdaten online durchsuchen können, um bestimmte Fundkategorien zu finden:

> „Imagine, for example, that each reference to each specific category of artefact was tagged. This would mean that future research could cross-search any number of integrated excavation archives looking for occurrences of pottery type X or brooch type Y."[129]

[127] Richards 2009, 32.
[128] Richards 2009, 33.
[129] Richards – Hardman 2008, 186.

Es reicht jedoch nicht, sich auf einheitliche Klassifizierungen zu einigen und diese zu verwenden. Für „meaningful", also bedeutungsvolles Querfragen über mehrere Datensätze ist eine Datenstrukturierung auf einem höheren Niveau essenziell. Die Ontologie CIDOC-CRM ist für dieses Level des Semantic Web geeignet.[130] J. D. Richards und C. S. Hardman sind davon überzeugt, dass die überwältigenden Vorteile der Standardisierung, wie das Querfragen über viele Datensätze, auch breitere Begeisterung in der Archäologie erzeugen werden.[131]

Interoperabilität und Integration, der Austausch und das Zusammenspiel von Daten, finden immer mehr Beachtung. Wenn Daten kombiniert verwendet werden können, kann dies Möglichkeiten für die Datennutzung eröffnen, die den Datenurheber*innen möglicherweise nicht bewusst waren. Im Folgenden wird auf Definitionen und auch Unterschiede zwischen den Begriffen Interoperabilität und Integration eingegangen und wie sie in der Archäologie genutzt werden können.

2.2.2.1 Definition von Interoperabilität

Für die Definition von Interoperabilität wird auf zwei Quellen verwiesen, welche sich für Definitionen von Standards und Hilfestellungen für Datenanbieter*innen einsetzen: Die NISO Primer Series Definition und die Definition der ALCTS. In der NISO Primer Series Definition[132] bedeutet Interoperabilität den unproblematischen Austausch zwischen Systemen: „*Interoperability, the effective exchange of content between systems.*"[133] Mehrere Systeme mit unterschiedlicher Hard- und Software, Datenstrukturen und Interfaces sollen Daten austauschen können. Dies soll so gut wie ohne Verluste der Inhalte oder der Funktionalität möglich sein.

Bei der ALCTS (Association for Library Collections and Technical Services)[134] lautet die Definition: „*Interoperability is the ability of two or more systems or components to exchange information and use the exchanged information without special effort on either system.*"[135] Interoperabilität soll also den Austausch von Informationen ohne großen Aufwand ermöglichen. Systeme sollen miteinander kommunizieren können und die Funktionalität eines anderen Systems benutzen können. Das Wort „interoperabel" deute an, ein System soll für das andere System Operationen ausführen können. In

[130] Richards – Hardman 2008, 186 f.
[131] Richards – Hardman 2008, 187 f.
[132] Siehe https://www.niso.org/what-we-do, abgerufen am 19.10.2021.
NISO steht für National Information Standards Organization. Die Organisation veröffentlicht technische Standards, die den ganzen Bereich des Datenmanagements umfassen, von der Erstellung bis zur Speicherung und Erhaltung von Daten und Metadaten. Ihr Ziel ist es, den Informationsaustausch global zu verbessern.
[133] Riley 2017, 6 f.
[134] Siehe http://www.ala.org/alcts/about, abgerufen am 19.10.2021.
Der ALCTS Verband ist der ALA, der American Library Association, zugehörig und gibt Hilfestellungen für Daten- und Informationsanbieter von Sammlungen und technischen Dienstleistungen, wie beispielsweise die Katalogisierung von Sammlungen, die Metadatenerstellung und die Sammlungsverwaltung.
[135] ALCTS Association for Library Collections and Technical Services 2000, 4.

der Computer-Technologie beispielsweise ist dies der Fall, wenn zwei heterogene Systeme gemeinsam funktionieren können und sich gegenseitigen Zugriff auf Ressourcen geben können.[136]

2.2.2.2 Ebenen der Interoperabilität

N. Beer u. a. gehen auf die unterschiedlichen Ebenen der Interoperabilität ein (siehe Abbildung 4), konkret die technische und die Informations-Ebene. Der technische Bereich soll Hardware und Software kompatibel machen. Hier werden technische Standards von Kommunikations-Protokollen und Datenformaten geklärt.[137] Auf der Informations-Ebene kommen syntaktische und semantische Kategorien hinzu. Syntaktische Interoperabilität betrifft die formale Struktur von Daten (wie etwa die XML-Sprache) und ihre Maschinenlesbarkeit. Sie bietet Schemata, um stabile Informations-Strukturen zu garantieren und ist notwendig, um einen Schritt weiterzugehen und die Bedeutung und Beziehung der Daten auf eine semantische Ebene zu bekommen. Auf der semantischen Ebene werden die Daten in einem Schema geregelt, wie beispielsweise kontrollierte Vokabularien oder Ontologien.[138] Ontologien sind hilfreich, um die syntaktischen und strukturellen Unterschiede in den unterschiedlichen Datensets zu lösen, welche für Interoperabilität notwendig sind.[139]

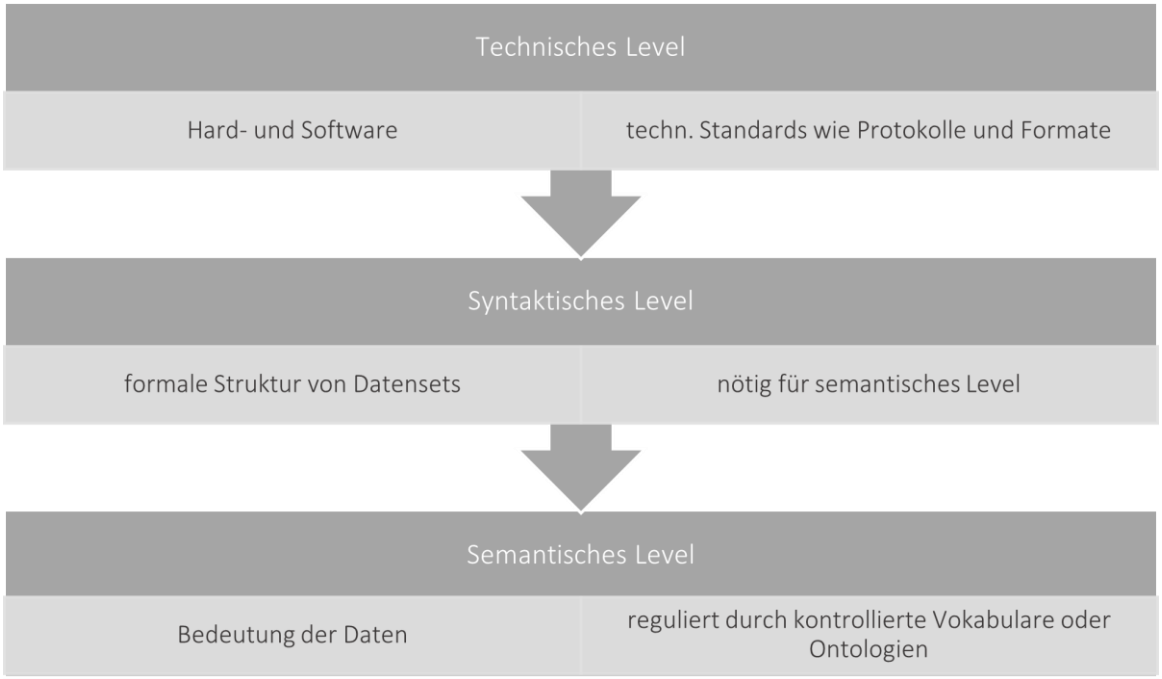


Abbildung 4 - Ebenen der Interoperabilität (Bild: M. Simon nach Inhalten von Beer u. a. 2014, 5-6).

[136] Chen u. a. 2008, 648.
[137] Beer u. a. 2014, 5.
[138] Beer u. a. 2014, 5 f.
[139] Kintigh 2006, 573.

In Bezug auf Metadaten geben B. Haslhofer und W. Klas einen Überblick über deren verschiedene Ansätze der Interoperabilität.[140] Auch Metadaten benötigen verschiedene Ebenen der Interoperabilität: Auf der technischen Ebene muss sichergestellt sein, dass Maschinen miteinander kommunizieren können, um Metadaten auszutauschen. In weiterer Folge müssen Metadaten prozessiert bzw. gelesen werden können. Das Ziel der semantischen Ebene ist, dass sowohl Mensch als auch die Maschine die Daten richtig interpretieren und ihre Bedeutung verstehen können.[141]

2.2.2.3 Interoperabilität versus Integration

Die Begriffe Interoperabilität und Integration werden oft synonym verwendet. Interoperabilität ist jedoch nicht gleich Integration, es sind unterschiedliche Konzepte.[142] Interoperabilität ist – wie oben bereits beschrieben – die Fähigkeit zweier unabhängiger, separater Systeme, Daten ohne menschliche Interaktion auszutauschen. Bei der Integration hingegen spielen verschiedene Anwendungen innerhalb eines größeren und komplexeren Systems zusammen.[143]

Datenintegration meint den Prozess, bei dem unterschiedlich aufgenommene Datensets in ein größeres vereintes Datenset transformiert werden, welches in weiterer Folge analytisch untersucht werden kann.[144] D. Chen u. a. meinen, zwei integrierte Systeme seien unweigerlich interoperabel, da sie die nötigen Übersetzer zur Kommunikation haben und einander verstehen können. Aber zwei interoperable Systeme seien nicht zwingend integriert[145]. Die Komponenten sind über ein gemeinsames Kommunikations-Netzwerk verbunden und können miteinander interagieren. Sie können Daten austauschen aber dabei nur lokal ihrer eigenen Logik folgen.[146]

Beispiele für technische Interoperabilität sind das Web, das Telefon, das Fax oder Emails. Sie können direkt miteinander kommunizieren, da sie die gleiche Sprache bzw. die gleichen Protokolle benutzen. Die Integration erfordert zusätzlich eine Übersetzung, wie beispielsweise gemeinsam benutzte Standards.

Eine Metapher[147] für die beiden Systeme wäre beispielsweise eine internationale Konferenz, auf der jeder und jede Teilnehmer*in eine andere Sprache spricht. Die Teilnehmer*innen haben drei Optionen, miteinander zu kommunizieren:

1. Jeder lernt alle Sprachen, die auf der Konferenz gesprochen werden

[140] Haslhofer – Klas 2010, 13.
[141] Haslhofer – Klas 2010, 2.
[142] Scholl u. a. 2011, 346.
[143] McManus u. a. 2012, 1164.
[144] Kintigh u. a. 2018, 31.
[145] Chen u. a. 2008, 648.
[146] Chen u. a. 2008, 648.
[147] Beispiel von https://about.caredove.com/blog/integration-vs-interoperability, abgerufen am 19.10.2021.

2. Für jede Sprache ist ein Übersetzer oder eine Übersetzerin anwesend

3. Jeder lernt eine universelle Sprache

Die ersten beiden Optionen stellen Integration dar. Punkt eins ist direkte Integration, jedes Element muss an das andere angepasst werden. Der zweite Punkt erfordert einen Zwischenschritt mithilfe des oder der Übersetzer*in. Jeder und jede neue Teilnehmer*in birgt nun neue Herausforderungen, da die neue Sprache wieder miteingebaut werden muss. Bei zu vielen Systemen wird die Integration zeitaufwendig und teuer. Die dritte Operation stellt die Interoperabilität dar. Hier ist der Vorteil, dass die Anzahl der Teilnehmer*innen nebensächlich ist, da jeder und jede diese eine Sprache lernt. Die Kommunikation läuft über eine Sprache, ohne Übersetzer*in.

2.2.2.4 Integration und Interoperabilität in der Archäologie

In den Digital Humanities ist es wichtig, Daten und Metadaten aus unterschiedlichen Kontexten direkt miteinander verwenden zu können ohne die Daten neu organisieren oder umformatieren zu müssen.[148]

K. Kintigh erklärt, welche Herausforderungen es bei archäologischer Datenintegration und der Interoperabilität geben kann. Mehrere Datensets sollen kombiniert werden können, auch wenn sie durch unterschiedliche Aufnahmebedingungen mit anderen Datenstrukturen zustande gekommen sind.[149] Wie bereits oben ausgeführt, sind archäologische Daten meist komplex, kontextuell, heterogen und inkonsistent und oft wenig kompatibel und schwer vergleichbar mit anderen archäologischen Daten.[150] Archäologische Datenbanken haben meist hochkomplexe Hierarchien und je nach Ausgangslage unterschiedliche Aufnahmekriterien der Verfasser*innen.[151]

Ein Teil der Lösung sind technische Infrastrukturen, wie Repositorien mit standardisierten Aufnahmekriterien und Metadaten. Dadurch können Daten archiviert werden, der Zugriff darauf wird möglich gemacht und heterogene Daten können integriert werden.[152] In England wurde dafür das ADS (Archaeological Data Service)[153], in den USA das tDAR (the Digital Archaeological Record)[154] und in Europa die ARIADNEplus Infrastruktur[155] etabliert.[156] Diese werden auch als Repositorien verwendet, um Datenverlust entgegenzuwirken. K. Kintigh ist der Meinung, für die

[148] Beer u. a. 2014, 5.
[149] Kintigh u. a. 2018, 30.
[150] Kintigh 2006, 570 f.
[151] Kintigh 2009, 81 f.
[152] Kintigh u. a. 2018, 30 f.
[153] Siehe https://archaeologydataservice.ac.uk/, abgerufen am 19.10.2021.
[154] Siehe https://core.tdar.org/, abgerufen am 19.10.2021.
[155] Siehe https://ariadne-infrastructure.eu/, abgerufen am 19.10.2021.
[156] Kintigh u. a. 2018, 31.

Datenintegration wären umfangreicher internationaler Einsatz und Zusammenarbeit notwendig.[157]

Der Aufwand würde sich auf längere Sicht jedenfalls lohnen. Wenn Datensets wiederverwendet und kombiniert werden können, eröffnet dies Möglichkeiten, die den Datenerfasser*innen möglicherweise gar nicht bewusst waren. Beispielsweise können neue Phänomene über große spatio-temporale und makro-regionale Bereiche schneller erkannt werden. Dies ist für neu gewonnene Daten ebenso sinnvoll wie für Altdaten, sogenannte „Legacy Data".[158] Im interdisziplinären Bereich kann die Archäologie unter anderem wertvolle Langzeitdaten über die Gesellschaft, die Bevölkerung und ihren Umgang mit der Umwelt beitragen, die hilfreich zur Erklärung von sozialen Dynamiken sind.[159]

2.2.3 Open Access, Open Data, Wiederverwendung und Langzeitarchivierung

Weitere Themen des Datenmanagements sind der offene Zugang zu Publikationen und Forschungsdaten, die Ermöglichung der Wiederverwendung bereits erstellter Daten und deren nachhaltige Langzeitarchivierung. Sie werden breit diskutiert und rücken immer näher ins Zentrum der Aufmerksamkeit.

2.2.3.1 Open Access

Einer der umstrittensten Aspekte im Datenmanagement ist der freie Zugang zu Forschungsliteratur. Aktuell zeigt sich im Licht der Covid-19 Pandemie, wie sehr geschlossene Bibliotheken das Arbeiten sowohl für Student*innen als auch für Forschende oder Mitarbeiter*innen in Sammlungen, Museen oder Grabungsfirmen beeinträchtigen. Viele Bücher sind nicht online erhältlich oder müssen gekauft werden. Die Situation gibt möglicherweise einen weiteren Impuls in Richtung freiem Zugang zu Fachliteratur.

Die häufigste verwendete Art der wissenschaftlichen Publikation sind Zeitschriftenartikel. Sie können durch die mehrmalige Erscheinung der Zeitschriften pro Jahr und ihre zu Büchern verhältnismäßige Kürze ermöglichen, dass Ergebnisse schnell veröffentlicht werden können. Die Entwicklung der letzten Jahrzehnte zeigt dabei einen Wandel von Zeitschriften in Print zu Online-Formaten.[160] Andere Formen der Publikation sind wissenschaftliche Bücher, wie Monografien, Sammelbände und Tagungsbände. Um Open Access handelt es sich, wenn digitale Fachzeitschriften bzw. Journals kostenlos für die Öffentlichkeit zugänglich sind. Dabei handelt es sich in erster

[157] Kintigh 2009, 81.
[158] Kintigh u. a. 2018, 31.
[159] Kintigh 2009, 81.
[160] Kaier – Edig 2020, 53 f.

Linie um Artikel in Zeitschriften, doch auch bei Büchern wird diese Option immer öfter gewünscht.[161] Die Transformation von Zeitschriften vom bisher üblichen Subskriptionsmodell zu Open Access ist von vielen Seiten positiv bewertet und Projekte wie das „Austrian Transition to Open Access" widmen sich im Zusammenwirken vieler österreichischer Universitäten der Analyse der Umstellung sowie Finanzierungsfragen.[162]

Zu ca. 80 % werden Forschungsergebnisse zurzeit noch in Zeitschriften von großen oder kleinen Verlagen mit dem Finanzierungsmodell der Subskription veröffentlicht. Bibliotheken, Unternehmen und Leser*innen müssen bei diesem Modell zahlen, wenn sie die Artikel lesen möchten. Doch der Wunsch nach einer Alternative wird stärker, denn die Preise steigen sehr schnell an und weitere Nebenkosten, wie beispielsweise *Submission Charges*, *Page Charges* und *Colour Charges* kommen hinzu. Diese Kosten übersteigen oft das knapp bemessene Budget von Bibliotheken und Institutionen.[163]

Open Access Zeitschriften gewinnen daher zunehmend an Bedeutung. Das wohl stärkste Argument für diese Form der Publikation ist der ermöglichte Zugang der Öffentlichkeit zu Forschungsergebnissen, welche durch öffentliche Fördergelder finanziert wurden.[164] Es wird im Allgemeinen zwischen zwei Kategorien von Open Access unterschieden: Goldenem und grünem Open Access. Beim goldenen Open Access wird der Artikel von Beginn an mit freiem Zugang veröffentlicht. Die grüne Variante erlaubt, den vorher bereits im Druck erschienenen Artikel nach einer „Embargozeit" elektronisch offen zur Verfügung zu stellen.[165]

Die Finanzierung der goldenen Open Access Modelle geschieht meist durch die Autor*innen selbst. Es gibt zwar einige Zeitschriften, die entweder vom Betreiber bzw. Herausgeber selbst, wie Universitätsverlage oder andere wissenschaftliche Einrichtungen, oder von Dritten finanziert werden, doch dies ist eher die Ausnahme. Die Autoren und Autorinnen tragen also den größten Teil der Kosten. Dies wird durch Artikelbearbeitungsgebühren, sogenannten *Article Processing Charges* (APCs), verrechnet, welche zwischen 500 bis 4.500 Euro betragen können.[166] Diesen Betrag müssen die meisten Autor*innen jedoch nicht aus eigener Hand bezahlen, da er größtenteils über Projekt- oder Institutsbudgets abgedeckt wird.[167] Dies führt zu vielen Einzelrechnungen von Autor*innen für Institutionen und zu hohen administrativen Kosten. Doch dieses Modell bringt erhebliche Nachteile für Forscher*innen, welche keiner Institution angehörig sind, am Beginn

[161] Stadler – Kaier 2020, 83.
[162] Kromp u. a. 2019, 29 f.
[163] Kaier – Edig 2020, 57–59.
[164] Geser 2014, siehe Abschnitt Open Access to academic works and scientific data.
[165] Schröter 2017, 17.
[166] Shamash 2016, 7.
[167] Kändler 2020, 183–185.

ihrer wissenschaftlichen Karriere stehen oder die Institute nicht über ausreichend budgetäre Mittel verfügen.[168,169,170] R. Salisbury ist positiv gegenüber Open Access eingestellt, warnt jedoch davor, dass sich die Publikationslandschaft verändern könnte, von einem System, in dem jeder publizieren kann und wenige die Artikel lesen können zu einem System, in dem jeder Zugang zu Publikationen hat, aber nur wenige schreiben können. Die Wissenschafter*innen, welche nicht in Open Access Zeitschriften veröffentlichen können, würden weniger Aufmerksamkeit für ihre Forschung erhalten und die Anforderungen für Forschungsprojekte nicht erfüllen.[171]

Eine zusätzliche Variante, welche zwar formal die Open Access Bedingungen von Forschungsförderern erfüllt, jedoch stark kritisiert wird, ist der Hybrid Open Access. Dabei werden einzelne Artikel in Subskriptionszeitschriften frei zugänglich gemacht durch die Bezahlung der Autor*innen. Durch die Subskription und die Open Access Artikel müssen Bibliotheken und Institutionen jedoch doppelt zahlen.[172] Doch Bibliotheken leiden finanziell ohnehin unter dem Open Access Modell, solange der Transformationsprozess noch nicht vollständig durchgeführt wurde, denn zurzeit müssen sie sowohl für die Kosten von Open Access Artikeln als auch weiterhin für die Subskriptionsgebühren aufkommen.[173]

Durch die erhöhte Nachfrage an Open Access Zeitschriften hat sich jedoch auch eine Sorte Zeitschriften herausgebildet, welche rein am Profit durch die Autoren und Autorinnen interessiert sind und einen niedrigen Qualitätsstandard aufweisen, die sogenannten *Predatory Journals*.[174] Sie führen wenig bis keine Peer Review durch, sind selten in Bibliotheksverzeichnissen oder anderen Suchmaschinen gelistet und nehmen nicht teil am akademischen Diskurs. Daher werden Autor*innen als Zielgruppe dieser Zeitschriften ausgewählt, die für ihre akademische Karriere viele Publikationen in kurzer Zeit durchführen müssen.[175] Oft wird von einer *Publish or Perish*-Mentalität gesprochen, in der es wichtiger ist, viele Artikel in möglichst kurzer Zeit zu publizieren, um die eigene Karriere voranzutreiben, als qualitativ hochwertigere, dafür weniger Artikel zu veröffentlichen.[176] Doch selbst wenn Autoren und Autorinnen versuchen, diese *Predatory Journals* zu vermeiden, ist dies oft schwierig. Zum einen gibt es bereits unzählige dieser Zeitschriften und zum anderen tarnen sie sich meist mit wohlklingenden Namen und professionell wirkenden Webseiten.[177]

[168] Nikolova 2015, 84.
[169] Canny 2015, 22.
[170] Kändler 2020, 190.
[171] Salisbury 2017, 12 f.
[172] Kaier – Edig 2020, 60 f.
[173] Kändler 2020, 192.
[174] Kurt 2018, 141 f.
[175] Vakil 2019, 92.
[176] Schilhan – Lackner 2020, 163 f.
[177] Weingart 2016, 284 f.

Um diesen Zeitschriften entgegenzutreten, wurden mehrere Kontrollinstanzen gegründet.[178] Das Directory of Open Access Journals (DOAJ)[179] führt eine Liste mit fast 17.000 Open Access Zeitschriften, welche ihre Qualitätskriterien erfüllen. Weiters wurde die Open Access Scholary Publishers Association (OASPA)[180] gegründet, welche ihre Mitglieder auf hohe Qualitätsstandards überprüft.

Abgesehen von oben genannten Kritikpunkten sind die Vorteile für Forschende, die Wissenschaft und die breite Öffentlichkeit vielfältig. F. Siegmund zeigt die relevantesten Vorteile des Open Access auf:[181]

- Die Ausdehnung der potenziellen Leser*innen wird auf die ganze Welt erweitert. Auch M. Effinger und A. Büttner beschreiben, wie durch das Online-Stellen von Artikeln diese eine sehr viel größere Reichweite und mehr Rezensionen erlangen[182].
- Durchsuchbare, öffentlich zugängliche PDF-Dateien ermöglichen Suchmaschinen über die Schlüsselwortsuche hinaus Inhalte zu finden. Texte können inhaltlich eruiert, auffindbar und zugänglich gemacht werden, während nicht frei zugängliche Texte weniger wahrgenommen werden bzw. nicht verwendet werden.
- Verlinkungen von Inhalten und Links sind einfacher möglich. Die Texte selbst oder Inhalte und Querverweise können verknüpft werden. Dies macht die Querverweise nützlicher und die Texte können in öffentlichen Online-Medien direkt hinzugefügt werden.
- In interdisziplinären Forschungen würden die Artikel mehr Aufmerksamkeit erhalten und neue und überraschende Verknüpfungen zu Tage bringen können.

In der Archäologie wenden bereits viele Fachzeitschriften eine Open Access Strategie an.[183] Im deutschsprachigen Gebiet ist dies jedoch noch eher seltener der Fall. Die DGUF[184] beispielsweise publiziert ihre Zeitschrift „Archäologische Informationen" und auch diverse Monografien offen zugänglich.[185] Sie setzt sich für das Konzept der Open Access Publikationen ein und versucht, die Interessen der Archäologie im breiten Wissenschaftssektor und Initiativen, wie beispielsweise der Initiative „Open Access 2020", miteinzubringen.[186,187]

[178] Weingart 2016, 288.
[179] Siehe unter https://doaj.org/, abgerufen am 19.10.2021.
[180] Siehe unter https://oaspa.org/, abgerufen am 19.10.2021.
[181] Siegmund 2013, 87–91.
[182] Effinger – Büttner 2015, 76.
[183] Schäfer u. a. 2015, 126–128.
[184] Deutsche Gesellschaft für Ur- und Frühgeschichte.
[185] Siegmund – Scherzler 2015, 12.
[186] Zur Initiative „Open Access 2020" siehe auch: https://oa2020.org/, abgerufen am 19.10.2021.
[187] Siegmund – Scherzler 2016.

Für Österreich wären hier zum Beispiel folgende Herausgeber zu nennen: Die Online-Zeitschrift „Historische Archäologie“[188] wird gemeinsam veröffentlicht vom Institut für Ur- und Frühgeschichte der Christian-Albrechts-Universität Kiel und des Instituts für Urgeschichte und Historische Archäologie der Universität Wien, und das Online-Magazin „Forum Archaeologiae“,[189] das unabhängig von Wien aus herausgegeben wird.[190] Der Verlag der Österreichischen Akademie der Wissenschaften bietet das Journal „Archaeologica Austriaca“[191] online mit offenem Zugang an. Andere Herangehensweisen von Herausgeber*innen sind auch, bereits gedruckte Ausgaben nach einer gewissen Frist online zur Verfügung zu stellen.[192] Der Verlag der Österreichischen Akademie der Wissenschaften verwendet diese Praxis beispielsweise bei Reihen wie „Archäologische Forschungen“[193], „Der römische Limes in Österreich“[194] und vielen weiteren. Die wissenschaftliche Zeitschrift „Beiträge zur Mittelalterarchäologie in Österreich“[195] (BMÖ) der „Österreichischen Gesellschaft für Mittelalterarchäologie“ veröffentlicht ebenfalls einige seiner Bände als Open Access, während andere zum Kauf zur Verfügung stehen.

Immer mehr europäische Forschungsförderer (darunter auch der Österreichische Wissenschaftsfonds - FWF) verlangen einen freien Zugang für Publikationen und Daten, die durch deren Fördergeld entstanden sind. Der European Research Council (ERC) ist fast seit seinem Beginn im Jahr 2007 Befürworter von Open Access für Publikationen und Forschungsdaten.[196] Die Europäische Kommission hat beschlossen, dass alle Publikationen, die durch Projekte über „Horzion 2020“ finanziert werden, sechs Monate nach Projektende offen verfügbar sein müssen. Für Sozial- und Geisteswissenschaften ist dieser Zeitraum zwölf Monate. Die Europäische Kommission setzt sich dafür ein, Lösungen zu finden, wie dieses Ziel auch mit den dazugehörigen Forschungsdaten umzusetzen ist. Sie finanziert das Projekt Open AIRE[197], das Unterstützung bieten soll für Betreiber von Repositorien, um Interoperabilität und Archivierung sicherzustellen.[198] Die prestigeträchtigeren „Hauptzeitschriften“ stehen dem Open Access Modell zwar aufgeschlossen, aber kritisch gegenüber, da sie Umsatzeinbußen befürchten.[199] Viele Autor*innen sehen Publikationen in den

[188] Siehe https://www.histarch.uni-kiel.de/, abgerufen am 19.10.2021.
[189] Siehe http://farch.net/, abgerufen am 19.10.2021.
[190] Schäfer u. a. 2015, 126.
[191] Siehe https://austriaca.at/archaeologia, abgerufen am 19.10.2021.
[192] Effinger – Büttner 2015, 74.
[193] Siehe https://verlag.oeaw.ac.at/kategorie_197.ahtml, abgerufen am 19.10.2021.
[194] Siehe https://verlag.oeaw.ac.at/kategorie_80.ahtml, abgerufen am 19.10.2021.
[195] Siehe https://www.univie.ac.at/oegm/publikationen/bmoe.html, abgerufen am 19.10.2021.
[196] Canny 2015, 21.
[197] Siehe https://www.openaire.eu/, abgerufen am 19.10.2021.
[198] Canny 2015, 23 f.
[199] Taubert 2016, 100.

Leitzeitschriften als wichtigen Karrierefaktor und haben wenig Wahlmöglichkeiten.[200] F. Siegmund und D. Scherzler machen auf Forscher*innen in weniger wirtschaftlich aufgeschlossenen Staaten aufmerksam. Diese könnten aufgrund des mangelnden Zugangs zu Fachliteratur bedeutend weniger am wissenschaftlichen Diskurs teilnehmen. Viele Verlage würden in ihrem derzeitigen Geschäftsmodell mehr Vorteile für sie sehen und würden die Diskussion über Open Access hinauszögern. Im Zuge der Initiative „Open-Access 2020" wurde eine alternative Grundlage zur Finanzierung vorgeschlagen. Staatliche Gelder sollen statt für Abonnements von Zeitschriften mit Zahlungsbarriere für die Publikationsgebühren in Open-Access Zeitschriften verwendet werden. F. Siegmund und D. Scherzler schlagen eine Umkehr der Finanzierung vor, um freien Lesezugang für die Öffentlichkeit zu erhalten.[201] Es gibt jedenfalls Handlungsbedarf, da sich auch Piratenplattformen gebildet haben, wie Sci-Hub, auf welchen Publikationen illegal verbreitet werden.[202]

Durch Open Access kann die breite Öffentlichkeit, darunter Bürger*innen, Politiker*innen und Journalist*innen, einfacher am wissenschaftlichen Diskurs teilhaben. Das derzeitige System schließt die Öffentlichkeit hinter Bezahlhürden aus. Darüber hinaus können durch diesen Einblick die Wichtigkeit und die Wahrnehmung der Archäologie besser verstanden werden.

2.2.3.2 Open Data und Wiederverwendung von Daten

Der Begriff Open Data wird für den freien Zugang zu den wissenschaftlichen Daten, die u. a. für Publikationen genutzt werden, verwendet und ist im daher auch im Zusammenhang mit Open Access zu sehen.[203] Zu diesen wissenschaftlichen Daten zählen Datenbanken, wie Fund- oder Befunddatenbanken, Grabungsdokumentationen, digitale Bilder und Fotos, GIS-Daten und andere Resultate, wie naturwissenschaftliche Untersuchungsergebnisse. Würden diese Daten offen zur Verfügung stehen, könnten sie von anderen Forscher*innen wiederverwendet oder eingesehen werden. Open Data sind dabei nicht unbedingt als FAIRe Daten zu verstehen (siehe Kapitel 2.1.2). Sie können diesen Status jedoch zusätzlich erreichen, wenn sie maschinenlesbar, gut dokumentiert, mit einer offenen Lizenz versehen werden und freie Standards verwenden.[204]

Forschungsdaten zu veröffentlichen, wird von einigen Wissenschafter*innen bereits als wesentlicher Teil der wissenschaftlichen Praxis angesehen und es existieren einige Open Access Publikationsplattformen, die mit Open Data verbunden sind.[205] Die Ergebnisse von Forschungsprojekten

[200] Siehe beispielsweise den Blogbeitrag zu Open Access und dessen Aktzeptanzprobleme: Scheloske 2012.
[201] Siegmund – Scherzler 2015, 13.
[202] Salisbury 2017, 12 f.
[203] Gewin 2016, 117 f.
[204] Sánchez Solís u. a. 2020, 106 f.
[205] Schäfer u. a. 2015, 128 f.

beinhalten immer öfter nicht nur Publikationen, sondern auch die entstandenen Forschungsdaten.[206] Einige europäische Staaten, insbesondere England, haben die Archivierung, Open Data und die Wiederverwendung digitaler Daten bereits in den wissenschaftlichen Alltag integriert. In anderen Ländern Europas wird dies mit Forschungsdaten und auch mit der „Grauen Literatur" seit der Mitte der 1990er Jahre umgesetzt.[207] In Österreich geben in etwa 11 Prozent der Forschenden Open Data heraus.[208] Doch immer mehr Forschungsförderer, darunter auch der FWF Wissenschaftsfonds, verlangen von geförderten Projekten, ihre Forschungsdaten mit entsprechenden Lizenzen und unter Beachtung ethischer Aspekte öffentlich zur Verfügung zu stellen.[209,210]

Es gibt mehrere Möglichkeiten, Open Data zu veröffentlichen. Repositorien bieten Wege, Daten zu archivieren, zu publizieren und zu finden.[211] Sogenannte Data Papers in Data Journals sind gedacht, um Artikel über veröffentlichte Daten zu publizieren. Darin wird beschrieben, warum und wie die Daten erhoben wurden. Sie geben die Möglichkeit, Links zu den Rohdaten in Repositorien anzugeben.[212,213] Eine weitere Option, Forschungsdaten zu publizieren, stellen die Enhanced Papers dar. Dies sind Publikationen, die mit weiteren Zugaben, wie Daten, Annotationen, Zusatzmaterial, Bilder, Videos u. a., erweitert werden können. Sie stehen jedoch oft vor Herausforderungen bezüglich der langfristigen Verfügbarkeit und Archivierung.[214]

Bei archäologischen Daten würde sich die Möglichkeit ergeben, viel größere Datenmengen wie beispielsweise Funddatenbanken mit Darstellungen zu veröffentlichen, die zu groß sind für einzelne Artikel.[215] Die meisten Projekte im deutschsprachigen Raum präsentieren die Resultate ihrer Forschung über analoge oder digitale Publikationen sowie durch möglichst öffentlichkeitswirksame Dissemination. Jedoch oft, ohne die primären Daten zu veröffentlichen. Selbst bei Open Access Plattformen, die Publikationen zur Verfügung stellen, fehlt meist überhaupt die Möglichkeit, diese Forschungsdaten hinzuzufügen[216]. F. Schäfer u. a. führen dies erstens darauf zurück, die Dienste würden sich am Beispiel der gedruckten Veröffentlichungen orientieren. Hier ist die

[206] Sánchez Solís u. a. 2020, 101.
[207] Richards 2015, 63.
[208] Bauer u. a. 2015, 42, im Jahr 2015.
[209] Siehe unter https://www.fwf.ac.at/de/forschungsfoerderung/open-access-policy/open-access-fuer-forschungsdaten, abgerufen am 19.10.2021.
[210] Gewin 2016, 118.
[211] Eine vom FWF Wissenschaftsfonds empfohlene Auflistung an geeigneten Repositorien findet sich unter https://www.re3data.org/, abgerufen am 19.09.2021.
[212] Eine Liste von Data Journals ist unter https://www.forschungsdaten.org/index.php/Data_Journals zu finden, abgerufen am 19.09.2021.
[213] Sánchez Solís u. a. 2020, 113.
[214] Kaier – Edig 2020, 76 f.
[215] Jacobs – Holland 2007, 198 f.
[216] Bei den Archäologischen Informationen der DGUF ist dies möglich, siehe https://dguf.de/publikationen/archaeologische-informationen, abgerufen am 19.10.2021.

Zugabe von Daten nicht praktikabel. Zweitens weisen sie auf die immer größer werdenden technischen Herausforderungen für die Archivierung hin. Die Plattformen müssten dafür öfter technische Partner wie Rechenzentren anheuern und Kooperationen mit Bibliotheken und Archiven eingehen.[217]

2.2.3.2.1 Vorteile von Open Data

Transparenz, Nachvollziehbarkeit und Wiederverwendung bieten ungeahnte neue Wege für archäologische Daten. Die Transparenz der wissenschaftlichen Publikationen wird durch die nachfolgende Veröffentlichung der zugehörigen Daten stark erhöht. Zitierfähige und -pflichtige Daten wiederzuverwenden, spielt eine immer größere Rolle.[218] Durch den Zugang zu den Primärdaten, den unveränderten Rohdaten, mit denen die Forscher*innen arbeiten und zu ihren Ergebnissen gelangen, können die Daten wiederverwendet werden und neue Fragestellungen, Methoden oder Erkenntnisse gewonnen werden. Außerdem wird auch die wissenschaftliche Praxis verbessert, Thesen und Interpretationen nachprüfen zu können. Dies macht die Ergebnisse nachvollziehbarer.[219] Die Primärdaten sieht auch K. Kintigh als wichtig an, um unabhängig von Verfasser*innen und deren Zusammenfassungen, Selektionen und Schlussfolgerungen auch in der Zukunft neue Fragen beantworten zu können und die Daten in wissenschaftlicher Weise verwerten zu können. Seiner Meinung nach werden noch zu wenige Primärdaten zugänglich gemacht.[220]

Doch nicht nur für aktuelle und zukünftige Forschungsfragen sind offene Forschungsdaten von Vorteil. Auch die Wissenschafter*innen, welche sie publizieren, können profitieren. Sie erreichen mehr Sichtbarkeit und Anerkennung in ihrem Fach und haben neue Möglichkeiten für Kooperationen und Kontakte. Je mehr Forscher*innen ihre Daten als Open Data veröffentlichen, desto früher werden die Daten als zusätzliche Publikation und Beitrag zur Forschung im Fachkreis angesehen.[221] Denn so hätten Wissenschafter*innen gleichzeitig mehr Publikationen zu verzeichnen, zum einen die Daten selbst und zum anderen die Publikation über ihre Ergebnisse. Open Data tragen so auch zur Transparenz und Offenheit der Publizierenden bei. Vor allem für angehende Forscher*innen könnte diese Vorgehensweise Vorteile bringen, da sie einerseits Zugang zu mehr Daten für ihre Forschung bekommen, ohne auf Kontakte im Umfeld angewiesen zu sein, und andererseits ihren Ruf als Beitragende*r schneller festigen können, wenn sie Daten frei zur Verfügung stellen.[222]

[217] Schäfer u. a. 2015, 128.
[218] Kansa – Whitcher Kansa 2013, 88.
[219] Jacobs – Holland 2007, 198 f.
[220] Kintigh 2009, 81.
[221] Bauer u. a. 2015, 50.
[222] Gewin 2016, 118 f.

2.2.3.2.2 Hindernisse und Bedenken auf dem Weg zu Open Data

Wissenschafter*innen sind noch nicht vollends überzeugt von den Vorteilen, die Open Data ihnen bringen könnten, wenn sie eigene oder fremde Forschungsdaten wiederverwenden möchten. Die Motivation im archäologischen Fachkreis hält sich ebenso in Grenzen. Die Meinung, der Prozess wäre zu aufwendig und übersteige die Vorteile, neben anderen Bedenken, überwiegt.[223]

Die Gründe für die zurückhaltende Entwicklung im deutschsprachigen Raum sehen F. Schäfer u. a. in der derzeitigen Praxis und der Grundanschauung einerseits und andererseits in den zusätzlichen Anforderungen. Um offene und nachhaltig gesicherte Daten zu erhalten, ist ein erheblicher Aufwand zu bewältigen. Die Langzeitarchivierung, die Wiederverwendbarkeit als auch die Nutzbarkeit der Daten soll sichergestellt werden. Dafür müssen die digitalen Daten teilweise aufwendig vorbereitet werden.[224]

Doch dies sind nicht die einzigen Bedenken, die bestehen. Forscher*innen wollen mit ihren Daten eventuell noch weitere Publikationen tätigen oder sind sich möglicherweise nicht sicher, ob die Daten der öffentlichen Kritik standhalten könnten. Zurzeit ist es eher üblich, sich im persönlichen Austausch mit Kolleg*innen gezielt Daten zu besorgen, die für die eigene Forschung nützlich sind.[225] Hierbei kann es jedoch selbst bei Zustimmung der Kolleg*innen zu Problemen kommen. Oft stellen nicht standardisierte Dateiformate und nicht klar definierte Datenbankinhalte Hindernisse dar, wenn die Daten wiederverwendet werden sollen. Dies kann zu Unsicherheiten, Missinterpretation der Daten oder Fehlern bis hin zur Unbrauchbarkeit der Daten führen.[226]

Auch Fragen zum Copyright können das Verwenden der Daten behindern. Diese Fragen sind jedoch von essenziellem Wert für alle veröffentlichten Forschungsdaten, um ohne Zweifel von anderen Forscher*innen wiederverwendet werden zu können. Durch die Verwendung von Creative Commons Lizenzen[227] kann beispielsweise festgelegt werden, wie die Daten weiterverwendet und auch bearbeitet werden können.[228] Das Digital Curation Centre veröffentlichte Richtlinien für die Lizenzierung von Forschungsdaten[229] als Hilfsmittel für Datenverantwortliche. Doch oft bleiben trotz Richtlinien rechtliche Unsicherheiten.[230] Daher wünschen sich Forscher*innen Ansprechpersonen an Institutionen, welche ihnen bei Fragen zur Lizenzierung ihrer Daten helfen können. Weiters sollen finanzielle und personelle Ressourcen geschaffen werden, um bei organi-

[223] Siegmund – Scherzler 2015, 14.
[224] Schäfer u. a. 2015, 130.
[225] Aspöck 2020, 13.
[226] Kansa – Whitcher Kansa 2013, 89 f.
[227] Siehe unter https://creativecommons.org/, abgerufen am 19.10.2021.
[228] Bauer u. a. 2015, 34 f.
[229] Ball 2014.
[230] Morgan – Eve 2012, 530 f.

satorischen, rechtlichen oder ethischen Unsicherheiten und Unklarheiten zum Datenschutz helfen zu können.[231] Wenn Forschungsdaten sensible, ethische oder personenbezogene Daten beinhalten, dürfen diese nicht veröffentlicht werden.[232,233] Ethische Fragen betreffen Daten, die die Menschenwürde oder Persönlichkeitsrechte betreffen. In der Archäologie könnten dies u. a. Daten von neuzeitlichen Ausgrabungen, wie beispielweise in Konzentrationslagern oder Friedhöfen, sein. Weitere Bedenken betreffen Prospektionsdaten, welche von Raubgräber*innen missbraucht werden könnten.

Diese Unsicherheiten sind jedoch nicht alle Bedenken, welche Forscher*innen bei der Verwendung von Open Data haben. Der zu leistende Mehraufwand, um rechtliche Unsicherheiten zu klären, ist ein großer Faktor. Denn Wissenschafter*innen verfügen meist nicht über genügend Zeit, sich mit all diesen Fragen beschäftigen zu können.[234] Hinzu kommen Bedenken über die weitere Verwendung ihrer Daten. Die Daten könnten missbraucht, gefälscht oder kommerzialisiert werden. Daher ist es sinnvoll, Systeme in den Institutionen einzuführen, welche den Aufwand auf ein Minimum reduzieren können und standardisierte Leitlinien anzubieten, an die sich Wissenschafter*innen halten können.[235]

Ein letzter Kritikpunkt von Forscher*innen stellt ein Gegenargument zu einem vermeintlich positiven Aspekt zu Open Data dar. Er betrifft die neuen Möglichkeiten zu Kooperationen. Denn beim Zusammenarbeiten muss immer im Vorfeld geklärt werden, ob der mögliche Kooperationspartner mit dem Konzept von Open Data einverstanden ist. Dies könnte Kooperationen auch verhindern. Vor Projektbeginn sollte jedenfalls geklärt werden, wer im Besitz der Daten ist und wann und mit welcher Lizenz bzw. Zugriffsrecht sie veröffentlicht werden.[236]

Im Jahr 2013 wurde eine Analyse mittels einer anonymen Online-Befragung durchgeführt, um die Interessen der Altertumswissenschaften im Hinblick auf ein Forschungsdatenzentrum und auch die generelle Bereitwilligkeit für Open Data zu erkunden. Dabei zeigte sich, dass kein generelles System für die Nutzung von Daten für Dritte existiert. Der Datenzugriff ist im Allgemeinen nicht leicht möglich. Auf die Frage, welche Komponenten für die Langzeitarchivierung und das Anbieten von Forschungsdaten wichtig seien, wurden als wichtigste Faktoren „Nachvollziehbarkeit der

[231] Bauer u. a. 2015, 34-35 und 50.
[232] Zozus 2017, 291.
[233] Die Plattform forschungsdaten.info gibt zu diesen Themen einen Überblick und weiterführende Links. Siehe unter https://www.forschungsdaten.info/themen/ethik-und-gute-wissenschaftliche-praxis/ethische-aspekte-im-fdm/, abgerufen am 19.10.2021.
[234] Kansa – Whitcher Kansa 2013, 88–90.
[235] Bauer u. a. 2015, 51 f.
[236] Gewin 2016, 119.

Ergebnisse“, „langfristiger Zugriff“, „Verlust von Fachwissen vorbeugen“ und „Wiederverwendbarkeit der Daten“ genannt. Die Punkte, die erfüllt werden müssten, damit Forschende ihre eigenen Daten bei IANUS archivieren würden, sind „Schutzmechanismus für sensible Daten“, „Eindeutige Zitierbarkeit“, „Differenzierte Zugriffsrechte“ und „Klare Lizenzvereinbarungen“. Die Bereitschaft für Open Data ist definitiv gegeben, es mangelt allerdings in der Praxis, vor allem an der meist fehlenden Infrastruktur und Aufklärung bzw. Beratung von Forscher*innen hinsichtlich von Lizenzen und Zitierung von Daten.[237]

Die für wichtig gehaltenen Aspekte sind interessanterweise gleichzeitig die oft genannten positiven Argumente für Open Data. Besonders oft und kontrovers diskutiert ist im Bereich der Archäologie der offene Zugang zu sensiblen und zu schützenden Daten wie beispielsweise Geoinformationen zu Fundstellen oder personenbezogene Daten.[238]

2.2.3.2.3 Open Data für archäologische Daten

Archäologische Daten bringen allerdings besondere Herausforderungen für das Datenmanagement und die Aufarbeitung mit, die auch für die Diskussion um Open Data relevant sind. Die Daten setzen sich aus verschiedenen Datentypen und -formaten zusammen. Es gibt sogenannte „born digital“ Daten, die bereits digital erfasst werden. Darunter fallen beispielsweise Tabellen, Datenbanken, Texte, Digitalfotografien, Vektorgrafiken, Vermessungsdaten und 3D-Modelle. Einige Daten werden zuerst in analoger Form erschaffen und erst im Nachhinein digitalisiert. Beispiele hierfür sind jegliche Art von Befund- und Fundzeichnungen, welche digitalisiert werden oder in einem Programm in eine Vektorgrafik übertragen werden, Pläne und handschriftliche Notizen.[239] Dieser Punkt betrifft vor allem Daten von älteren Forschungen und Ausgrabungen. Analoge Archive, welche digitalisiert und online frei zur Verfügung gestellt werden, eröffnen Forscher*innen weltweit Möglichkeiten mit den Daten zu arbeiten, ohne die Archive in möglicherweise weit entfernten Ländern zu besuchen.[240]

Die Methoden der Datengewinnung können auch zerstörerisch (mikro- oder makro-invasiv) sein, wie beispielsweise Ausgrabungen und Bohrungen. Die Daten, die von Ausgrabungen stammen, sind nicht ersetzbar, denn die ausgegrabenen Komplexe werden dabei abgetragen.[241] Die Funde werden also aus ihrem Auffindungskontext entfernt. Daher ist es für Forscher*innen selbstverständlich, eine gute Dokumentation und die zugehörige Datenerfassung anzufertigen. Diese Primärdaten sind das „objektivste“ Abbild der ausgegrabenen Fundstellen ohne die miteingeflossene

[237] Heinrich u. a. 2014, 3–22.
[238] Schäfer u. a. 2015, 132.
[239] Hagmann 2018, 53–55.
[240] Aspöck 2020, 14.
[241] Kansa – Whitcher Kansa 2013, 88.

Interpretation in den Publikationen. Sie sind Startpunkt für jede weitere Auseinandersetzung mit den Befunden und dem Fundmaterial. Dritte bzw. andere Archäolog*innen benötigen die Rohdaten ebenso wie die Ergebnisse und Interpretationen.[242] Durch diese Rohdaten können neue Details entdeckt werden, die für aktuelle Forschungsfragen vielleicht nicht relevant scheinen und daher nicht publiziert wurden, doch für zukünftige Fragen enorm wichtig wären.[243] Um die Daten möglichst gut wiederverwenden zu können, ist es wichtig, auf die Angabe von Forschungsmethoden und dem Prozess der Dokumentation zu achten. Dies erhöht die Möglichkeit, den Wert der Daten für die eigene Forschung einzuschätzen, die Daten zu verstehen und sie als transparent einstufen zu können.[244]

Ein wichtiger Punkt bei Open Data für die Archäologie ist, dass bei archäologischen Daten in den meisten Fällen ein Ortsbezug vorhanden ist. Laut der INSPIRE Richtlinien der Europäischen Kommission[245] sind Mitwirkende verpflichtet, die Geodaten interoperabel auszuweisen. Dies könnte bei der Offenlegung der genauen Koordinaten von archäologischen Fundstellen Bedenken geben aus Angst vor Raubgräber*innen. Eine Diskussion in Fachkreisen ist hier noch im Gange. Man sollte jedenfalls bei der Veröffentlichung der Daten berücksichtigen, ob die Fundstelle nach bestem Wissen vollständig ausgegraben wurde oder ob Raubgräber*innen dort noch erheblichen Schaden bereiten könnten.

Auch Grabungsberichte und die Dokumentation der Rettungsgrabungen sollten in die Diskussion miteingebunden sein. Hier liegt jahrzehntelange Arbeit in Archiven, oftmals ohne Aussicht auf wissenschaftliche Aufarbeitung. Es könnten hier neue Perspektiven sowie die Voraussetzungen für Big Data in der Archäologie, also die Möglichkeit zur Verarbeitung und Auswertung riesiger Datenmengen, entstehen.[246] Dafür müsste das Bundesdenkmalamt die Richtlinien für archäologische Maßnahmen[247] eventuell anpassen und mit einem online abrufbaren Repositorium zusammenarbeiten.

Eine weitere Kategorie von archäologischen Daten eignet sich langfristig für Open Data, Daten von bereits abgeschlossenen Projekten und Ausgrabungen. Wenn Altdaten aufgearbeitet und Archive digitalisiert werden, eröffnen sich neue Möglichkeiten, diese zu nutzen.[248] Ideen und Projekte können entstehen, wie die Verwendung der Altdaten mit GIS-Software für spatio-temporale

[242] Schäfer u. a. 2015, 125.
[243] Jacobs – Holland 2007, 199.
[244] Faniel u. a. 2013, 295.
[245] Siehe https://inspire.ec.europa.eu/, abgerufen am 19.10.2021.
[246] Siegmund – Scherzler 2015, 14.
[247] Bundesdenkmalamt 2018.
[248] Aspöck 2020, 13.

Analysen oder Fundanalysen. Durch die Digitalisierung können Informationen von fragilem Archivmaterial erhalten bleiben. Sie bietet aber auch neue Möglichkeiten, auf vorher schwer zugängliche Archive online zuzugreifen. Dies ermöglicht Forscher*innen weltweit mit den Daten zu arbeiten.[249]

2.2.3.2.4 Open Data in Projekten und Institutionen

Einige internationale Projekte und Institutionen verwenden bereits Open Data in der Archäologie. Beispiele hierfür sind „Internet Archaeology“[250] und der „Archaeology Data Service“[251] aus Großbritannien, „Data Archiving and Networked Services“[252] aus den Niederlanden, „Open Context“[253] aus den USA, das „ARIADNE“ Projekt[254,255] und „PANGAEA“[256] aus Deutschland. Die dafür notwendige Infrastruktur wurde in den vergangenen Jahren aufgebaut.

In Deutschland hat sich das IANUS Forschungsdatenzentrum[257] etabliert. Das Ziel ist kostenlose und leicht zugängliche Dienstleistungen für die Kuratierung und Archivierung von digitalen Primärdaten anzubieten. Forschungsprojekte und Institutionen sollen darüber hinaus im Datenmanagement beraten werden. Die Leistungen wurden speziell für Forscher*innen im Bereich der Archäologie und antiken Kulturen entwickelt. IANUS unterstreicht, wie wichtig der „Datenlebenszyklus“ (siehe Abbildung 5) ist. Daten werden nach ihrer Erstellung analysiert und archiviert, doch ihr Lebenszyklus sollte durch einen ermöglichten Zugang verlängert werden, damit sie weiterverwendet werden können. IANUS möchte eine allgemeine positive Resonanz für

[249] Aspöck 2020, 11–17.
[250] Siehe https://intarch.ac.uk/, abgerufen am 19.10.2021.
[251] Siehe https://archaeologydataservice.ac.uk/, abgerufen am 19.10.2021.
[252] Siehe https://dans.knaw.nl/en, abgerufen am 19.10.2021.
[253] Siehe https://opencontext.org/, abgerufen am 19.10.2021.
[254] Meghini u. a. 2017.
[255] Siehe die Ressourcendatenbank unter http://ariadne2.isti.cnr.it/, abgerufen am 19.10.2021.
[256] Siehe https://www.pangaea.de/, abgerufen am 19.10.2021.
[257] Siehe https://www.ianus-fdz.de/, abgerufen am 19.10.2021.

das Wiederverwenden von Forschungsdaten erzeugen. Dabei ist es wichtig, über die komplette Projektlaufzeit ein Auge auf das Datenmanagement zu haben. Von der Antragstellung an sollte klar sein, welche Kosten von der Datenerstellung bis zur Veröffentlichung und Archivierung einzurechnen sind. Hinzu kommt das Festlegen der Lizenzen, die Dokumentation und die Publikation und Archivierung.[258]

Abbildung 5 - Der Lebenszyklus von Forschungsdaten (Quelle: IANUS, https://ianus-fdz.de/lebenszyklus, abgerufen am 19.10.2021).

2.2.3.3 Repositorien und Langzeitarchivierung

Die Archäologie besitzt die Eigenschaft, besonders engagiert neue digitale Technologien aufzugreifen. Bereits in den 1980er Jahren konnte man dies beobachten und seit den 2000er Jahren ist deshalb ein enormer Anstieg an digitalen Daten zu verzeichnen. Der Großteil der Dokumentation archäologischer Forschungen wird in digitaler Form aufgenommen, oft ohne breit anerkannte Richtlinien und Abläufe für den weiteren Umgang mit den Daten.[259] Doch schon jetzt zeichnen sich Probleme ab im Umgang mit Daten, welche über Jahrzehnte gesammelt wurden, den sogenannten „Legacy data". Die Erhaltung und weitere Verwendung sind oftmals bereits gefährdet.

[258] Sánchez Solís u. a. 2020, 108–110.
[259] Richards u. a. 2021.

Einige veraltete Datenträger (wie zum Beispiel Floppy Disks) und Dateiformate können teilweise nur mehr mit erheblichem Mehraufwand gelesen werden. Dateiformate, welche veraltet sind oder mit proprietärer Software produziert wurden, laufen Gefahr, in naher Zukunft nicht mehr lesbar zu sein.[260] Viele Experten im Umgang mit archäologischen Daten warnen vor einem herannahenden digitalen dunklen Zeitalter, dem „Digital Dark Age", in dem viele digitale Daten gefährdet sind, aufgrund fehlender Archivierung verloren zu gehen.[261] D. Hagmann weist auf die Herausforderungen hin, die es zu meistern gilt, um dies verhindern zu können.[262] Der Begriff „Digital Dark Age" wird oft für das Phänomen des Datenverlustes verwendet. Dies passiert unter anderem durch die schnelle Adaption von neuen Technologien und die vergleichsweise langsame Erweiterung der Infrastruktur, um die entstehenden digitalen Daten zu sichern.[263] Es gilt, die inhomogenen Daten, die durch unterschiedliche Methoden und teilweise fehlende Dokumentations- und Terminologiestandards entstanden sind und laufend weiter entstehen, langfristig und nachhaltig zu sichern.[264]

Die Archäologie sollte besonders an der Archivierung ihrer Forschungsdaten interessiert sein, da, wie bereits weiter oben erwähnt, die Grundlage der Forschung bei archäologischen Untersuchungen wie Ausgrabungen zerstört wird.[265] Es ist daher besonders wichtig, die Daten vor möglichem Verlust zu schützen. „Legacy Data", die Daten aus bereits abgeschlossenen Projekten und Ausgrabungen, die nicht mehr aktiv verwendet werden, müssen in Repositorien archiviert und gesichert werden. Durch jahrzehntelange Ausgrabungstätigkeit hindurch wurden viele Archive mit analoger Dokumentation gefüllt. Doch auch digitale Daten gingen daraus hervor und müssen gesichert werden. Ebenfalls gefährdet sind die Primärdaten der Forschenden, die sich meist in Datenbanken befinden, und die sogenannte „Graue Literatur". Sie betrifft alle Daten, welche nicht in der Publikation vorhanden sind und nicht veröffentlicht wurden. Meist beinhaltet sie unstrukturierten Text, Aufzeichnungen oder andere Dokumentationsformen und diverse Grafiken, die zur Dokumentation verwendet werden, wie Karten, Fotos, GIS-Dateien etc.[266] An dieser Stelle sei auf das Projekt SEADDA (Saving European Archaeology from the Digital Dark Age)[267] hingewiesen, welches sich einsetzt, das „Digital Dark Age" in Europa zu verhindern.

[260] Aspöck 2020, 11–16.
[261] U. a. Richards u. a. 2021; Aspöck 2020, 11–16; Kansa 2016, 466-168; Moore – Richards 2015, 40.
[262] Hagmann 2018, 57.
[263] Jeffrey 2012, 553.
[264] Hagmann 2018, 57.
[265] Geser 2019, 1.
[266] Kintigh 2009, 81 f.
[267] Siehe https://www.seadda.eu/, abgerufen am 19.10.2021.

Die Langzeitarchivierung birgt jedoch große Herausforderungen. Digitale Daten sind in vielerlei Hinsicht schwieriger zu sichern, als es noch mit der analogen Dokumentation der Fall war.[268] Die Daten sind nicht von Anfang an archivierbar, sie müssen vorbereitet und bearbeitet werden.[269] Wenn die Daten einmal archiviert sind, müssen sie jedoch laufend weiter kuriert werden. Es muss darauf geachtet werden, ob die Dateiformate aktuell sind und ob die Server gewartet werden.[270] Doch neben diesen Herausforderungen ergeben sich viele neue Möglichkeiten, mithilfe von Repositorien Daten langfristig für weitere Forschungen zur Verfügung zu stellen und wiederzuverwenden.[271]

Die archäologische Forschung in Österreich setzt zurzeit noch keinen besonderen Fokus auf diese Thematik. Nur ein kleiner Teil der Projekte sichert seine Daten in einem geeigneten Repositorium[272] und das Bewusstsein für die Herausforderungen des Datenmanagements und der Archivierung setzt nur langsam ein.[273]

Die Sicherung digitaler Daten wird nicht vom Denkmalschutzgesetz vorgeschrieben[274], jedoch gibt es Vorgaben des Bundesdenkmalamtes, wie die Dokumentation archäologischer Maßnahmen abgegeben werden muss. Die digitale Dokumentation muss in einer vorgegebenen Ordnerstruktur mit ausgewählten Dateiformaten auf einem Speichermedium an das Bundesdenkmalamt übergeben werden, welches die Daten anschließend sichert.[275] Für Forschungsprojekte wird der Druck für Archivierungspläne jedoch aufgrund der Forderungen von Forschungsförderern[276], wie beispielsweise dem FWF für Open Access, Open Data und die Langzeitarchivierung immer größer. Dabei gestaltet sich die Suche nach einem geeigneten Repositorium in Österreich nicht einfach und wird auf Konferenzen wie beispielsweise der CHNT (Conference on Cultural Heritage and New Technologies)[277] diskutiert. Eine Studie aus dem Jahr 2015 des e-Infrastructures Austria-Projekts, welches sich für die Erstellung von universitären und anderen institutionellen Repositorien einsetzt, zeigt, dass fast alle Forschenden selbst für die Speicherung ihrer Daten verantwortlich sind und nur in etwa ein Drittel ihre Daten in Repositorien archiviert. Es besteht ein hoher

[268] Richards 2002, 343.
[269] Allison 2008, Kapitel 3.
[270] Richards 2002, 343.
[271] Aspöck 2020, 11.
[272] Trognitz 2021, Kapitel 2.3.
[273] Aspöck 2019, 51.
[274] DMSG.
[275] Bundesdenkmalamt 2018.
[276] Tonto u. a. 2015, 3 f.
[277] Im Zuge der CHNT 2016 in Wien wurde beispielsweise ein Runder Tisch zum Thema „Long-term preservation and access: Where is an archive for my data?“ von E. Aspöck und G. Geser veranstaltet. Darin wurde verdeutlicht, wie wenig Möglichkeiten Archäologen und Archäologinnen in Österreich haben, ein geeignetes Repositorium für ihre Daten zu finden (siehe unter https://www.chnt.at/long-term-preservation-and-access-where-is-an-archive-for-my-data/, abgerufen am 19.10.2021).

Bedarf an technischer Infrastruktur in Form von Archiven und zusätzlich geschultem Personal, welches unterstützend Hilfestellungen anbieten kann.[278]

2.2.3.3.1 Anforderungen an Repositorien

Die Anforderungen an ein geeignetes Repositorium sind vielfältig. Die Hauptaufgaben sind das langfristige Archivieren und Sammeln von Daten, den Zugang, das Auffinden und die Wiederverwendung der Daten zu gewährleisten und die Dissemination der Daten zu ermöglichen.[279] Es muss unterschiedliche archivkonforme Datenformate fassen und diese migrieren, um aktuelle Formate zu behalten, die Daten innerhalb miteinander in Beziehung setzen und langfristig zur Verfügung stellen können.[280] Weiters sollen Forscher*innen den Zugang zu den Daten frei wählen und ändern können und die Daten sollen mit einer eigenen Signatur bzw. persistenten Identifikatoren versehen werden, um zitierbar zu sein.[281] D. Hagmann sieht das Repositorium als eine weitere Möglichkeit für Publikationen, um die Daten auch weiter zugänglich halten zu können. Ein weiterer Vorteil einer solchen Publikationsweise besteht darin, dass digitale Daten ohne die drucktechnischen und kostenbasierten Einschränkungen, die bei gedruckten Publikationen meist auftreten, auskommen können.[282] Die Identifikatoren, definierte Nutzungsbedingungen und frei einsehbaren, im besten Fall maschinenlesbaren Metadaten erhöhen die Chance der Daten, langfristig zitierbar und wiederverwendbar zu bleiben.[283]

Die Herausforderungen, die Repositorien bewältigen müssen, sind oft organisatorischer und finanzieller Natur. Eine tatsächliche Langzeitarchivierung über einen unbegrenzten Zeitraum ist in den meisten Fällen jedoch noch nicht gewährleistet und bedarf weiterer Abklärungen.[284,285] Wenn die Infrastruktur aufgebaut ist und die technischen Gegebenheiten eingeräumt sind, hängt der reibungslose Ablauf der Archivierung davon ab, ob geschultes Personal verfügbar und leistbar ist.[286] Das Personal eines Repositoriums muss breitgefächerte Qualifikationen aufweisen. Kenntnisse über Datenmanagement, den Forschungsdatenzyklus, Lizenzen, Open Data, ethische und rechtliche Fragen und technisches Wissen über Langzeitarchivierung, Metadatenstandards und andere Richtlinien sind Voraussetzung. Darüber hinaus müssen sie über die Forderungen von Fördergebern Bescheid wissen und den Forschern und Forscherinnen Unterstützung bieten können.[287]

[278] Bauer u. a. 2015, 64–67.
[279] Stigler – Steiner 2018, 213 f.
[280] Trognitz – Ďurčo 2018, 218 f.
[281] Blumesberger 2020, 503–505.
[282] Hagmann 2018, 58.
[283] Sánchez Solís u. a. 2020, 111.
[284] Torggler – Andrae 2018, 119.
[285] Knoche 2017, 120 f.
[286] Stigler – Steiner 2018, 213 f.
[287] Blumesberger 2018, 152.

Fördergeber fordern meist vertrauenswürdige Repositorien für die Archivierung von Forschungsdaten. Der FWF beispielsweise verlangt, dass ein verwendetes Repositorium in re3data[288], der Webseite für die Registrierung von Datenrepositorien (Registry of Research Data Repositories), gelistet ist.[289] Repositorien, welche eine Zertifizierung erlangt haben, gelten als besonders geeignet.[290] Zertifizierungen für Repositorien sind beispielsweise CoreTrustSeal[291] oder Nestor Seal[292].

2.2.3.3.2 Auswahl an Repositorien

Das Registry of Research Data Repositories (re3data) ist hilfreich bei der Suche nach einem passenden Archiv[293]. Eine aktuelle, weltweite Aufstellung von Open Access Repositorien ist das Directory of Open Access Repositories[294] (OpenDOAR). Da es die umfangreichste Datenbank an Repositorien ist, wird OpenDOAR oft für Vergleiche und die Entwicklung von Archiven herangezogen.[295]

Es existieren mehrere Arten von Repositorien. Institutionelle Repositorien werden von Universitäten oder anderen Instituten betrieben. Sie unterscheiden sich untereinander in ihrer Kapazität, den akzeptierten Dateiformaten und dem Archivieren von Software und Datenbanken. Daneben gibt es Fachrepositorien, welche Daten von spezifischen Forschungsgebieten aufnehmen. Sie können den Vorteil bieten, die Daten innerhalb des eigenen Faches besser auffindbar zu machen. Weiters existieren interdisziplinäre Repositorien, welche oft über gemeinnützige oder internationale Organisationen betreut werden.[296]

Als österreichisches Beispiel eines institutionellen Repositoriums ist PHAIDRA[297] der Universität Wien zu nennen. Das Repositorium PHAIDRA gründet auf dem Fedora Commons Repository und wird von Universitäten in unterschiedlichen Ländern Europas verwendet. Sein Forschungsdatenmanagement folgt den Richtlinien der FAIR Richtlinien.[298,299] Institutionelle Repositorien sollen Forschungsergebnisse einer Institution sichern und sichtbar machen. Sie werden jedoch wenig genutzt und oft wissen Forschende nicht, dass solch ein Repositorium an ihrem Institut

[288] Siehe unter https://www.re3data.org/, abgerufen am 10.09.2021.
[289] Zu den Kriterien für offene Forschungsdaten, welche im Zuge von FWF geförderten Projekten entstehen, siehe https://www.fwf.ac.at/de/forschungsfoerderung/open-access-policy/
open-access-fuer-forschungsdaten, abgerufen am 10.09.2021.
[290] Sánchez Solís u. a. 2020, 112.
[291] Siehe unter https://www.coretrustseal.org/why-certification/certified-repositories/, abgerufen am 10.09.2021.
[292] Siehe unter https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html, abgerufen am 10.09.2021.
[293] Siehe https://www.re3data.org/, abgerufen am 19.10.2021.
[294] Siehe unter https://v2.sherpa.ac.uk/opendoar/, abgerufen am 10.09.2021.
[295] Bauer – Ferus 2018, 75.
[296] Sánchez Solís u. a. 2020, 111.
[297] Siehe https://phaidra.univie.ac.at/, abgerufen am 19.10.2021.
[298] Siehe https://datamanagement.univie.ac.at/forschungsdatenmanagement/, abgerufen am 19.10.2021.
[299] Hagmann 2018, 58.

existiert.[300] Dies kann unter anderem daran liegen, dass diese Repositorien oft nur Hochschulschriften oder andere Literaturbeiträge verwalten, während die Archivierung von Forschungsdaten und Datenbanken oft nicht möglich ist. Wenn der Zugang eines Repositoriums nur Angehörigen des jeweiligen Instituts gestattet ist, kann dies ebenfalls ein Grund sein, warum Forschende sich stattdessen für Fachrepositorien entscheiden. Ein eingeschränkter Zugang spricht gegen die Prinzipien von Open Science und des Semantic Web.[301]

Als Fachrepositorien für archäologische Daten gibt es auf internationaler Ebene beispielsweise das Open Context Repositorium[302]. Es verlinkt die Forschungsdaten mit anderen Quellen und arbeitet mit dem Deutschen Archäologischen Institut zusammen. Die Finanzierung erfolgt ähnlich wie bei Open Access Journals über Publikationsgebühren. Dies schließt Forscher*innen aus, welche nicht über ein Institut oder Projekt forschen und ihre Daten veröffentlichen möchten.[303] Die Betreiber von Open Context setzen sich sehr stark für die Möglichkeit der Wiederverwendung der gesicherten Daten ein.[304] Weitere Repositorien für archäologische Daten sind das Archaeology Data Service (ADS)[305] und das tDAR (The Digital Archaeological Record)[306]. In Österreich existieren zwei geisteswissenschaftliche Fachrepositorien, das ARCHE (A Resource Centre for the HumanitiEs)[307] an der Österreichischen Akademie der Wissenschaften und das GAMS (Geisteswissenschaftliches Asset Management System)[308] an der Universität Graz. Das ARCHE Repositorium stellt maschinenlesbare Metadaten zur Verfügung und wurde 2018 mit dem CoreTrustSeal ausgezeichnet.[309] Im Zusammenhang mit dem Projekt „A Puzzle in 4D" wurde in einer Case Study ein Repository für archäologische Daten weiterentwickelt.[310] Das GAMS Archiv stellt ebenso maschinenlesbare Metadaten zur Verfügung. Das CoreTrustSeal wurde dem Repositorium 2019 ausgewiesen.[311] M. Trognitz sieht einen Mangel an archäologischen Fachrepositorien für Forschungsdaten in Österreich, da die meisten Archive nur für Fachzeitschriften und nicht für Forschungsdaten zur Verfügung stehen. Die Repositorien ARCHE, GAMS und PHAIDRA wären die einzigen Möglichkeiten, archäologische Daten in Österreich zu archivieren, wobei nur zwei davon, ARCHE und GAMS, das CoreTrustSeal besitzen.[312] Viele weitere Fachkolleg*innen

[300] Novotny – Seyffertitz 2018, 88–92.
[301] Torggler – Andrae 2018, 114–120.
[302] Siehe unter https://opencontext.org/, abgerufen am 19.10.2021.
[303] Kansa 2016, 445–458.
[304] Kansa – Kansa 2018, 91 f.
[305] Siehe http://archaeologydataservice.ac.uk/, abgerufen am 19.10.2021.
[306] Siehe https://www.tdar.org/, abgerufen am 19.10.2021.
[307] Siehe https://arche.acdh.oeaw.ac.at/browser/, abgerufen am 19.10.2021.
[308] Siehe http://gams.uni-graz.at/, abgerufen am 19.10.2021.
[309] Trognitz – Ďurčo 2018, 218 f.
[310] Hiebel u. a. 2021b, 2.
[311] Trognitz 2021, Kapitel 6.1.
[312] Trognitz 2021, Kapitel 5.1-6.1.

sehen einen Mangel an archäologischen Repositorien, welche Daten sichern und für die Zukunft verfügbar machen.[313] Die ARIADNE Infrastruktur[314] für das Finden von Daten in archäologischen Repositorien in Europa arbeitet laufend an der Erweiterung der Aggregation von Datensets. Ihr Fortschritt steht in direktem Zusammenhang mit den verfügbaren Repositorien in den einzelnen Ländern.[315]

Für die Kategorie der interdisziplinären Repositorien können beispielsweise Zenodo[316], Dryad[317] oder das Open Science Framework (OSF)[318] genannt werden. Sie sind offen für alle Arten von Forschungsdaten und werden vom FWF empfohlen.

2.2.3.3.3 Zukunft der Repositorien

Die vielen Open Science Initiativen[319] und die Forderung vieler Fördergeber für offene Forschungsdaten sind Indizien dafür, dass Forschungsdaten und deren Archivierung und Rolle für die Zukunft immer wichtiger werden. Infrastrukturprojekte wie ARIADNE weisen darauf hin, dass der Schlüsselmoment, zu handeln und die Archivierung von Forschungsdaten fest im Datenmanagement zu verankern, gekommen ist, um Datenverlust zu vermeiden.[320]

Daher ist es wichtig, die Repositorien als immer größer werdenden Teil der Infrastruktur miteinzubauen. Der Archivierungsschritt soll als Element und Werkzeug des Datenmanagements verstanden werden.[321] Es ist notwendig, ein größeres Bewusstsein hierfür unter den Forscher*innen und den Verantwortlichen für finanzielle und organisatorische Fragen zu bilden.[322] Forscher*innen sollten mehr Unterstützung erhalten und den gesamten Forschungsdatenzyklus hindurch begleitet werden. S. Blumesberger betont den vermehrten Personalbedarf für Repositorien in der Zukunft, da Mitarbeiter*innen mit technischem Wissen und einem Einblick in den Alltag von Forscher*innen so mehr Ressourcen für die Kommunikation mit ihnen und deren Betreuung hätten.[323] Neben der direkten Betreuung ist auch der Aufbau von Richtlinien für die Datenkuration und die Harmonisierung der Prozesse von der Kreierung bis zur Archivierung der Daten wichtig.[324]

---

313 Beispielsweise Richards u. a. 2021; Geser 2019, 1.
314 Siehe http://legacy.ariadne-infrastructure.eu/about-ariadne/introduction/, abgerufen am 19.10.2021.
315 Geser 2019, 1 f.
316 Siehe unter https://zenodo.org/, abgerufen am 19.10.2021.
317 Siehe unter https://datadryad.org/stash, abgerufen am 19.10.2021.
318 Siehe unter https://osf.io/, abgerufen am 19.10.2021.
319 Open Science Initiativen sind z. B. das Open Science Network Austria (OANA, siehe unter https://oana.at/, abgerufen am 19.10.2021), OpenAIRE (siehe unter https://www.openaire.eu/, abgerufen am 19.10.2021) oder die European Open Science Cloud (EOSC, siehe unter https://eosc-portal.eu/, abgerufen am 19.10.2021).
320 Richards u. a. 2021.
321 Stigler – Steiner 2018, 214 f.; Blumesberger 2020, 509; Kansa – Kansa 2018, 91 f.; Zozus 2017, 299.
322 Trognitz 2021, Kapitel 5.
323 Blumesberger 2020, 506–509.
324 Niccolucci – Richards 2013.

Die Ansprüche wissenschaftlicher Daten steigen weiter an und die Summe an Daten wird immer größer werden.[325] Zurzeit liegt die Entwicklung und Verfügbarkeit von Archiven noch hinter der Menge an digitalen Daten.[326] Vor allem institutionelle Repositorien wurden schnell erschaffen, um den Anforderungen der Fördergeber gerecht zu werden. Für die Langzeitarchivierung würden diese Archive von einer Zusammenarbeit mit Fachrepositorien profitieren.[327] Einige wichtige Schritte zur technischen Weiterentwicklung sind die Verbesserung von Software, um beispielsweise schnellere Uploads zu erreichen, die Einführung maschinenlesbarer Metadaten, um Suchmaschinen besser nutzen zu können und den Austausch mit Plattformen wie ARIADNE[328] oder Europeana[329] zu erleichtern und die FAIR Richtlinien für die Daten anzuwenden.[330] Um die Wartung eines Archivs zu gewährleisten, muss auf die Qualität der Daten geachtet werden. Es sollten keine undokumentierten und unsortierten Daten archiviert werden.[331] Weiters sollten die Formate, welche archiviert werden, limitiert sein, um die Migration in aktuelle Formate bewerkstelligen zu können.[332] Zur gleichen Zeit muss die dauerhafte Verfügbarkeit von Objekten und archivierten Datenbanken sichergestellt werden. Zertifizierungen von Repositorien und die Vernetzung mit der European Open Science Cloud (EOSC)[333] werden zukünftig an Bedeutung gewinnen.[334]

Es sind jedoch noch einige finanzielle und organisatorische Fragen offen. Die Langzeitarchivierung von Daten, welche beispielsweise durch Drittmittelprojekte entstanden sind, ist oft zeitlich begrenzt und die weitere Finanzierung ist ungeklärt.[335] Um die immer mehr werdenden Daten umsichtig archivieren zu können, schlägt G. Geser ein archäologisches Repositorium auf nationaler Ebene vor. Internationale Beispiele hierfür sind das ADS (Archaeology Data Service)[336] in Großbritannien, das DANS (Data Archiving and Networked Services)[337] in den Niederlanden oder das tDAR (The Digital Archaeological Record)[338] in den USA. Durch ein zentralisiertes Repositorium entsteht ein Knotenpunkt für Informationen und Daten, welches eine einfache,

[325] Bauer – Ferus 2018, 82.
[326] Trognitz 2021, Kapitel 5.
[327] Moore – Richards 2015, 40.
[328] Der Ariadnekatalog ist unter http://ariadne2.isti.cnr.it/ erreichbar, abgerufen am 12.09.2021.
[329] Archäologische Sammlungen in Europeana können unter folgendem Link angesehen werden: https://www.europeana.eu/en/collections/topic/80-archaeology, abgerufen am 12.09.2021.
[330] Blumesberger 2020, 506–509.
[331] Kansa 2016, 466–468.
[332] Stigler – Steiner 2018, 214 f.
[333] Zur Deklaration der European Open Science Cloud siehe EOSC 2017.
[334] Blumesberger 2018, 155–158; Bauer – Ferus 2018, 82.
[335] Blumesberger 2018, 158 f.; Kansa 2016, 466–468.
[336] Siehe unter https://archaeologydataservice.ac.uk/, abgerufen am 12.09.2021.
[337] Siehe unter https://dans.knaw.nl/en, abgerufen am 12.09.2021.
[338] Siehe unter https://www.tdar.org/, abgerufen am 12.09.2021.

schnelle und kostengünstige Lösung für die Forschung darstellen würde.[339] M. Trognitz sieht hierfür noch große Hürden und schlägt als Alternative ein zentrales Registrierungs- oder Aggregationsservice für archäologische Daten vor, welches es erleichtert, Daten aufzufinden und wiederzuverwenden. Dieses Service könnte ebenfalls als Knotenpunkt für internationale Aggregationsservices wie ARIADNE dienen.[340] J. D. Richards u. a. sehen einen ähnlichen Lösungsansatz für die zukünftige Archivierung. Repositorien sollten auf lokaler oder nationaler Ebene organisiert werden und die Metadaten sollten für Aggregatoren verfügbar gemacht werden. Neben der besseren Auffindbarkeit bleiben die Forschungsdaten im Besitz der jeweiligen Gemeinschaft. Die Wiederverwendung von Daten wird in der näheren Zukunft eine immer größer werdende Rolle spielen.[341]

Der Prozess der Archivierung sollte in die Praxis als normale Vorgehensweise nach jeder Forschungstätigkeit eingegliedert werden. Andernfalls drohen viele wertvolle Daten auf Projektfestplatten oder einzelnen Rechnern verloren zu gehen. Der Mehraufwand, die Daten in archivgeeignete Formate zu bringen und mit genügend Metadaten auszustatten, sollte dabei nicht unterschätzt werden. Doch es lohnt sich, wenn sichergestellt werden kann, dieses enorme Wissen für eine möglichst lange Zeit bewahren zu können. Im Zuge dessen wäre die Überlegung passend, die Einführung von Standards, Open Access und Open Data mit in die Diskussion miteinzubeziehen. So könnten Daten auch für zukünftige Forschungen zur Verfügung stehen und noch nicht vorstellbare Chancen entstehen.

Im nachfolgenden Kapitel werden die Aussichten für Forschungsdaten noch einen Schritt weitergesponnen. Es wird die Frage, welche Technologien und Konzepte es gibt, um das Potential von Open Data sogar noch weiter zu vergrößern, erläutert. Dies kann mit untereinander vernetzten, verlinkten und computerlesbaren Daten erreicht werden. Das Konzept dazu ist unter dem Begriff „Semantic Web“ oder dem in letzter Zeit häufiger verwendeten Begriff „Linked (Open) Data“ bekannt.

[339] Geser 2019, 1 f.
[340] Trognitz 2021, Kapitel 7.
[341] Richards u. a. 2021.

3 Semantic Web und Linked (Open) Data

> „In addition, if properly designed, the Semantic Web can assist the evolution of human knowledge as a whole."[342] Tim Berners-Lee, Vorreiter des World Wide Web und des Semantic Web

Semantic Web, Linked (Open) Data, eine Wissensevolution der Menschheit: was verbirgt sich hinter diesen Begriffen und wie kann Wissen dadurch weiterentwickelt werden? Welche Ideen und Konzepte sich hinter diesen Begriffen verbergen, wird in diesem Kapitel veranschaulicht. Die nötigen Grundbausteine, die Entstehungsgeschichte und die vielfältigen Möglichkeiten der Anwendung sollen eine Einführung in das breite Feld der technologischen Entwicklung von vernetzten Daten bieten.

3.1 Das Semantic Web

„*The Semantic Web is a web of data, in some ways like a global database*." [343] Eine globale Datenbank, diese Vision hatte T. Berners-Lee, Erfinder des World Wide Web[344], zu Beginn seiner Entwicklung des Semantic Web. Im bekannten Internet können menschenlesbare Informationen, wie Texte und Bilder, ausgetauscht werden. Man könnte es auch als „Web of Documents" bezeichnen, da die menschenlesbaren Informationen in Form von Dokumenten vorliegen. Die Grundidee des Semantic Web ist, Informationen in maschinenlesbare Form zu bringen[345] und ein „Web of Data", ein Datennetz, zu erschaffen, welches wie eine globale Datenbank funktionieren könnte.[346] Oder, anders ausgedrückt, den Zugang zum Netz für Computer zu vereinfachen.[347]

Die Definition des Wortes Semantik[348] besagt, sie sei ein „Teilgebiet der Linguistik, das sich mit den Bedeutungen sprachlicher Zeichen und Zeichenfolgen befasst" aber auch für „Bedeutung" oder „Inhalt (eines Wortes, Satzes oder Textes)" steht. Im Semantic Web sollen Maschinen die Bedeutung der Daten einordnen können. Diese maschinenlesbaren Daten und der verbesserte Zugang für Computer können einige Vorteile mit sich bringen. Das Semantic Web soll durch die mit Bedeutung versehenen Daten, Informationen und Inhalten Mensch und Computer ermöglichen, besser und effizienter zusammenarbeiten zu können.[349]

---

[342] Berners-Lee u. a. 2001, Evolution of Knowledge.
[343] Berners-Lee 1999, Machine-Understandable information: Semantic Web.
[344] Für Hintergründe zur Entstehung des World Wide Web siehe Berners-Lee – Fischetti 1999.
[345] Berners-Lee 1998a, 1.
[346] Berners-Lee 1999, Machine-Understandable information: Semantic Web.
[347] Antoniou u. a. 2012, 1.
[348] Nach Duden, https://www.duden.de/rechtschreibung/Semantik, abgerufen am 19.10.2021.
[349] Berners-Lee u. a. 2001, 34.

Die grundlegenden Aufgaben von Computern im Netz sind die Zuweisung von Such- oder Schlagwörtern und das Austauschen von Informationen zwischen Servern und Clients.[350] Hypertext-Links können Webseiten und Dokumente verlinken. T. Berners-Lee sieht in der Verlinkung einen der großen Vorteile des World Wide Web, da es zur Dezentralisierung beiträgt. Das Semantic Web soll nunmehr zum bestehenden „Web of Documents" von Computern prozessierbare Inhalte hinzufügen. Daraus soll das „Web of Data" wachsen. Die Dezentralisierung stellt beim Semantic Web, wie schon beim Internet, einen wichtigen Aspekt dar. Dadurch entstehen für alle Nutzer*innen Vorteile, die man im Vorfeld noch gar nicht genau abschätzen kann. Der Preis, den die Dezentralisierung mit sich bringt, musste auch schon beim World Wide Web bezahlt werden. Denn nicht alle Inhalte und Links sind einheitlich und konsistent, da jede*r Nutzer*in die Möglichkeit hat, sie zu erstellen. Aber genau dadurch wurde das exponentielle Wachstum des Internets überhaupt erst ermöglicht.[351]

Die Vorteile des Semantic Web können vielseitig sein. Sie zeigen sich zum Beispiel bei Suchanfragen. Man muss sich nicht mehr auf Schlüssel- oder Suchwörter beschränken, sondern kann Synonyme und Kontexte in Suchen miteinschließen. Personalisierte Programme, die den Inhalt von Webseiten prozessieren können, ermöglichen die Auswahl von nützlichen Links und neuen Seiten, abhängig von den jeweiligen Benutzer*innen und ihren Bedürfnissen. Ebenso ließen sich die Inhalte von unterschiedlichen Websites integrieren und Zusammenhänge müssten so nicht mehr manuell gesucht werden.[352] Im „Web of Documents" muss die intellektuelle Arbeit, wie Seiten und Informationen zusammentragen oder selektieren, von den Nutzer*innen selbst durchgeführt werden. Dies kann oft ein zeitaufwendiger und auch fruchtloser und frustrierender Prozess sein.[353] Für die Archäologie stellt allein dieser Punkt der verbesserten Suchanfragen bereits immense Vorteile dar, vor allem, wenn am Ende auch benutzbare Daten gefunden werden können, von deren Existenz der Suchende im Vorfeld noch nichts ahnte. Weitere Aspekte semantischer Technologien für die Archäologie werden in Kapitel 7 genauer beleuchtet. T. Berners-Lee, der Vorreiter des World Wide Web, sieht großes Potential im Semantic Web, Wissen global zu vernetzen und zu verbreiten und als Ganzes weiterzuentwickeln[354].

[350] Antoniou u. a. 2012, 1.
[351] Berners-Lee u. a. 2001, 34.
[352] Antoniou u. a. 2012, 1 f.
[353] Antoniou u. a. 2012, 1.
[354] Berners-Lee u. a. 2001, 35 f.

3.1.1 Grundbausteine des Semantic Web

Die Technologien, die grundlegend sind für das Semantic Web, werden in Kapitel 4 im Detail erläutert. Die Grundbausteine und fundamentalen Konzepte, auf denen das Semantic Web aufbaut, sind allerdings essenziell für das Verständnis, weshalb sie an dieser Stelle kurz umrissen werden (siehe Abbildung 6). Hier werden die Komponenten und Grundideen sowie die Grenzen des Semantic Web vorgestellt.

3.1.1.1 Standardisierte Sprache und Datenmodell

T. Berners-Lee betont, wie wichtig eine standardisierte Sprache für das Semantic Web ist. Diese soll dafür verwendet werden, sowohl die Daten selbst als auch ihre Struktur auszudrücken.[355] Dabei müssen grundlegende Regeln beachtet werden. Denn jede Sprache, die dafür verwendet wird, Daten auszutauschen, muss nach G. Antoniou u. a. drei Bestandteile vorweisen:[356] eine Syntax, ein Datenmodell und Semantik. Dabei bestimmt die Syntax, wie die Daten formuliert sind, das Datenmodell die Struktur und Ordnung der Daten und die Semantik, wie die Daten zu interpretieren sind.

HTML (Hypertext Markup Language) zum Beispiel ist die Sprache, in der Webseiten geschrieben werden. Bei dieser Sprache wird die Syntax mittels HTML-Tags, die in Pfeilsymbolen geschrieben sind, formuliert. Das Datenmodell ist das sogenannte „Document Object Model". Es gibt die Struktur des hierarchischen Baumes vor, in welchem die mit den HTML-Tags definierten Objekte eingeordnet werden. Die Semantik gibt dem Web-Browser vor, wie die Webseite einzuordnen ist.[357]

Auch für die Sprache des Semantic Web ist eine Syntax erforderlich. Ebenso ist ein Datenmodell nötig, das flexibel ist und von unterschiedlichen Branchen benutzt werden kann. Darüber hinaus müssen Computer-Anwendungen wissen, wie die Informationen, die durch das Datenmodell dargestellt sind, semantisch einzuordnen und zu interpretieren sind. Die Rolle des Datenmodells übernimmt das „Resource Description Framework", kurz RDF genannt. Es bietet den Rahmen für die Beschreibung der Daten. Die Syntax wird durch Serialisierungen bzw. Repräsentationen des Datenmodells Resource Description Frameworks, wie Turtle, RDF-XML, RDFa und weitere, ermöglicht. Die Semantik wird erfüllt, wenn Ontologien und kontrollierte Vokabulare verwendet werden.[358]

[355] Berners-Lee u. a. 2001, 35.
[356] Antoniou u. a. 2012, 23–25.
[357] Antoniou u. a. 2012, 23–25.
[358] Antoniou u. a. 2012, 23–25.

Detaillierte Erläuterungen über das RDF und der Serialisierungen werden in Kapitel 4.3 gegeben, jedoch soll die Grundfunktion des Datenmodells hier umrissen werden. RDF ist ein Datenmodell, das den Anforderungen des Semantic Web gerecht wird. Das RDF wird aus „Statements", also einzelnen Aussagen, aufgebaut. Dabei verwendet es sogenannte „Triples" (Dreiergruppen), um die Informationen in den Daten zu codieren. Ein Triple ist aufgebaut wie ein einfacher Satz, mit Subjekt, Prädikat und Objekt. Das Grundprinzip besteht daraus, Objekte, ihre Attribute und ihre Werte miteinander zu verbinden bzw. in Beziehung zueinander zu bringen.[359] Die Objekte, die in den Daten beschrieben werden, bilden das Subjekt des Satzes. Sie können reale Objekte, Webseiten, Personen, Konzepte usw. darstellen. Die Attribute dieser Objekte als Prädikate sind Eigenschaften wie „hat Material", „hat Erhaltungszustand", „besteht aus" etc. Die Werte als Objekte der Aussagen können beispielsweise eine andere Person, ein Material oder Ähnliches sein. Ein Beispiel eines solchen Statements könnte lauten: [Ring] (Subjekt; zu beschreibendes Objekt) [besteht aus] (Prädikat; Eigenschaft) [Gold] (Objekt; Wert). Durch diese Ausdrucksform können nahezu alle Daten auf sehr intuitive und einfache Weise beschrieben werden.[360]

3.1.1.2 Identifizierung mit URIs, Uniform Resource Identifier

Als wohl wichtigsten Grundbaustein des Semantic Web zählt T. Berners-Lee die Identifikation der einzelnen Elemente der Triples im RDF. Jedes Element wird dabei mit einem universellen Ressourcen Identifikator, einem sogenannten „Uniform Resource Identifier", kurz URI genannt, identifiziert.[361] Der URI ist eine Abfolge von Zeichen, welche eine Ressource eindeutig wieder erkennen lässt.[362] Meist handelt es sich dabei um einen „Uniform Resource Locator" oder URL (z. B. ein Link zu einer Website). Dieser zeigt den Ort im Web an, an welchem Informationen zu der Ressource zu finden sind. Durch die Identifikation wird sichergestellt, dass die Konzepte in einem Dokument nicht nur Worte sind, sondern durch Definitionen, die jeder online einsehen kann, genau beschrieben werden.[363] Alle Elemente in den Daten müssen eine solche URI besitzen. Nur so können Daten transparent, nachvollziehbar und richtig interpretierbar sein.[364]

3.1.1.3 Ontologien

T. Berners-Lee gibt noch eine weitere Komponente für das Funktionieren des Semantic Web an: Ontologien. Wenn Daten aus verschiedenen Quellen zusammengeführt werden, könnten sie verschiedene Identifikatoren benutzen, um das Gleiche zu beschreiben (wie z. B. eine Ortsangabe

[359] Antoniou u. a. 2012, 24 f.
[360] Berners-Lee u. a. 2001, 35.
[361] Berners-Lee 1999, The fundamentals: The Universal Web.
[362] Berners-Lee u. a. 2005, 1.
[363] Berners-Lee u. a. 2001, 35.
[364] Berners-Lee 1999, The fundamentals: The Universal Web.

oder eine bestimmte Ansprache eines Fundobjektes) und es kann nicht erkannt werden, ob es sich hier um das gleiche Konzept handelt. Ontologien können dieses Problem lösen.[365] Wenn mehrere archäologische Funddatenbanken mithilfe einer Ontologie modelliert werden, können gleiche Konzepte, wie beispielsweise der Fundort, besser identifiziert werden. Dies ist auch dann der Fall, wenn in verschiedenen Datenbanken die Spalte, in welcher der Fundort steht, anders betitelt wurde. Durch die Ontologie wird diese Spalte mit der Bedeutung „das Objekt wurde an diesem Ort gefunden" versehen. Dies ist ein wesentlicher Vorteil für die Integration und Interoperabilität über mehrere Datenbänke hinweg.[366]

In der Philosophie wird die Ontologie mit der Lehre des Seins beschrieben. In der Informatik versteht man darunter ein System, welches die Beziehungen von Informationen zueinander formal beschreibt und definiert. Die Beschreibungen verwenden jeweils wieder URIs, um eindeutige Definitionen für ihre Bedeutungen bieten zu können. Welche Ontologien im Semantic Web und respektive für archäologische Daten verwendet werden können, wird in Kapitel 5 dargelegt.

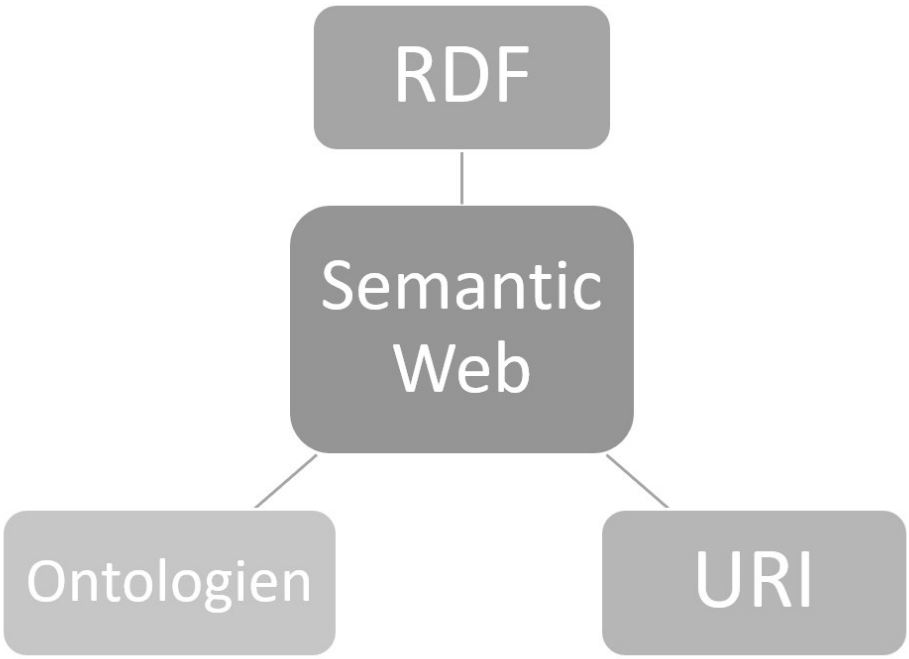


Abbildung 6 - Bausteine des Semantic Web.

[365] Berners-Lee u. a. 2001, 35.
[366] Binding u. a. 2008, 286–289.

Zusammenfassend können drei Konstruktionsprinzipien als wichtigste Bausteine für das Semantic Web genannt werden (siehe Abbildung 6)[367]:

1. RDF – Das Resource Description Framework wird verwendet als Datenmodell für die Objekte und ihre Beziehungen zueinander. Serialisierungen davon werden als Sprache genutzt.

2. URI – Der Uniform Resource Identifier dienen als Identifizierungsgrundlage für einzelne Daten als auch für ihre Relationen.

3. Ontologien – Die hierarchisch aufgebauten Vokabulare werden als Daten-modell zur formalen Repräsentation der Semantik verwendet. Es gibt unter-schiedliche Ontologien (z. B. RDFS – RDF Schema, OWL – die Web Ontology Language, Dublin Core, CIDOC-CRM etc.) die ebenso URIs benutzen, um die Objekte und Eigenschaften zu definieren.

3.1.1.4 Bausteine des Semantic Web

Die Konstruktionsprinzipien des Semantic Web, RDF, URI und Ontologien, bauen aufeinander auf und stehen miteinander in Beziehung. Um Daten semantisch aufzubereiten, werden bestimmte Technologien verwendet. Das semantische Anreichern der Daten muss (semi-)manuell durchgeführt werden. Dazu gab es 2000 vom W3C formale Empfehlungen von T. Berners-Lee, um die Technologien, die dafür notwendig sind, als Standards zu deklarieren. Er entwickelte den sogenannten Semantic Web „Layer Cake", der die Technologien sowohl als Überblick als auch in ihrem Bezug zueinander darstellt. Der Layer Cake entwickelte sich über Zeit von einer konzeptuellen Darstellung (siehe Abbildung 7) zu konkret definierten Technologien hin (siehe Abbildung 8).[368]

Die drei unteren Schichten stellen Technologien dar, die schon im Hypertext-Web verwendet werden. Der IRI, „Internationalized Resource Identifer"[369], ist die international abgewandelte Form des URI. Er wird ähnlich wie der URI verwendet, benutzt jedoch auch Buchstaben und Zeichen des universellen Zeichensets, des Unicodes[370]. XML, die „Extensible Markup Language", wird zum Beschreiben für hierarchisch aufgebaute Daten verwendet[371] und XML Namespaces, sogenannte Namensräume, sind ein Weg, um Elemente in einer XML-Datei zu definieren[372]. Da die Technologien, die im World Wide Web genutzt werden, auch für das Semantic Web verwendet

[367] Antoniou u. a. 2012, 4.
[368] Isaksen 2011, 22 f.
[369] Dürst – Suignard 2005, 1.
[370] Definition im ISO 10646, siehe ISO 2010.
[371] Bray u. a. 2008, Kapitel 1.
[372] Bray u. a. 2009, Kapitel 2.

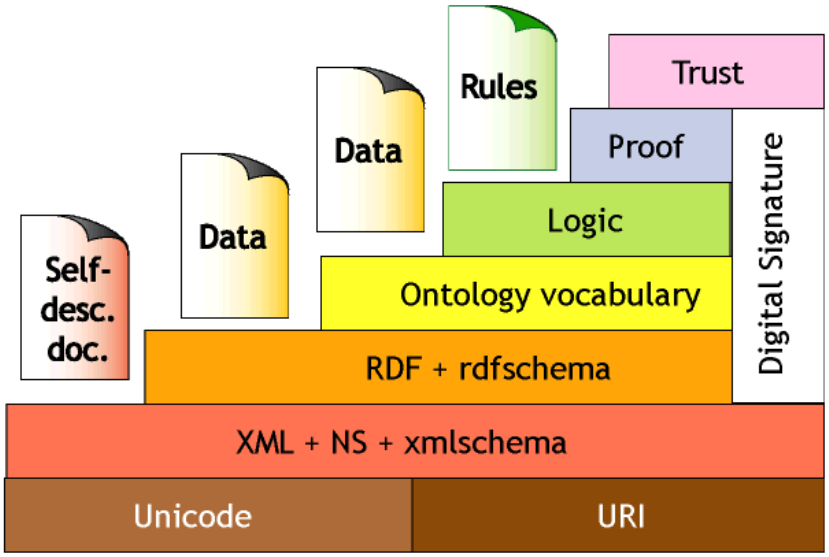


Abbildung 7 - Semantic Web Layer Cake 2000 (Quelle: https://www.w3.org/2001/09/06-ecdl/slide17-0.html, abgerufen 19.10.2021).

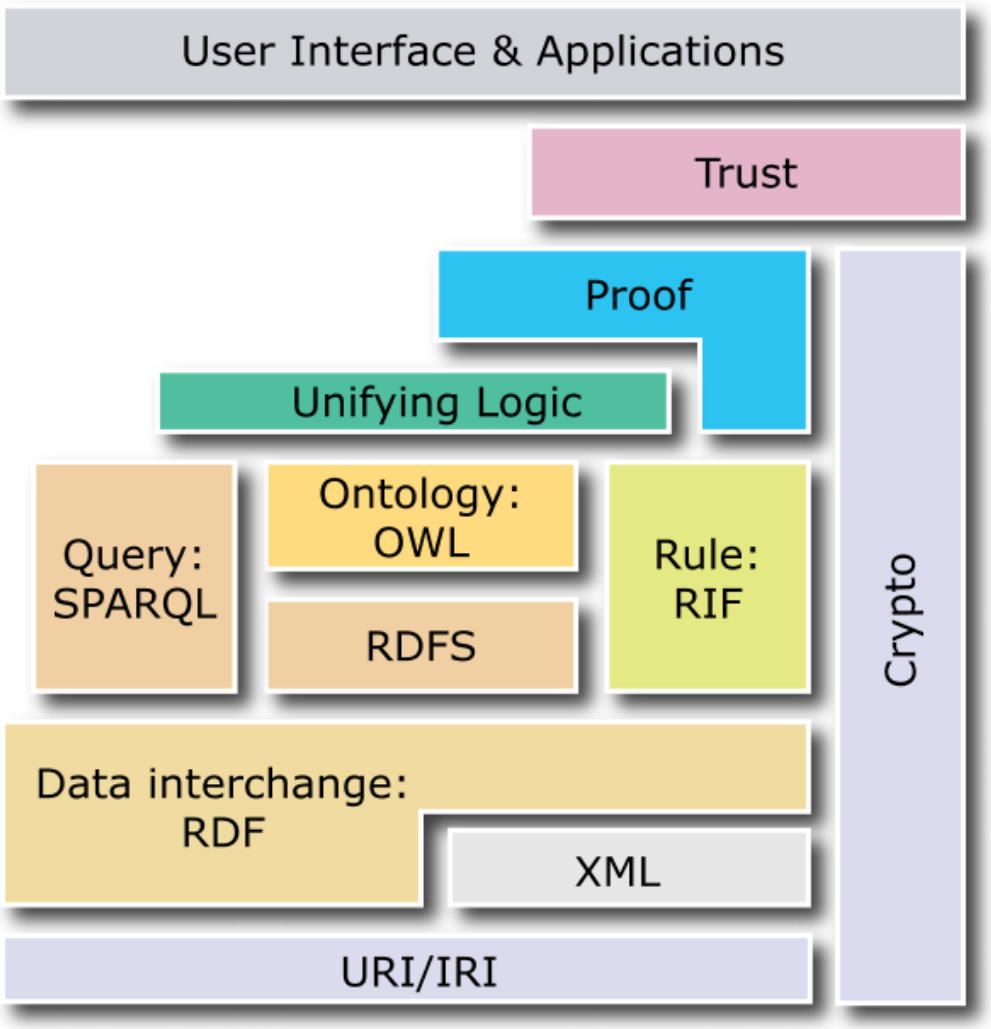


Abbildung 8 - Semantic Web Layer Cake 2007 (Quelle: https://www.w3.org/2007/03/layerCake.png, abgerufen 19.10.2021).

werden, kann es als Erweiterung des bestehenden Webs gesehen werden und wird nicht als eigenständiges System verstanden.

Die Schichten, die in der Mitte dargestellt werden, sind standardisierte Technologien des Semantic Web. RDF wird als Datenmodell genutzt und RDFS, das RDF Schema, bietet zusätzliches Vokabular für RDF Daten, um sie modellieren zu können[373]. OWL, die Web Ontology Language, kann als Ontologie zur Definition von Begriffen in den Daten verwendet werden[374] und SPARQL als Abfragesprache für RDF Daten[375]. Die oberen Schichten (Cryptography, Proof, Trust) stehen für die Sicherheit und Kontrolle. Sie sind noch nicht vollständig geklärt.[376] Die Technologien des Semantic Web werden in Kapitel 4 genauer beschrieben.

3.1.2 Fundamentale Konzepte des Semantic Web

Die Idee des World Wide Web war bahnbrechend für die Art und Weise, wie Menschen Informationen geteilt, ausgetauscht und erhalten haben. Dabei waren die niederschwelligen Möglichkeiten, sich zu beteiligen, ein Hauptfaktor des Erfolgs. Daher wurden einige dieser Konzepte bei der Entwicklung des Semantic Web ebenfalls berücksichtigt.[377] Im Folgenden werden die grundlegenden Konzepte des Semantic Web vorgestellt, welche Möglichkeiten sich ergeben können und auf welche Umstände geachtet werden sollte (siehe Abbildung 9).

Semantic Web ist nicht künstliche Intelligenz

T. Berners-Lee betont, das Semantic Web und maschinenlesbare Daten seien nicht mit künstlicher Intelligenz gleichzusetzen. Maschinen können nicht wirklich „verstehen", was geschrieben wird. Die Grundidee von maschinenlesbaren Daten ist das Anreichern von Daten mit Definitionen. Maschinen können daraufhin Probleme lösen, die zuvor definiert wurden.[378] Der Computer kann keine menschlichen Fehler oder falsche Interpretationen, die eingegeben wurden, korrigieren. Der Mehraufwand zu Beginn, die Daten entsprechend aufzubereiten, soll am Ende belohnt werden.[379] T. Berners-Lees Idee stellt das Web als dezentralisiertes Netzwerk dar, in dem maschinenlesbare Daten, die miteinander verlinkt sind, Probleme lösen und Daten integrieren können.[380]

[373] Brickley u. a. 2014, Kapitel 1.
[374] W3C OWL Working Group 2012, Kapitel 1.
[375] W3C SPARQL Working Group 2013, Kapitel 1.
[376] Thiery 2013, 33.
[377] Allemang – Hendler 2012, 6.
[378] Berners-Lee 1998b, 1.
[379] Isaksen 2011, 18 f.
[380] Gerth 2017, 3.

Logische Schlussfolgerungen - Inferenz

Durch die Verwendung von Ontologien, die die Beziehungen zwischen den einzelnen Daten definieren, können auch hierarchische Einordnungen zugewiesen werden. Dies ermöglicht die Anwendung von Inferenzregeln. Dies sind Schlussfolgerungen, die logisch aus den angegebenen Aussagen ohne zusätzliche Angaben dazu abgeleitet werden können. Sie werden vom Computer benutzt, um Schlüsse zu ziehen, die von den Datenherausgeber*innen nicht direkt eingegeben wurden.[381]

Die Taxonomie, also die Einordnung in ein bestimmtes System, beschreibt die Klassen und Subklassen der Objekte und ihre Beziehungen zueinander. Diesen Klassen können Eigenschaften, sogenannte „Properties", zugeordnet werden, welche die Subklassen somit ebenfalls übernehmen. Wenn zum Beispiel in den Datensätzen eine Postleitzahl einem Bundesland zugeordnet wurde, kann eine Adresse mit dieser Postleitzahl automatisch auch dem Bundesland zugeordnet werden. Ontologien können auch Lösungen für Terminologien bieten, Suchanfragen im Internet effektiver machen und Verbindungen von Informationen bereithalten. Fragen, die sonst nicht von einer einzigen Quelle beantwortet werden können, können bessere Ergebnisse erzielen, die ansonsten mühselig per Hand durchsucht und zusammengesammelt werden müssten.[382]

Unvollständigkeit der Information – die Open World Assumption

Ein weiteres fundamentales Konzept des Semantic Web bzw. einem Web of Data wurden von D. Allemang und J. Hendler mit Hilfe des AAA-Leitspruchs zusammengefasst, der besagt, „*Anyone can say Anything about Any topic*."[383] Jeder kann alles über jedes Thema sagen. Schon das bekannte Web hat diese Form der dezentralen Freiheit. Jeder kann eine Webseite erstellen und über jedes Thema schreiben. Beim Semantic Web bedeutet dies, jedem muss es möglich sein, Daten so zu beschreiben, dass sie mit anderen Daten verbunden werden können. Das Web muss als offene Welt (engl. *Open World*) gesehen werden. Folglich kann niemand annehmen, er habe alle Informationen im Netz zur Verfügung. Es können jederzeit neue Informationen auftauchen und es gibt Informationen, die nicht gefunden worden sind. Dieses Konzept wird auch „Open World Assumption" genannt. Es bietet den Vorteil, dass jeder sich beteiligen kann. Der Preis, der dafür gezahlt wird, ist, dass angenommen werden muss, auch fehlerhafte oder falsche Daten zu finden.[384]

[381] Berners-Lee u. a. 2001, 35 f.
[382] Berners-Lee u. a. 2001, 35 f.
[383] Allemang – Hendler 2012, 6.
[384] Allemang – Hendler 2012, 6–11.

Fehlen einheitlicher Namensgebung - Nonunique Naming Assumption

Zusätzlich zur „Open World" muss angenommen werden, dass unterschiedliche Namen für die gleichen Dinge verwendet werden können. Um Objekten oder Konzepten eine eindeutige Zuweisung zu geben und zu benennen, gibt es im Semantic Web die URIs. Doch es gibt keinen weltweiten Katalog von Namen, um sichergehen zu können, ein Konzept habe nur diese eine Bezeichnung. Darum kann es für dasselbe Konzept oder dieselbe Sache auch mehrere URIs geben. Dies ist nicht zu vermeiden, denn die Namensgeber*innen koordinieren sich nicht untereinander. Dieser Aspekt wird „Nonunique Naming Assumption", die Annahme nicht einzigartiger Namen, genannt.[385] Im Rahmen eines Semantic Web stellt dies ein Problem dar, das es zu lösen gilt; unter anderem versucht man, dies über Ontologien zu bewerkstelligen, wie weiter unten noch ausführlich dargestellt wird (siehe Kapitel 5).

Netzwerk-Effekt

Der Netzwerk Effekt (engl. *Network Effect*) hat auch dem Web of Documents schon zu seiner weiten Verbreitung geholfen. Er besagt, je mehr Menschen Netzwerke verwenden würden, desto mehr Nutzen haben sie auch für neue Anwender*innen. Dies lässt das Netz immer weiterwachsen. Zu Beginn des World Wide Webs hatte man Bedenken, wer denn das Web mit Inhalten und Webseiten füllen sollte. Doch als die Infrastruktur erst einmal geschaffen war und einem breiten Publikum zur Verfügung stand, konnte man sehen, wie eine Unzahl an Leuten dazu beitragen konnte und das Web wuchs schneller als erwartet. Diesen Effekt erhofft man sich auch für das Semantic Web.[386]

Wildwuchs der Daten

Durch all die vorhergehenden Effekte entsteht eine gewisser Datenwildwuchs, in der sich viele wertvolle Informationen finden lassen, aber auch vieles an nicht verwertbaren Daten auftaucht.[387] Wenn alle Nutzer*innen Daten ins Semantic Web einfügen können, werden auch einige nicht verständlich oder unbrauchbar sein. Dies ist allerdings notwendig, um die Freiheit im Web zu wahren, denn nur so kann das Semantic Web auch wachsen und erfolgreich sein.[388]

[385] Allemang – Hendler 2012, 9–11.
[386] Allemang – Hendler 2012, 7–11.
[387] Allemang – Hendler 2012, 7.
[388] Allemang – Hendler 2012, 11 f.

Einige hier vorgestellte Konzepte gelten auch schon für das bekannte Web. Sich in der Fülle der Informationen zurecht zu finden, erfordert einige Zusammenarbeit und die Weiterentwicklung von Navigationstools und Hilfsmitteln wie beispielsweise Suchmaschinen und Enzyklopädien. [389] Bei einem Web of Data wäre es noch anspruchsvoller, sich zurechtzufinden. Fehler fallen weniger offensichtlich ins Auge und der Überblick fällt schwer. Wie also können Nutzer*innen darauf vertrauen, Informationen, die von Maschinen zusammengesammelt und zusammengestellt wurden, würden auch stimmen? Diesen Dschungel an Daten können nur Ontologien in die richtigen Bahnen lenken. Wenn Daten mit Semantic Web Technologien modelliert werden, können nützliche Informationen aus den zuvor nicht strukturierten Daten hervortreten. Ohne ein Datenmodell sind die Massen an Daten jedoch nicht zu bewältigen und besitzen keinen großen Mehrwert. Denn dann kann nicht festgestellt werden, wie diese Daten zusammenhängen. Daten in einem durchdachten Datenmodell formen ein „Web der Informationen". Diese Informationen können fließen und sollen eine weltweite Informationsquelle ergeben, die laufend wachsen kann.[390]

Abbildung 9 - Die Konzepte des Semantic Web (Quelle: M. Simon nach Allemang – Hendler 2012, 6-12).

[389] Allemang – Hendler 2012, 11.
[390] Allemang – Hendler 2012, 11.

Zusammenfassend können folgende Konzepte des Semantic Web genannt werden (siehe Abbildung 9):

- Maschinenlesbare Daten sind nicht mit künstlicher Intelligenz zu verwechseln
- Inferenzregeln können durch Ontologien angewendet werden
- AAA-Slogan - Anyone can say Anything about Any topic
- Open World Assumption – jeder kann sich beteiligen
- Nonunique Naming Assumption – gleiche Dinge können unterschiedlich benannt werden
- Der Network Effekt – je mehr Nutzer teilnehmen, desto schneller wächst das Web
- Wildwuchs der Daten – der Preis für die Freiheit im Web sind auch unbrauchbare Daten

3.1.3 Forschungsgeschichte und Entwicklung des Semantic Web

Der Vordenker des Semantic Web ist auch jener des World Wide Web im Jahr 1989, Sir Tim Berners-Lee.[391] Er ist Direktor des World Wide Web Consortiums (W3C), dessen Ziel es ist, Webstandards zu entwickeln und laufend zu aktualisieren.[392] Die erste Idee des Semantic Web äußerte T. Berners-Lee bereits auf der ersten World Wide Web Conference 1994 in Genf.[393] Als offizielle W3C Aktivität[394] findet das Semantic Web 1998/1999 in Form von drei W3C „Design Issues", die T. Berners-Lee verfasst hat, statt.[395,396,397] Die Design Issues „The Semantic Web Road Map", „What the Semantic Web is Not" und „Web Architecture from 50,000 feet" bilden den Auftakt für das Semantic Web.

T. Berners-Lee beschreibt die Grundidee als eine Art globale Datenbank, in der die Daten miteinander verwoben sind. Die Methode, die dem Semantic Web dabei zu Grunde liegt, ist die Entwicklung von Sprachen, die entwickelt werden, um Informationen in maschinenlesbarer Form auszudrücken. [398] Die Entwicklung des Semantic Web teilt L. Isaksen in zwei grob gefasste Perioden ein: die Semantic Web Ära (2001-2005) und die Linked Data Ära (2006-laufend[399]).[400]

[391] Zu seiner Biografie auf der W3C-Webseite siehe https://www.w3.org/People/Berners-Lee/, abgerufen am 21.10.2021.
[392] Siehe auch https://www.w3.org/Consortium/, abgerufen am 21.10.2020.
[393] Shadbolt – Berners-Lee 2006, 96.
[394] Siehe https://www.w3.org/2013/data/, abgerufen am 21.10.2021.
[395] Berners-Lee 1998a, 1–9.
[396] Berners-Lee 1998b, 1–6.
[397] Berners-Lee 1999, 1–17.
[398] Berners-Lee 1998a, 1.
[399] Seine Dissertation ist 2011 erschienen.
[400] Isaksen 2011, 20.

Die *Semantic Web Ära* wird durch den vielzitierten Artikel von T. Berners-Lee „The Semantic Web“[401], der 2001 im Scientific American erschienen ist, eingeleitet. Darin erklärt er, das Semantic Web soll eine Erweiterung des bekannten Webs sein, in dem die Informationen untereinander in Beziehung stehen und eine Bedeutung haben. Er sieht die größte Herausforderung für die Semantic Web Community, die Semantic Web Technologien, wie die Sprache, zu entwickeln und dem Web statt nur Dokumenten auch Logik hinzuzufügen.[402]

In der Forschung werden nachfolgend Fachzeitschriften herausgegeben und Symposien organisiert. 2001 entstand die International Semantic Web Conference (ISWC), 2004 das europäische Pendant, die Extended Semantic Web Conference (ESWC)[403], 2005 fand die erste Semantic Technologies Conference (SemTech) statt und 2003 veröffentlichte das Journal of Web Semantics ihre erste Ausgabe. Es werden Bücher publiziert, die eine Einführung ins Semantic Web geben, wie zum Beispiel „A Semantic Web Primer“[404].[405] Die formale Empfehlung des W3C für einige Technologien des Semantic Layer Cake aus dem Jahr 2000 (siehe Abbildung 7) war der nächste große Schritt.[406]

Die Forschung und das W3C bemühen sich, das Semantic Web zu etablieren, doch es gab einige Startschwierigkeiten. Das Semantic Web wächst, so wie auch das World Wide Web, wenn viele Parteien daran teilnehmen und ihre Inhalte zugänglich machen.[407] 2006 publizieren N. Shadbolt und T. Berners-Lee den Artikel „The Semantic Web Revisited“, in dem sie einen Zwischenstand des Semantic Web festhalten. Sie merken an, das Web of Data wäre immer noch nicht realisiert worden, aber ein wachsender Anteil an unterschiedlichen Wissenschaften hätten Ontologien für sich entdeckt. Daher sehen sie das Bedürfnis von Datenintegration in der Wissenschaft generell im Steigen. Das Semantic Web ist 2006 an dem Punkt, an dem die Standards weitestgehend feststehen und eine Wiederverwendung der eigenen Daten sowie die von anderen Teilnehmern möglich ist und langsam beginnen kann. Die durchgeführten Projekte haben jedoch noch nicht die revolutionierende Reichweite erreicht, die wünschenswert wäre.[408]

S. Auer u. a. betonen in einem Artikel aus dem Jahr 2009, die eindeutig größte Anzahl der Beiträge würde immer noch überwiegend in Form von traditionellen Webseiten erfolgen. Sie sehen den potenziellen Durchbruch von semantisch repräsentierten Daten bis 2009 noch nicht. Es fehlt

[401] Siehe Berners-Lee u. a. 2001.
[402] Berners-Lee u. a. 2001, 34.
[403] Zu Beginn wurde die Konferenz „European Semantic Web Conference“ genannt.
[404] Antoniou u. a. 2012.
[405] Isaksen 2011, 26 f.
[406] Isaksen 2011, 22.
[407] Isaksen 2011, 26 f.
[408] Shadbolt – Berners-Lee 2006, 96–99.

sowohl an semantisch zugänglich gemachten Beiträgen als auch an Suchmaschinen oder ähnlichen Möglichkeiten, Inhalte zu finden, auch wenn sie öffentlich gemacht wurden. Fachkreise hätten den Mehrwert anerkannt, der entsteht, wenn relationale Daten als RDF veröffentlicht werden. Doch obwohl bis 2009 bereits einige Werkzeuge dafür veröffentlicht worden waren, war kein signifikanter Anstieg zu erkennen. Sie führen dies auf die Schwierigkeit zurück, Modellierungen erfolgreich durchzuführen sowie die generell hohe „Einstiegsbarriere".[409] Technische Herausforderungen konnten im Allgemeinen von den Entwicklern überwunden werden. Allerdings führte dies aufgrund der steigenden Komplexität zu einem noch größeren Hindernis für Neueinsteiger*innen. L. Isaksen spricht vom „Elfenbeinturm" der Wissenschaft und meint, das Semantic Web müsse sich dringend an den Mainstream anpassen, um Erfolge verzeichnen zu können.[410]

2012 erscheint die zweite Auflage des Buches „Semantic Web for the Working Ontologist" von D. Allemang und J. Hendler. Darin stellen sie fest, das Semantic Web hätte regeren Anklang gefunden, sobald mehr Daten im RDF-Format verfügbar wären. Zu diesem Zeitpunkt gibt es bereits mehr öffentlich zugängliche Daten im RDF Format, wie beispielsweise die Wikipedia-Infoboxen (DBpedia[411]) und Regierungsdaten.[412]

Die *Linked Data Ära* nach L. Isaksen beginnt 2006 mit dem Design Issue „Linked Data"[413] von T. Berners-Lee. Darin stellt T. Berners-Lee vier Regeln auf, oder wie er sie nennt, „expectations", also Erwartungen. Diese und die weitere Entwicklung der Linked Data werden im nachfolgenden Kapitel beleuchtet. Diese Prinzipien sollen einen schnelleren Einstieg für strukturierte, verlinkte Daten zu bieten und helfen das Web of Data zum Wachsen zu bringen und seine Reichweite zu erhöhen.[414]

3.2 Linked (Open) Data

Der erhoffte Durchbruch des Semantic Web fiel geringer aus als erwartet. Doch viele der Ideen und Technologien werden unter dem Stichwort „Linked Data" weitergeführt und weiterentwickelt. Immer mehr Organisationen veröffentlichen ihre Daten nach den Linked Data Prinzipien (siehe Abbildung 10) und führten so die Idee des Semantic Webs weiter. [415]

[409] Auer u. a. 2009, 621 f.
[410] Isaksen 2011, 27.
[411] Siehe https://wiki.dbpedia.org/about, abgerufen am 21.10.2021.
[412] Allemang – Hendler 2012, 7 f.
[413] Berners-Lee 2006.
[414] Isaksen 2011, 27.
[415] Stuart 2012, 175.

3.2.1 Vernetzte Daten - Linked Data

Der Begriff „Linked Data“ beschreibt eine Anleitung, wie strukturierte Daten bestmöglich veröffentlicht und miteinander in Beziehung gesetzt werden können.[416] Sie sind als Empfehlungen für Veröffentlicher*innen von strukturierten Daten zu verstehen.[417] Das Prinzip der Linked Data soll helfen, das Web of Documents in ein verlinktes Web, in das Web of Data, und in Folge in das Semantic Web zu transformieren. Die Bezeichnung Linked Data stellt die Regeln dar, die beachtet werden sollten, wenn strukturierte Daten im Netz veröffentlicht und verlinkt werden.[418] Im Web of Documents sind die Links zwischen den Dokumenten in HTML geschrieben, während sie im Web of Data in RDF verfasst sind. Links zwischen den Dokumenten und Daten sind wichtig, damit das Web immer weiterwachsen kann. Die Prinzipien der Linked Data sind von T. Berners-Lee 2006 in einem der W3C Design Issues mit dem Titel „Linked Data“ veröffentlich worden (siehe Abbildung 10):[419,420]

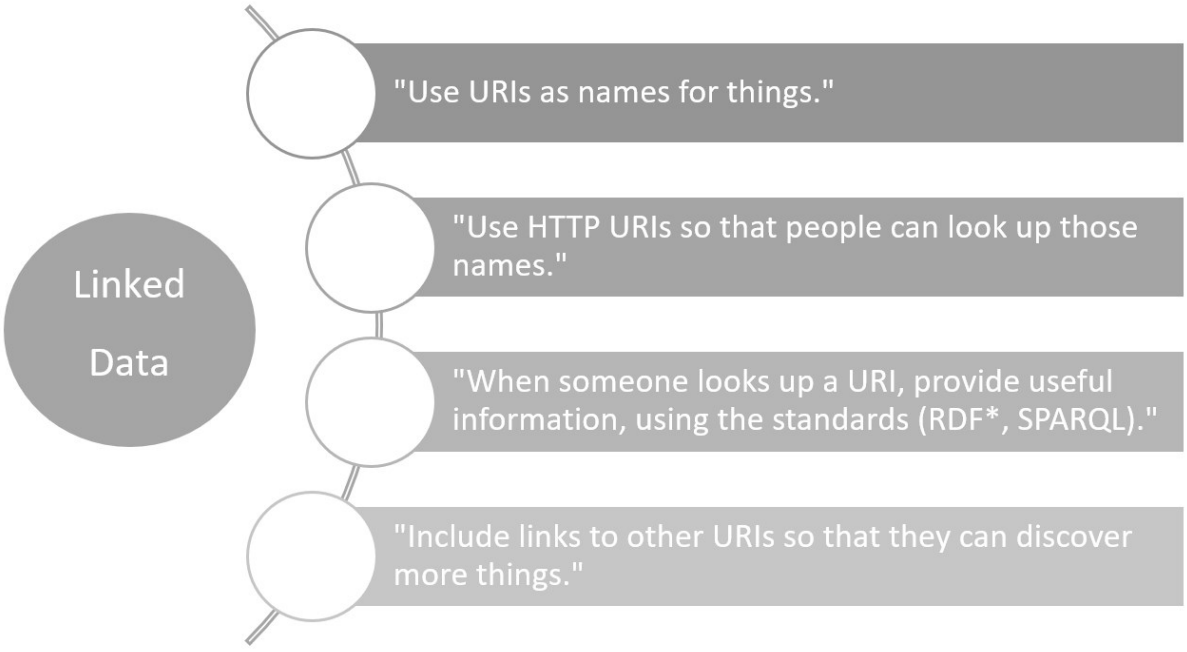


Abbildung 10 – Linked Data Empfehlungen nach Berners-Lee 2006, 1.

[416] Bizer u. a. 2009, 1.
[417] Bizer u. a. 2009, 1.
[418] Auer 2014, 1.
[419] Berners-Lee 2006, 1.
[420] Zu Punkt 3, *RDF: Die Erwähnung der Standards RDF und SPARQL wurden bei der Überarbeitung 2009 hinzugefügt. Quelle: WayBack Machine http://web.archive.org/web/20090526040403/http://www.w3.org/DesignIssues/LinkedData.html, abgerufen am 21.10.2021.

Die erste Regel besagt, URIs sollen verwendet werden, um Objekte eindeutig zu identifizieren. T. Berners-Lee nennt diese Objekte Ressourcen (engl. *Resources*). Sie können alles repräsentieren, das mit einer URI identifiziert werden kann. Dies können zum Beispiel digitale Objekte, Bilder oder Informationen sein. Eine Ressource muss dabei aber nicht direkt im Web aufrufbar sein, da es ebenso Dinge in der realen Welt sein können, wie Personen oder auch Konzepte, die über die URI definiert werden.[421]

In der zweiten Regel wird die Verwendung von URIs weiter spezifiziert. Es sollen bevorzugt HTTP URIs verwendet werden. Dies ist eine Subkategorie von URIs und bezeichnet meist eine Webseitenadresse. Diese Form von URIs wird auch URL, also „Uniform Resource Loacator" genannt. HTTP URIs sollen in diesem Zusammenhang jedoch als Namen und nicht als Adressen verstanden werden.[422] Das HTTP Protokoll (Hypertext Transfer Protocol)[423] ist der allgemeine Aufruf- bzw. Zugriffsmechanismus des World Wide Web.[424] Das Protokoll macht es für Webbrowser möglich, Anfragen zu einer URI zu beantworten und die Nutzer*innen zu einem Dokument mit Informationen zu dieser Ressource zu führen. Dieser Vorgang wird auch Dereferenzierung genannt. HTTP URIs sind folglich dereferenzierbar, wenn die Anfrage des Browsers zu einem Dokument mit Informationen über die definierte Ressource führt.[425] Die HTTP URI zu dem Terminus „Fingerring" im Vokabular des Deutschen Archäologischen Instituts beispielsweise ist <http://archwort.dainst.org/de/term/1789>. Der Webbrowser führt zu dieser Webseitenadresse, an der Informationen zu dieser URI hinterlegt sind. Weitere Beispiele und die Anleitung zur Erstellung von HTTP URIs sind in Kapitel 4.1 näher ausgeführt.

Bei der dritten Regel wird darauf hingewiesen, URIs sollen auch sinnvolle Informationen für Benutzer*innen liefern, wenn sie aufgerufen werden. URIs ergeben nicht unbedingt Sinn nur aufgrund ihres Aufbaus und können so von menschlichen Benutzer*innen häufig nicht sofort eingeordnet werden. Die Informationen hinter URIs sollen für Ontologien ebenso zur Verfügung gestellt werden, wie für die Daten selbst. URIs für Ontologien geben Informationen und Definitionen über die Eigenschaften und Klassen, die in der Ontologie verwendet werden.[426] T. Berners-Lee weist ebenfalls darauf hin, es solle nur ein Datenmodell, das RDF Graphenmodell[427], als Stan-

[421] Berners-Lee u. a. 2005, 5.
[422] Berners-Lee 2006, 2.
[423] Belshe u. a. 2015.
[424] Auer 2014, 7.
[425] Isaksen 2011, 24.
[426] Berners-Lee 2006, 2.
[427] Siehe auch Lassila – Swick 1999 Kapitel 2.1 zur Basis-Syntax von RDF.

dard für strukturierte Daten verwendet werden. Die Abfragen der Daten sollen in der Abfragesprache SPARQL[428] ermöglicht werden. Näheres zu RDF und SPARQL wird in den Kapiteln 4.3 und 4.4 erläutert.

Die vierte Regel, die besagt, Links zu anderen Ressourcen zu bilden, ermöglicht es erst ein Web of Data zu erstellen. Das Web of Documents konnte ebenfalls wachsen durch die Absichten aller, möglichst viele Links zu anderen Informationen (z. B. Webseiten) zur Verfügung zu stellen. Durch vermehrte Verlinkungen steigt auch der Wert der eigenen Inhalte, da sie leichter gefunden werden können. Dieses Prinzip kann ebenso auf das Semantic Web angewandt werden.[429] Hier werden nun aber nicht nur Dokumente, sondern alle Ressourcen und Dinge miteinander verlinkt und die Beziehung zwischen diesen Dingen beschrieben. Wenn RDF-Links unterschiedliche URIs miteinander verbinden, dann verlinken sie im Endeffekt verschiedene Ressourcen in anderen Datensets.[430]

3.2.2 Vernetzte offene Daten - Linked Open Data

Alle, die Daten ins Netz bringen, können entscheiden, ob diese Daten öffentlich einsehbar sind oder nicht. Daten, die offen zugänglich sind, werden auch Offene Daten bzw. Open Data genannt. Open Data Lizenzen regeln, in welcher Form die Daten von anderen Nutzer*innen verwendet werden können.[431] Die Lizenzen gehen von keinen Einschränkungen bei einer CC-0 (CC steht für Creative Commons) Lizenz bis zur Urhebernennung bei der CC-BY Lizenz. Die CC-BY Lizenz hat noch sechs weitere Spezialisierungen.[432]

Linked Data kann einen Umschwung der Datennutzung hervorbringen. Die Revolution, wie Dokumente über das Web verbreitet und genutzt werden, soll in ähnlicher Form auch mit verlinkten Daten möglich sein. C. Bizer, T. Heath und T. Bernes-Lee sind der Meinung, mit Linked Data könne sowohl das Potential aus Daten im Netz als auch das Potential des Webs besser ausgeschöpft werden.[433]

Der Begriff Linked Open Data kombiniert nun diese Konzepte. Er bezeichnet Linked Data, die unter einer offenen Lizenz veröffentlicht werden. T. Berners-Lee erweiterte 2010 seinen Design

[428] Siehe auch Harris – Seaborne 2013 als Überblick über SPARQL.
[429] Berners-Lee 2006, 2 f.
[430] Heath – Bizer 2011, 15 f.
[431] Gerth 2017, 4.
[432] Siehe https://creativecommons.org/about/cclicenses/ für Creative Common Licences, abgerufen am 21.10.2021.
[433] Bizer u. a. 2009, 18.

Issues-Beitrag von 2006 mit Informationen zu Linked Open Data. Er erstellte ein 5-Sterne Modell, um den Grad der Offenheit von Daten bewerten zu können (siehe Abbildung 11):[434]

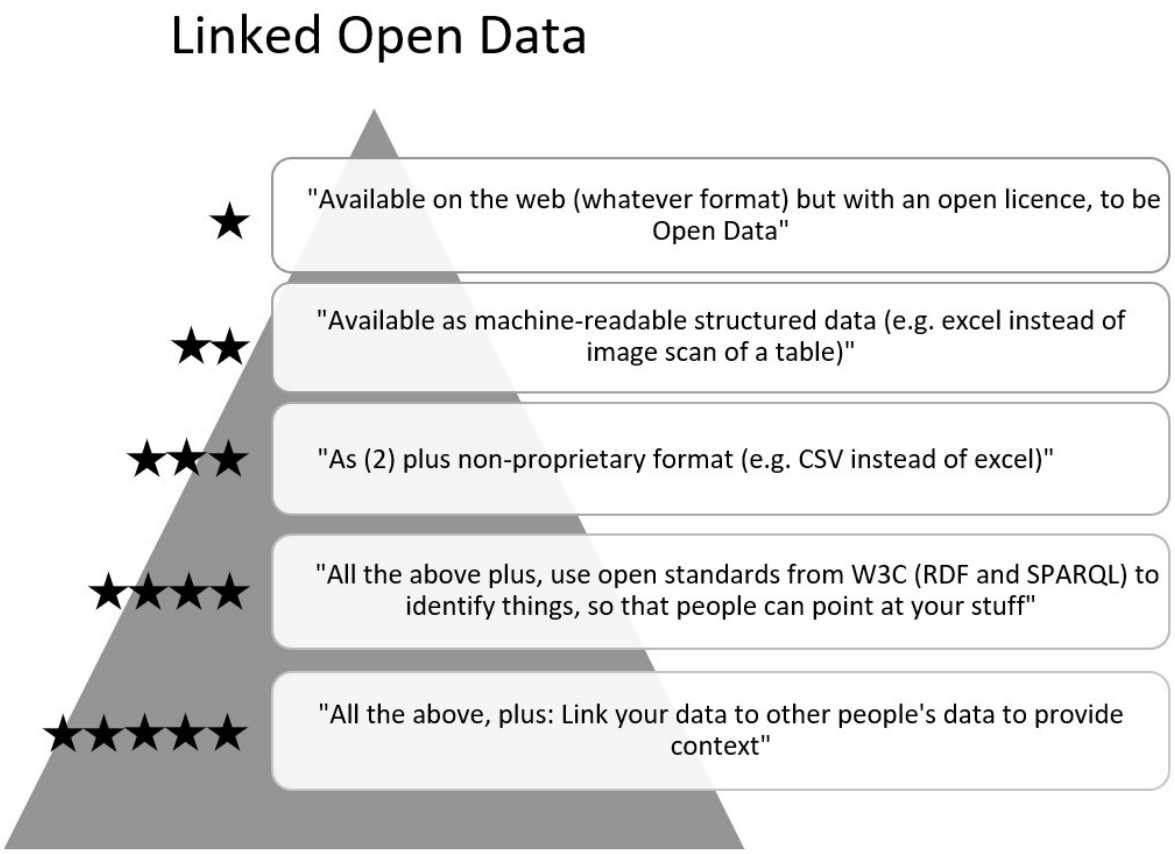


Abbildung 11 - Linked Open Data Sterne Schema (Quelle: Grafik M. Simon nach Berners-Lee 2006, 6-7).

Damit die Daten zumindest einen Stern im Linked Open Data Sterne-Schema erhalten und somit als Open Data gelten können, müssen sie unter einer offenen Lizenz veröffentlicht werden. Je mehr Sterne die Daten haben, desto einfacher ist es für andere Personen, diese Daten wiederzuverwenden. Jeder Stern muss die Kriterien aller Sterne zuvor erfüllen.[435] Jede Stufe im 5-Sterne Modell bringt neue Vorteile für Nutzer*innen und Dateninhaber*innen mit sich (siehe Abbildung 12). Der Aufwand, die Daten in das vorgesehene Format zu bringen, kann sich naturgemäß steigern, abhängig je nach der Ausgangslage der Daten.

Ein Stern erfordert zumindest eine offene Datenlizenz.[436] Das Dateiformat ist hier nicht relevant. Dies können zum Beispiel Bilder oder Scans von Tabellen sein. Die Daten können als PDF- oder Bilddateien veröffentlicht werden. Vorteile entstehen für Nutzer*innen, weil die Daten angesehen, ausgedruckt, gespeichert, weiterverwendet und in andere Formate gebracht werden können.

[434] Berners-Lee 2006, 6 f.
[435] Berners-Lee 2006, 7.
[436] Für Vorteile der 5-Star Daten siehe: https://5stardata.info/de/, abgerufen am 21.10.2021.

Dateninhaber*innen profitieren, da die Veröffentlichung einfach ist. Um beispielsweise an die Daten, die in einer Tabelle zu finden sind, zu kommen, müssen Nutzer*innen die Daten allerdings händisch extrahieren.

Zwei Sterne besagen, das Datenformat müsse strukturierte Daten aufweisen. Dies kann ein proprietäres Format, wie zum Beispiel eine Excel-Tabelle sein. Nutzer*innen können mit proprietärer Software diese Daten weiterbearbeiten, wiederverwenden oder in ein anderes Format übertragen. Für die Dateninhaber*innen ist es kein großer Aufwand, die Daten zu veröffentlichen.

Bei drei Sternen werden strukturierte Daten mit nicht-proprietären Formaten, wie zum Beispiel CSV statt Excel, veröffentlicht. Für Dateninhaber*innen bringt dies den Aufwand, die Daten aus dem proprietären Format zu exportieren.

Vier Sterne erfordern, die Daten im RDF-Format zu veröffentlichen. Den Ressourcen bzw. Datensätzen sollen eindeutige URIs zugewiesen werden, so dass auf diese Daten verwiesen werden kann. Die Daten können von anderen Nutzer*innen nun verlinkt werden. Auf die Daten der Dateninhaber*innen kann so verwiesen werden und damit erhöht sich die Datenpräsenz.

Für fünf Sterne müssen die eigenen Daten bereits mit anderen Daten verlinkt sein. Dadurch können Nutzer*innen neue Daten entdecken oder ein neues Datenschema finden, welches sie weiterverwenden könnten. Dateninhaber*innen machen ihre Daten besser auffindbar im Netz, müssen jedoch ihre Daten mit anderen Datensätzen via URIs verlinken.

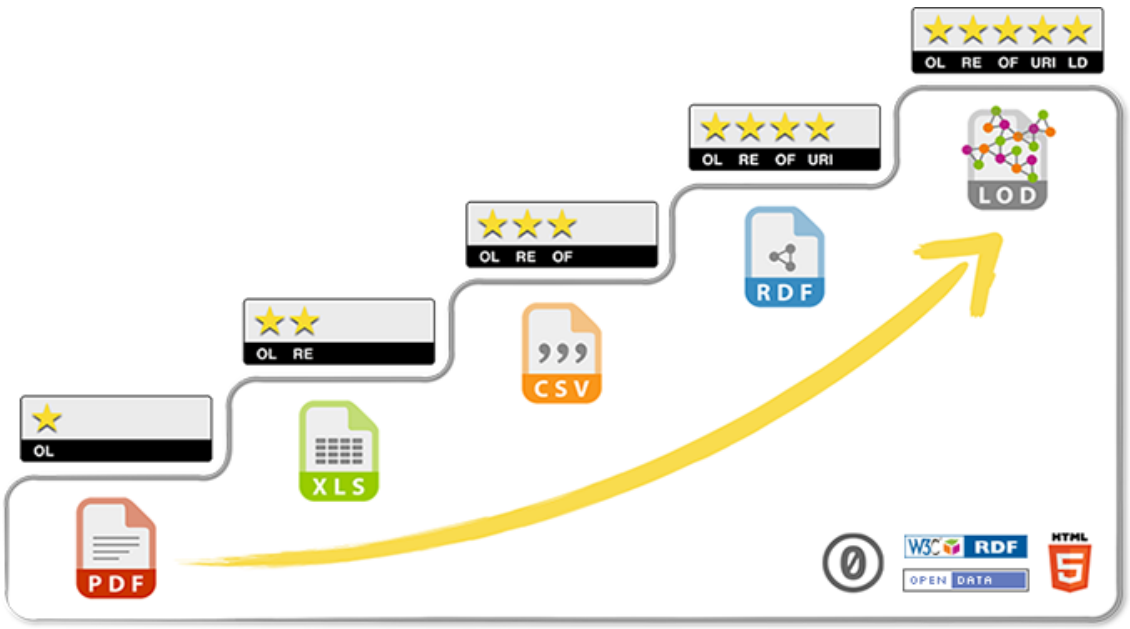


Abbildung 12 - 5 Star Linked Open Data (Quelle: https://5stardata.info/en/, abgerufen am 21.10.2021).

3.2.3 Weitere Entwicklung von Linked Open Data

2007 ist die Linked Open Data Cloud[437] durch das „Linking Open Data“ Community Projekt[438] der W3C Arbeitsgruppe „Semantic Web Education and Outreach Group“ entstanden, in der öffentlich zugängliche Datensätze herausgegeben und miteinander verlinkt werden.[439] Im Mai 2020 zählt die Cloud 1.255 Datensets mit 16.174 Links. Die Inhaber*innen kommen aus verschiedenen Disziplinen, unter anderem ist hier die Collection des British Museum und einige andere Datensets aus dem Bereich der Kulturgüter inkludiert.[440] Die Linked Open Data Cloud hat den Semantic Layer Cake als Symbol für das Semantic Web seit einiger Zeit abgelöst und wird oft auch als Bemessungsgrundlage verwendet, wie viele Nutzer*innen das Semantic Web verwenden.[441] Abbildung 13 und Abbildung 14 zeigen die Entwicklung der Datensets von 2007 bis 2020 und veranschaulichen das enorme Wachstum der neu hinzugekommenen Datensets. Jeder Kreis verdeutlicht ein Datenset, welches Teil der Linked Open Data Cloud ist. Abbildung 14 zeigt die Datensets und ihre Verlinkungen an. Die Farben visualisieren die Zuordnung zu einer Domäne.[442] Die Benutzer*innen der LOD-Cloud können darin navigieren und jedes Datenset auswählen, Informationen darüber finden und downloaden oder über einen SPARQL-Endpoint (siehe Kapitel 4.4) durchsuchen.

[437] Siehe auch https://lod-cloud.net/, abgerufen am 21.10.2021.
[438] Siehe unter https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData, abgerufen am 23.09.2021.
[439] Bizer u. a. 2009, 4–6.
[440] Siehe auch https://lod-cloud.net/datasets, um die Datensets zu durchsuchen, abgerufen am 21.10.2021.
[441] Isaksen 2011, 29.
[442] Als Erläuterung der verschiedenen Farben der Legende in Abbildung 14 werden sie hier beschrieben. Braun: Interdisziplinäres Gebiet; Türkis: Geographie; Orange: Regierungsdaten; Rot: Biowissenschaften; Grün: Linguistik; Blau: Medien; Hellbraun: Publikationen; Grau: Soziale Netzwerke; Rosa: benutzergenerierte Daten.

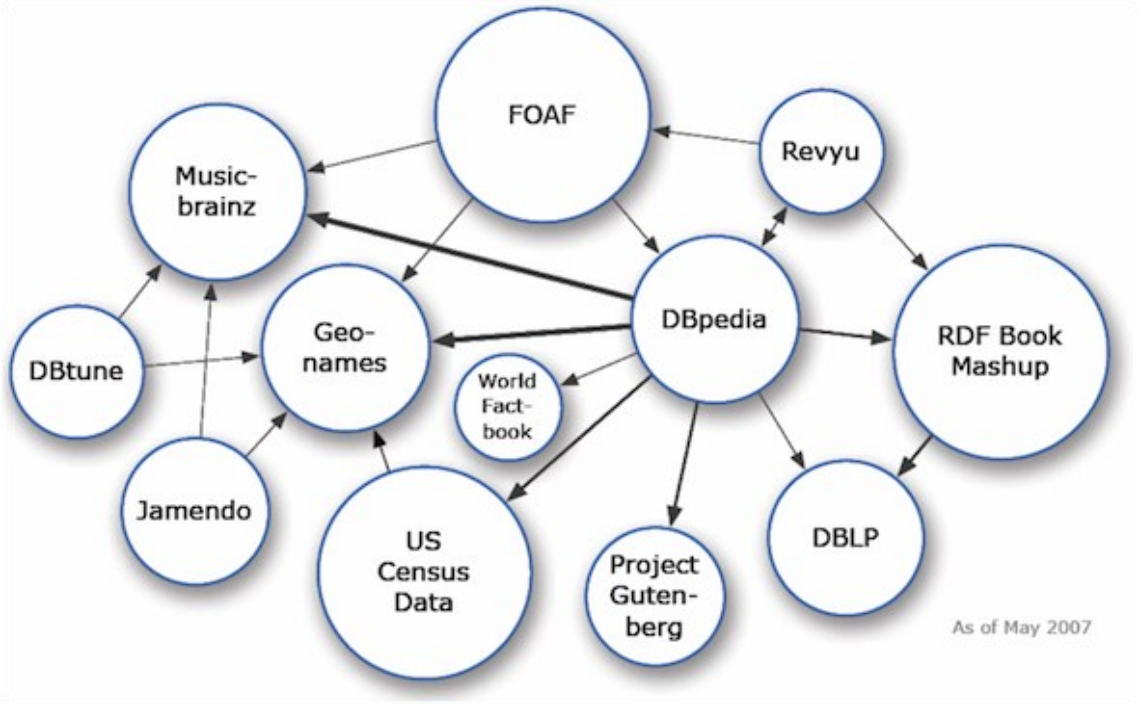


Abbildung 13 - Die Linked Open Data Cloud 2007 (Quelle: https://lod-cloud.net/versions/2007-05-01/lod-cloud.png, abgerufen am 21.10.2021).

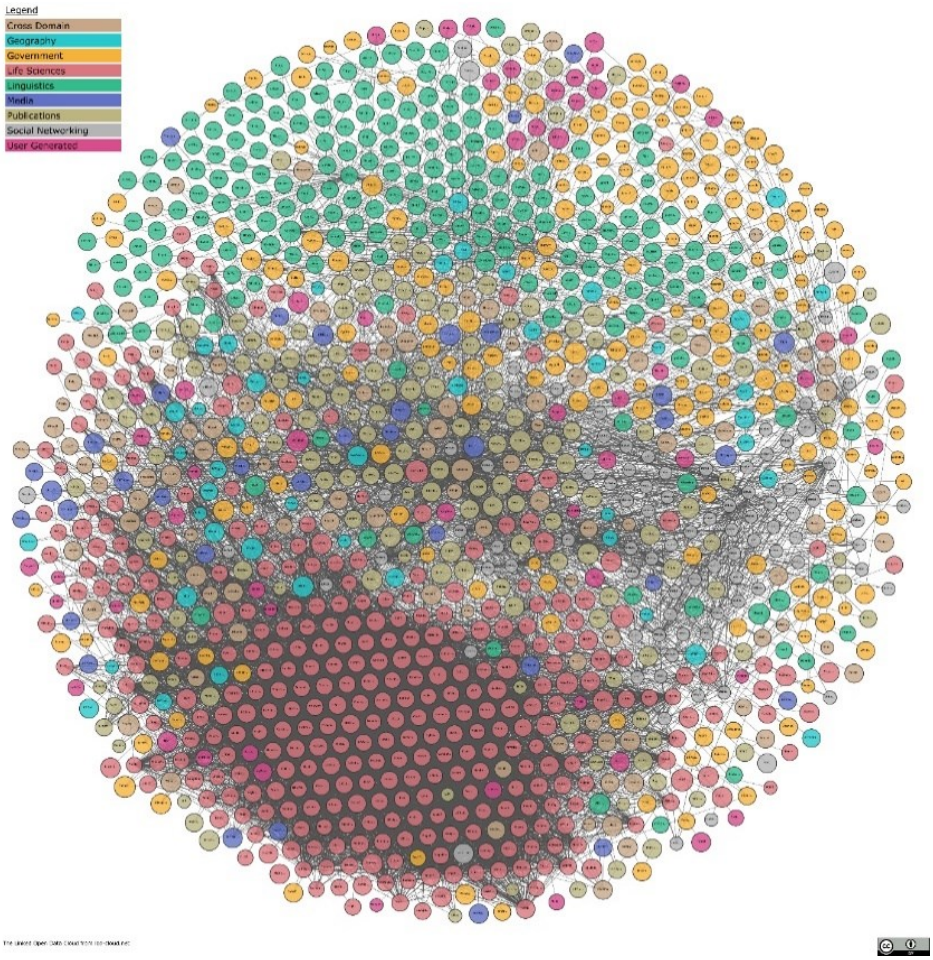


Abbildung 14 - Die Linked Open Data Cloud 2020 (Quelle: https://lod-cloud.net/versions/2020-05-20/lod-cloud.png, abgerufen am 21.10.2021).

Ein Teil der LOD-Cloud ist DBpedia[443]. Es wurde 2007 im Zuge des DBpedia Projekts gegründet und ermöglicht es, strukturierte Informationen aus Wikipedia zu extrahieren. Diese Informationen können abgefragt und mit anderen Daten verlinkt werden. DBpedia bietet so URIs für Millionen von Konzepten und Dingen, welche von Datenanbieter*innen zu ihren Daten durch RDF-Links verlinkt werden können.[444] Mittlerweile fungiert DBpedia als einer der zentralen Knotenpunkte im Web of Data.[445]

Ein weiteres wichtiges Projekt stellt das LOD2 – Creating Knowledge out of Interlinked Data Projekt[446] dar, welches von 2010 bis 2014 aktiv war. Darin wurden Werkzeuge, Software und Methoden entwickelt, um Linked Open Data Anwender*innen zu unterstützen.[447] Das DIACHRON – Managing the Evolution and Preservation of the Data Web Projekt[448] hat an diese Technologien angeknüpft und weitere Entwicklungen hinsichtlich der Datenqualität, der Zitation von bestimmten verlinkten Daten und der Archivierung vorangetrieben.[449]

Währenddessen werden Bücher zu verlinkten Daten und Anleitungen für Anwender*innen und Neueinsteiger*innen herausgegeben, wie beispielsweise „Linked Data: Evolving the Web into a Global Data Space“[450] von T. Heath und C. Bizer oder „Linked Open Data: The Essentials“[451] von F. Bauer und M. Kaltenböck. Aber auch Workshops und Konferenzen zum Thema Linked Open Data und Semantic Web werden laufend weitergeführt. Die International Semantic Web Conference (ISWC)[452] findet im Jahr 2021 zum 20. Mal und die Extended Semantic Web Conference (ESWC)[453] zum 18. Mal statt.

Viele Herausgeber*innen von Daten wenden die Linked Open Data Kriterien bereits an.[454] Diese finden sich in vielen Disziplinen und Branchen, wie beispielsweise Bibliotheken[455], der Biomedizin, kulturhistorischen Einrichtungen oder Regierungen und öffentlichen Ämtern.[456] 2011 wurde das „Linked Data Cookbook – Cookbook for Open Government Linked Data“[457] herausgegeben,

[443] Siehe unter https://www.dbpedia.org/, abgerufen am 23.09.2021.
[444] Auer u. a. 2007, 722–724.
[445] Lehmann u. a. 2015, 167.
[446] Siehe unter https://cordis.europa.eu/project/id/257943, abgerufen am 23.09.2021.
[447] Auer 2014, 1–3.
[448] Das DIACHRON Projekt lief von 2013 bis 2016. Siehe unter https://cordis.europa.eu/project/id/601043, abgerufen am 23.09.2021.
[449] Hasapis u. a. 2015, 1 f.; Meimaris u. a. 2015, 16.
[450] Heath – Bizer 2011.
[451] Bauer – Kaltenböck 2012; und ihre zweite Auflage: Bauer – Kaltenböck 2016.
[452] Siehe unter https://iswc2021.semanticweb.org/, abgerufen am 23.09.2021.
[453] Siehe unter https://2021.eswc-conferences.org/, abgerufen am 23.09.2021.
[454] Auer 2014, 1 f.
[455] Hannemann – Kett 2010, 1.
[456] Bauer – Kaltenböck 2016, 35 f.
[457] Hyland – Villazón-Terrazas 2011.

um öffentlichen Stellen die Veröffentlichung von Regierungsdaten zu erleichtern. Daten von europäischen Institutionen werden im EU Open Data Portal[458] veröffentlicht. In Österreich wurde 2011 die Kooperationsvereinbarung Cooperation Open Government Data Österreich (Cooperation OGD Österreich) zwischen Bund, Städten, Ländern und Gemeinden gegründet und Bestrebungen für offene Regierungsdaten diskutiert. Als Folge wurde 2012 der nationale Datenkatalog unter data.gv.at online gestellt.[459] Als Ergänzung für die Regierungsdaten wurde 2014 das Open Data Portal (ODP) für Daten aus Wirtschaft, Kultur, NGOs, Forschung und Zivilgesellschaft eröffnet.[460] Im gleichen Jahr wurde auf Basis dieser Daten, also dem nationalen Datenkatalog und dem Open Data Portal, sowie den Daten der Wiener Stadtregierung[461] das Linked Open Data Pilotprojekt Österreich (LOD PILOT AT)[462] gestartet. Unter <http://www.linkeddata.gv.at> und <http://www.lodpilot.at/> soll eine frei verfügbare Linked Open Data Infrastruktur für Österreich entstehen.

[458] Siehe unter https://data.europa.eu/de, abgerufen am 23.09.2021.
[459] Eibl u. a. 2019.
[460] Siehe unter https://www.opendataportal.at/, abgerufen am 23.09.2021.
[461] Siehe unter data.wien.gv.at oder https://www.data.gv.at/auftritte/?organisation=stadt-wien, abgerufen am 23.09.2021.
[462] Kaltenböck 2013.

4 Technische Grundlagen – Technologien für das Semantic Web und Linked (Open) Data

Wie bereits im vorhergehenden Kapitel beschrieben, verwenden Linked Data die wichtigsten Technologien des Semantic Web (siehe Abbildung 15)[463], um zum Ziel des Web of Data zu gelangen.[464] Mit diesen Werkzeugen, die in diesem Kapitel im Detail erklärt werden, können Daten in Linked Data umgebaut werden.

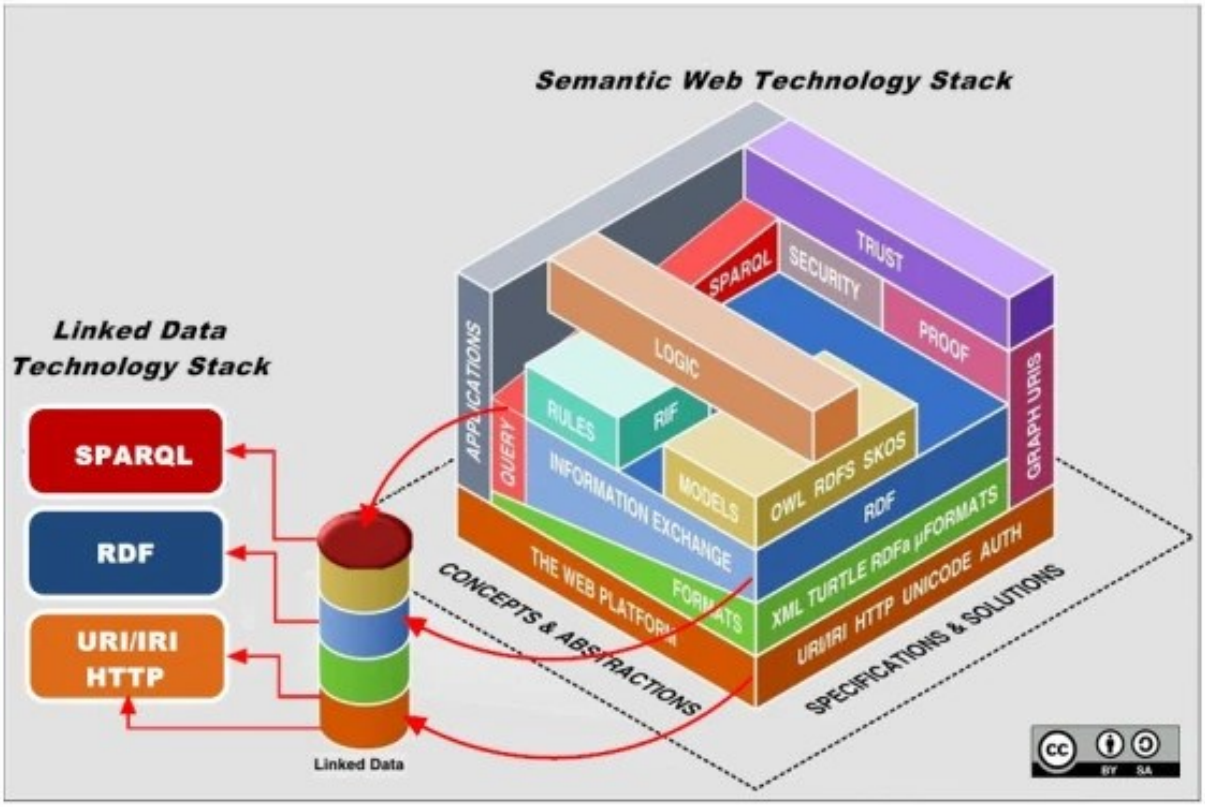


Abbildung 15 - Semantic Web Technologien und Linked Data Technologien (Quelle: Parapontis 2012, adaptiert nach Nowack 2009).

4.1 URI – Uniform Resource Identifier

Das erste Prinzip der Linked Data besagt, URIs zu benutzen, um Objekte, Dinge, Konzepte und Ressourcen zu identifizieren. Dies sind nicht nur Objekte, die online existieren, sondern auch Dinge oder Konzepte aus der realen Welt.[465]

[463] Quelle: Parapontis 2012; adaptiert nach Nowack 2009.
[464] Machado u. a. 2019, 710.
[465] Heath – Bizer 2011, 15.

URIs sind Uniform Resource Identifier und können Web-Dokumente, Objekte aus der realen Welt und abstrakte Konzepte beschreiben.[466] Sie bestehen aus einer Folge von Zeichen, die eine Ressource, ob abstrakt oder physisch, identifiziert.[467]

4.1.1 Aufbau von URIs

Ein URI setzt sich aus einem Schema-Namen, wie „http" bzw. „https" oder „urn", und einem schema-spezifischen Teil zusammen. HTTP-URIs werden aus einer hierarchischen Folge von Komponenten gebildet:[468]

[Schema:] [//Autorität] [Pfad] [?Abfrage] [#Fragment]

Die Komponenten des Schemas und des Pfades müssen immer angegeben werden.[469] Es gibt mehrere mögliche Schemata, wobei die bekanntesten sind URN, DOI oder HTTP sind. Bei Linked Data sollen als Best Practice immer URIs mit dem HTTP-Schema verwendet werden, also HTTP-URIs, um sie wie bei einer URL zu verwenden, um die Ressourcen aufzufinden. Andere Schemata, wie URNs oder DOIs, sollten nicht verwendet werden, da dies dann nicht möglich wäre.[470] Dies stellt das zweite Prinzip der Linked Data dar.[471] Ein Beispiel für die URI des Ortes Oberleis in Niederösterreich als HTTP-URI wäre:

<https://www.geonames.org/2769960/oberleis.html>

[https:][//www.geonames.org][/2769960/oberleis.html]

[Schema = **https:**] [Autorität = **//www.geonames.org**] [Pfad = **/276996/oberleis.html**]

In diesem Beispiel steht „https:" für das Schema der URI, die Autorität ist „//www.geonames.org" und der Pfad wird als „/2769960/oberleis.html" angegeben.

Mit URIs können weltweit beliebig viele einzigartige Namen gebildet werden. Jede*r Besitzer*in eines Domainnamens kann neue Identifier kreieren. Die URIs werden allerdings nicht nur als Namen für Ressourcen verwendet, sondern sollen Nutzer*innen auch Informationen über die beschriebenen Einheiten bieten, wenn sie aufgerufen werden.[472] Das ist die Basis der dritten Empfehlung für Linked Data.[473]

[466] Auer 2014, 7.
[467] Berners-Lee u. a. 2005, 5.
[468] Berners-Lee u. a. 2005, 16.
[469] Berners-Lee u. a. 2005, 16.
[470] Auer 2014, 7.
[471] Berners-Lee 2006, 1.
[472] Auer 2014, 7.
[473] Berners-Lee 2006, 1.

URIs sind unabhängig vom Datensystem, in dem sie gespeichert sind, und können in mehreren Datensystemen verwendet und verlinkt werden.[474] An dieser Stelle soll noch einmal auf die „Non-unique Naming Assumption“ (siehe Kapitel 3.1.2) hingewiesen werden. Da sich die Autor*innen von URIs nicht absprechen, muss angenommen werden, dass gleiche Dinge unterschiedliche oder mehrere Namen bzw. URIs haben können. Sollten diese beiden Namen gleichzeitig auftauchen, weiß die Webinfrastruktur nicht, dass es sich um die gleiche Ressource handelt. Andersherum gedacht, kann ebenfalls nicht angenommen werden, es handle sich um unterschiedliche Ressourcen, wenn sie unterschiedliche URIs haben.[475] Der Ort Oberleis beispielsweise kann durch mehrere URIs identifiziert werden. Auf Geonames ist die URI wie im obigen Beispiel <https://www.geonames.org/2769960/oberleis.html>. Auf der Nominatim OpenStreetMap wird Oberleis unter dieser URI angegeben: <https://nominatim.openstreetmap.org/ui/details.html?osmtype=N&osmid=1583733718&class=place>. Beide URIs bezeichnen jedoch den gleichen Ort.

4.1.2 Dereferenzieren von URIs

Jede HTTP-URI sollte „dereferenzierbar“ sein. Dies ist der Fall, wenn HTTP-Clients (zum Beispiel Webbrowser) diese über das HTTP-Protokoll abrufen und Informationen über die Ressource, die von dem URI identifiziert ist, finden können. Wenn ein im obigen Beispiel genannter URI des Ortes Oberleis in einen Webbrowser eingegeben wird, gelangt der oder die Nutzer*in zu einer Seite, auf der Informationen wie das Land, das Bundesland und die Koordinaten des Ortes hinterlegt sind. Dies gilt sowohl für HTML-Dokumente als auch für URIs, die zur Beschreibung von realen Dingen und Konzepten fungieren. Ressourcen werden durch Webdokumente beschrieben. Die Beschreibungen können in HTML-Form (menschenlesbar) oder RDF-Form (maschinenlesbar) geschehen.[476]

Ein wichtiger Faktor des Web of Data ist, Verwechslungen zwischen den Beschreibungen der realen Dinge und den Dokumenten über diese Dinge zu vermeiden. Hierfür gibt es unterschiedliche URIs. So ist es möglich, verschiedene Aussagen über Dokumente, die zur Beschreibung dienen, und über die Objekte selbst zu tätigen.[477] Der Zeitpunkt der Fertigung eines archäologischen Fundes ist ein anderer als der der Fertigung eines Artikels über diesen Fund. Darum wurden zwei Strategien entwickelt, um reale Dinge identifizieren zu können.[478] Diese Vorgehensweise wird meist bei selbst erzeugten Daten gewählt, wenn Dateninhaber*innen die Möglichkeit haben,

[474] Isaksen 2011, 24.
[475] Allemang – Hendler 2012, 9.
[476] Heath – Bizer 2011, 18.
[477] Auer 2014, 8.
[478] Sauermann – Cyganiak 2008, Kapitel 4.

die URIs selbst zu erstellen. Sie finden sich oft nicht bei URIs von Seiten wie Geonames oder anderen URI-Quellenseiten (zu den verschiedenen URI-Quellen siehe Kapitel 6). Ein reales Objekt hat demnach nicht automatisch eine andere URI. Die Strategien sind die Hash Strategie mit Hash URIs und die 303 Strategie mit 303 URIs und werden im Folgenden erläutert.

4.1.2.1 Die Hash Strategie

Bei dieser Strategie werden den URIs Fragmente hintangestellt, die vom restlichen Teil des URI durch ein #-Symbol getrennt werden. Wenn der Client diesen URI abrufen möchte, zwingt ihn das HTTP-Protokoll, den letzten Teil abzutrennen, bevor der Client sie erreichen kann. Da der URI mit dem Hashsymbol nicht direkt abrufbar ist, kann es sich nicht um einen URI eines Webdokuments handeln, sondern ist der URI eines realen Objektes. Aus diesem Grund können sie für reale Dinge oder Konzepte verwendet werden. Diese Strategie ist eine Möglichkeit, URIs für reale Objekte zu generieren, aber nicht alle realen Objekte haben ein #-Symbol in dem URI.

Das Beispiel <http://www.fundkatalog.at/Fundliste#1234> (siehe Abbildung 16 oben) würde einen Fund mit der Fundnummer 1234 innerhalb einer Fundliste mit einem Hash-URI identifizieren. Das Beispiel <http://www.fundkatalog.at/Fundliste/1234> (siehe Abbildung 16 unten) würde den Fund mit der Fundnummer 1234 mit dem herkömmlichen HTTP-URI identifizieren.[479]

Abbildung 16 – Unterschied zwischen einer Hash-URI und einer HTTP-URI.

4.1.2.2 Die 303 Strategie

Die 303 Strategie verwendet den HTTP Statuscode „303 see other“[480]. Dieser dient bei der Abfrage des URI als Hinweis, dass es sich bei dem URI um einen Identifikator eines realen Objektes und nicht um das eines Webdokuments handelt. Anstatt das reale Objekt über das Netz zu senden, denn dies wäre unmöglich, antwortet der Server dem Client mit dem 303 Code und einer

[479] Sauermann – Cyganiak 2008, Kapitel 4.1.
[480] Fielding – Reschke 2014, Kapitel 6.4.4.

URI, die das reale Objekt beschreibt. Dies wird „303 redirect“ genannt. Anschließend kann der Client die erhaltene URI dereferenzieren und erhält als Antwort ein Webdokument, welches das reale Objekt beschreibt.[481] Der Beispiel-URI des Fundobjektes mit der Fundnummer 1234 <http://fundkatalog.at/Fundliste/1234> würde Nutzer*innen zu einer Seite mit Informationen über den Fund führen, wenn diese ihn in die Adresszeile des Webbrowsers eingeben.

Jede Strategie hat Vor- und Nachteile. Die Hash Strategie vermindert die HTTP An- und Rückfragen. Dies verkürzt die Zugriffszeit. Mehrere Hash-URIs können den gleichen Teil eines URI ohne den Hash-Teil haben und daher gleichzeitig abgerufen werden. Wenn allerdings nur eine Ressource erwünscht ist, werden trotzdem alle Daten mitgeladen, da sie sich in derselben Datei befinden. Im oben genannten Beispiel würden alle Funde in der Fundliste geladen werden. Dies kann zu großen Datenmengen führen. Die 303 Strategie ist flexibler, da die Rückführung individuell eingestellt werden kann. Es kann ein beschreibendes Dokument für jede Ressource, für alle Ressourcen oder für eine Mischung daraus abgerufen werden. Bei der Nutzung von Ontologien kann es aber zu großen Netzwerkbelastungen kommen.[482]

Für die Ontologie CIDOC-CRM wird die 303 Strategie angewandt. Jede Einheit in einem Datenset führt zu mindestens drei URIs. Ein URI, um das reale Objekt oder Konzept zu identifizieren, ein URI, der zu einem beschreibenden, von Menschen lesbaren HTML-Dokument führt und ein URI, der zu einem beschreibenden, von Maschinen lesbaren RDF/XML-Dokument führt. Je nach Client-Anforderung wird die HTML- oder RDF-Repräsentation übermittelt.[483]

4.2 Namensräume

An dieser Stelle soll auf die Namensräume hingewiesen werden, da sie sehr oft verwendet werden. Namensräume, oder auch Präfixe bzw. Namespaces genannt, sind Abkürzungen von URIs. Sie werden verwendet, um Ergebnisse übersichtlicher zu gestalten und das Datenvolumen zu reduzieren. URIs können oftmals sehr lange Zeichenketten bilden. Daher sind die Abkürzungen bei der Beschreibung von Daten nützlich, da sich in einem Dokument oftmals auf denselben URI bezogen wird.[484] Die Namensräume sind nicht global definiert, daher müssen sie immer am Beginn des Dokuments angegeben werden.[485] Einige Beispiele von Namensräumen, welche öfters verwendet werden, sind:

- **crm:** http://www.cidoc-crm.org/cidoc-crm/

[481] Sauermann – Cyganiak 2008, Kapitel 4.2.
[482] Sauermann – Cyganiak 2008, Kapitel 4.4.
[483] Georgis 2018.
[484] Gerth 2017, 10 f.
[485] Isaksen 2011, 24.

- **skos:** http://www.w3.org/2004/02/skos/core#
- **dcterms:** https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#
- **aat:** http://vocab.getty.edu/aat/
- **iDAIvocab:** https://archwort.dainst.org/de/term/

Der URI setzt sich dann zusammen aus Präfix und Suffix, getrennt von einem Doppelpunkt. Die vollständige URI der Abkürzung dcterms:creator würde <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#creator> lauten. Hier steht „dcterms" für den Präfix (https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#) und „creator" für den Suffix. Da in einem Dokument zu Beginn immer die Namensräume angegeben werden, kann der Nutzer oder die Nutzerin nachvollziehen, welche konkrete Adresse die Abkürzung hat. Die Adresse führt den Nutzer oder die Nutzerin zur Definition des Begriffs „creator".

4.3 RDF Datenmodell – Resource Description Framework

RDF ist ein Graphen-basiertes Datenmodell und wird zur Repräsentation von Linked Data benutzt.[486] Es bietet den Rahmen, um Informationen über Ressourcen beschreiben zu können.[487] L. Isaksen beschreibt es in einer Metapher über den Satzbau:[488] Die URIs als Bausteine des Semantic Web stehen dabei metaphorisch im Satzbau für die Wörter und das RDF als Rahmen für die Bausteine als Sätze.

4.3.1 Triples – Die Bestandteile eines Graphen

Die Kernstruktur des RDF besteht aus Sets von Dreiergruppen (Triples). Diese werden so genannt, weil ein Triple aus drei Elementen besteht. Von der Metapher hergeleitet, würde es bedeuten, ein Satz würde aus drei Wörtern gebildet werden. Das sind jeweils ein Subjekt, ein Prädikat und ein Objekt. Triples können durch Knoten (engl. *node*) und Pfeile (engl. *arc*) visualisiert werden. Dabei stellen die Knoten die Ressourcen und die Pfeile die Beziehung zwischen den Ressourcen dar (siehe Abbildung 17). Eine durch ein Triple gemachte Aussage wird auch RDF Statement

[486] Cyganiak u. a. 2014, Kapitel 1.1.
[487] Schreiber – Raimond 2014, Kapitel 1.
[488] Isaksen 2011, 24.

genannt. Ein Set von mehreren Triples nennt man RDF Graph. Eine Kollektion von RDF Graphen ist ein RDF Datenset.[489]

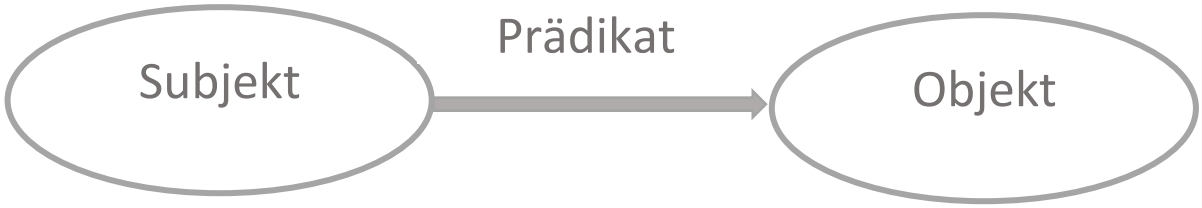


Abbildung 17 – Visualisierung eines RDF Triples (Quelle: M. Simon nach Cyganiak u. a. 2014, Kapitel 1.1).

Es können drei Arten von Knoten in einem RDF Graphen vorkommen:[490]

- URIs/IRIs[491],
- Literale (Zeichenfolgen, Nummern, Datumsangaben)
- und leere Knoten (Knoten ohne URIs zur Überbrückung).

Das Prädikat ist ebenfalls ein URI und stellt eine Eigenschaft (engl. *property*) dar.[492] Das bedeutet, dass sich die drei Komponenten eines RDF Triples folgendermaßen zusammensetzen können:

- Subjekt: URI oder leerer Knoten
- Prädikat: URI (von einem Vokabular)
- Objekt: URI, Literal oder leerer Knoten

Im folgenden Beispiel (siehe Abbildung 18) wird eine Aussage über ein Fundobjekt getätigt. Das Fundobjekt im ersten Knoten bildet das Subjekt des Satzes. Es besteht aus einem Material. Dies wird durch den Pfeil dargestellt. Das Objekt des Triples ist der Knoten auf der rechten Seite, welcher das Material beschreibt.

[489] Cyganiak u. a. 2014, Kapitel 1.1.
[490] Cyganiak u. a. 2014, Kapitel 1.1.
[491] IRIs sind Internationalized Resource Identifier und stellen eine Generalisierung von URIs dar. Siehe dazu Cyganiak u. a. 2014, Kapitel 3.2.
[492] Cyganiak u. a. 2014, Kapitel 1.2.

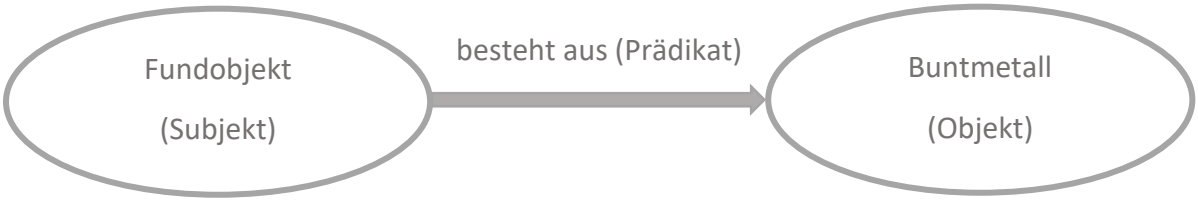


Abbildung 18 – Beispiel eines Triples, welches das Material eines Fundobjektes beschreibt (Grafik: M. Simon).

Das Subjekt ist immer ein URI oder ein leerer Knoten, das Prädikat ist immer ein URI und das Objekt ist ein URI, ein Literal oder ein leerer Knoten.[493] Die URIs von Prädikaten stammen dabei von Vokabularen, welche dazu dienen, die Beziehung zwischen Subjekt und Objekt zu beschreiben.[494] Das oben beschriebene Beispiel kann mit den drei URIs folgendermaßen ausgedrückt werden (siehe Abbildung 19):

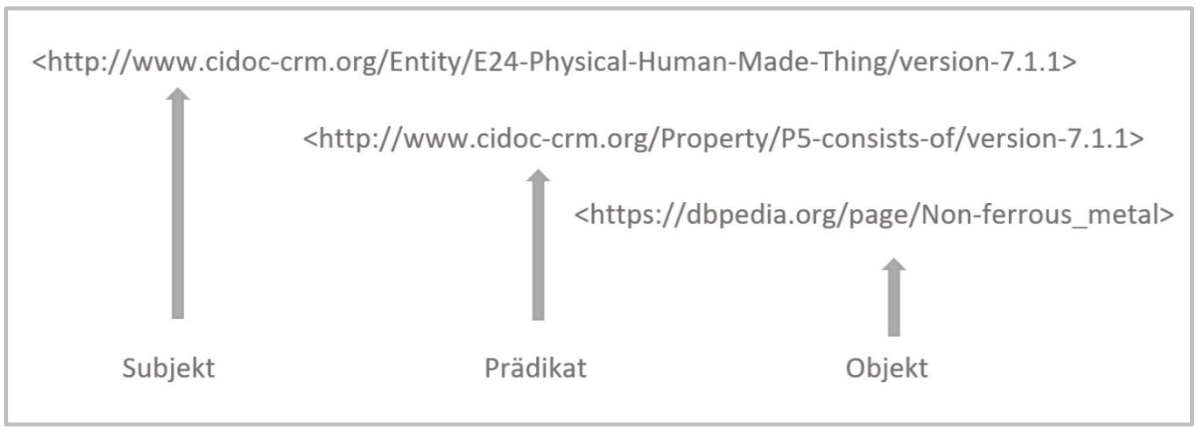


Abbildung 19 – Triple mit URIs beschrieben (Grafik: M. Simon).

Ein Beispiel mit einem Literal als Objekt wird im Folgenden beschrieben (siehe Abbildung 20). Dabei wird dem Fundobjekt eine Fundnummer zugewiesen. Der erste Knoten stellt das Fundobjekt (Subjekt) dar, das Prädikat gibt an, dass es sich um eine Identifikation handelt und das Objekt ist ein Literal mit der Fundnummer des Fundobjektes. Das Triple mit dem Literal der Fundnummer „1234“ als Objekt kann mit URIs ausgedrückt werden und wird in Abbildung 21 dargestellt.

[493] Cyganiak u. a. 2014, Kapitel 3.1.
[494] Heath – Bizer 2011, Kapitel 2.4.1.

Abbildung 20 - Graphische Darstellung eines Triples, welches die Fundnummer eines Fundobjektes beschreibt (Grafik: M. Simon).

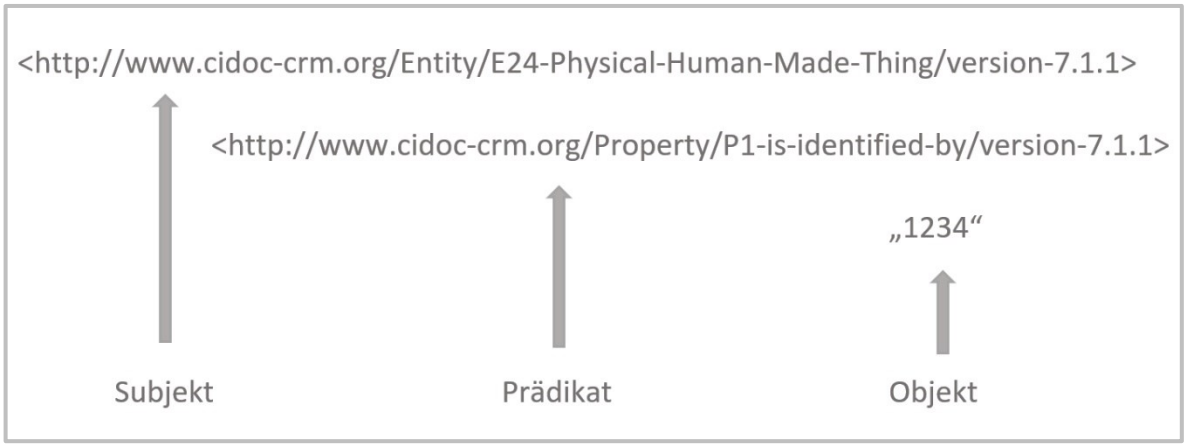


Abbildung 21 – Triple mit einem Literal als Objekt (Grafik: M. Simon).

Die Verwendung von RDF in Kombination mit Vokabularen ist notwendig, um die semantischen Informationen über die Ressourcen ausdrücken zu können. Beispiele für Vokabulare sind RDF Schema, Dublin Core, schema.org, SKOS, OWL oder andere Ontologien wie CIDOC-CRM.[495]

Die beiden Arten von Triples, welche in den zwei oben beschriebenen Beispielen dargestellt wurden, werden Literal Triples und RDF Links genannt:[496]

- *RDF Links*: Sie beschreiben die Beziehung von zwei definierten Ressourcen. Diese Triples haben immer drei URIs. Das Beispiel des Fundobjektes, welches aus einem bestimmten Material besteht, ist ein RDF Link (siehe Abbildung 19). Alle drei Komponenten werden mit einem URI beschrieben.
- *Literal-Triples:* Bei dieser Art von Triples ist das Objekt ein Literal. Es wird verwendet, um Eigenschaften von Ressourcen frei zu beschreiben wie zum Beispiel ein Name, ein Geburtsdatum einer Person, Zahlen oder Messdaten. Das Beispiel, in welchem dem Fundobjekt eine Fundnummer zugeordnet wurde (siehe Abbildung 20), wird als Literal-Triple bezeichnet, da das Objekt ein Literal ist.

[495] Schreiber – Raimond 2014, Kapitel 4.
[496] Heath – Bizer 2011, Kapitel 2.4.1.

4.3.2 RDF Serialisierungen

Zur Darstellung eines RDF Graphen benötigt man eine RDF Syntax, welche die Triples beschreibt.[497] Die Triples werden einer gewählten Syntax folgend in einer Datei ausformuliert, bzw. serialisiert, d.h. aneinandergereiht. Es gibt unterschiedliche Serialisierungs-Formate zum Schreiben von RDF Graphen. Die Bedeutung der Triples bleibt jedoch ident.[498] Die geläufigsten Serialisierungsformen werden im Folgenden anhand eines Beispiels zu einer Nadel (siehe Abbildung 22) näher erläutert.

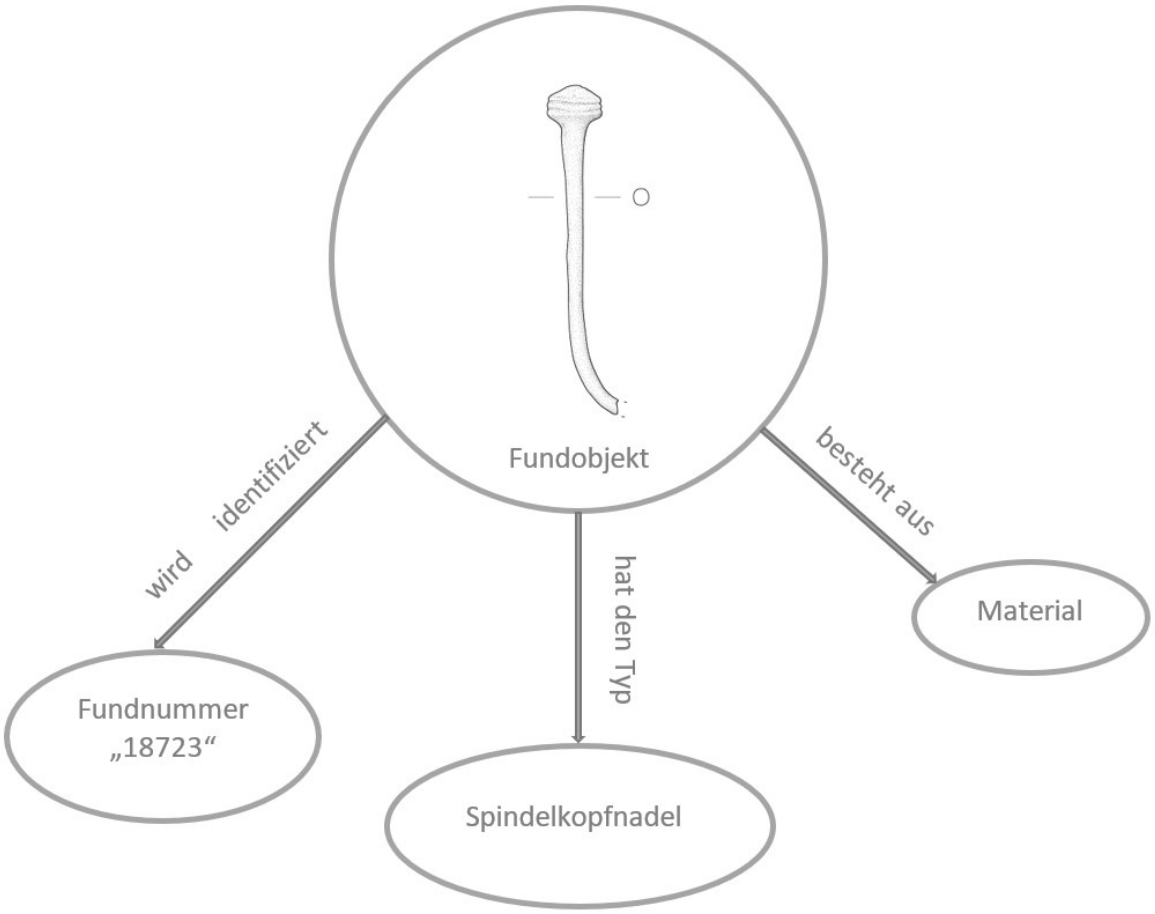


Abbildung 22 – Exemplarischer RDF Graph zu einer Spindelkopfnadel graphisch dargestellt (Grafik: M. Simon; Zeichnung der Nadel: Copyright Institut für Urgeschichte und Historische Archäologie der Universität Wien).

4.3.2.1 RDF/XML

RDF/XML ist eine Syntax, um RDF Graphen in XML (Extensible Markup Language) auszudrücken. Sie war die erste Serialisierung von RDF. Triples mit Subjekt, Prädikat und Objekt werden innerhalb eines XML-Elements (engl. *term*) geschrieben. In RDF/XML können Namensräume

[497] Heath – Bizer 2011, 28.
[498] Schreiber – Raimond 2014, Kapitel 5.

für Abkürzungen von URIs verwendet werden.[499] Zur einfacheren Darstellung kann ein Subjekt mit mehreren Prädikaten untereinander beschrieben werden.[500]

Die Syntax soll anhand eines Beispiels beschrieben werden: Um ein RDF/XML Dokument als solches zu deklarieren, werden die XML-Anfangs- und Endbezeichnungen <rdf:RDF> und </rdf:RDF> als Klammern am Beginn und am Ende des XML-Elements gesetzt. Innerhalb dieser Klammern werden die Triples beschrieben. Zu Beginn dieses Elements können auch die Namensräume definiert werden. Hierfür wird durch die Abkürzung „xmlns:“ (XML-Namespace) angezeigt, dass ein Namensraum angegeben wird. Die verwendete XML-Version kann optional noch vor diesem XML-Element gesetzt werden.[501]

Das Beispiel unten zeigt einen RDF-Graphen mit drei Triples, welche in Abbildung 22 visualisiert sind. Die Triples lauten: „Das Fundobjekt wird identifiziert durch die Fundnummer 18723“, „Das Fundobjekt hat den Typ „Spindelkopfnadel“ und „Das Fundobjekt besteht aus einem Material“. Im Beispiel werden zuerst die Namensräume angegeben. Dabei steht „rdf:“ für die Adresse „http://www.w3.org/1999/02/22-rdf-syntax-ns#“. Dort ist definiert, wie das Vokabular verwendet werden kann. Der Namensraum „CRM:“ steht für „http://www.cidoc-crm.org/“. In Zeile vier beginnt die Beschreibung der Triples. „rdf:Description“ besagt, welches Subjekt im Folgenden beschrieben wird. In diesem Beispiel ist das Subjekt das Fundobjekt mit dem URI „http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/18723”.
Die nachfolgenden Zeilen fünf bis sieben beziehen sich auf das angegebene Subjekt und geben das jeweilige Prädikat und Objekt wieder. Zeile fünf besagt, dass das Subjekt identifiziert wird durch die Fundnummer 18723. Dabei wird das Prädikat vor und nach dem Objekt angegeben (<CRM:Property> 18723 </CRM:Property>). Das zweite Triple wird in Zeile sechs angegeben. Dieses besagt, dass das Subjekt den Typ „Spindelkopfnadel“ hat. Das dritte Triple in Zeile sieben gibt an, dass das Fundobjekt aus einem Material besteht. Dieses kann in weiterer Folge wiederum mit weiteren Triples beschrieben werden. Unter den Triples, in Zeile neun, wird durch </rdf:Description> angezeigt, dass die Beschreibung des Subjekts aus Zeile vier beendet ist. Zur optischen Unterstützung ist das Subjekt in schwarz, das Prädikat in rot und das Objekt fett geschrieben.

[499] Gandon – Schreiber 2014, Kapitel 2.1.
[500] Gandon – Schreiber 2014, Kapitel 2.3.
[501] Gandon – Schreiber 2014, Kapitel 2.6.

Beispiel RDF/XML:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:CRM="http://www.cidoc-crm.org/">
     <rdf:Description
     rdf:about="http://univie.ac.at/Oberleiserberg/Metalle/
     Masterarbeit-Simon-2021/E22-object/18723">

         <CRM:Property/P1-is-identified-by/version-7.1.1> "18723"
         </CRM:Property/P1-is-identified-by/version-7.1.1>
         <CRM:Property/P2-has-type/
         version-7.1.1>"Spindelkopfnadel"</CRM:Property/
         P2-has-type/version-7.1.1>
         <CRM:Property/P5-consists-of/version-7.1.1> CRM:Entity/
         E57-Material/version-7.1.1 </CRM:Property/
         P5-consists-of/version-7.1.1>

     </rdf:Description>
</rdf:RDF>
```

4.3.2.2 Turtle Familie – N-Triples/Turtle

Im Gegensatz zur hierarchischen Beschreibung in RDF/XML werden bei der Turtle Familie die Triples einzeln nacheinander geschrieben. Jedes Triple wird in einer Zeile geschrieben und mit einem Punkt vom nächsten Triple getrennt. Zur Turtle Familie (Terse RDF Triple Language) werden vier Syntaxen gezählt. „N-Triples“ hat einen einfachen Satzbau, „Turtle“ stellt eine Erweiterung davon dar. Die anderen beiden Syntaxen „TriG“ und „N-Quads“, die wiederum eine Erweiterung der beiden ersten sind, sind geeignet, um mehrere Graphen zu codieren.[502] Nachfolgend werden die drei Syntaxen N-Triples, Turtle und N-Quads näher erläutert.

4.3.2.2.1 N-Triples

N-Triples ist eine zeilenbasierte Syntax, die nicht hierarchisch aufgebaut ist. Dabei wird pro Zeile nur ein Triple beschrieben. Jedes Triple ist unterteilt in Subjekt, Prädikat und Objekt. Die einzelnen Elemente sind mit einem Leerzeichen getrennt und das Triple wird mit einem Punkt am Ende der Zeile geschlossen. Die URIs müssen hier ausgeschrieben werden, ohne Namensräume, und sind mit Pfeilsymbolen (< >) eingerahmt. Literale werden mit Anführungszeichen gekennzeichnet und Blank Nodes (leere Knoten) werden mit einem Unterstrich und einem Doppelpunkt geschrieben (_:).[503]

[502] Schreiber – Raimond 2014, Kapitel 5.1.
[503] Beckett 2014, Kapitel 1-2.4.

Beispiel für N-Triples:

```
1     <http://univie.ac.at/Oberleiserberg/Metalle/
Masterarbeit-Simon-2021/E22-object/18723> <http://
www.cidoc-crm.org/Property/P1-is-identified-by/version-7.1.1> "18723" .
2     <http://univie.ac.at/Oberleiserberg/Metalle/
Masterarbeit-Simon-2021/E22-object/18723>  <http://
www.cidoc-crm.org/property/P2-has-type/version-7.1.1> "Spindelkopfnadel" .
3     <http://univie.ac.at/Oberleiserberg/Metalle/
Masterarbeit-Simon-2021/E22-object/18723> <http://
www.cidoc-crm.org/Property/P5-consists-of/version-7.1.1>
<http://www.cidoc-crm.org/Entity/E57-Material/version-7.1.1> .
```

Das Beispiel zeigt drei Triples. Die ersten beiden Triples sind Literal-Triples, während das dritte ein RDF Link ist. Zur optischen Unterstützung wird hier das Subjekt schwarz, das Prädikat rot und das Objekt fett geschrieben.

In der ersten Zeile wird angegeben, dass das Fundobjekt als Identifikator eine Fundnummer zugewiesen bekommen hat. Dabei wird das Fundobjekt durch den URI aus den eigenen Daten definiert <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/18723> und bildet das Subjekt. Das Prädikat <http://www.cidoc-crm.org/Property/P1-is-identified-by/version-7.1.1> ist aus dem CIDOC CRM entnommen und zeigt die Beziehung zwischen Subjekt und Objekt an. Das Objekt in diesem Triple ist die Fundnummer, welche als Literal angegeben wird.

Das zweite Triple in Zeile zwei gibt den Typ des Fundobjekts an. Das Fundobjekt als Subjekt mit dem URI <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/18723> hat die Beziehung „hat den Typ“ mit der URI <http://www.cidoc-crm.org/property/P2-has-type/version-7.1.1>. Das Objekt ist als Literal eingegeben als „Spindelkopfnadel“.

Das dritte Triple in Zeile drei zeigt an, dass das Fundobjekt aus einem Material besteht. Das Fundobjekt als Subjekt hat das Prädikat „besteht aus“ mit dem URI <http://www.cidoc-crm.org/Property/P5-consists-of/version-7.1.1>. Das Objekt ist das Material, aus dem das Fundobjekt besteht, welches durch eine Definition aus CIDOC CRM angegeben wird <http://www.cidoc-crm.org/Entity/E57-Material/version-7.1.1>. Mit weiteren Beschreibungen kann hier die Art des Materials angegeben werden, z. B. „Buntmetall“.

4.3.2.2.2 Turtle

Turtle ist N-Triples sehr ähnlich. Auch hier werden die Triples als Statement mit Subjekt, Prädikat und Objekt geschrieben. Die Syntax ist jedoch kompakter durch die Verwendung von Präfixen und die Verkettung von Prädikaten. Dabei zeigt ein Strichpunkt die Wiederholung des zuletzt

verwendeten Subjekts als Subjekt des nächsten Triples an. So kann ein Subjekt mit mehreren Prädikaten und Objekten verbunden werden. Wenn das gleiche Subjekt und Prädikat verwendet wird, ist es ebenso möglich, das Objekt zu wiederholen. Dies geschieht durch eine Trennung mithilfe eines Strichpunkts.[504]

Beispiel für Turtle:

```
1    PREFIX Fundobjekt: <http://univie.ac.at/Oberleiserberg/Metalle/
     Masterarbeit-Simon-2021/E22-object/>
2    PREFIX crm: <http://www.cidoc-crm.org/>
3
4    Fundobjekt:18723 crm:P1-is-identified-by/version-7.1.1 "18723";
5                     crm:P2-has-type/version-7.1.1 "Spindelkopfnadel";
6                      crm:P5-consists-of/version-7.1.1 crm:Entity/
                       E57-Material/version-7.1.1.
```

In diesem Beispiel wird der RDF Graph der Abbildung 22 in Turtle geschrieben. Zu Beginn sind zwei Präfixe angegeben: „Fundobjekt:“ für die Adresse <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/> und „crm:“ für <http://www.cidoc-crm.org/>. Das Beispiel zeigt drei Triples. Das Subjekt ist schwarz, das Prädikat rot und das Objekt fett geschrieben, um optisch eine bessere Einordnung zu ermöglichen.

Das erste Triple in Zeile vier besitzt als Subjekt das Fundobjekt mit dem URI <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/18723>. Der URI ist durch den vorher definierten Präfix abgekürzt und lautet nun „Fundobjekt:18723“. Anschließend wird das Prädikat mit dem abgekürzten URI „crm:P1-is-identified-by/version-7.1.1“ angegeben. Durch die Verwendung des Präfixes wissen Nutzer*innen, dass der vollständige URI <http://www.cidoc-crm.org/Property/P1-is-identified-by/version-7.1.1> lautet. Das Prädikat gibt die Beziehung „wird identifiziert durch“ zwischen Subjekt und Objekt an. Das Objekt ist das Literal mit der Fundnummer „18723“.

Anders als bei N-Triples wird hier das Ende des ersten Triples durch einen Strichpunkt anstatt eines Punktes angegeben, da das Subjekt im nächsten Triple identisch ist. So kann angezeigt werden, dass das Subjekt des nächsten Triples ebenfalls das Fundobjekt ist. Das zweite Triple in Zeile fünf lautet daher: Das Fundobjekt hat den Typ „Spindelkopfnadel“. Das Fundobjekt stellt das Subjekt dar, das Prädikat wird gebildet durch „hat den Typ“ und das Objekt ist das Literal „Spindelkopfnadel“.

Da beim zweiten Triple am Ende ebenso wie beim ersten Triple ein Strichpunkt gesetzt wird, ist das Subjekt des dritten Triples ebenso das Fundobjekt. Das dritte Triple in Zeile sechs lautet

[504] Beckett u. a. 2014, Kapitel 2.

daher: Das Fundobjekt besteht aus einem Material. Hier wird am Ende ein Punkt gesetzt, welcher anzeigt, dass dies entweder das letzte Triple ist oder das nächste Subjekt ein anderes sein wird als das Fundobjekt.

4.3.2.2.3 N-Quads

N-Quads ist ein zeilenbasiertes, einfaches Textformat zur Serialisierung von RDF Datensets. Es ist ähnlich aufgebaut wie N-Triples. N-Quads ermöglicht es jedoch, mehrere Graphen zu codieren. N-Quads Statements bestehen aus einer Sequenz von RDF Ausdrücken (engl. *terms*). Diese setzen sich jeweils aus dem Subjekt, Prädikat und Objekt zusammen mit dem Zusatz des Graphen-Labels. Das Graphen-Label gibt an, zu welchem Graphen das RDF-Triple gehört. Die einzelnen Elemente sind mit einem Leerzeichen oder einem Tab voneinander getrennt und werden mit einem Punkt und einer neuen Zeile beendet.[505] Dabei kann das Graphen-Label auch weggelassen werden. In diesem Fall wird das Triple dem Standard-Graphen des Datensets zugeordnet.[506] URIs dürfen nur in ausgeschriebener Form ohne Präfixe verwendet werden und werden von Pfeilsymbolen eingesäumt (< >).[507]

Beispiel für N-Quads:

```
1  <http://univie.ac.at/Oberleiserberg/Metalle/
   Masterarbeit-Simon-2021/E22-object/18723> <http://
   www.cidoc-crm.org/Property/P1-is-identified-by/version-7.1.1> "18723"
   <http://www.repositorium.org/Graph1>.
2  <http://univie.ac.at/Oberleiserberg/Metalle/
   Masterarbeit-Simon-2021/E22-object/18723>  <http://
   www.cidoc-crm.org/property/P2-has-type/version-7.1.1>
   "Spindelkopfnadel" <http://www.repositorium.org/Graph1>.
3  <http://univie.ac.at/Oberleiserberg/Metalle/
   Masterarbeit-Simon-2021/E22-object/18723> <http://
   www.cidoc-crm.org/Property/P5-consists-of/version-7.1.1>
   <http://www.cidoc-crm.org/Entity/E57-Material/
   version-7.1.1><http://www.repositorium.org/Graph4>.
```

In diesem Beispiel wird der Graph, welcher im Beispiel der N-Triples gezeigt wurde, erweitert. Die vierte Komponente, der Graph, in dem sich das Triple befindet, wird mitangegeben. Die Graphen wurden in diesem Beispiel blau geschrieben, um optisch besser erkennbar zu sein. Die ersten beiden Triples befinden sich in Graph 1, während sich das dritte Triple in Graph 4 befindet.

[505] Carothers 2014, Kapitel 1.
[506] Carothers 2014, Kapitel 2.1.
[507] Carothers 2014, Kapitel 2.2.

4.3.2.3 JSON-LD

JSON ist ein Format zum Austausch von Daten zwischen Webseiten und dem Browser. Es steht für „Java Script Object Notation“. JSON-LD ist eine Erweiterung davon, wobei LD für Linked Data steht. Es kann zur Serialisierung von RDF-Daten verwendet werden.

Bei JSON werden Name-Wert-Paare (engl. *name/value pairs*) verwendet. Hierbei wird der Name in Anführungszeichen geschrieben, gefolgt von einem Doppelpunkt und einem Wert (z.B. “Name“: “Anna“ oder “Nummer“: “1234“). Diese werden bei der Erweiterung JSON-LD um sogenannte Schlüsselwörter (engl. *keywords*) ergänzt, die mit einem vorangestellten @-Zeichen geschrieben werden können. Das ist wichtig, um Mehrdeutigkeiten zu vermeiden und Objekte eindeutig identifizieren zu können.[508]

Die Grundidee von JSON-LD ist, seine Daten in einen Kontext zu stellen und sie zu identifizieren. Das Schlüsselwort *@context* definiert das für den Kontext zugrundeliegende Vokabular (ähnlich den Präfixen). So können Abkürzungen für URIs oder Prädikate verwendet werden. Zur Identifikation dient das Schlüsselwort *@id*, welches den URI der zu beschreibenden Ressource, also des Subjektes, beinhaltet. Die Eigenschaften in den nachfolgenden Zeilen beziehen sich auf dieses Subjekt.[509] Die geschwungenen Klammern {} rahmen ein JSON Objekt ein.[510] Die folgende Abbildung 23 zeigt eine JSON-LD Syntax als ein Beispiel-Muster mit Subjekt, Prädikat und Objekt:[511]

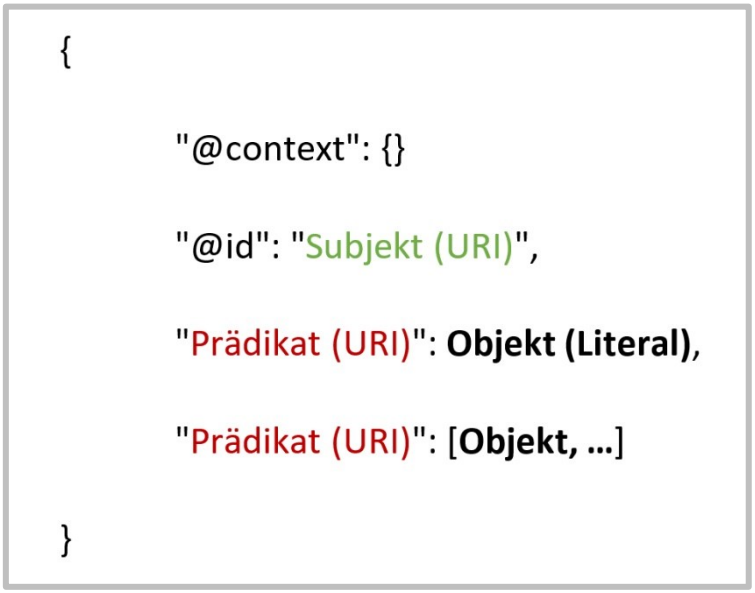


Abbildung 23 - Darstellung der JSON-LD Syntax mit Subjekt (grün), Prädikat (rot) und Objekt (fett) (Grafik: M. Simon nach Gerth 2017, 12).

[508] Sporny u. a. 2020, Kapitel 1.
[509] Sporny u. a. 2020, Kapitel 3.1, 3.2.
[510] Sporny u. a. 2020, Kapitel 1.4.
[511] Gerth 2017, 12.

Beispiel für JSON-LD:

```
{
      "@context": {
        "Fundobjekt": "http://univie.ac.at/Oberleiserberg/Metalle/
        Masterarbeit-Simon-2021/E22-object/",
        "crm": "http://www.cidoc-crm.org/property/",
      },
      "@id": "Fundobjekt:18723",
      "crm:Property/P1-is-identified-by/version-7.1.1": "18723",
      "crm:Property/P2-has-type/version-7.1.1": "Spindelkopfnadel",
      "crm:Property/P5-consists-of/version-7.1.1": "crm:Entity/
      E57-Material/version-7.1.1"
}
```

Der Kontext in diesem Beispiel ist „http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/“ und „http://www.cidoc-crm.org/“. Der Kontext wird unten mit „Fundobjekt“ und „crm“ abgekürzt. Hier wird definiert, welche Bedeutung die Abkürzungen „Property/P1-is-identified-by/version-7.1.1“, „Property/P2-has-type/version-7.1.1“ und „Property/P5-consists-of/version-7.1.1“ haben. Die @id (das Subjekt, also das Fundobjekt) wird von den nachfolgenden Prädikaten („Property/P1-is-identified-by“, etc.) beschrieben. Die Ressource Fundobjekt wird mit dem Identifikator Fundnummer „18723“, dem Typ „Spindelkopfnadel“ und eines Materials beschrieben.

4.4 SPARQL – Abfragesprache für RDF Graphen

SPARQL (SPARQL Protocol and RDF Query Language) ist eine Abfragesprache für RDF Graphen und die Abbildung ihrer Resultate. Sie wurde speziell für RDF entwickelt und bietet Möglichkeiten, um RDF-Daten zu selektieren, extrahieren und mit ihnen zu interagieren.[512] T. Berners-Lee betont die Bedeutung von SPARQL bei semantischen Technologien.[513]

Mit der Abfragesprache können RDF-Graphen im Web oder in einem RDF-Store (Triple Store, siehe auch Kapitel 4.5) abgefragt oder manipuliert werden. Die Bandbreite der Abfragen reicht von einfachen Abfragen, welche übereinstimmende Muster zwischen den RDF Daten und der Abfrage erkennen bis hin zu sehr komplexen Abfragen.[514]

Die benötigte Infrastruktur, um solche Abfragen durchführen zu können, bildet eine Software. Meist handelt es sich um einen sogenannten Triple Store. Ein Triple Store ist eine Datenbank für RDF-Daten und muss mit diesen Daten befüllt werden. Diese können abgefragt werden, indem

[512] Antoniou u. a. 2012, 69.
[513] W3C World Wide Web Consortium 2008.
[514] W3C SPARQL Working Group 2013, Kapitel 1-2.

die SPARQL-Abfragen über das HTTP-basierte SPARQL-Protokoll gesendet werden. Es gibt meist einen sogenannten „Endpoint“, einen Endpunkt, in welchem die Abfragen stattfinden. Clients senden ihre Abfragen zu einem Endpoint, indem sie das HTTP-Protokoll benutzen. Dies wäre bereits in der URL-Leiste jedes Browsers möglich, es wird jedoch empfohlen, spezielle Clients für SPARQL zu benutzen.[515]

Als Basis, um die Abfragen zu formulieren, dient die Turtle-Syntax (siehe Kapitel 4.3.2.2). SPARQL bietet zusätzlich Abfragevariablen, die das Resultat der Abfrage bestimmen sollen. Die Grundidee hinter SPARQL-Abfragen ist, Sets von Triples in den RDF-Daten zu finden, die zu dem gewünschten Graphenmuster (engl. *Graph pattern*) in der Anfrage passen.[516]

Abfragen beinhalten Basis-Graphenmuster (engl. *Basic Graph patterns*), welche wie RDF-Triples zu verstehen sind. Der Unterschied ist, Subjekt, Prädikat und Objekt können als Variablen dargestellt werden. Das Basis-Graphenmuster stimmt mit den RDF-Daten überein, wenn einzelne Elemente des RDF-Graphen mit der Variablen ausgetauscht werden können.[517] Eine Abfrage setzt sich aus drei wichtigen Elementen zusammen: PREFIX, SELECT und WHERE.[518] Im Folgenden (Abbildung 24) ist das Grundgerüst einer SPARQL-Abfrage zu sehen:

```
PREFIX

PREFIX

SELECT ?Variable1 ?Variable2

WHERE {

        Graph

}
```

Abbildung 24 – Grundgerüst einer SPARQL-Abfrage (Grafik: M. Simon nach Antoniou u. a. 2012, 70-72).

[515] Antoniou u. a. 2012, 70.
[516] Antoniou u. a. 2012, 70–72.
[517] Harris – Seaborne 2013, Kapitel 2.
[518] Antoniou u. a. 2012, 70–72.

4.4.1 Schlüsselwörter

Das erste Schlüsselwort PREFIX bietet die Möglichkeit, Abkürzungen für URLs anzugeben und zu definieren, ähnlich der Präfixe bei einigen Serialisationen von RDF.[519] Beim zweiten Schlüsselwort SELECT wird bestimmt, welche Variablen ausgegeben und in der Trefferliste angezeigt werden sollen. Dies können auch mehrere Variablen sein. Die Variablen dienen als Platzhalter und werden im Resultat dann mit Werten versehen.[520] Mit der WHERE Klausel wird die Abfrage eingeleitet. In geschwungenen Klammern steht ein Graphenmuster bestehend aus Triples in der Turtle Syntax. Jedes Element eines Triples kann mit einer Variablen ersetzt werden. Die Variable ist gekennzeichnet durch ein vorangeschriebenes Fragezeichen (z.B. ?location). In der Abfrage werden nun Triples, die zum angegebenen Graphenmuster mit der Variablen übereinstimmen, gesucht.[521]

Wenn in der Angabe das Triple "Fundstelle dbpo:location ?location" steht, werden die Triples gefunden, die darauf passen. Also alle, die "Fundstelle" als Subjekt und "dbpo:location" als Prädikat haben. Die Abfrage soll demnach alle Fundstellen aus einer Datenbank ausgeben.

Beispiel einer SPARQL-Abfrage:

```
1     PREFIX agd: <http://www.ausgrabungsdaten.at/Fundstellen.ttl#>.
2     PREFIX dbpo: <http://dbpedia.org/ontology/>.
3     SELECT ?location
4     WHERE {
5         agd:Fundstelle dbpo:location ?location.
6     }
```

Die beiden ersten Zeilen definieren die Präfixe, welche verwendet werden. „agd" steht demnach für <http://www.ausgrabungsdaten.at/Fundstellen.ttl#> und „dbpo" wird stellvertretend für <http://dbpedia.org/ontology/> angegeben. Die dritte Zeile mit dem Schlüsselwort SELECT bestimmt die Variable, welche im Ergebnis ausgegeben werden soll. Nach dem Schlüsselwort WHERE wird das Graphenmuster angegeben, auf welches die Abfrage zutreffen soll. In diesem Fall werden alle Triples ausgegeben, welche als Subjekt <http://www.ausgrabungsdaten.at/Fundstellen.ttl#Fundstelle> und als Prädikat <http://dbpedia.org/ontology/location> aufweisen.

[519] Antoniou u. a. 2012, 70–72.
[520] Antoniou u. a. 2012, 70–72.
[521] Antoniou u. a. 2012, 70–72.

4.4.2 Resultat

Das Resultat wird bestimmt durch die Variable, die nach dem SELECT Schlüsselwort angegeben wird, und in einer Tabelle angezeigt (siehe Abbildung 25). Dabei beinhaltet jede Zeile ein Resultat.[522] Je nachdem, wie das Graphenmuster mit den RDF-Daten übereinstimmt, kann die Abfrage null, einen oder mehrere Treffer ergeben. Das Resultat zeigt hier alle möglichen Treffer von Orten (durch die Variable ?location) von Fundstellen an, welche sich in diesem durchsuchten Graphen befinden.[523] Das Ergebnis der oben geschriebenen Abfrage würde folgendermaßen aussehen, wenn in den Daten nur diese beiden Fundstellen vorhanden wären:

?location
http://dbpedia.org/resource/Oberleiserberg.
http://dpbedia.org/resource/ThunauAmKamp.

Abbildung 25 – Resultat der SPARQL-Abfrage (Grafik: M. Simon).

4.4.3 Weitere Funktionen von SPARQL

SPARQL bietet noch andere Funktionen, um die Abfragen flexibel gestalten zu können. Das Schlüsselwort LIMIT mit nachgestellter Zahl kann ans Ende nach der geschwungenen Klammer gesetzt werden, um die Ausgabe auf die angegebene Zahl zu reduzieren.[524] Um nach einem gewissen Zahlenbereich (wie z. B. die Datierung) zu suchen, können in der Anfrage noch zusätzliche Bedingungen hinzugefügt werden. Mit dem FILTER Schlüsselwort können die Suchfelder mit verschiedenen Operatoren (wie etwa <, >, =, isLiteral(X) etc.) eingegrenzt werden.[525]

Literale werden mit Anführungszeichen ("x", 'x') geschrieben. Optional können Sprachen-Tags hintangestellt werden, die mit dem @-Symbol gekennzeichnet sind (z.B. 'x'@en für den Sprachen-Tag Englisch). Zahlen können mit oder ohne Anführungszeichen geschrieben werden.[526]

4.4.3.1 Abfrageformen

Es gibt vier Abfrageformen bei SPARQL:[527]

[522] Antoniou u. a. 2012, 70–72.
[523] Harris – Seaborne 2013, Kapitel 2.2.
[524] Antoniou u. a. 2012, 74.
[525] Antoniou u. a. 2012, 75–78.
[526] Harris – Seaborne 2013, Kapitel 2.3 und 4.1.2.
[527] Harris – Seaborne 2013, Kapitel 16.

- SELECT: Bei dieser Abfrage werden alle Ergebnisse in einer Tabelle angezeigt, die die gewünschten Bedingungen erfüllen.[528]
- CONSTRUCT: Hier wird als Resultat anstatt einer Tabelle ein RDF-Graph ausgegeben. Die Auswahl ist wie bei der SELECT-Abfrage.[529]
- DESCRIBE: Diese Abfrage ergibt einen RDF-Graphen mit Informationen/Daten über die angefragten Ressourcen. Zum Beispiel alle Triples über eine Ressource.[530]
- ASK: Diese Form der Abfrage resultiert in einem booleschen Ausdruck (TRUE/FALSE), um anzuzeigen, ob eine Abfrage Treffer ergeben würde oder nicht. Es werden keine Informationen ausgegeben.[531]

Mittels SPARQL können die Daten im Triple Store auch bearbeitet werden. Abseits der Abfragen können sowohl Updates der RDF-Graphen durchgeführt werden als auch neue RDF-Graphen kreiert oder andere in einem Triple Store gelöscht werden.[532]

4.5 Veröffentlichung und Speicherung von Linked Data - Triple Store – Datenbank für RDF

Nachdem die Daten in ein RDF Format gebracht wurden, sollen sie auch für die Öffentlichkeit zugänglich gemacht werden. Hierfür gibt es mehrere Wege. Der gebräuchlichste und für Anwender nützlichste ist über einen Triple Store, einen Speicherort für RDF Graphen.

4.5.1 Veröffentlichung

Die Veröffentlichung erfolgt je nach Art der Daten, die als Linked Data publiziert werden sollen. Überlegungen zum Datenvolumen und der voraussichtlichen Dynamik, also wie oft die Daten aktualisiert werden müssen, sollten miteinbezogen werden. Bei archäologischen Daten, die abgeschlossene Ausgrabungen oder Sammlungen beschreiben, ist dies wahrscheinlich eher selten der Fall.

T. Heath und C. Bizer beschreiben die verschiedenen Ausgangsdaten, wie die Datentypen als Linked Data veröffentlicht werden können. Sie teilen sie grob in folgende Kategorien ein:[533]

[528] Harris – Seaborne 2013, Kapitel 16.1.
[529] Harris – Seaborne 2013, Kapitel 16.2.
[530] Harris – Seaborne 2013, Kapitel 16.4.
[531] Harris – Seaborne 2013, Kapitel 16.3.
[532] W3C SPARQL Working Group 2013, Kapitel 6.
[533] Heath – Bizer 2011, Kapitel 5.1.1.

- *Von abfragbaren und strukturierten Daten zu Linked Data:* Hier sind relationale Datenbanken, wie MySQL, PostgreSQL oder ORACLE, gemeint, welche als Linked Data publiziert werden können mit Hilfe von sogenannten „Relational-Database to RDF wrappers". Diese bieten Möglichkeiten zum Mapping der Daten in der Datenbank, um daraus RDF-Graphen zu erstellen. Ein Beispiel für dieses Tool wäre der D2R Server.[534]
- *Von statischen, strukturierten Daten zu Linked Data:* Hier können als Beispiele CSV-Dateien, Excel-Dateien, XML-Dateien oder aus Datenbank extrahierte Daten (engl. *dumps*) genannt werden. Diese Daten müssen in RDF konvertiert werden. Dies kann mit Hilfe von Online-Tools wie Karma[535] durchgeführt werden. Daraus resultieren statische RDF-Dateien, die auf einen Webserver oder direkt in einen Triple Store geladen werden können.[536]
- *Von Textdokumenten zu Linked Data:* Mit einem „Linked Data entity extractor" können Dokumente annotiert werden. Dabei werden die URIs von Objekten, die in einem Text vorkommen, annotiert und können gemeinsam mit dem Textdokument publiziert werden. Pelagios hat die Online-Software „Recogito"[537] entwickelt, mit der „semantische Annotationen ohne die spitzen Klammern", also besonders benutzerfreundlich, an Texten und Bildern durchgeführt werden können.[538]

Es gibt viele Möglichkeiten, die in RDF-Formate gebrachten Daten im Anschluss zu veröffentlichen.[539] Die Basisschritte zur Veröffentlichung von Linked Data sind: URIs für die beschriebenen Einheiten vergeben und diese zu dereferenzieren (siehe Kapitel 4.1.2), RDF-Links zu anderen Datenquellen herzustellen, um das Web of Data als Ganzes zu unterstützen, und Metadaten für die Daten bereitzustellen.[540] Am nützlichsten für andere Nutzer*innen sind Daten, die über einen SPARQL-Endpoint abzufragen sind und ebenfalls über einen sogenannten RDF-Dump zur weiteren Verarbeitung heruntergeladen werden können.[541,542]

4.5.2 Triple Stores – Speicherung von RDF-Daten

Ein Triple Store ist im Wesentlichen eine Datenbank für RDF-Daten. Er dient dazu, Triples zu speichern. Der Triple Store kann abgefragt werden, wenn die Triples hineingeladen wurden. Dies

534 Heath – Bizer 2011, Kapitel 5.4.2.
535 Siehe https://usc-isi-i2.github.io/karma/, abgerufen am 21.10.2021.
536 Eine Liste von RDF-Konvertierungs-Tools steht auf dieser Webseite: https://www.w3.org/wiki/ConverterToRdf, abgerufen am 21.10.2021.
537 Siehe https://recogito.pelagios.org/, abgerufen am 21.10.2021.
538 Simon 2017.
539 Heath – Bizer 2011, Kapitel 5.1 und 5.2.
540 Bizer u. a. 2009, 6–8.
541 Bizer u. a. 2009, 8 f.
542 Heath – Bizer 2011, Kapitel 5.3.

kann auf unterschiedliche Weise durchgeführt werden. Es gibt beispielsweise „Bulk Upload" Optionen oder die SPARQL-Updates. Hierbei können Triples sowohl in den Triple Store übertragen als auch gelöscht werden. Über einen SPARQL-Endpoint können Abfragen über die Daten durchgeführt werden (siehe Kapitel 4.4).[543]

Es gibt viele unterschiedliche Triple Stores.[544] Am besten geeignet sind nach T. Heath und C. Bizer Triple Stores, die außer der Speicherung von RDF außerdem einen SPARQL-Endpoint sowie ein Linked Data Interface anbieten.[545] Im Folgenden werden zwei Beispiele für Triple Stores erläutert. Die Wahl basiert auf den jeweiligen Bedürfnissen der Datenpublizierer*innen.[546]

OpenLink Virtuoso - Virtuoso Universal Server

Virtuoso Universal Server[547] von OpenLink Software bietet einige Softwarefunktionen an. Es wird sowohl als Open-Source als auch in einer kommerziellen Form angeboten. Das Paket beinhaltet ein RDBMS (Relational Database Management System), einen Triple Store und einen Web Applikation Server.[548] Der Triple Store kann RDF-Daten über ein Linked Data Interface und einen SPARQL-Endpoint bereitstellen. RDF-Daten können in Virtuoso gespeichert werden oder dort basierend auf einem Mapping kreiert werden.[549]

Blazegraph

Blazegraph ist eine Open-Source Graphdatenbank (welche unter anderem auch andere Abfragesprachen als SPARQL verwenden und unterschiedliche Arten von Graphen speichern kann, nicht nur RDF-Graphen) und ein Triple Store, der SPARQL-Abfragen und -Updates unterstützt. Es kann auch eine erweiterte Lizenz erworben werden.[550]

---

[543] Antoniou u. a. 2012, 70.
[544] Eine Liste von Triple Stores findet sich auf dieser Webseite: https://www.w3.org/wiki/SemanticWebTools#RDF_Triple_Store_Systems, abgerufen am 21.10.2021.
[545] Heath – Bizer 2011, Kapitel 5.2.5.
[546] Für weitere Beispiele und Erläuterungen siehe Gerth 2017, 18 f.
[547] Siehe https://virtuoso.openlinksw.com/, abgerufen am 21.10.2021.
[548] Gerth 2017, 18.
[549] Bizer u. a. 2009, 8 f.
[550] Siehe https://blazegraph.com, abgerufen am 21.10.2021.

5 Ontologien – Beziehungen zwischen den Daten definieren

T. Berners-Lee bringt Ontologien als Lösung für ein spezielles Problem ins Spiel, die Nonunique Naming Assumption (siehe Kapitel 3.1.2). Wenn zwei Datenbanken einen unterschiedlichen Identifikator für ein und dasselbe Konzept, wie z.B. Adresse, verwenden, ist es schwierig zu erkennen, ob es sich um das gleiche Konzept handelt. T. Berners-Lee sieht die Lösung hierfür bei Ontologien, welche die Informationen zu den Daten liefern und die Beziehungen zwischen ihnen definieren können.[551]

Das Wort Ontologie hat unterschiedliche Bedeutungen in verschiedenen Bereichen. Am deutlichsten sind diese Unterschiede zwischen der Bedeutung in der Philosophie und der Computertechnik ersichtlich.[552] Der Ausdruck Ontologie stammt ursprünglich aus der Philosophie.[553] Sie versteht die Ontologie als Zweig, der sich mit der Natur und der Grundstruktur der Wirklichkeit oder des Seins beschäftigt und wird oft auch als die „Lehre des Seins" betitelt. Schon Aristoteles definierte die Ontologie als Studie der Eigenschaften, die aufgrund ihrer Natur Dingen in der Welt zugeordnet werden.[554] Ontologien beschreiben in der Philosophie also Dinge, die in der Welt existieren, und wie sie zusammenhängen.[555]

In der Computertechnik sollen Ontologien die Struktur eines Systems formal modellieren.[556] Sie bieten ein gemeinsam genütztes und verstandenes Vokabular, welches Konzepte, Eigenschaften, Beziehungen sowie alle Definitionen dazu liefert.[557] Als Struktur eines Systems können beispielsweise relevante Einheiten und deren Beziehungen in einem System verstanden werden. Einheiten werden Konzepte, Beziehungen und Eigenschaften zugeordnet. Als Konzepte könnten als Beispiel „Person", „Wissenschafter*in" oder „Autor*in" dienen. Dabei würde die „Person" als übergeordnetes Konzept oder Klasse gelten und die letzten beiden als hierarchisch untergeordnetes Konzept oder Subklasse. „Kooperiert mit" könnte als Beziehung zwischen den Klassen oder Subklassen dienen.[558]

Die Definitionen der Ontologie in der Computertechnik wurden mehrfach spezifiziert. T. R. Gruber definierte 1993 eine Ontologie als exakte „Spezifizierung einer Konzeptualisierung".[559] 1997 erweiterte W. N. Borst die Definition um den Aspekt eines Konsenses über die Konzepte, anstelle

[551] Berners-Lee u. a. 2001, Ontologies.
[552] Guarino u. a. 2009, 1.
[553] Pan 2009, 74 f.
[554] Guarino u. a. 2009, 1.
[555] Pan 2009, 74 f.
[556] Guarino u. a. 2009, 2.
[557] Pan 2009, 75.
[558] Guarino u. a. 2009, 2.
[559] Gruber 1993, 199.

von einzelnen Ansichten. Denn eine Ontologie wird nur dann von anderen wiederverwendet, wenn ihre Konzepte generell akzeptiert werden.[560] Zu den geteilten Ansichten ist für ihn zusätzlich ein formales, maschinenlesbares Format notwendig. 1998 führten R. Studer u. a. diese Definitionen zusammen.[561] Sie kommen zum Ergebnis, eine Ontologie sei eine „formale, explizite Spezifizierung einer geteilten Konzeptualisierung".

Dabei gehen sie näher auf die verwendeten Begriffe ein. Eine „Konzeptualisierung" nimmt Bezug auf ein abstraktes Modell eines Phänomens in der Welt, indem die relevanten Konzepte dieses Phänomens identifiziert werden.[562] Sie gibt einen abstrakten, einfachen Blickwinkel auf die Welt wieder, der repräsentiert werden soll.[563] Mit „explizit" ist gemeint, dass der Typ der Konzepte und ihrer Beschränkungen, die verwendet werden, genau definiert werden sollen. Als „formal" wird eine Ontologie beschrieben, die in einem maschinenlesbaren Format verfasst ist. „Geteilt" soll bedeuten, eine Ontologie soll Wissen verwenden, über welches Konsens in einer Gruppe herrscht und welches nicht nur von einem Individuum akzeptiert wird.[564]

Das Angebot von unterschiedlichsten Ontologien ist sehr groß, doch es gibt innerhalb einzelner Disziplinen einen Konsens über die Ontologien, welche am besten für das jeweilige Fach geeignet sind. So kann die Anwendung dieser Ontologien das Problem der Nonunique Naming Assumption lösen. Daten, welche mit Informationen über ihre Konzepte hinterlegt sind, können von anderen Nutzer*innen auch unmissverständlich interpretiert werden. Im Folgenden werden zwei stark verbreitete Ontologien, die verwendet werden können, um Metadaten und archäologische Datensätze zu modellieren, vorgestellt.

5.1 Dublin Core – Ontologie zur Beschreibung von Metadaten

Dublin Core ist eine einfache Ontologie mit wenig Klassen und Eigenschaften, die meist verwendet wird, um Metadaten von Dokumenten und Datensets zu definieren.[565] Sie wurde von der Dublin Core Metadata Initiative (DCMI)[566] entwickelt. Diese ist eine Organisation, die internationale Standards für Metadaten erarbeitet. Mitwirkende einigen sich über die Metadaten, die verwendet werden können, um Ressourcen zu beschreiben.

Das originale Dublin Core Metadata Elements Set beinhaltete 15 Kernelemente. Dieser Kern wurde durch die DCMI Metadata Terms erweitert, die komplexer sind und 22 Klassen und 55

[560] Borst 1997, 12.
[561] Studer u. a. 1998, 184.
[562] Studer u. a. 1998, 184.
[563] Guarino u. a. 2009, 3.
[564] Studer u. a. 1998, 184.
[565] Siehe https://www.w3.org/wiki/Good_Ontologies, abgerufen am 21.10.2021.
[566] Siehe https://dublincore.org/, abgerufen am 21.10.2021.

Eigenschaften beinhalten. Sie sind publiziert als ISO 15836-2:2019[567]. Seit 2017 wird empfohlen, sich bei der Beschreibung von Metadaten auf diese DCMI Metadata Terms[568] zu beziehen. Die Core Elements werden in der Erweiterung weitergeführt und uneingeschränkt weiter unterstützt, damit bereits verlinkte Metadaten erhalten bleiben. Die am meisten verwendeten Namensräume für die beiden Versionen des Dublin Core sind <http://purl.org/dc/elements/1.1/>, mit dem Präfix „dc:“ und <http://purl.org/dc/terms/>, mit dem Präfix „dcterms:“.[569]

Der Term *creator* aus dieser Ontologie beschreibt beispielsweise eine Einheit, die verantwortlich ist für die Erstellung einer Ressource. Der Präfix für diesen Term ist *dcterms:creator* als Abkürzung für <http://purl.org/dc/terms/creator>.

Als Beispiel könnte ein Datenset mit der URI <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/> mit der URI <http://purl.org/dc/terms/creator> für die Erschaffung der Ressource das Institut für Urgeschichte und Historische Archäologie <http://viaf.org/viaf/245829558> angeben. Mit der Abkürzung durch den Präfix wäre es: <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/> dcterms:creator <http://viaf.org/viaf/245829558>.

Ein weiteres Beispiel könnte einem archäologischem Datenset ein Thema zuweisen, um Interessent*innen mit einem Blick auf die Metadaten zu vermitteln, ob die Daten für sie relevant sind. Hierzu kann die Eigenschaft *subject* gewählt werden und mit einem URI spezifiziert werden. Die URI der Eigenschaft ist <http://purl.org/dc/terms/subject>, mit dem Präfix „dcterms:“ für <http://purl.org/dc/terms/>. Die Abkürzung würde „dcterms:subject“ lauten. Das Datenset <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/> hat das Thema <http://purl.org/dc/terms/subject> „Urnenfelderzeit“) <https://dbpedia.org/page/Urnfield_culture>. Mit der Abkürzung durch den Präfix wäre es: <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/> dcterms:subject <https://dbpedia.org/page/Urnfield_culture>.

[567] ISO 2019.
[568] Siehe https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, abgerufen am 21.10.2021.
[569] Siehe https://www.w3.org/wiki/Good_Ontologies, abgerufen am 21.10.2021.

5.2 CIDOC-CRM – Ontologie für Daten aus dem Bereich des Kulturerbes

CIDOC ist das International Comittee for Documentation des internationalen Museumsverbandes ICOM (International Council of Museums)[570], welches Standards und Konzepte für die Dokumentation von Museumskollektionen entwickelt. Das Komitee veröffentlicht das CIDOC-CRM, das Conceptual Reference Model. Das CRM ist eine formale Ontologie, welche die Integration und den Austausch von heterogenen Informationen und Begriffen aus dem Kulturerbe erleichtern soll.[571]

Das CRM bietet Definitionen, eine formale Struktur und Beschreibungen von Konzepten und Beziehungen zur Datenintegration. Informationen im Bereich des Kulturerbes sollen besser verstanden werden, indem ein semantischer Rahmen geschaffen wird. Dabei kann jede Information aus dem gesamten Kulturerbe mit der Ontologie gemappt werden. Sie wird beispielsweise von Museen, Bibliotheken, Forschungsinstituten, Galerien, Archiven und in der Archäologie genutzt. Das Ziel ist, lokale, unterschiedliche Quellen integrieren und so gemeinsam in zusammenhängenden, globalen Informationsressourcen nutzen zu können.[572]

Das CIDOC CRM wurde 2006 zur ISO Norm, welche im Jahr 2014 erneuert wurde (ISO 21127:2014)[573]. Die letzte offizielle Version 7.1.1[574] beinhaltet als CRMbase (Basisversion) 99 Klassen und 198 Eigenschaften. Die vorige offizielle Version 5.4.0[575] aus dem Jahr 2011 enthielt 90 Klassen und 149 Eigenschaften.[576]

5.2.1 Methodik des CIDOC CRM

CIDOC CRM ist eine sogenannte Top-Level Ontologie, welche gängige Datenstrukturen der abzubildenden Forschungsbereiche analysiert und seine Ontologie darauf aufbaut. Top-Level Ontologien sind charakterisiert durch allgemeine Konzepte und eine breitgefächerte Ansicht auf die

[570] Siehe https://icom.museum/en/, abgerufen am 28.09.2020.
[571] Siehe https://cidoc.mini.icom.museum/standards/cidoc-standards-guidelines/, abgerufen am 21.10.2021.
[572] Siehe http://www.cidoc-crm.org/, abgerufen am 21.10.2021.
[573] ISO 2014.
[574] Bekiari u. a. 2021.
[575] Crofts u. a. 2011.
[576] Zu den Versionen des CIDOC CRM siehe http://www.cidoc-crm.org/versions-of-the-cidoc-crm, abgerufen am 21.10.2021.

Welt und sind daher für besonders viele Bereiche geeignet, was man an den zahlreichen Erweiterungen für spezielle Forschungsbereiche erkennen kann.[577] Dabei sind die Beziehungen, die zwischen den Einheiten liegen, wichtiger für die daraus resultierende Logik als die individuellen Daten selbst.[578]

Die Top-Level Ontologie ist nützlich, da Daten aus dem Kultur- und Museumsbereich so divers und vielfältig sind und es wohl nur schwer zu einer einheitlichen Datenstruktur kommen würde, welche großflächig anwendbar wäre und die alle Kerndaten erfassen könnte. Eine solche Ontologie würde zwangsläufig auch sehr groß werden. Die Datenstrukturen müssen flexibel sein, da Daten eines Museums der Modernen Kunst und einem archäologischen Archiv andere Bedürfnisse aufweisen.[579]

Die Methodik hinter dem CIDOC CRM soll interdisziplinär zwischen dem Kulturerbe und der Informatik funktionieren. Für IT-Spezialist*innen ist es oft schwer, die Logik von kulturhistorischen Konzepten nachzuvollziehen. Gleichermaßen schwer ist es für Expert*innen aus dem Kulturerbe, diese Konzepte den IT-Spezialist*innen zu kommunizieren. CIDOC CRM möchte diese Lücke schließen.[580]

Um eine kompakte Ontologie zu schaffen, ist es sinnvoll, sich auf die ontologischen Strukturen zu konzentrieren und weniger auf die domänen-spezifischen Terminologien.[581] Das CIDOC CRM wird nach Events, bzw. Ereignissen, modelliert. Zeitliche Einheiten und Events spielen eine zentrale Rolle und sind der Ausgangspunkt für die Modellierung.[582] Das CIDOC CRM versteht sich somit auch als eine „bottum-up" Ontologie, welche die allgemeine ontologische Struktur aus unterschiedlichen Fachbereichen gefunden hat, indem sie unterschiedliche Forschungsszenarien und Informations-Managementtaktiken analysiert hat und die Ontologie „von unten" aufgebaut hat.[583]

Die „Core"-Ontologie (Kern-Ontologie) des CIDOC CRM versucht, die in den Daten liegende Datenstruktur einzufangen. So soll ein allgemein gültiger Weg gefunden werden, im Gegensatz zu sogenannten „terminologischen Ontologien". Diese bauen auf individuellen Konzepten auf und werden in der Regel sehr viel größer. Beim CIDOC CRM sind weniger als fünf Prozent der Eigenschaften und Klassen museumsspezifisch zuzuordnen. Core-Ontologien sind kompakt, klein und sie konzentrieren sich auf die Beziehungen der Objekte und nicht auf die Objekte

[577] Guarino u. a. 2009, 16.
[578] Doerr 2003, 76, 79, 90.
[579] Doerr 2003, 77.
[580] Doerr 2003, 79 f.
[581] Doerr – Iorizzo 2008, 4.
[582] Doerr 2003, 84.
[583] Doerr – Iorizzo 2008, 4.

selbst.[584] Die wenigen Klassen und Eigenschaften können die Semantik von vielen unterschiedlichen Datenschemen einfangen und decken sehr allgemeine Konzepte ab.[585]

5.2.1.1 Details der Modellierung

Im Folgenden sollen die Methoden der Modellierung näher erklärt werden.[586] Sie unterscheiden sich von der üblichen Datenaufnahme in einer Datenbank oder Excel-Datei. Jede Instanz einer Klasse kann mit einer Appellation (wie beispielsweise der Name, ein Label oder ein Titel) benannt werden. Um die Terminologie näher zu bestimmen, können die Instanzen der Klassen auch in Typen (engl. *types*) eingeteilt werden.[587] Zeitliche Einheiten werden nicht als Zeitpunkte beschrieben, sondern als Zeitintervalle.[588] Eine Ortsbestimmung kann mittels eines geometrischen Areals (wie z.B. Frankreich, Wien oder Koordinaten) oder eines Objektes (z.B. ein Schiff oder auch Landschaften wie ein Berg oder ein Fluss) geschehen.[589] Eigenschaften, die ein Objekt beeinflusst haben, können ebenfalls eingebunden werden. Dies können physikalische Einflüsse wie Modifikationen, Produktion oder Kopien durch Werkzeuge oder Personen sein.[590] M. Doerr und D. Iorizzo beschreiben drei zentrale Ideen, die dem CIDOC CRM zugrunde liegen:[591]

- Es wird zwischen der Repräsentation des realen Objekts in der Welt und der Beschreibung, welche den Namen dieses Objekts repräsentiert, unterschieden.
- Die zweite Idee besagt, alle Typen und Klassifikationen von Objekten und Konzepten seien vom Menschen erfunden worden und deshalb soll die Dokumentation dieser mitbeschrieben werden.
- Die dritte zentrale Idee ist die ereigniszentrierte (bzw. eventzentrierte) Art und Weise der Beschreibung von Geschichte. Um die Geschichte zu analysieren, kann sie in einzelne, separate Ereignisse (engl. *Events*) eingeteilt werden. Im CIDOC CRM wird die Klasse dafür *E2 Temporal Entity* genannt. Diese Events involvieren wiederum persistente Items materieller und immaterieller Natur. Persistente Items sind dauerhaft existierende Dinge, im CIDOC CRM wird hierfür die Klasse *E77 Persistent Item* verwendet. Unter materiellen „Dingen" sind beispielsweise Personen oder Fundobjekte zu verstehen. Immaterielle Dinge schließen Konzepte oder Bezeichnungen wie z.B. das englische Königreich als *E28*

[584] Doerr – Iorizzo 2008, 5.
[585] Doerr – Iorizzo 2008, 7.
[586] Dabei werden die Klassen und Eigenschaften des CIDOC CRM in kursiv angegeben, um ein besseres Verständnis zu ermöglichen.
[587] Doerr 2003, 85.
[588] Doerr 2003, 87.
[589] Doerr 2003, 87 f.
[590] Doerr 2003, 88 f.
[591] Doerr – Iorizzo 2008, 7.

Conceptual object, ein. Die persistenten Items können bei Ereignissen als Konzepte oder durch physikalische Informationsträger präsent sein.[592] Jede dieser Klassen kann auch Subklassen haben, um sie näher zu beschreiben. Mit dieser eventzentrierten Betrachtungsweise kann die Geschichte als ein Netzwerk von Lebenslinien von persistenten Items gesehen werden, die in den Events aufeinandertreffen. Hierbei kann jedes Event vom CIDOC CRM beschrieben werden und die Beziehungen werden weniger komplex. Das CRM kann als eventbasiertes Modell bezeichnet werden, das Interaktionen beschreibt und keine Zustände.[593] In Abbildung 26 soll dies veranschaulicht werden. Caesars Geburt und sein Tod sind beides Events. Die involvierten Personen, wie Caesar, seine Mutter bei der Geburt oder Brutus bei seinem Tod, und Dinge, wie der Dolch, werden durch Beziehungen zu diesem Event beschrieben.

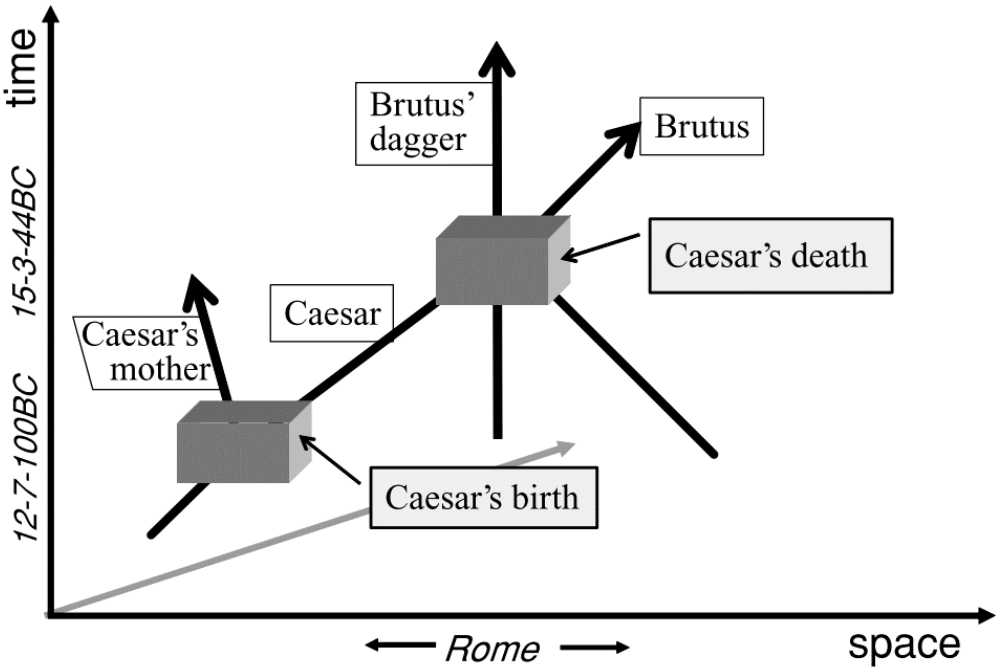


Abbildung 26 - Events bzw. Ereignisse in denen Personen und Dinge zusammentreffen (Quelle: Doerr – Iorizzo 2008, 8, Fig. 1).

Aus der „Definition des CIDOC CRM" ist zu entnehmen, dass die Ausdruckskraft der Ontologie daraus kommt, die Beziehungen der historischen Objekte durch definierte Eigenschaften zu dokumentieren.[594]

[592] Doerr – Iorizzo 2008, 7.
[593] Doerr – Iorizzo 2008, 8-11.
[594] Doerr u. a. 2020a, xviii.

5.2.2 Erweiterungen des CIDOC CRM

Die Ontologie ist in Modulform aufgebaut und bietet mehrere Erweiterungen. Diese wurden für unterschiedliche Forschungsfragen oder Dokumentationsweisen entwickelt, wie beispielsweise in der Geoinformatik oder Archäologie, und sind kompatibel mit der CRMbase (Basis) Ontologie. Folgende 10 Erweiterungen sind bis zum aktuellen Zeitpunkt entwickelt worden:[595]

- *FRBRoo – Functional Requirements for Bibliographic Records*[596] ist eine Erweiterung für bibliographische Informationen, um Bibliotheks- und Museumsinformationen zu integrieren.
- *PRESSoo – Modeling of bibliographic information*[597] ist eine Erweiterung von FRBRoo für bibliographische Informationen mit dem Schwerpunkt auf Zeitschriften (z.B. Fachzeitschriften, Zeitungen, Magazine).
- *CRMinf – Argumentation model*[598] verwendet und erweitert das CIDOC CRM, um eine formale Ontologie über Argumentationen und Inferenzen zu schaffen. Es wird vor allem bei beschreibenden und empirischen Studien verwendet und soll helfen, Daten über Beschreibungen von semantischen Beziehungen und die daraus gezogenen Schlussfolgerungen zu integrieren und auszutauschen.
- *CRMarchaeo – Excavation model*[599] unterstützt die semantische Beschreibung von archäologischer Felddokumentation. Es soll helfen, alle Beobachtungen und Dokumentationen, die Archäolog*innen während einer Ausgrabung tätigen, zu organisieren, integrieren und semantisch aufzubereiten. Dies soll Interoperabilität abseits der verwendeten Dokumentationsstandards ermöglichen. Dabei orientiert sich das Modell an der Harris Matrix, um die Stratigraphie archäologischer Fundstellen aufzunehmen.
- *CRMsci – Scientific observation model*[600] ist eine formale Ontologie für semantische Beschreibungen von wissenschaftlichen Beobachtungen und Messdaten in beschreibenden und empirischen Wissenschaften (wie etwa Geologie, Geografie, Archäologie, Informatik oder Konservierung).

[595] Siehe http://www.cidoc-crm.org/collaboration_resources, abgerufen am 21.10.2021.
[596] Siehe http://www.cidoc-crm.org/frbroo/, abgerufen am 21.10.2021.
[597] Siehe http://www.cidoc-crm.org/pressoo/, abgerufen am 21.10.2021.
[598] Siehe http://www.cidoc-crm.org/crminf/, abgerufen am 21.10.2021.
[599] Siehe http://www.cidoc-crm.org/crmarchaeo/, abgerufen am 21.10.2021.
[600] Siehe http://www.cidoc-crm.org/crmsci/, abgerufen am 21.10.2021.

- *CRMgeo – Spatiotemporal model*[601] bietet die Möglichkeit, eine formale Ontologie im raumzeitlichen Kontext zu verwenden. Das Modell verbindet CIDOC-CRM mit GeoSPARQL[602], um zeitliche Einheiten und Objekte sowohl räumlich als auch zeitlich einzuordnen. Archäologische Daten oder Daten aus dem Kulturerbe können mit identifizierbaren Ortsdaten und Geometrien von Fundstellen und Objekten bereichert werden. Es erleichtert den historischen Kontext mit dem räumlichen zu verbinden.
- *CRMdig – Model for provenance metadata*[603] erleichtert es, Metadaten zu Digitalisierungsprozessen, wie 2D oder 3D Modellen oder Scans zu beschreiben. Es beinhaltet die Beschreibung der Provenienz von digitalen Daten, wie beispielsweise Messprozesse und Parameter der Messungen oder der Produktionsmethoden.
- *CRMba – Model for Archaeological buildings*[604] ist geeignet, um archäologische Gebäudekomplexe zu dokumentieren und semantisch zu beschreiben. Gebäudeteile, Räume und die Stratigrafie können mit dem ganzen Gebäudekomplex in Kontext gestellt werden. So kann die Struktur, die Entwicklungsphasen und die Interpretation erfasst werden.
- *CRMtex – Model for the study of ancient texts*[605] ist eine Erweiterung des CIDOC-CRM, um antike Texte besser zu verstehen, indem die Texteinheiten identifiziert werden und der wissenschaftliche Prozess der Untersuchungen dokumentiert wird. Besonders die Papyrologie, die Paläografie, die Kodikologie und die Epigrafik profitieren davon, Texte mit allen Aspekten, von der Kreierung bis zur Konservierung in der heutigen Zeit und aller miteinhergehenden Untersuchungen und Übersetzungen semantisch beschreiben zu können.
- *CRMsoc – Model for Social Phenomena*[606] ist eine Ontologie, um soziale Phänomene und Konstrukte zu beschreiben, welche für die Geisteswissenschaften und Sozialwissenschaften relevant sind. Das Modell definiert wirtschaftliche Transaktionen, Rechte von Personen und Gruppen, historische Phasen und die Beschreibung von Plänen. Die Ontologie ist hilfreich in historischen „Hilfswissenschaften" wie etwa Politik- und Wirtschaftsgeschichte oder Sozialgeschichte.

[601] Siehe http://www.cidoc-crm.org/crmgeo/, abgerufen am 21.10.2021.
[602] GeoSPARQL ist eine geographische Abfragesprache für RDF-Daten, siehe dazu https://www.ogc.org/standards/geosparql, abgerufen am 21.10.2021.
[603] Siehe http://www.cidoc-crm.org/crmdig/, abgerufen am 21.10.2021.
[604] Siehe http://www.cidoc-crm.org/crmba/, abgerufen am 21.10.2021.
[605] Siehe http://www.cidoc-crm.org/crmtex/, abgerufen am 21.10.2021.
[606] Siehe http://www.cidoc-crm.org/crmsoc/, abgerufen am 21.10.2021.

5.2.3 CIDOC CRM und das Semantic Web

Das CIDOC CRM wurde ursprünglich nicht als Semantic Web-Projekt konzipiert, sondern zur Begriffsbildung von Konzepten. Diese werden benötigt, um Interaktionen von Menschen und kulturellen Einheiten zu dokumentieren.[607] Das CIDOC CRM wurde erstmals 1999 veröffentlicht und somit noch vor dem Semantic Web.[608] Seither entstanden jedoch viele konstruktive Zusammenarbeiten. Bei der semantischen Aufbereitung kulturhistorischer Daten ist das CIDOC CRM die wohl am häufigsten verwendete Ontologie.

5.2.3.1 RDF und CIDOC CRM

1998, zur gleichen Zeit als RDF von der W3C empfohlen wurde, wurde CIDOC CRM als ISO-Standard angemeldet. Es ist in seiner logischen Form als Ontologie unabhängig von RDF. Die weitreichendste logische Basis der Ontologie soll weiterhin erhalten und auch mit unterschiedlichen Codierungen verwendbar bleiben. Dies ist deshalb wichtig, um unabhängig von Codierungs-Modellen, welche sich im Laufe der Zeit ändern können, mit dem CIDOC CRM unterschiedliche Standards und Systeme vergleichen und analysieren zu können. Daher wird auf Codierungs-Modelle in der „Definition des CIDOC CRM" keine Rücksicht genommen.

Zurzeit ist RDF jedoch das am weitesten verbreitete „Knowledge Encoding Paradigm" (Wissenscodierungs-Modell) und auch das Modell, welches CIDOC CRM am nächsten steht. Daher wird bei jeder Veröffentlichung einer CIDOC CRM-Version auch ein kompatibles RDF-Schema auf der Homepage als offizielles Dokument zusätzlich zur „Definition des CIDOC CRM" mitveröffentlicht.[609]

Die meisten Klassen und Eigenschaften werden im CRM RDF-Schema definiert und die vorhandenen URLs sind dereferenzierbare Linked Data Identifikatoren. Der CRM-Präfix der Version 5.0.4 lautet folgendermaßen:[610] @PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>. Der Präfix der neuesten Version 7.1.1 lautet crm: <http://www.cidoc-crm.org/version/version-7.1.1>. Mit diesem Präfix können die Identifier wie beispielsweise „crm:E53_Place" in den Daten verwendet werden.[611] Es gibt allerdings einige Ausnahmen und Sonderfälle, die beachtet werden müssen, wenn eine Implementierung mit RDF vorgenommen werden soll. Hier soll auf das Dokument von M. Doerr, R. Light u. a. verwiesen werden.[612]

---

[607] Isaksen 2011, 34 f.
[608] Binding u. a. 2015, 7.
[609] Doerr u. a. 2020c, Introduction.
[610] Dieser verweist auf die Definition des CIDOC CRM der Version 5.0.4 von November 2011.
[611] Doerr u. a. 2020c, The CRM RDF Schema.
[612] Doerr u. a. 2020c.

5.2.3.2 URIs und CIDOC CRM

Um URIs bei der Verwendung des RDF-Schemas zu verwenden, wird der 303-redirect namespace mechanism (303-Umleitungsmechanismus) verwendet (siehe Kapitel 4.1.2.2). Jede Einheit, die in einem Datenset dargestellt wird, bringt mindestens drei URIs hervor:[613]

1. Die URI für das Objekt in der realen Welt selbst.
2. Die URI einer HTML Repräsentation, welche Informationen über das Objekt in der realen Welt liefert (menschenlesbar).
3. Die URI einer zugehörigen Informationsressource, welche Informationen über das Objekt in der realen Welt liefert und eine Repräsentation in RDF/XML beinhaltet (maschinenlesbar).

Das CIDOC CRM-Vokabular wird mittels eines HTML-Dokuments zur Verfügung gestellt. Je nach Client-Anfrage, welches Format erwünscht ist, kann der Inhalt in menschenlesbarer (HTML) oder maschinenlesbarer (RDF) Form abgerufen werden.

Die URI <http://www.cidoc-crm.org/cidoc-crm> wird sowohl verwendet, um die URIs des Vokabulars zu dereferenzieren, wenn ein HTML- oder RDF-Inhalt abgerufen wird und auch als Präfix für das CIDOC CRM-Vokabular. Um URIs von Klassen oder Eigenschaften des CIDOC CRM zu dereferenzieren, wie beispielsweise <http://www.cidoc-crm.org/cidoc-crm/E5_Event> oder <http://www.cidoc-crm.org/cidoc-crm/P148_has_component>, können jeweils entweder ein HTML- oder RDF-Inhalt abgerufen werden.[614] Eine archäologische Ausgrabung wäre zum Beispiel ein *E5 Event*. Wenn der Nutzer oder die Nutzerin dem Link dieser Klasse folgt, findet er die Definition der Klasse.

5.2.4 CIDOC CRM und die Anwendung in der Archäologie

Altdaten, Katalogdaten, Ausgrabungsdaten – archäologische Daten können vielfältig sein, ebenso wie die Möglichkeiten der Datenmodellierung mit dem CIDOC CRM. Wie die Ausgangsdaten modelliert werden können, welche Ziele sich ergeben und auf welche Herausforderungen Dateninhaber*innen stoßen können, wird im folgenden Kapitel diskutiert.

5.2.4.1 Semantische Interoperabilität in archäologischen Daten

Derzeit finden sich in der Archäologie oft fragmentierte Daten, welche unterschiedliche Datenschemas und auch Terminologien verwenden. Wenn Daten mit CIDOC CRM modelliert werden,

[613] Georgis 2018, Description of mechanism.
[614] Georgis 2018, Description of mechanism.

können sie inhaltlich richtig interpretiert werden und so semantische Interoperabilität, also das Zusammenspiel der Daten, erreicht werden.[615] Das Ziel von semantischer Interoperabilität sind unter anderem datenübergreifende Suchanfragen. Dies wird möglich, wenn die Daten in RDF repräsentiert sind und in einem konzeptionellen Datenstrukturschema bzw. einer allgemeinen Ontologie wie dem CIDOC CRM modelliert sind. Die Suche quer über mehrere Datenbanken soll so effektiver werden.[616] Um CRM anzuwenden, müssen die gewünschten Daten nach den Regeln des CRM-Datenmodells modelliert werden. Dieser Vorgang wird auch Mapping oder Modellierung genannt.

In zahlreichen Projekten[617] wurde untersucht, wie semantische Interoperabilität von archäologischen Daten und Datenbanken mit Hilfe von Mappings zur Ontologie CIDOC CRM erreicht werden kann. C. Binding u. a. haben im Zuge des STAR Projekts[618] die potenziellen Vorteile untersucht, indem sie fünf unterschiedlich strukturierte relationale Datenbanken mit Resultaten aus archäologischen Ausgrabungen mit der Ontologie CIDOC CRM gemappt haben.[619]

Diese Mappings sind zeitaufwendig, nicht einfach durchzuführen und bergen viele Fehlerquellen. Ein Hindernis sind oft die abstrakten und langen Beziehungsketten, die sich durch das eventbasierte Modell ergeben. Das Mapping ähnlicher Datensets durch unterschiedliche Fachexpert*innen resultieren oft in unterschiedlichen Ergebnissen. Dadurch können sich Inkonsistenzen ergeben. Beispielsweise können Datierungen oder die Ansprache von Fundobjekten auf mehrere Arten eingegeben werden. Diese Unterschiede und Inkonsistenzen erschweren die Suchanfragen. Sie können jedoch mit semi-automatisierten Werkzeugen (engl. *Tools*) eingeschränkt werden. Diese Werkzeuge geben Vorschläge zu passenden Klassen und Eigenschaften und können fehlerhafte Verbindungen aufzeigen.[620] Events, die Objekte und Orte miteinander verbinden, wie beispielsweise das Event einer archäologischen Ausgrabung, sind oft nur implizit in den Original-Datenbankenstrukturen und im nachfolgenden Mapping vorhanden. Beispiele hierfür sind auch Funde, die in einem Mess-Event gemessen werden und aus diesen Events ergeben sich in weiterer Folge die Messdaten. Bei der Übersetzung von relationalen Datenbanken zu einem RDF Graph muss die Information zu diesem Event erst kreiert werden, weil sie nicht explizit in der Datenbank vorhanden ist. Dies ist notwendig, um die Beziehungen im RDF Graph zu modellieren.[621]

---

615 Binding u. a. 2008, 281.
616 Binding u. a. 2008, 280 f.
617 Siehe zu den Projekten auch die Use Cases auf der CIDOC CRM-Webseite unter http://www.cidoc-crm.org/useCasesPage, abgerufen am 21.10.2021.
618 STAR ist ein Projekt des Archaeology Data Service, siehe unter https://archaeologydataservice.ac.uk/research/star.xhtml, abgerufen am 21.10.2021.
619 Binding u. a. 2008, 283.
620 Binding u. a. 2008, 282 f.
621 Binding u. a. 2008, 284 f.

Semi-automatisierte Werkzeuge erleichtern konsistentes Modellieren und Extrahieren. Vor der Modellierung müssen die Daten bereinigt werden und Identifikatoren via eines persistenten Domänen-Präfixes erstellt werden.[622] Archäologische Fachexpert*innen sind unerlässlich, um die passenden Mappings zu den archäologischen Daten zu finden. IT-Fachleute können den Vorgang nicht ohne sie bewerkstelligen, da sie die Daten oft nicht richtig interpretieren können.

C. Binding u. a. kommen zum Ergebnis, dass datenübergreifende Suchanfragen über miteinander verschmolzene Daten mit CIDOC CRM möglich sind und neue Ergebnisse hervorbringen können. Der Weg sollte gemeinsam von archäologischen und IT-Fachleuten beschritten werden. Die Anwendung sollte jedoch wesentlich benutzerfreundlicher werden, zum Beispiel durch semi-automatisierte Tools.[623]

5.2.4.2 Modellierung von Katalogdaten

Semantische und Linked Open Data Technologien gewinnen laut A. Deicke an Wichtigkeit in der Archäologie. Die Digitalisierung wurde im letzten Jahrzehnt immer präsenter und viele Forscher*innen publizieren ihre Rohdaten und Datenbanken online. A. Deicke sieht für die Archäologie einen großen Nutzen durch das Modellieren von Katalogdaten mit CIDOC CRM. Sie hat ein exemplarisches Datenmodell für archäologische Katalogdaten entworfen mit dem Fokus auf Fundstellen, Funden, Objekten und ihren Beziehungen zueinander (siehe Abbildung 27).

Sie betont, archäologische Daten können ihren Wiederverwendungs-Wert erhöhen, da die modellierten Einheiten im CIDOC CRM klar definiert sind. Der Vorteil sei, dass CIDOC CRM als Ontologie im Kulturerbe weithin angenommen wurde.

> „One important use case in this context concerns the growing importance of Linked Open Data and the Semantic Web, especially in archaeology. […] This potential for comprehensive analysis beyond the context of one specific project can only be unlocked if the underlying data is well structured according to a commonly used ontology.“[624]

Das CIDOC CRM bietet eine solche Ontologie und fördert dieses Potenzial.[625] A. Deicke gibt zu bedenken, die Vorteile einer zeitaufwändigen und komplexen Modellierung seien nicht sofort ersichtlich. Dies trifft insbesondere auf individuelle Forscher*innen oder kleinere Projekte zu, die nicht vorhaben, weitere Aufarbeitungen mit den gewonnenen Daten durchzuführen. Sie weist

[622] Binding u. a. 2008, 282–284.
[623] Binding u. a. 2008, 286.
[624] Deicke 2016, Kapitel 1.
[625] Deicke 2016, Kapitel 3.

jedoch auf die positiven Effekte hin: Ein Vorteil besteht in der einfacheren Vermittlung und den entstehenden Diskurs über die Forschungsergebnisse. Zusätzlich ist die Datenmodellierung als eigenständiger Prozess eine Methode, um die Objekte besser zu verstehen, potenzielle Lücken oder Fehler zu erkennen und neue Ideen oder Sichtweisen zu ermitteln.[626]

Der Katalog ist eine breit angewandte Form der Publikation von archäologischen Ergebnissen. Dieser wird in unterschiedlichen Ausführungen publiziert. Enthalten sind oft detailliert beschriebene Objekte aus dem Ausgrabungskontext und Informationen zur Typologie in Form von bibliographischen Zitaten.[627] Katalogdaten haben andere Publikationsansprüche als beispielsweise Museumsdaten. Das eventbasierte Modell des CIDOC CRM kann für Katalogdaten verwendet werden, indem auf Kontexte und die Art und Weise, wie sie dokumentiert sind, geachtet wird. Das Datenmodell von A. Deicke geht von der Modellierung von Konzepten aus, die für die archäologische Forschung wichtig sind, wie bibliographische Angaben oder Fundkontexte. Dabei wird von einem zentralen Objekt als geschlossener Fund ausgegangen (wie beispielsweise ein Grab, Befund oder ein Fundobjekt). Diesem werden weitere Kontexte zugeordnet. Das zentrale Objekt kann Teil eines anderen Kontexts sein, wie eine Fundstelle, es kann andere Objekte enthalten, wie Knochen, Skelette oder Funde. Diesen können wieder Klassen zugeordnet werden wie Material oder Zustand.

Die Fundobjekte können durch einen Typ (*E55 Type*) typologisch klassifiziert werden. Mit diesem Typ können wiederum die zur wissenschaftlichen Analyse notwendigen bibliographischen Daten verbunden werden. Die Zuteilung von Perioden dient zum einen dem Zweck, Objekte einer Zeitspanne zuzuordnen, aber noch mehr der relativchronologischen Einordnung des Objektes. Es können im Zuge der Datierung auch regionale Gruppen bzw. Perioden zugeteilt werden, indem der Ort (*E53 Place*) angegeben wird. Zeitperioden können auch benannt werden (*E49 Time Appellation*) und absolut-chronologisch eingegeben werden. Angaben, die mit einem Dokument verknüpft sind, sind von Vorteil, um z.B. typenchronologische Entscheidungen nachvollziehbar machen zu können. Identifikatoren und Notizen können ebenfalls eingebaut werden.[628]

Dies zeigt eine der vielfältigen Möglichkeiten der Anwendung des CIDOC CRM für archäologische Daten. Weitere Überlegungen könnten die Beziehungen der Fundgegenstände zueinander, vielschichtige Befunde und Kulturkreiszuordnungen sein.[629]

[626] Deicke 2016, Kapitel 1.1.
[627] Deicke 2016, Kapitel 1.1.
[628] Deicke 2016, Kapitel 2.
[629] Deicke 2016, Kapitel 3.

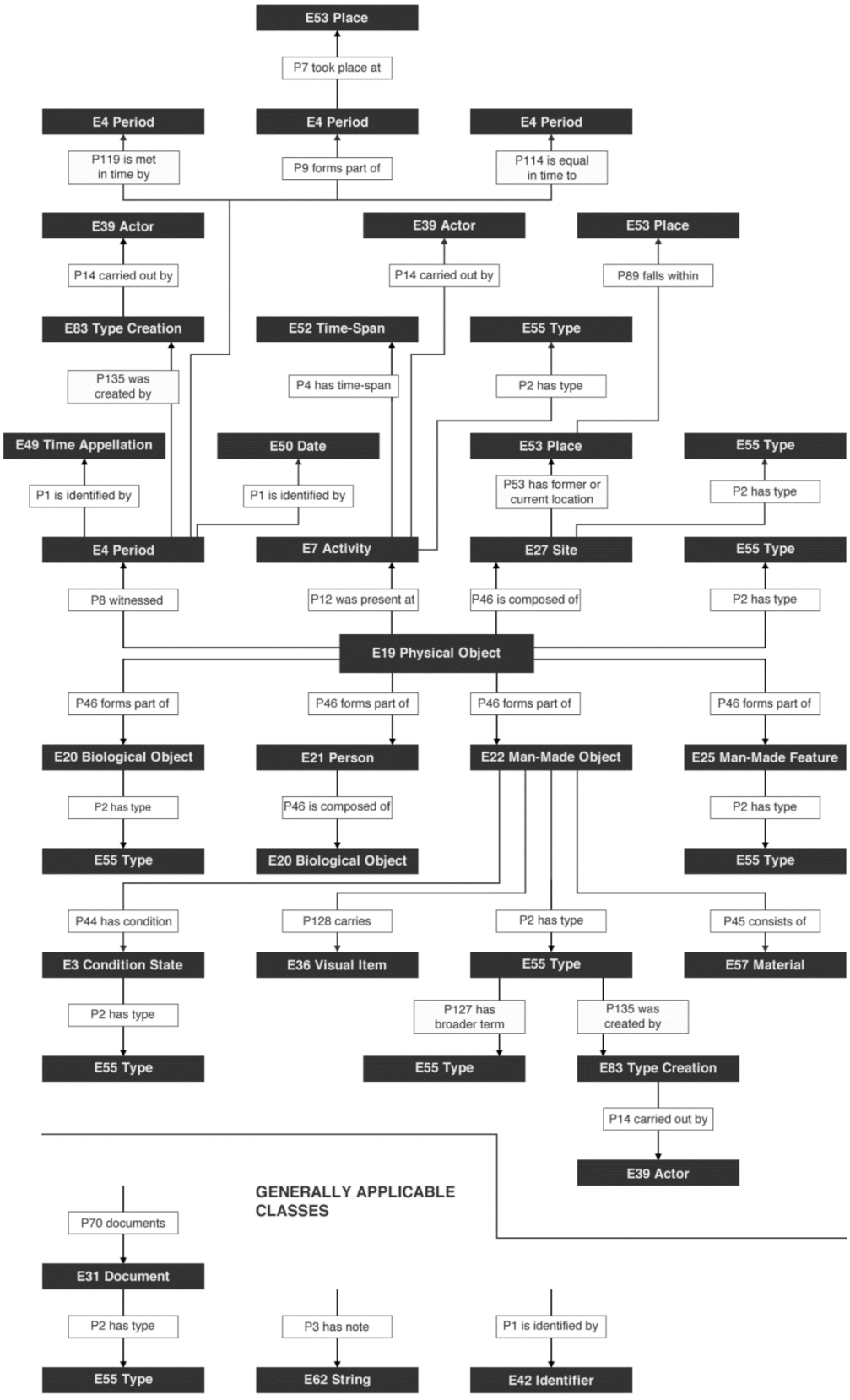


Abbildung 27 - Beispiel eines Objekts aus einem Katalog in einem CIDOC CRM basierten Datenmodell nach Aline Deicke (Quelle: Deicke 2016, Fig. 1).

5.2.4.3 Altdaten semantisch aufbereiten

Es gibt einige Möglichkeiten, um archäologische Altdaten bzw. bereits existierende Daten und Datenbanken semantisch aufzubereiten. Dabei nützt das Modellieren mit CIDOC CRM in erster Linie dem Informationsaustausch und der Integration zwischen unterschiedlichen bzw. heterogenen Daten. Auf diese Weise können die zahlreichen und für die Forschung wertvollen Daten in größeren Kontexten wiederverwendet werden.

M. Doerr u. a. haben Beispiel-Mappings von Datenbanken mit typisch archäologischen Datensituationen zu CIDOC CRM veröffentlicht und dabei das ARIADNE-Reference Model verwendet.[630] Sie erläutern, worauf geachtet werden sollte und welche Herausforderungen zu bewältigen sind. Dabei sind Werkzeuge notwendig. Sie verwenden dafür den 3M Mapping Memory Manager.[631,632] Ein anderes oft verwendetes Werkzeug ist das Online-Datenintegrations-Tool Karma[633].

Wenn die Daten zum Zielmodell transferiert werden, muss darauf geachtet werden, keine Informationen aus den Originaldaten zu verlieren. Oft sind Informationen versteckt oder nur implizit in der Datenbank enthalten. Dies könnte beispielsweise für den Fundort von Objekten in einer Datenbank passieren, wenn dieser nur als Datenbanktitel erhalten ist, weil die Datenbank nur für diesen Fundort gefertigt wurde. M. Doerr u. a. sind der Meinung, es werde daher auch immer ein* archäologische*r Expert*in benötigt und nicht nur die technischen Expert*innen und umgekehrt. Das Kombinieren von Computertechnologie und Wissen über semantische Technologien und der Archäologie ist wichtig.[634]

Das erste Beispiel-Mapping von M. Doerr u. a. zeigt eine Datenbank, welche sowohl Daten über die Objekte selbst, wie etwa Messdaten, Maße, Metallart und Zustand, als auch Kontext-Daten, die mit dem Objekt in Verbindung stehen, wie die Fundstelle, der archäologische Kontext und die Datierung, beinhaltet. Dieses relationale Datenbankschema soll in ein semantisches Modell umgewandelt werden. Gestartet wird mit der Feststellung des zentralen Objekttyps, welcher in der Originaldatenbank beschrieben wird. Dies können beispielsweise Funde, Befunde oder Fundorte sein. In diesem Fall sind dies Fundobjekte. Die Projektmitarbeiter*innen einigten sich für die Repräsentation der Fundobjekte auf die CIDOC CRM-Klasse *E22 Man-Made Object*[635,636].

---

630 Doerr u. a. 2016, 443.
631 Siehe hierfür: Oldman u. a. 2010.
632 Siehe unter: http://139.91.183.3/3M/, oder https://github.com/isl/Mapping-Memory-Manager, abgerufen am 21.10.2021.
633 Siehe https://usc-isi-i2.github.io/karma/, abgerufen am 21.10.2021.
634 Doerr u. a. 2016, 445.
635 Die Definitionen zu allen CIDOC CRM Entitäten der Version 5.0.4 sind im Dokument Crofts u. a. 2011 aufgelistet.
636 Die Entität ist in der neueren Version 6.2.9 unter E22 Human-Made Object zu finden.

In einem weiteren Schritt werden die Spaltennamen der Datenbank mit den Eigenschaften (engl. *properties*) gleichgestellt. Sie stellen die Beziehungen zwischen den Einheiten dar. Die Feldinhalte stehen für die Entitäten. Ein Beispiel hierfür wäre der Identifikator. In der Funddatenbank gibt es eine Spalte mit „Fundnummer". Die Feldinhalte führen die Fundnummern auf (siehe Abbildung 28).

Fundobjekt	Fundnummer	Gewicht in Gramm
Ring (*E22 Man-Made Objekt*)	1234 (*E42 Identifier*)	2 (*E60 Number*)

Abbildung 28 – Beispieldaten aus einer Funddatenbank (Grafik: M. Simon).

Der Ring ist das Subjekt, die Fundnummer das Objekt. Die Beziehung zwischen diesen beiden würde lauten: *P1 is identified by*. Die Aussage würde folgendermaßen lauten: *E22 Man-Made Object – P1 is identified by – E42 Identifier*. Der Ring wird identifiziert durch die Nummer „1234".

Es ergeben sich jedoch meist keine einfachen eins-zu-eins Beziehungen, wie im vorhergehenden Beispiel. Ein Feld der Datenbank wird dann zu einem komplexen Pfad von Informationen im Zielschema. Ein Beispiel hierfür ist das Gewicht oder die Maße eines Fundobjektes. Die CIDOC CRM-Klasse *E54 Dimension* hat zwei Eigenschaften: *P90 has value* (hat den Wert): *E60 Number*, *P91 has unit* (hat die Einheit): *E58 Measurement Unit*.[637] Der Pfad zur Information „Der Ring hat 2 Gramm Gewicht" führt zu einem komplexen Pfad, der folgendermaßen aussehen könnte:

E22 Man-Made Object – P43 has dimension – E54 Dimension
E54 Dimension – P91 has unit – E58 Measurement Unit = „Gramm"
E54 Dimension – P2 has value – E60 Number = „2"

Beim zweiten Beispiel-Mapping des Artikels von M. Doerr u. a. handelt es sich um eine Datenbank über ein urnenfelderzeitliches Gräberfeld. Die Gräber beinhalten meist eine Urne mit menschlichen Überresten, Tierknochen und Fundobjekte aus Kupferlegierungen oder Keramik. Die Datenbank enthält Informationen über die Befunde, die Funde und die Analyseresultate der anthropologischen Untersuchungen der menschlichen Überreste. Die Ausgrabung wird hier als *E7 Activities* (Aktivitäten) gemappt (siehe Abbildung 29). Diese Aktivitäten bestehen aus mehreren *S19 Encounter Events* (Begegnungsevents, aus der Erweiterung des CIDOC CRM CRMsci), die wiederum zu jedem Grab verlinkt werden. Das Grab gilt als *E25 Man-Made-Feature*, und wird zu allen Objekten (wie beispielsweise Keramik*, E25 Man-Made-Object*), die im Grab gefunden wurden, verlinkt. Die Information über die Dekoration auf den Keramikgefäßen wurde als *E26 Physical*

[637] Doerr u. a. 2016, 445–449.

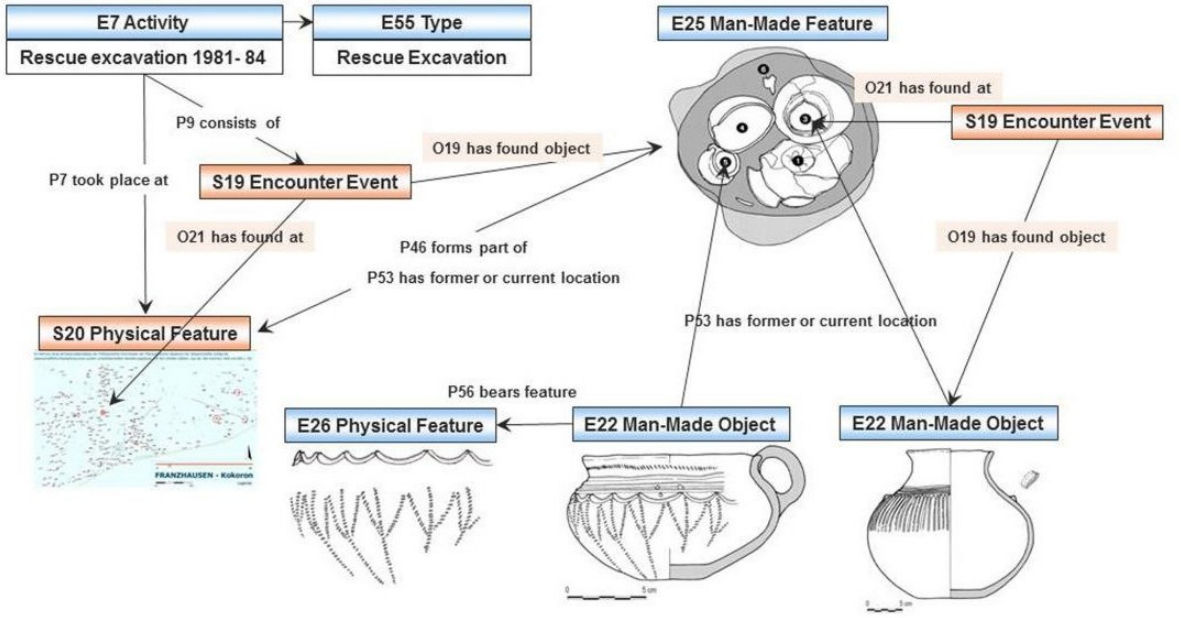


Abbildung 29 - Ausschnitt des semantischen Graphen des Gräberfeldes Franzhausen Kokoron bei M. Doerr u. a. (Quelle: Doerr u. a. 2016, 450).

Feature (physikalisches Merkmal) zur jeweiligen Keramik mittels *P56 bears feature* (trägt Eigenschaft) gemappt.[638]

Die Vorteile des Mappings mit CIDOC CRM sehen M. Doerr u. a. vor allem darin, sehr fachspezifische, unterschiedliche Datenbanken ohne Verlust von Details oder Bedeutung zu einem semantischen Modell bringen zu können. Es stellt einen Leitfaden für Datenstrukturen dar, die sich bewährt haben. CIDOC CRM bietet ein semantisches Wissensnetz für den kulturhistorischen Bereich. Daten können besser mit anderen Daten verglichen werden. Der Erkenntnisgewinn wird durch das Mapping selbst, die Mustererkennung (z.B. Gräber von Altersgruppen in einer spezifischen Position im Grab) und die Definition der Entitäten gestärkt. Außerdem können statistische Studien und Datenanalysen einfacher vorgenommen werden. CIDOC CRM hilft dabei, Altdaten zu erhalten, wiederzuverwenden, zukünftig zu erweitern und zu integrieren.[639]

5.2.4.4 Archäologische Ausgrabungen und CIDOC CRM

Bei Daten von Ausgrabungen und Felduntersuchungen kann das CIDOC CRM helfen, unterschiedliche Datenstrukturen zu harmonisieren und Daten auszutauschen. Doch es gibt einige Herausforderungen zu bewältigen. A. Masur, K. May u.a. haben diese genauer betrachtet. Dafür haben sie nach Gemeinsamkeiten der Ausgrabungsdaten hinsichtlich der verwendeten Konzepte

[638] Doerr u. a. 2016, 449 f.
[639] Doerr u. a. 2016, 450.

und Begriffe gesucht und die unterschiedlichen Methoden der Dokumentation und Analyse erfasst. Dabei seien das CIDOC CRM und dessen Erweiterung CRMarchaeo[640] geeignet, um unterschiedliche Ausgrabungsdaten von verschiedenen Fundstellen zu vergleichen und zu analysieren und dies ohne den Verlust von Zusammenhängen oder Bedeutungen.[641]

Die Schwierigkeiten für die Integration der Daten betreffen vor allem die unterschiedlich verwendeten Konzepte, Begriffe und Vokabulare. Ausgrabungen und die dahinterstehenden Prozesse werden durch unterschiedliche Methoden dokumentiert. Einzelne Länder haben sich unterscheidende standardisierte oder übliche Vorgehensweisen für die Aufnahme erarbeitet. In Österreich gilt beispielsweise der Dokumentationsstandard für Ausgrabungen nach den Richtlinien des Bundesdenkmalamtes[642].[643]

Die verschiedenen Dokumentationsweisen teilen sich jedoch einige konzeptuelle Elemente. Das Problem hierbei ist, dass es oft unterschiedliche Bedeutungen bzw. Synonyme für dieselben Begriffe und Konzepte gibt. Oder aber es gibt ähnliche Konzepte oder Bedeutungen, für welche allerdings andere Begriffe bzw. Homonyme verwendet werden. Oft werden Konzepte nicht ausreichend definiert oder die Ausgrabungsmethoden, Dokumentationsweisen oder Hinweise auf Interpretationen werden nicht explizit angeführt. Die Unterscheidung zwischen einer Beschreibung oder einer Interpretation wird ebenfalls als problematisch angeführt.[644] Archäologische Ausgrabungen haben drei Kategorien gemeinsam:

- raumbezogene Daten
- Informationen über stratigraphische Einheiten
- Informationen über Fundgegenstände

Die raumbezogenen Daten beziehen sich auf die Ortsangaben der Fundstelle. Meist kommen noch zusätzliche Unterteilungen der Fundstelle hinzu und künstlich erzeugte Grabungseinheiten, wie Raster, Grabungsschnitte, Teile von Befunden etc., werden angeführt. Die raumbezogenen Daten befinden sich auf unterschiedlichen Detailebenen.[645] Die Informationen über stratigraphische Einheiten können sehr vielseitig sein. Es werden grob Deposits und (Feature-)Interfaces als Stratigraphie-Einheiten unterschieden und diese werden unterschiedlich in der Dokumentation erfasst. Bei Deposits werden die Farbe, Textur und der Typ der Erde sowie die Zusammensetzung

[640] Doerr u. a. 2018a.
[641] Masur u. a. 2014, 1-2.
[642] Siehe hierzu die Standards, Leitfäden und Richtlinien des Bundesdenkmalamtes unter https://bda.gv.at/de/publikationen/standards-leitfaeden-richtlinien/, abgerufen am 21.10.2021.
[643] Masur u. a. 2014, 2.
[644] Masur u. a. 2014, 2-8.
[645] Masur u. a. 2014, 3.

und Einschlüsse beschrieben. Bei (Feature-)Interfaces wird die Form, die Form der Kontur, die Seiten und der Boden beschrieben und zusätzliche Messungen der Länge, Breite und Tiefe durchgeführt. Weiters sind zusätzliche Informationen über die relative Position und die Abfolge der Schichten bzw. stratigraphischen Einheiten wichtig. Aber auch wie die Beziehungen der stratigraphischen Einheiten zueinander beschrieben werden, wird dokumentiert. Es wird festgestellt, ob eine Einheit eine andere schneidet oder füllt, ob sie die Position über oder unter einer anderen Einheit einnimmt und Gruppierungen von Einheiten zu einem größeren archäologischen Befund, wie beispielsweise eine Gruppe von Pfostenlöchern, welche in der Gesamtheit einen Hausgrundriss darstellen, werden vorgenommen.[646] So können Parallelen zu anderen Fundstellen einer spezifischen Zeitstellung gefunden werden. Fundgegenstände werden meist separat dokumentiert. Wobei auch in dieser Kategorie zusätzliche Informationen zur Situation der Bergung aufgenommen werden. Wichtig ist hier beispielsweise die Position, die Orientierung, der Zustand, das Material und die Zugehörigkeit zu einer stratigraphischen Einheit der Funde.[647] In weiterer Folge werden bei der Aufarbeitung weitere Informationen, wie die typo-chronologische Einordnung oder Ergebnisse von naturwissenschaftlichen Datierungsmethoden oder Analysemethoden hinzugefügt.

Ontologien wurden entwickelt, um die gemeinsamen Konzepte zu beschreiben, die unterschiedlichen Schemata zugrunde liegen.[648] Die Schemata der Ausgrabungen können die drei oben genannten Konzepte beschreiben. P. Cripps u. a. betonen die Archäologiefreundlichkeit der eventzentrierten Ontologie CIDOC CRM. Alle physikalischen Objekte oder stratigraphischen Einheiten werden durch Events, die entweder in archäologischer Zeit oder in der Zeit der archäologischen Ausgrabung geschehen, zu Zeitspannen, Akteuren etc. verlinkt.[649] Hier sehen auch A. Masur, K. May u. a. den Vorteil des Modells für archäologische Ausgrabungen. Stratigraphische Einheiten wurden bei Events in der Vergangenheit kreiert und verändert. Bei Events wie Störungen oder Ausgrabungen werden sie zerstört. Diese Events wiederum können jeweils Zeitspannen zugeordnet werden.[650]

Das CRMarchaeo, eine Erweiterung des CRM (siehe dazu Kapitel 5.2.2), bietet neue Klassen, um die einer archäologischen Ausgrabung zugrundeliegenden Konzepte miteinbeziehen zu können.

[646] Masur u. a. 2014, 3-4.
[647] Masur u. a. 2014, 4.
[648] Doerr 2009, 467.
[649] Cripps u. a. 2004, 11 f.
[650] Masur u. a. 2014, 8.

Es wurde im Zuge des ARIADNE Projektes entwickelt, welches sich zum Ziel gesetzt hat, archäologische Ausgrabungsdaten besser integrieren zu können.[651] Die neueste Version 1.5 des CRMarchaeo ist durch das Projekt ADED (Archaeological Digital Excavation Documentation)[652] entstanden. In diesem Projekt werden 1200 Ausgrabungsprojekte in eine CRMarchaeo-Repräsentation umgewandelt.[653]

Im CRMarchaeo gibt es die *A8 Stratigraphic Unit* (stratigraphische Einheit) und die *A1 Excavation Process Unit* (Ausgrabungsprozess-Einheit).[654] Die *Stratigraphic Unit* bezeichnet die ungestörte Stratigraphie vor einer Ausgrabung, welche von einem *A4 Stratigraphic Genesis* Event kreiert wird. Dabei werden die *Stratigraphic Units* und die *Stratigraphic Genesis* Events getrennt dokumentiert und den Events werden Zeitspannen zugeordnet. Durch die zeitlichen Beziehungen zwischen den *Stratigraphic Units* kann von Archäolog*innen die stratigraphische Matrix eingebunden werden.

Der Vorteil am übergeordneten Begriff *Stratigraphic Unit* ist, dass alle in der Archäologie gebräuchlichen Synonyme, wie Befund, Kontext, Schicht, etc., damit bezeichnet werden können. Ob eine *Stratigraphic Unit* ein Deposit oder Interface bezeichnet, wird durch zusätzliche Klassen definiert. Die *Excavation Process Unit* beschreibt das Event der Ausgrabung einer *Stratigraphic Unit* durch Archäolog*innen. Auch die *Excavation Process Unit* kann einer Zeitspanne zugeordnet werden, wie beispielsweise einem Arbeitstag oder einem Jahr. Zu dieser Aktivität der *Excavation Process Unit* können in weiterer Folge auch beispielsweise das Ausgraben einer *Stratigraphic Unit* oder das Finden von Fundgegenständen verlinkt werden.

Da die Events in der Vergangenheit und jener der Archäolog*innen in heutiger Zeit differenziert werden, können Arbeitsprozesse zielgerichtet dokumentiert werden und so auch die verwendeten Methoden festgehalten werden. Ein weiterer Vorteil des Mapping-Prozesses ist, eine Grabungsdokumentation muss nur einmal gemappt werden. Dies stellt zwar zunächst einen zeitlichen Mehraufwand dar, allerdings kann die Dokumentation zukünftig interoperabel zur Verfügung gestellt werden und mit anderen gemappten Ausgrabungsdaten durchsucht und analysiert werden.[655]

Als Anwendungsbeispiel des CRMarchaeo kann das Projekt „A Puzzle in 4D“[656] genannt werden. In diesem Projekt wurden große Teile das Archivs der seit 1966 ausgeführten Ausgrabung in Tell el-Daba, Ägypten, digitalisiert, für die Langzeitarchivierung vorbereitet und semantisch aufberei-

[651] Niccolucci – Richards 2013, 85 f.
[652] Zum Projekt (Laufzeit 2018-2025) siehe: https://www.khm.uio.no/forskning/prosjekter/archaeological-digital-excavation-documentation/, abgerufen am 21.10.2021.
[653] Uleberg u. a. 2020, 1.
[654] Doerr u. a. 2018b, 13–18.
[655] Masur u. a. 2014, 8-9.
[656] Siehe https://4dpuzzle.orea.oeaw.ac.at/index/, abgerufen am 21.10.2021.

tet. Auch die digitale Dokumentation wurde aufbereitet. Dies soll die Menge an heterogenen Daten zukünftig für andere Forscher*innen zugänglich und integrierbar machen.[657] Es handelt sich hier um die Aufarbeitung einer Altgrabung, bei der ein Ausgrabungssystem mit Plana verwendet wurde. Für das Datenmodell der Ausgrabungen, der Dokumentation und des Digitalisierungsprozesses wurde das CIDOC CRM und die Erweiterungen CRMarchaeo, CRMsci und CRMdig verwendet. Dabei wurden drei Hauptklassen identifiziert, welche im Datenmodell verarbeitet wurden, *Physical Things (E18), Events (E5)* und *Information Objects (E73)* (siehe Abbildung 30):[658]

- *Physical Things (E18)* stehen für Dinge, welche physikalisch fassbar sind und entweder in der Vergangenheit oder während den Ausgrabungen kreiert wurden. Sie werden in Subkategorien aufgeteilt, die Ausgrabungseinheit (*S20 Rigid Physical Feature*; physikalische Strukturen, die während der Ausgrabung kreiert wurden), die archäologische Einheit *(S20 Rigid Physical Feature*; Strukturen, welche durch menschliches Einwirken in der Vergangenheit entstanden sind) und archäologische Fundobjekte (*E19 Physikal Objekt*). Jedem *Physical Thing* (E18) kann ein *E55 Type* zugeordnet werden, um die Art des physikalischen Objektes weiter zu spezifizieren. Ein Ausgrabungsareal kann beispielsweise den Typ „Areal" erhalten. *Events (E5)* sind *Activities (E7)* und können *Physical Things (E18)* beeinflussen. Die weiteren Kategorien sind die Ausgrabungsaktivität (*A1 Excavation Process Unit*; alle Aktivitäten während einer Ausgrabung, wie beispielsweise das Entfernen von Materie oder das Kreieren von Ausgrabungseinheiten), die Dokumentationsaktivität (*E65 Creation*; Aktivitäten, welche Dokumentation erstellen) und die digitalen Aktivitäten (*D7 Digital Machine Event*; Aktivitäten mit Geräten, welche digitale Dokumente erstellen, wie die Digitalisierung von analoger Dokumentation).
- *Information Objects* (E7) können weiter unterteilt werden in analoge Objekte (*E31 Document*) und digitale Objekte (*D1 Digital Object*), welche beide die Dokumentation der Ausgrabungen beschreiben.

Die Beziehungen zwischen den Ausgrabungseinheiten (*E18 Physical Thing/S20 Rigid Physical Feature*) wurden mit der Eigenschaft *P89 falls within* ausgedrückt (siehe Abbildung 31). Der Schnitt (F-I_j21) liegt im Grabungsareal (F-I) und das Planum (F-I_j21_Planum3) befindet sich innerhalt des Schnittes. Das Grab 8 (TD_F-I_j21_Tomb8) wird mit der Eigenschaft *P53 has former or current location* dem Planum 3 zugewiesen.[659]

[657] Aspöck u. a. 2020, 86.
[658] Hiebel u. a. 2021b, 4–8.
[659] Hiebel u. a. 2021b, 8.

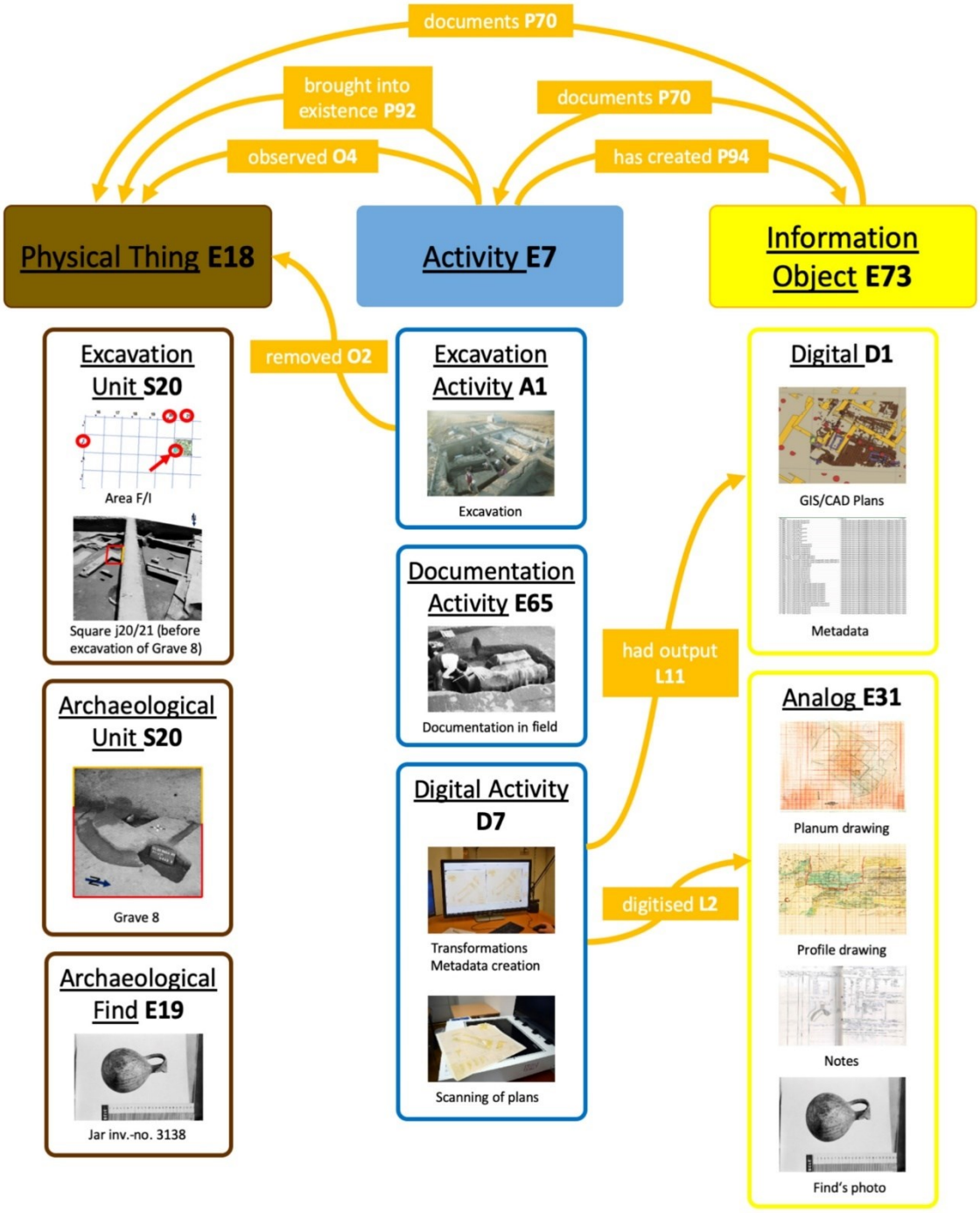


Abbildung 30 - CIDOC CRM-Klassen des Datenmodells für die Dokumentation. (Quelle: Hiebel u. a. 2021b, 7, Figure 6, Illustration von Birgit Danthine, Gerald Hiebel und Edeltraud Aspöck).

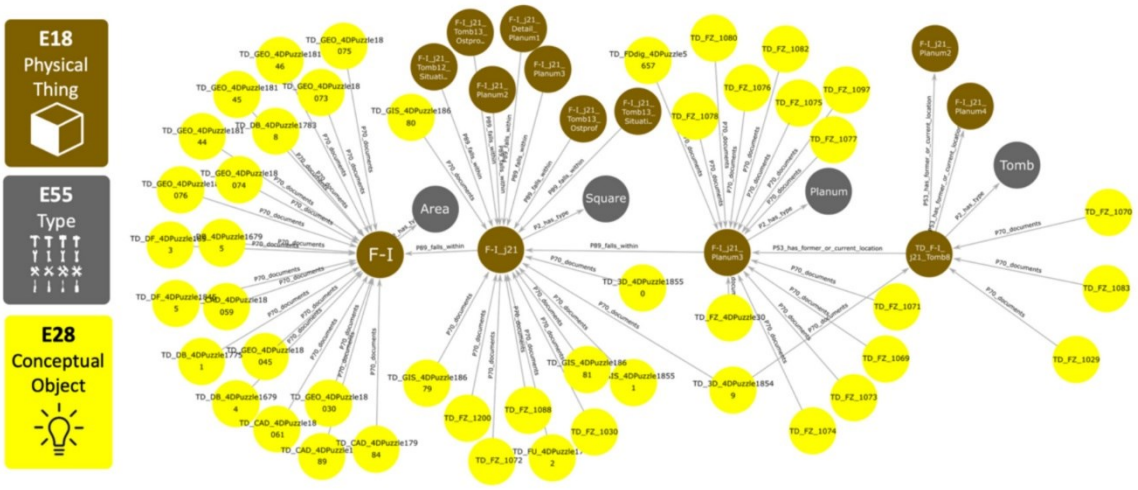


Abbildung 31 - Ausschnitt der Repräsentation der Ausgrabungs- und Dokumentationsobjekte. (Quelle: Hiebel u. a. 2021b, 8, Figure 7, Illustration von Birgit Danthine und Gerald Hiebel).

5.2.4.5 Archäologische Datierungen und Zeitperioden implementieren

Archäologische Daten beinhalten meist Datierungen der Fundobjekte bzw. Befunde von Ausgrabungen. Diese nach der Ontologie CIDOC CRM zu mappen, benötigt einige Vorarbeit. In Datenbanken können die Formate, die die Datierungen angeben, variieren. Bei Angaben mit absoluten Zahlen sind es eher angegebene Zeitspannen als absolute Datumsangaben. In Textform finden sich Formate wie das Jahrhundert oder das Jahrtausend, v. Chr./n. Chr. mit Jahresangabe oder Perioden und Kulturgruppenbezeichnungen.

Zuerst muss eine Datenbereinigung durchgeführt werden. In einer Spalte bzw. Kategorie einer einzelnen Datenbank sollen keine unterschiedlichen Formate zu finden sein oder uneinheitliche Datierungsangaben pro Eintrag. Für die Periodenangabe sollten möglichst kontrollierte Vokabulare[660] bzw. Definitionen, die Identifikatoren für Zeitperioden liefern und Definitionen der geographischen und zeitlichen Einordnung bieten, verwendet werden (siehe Kapitel 6). In CIDOC CRM gibt es *E2 Temporal Entity* mit der Subklasse *E4 Period* und relevanten Properties, um sowohl Datierungen als auch relativ-chronologische Zusammenhänge darstellen zu können.[661] Auch die Einteilung in *E92 Spacetime Volumes* kann hergenommen werden. Mit ihr können sowohl zeitliche wie auch geographische Anordnungen zugeteilt werden.[662] L. Isaksen empfiehlt, absolute Datierungen anzugeben, da diese über große Datenmengen besser digital visualisiert werden können.

[660] Beispielsweise Periodo, siehe https://perio.do/en/, abgerufen am 21.10.2021.
[661] Binding 2010, 280–285.
[662] Siehe hierzu auch die Ontologie des iDAI.chronontology in Kapitel 6.2.

Um unsichere Datierungen bestmöglich eingrenzen zu können, gibt es in CIDOC CRM die Möglichkeit Zeitspannen (*E52 Time Span*) anzugeben.[663]

5.2.4.6 CIDOC CRM im Zusammenhang mit räumlichen Daten

Geometrische Daten von archäologischen Untersuchungen, wie beispielsweise GIS-Daten, spielen eine wesentliche Rolle in der Archäologie. Im Carare Projekt[664] waren geographische Daten von Monumenten und Gebäuden bereits im Fokus. Es wurde ein Ortsverzeichnis entwickelt, in welchem die Koordinaten der Monumente des Europeana Projekts[665] eingebunden wurden. So konnten Geodaten auf der Ebene der Fundstelle in das Datenmodell des Europeana Projekts modelliert werden.[666]

Im Zuge des Spezialforschungsbereichs HiMAT[667] der Universität Innsbruck wurden die gemessenen Geodaten innerhalb der Fundstellen detailliert eingebunden.[668] Der multidisziplinäre Spezialforschungsbereich erforscht die Geschichte des prähistorischen und historischen Bergbaus in den Ostalpen. Aus den zehn verschiedenen Disziplinen wurden Daten generiert, welche in einem Datenmodell in CIDOC CRM modelliert wurden.[669] Um die Geoinformationen in die ontologischen Modelle einzubinden und sie in den Kontext der anderen Forschungsdaten zu stellen, wurde das CRMgeo entwickelt (siehe 5.2.2), eine Erweiterung des CIDOC CRM.[670] Es bietet die Möglichkeit, Daten, welche mit CIDOC CRM modelliert sind, mit GIS Daten zu integrieren. Das Open Geospatial Consortium (OCG) bietet Standards für geografische Informationen, darunter die GeoSPARQL-Abfragesprache[671] („A Geographic Query Language for RDF Data“) für Geodaten. Das CRMgeo integriert diese beiden ontologischen Modelle, das CIDOC CRM mit dem OCG-Standard, indem zwei Konzepte für Orte neu hinzugefügt wurden; der *Phenomenal Place* und der *Declarative Place*. Der *Phenomenal Place* (im CRMgeo *SP2*) repräsentiert den realen Ort, an dem Events stattgefunden haben. Er wird angenommen durch Objekte aus der realen Welt und der *Spacetime Volume*s (im CIDOC CRM *E92*), die sie einnehmen.[672,673] Eine *Spacetime Volume* besteht

[663] Isaksen 2011, 113 f.
[664] Siehe https://pro.carare.eu/en/, abgerufen am 21.10.2021.
[665] Siehe https://www.europeana.eu/de, abgerufen am 21.10.2021.
[666] Fernie 2013, 88.
[667] Aus dem Spezialforschungsbereich wurde das Forschungszentrum HiMAT gegründet, siehe https://www.uibk.ac.at/himat/index.html.de, abgerufen am 21.10.2021.
[668] Hiebel u. a. 2014, 559.
[669] Hiebel – Hanke 2009, 653 f.
[670] Hiebel u. a. 2015a, 4 f.
[671] Perry – Herring 2012.
[672] Hiebel u. a. 2015b, 303–305.
[673] Als Beispiel für einen Phenomenal Place würde der Ort, an dem Caesar ermordet wurde, gelten. Hiebel u. a. 2015a, 8.

aus einem vierdimensionalen Punkteset, welches den Ort und die Zeit mit einbindet.[674,675] Ein *Declarative Place* (im CRMgeo *SP6*) hingegen wurde durch Menschen definiert und beispielsweise mittels Koordinaten eingegrenzt und nähert sich an den *Phenomenal Place* an.[676]

Im HiMAT Projekt wurden unter anderem Daten modelliert, welche von Forschungsaktivitäten stammen, die geologische oder zeitliche Informationen über bronzezeitliche Minenaktivitäten der Fundstelle Mauken bieten. Um sich den realen Orten, den *Phenomenal Places*, an denen diese Aktivitäten stattgefunden haben, anzunähern, werden die *Declarative Places*, die gemessenen räumlichen Ausdehnungen, verwendet. Das gleiche Prinzip wird auch für *Phenomenal* und *Declarative Time Spans* und *Phenomenal* und *Declarative Spacetime Volumes* angewandt (siehe Abbildung 32).

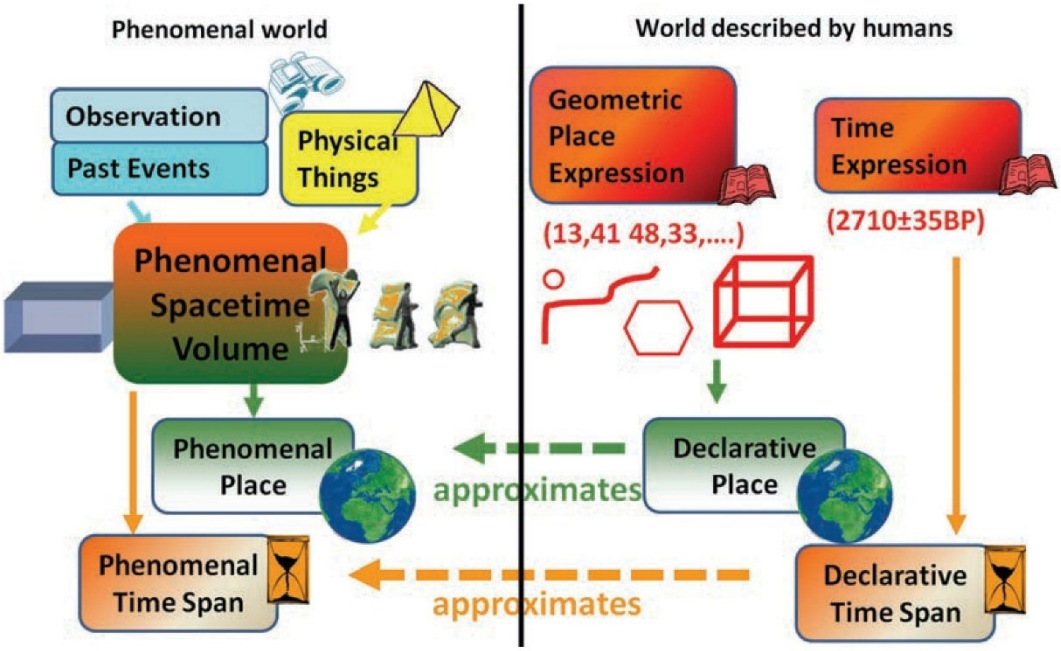


Abbildung 32 - Beziehungen der Phenomenal Spacetime Volume und Declarative Places (Quelle: Hiebel u. a. 2014, 564, Figure 3).

Die Modellierung von Prospektionsaktivitäten der Forscher*innen ergibt mehrere *Declarative Places* von prähistorischen Bergbauaktivitäten, welche gemessen wurden. Eine vermutete Mine in einer Felsformation namens Moosschrofen wurde in den 1980er Jahren prospektiert und in einem LIDAR Scan mit Koordinaten eingegrenzt. 2012 wurde die Mine mit GPS-Punkten vor Ort eingemessen. Diese Punkte kreieren einen weiteren *Declarative Place* für die Mine. Diese beiden *Declarative Places* können in Beziehung zueinander einen *Phenomenal Place* bilden. Diese Verbindung erlaubt

[674] Bekiari u. a. 2021, 108.
[675] Das Event von Caesars Ermordung mit seiner Ausdehnung in Raum und Zeit gilt als Beispiel für eine Spacetime Volume. Bekiari u. a. 2021, 108.
[676] Hiebel u. a. 2015b, 303–305.

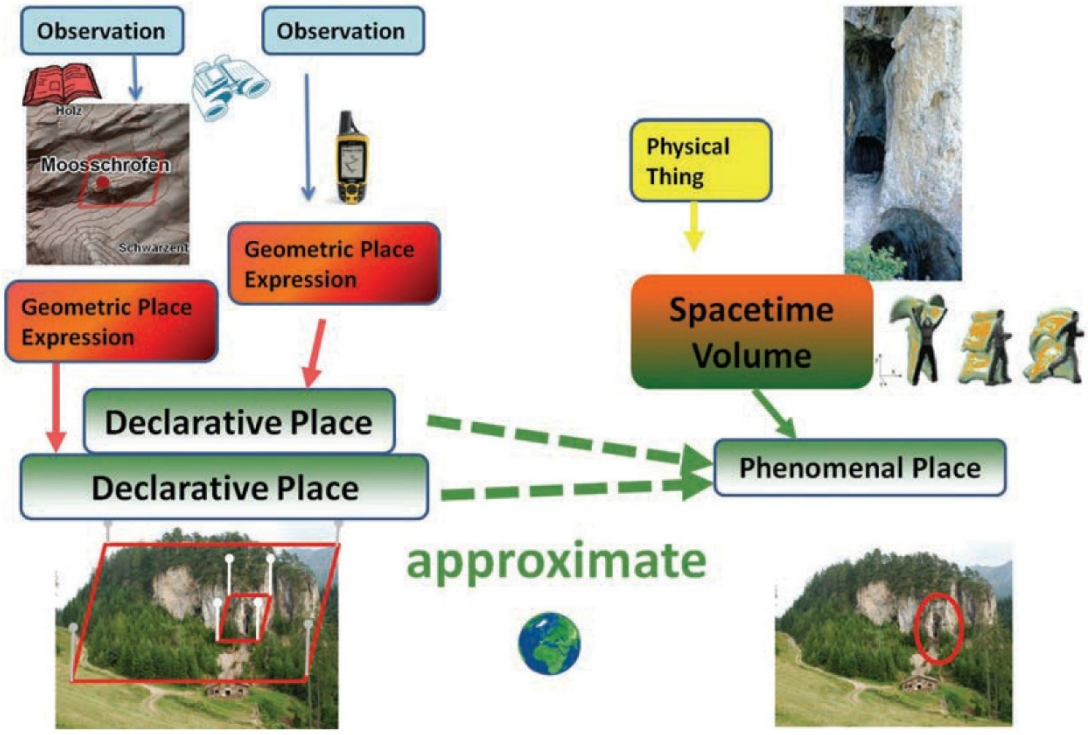


Abbildung 33 - Darstellung der Prospektionsaktivitäten (Quelle: Hiebel u. a. 2014, 566, Figure 4).

nun im ontologischen Modell die jeweiligen Informationen der Prospektionen miteinander in Beziehung zu setzen (siehe Abbildung 33).[677]

Archäologische Ausgrabungen in der Region Mauken fanden an verschiedenen Orten, zu verschiedenen Zeiten und mit unterschiedlichen Forschungsfragen statt und zeigen jeweils andere Stadien der Erzverarbeitung. Zur Modellierung dieser Daten wurde neben der Erweiterung des CIDOC CRM CRMgeo auch die Erweiterung CRMarchaeo[678] verwendet. Um prähistorische Aktivitäten nachweisen zu können, werden die Überreste in den stratigrafischen Schichten herangezogen, welche im CRMarchaeo als *Stratigraphic Units* dargestellt werden. Informationen über die Vergangenheit stammen aus diesen *Stratigraphic Units* und ihrer Beziehung zueinander. Forschende im Bereich der Archäologie nähern sich an die Grenzen dieser Schichten an und dokumentieren sie. Doch die Ausdehnungen werden nur annähernd erfasst, so wie beispielsweise manchmal nur Teile der Schichten ausgegraben werden, wenn Profile angelegt werden. Es wird bei der Modellierung daher zwischen den *Stratigraphic Units* als *Feature* und der Arbeit der Archäolog*innen unterschieden. Da das CIDOC CRM ein eventzentriertes Modell ist, wird die Arbeit der Forschenden als Event, die *Excavation Process Unit*, modelliert. Gefundene Objekte in den Schichten werden in diesem Ansatz zur *Excavation Process Unit* verlinkt, in der sie gefunden wurden. In Abbildung

[677] Hiebel u. a. 2014, 565 f.
[678] Doerr u. a. 2018a.

34 wird dargestellt, wie die Bergung eines Holztroges zur Aufbereitung des Kupfererzes modelliert wird. Bei einer *Excavation Process Unit* wurde der Trog in der *Stratigraphic Unit* Nummer drei gefunden und in der Fundsituation C dokumentiert. Das Profil zeigt die *Stratigraphic Units* über und unter der *Stratigraphic Unit* Nummer drei. Das Event in der Vergangenheit ist die Aufbereitungsaktivität an einem *Phenomenal Place*, welchem sich angenähert wird durch die *Declarative Places* der Bergung des Troges, welcher als *Physical Object* verlinkt wird.[679] Durch die Annäherung der Datierung des Trogs mithilfe der Dendrochronologie wird die zeitliche Komponente miteinbezogen.

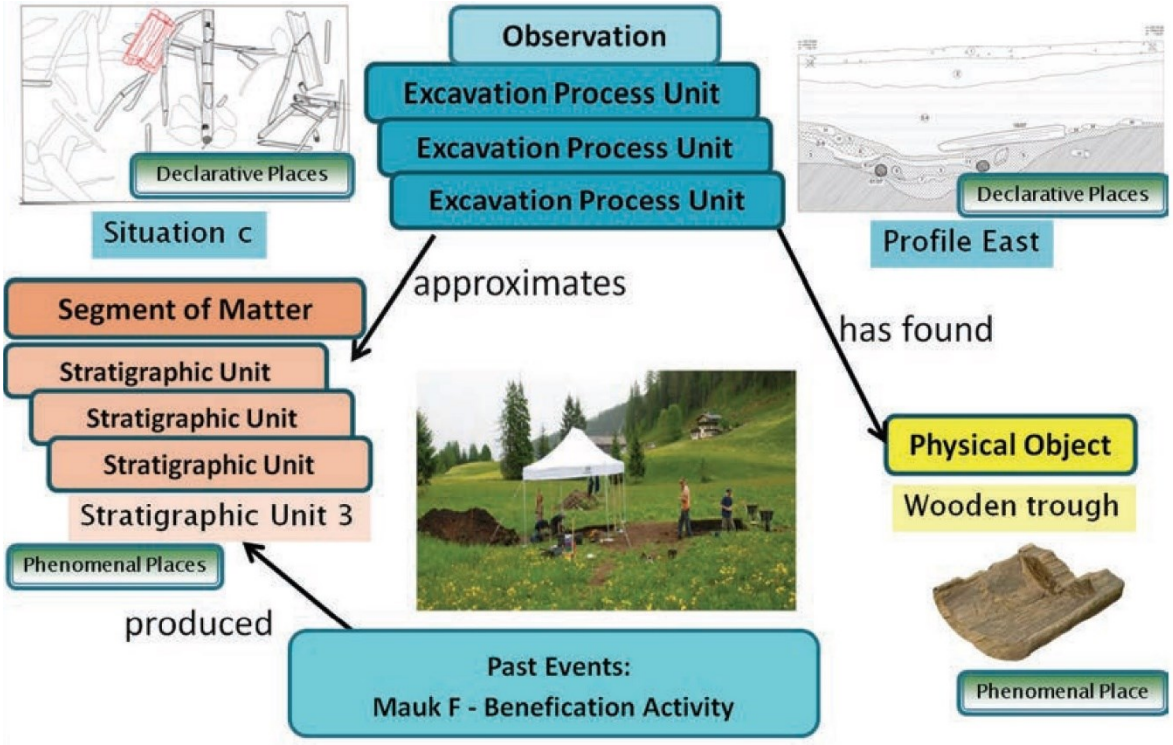


Abbildung 34 - Modellierung der Bergung eines Holztroges mit CRMarchaeo (Quelle: Hiebel u. a. 2014, 568, Figure 5).

5.2.4.7 Naturwissenschaftliche Daten mit CIDOC CRM

Daten, welche durch naturwissenschaftliche Datierungsmethoden gewonnen werden, wie C14-Datierungen oder Dendrochronologiedaten, können mithilfe der CRMsci[680] Erweiterung des CIDOC CRM modelliert werden (siehe Abbildung 35). Auch hier werden die Aktivitäten bzw. Events herangezogen, um das Modell zu generieren. Die Aktivität der Probenentnahme des soeben erwähnten Holztrogs dient zur Bestimmung des Alters durch das Zählen und Abgleichen der Baumringe mit anderen Sequenzen von Baumringen. Die Probenentnahme ist eine *Observation*, die Probe wurde entnommen aus dem *Physical Object*. Die Aktivität der Zählung der Baumringe

[679] Hiebel u. a. 2014, 566–568.
[680] Doerr u. a. 2020b.

wird als *Inference Making* modelliert, da es sich hier um eine Kalkulation handelt. Die Abgleichung mit anderen Baumringproben wird mit der Klasse *Design or Procedure* beschrieben. Die Bestimmung der Zeitspanne (*Timespan)* wird ebenfalls als *Design or Procedure* modelliert, welche auch die *Declarative Timespan* mit absoluten Zeitangaben definiert und einen Terminus Postquem, den frühesten Zeitpunkt einer Datierung, gibt.[681]

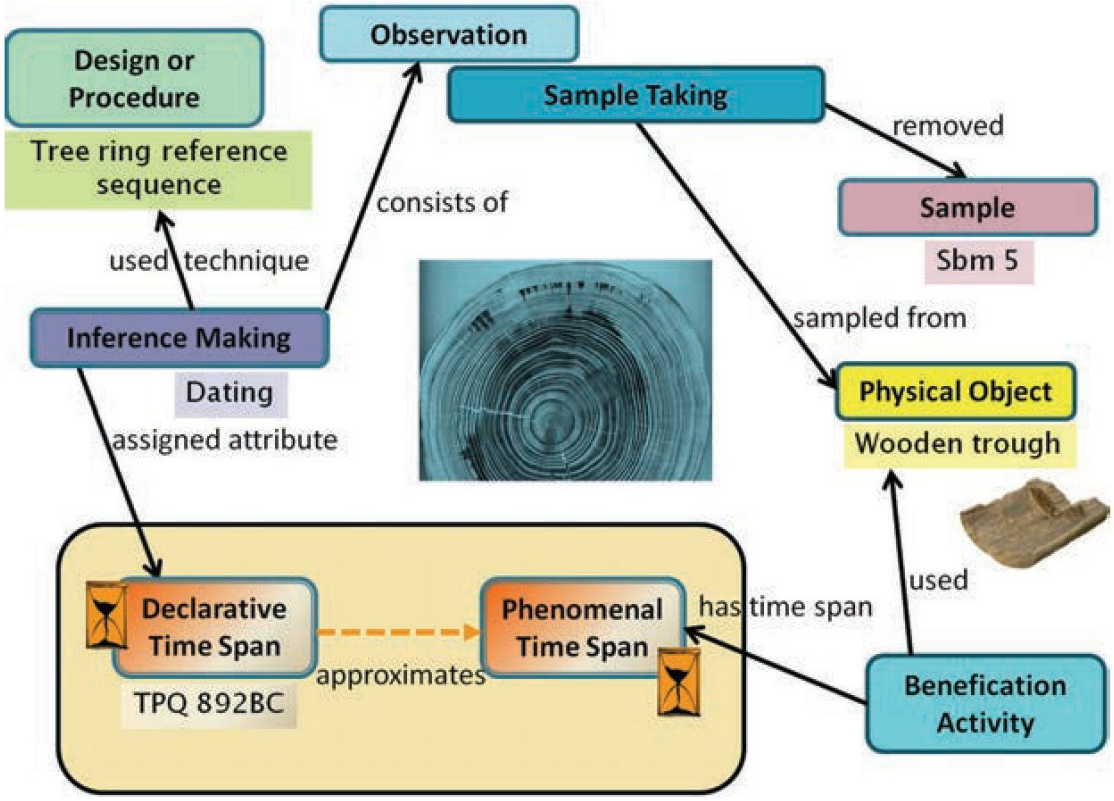


Abbildung 35 - Modellierung von dendrochronologischen Daten eines Holztroges (Quelle: Hiebel u. a. 2014, 572, Figure 9).

5.2.4.8 Vorlagenbasierte semantische Integration

Um semantisch angereicherte Daten zu erhalten und sie zu einem Teil des Linked Data Netzes oder Semantic Webs zu machen, mussten bisher üblicherweise immer Semantic Web- und Ontologie-Spezialisten mitarbeiten, da die Umsetzung sehr komplex ist. Dies war oftmals ein Grund für Archäolog*innen, Ideen in diese Richtung wieder zu verwerfen. Die Lösung, um ihnen den Zugang zu diesen Technologien zu erleichtern, sehen C. Binding u. a. in Vorlagen von Mappings, welche von den in der Archäologie Tätigen verwendet werden können.[682]

Bereits vorhandene oder neue Datensets von archäologischen Ausgrabungen sind meist auf einer Fundstelle basierend. Es ist aufgrund von unterschiedlichen Schemata und Vokabularen schwer,

[681] Hiebel u. a. 2014, 571–573.
[682] Binding u. a. 2015, 1 f.

sie datenübergreifend zu durchsuchen, zu vergleichen oder sie wiederzuverwenden. Die Voraussetzungen für das Modellieren in die CIDOC CRM-Ontologie beinhalten für C. Binding u. a. mehrere Faktoren. Das Datenschema der Ursprungsdaten und das Ontologie-Modell muss verstanden werden und ebenso die Techniken, die benötigt werden, um das Mapping auszudrücken. Im archäologischen Fachbereich finden sich selten Personen, die alle Fähigkeiten vereinen und so kommt es zu ressourcenintensiven und fehleranfälligen Ausführungen.[683]

C. Binding, u. a. sprechen sich für Werkzeuge aus, die Personen, welche keine Semantic Web Expert*innen sind, dabei zu unterstützen, Linked Data zu generieren und stellen diese im STELLAR Projekt vor.[684] Beim Überführen eines relationalen Datenmodells hin zu einer eventbasierten Ontologie wie CIDOC CRM kommt es oft zu Unklarheiten bei der Modellierung. Beispielsweise kommen Informationen über Events in den Ursprungsdaten oft nur implizit vor, müssen jedoch explizit modelliert werden. In manchen Fällen können die Anwender*innen nicht feststellen, ob ein Event kreiert werden muss oder soll, wenn beispielsweise einem Objekt ein Attribut zugeordnet werden möchte. Das CIDOC CRM gibt keine spezifischen Implementierungsdetails aus und so kann es vorkommen, dass ähnliche Ursprungsdaten zu unterschiedlichen Modellen gemappt werden können.[685] Dies stellt in weiterer Folge die Implementierung und Applikationen vor interoperative Herausforderungen.

„Mapping patterns“ oder „templates“, also Vorlagen für besonders oft vorkommende Daten von archäologischen Ausgrabungen könnten hier helfen (siehe Abbildung 36). Zum einen ermöglicht es Archäolog*innen ohne die nötigen technischen oder ontologischen Kenntnisse, Linked Data zu produzieren und die Ontologie einfacher anzuwenden. Zum anderen hilft es, die Lernkurve flacher zu halten und so größere Akzeptanz in Archäologie zu schaffen.[686] Das Projekt STELLAR hat einige Applikationen[687,688,689] entwickelt, die Vorlagen anbieten und die STELETO-Applikation[690] im Rahmen des ARIADNE Projektes[691] hat diese Anwendungen erweitert.

[683] Binding u. a. 2015, 2.
[684] Binding u. a. 2015, 1 f.
[685] Tudhope u. a. 2013, Kapitel 3.
[686] Binding u. a. 2015, 4–7.
[687] STELLAR 2011.
[688] Binding 2011.
[689] Siehe auf GitHub unter https://github.com/cbinding/stellar, abgerufen am 21.10.2021.
[690] Siehe die Repräsentation auf GitHub unter https://github.com/cbinding/STELETO, abgerufen am 21.10.2021.
[691] Siehe https://ariadne-infrastructure.eu/, abgerufen am 21.10.2021.

Die Verwendung einer gemeinsamen Ontologie ist der erste Schritt zu Interoperabilität.[692] Ein weiterer Faktor, so C. Binding u. a., um dahin zu gelangen, sind Terminologien, Vokabulare und Thesauri. Diese werden im nachfolgenden Kapitel im Detail erläutert.

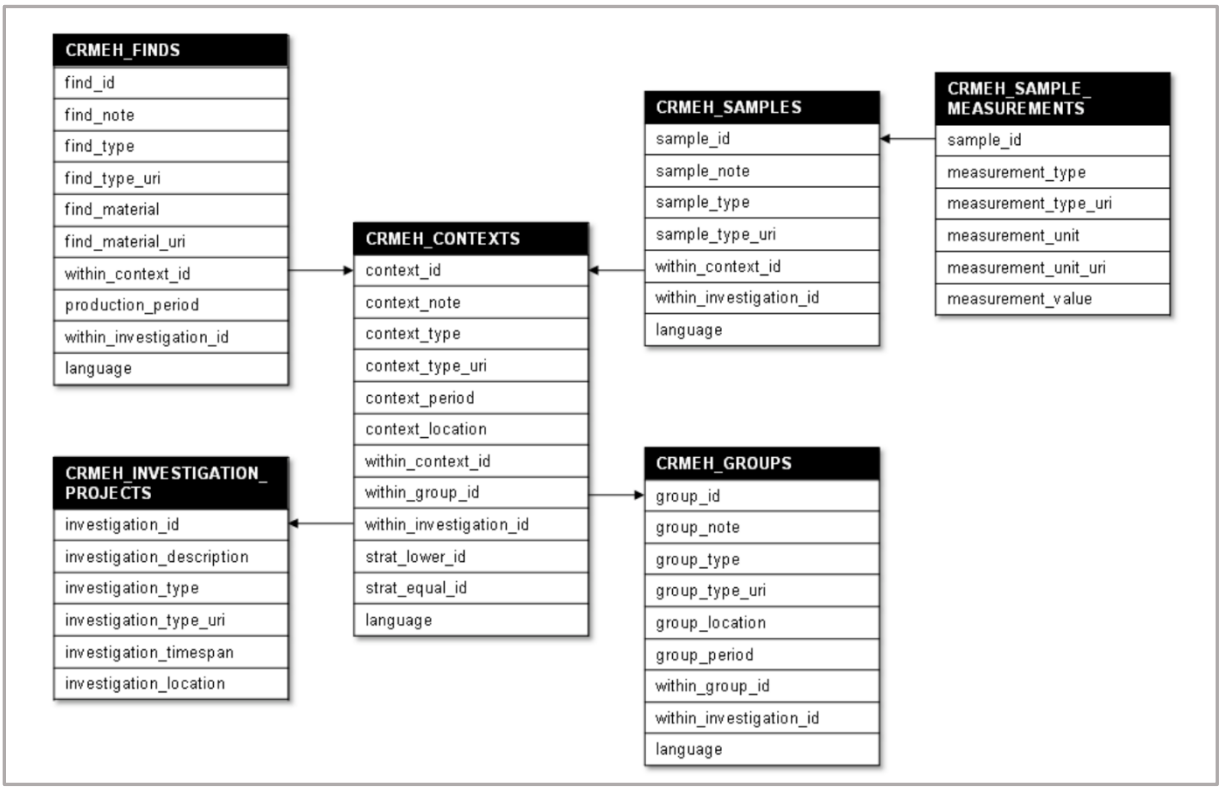


Abbildung 36 - Die Vorlagen des STELLAR Projekts für ein Mapping mit archäologischen Daten (Quelle: STELLAR 2011, 3, Figure 1).

[692] Binding u. a. 2015, 25.

6 Standards für die Integration archäologischer Daten

Um das gewünschte Ziel der Datenintegration heterogener Datensets zu erreichen, ist eine Modellierung mit einer Ontologie allein noch nicht ausreichend. Diese gibt lediglich die Semantik des Datenmodells und die Beziehungen und Eigenschaften der beschriebenen Objekte wieder. Die Werte der Daten selbst haben unterschiedliche Schreibweisen, Ansprachen oder Terminologien. Beispielsweise kann die Fundstelle in unterschiedlichen Formaten vorkommen und es kann nicht sichergestellt werden, ob es sich um die gleiche Fundstelle handelt.

Deshalb werden Online-Ressourcen, die breite Annahme und Akzeptanz genießen, herangezogen, um die Werte zu identifizieren. Selbst wenn Datenbestände mit der gleichen Ontologie wie z.B. dem CIDOC CRM modelliert wurden, kann ihre Maschinenlesbarkeit nur sichergestellt werden, wenn URIs verwendet werden.[693] URIs von Online-Ressourcen können bei räumlichen und zeitlichen Begriffen, Literaturzitaten, Terminologien, Personen, Kampagnen und Institutionen verwendet werden. Die meisten Webseiten von Online-Ressourcen verlinken sich gegenseitig, um ein Netz von Informationen zu generieren. Auf der Seite von PeriodO zur späten Bronzezeit in Österreich[694] wird der Link für Österreich zur Wikidata-Quelle zu Österreich[695] angegeben und die Epochen, welche im iDAI.chronontology beschrieben werden, verwenden als weitere Quelle jeweils Links zum Getty AAT Thesaurus[696], um nur zwei Beispiele zu nennen. Die verschiedenen Online-Ressourcen werden in diesem Kapitel näher beschrieben.

6.1 Räumlich

Die Information, an welchem Fundort ein Fundgegenstand gefunden wurde, bzw. welche geographischen Angaben zu einer Fundstelle gemacht werden, ist in der Archäologie von großer Bedeutung. Sie werden, wenn bekannt, in der Dokumentation angegeben. Um die Fundorte eindeutig zuordnen zu können, müssen dereferenzierbare URIs (siehe Kapitel 4.1) verwendet werden. Diese URIs sollen einen Namen für den Ort angeben, bestmöglich in verschiedenen Sprachen, eventuell Angaben zu antiken Namen des Ortes und eine geographische Zuordnung mit Koordinaten, wenn sie dereferenziert werden, sprich, wenn der Nutzer oder die Nutzerin der URL folgt, um Informationen zu erlangen über den URI.[697]

[693] Gerth 2017, 21 f.
[694] Siehe unter http://n2t.net/ark:/99152/p0nxc78x87x, abgerufen am 21.10.2021.
[695] Siehe unter http://www.wikidata.org/entity/Q40, abgerufen am 21.10.2021.
[696] Siehe auf der iDAI.chronontology Webseite https://chronontology.dainst.org/period/XXnNbsZApzO4, abgerufen am 21.10.2021, unter „digitale Provenienz".
[697] Gerth 2017, 23.

Angaben zu Fundstellen werden oftmals aus Angst vor Raubgräbern nicht öffentlich publiziert. Doch bereits ausgegrabene Fundstellen bringen diesen keinen großen Nutzen mehr, sofern die Fundstelle zur Gänze ausgegraben wurde. Dies ist bei der großen Anzahl an Rettungsgrabungen, welche nur den gefährdeten Bereich ausgraben, jedoch eher die Ausnahme. Die frei zugänglichen Informationen zu Fundstellen sind für die Forschung jedoch unabdingbar. Die INSPIRE Richtline (Infrastructure for Spatial Information in the European Community)[698] hat das Ziel, eine europäische Geodatenstruktur aufzubauen, doch in Hinblick auf Fundstellen hat sich noch kein Standard durchgesetzt.[699] Das österreichische Bundesdenkmalamt hat das Denkmalverzeichnis[700] für jedes Bundesland öffentlich zugänglich auf der Webseite mit Angaben zur Gemeinde und Katastralgemeinde, jedoch sind darin nicht alle Fundstellen erfasst.

In Datenbanken zu Ausgrabungen wird die Ortsangabe oft nur in der Überschrift oder als Titel der Datenbank verwendet und die Fundgegenstände werden den stratigraphischen Einheiten zugewiesen. Wenn Fundstellendaten integriert werden, ist es wichtig, jedem Fundgegenstand geographische Informationen zuzuweisen. Diese Geodaten mögen unzureichend für detaillierte Fundverteilungsanalysen innerhalb einzelner Fundstellen sein, sind jedoch sehr hilfreich für Analysen in weiter gefassten geographischen Analysen.[701]

Es gibt einige Online-Ressourcen für Ortsangaben, die breit etabliert sind und dereferenzierbare URIs anbieten. Viele davon stellen die Informationen ebenfalls im RDF-Format zur Verfügung. So können Ortsangaben eindeutig von Nutzer*innen identifiziert werden. Es handelt sich hier jedoch meist um Ortsangaben, wie Gemeindeflächen oder breiter gefasste Gebiete. Die Fundstelle selbst wird zurzeit nur selten mit Koordinaten veröffentlicht. Im Zuge des Interreg Projekts wurden auch Koordinaten der einzelnen Forschungsaktivitäten, wie Ausgrabungen und Begehungen zu Fundstellen der Bronze- und Eisenzeit, veröffentlicht. Die Online-Ressourcen für Ortsangaben sollen im Folgenden dargestellt werden:

- *Geonames*[702] ist eine geographische, globale, frei zugängliche Datenbank mit über 11 Millionen Orten. Die Orte werden in neun Haupt- und 645 Subkategorien eingeteilt, worunter sich auch archäologische Subkategorien befinden. Die Mehrheit der erfassten Orte ist jedoch rezent. Geonames verfügt über eine Ontologie, die „GeoNames Ontology“[703], und verwendet die 303 redirect-Strategie für URIs mit einer HTML- und einer RDF-Repräsentation (siehe Kapitel 4.1.2.2).

[698] Siehe https://inspire.ec.europa.eu/, abgerufen am 21.10.2021.
[699] Gerth 2017, 23.
[700] Siehe https://bda.gv.at/denkmalverzeichnis/, abgerufen am 21.10.2021.
[701] Isaksen 2011, 124.
[702] Siehe https://www.geonames.org/, abgerufen am 21.10.2021.
[703] Siehe https://www.geonames.org/ontology/documentation.html, abgerufen am 21.10.2021.

Der Ort Oberleis beispielsweise hätte folgende URIs: <https://www.geonames.org/2769960/oberleis.html> (HTML-Repräsentation) und <https://sws.geonames.org/2769960/about.rdf> (RDF-Repräsentation).

- *Pleiades*[704] ist ein Ortsverzeichnis von antiken Orten mit Daten aus unterschiedlichen Datensammlungen. Es besitzt über 37.000 Ortseinträge und wird laufend ergänzt. Die Orte sind zum größten Teil aus der griechischen und römischen Antike. Sie werden jedoch noch erweitert mit Einträgen aus dem antiken Nahen Osten, byzantinischen, keltischen und frühmittelalterlichen geographischen Informationen. Die Daten werden auch als RDF-Repräsentation angeboten und die Eigenschaften werden mit Ontologien wie SKOS, owl und einer eigenen Pleiades Ontologie[705] dargestellt.
  Vindobona (Wien) hat zum Beispiel die URIs <https://pleiades.stoa.org/places/128537> (HTML-Repräsentation) und <https://pleiades.stoa.org/places/128537/rdf> (RDF-Repräsentation).
- *iDAI.gazetteer*[706] des Deutschen Archäologischen Instituts ist ein Ortsverzeichnis mit über 148.000 rezenten und historischen bzw. antiken Orten. Die angegebenen Eigenschaften der Orte umfassen die Bezeichnung in mehreren Sprachen und die geographischen Angaben, wie Koordinaten und hierarchische geographische Verortung. Sie verfügen aber auch über zusätzliche Verlinkungen zu Orten, Identifikatoren von anderen Gazetteers, wie beispielsweise GeoNames, und andere Verweise. Auch hier ist eine Repräsentation in RDF vorhanden (weitere mögliche Repräsentationen sind RDF/XML, JSON, GeoJSON, KML und Shapefile). In der iDAI.welt[707] befinden sich noch andere Thesauri, kontrollierte Vokabulare und beispielsweise die Objektdatenbank des Deutschen Archäologischen Instituts. Die Daten werden bei bestehenden Beziehungen miteinander verlinkt. Die Fundorte der Objekte führen beispielsweise zu den Einträgen der Orte in den iDAI.gazetteer.
- *Getty Thesaurus of Geographic Names*[708] ist ein strukturierter Thesaurus für Namen geographischer Orte vornehmlich in den Bereichen Kunst, Architektur und materielle Kultur. Alle Vokabulare (Art & Architecture Thesaurus, Cultural Objects Name Authority und der Thesaurus of Geographic Names) sind für Linked Open Data aufbereitet und werden in unterschiedlichen Formaten (JSON, JSON-LD, RDF, N3-Turtle und N-Triples) angeboten.

[704] Siehe https://pleiades.stoa.org/, abgerufen am 21.10.2021.
[705] Siehe https://pleiades.stoa.org/help/conceptual-overview, abgerufen am 21.10.2021.
[706] Siehe https://gazetteer.dainst.org/app/#!/home, abgerufen am 21.10.2021.
[707] Siehe unter https://www.idai.world/, abgerufen am 21.10.2021.
[708] Siehe http://www.getty.edu/research/tools/vocabularies/tgn/, abgerufen am 21.10.2021.

- *Open Street Map*[709] ist das größte aus einer Gemeinschaft entstandene Ortsverzeichnis und bietet freie Geodaten. Im Datenmodell gibt es die Kategorie „historic" mit weiteren 39 Subkategorien, wie „archaeological site" oder „monument". Die Daten der Open Street Map werden als Linked Open Data von LinkedGeoData[710] angeboten mit einem SPARQL-Endpoint und vielen Export-Formaten.
- *Interreg – Danube Transnational Programme – Iron-Age-Danube*[711] bietet Daten zu Ausgrabungskampagnen für Teile von Österreich, Ungarn, Kroatien und Slowenien. Im Zuge des Iron-Age-Danube-Projekts wurden die Koordinaten der Forschungsaktivitäten veröffentlicht.[712]

Welches Online-Ortsverzeichnis gewählt wird, hängt von den Ausgangsdaten ab. Wenn in den Daten antike Ortsnamen vorkommen, bieten Pleiades oder iDAI.gazetteer beispielsweise mehr mögliche Verlinkungen. Auch die Datenformate können zur Entscheidung beitragen.

6.2 Zeitlich

Informationen zu den Datierungen in den Datensätzen gehören in der Archäologie ebenfalls zu den zentralen Angaben in den Daten. Für die Datenintegration sind standardisierte Definitionen der Perioden mit Quelleninformationen aus der Fachliteratur und dereferenzierbaren Identifikatoren daher ein wichtiges Werkzeug. Im Folgenden wird auf die wichtigsten Ressourcen dafür hingewiesen:

- *PeriodO*[713] ist ein öffentlich zugängliches, gemeinschaftliches, globales Zeitperioden-Verzeichnis. Es bietet Definitionen von Zeitperioden aus historischen, kunstgeschichtlichen und archäologischen Perioden an. Das PeriodO-Verzeichnis findet mittlerweile breite Anwendung bei Forschungsprojekten in Europa. Das verwendete Datenmodell ist einfach gestaltet mit einer Kombination der SKOS-Ontologie und der Time-Ontology in OWL[714]. Dabei stellen die Zeitperioden SKOS-Konzepte dar. Die Perioden können in JSON-LD und das ganze Datenset in JSON, Turtle oder CSV heruntergeladen werden. Jede Periode hat eine dereferenzierbare URI und muss im Verzeichnis folgende vier Punkte vorweisen:[715]
    - Sie hat einen Namen.

[709] Siehe https://www.openstreetmap.org/about, abgerufen am 21.10.2021.
[710] Siehe http://linkedgeodata.org, abgerufen am 21.10.2021.
[711] Siehe https://www.iron-age-danube.eu/, abgerufen am 26.09.2021.
[712] Siehe hierzu auch Czajlik u. a. 2019.
[713] Siehe https://perio.do/en/, abgerufen am 21.10.2021.
[714] Cox – Little 2020.
[715] Siehe https://perio.do/technical-overview/, abgerufen am 21.10.2021.

- Sie hat zeitliche Grenzen, welche durch Textquellen und einer strukturierten Repräsentation in der OWL-Time Ontology definiert werden.
- Sie hat Assoziationen zu einer geographischen Region, welche ebenfalls durch Textquellen und einem Identifikator eines Ortsverzeichnisses (z.B. Geonames, Pleiades oder DBPedia) identifiziert werden.
- Sie ist in einer zitierfähigen Quelle publiziert worden.

- *iDAI.chronontology*[716] ist durch das ChronOntology Projekt entstanden und wird vom Deutschen Archäologischen Institut im Rahmen der iDAI.welt langfristig betreut. Der Webservice verbindet Zeitperioden mit Datierungen, dient als Normdatenvokabular für zeitliche Informationen und verbindet die Informationen auch mit anderen Zeitverzeichnis-Systemen (wie PeriodO oder Getty AAT). Für die Definition der Datierungen werden mehrere Quellen herangezogen. Im Datenmodell, welches stark auf der CIDOC CRM-Ontologie basiert, werden Zeitperioden nach der CIDOC CRM-Klasse als *E92 Spacetime Volumes* geführt. Die Klasse erlaubt es, den Perioden eine zeitliche und eine geographische Ausdehnung zuzuordnen. Das Datenmodell kann den Forschungsstand exakt wiedergeben, weil Zeitbegriffe auch mit unvollständigen Angaben eingegeben werden können. Des Weiteren werden die Perioden in spezifische Periodentypen bzw. Epochentypen mit mehreren Untertypen eingeteilt (wie geologisch, politisch, kulturell, materielle Kultur oder kein Typ). Die Perioden besitzen dereferenzierbare URIs und können im JSON-Format heruntergeladen werden.
- *Getty Arts and Architecture*[717] (Getty AAT) ist ein Thesaurus für Kunst, Architektur und materielle Kultur. Im Bereich der Zeitperioden ermöglicht es die sogenannte Facette „Styles and Periods Facet" (Stile und Perioden Facette), Perioden zuzuordnen. Sie gibt Definitionen von kunst- und kulturgeschichtlichen, archäologischen oder historischen Zeitepochen wieder.

Bei der Wahl der Identifizierung der Datierungen in den Daten spielen ebenfalls die Ausgangsdaten eine Rolle. Für urgeschichtliche Zeitangaben eignen sich beispielsweise andere Verzeichnisse als für Angaben für römische Kaiser oder Kunstepochen der frühen Neuzeit.

6.3 Literaturzitate

Der Prozess der Standardisierung bei bibliographischen Daten hat bereits gegen Ende des 19. Jahrhunderts begonnen. Einzelne Regelwerke für Bibliotheken wurden ausgeweitet und führten

[716] Siehe https://chronontology.dainst.org/, abgerufen am 21.10.2021.
[717] Siehe http://www.getty.edu/research/tools/vocabularies/aat/index.html, abgerufen am 21.10.2021.

zu einer heute internationalen engen Zusammenarbeit von Bibliotheken und Bibliotheksverbänden. Die regulierenden Vorgehensweisen waren in Bibliotheken und anderen Informationssystemen auch früher schon beliebt. Mit der weiterschreitenden Digitalisierung bringen sie viele Vorteile, von der Arbeitseffizienz bis hin zu vernetzten Katalogsystemen.[718] Es haben sich einige Standards für bibliographische Informationen etabliert, welche sich eignen, die verwendete Literatur in den Daten zu identifizieren:

- Das *Marc 21*[719] (Machine Readable Cataloging) Format ist ein Standard, um bibliographische und verwandte Informationen in maschinenlesbarer Form zu repräsentieren und wird für Metadaten verwendet.
- 1972 wurde die *ISBN* (International Standard Book Number) als ISO-Norm[720] veröffentlicht. Da jedoch ältere Publikationen vor 1972 noch keine ISBN erhalten haben und viele kleinere Veröffentlichungen keine ISBN erhalten, ist es für den generellen Gebrauch in der Archäologie nicht umfangreich genug.
- Die *ISSN* (International Standard Serial Number)[721] bietet einen Identifikator für seriell erscheinende und fortlaufende Publikationen in analoger oder digitaler Form, wie Zeitungen, (Fach-)Zeitschriften, Konferenzbeiträge oder Serien von Monographien. Die sechste Auflage der ISO-Norm erschien im Jahr 2020.
- Globale Bibliothekskataloge eignen sich als Identifikationswerkzeug mit dereferenzierbaren URIs. Einer der größten Kataloge ist *WorldCat*[722]. Er vereint über 10.000 Bibliothekskataloge und weist zwei Milliarden Dateneinträge auf.
- Der Bibliothekskatalog *iDAI.bibliography zenon*[723] des Deutschen Archäologischen Instituts stammt aus der iDAI.welt. Der Katalog kombiniert alle Bibliothekskataloge des DAI sowie einige externe Kataloge und ist die größte Sammlung bibliographischer Informationen für Altertumswissenschaften.

6.4 Terminologie und Typologie

Kontrollierte Vokabulare sind wichtig, um archäologische Daten zu integrieren und stellen den komplexesten Bereich für Standardisierungen dar. Archäologische Fachtermini sind sehr detail-

[718] Wimmer 2017, 281 f.
[719] Siehe https://www.loc.gov/marc/, abgerufen am 21.10.2021.
[720] ISO 2017a.
[721] ISO 2020.
[722] Siehe https://www.worldcat.org/, abgerufen am 21.10.2021.
[723] Siehe https://zenon.dainst.org/, abgerufen am 21.10.2021.

reich und werden oft von spezifischen Forschungsgruppen innerhalb eines Forschungsschwerpunktes erstellt. Sie unterscheiden sich oft selbst innerhalb eines Fachgebietes. Die globalen autoritativen Thesauri (wie beispielsweise Getty AAT oder Heritage Data[724]) sind für die meisten archäologischen Zwecke daher zu unspezifisch.

Die Web-Applikation Labeling System[725] des Instituts für Raumbezogene Informations- und Messtechnik (i3mainz) und des Römisch Germanischen Zentralmuseums (RGZM) in Kooperation mit dem Mainzer Zentrum für Digitalität in den Geistes- und Kulturwissenschaften (mainzed) befindet sich in der Betaversion. Sie soll es ermöglichen, von Forscher*innen eigens kreierte kontrollierte Vokabulare zu erzeugen. Die Vokabulare sollen mit anderen Online-Thesauri verlinkt werden können. Zusätzlich sollen URIs für die eigenen Begriffe generiert werden können.[726] Beispiele für kontrollierte Vokabulare und Thesauri sind im Folgenden aufgeführt:

- Der *Getty Arts & Architecture Thesaurus*[727] ist ein kontrolliertes Vokabular für Kultur-, Kunst- und Altertumswissenschaften. Die hierarchische Struktur des Vokabulars wird in acht Hauptfacetten unterteilt (wie beispielsweise „Materials", „Objects", „Styles and Periods"), welche weitere Kategorisierungen vorweisen. Nachteile für archäologische Daten sind einerseits die geringe Spezialisierung in der Terminologie und andererseits die teilweise noch nicht fertiggestellte, sprachliche Übersetzung der Begriffe aus dem Englischen.
- Das *iDAI.vocab*[728] aus der iDAI.welt des Deutschen Archäologischen Instituts bietet ein gänzlich archäologisches Vokabular. Dieses online zugängliche Vokabular bietet 14 Thesauri mit je einer Sprache, wobei der deutsche Thesaurus die Ausgangsbasis darstellt. Jeder Begriff kann eine URI vorweisen mit entsprechender RDF Repräsentation im JSON-LD-Format. Verweise auf andere Vokabulare gibt es zu Getty AAT und DBPedia[729].

Zahlreiche Forschungsprojekte und Institute haben sehr spezielle fachspezifische Vokabulare in ihren Projekten entwickelt und stellen diese zur Verfügung. Beispiele für Vokabulare im Bereich der Archäologie mit Veröffentlichung im Linked Open Data Format sind:

- *Nomisma*[730] ist ein Vokabular und eine Ontologie für Münzen.
- *Kerameikos*[731] ist ein spezialisiertes Vokabular für griechische Keramik.

[724] Siehe https://www.heritagedata.org/blog/vocabularies-provided/ für die verfügbaren Vokabulare, abgerufen am 21.10.2021.
[725] Siehe https://www.rlp-forschung.de/public/facilities/400/research_projects/20489, abgerufen am 21.10.2021.
[726] Thiery – Mees 2017.
[727] Siehe https://www.getty.edu/research/tools/vocabularies/aat/about.html, abgerufen am 21.10.2021.
[728] Siehe https://archwort.dainst.org/de/vocab/, abgerufen am 21.10.2021.
[729] Eine Linked Open Data Repräsentation von Wikimedia, siehe https://wiki.dbpedia.org/, abgerufen am 21.10.2021.
[730] Siehe http://nomisma.org/, abgerufen am 21.10.2021.
[731] Siehe http://kerameikos.org/, abgerufen am 21.10.2021.

- Das Austrian Centre for Digital Humanities (ACDH) der Österreichischen Akademie der Wissenschaften stellt auf der Webseite *Vocabs*[732] eine Vielzahl an kontrollieren Vokabularen zur Verfügung. In der Kategorie Archäologie wird der *DEFC Thesaurus*[733] zur Verfügung gestellt, der für das Neolithikum in Griechenland und Anatolien im Rahmen des DEFC Projektes[734] entwickelt wurde. Ebenso in dieser Kategorie befindet sich der *Iron-Age-Danube Thesaurus*[735], der ein kontrolliertes Vokabular der Eisenzeit in den Regionen Österreich, Ungarn, Kroatien und Slowenien sowie von Ausgrabungsmethoden darstellt und im Zuge des Interreg Iron-Age-Danube Projektes[736] entwickelt wurde[737].
- Das *Thanados*-Projekt[738] hat ein spezielles Vokabular für Artefakte von frühmittelalterlichen Gräberfeldern[739] entwickelt, welches auf der Webseite verfügbar ist und permanente Links bietet.

6.5 Personen, Institute und Ausgrabungskampagnen

Für archäologische Daten sind Definitionen und eine eindeutige Zuweisung von Personen, Wissenschafter*innen, Instituten und Ausgrabungskampagnen nützlich. In manchen Fällen können auch historische Personen, wie bekannte historische Persönlichkeiten in Inschriften, Herrscher oder Prägeherren von Münzen, von Bedeutung sein oder mythologische Kreaturen oder Charaktere.[740]

Die Zuweisung von URIs ist dabei nicht immer einfach. Für reale Personen existieren Möglichkeiten zur Identifikation über das semantische Vokabular FOAF (Friend Of A Friend)[741,742] oder das VIAF (Virtual International Authority File)[743]. Das VIAF eignet sich ebenfalls für Institutionen. Um Ausgrabungskampagnen eindeutig zu identifizieren, gibt es noch kein großflächig abdeckendes Informationssystem. Für Teile von Österreich, Ungarn, Kroatien und Slowenien wurden

[732] Siehe https://vocabs.dariah.eu/de/, abgerufen am 21.10.2021.
[733] Siehe https://vocabs.dariah.eu/defc_thesaurus/de/, abgerufen am 21.10.2021.
[734] Siehe zum DEFC (Digitizing Early Farming Cultures) https://defc.acdh.oeaw.ac.at/, abgerufen am 21.10.2021.
[735] Siehe https://vocabs.dariah.eu/iad_thesaurus/de/, abgerufen am 21.10.2021.
[736] Siehe www.iron-age-danube.eu, abgerufen am 21.10.2021.
[737] Siehe dazu Czajlik u. a. 2019.
[738] Siehe unter https://thanados.net/about, abgerufen am 30.09.2021.
[739] Das Vokabular ist verfügbar unter https://thanados.net/vocabulary/, abgerufen am 30.09.2021.
[740] Gerth 2017, 30.
[741] Stuart 2012, 173.
[742] Siehe unter http://xmlns.com/foaf/spec/, abgerufen am 21.10.2021.
[743] Siehe unter http://viaf.org/, abgerufen am 21.10.2021.

vom Iron-Age-Danube-Projekt[744] IDs für archäologische Maßnahmen und Ausgrabungskampagnen vergeben, die als URIs verwendet werden können. Identifikatoren für historische Personen oder mythologische Kreaturen und Personen sind über VIAF oder Wikidata zu finden.

- *FOAF (Friend Of A Friend)*[745] ist im Jahr 2000 veröffentlicht worden und ist eines der Vokabulare, das am längsten verwendeten wurde. Hier können Personen in Online-Profilen Informationen zu ihrer individuellen Person (wie Name, Alter etc.), Institution (z.B. Universität oder Institut) und Online-Präsenz (E-Mail, Homepage etc.) angeben. Profile können mit anderen Profilen semantisch verknüpft werden (z.B. über den Begriff „knows"). Trotz der langen Laufzeit ist die Benutzung für Wissenschafter*innen nicht sehr breit angenommen worden.[746]
- *VIAF (Virtual International Authority File)*[747] ist ein von OCLC bereit gestellter Personennamensdienst, welcher Personennamen-Dateien vereint und vernetzt. Benutzer*innen haben Zugang zu globalen Normdaten von Nationalbibliotheken, Kulturagenturen und anderen Institutionen. Dadurch werden Namen, Orte, Werke und Ausdrücke mit Möglichkeit der sprachlichen Präferenz mit dereferenzierbaren URIs zur Verfügung gestellt. Das Institut für Urgeschichte und Historische Archäologie Wien hat beispielsweise folgende URI: <http://viaf.org/viaf/245829558>.
- *Wikidata*[748] ist eine offene und freie Datenbank mit strukturierten Informationen aus den Wikimedia-Projekten (z.B. Wikipedia). Diese Informationen sind als Linked Open Data in unterschiedlichen Formaten (JSON, RDF oder XML) zugänglich. Benutzer*innen können auch Einträge neu erstellen und mit anderen Ressourcen oder Datensätzen verlinken.
- *DBPedia*[749]ist ein Projekt, welches die strukturierten Wikipedia-Einträge aus den Infoboxen in einen Knowledge-Graphen umwandelt. Es eignet sich darum für alle Kategorien. Das Projekt verwendet eine Ontologie mit 320 Klassen und 1650 Eigenschaften. Die Wissenssammlungen können per Download auf den eigenen Rechner geladen werden und es werden SPARQL-Abfragen angeboten. DBPedia ist durch seine vielen Verlinkungen ein großer Teil der LOD-Cloud (siehe Kapitel 3.2.2) und kann verwendet werden, um seine eigenen Daten dorthin zu verlinken.[750]

[744] Siehe www.iron-age-danube.eu, abgerufen am 21.10.2021.
[745] Siehe http://www.foaf-project.org/ und http://xmlns.com/foaf/spec/, abgerufen am 21.10.2021.
[746] Stuart 2012, 177.
[747] Siehe http://viaf.org/, abgerufen am 21.10.2021.
[748] Siehe https://www.wikidata.org, abgerufen am 21.10.2021.
[749] Siehe https://www.dbpedia.org/about/, abgerufen am 21.10.2021.
[750] Lehmann u. a. 2015, 167.

Teil eins dieser Arbeit zeigt die grundlegenden Konzepte, Ideen und Visionen eines Semantic Web und Linked (Open) Data. Mithilfe der vorgestellten semantischen Technologien können Daten interoperabel, integrierbar und wiederverwendbar transformiert werden. Diese Vorteile sind für archäologische Daten besonders interessant, da sie meist aufgrund unterschiedlicher Aufnahme- und Dokumentationsverfahren und durch die Verwendung von verschiedenen Standards und Terminologien heterogen sind.

Semantische Technologien können helfen, archäologische Daten, wie beispielsweise Katalogdaten, Daten von bereits abgeschlossenen Ausgrabungen, sowie neu generierte und naturwissenschaftliche Daten so aufzubereiten, dass ein Netz aus Wissen gesponnen werden kann. Die verwendbaren Standards, Vokabulare und Identifikatoren, sowie ein maschinenlesbares Format und Ontologien können Daten aus unterschiedlichen Quellen zusammenbringen und bergen ein enormes Potential für archäologische Daten.

Daher soll im zweiten Teil dieser Arbeit untersucht werden, wie diese Technologien im Detail in der Archäologie bereits angewendet werden und wie Archäolog*innen sie selbst an ihren Daten verwirklichen können. Die theoretischen Grundlagen werden anhand der Case Study zu einer Funddatenbank zu Funden des Oberleiserbergs gezeigt und sollen Archäolog*innen helfen, ihre Daten interoperabel und wiederverwendbar zu konvertieren. Denn je mehr archäologische Daten auf diese Weise veröffentlicht werden können, desto näher rückt die Vision eines archäologischen Wissensnetzwerks.

Teil 2 –Anwendung von Linked Open Data und semantischer Technologien in der Archäologie

Im zweiten Teil dieser Arbeit soll die praktische archäologische Anwendung semantischer Technologien im Fokus stehen. Die Archäologie verwendet semantische Technologien bereits seit Jahren, deshalb werden in Kapitel 7 Beispiele archäologischer Projekte gezeigt, welche diese Technologien zu ihrem Vorteil verwenden. Im Anschluss soll eine Schritt-für-Schritt Anleitung folgen, um archäologische Daten aus einem Fundkatalog für das Semantic Web aufzubereiten und nutzen zu können.

7 Linked Open Data und die Archäologie

Aktuell werden Linked Open Data Technologien in der Archäologie noch wenig angenommen und verwendet. In den letzten zehn bis fünfzehn Jahren hat sich das Basiswissen darum, wie verlinkte Daten produziert, veröffentlicht und verlinkt werden sollen, erhöht. Doch dies ist mehrheitlich in einzelnen progressiven Projekten angewandt worden. In der Archäologie existieren bisher wenige Datensets, welche durch Linked Open Data verlinkt sind, und fast keine davon finden sich in der LOD-Cloud wieder. Ein Beispiel für archäologische Daten, welche im Netz publiziert sind und auf externe Online-Ressourcen verweisen, sind die Daten des Thanados-Projekts[751]. Der Datenkatalog wurde mit der Ontologie CIDOC CRM modelliert, online zur Verfügung gestellt und verweist dabei auf andere Quellen bzw. URIs, wie beispielsweise von GeoNames[752]. Für die Vision eines Web of Data in der Archäologie müssten dies jedoch auf jeden Fall bedeutend mehr werden.[753] Dabei würde die Archäologie als multi-disziplinäres Forschungsfeld von einem weit verlinkten Netz von Daten profitieren. Bis es zu einem weitreichenden Web of Data kommt, welches mit anderen Gebieten vernetzt werden kann, gibt es noch viele Hürden zu überwinden.[754]

Das Ziel der Linked Open Data ist es, vereinzelte, alleinstehende Datensets mit anderen zu verlinken, um die Daten wiederum für andere Optionen verwenden zu können. Ebenso sollen ansonsten isolierte Daten entdeckt werden können.[755] Doch es wäre nicht sehr sinnvoll, einfach seine Daten als Linked Open Data zu veröffentlichen. Dateninhaber*innen sollten sich überlegen, welchen Nutzen die Daten haben könnten, wenn sie wiederverwendet werden und auch, wer sie

[751] Siehe unter https://thanados.net/sites, abgerufen am 21.10.2021.
[752] Siehe unter https://www.geonames.org/, abgerufen am 21.10.2021.
[753] Geser 2016, 5.
[754] Geser 2016, 7.
[755] Geser 2016, 8.

gebrauchen könnte[756]. Wichtig ist, zu überlegen, zu welchen anderen Daten die eigenen Daten verlinkt werden können. Dies ist aktuell noch eine schwierige Überlegung, da es nur wenige gibt. Gemeinsame Initiativen im Feld der Archäologie würden die Adaptierung voranschreiten lassen. Dafür müssten Linked Open Data präsenter werden, die Vorteile und Kosten dieser Technologien transparenter kommuniziert werden, die Einstiegsschwelle erniedrigt und so für die Forschung attraktiver gemacht werden.[757]

7.1 Die Adaptierung von Linked Open Data in der Archäologiepraxis

Bei der Überlegung, wie semantische Technologien in die Praxis und den Alltag der Archäologie passen, betont L. Isaksen wie schwierig eine Prognose sei, ob und wie die Archäologie das Semantic Web verwenden wird. Es würde nur Stück für Stück und unter Mitarbeit von vielen unterschiedlichen Personen mit sich unterscheidenden Ansichten möglich sein. Um dies besser herausfinden zu können, schlägt er vor, sich anzusehen, wie attraktiv das Semantic Web für die Archäologie ist und die Fähigkeit der Archäologie, diese miteinzubinden.[758]

Für die Archäologie sind nicht alle Vorteile gleich relevant. Wobei L. Isaksen betont, höchstwahrscheinlich seien die meisten, wenn nicht sogar alle davon, für Archäologen oder Archäologinnen zumindest interessant wären. Einige würden unmittelbar Nutzen bringen, andere sind eher längerfristig relevant:

- Am interessantesten, beschreibt er, seien interaktive Daten, Datenintegration, Kontextualisierung, Herkunftsbestimmung und Dokumentenzerlegung.
- Immer noch von Interesse, aber nicht mehr so unmittelbar, sieht er die Maschinenlesbarkeit, Datenintegrität (Inkonsistenzen), selbstbeschreibende Daten und die Wiederverwendung.
- Als letzten Punkt nennt er die Logikschlussfolgerungen bzw. Inferenzen, die Personalisierung und die Web-Persistenz der Daten.[759]

Interaktive Daten würden in der Archäologie eine große Zeitersparnis bedeuten. In der Archäologie Tätige bringen viel Mühe auf, um Konzepte nachzuschlagen und Ähnlichkeiten in der Literatur zu finden. Ein direkter Zugang zu diesen Daten wäre ein Bonus. Datenintegration ist ein

[756] Dieser Punkt gilt jedoch nicht für die Metadaten. Diese sollten so genau wie möglich generiert werden (siehe Kapitel 2.1.2).
[757] Geser 2016, 8 f.
[758] Isaksen 2011, 10.
[759] Isaksen 2011, 10.

viel diskutiertes Thema in der Archäologie und birgt viele Hürden. Der Großteil an archäologischen Daten ist heterogen und isoliert. Daten zusammen bearbeiten zu können, würde einen großen Vorteil darstellen. Kontextualisierung bringt zusätzliche Informationen und unerwartete neue Entdeckungen. Die Herkunft von archäologischen Daten ist von besonderer Wichtigkeit. Sie gibt zum einen Auskunft über die Qualität und Vertrauenswürdigkeit von gefundenen Daten und stellt klar, wie sie verwendet werden können. Bei der Dokumentenzerteilung kann Graue Literatur und auch einzelne Datensätze, wie beispielsweise ein Fund einer bestimmten Kategorie, durchforstet und genutzt werden.[760]

Vorteilhaft, aber in der Archäologie weniger interessant ist die Automatisierung durch Maschinenlesbarkeit. Rohdaten können durch Algorithmen durchforstet und prozessiert werden und einfache Prozesse automatisiert werden. Da nur wenige Archäolog*innen Software-Programmierkenntnisse besitzen, wäre meist zusätzliches Personal erforderlich. Datenintegrität ist nicht immer einfach in einem so breit aufgestellten Feld. Inkonsistenzen und Ungereimtheiten einfacher finden zu können, hätte jedenfalls einen positiven Effekt. Es ist allgemein anerkannt, dass Metadaten einen wichtigen Bestandteil des archäologischen Datenmanagements darstellen. In der Praxis wird dies oft vernachlässigt, sei es durch Zeitmangel oder anderen Ursachen. Selbstbeschreibende Daten würden diesen Prozess erleichtern. Die archäologische Forschung ist stark abhängig von vorangegangener Arbeit (wie Datierungen, Fundkategorisierungen, Fundorte etc.). Die Forschungsergebnisse tendieren eher als fertig interpretierte Publikationen veröffentlicht zu werden und die Rohdaten werden selten mitpubliziert (siehe Kapitel 2.2.3.2). Die Daten wiederverwenden zu können kann zu neuen Fragestellungen und Ergebnissen führen.[761]

Doch es können weitere Vorteile genannt werden. A. Deicke verweist auf die positiven Effekte, welche entstehen, wenn Katalogdaten mit einer Ontologie modelliert werden.[762] Die häufigste Form der archäologischen Publikation geschieht über Kataloge. Fundobjekte aus einem bestimmten Kontext werden aufgelistet, klassifiziert und beschrieben. Dabei geht es von detaillierten Angaben zu den Funden und ihren Eigenschaften, wie etwa bei Ausgrabungsberichten, bis hin zur weiteren Typologisierung, chronologischen Einteilung und bibliographischen Informationen. Diese Daten wären besonders interessant für alle Arten von Analysen, wie Netzwerkanalysen oder Verteilungsanalysen, oder als Beiträge zur Linked Data Cloud (siehe Kapitel 3.2).

Der Gewinn, den dieses komplexe und zeitaufwändige Modellieren mit sich bringt, ist oft nicht unmittelbar zu erkennen. Vor allem bei kleineren Projekten oder Arbeiten von Einzelpersonen,

[760] Isaksen 2011, 11.
[761] Isaksen 2011, 11.
[762] Deicke 2016, Kapitel 1.1.

welche nicht unbedingt die Absicht verfolgen, ihre Rohdaten der Öffentlichkeit zur Verfügung zu stellen. A. Deicke weist allerdings darauf hin, wie eine Modellierung und die anschließende Publikation der strukturierten Daten nutzen kann. Die Publikation erleichtert die Kommunikation über die Daten selbst und über die Forschungsergebnisse und ebenso nachfolgende Diskussionen darüber. Auch das Verständnis des Themas und der komplexen Zusammenhänge wird verbessert und bringt eventuell neue Ideen und Perspektiven. Dadurch können eventuelle Lücken in den Daten besser erkannt werden. Archäologisches Ausgangsmaterial weist oft von Natur aus Lücken auf, die nicht immer gut erkennbar sind. Die Modellierung kann helfen, fehlende Elemente zu finden und die logischen Schlussfolgerungen, die aus den Daten gezogen werden, zu stärken.[763] Weil die Ontologie CIDOC-CRM (siehe Kapitel 5.2.1) eventbasiert aufgebaut ist, ist es einfacher, zu entdecken, wenn einem Objekt kein Event zugeordnet werden kann. Diese Events fehlen dann im Modell.[764]

Wesentliche Kriterien für eine größere Akzeptanz innerhalb der Archäologie sind für L. Isaksen verbesserte Benutzeroberflächen, um die Datenproduktion und Datenkonsumation zu erleichtern. Dies könnte helfen, Kosten zu reduzieren und die hohen technologischen Anforderungen zu reduzieren. Kontrollierte Vokabulare und Standards wären jedenfalls gute Alternativen, solange diese Kriterien nicht erfüllt sind.[765] Für die Datenproduktion und -konsumation gibt es noch keine generellen und vor allem einfachen Methoden und Anleitungen zur Anwendung. Die jeweiligen Schritte, die nötig sind, um zum gewünschten Ergebnis zu kommen, müssen recherchiert und angelernt werden. Bis zum jetzigen Zeitpunkt muss in den meisten Fällen ein Spezialist hinzugezogen werden. Dies führt zum Punkt der Kosten. Diese sind schwierig zu kalkulieren und hängen von den gewünschten Ergebnissen ab. Es ist jedenfalls mit zusätzlichen Personal- und Infrastrukturkosten zu rechnen. Auch die längerfristigen Pläne für die Daten bezüglich der Onlinepräsenz sollten beachtet werden. Ausgerechnet der Punkt der Kosten trifft die Archäologie hart, denn das Feld ist im Allgemeinen nicht besonders hoch dotiert und muss bei Projekten seine Ausgaben gut abwägen. Hinzugezogene Spezialisten und ein zeitlicher Mehraufwand könnten wichtige andere Projektteile erschweren. Jedenfalls sind ein größerer Aufwand und zurzeit noch spezielle Hilfe notwendig, um die Technologien anzuwenden. Kostengünstigere Alternativen, die zumindest ein gewisses Level an Semantik ermöglichen, sollten zurzeit jedenfalls noch in Betracht gezogen werden.[766] Dies wären beispielsweise kontrollierte Vokabulare und Standards zu verwenden.

[763] Deicke 2016, Kapitel 1.1.
[764] Cripps u. a. 2004, 7, 26.
[765] Isaksen 2011, 10.
[766] Isaksen 2011, 10.

7.1.1 Schritte zur Adaptierung von Linked Open Data Technologien

G. Geser hat im Zuge des ARIADNE Projektes eine Studie zusammengefasst, die ermittelt, welche Schritte für die nötige Adaptierung von Linked Open Data in der Archäologie notwendig wären. Die wichtigsten sind: Bewusstsein schaffen für Linked Open Data, Kosten-Nutzen darstellen, nicht IT-Spezialisten ermöglichen, Linked Open Data Tools zu verwenden, mehr Linked Data zum Verlinken generieren und Anklang in der Forschung zu finden.[767]

Die Vorbehalte, welche Linked Open Data gegenüber geäußert werden, sind vielseitig. Archäolog*innen erhoffen sich zwar mehr Sichtbarkeit für ihre Daten und Sammlungen, fürchten jedoch Kontroll- und Qualitätsverlust ihrer Daten durch Verlinkungen mit weniger hochwertigen Daten. Ein weiterer Grund für die langsame Adaptierung ist die gängige Meinung, diese Technologien würden hohe Kosten verursachen. Dies liegt unter anderem auch daran, dass die Linked Data Projekte, welche bisher durchgeführt wurden, die Vorteile aufführen, die Kosten für die Adaptierung bleiben jedoch im Dunkeln. Es fehlen auch Beispiele, ob sich ein längerfristiger Nutzen erkennen lässt. Es gibt einige Methoden und Tools, welche Kosten verringern können, da sie die Effizienz der Arbeit steigern können[768]. Doch es gibt bereits so viele Tools, dass die Nutzer*innen nicht mehr wissen können, welche für sie sinnvoll sind. Dabei ist es wichtig, nicht IT-Spezialisten zu ermöglichen, Tools für die Anwendung dieser Technologien zu benutzen. Die meisten Projekte, welche Linked Data verwenden, sind abhängig von Expert*innen in diesem Gebiet. Denn die Tools, die zur Verfügung stehen, sind oft nicht sehr benutzerfreundlich und für nicht IT-Spezialisten nicht verständlich und daher nicht verwendbar. Es könnte beispielsweise Tools geben, welche bereits auf der Ausgrabung oder bei der Bearbeitung der Daten verwendet werden könnten. Diese können Linked Data Vokabular verwenden, damit dieser Schritt automatisch passiert. Datenbanken, welche mit Tablets auf der Ausgrabung verwendet werden, werden oft schon angewendet (z. B. die proprietäre Software von INARI[769]). Diese könnten mit einer Ontologie und einer Möglichkeit für RDF-Export der Daten versehen werden. Die Archäologie ist im Allgemeinen eher zurückhaltend, wenn es um die Verwendung von neuen Standards oder Terminologien geht. Die Schwierigkeiten und die fehlende Zeit sind hier meist das größte Hindernis. Somit sollten sich neue Tools in die normalen Tätigkeiten der Archäologie einfügen. Im Idealfall würden sie den semantischen Aspekt im Hintergrund mitlaufen lassen, so aber Interoperabilität ermöglichen, wenn die Daten publiziert werden. Nur wenn die Einstiegsbarriere möglichst niedrig gehalten wird, kann mehr Linked Data produziert werden, welche wiederum wichtig sind, um neue

[767] Geser 2016, 11–16.
[768] Beispiele hierfür sind OpenRefine, siehe unter https://openrefine.org/, abgerufen am 21.10.2021, oder RDF-Konverter.
[769] Siehe https://www.inari-software.com/de/, abgerufen am 21.10.2021.

Daten damit zu verlinken. Eine Stelle, an die sich Archäolog*innen wenden könnten und die Hilfestellung und Anleitungen anbietet, wäre nützlich für die Etablierung dieser Technologien.[770]

L. Isaksen hat in einer Umfrage unter Archäolog*innen, Museums- und Archivmitarbeiter*innen, die semantische Technologien in einem Projekt ausprobiert haben, ähnliche Resultate erhalten. Es wurden jeweils die größten Vorteile und Schwierigkeiten herausgefiltert.[771] Die positiven Aspekte seien der webbasierte Zugang, das Teilen und Wiederverwenden von Daten und die Datenintegration. Als Herausforderungen gelten die Umsetzung, die Schwierigkeit der Technologien, die Anwendung der Tools, das Training der Mitarbeiter*innen und der Konsum von Linked Data.[772]

7.2 LOD in der Archäologie – Institute, Projekte, Plattformen und Aggregatoren mit LOD oder semantischen Technologien

Besonders die Antikenforschung hat bereits sehr viele Projekte, die Linked Open Data Technologien verwenden, wie die iDAI.world und die Gazetteers, realisiert. Es haben sich Standards, von der Numismatik über die Epigraphie, etabliert. Auch das Pelagios-Projekt hat viel dazu beigetragen. Die Plattform und Tools zur Annotation sind benutzerfreundlicher als vorangegangene. Archäologische Datensets wurden bisher noch eher selten als Linked Open Data umgesetzt.[773] Es gibt einige Punkte, bei denen sich die Möglichkeiten für Projekte und Motivationen des zusammengefassten Gebietes der Institute, Galerien, Bibliotheken, Archiven und Museen, mit denen der Archäologie unterscheiden.[774]

Institute, Museen etc. tendieren eher zu langfristigen Zielen, wobei Investitionen in Richtung Ausbildung und Ausstattung oder in die Verbesserung der Informationsvermittlung geht. Die Verantwortung dieses Sektors liegt vor allem in der Kontextualisierung ihrer Objekte, bevor sie der Öffentlichkeit präsentiert werden und nicht in der Freigabe der Rohdaten.[775] Sie möchten ihre Inhalte für die Öffentlichkeit und für Forscher*innen zur Verfügung stellen.[776] Die Finanzierung geschieht teilweise durch Einnahmen über den Inhalt der Ausstellungen wie beispielsweise Postkarten, digitale Bilder, Veranstaltungen und Eintrittsgelder. Daher wird die freie Veröffentlichung von Fotos und Informationen öfter als Risiko der Einnahmen gesehen. Andere Hindernisse können Urheberrechtsregeln oder die nationale Rechtslage bei nationalen Institutionen sein. Die

[770] Geser 2016, 11–16.
[771] Isaksen 2011, 64–66.
[772] Isaksen 2011, 65 f.
[773] Geser 2016, 10.
[774] Isaksen 2011, 36.
[775] Isaksen 2011, 36.
[776] Geser 2016, 10.

Grundideen der Dezentralisierung und Offenheit können daher schwierig umzusetzen zu sein für diesen Sektor.[777] Als Beispiel für die Dissemination mit Online-RDF- und Bilddaten kann die Umsetzung der Niederösterreichischen Landessammlungen dienen.[778]

In der Archäologie hingegen sind Projekte oft abhängig von fixierten Förderungen über einen kurzen Zeitraum, wobei der Output im Vorfeld meist schon abgesteckt ist. Hier bleiben oft wenige Ressourcen übrig für die anschließende Dissemination. Es ist des Öfteren wenig Budget für Datenmanagement-Technologien oder langfristige Datenmanagement-Pläne übrig. Auf der anderen Seite gibt es auch Langzeitausgrabungen, die in internationalen Projekten sehr wohl ihre Ressourcen für digitale Investitionen verwenden wollen und dies auch tun.[779] Doch die Ausgangssituation ist eine kompliziertere. Die Basis der meisten Forschung bilden Ausgrabungen und Fundstellen, wo Strukturen, Fundobjekte und biologisches Material erforscht werden. Bisher wurden wenige alleinstehende Ausgrabungen oder Prospektionen als semantische Daten publiziert. Man findet sie eher bei Datenbanken von Forschungsinstituten oder Datenrepositorien.

> „[…] much more needs to be done to raise awareness of the approach, promote uptake, and provide practical guidance and easy to use tools for the generation, publication and interlinking of Linked Data.“[780]

Linked Data Expert*innen in der Archäologie möchten die Technologien präsenter machen, indem verbesserte Tools zur Anwendung für Archäolog*innen entwickelt werden.[781] Die Verwendung von semantischen Technologien in der Archäologie teilt L. Isaksen in die Perioden 2001-2005 und Post-2005 ein.[782] Die Ontologie CIDOC-CRM wurde bereits in der ersten Zeitperiode von 2001-2005[783] regelmäßig verwendet und gilt als der wichtigste Entwicklungspunkt. Das möglicherweise erste semantische Projekt, das URIs und RDFs verwendet hat, war das VBI-ERAT-LUPA[784]. Dies ist eine online zugängliche Datenbank zu antiken Steindenkmälern in Mittel- und Südosteuropa.

Ab dem Jahr 2005 gab es vermehrt Workshops zu CIDOC-CRM. Es entwickelten sich viele unterschiedliche und vielseitige Projekte. L. Isaksen teilt die Projekte grob in folgende Kategorien

[777] Isaksen 2011, 36.
[778] Siehe https://visualisierung.landessammlungen-noe.at/, abgerufen am 21.10.2021.
[779] Isaksen 2011, 41.
[780] Geser 2016, 11.
[781] Geser 2016, 11.
[782] Isaksen 2011, 40–46.
[783] Für eine detaillierte Auflistung der Entwicklung von semantischen Technologien in der Archäologie siehe Isaksen 2011, 40–46.
[784] Siehe http://lupa.at/, abgerufen am 21.10.2021.

ein, wobei die Übergänge oft fließend sind: technische Plattformen, archäologische Forschungsprojekte und Aggregatoren.[785] An dieser Stelle sei darauf hingewiesen, dass die Vielzahl an Projekten beinahe unüberblickbar geworden ist und hier nur Beispiele zur Anschauung genannt werden sollen.

7.2.1 Technische Plattformen

Projekte, die technische Plattformen hervorbringen, versuchen, beliebige archäologische Informationen und Daten aufzunehmen, zu organisieren und deren Abfrage möglich zu machen. Ein Beispiel hierfür wäre ArcheoInf[786], ein Projekt, welches zum Ziel hatte, das gleichzeitige Durchsuchen von wissenschaftliche Primärdaten aus heterogenen Datenstrukturen möglich zu machen. Es sollte als Mediator fungieren, um archäologische Primärdaten, bibliografische Informationen und Geodaten einsehen zu können und mit nur einer Benutzeroberfläche abfragen zu können.

7.2.2 Archäologische Forschungsprojekte

Archäologische Forschungsprojekte, die semantische Technologien verwenden, sind sehr vielseitig und die Gründe für ihre Verwendung variieren. Einige möchten zwischen Funden, Berichten und Daten bessere Verbindungen herstellen, um effizienter zu suchen, zu forschen und Daten öffentlich zugänglich zu machen (wie beispielsweise die Athenian Agora Excavations[787]). Andere Projekte ermöglichen Spezialisten, direkt auf der Webseite Abfragen durchzuführen (zum Beispiel Ilion (Troia)[788]). Das Projekt A Puzzle in 4D[789,790] vereinte die Digitalisierung, Langzeitarchivierung und Dissemination der Ausgrabungen in Tell el-Daba und generierte einen SPARQL-Endpoint zum Abfragen der Daten.[791]

Das Forschungszentrum HiMAT[792] der Universität Innsbruck ging aus dem interdisziplinären Spezialforschungsbereich für die Erforschung des spätbronzezeitlichen bis früheisenzeitlichen Kupfererzbergbaus im Maukental bei Radfeld und Brixleg im Bezirk Kufstein hervor.[793] Dabei wurden Daten und Geoinformationen von durchgeführten Prospektionen, archäologischen Ausgrabungen, Begehungen und dendrochronologischen Analyseergebnissen mithilfe von CIDOC

[785] Isaksen 2011, 41 f.
[786] Siehe http://archeoinf.tu-dortmund.de/, abgerufen am 21.10.2021.
[787] Siehe http://agora.ascsa.net/research?v=default, abgerufen am 21.10.2021.
[788] Siehe https://classics.uc.edu/troy/grbpottery/, abgerufen am 21.10.2021.
[789] Siehe https://4dpuzzle.orea.oeaw.ac.at/index/, abgerufen am 21.10.2021.
[790] Aspöck u. a. 2015, 675.
[791] Aspöck u. a. 2020, 79–81.
[792] Siehe https://www.uibk.ac.at/himat/, abgerufen am 21.10.2021.
[793] Goldenberg u. a. 2011, 61.

CRM modelliert und in Kontext zueinander gestellt (siehe Kapitel 5.2.4.6 und 5.2.4.7).[794,795,796] Die RDF-Repräsentation dieser Daten ermöglicht die weitere Verarbeitung mit semantischen Technologien und Integration mit anderen Daten.[797] Diesem Projekt nachfolgend wurden weitere Projekte verfolgt, welche sich mit der Integration von Forschungsdaten beschäftigen. Das Open Research Data for Prehistoric Mining Archaeology (Offene Forschungsdaten für prähistorische Bergbauarchäologie)[798] hat es sich zum Ziel gesetzt, Daten aus dem multinationalen DACH-Projekt für prähistorische Kupferproduktion (Prehistoric copper production in the eastern and central Alps)[799] semantisch aufzubereiten und frei zugänglich zur Verfügung zu stellen. Das Projekt haltet sich dabei an die Empfehlungen der FAIR Richtlinien (siehe Kapitel 2.1.2), um dieses Inventar von prähistorischen Fundstellen und ihrer Forschungsergebnisse für neue Forschungen wiederverwendbar zu machen.[800] Darüber hinaus werden Daten über Fundstellen, Fundobjekte, ihrer Interpretationen und ihrer geographischen Daten mithilfe der Ontologie CIDOC CRM miteinander verlinkt. So können neue Sichtweisen gewonnen oder neue Fundstellen entdeckt werden, welche durch weiterführende Prospektionen untersucht werden können.[801] In dem laufenden Projekt Information Integration for Prehistoric Mining Archaeology (Informationsintegration für prähistorische Bergbauarchäologie)[802] werden Daten montanarchäologischer Forschung des Forschungszentrums HiMAT und Fernerkundungsdaten mit semantischen Technologien integriert. Gemeinsam mit digitalen Geländemodellen und den integrierten Daten sollen neue Fundplätze mit prähistorischen Bergbauspuren entdeckt und untersucht werden.[803,804,805] Die Daten dieser Projekte eignen sich gut für die Modellierung mit CIDOC CRM, da es sich um ein eventzentriertes Modell handelt. In den Datenmodellen werden prähistorische Aktivitäten des Bergbaus und Forschungsaktivitäten mit ihrer Dokumentation behandelt. Erweiterungen des CIDOC CRM (siehe Kapitel 5.2.2) werden hinzugenommen für naturwissenschaftliche Daten (hierfür CRMsci), Interpretationen (CRMinf), geometrische Informatrionen (CRMgeo) und bereits vorhandene digitale Dokumentation (CRMdig).[806] Andere Projekte mit Anwendung semantischer Technologien des Forschungszentrums orientieren sich beispielsweise an ontologischen Modellen für Sprache. Das

[794] Hiebel – Hanke 2009, 653–657.
[795] Hiebel u. a. 2010, 405–408.
[796] Hiebel u. a. 2014, 558–560.
[797] Hiebel u. a. 2013, 8 f.
[798] Siehe https://www.uibk.ac.at/projects/ord4mining-archaeo/, abgerufen am 21.10.2021.
[799] Siehe https://www.uibk.ac.at/himat/projekte/dach/index.html.de, abgerufen am 21.10.2021.
[800] Hiebel u. a. 2021a, 267-269.
[801] Hiebel u. a. 2018, 192 f.
[802] Siehe https://www.uibk.ac.at/projects/inf-integ4pre-min-archaeo/, abgerufen am 21.10.2021.
[803] Hiebel – Hanke 2017, 260.
[804] Hiebel u. a. 2017, 692–697.
[805] Hiebel u. a. 2018, 183 f.
[806] Hiebel u. a. 2017, 695.

Semantics for Mountaineering History Projekt[807] hat den Alpenwortkorpus[808] mit Ortsnamen, Personen und Erstbesteigungen mit semantischen Annotationen versehen. So können Suchanfragen gezielt eingegeben werden, wie beispielsweise alle Personen, welche im Jahr 1914 den Großglockner bestiegen haben. Das Projekt Text Mining Medieval Mining Texts[809] sammelt Informationen aus historischen Textquellen zu mittelalterlichen Bergbaugebieten in Tirol. Die Beziehungen der Informationen wie Personennamen, Ortsnamen und Grubennamen werden verlinkt. Dadurch entsteht ein Netzwerk, welches mithilfe von geographischen Informationssystemen die Entwicklung des Bergbaugebietes beleuchten kann.

Ein weiteres innovatives Forschungsprojekt, welches die Ontologie CIDOC CRM für die semantische Aufbereitung von Forschungsdaten verwendet, ist mit dem Naturhistorischen Museum Wien und dem Österreichischen Archäologischen Institut der Österreichischen Akademie der Wissenschaften verbunden. Thanados[810] (The Anthropological and Archaeological Database of Sepultures) ist ein Projekt, welches eine digitale Sammlung von Gräberfeldern des Frühmittelalters in Österreich aufbereitet hat. Die Daten wurden mithilfe der CIDOC-CRM Ontologie modelliert und sind neben interaktiven Karten und anderen Visualisierungen frei zugänglich. Die Daten können sowohl abgefragt als auch exportiert werden.[811] Dabei wurde das Datenbanksystem OpenAtlas verwendet, welches im Zuge vorangegangener Projekte entwickelt wurde. Darin werden die Daten eingegeben und durch das im Hintergrund mitlaufende Datenmodell gleichzeitig in CIDOC CRM modelliert. Die Software ist frei zugänglich und kann von anderen Projekten weiterverwendet werden.[812]

Einige Forschungsprojekte haben ihren Schwerpunkt auf der Aufarbeitung und Dissemination geschlossener Ausgrabungen während andere mehrere inhomogene Datensätze kombinieren, um eingegrenzte Themengebiete abzudecken.[813]

[807] Siehe http://sprawi.at/de/semohi, abgerufen am 21.10.2021.
[808] Der Alpenwortkorpus wurde im Projekt „Alpenwort" der Österreichischen Akademie der Wissenschaften aus der Zeitschrift des Deutschen und Österreichischen Alpenvereins (ZAV) erstellt, siehe unter: http://alpenwort.at/, abgerufen am 21.10.2021.
[809] Siehe http://sprawi.at/de/tmmmt, abgerufen am 21.10.2021.
[810] Siehe https://thanados.net/, abgerufen am 21.10.2021.
[811] Eichert 2021, 2–4.
[812] Filzwieser – Eichert 2020, 1385–1390.
[813] Isaksen 2011, 43.

7.2.3 Aggregation und Dissemination

Viele Projekte haben sich die Aggregation und Dissemination von archäologischem Material zum Ziel gemacht. Es werden Daten aus unterschiedlichen Sammlungen, Datenbanken und institutionellen Beständen zusammengesammelt und meist mittels RDF und URIs der Öffentlichkeit für Untersuchungen und Abfragen zur Verfügung gestellt.

Das Nomisma Projekt[814] bietet stabile digitale Repräsentationen nach den Linked Open Data Prinzipien und Methoden des Semantic Web. Es verwendet RDF und HTTP-URIs von Münzdaten und numismatischen Konzepten von vielen unterschiedlichen Sammlungen und kreiert Verlinkungen, eine Ontologie und bietet eine SPARQL-Abfragefunktion auf der Webseite[815].[816] Die Daten des Nomisma Projekts können ebenfalls, neben vielen anderen Datenquellen[817], in der Peripleo-Suchmaschine gesucht werden.

Peripleo[818] ist eine spatio-temporale Suchmaschine und ein Visualisierungstool. Es wird verwaltet vom Pelagios Projekt. Das Pelagios Netzwerk[819] verlinkt Informationen online durch semantische Annotationen.[820]

Daneben gibt es noch einige nationale Archivierungsinitiativen, die Daten oder Metadaten in semantischen Formaten bereitstellen und archivieren.[821] Das sind beispielsweise das ADS (Archaeology Data Service)[822], das ein digitales Langzeitrepositorium für Kulturgut-Daten ist und diese präserviert und disseminiert. Sie geben Ratschläge und Unterstützung für digitale Good Practice Methoden. Unter die Archivierungsinitiativen fällt auch tDAR (Digital Archaeological Record)[823], ein internationales digitales Repositorium.

Die zentrale Bilddatenbank Arachne[824] des Deutschen Archäologischen Instituts und des Archäologischen Instituts der Universität zu Köln bietet ein freies Werkzeug zur Recherche über hunderttausende von Datensätzen archäologischer Objekte. Analoge Dokumentationsbestände werden digitalisiert und eingepflegt und neue digitale Daten kommen laufend hinzu. Maschinenlesbare Metadaten und semantische Technologien bilden die Struktur des Systems.

---

[814] Siehe http://nomisma.org/, abgerufen am 21.10.2021.
[815] Siehe http://nomisma.org/sparql, abgerufen am 21.10.2021.
[816] Wigg-Wolf – Duyrat 2017, 1.
[817] Für eine Liste der Partner und Datenquellen von Peripleo siehe https://peripleo.pelagios.org/about, abgerufen am 21.10.2021.
[818] Siehe https://peripleo.pelagios.org/, abgerufen am 21.10.2021.
[819] Siehe https://pelagios.org/, abgerufen am 21.10.2021.
[820] Simon u. a. 2016, 1.
[821] Isaksen 2011, 44.
[822] Siehe https://archaeologydataservice.ac.uk/about.xhtml, abgerufen am 21.10.2021.
[823] Siehe https://www.tdar.org/about/, abgerufen am 21.10.2021.
[824] Siehe https://arachne.uni-koeln.de/drupal/, abgerufen am 21.10.2021.

Unter dem ADS laufen viele weitere Projekte.[825] Das SEADDA Projekt (Saving European Archaeology from the Digital Dark Age) setzt sich dafür ein, digitale archäologische Daten zu erhalten, verbreiten und wiederzuverwenden. Sie agieren nach den FAIR-Prinzipien.[826] Das ARIADNE-Projekt und das Nachfolgeprojekt ARIADNE Plus[827] haben das Ziel, europäische, archäologische Datenspeicher zu integrieren. Sie haben einen durchsuchbaren Katalog von Datensets nach den Suchkriterien „Where", „When" und „What" mithilfe von Keywords und einem kontrollierten Vokabular entwickelt. ARIADNE bildet keine Aggregation von Daten selbst, sondern nur der Metadaten der einzelnen Datensets.[828] Das Archaeotools Projekt[829] des ADS hatte einen ähnlichen Ansatz. In diesem Projekt wurden über eine Million Metadateneinträge von archäologischen Fundstellen des ADS nach den Kriterien „When, What, Where" indexiert und öffentlich zugänglich gemacht. Informationen aus der Grauen Literatur wurden ebenfalls extrahiert und hinzugefügt.[830] Die STAR (Semantic Technologies for Archaeological Resources) und STELLAR (Semantic Technologies Enhancing Links and Linked data for Archaeological Resources) Projekte haben wissenschaftliche archäologische Daten semantisch verfügbar gemacht. Das Ziel des STAR Projektes war die semantische Interoperabilität für Datenbanken und Graue Literatur zu erreichen. Dafür wurde die CIDOC-CRM Ontologie und andere Knowledge Organisation Systeme (wie beispielsweise Klassifizierungssysteme, Thesauri, Ontologien, Ortsverzeichnisse) verwendet. Das STELLAR Projekt hat die Repräsentation von Ausgrabungsdaten mit CIDOC-CRM für Linked Data entwickelt.[831]

7.2.4 Linked Open Data und Wissens-Repräsentation aus mehreren Quellen

L. Isaksen nennt zwei Visionen, die zur Weiterentwicklung des Semantic Web beigetragen haben. Die Terminologie ist nicht ganz festgelegt, es ist eher eine Tendenz in der Ausführung von Projekten und die Motivation und Ziele der Anwender von semantischen Technologien.

Die erste Vision lässt sich mit T. Berners-Lees Idee der Linked Open Data beschreiben. Dieser Ansatz setzt sich für Offenheit, Dezentralisierung und die Verbindung von unabhängigen Ressourcen ein. L. Isaksen bezeichnet die zweite Vision als „Mixed-Source Knowledge Representation" (MSKR), also Wissens-Repräsentation aus mehreren gemischten Quellen. Es werden meist formale Beschreibungen und Schlussfolgerungen von mehreren heterogenen Datenbeständen

[825] Für eine komplette Liste der Projekte des ADS siehe https://archaeologydataservice.ac.uk/research/projects.xhtml, abgerufen am 21.10.2021.
[826] Siehe https://www.seadda.eu/, abgerufen am 21.10.2021.
[827] Siehe https://whatis.ariadne-infrastructure.eu/de/, abgerufen am 21.10.2021.
[828] ARIADNE Plus 2019, 2.
[829] Siehe https://archaeologydataservice.ac.uk/research/archaeotools.xhtml, abgerufen am 21.10.2021.
[830] Jeffrey u. a. 2009, 2507–2512.
[831] Tudhope u. a. 2011, 15.

ausgeführt und in einem eher geschlossenen digitalen Umfeld gehandhabt. Beide Ansätze werden in der Literatur und in der Praxis oft miteinander verwoben und fast jedes Projekt, das mit semantischen Techniken arbeitet, verwendet in irgendeiner Art beide Visionen.[832]

J. D. Richards hat zu diesem Thema geschrieben, das Semantic Web würde zwar nicht zwingend Open Data benötigen, die Weiterentwicklung und die Verbreitung dieser Technologien aber sehr wohl erheblich erleichtert. Die Debatte sollte sich nicht nur um Open Access Publikationen drehen, sondern auch um Open Repositories.[833]

[832] Isaksen 2011, 45 f.
[833] Richards 2006, 971.

8 Case Study - Anwendungsschritte zur Datenmodellierung und Datenintegration

In diesem Kapitel wird anhand einer Funddatenbank zu Metallfunden des Oberleiserbergs ein Weg aufgezeigt, diese Daten in ein Linked Data-Format zu konvertieren. Zunächst werden die Daten, welche die Basis für die Anwendung bilden, näher betrachtet. Im Anschluss werden die Schritte im Detail erläutert, welche zum RDF-Graphen führen.

8.1 Datenbasis für die Modellierung mit CIDOC CRM

8.1.1 Archäologische Beschreibung des Fundortes Oberleiserberg

Der Oberleiserberg bei Ernstbrunn im Weinviertel (Niederösterreich) ist mit 457 m Meereshöhe die vierthöchste Erhöhung der Leiserberge. Er bildet ein in etwa 6.5 – 7 ha großes Plateau, welches leicht nach Westen geneigt ist (siehe Abbildung 37). Die verkehrstechnisch vorteilige Lage und die erhöhte Position machten ihn durch die Zeitepochen hindurch zu einem oft genutzten Siedlungsplatz, einer Handels- und Produktionsstätte, einem Bestattungsplatz und einem religiösen Versammlungsort.[834]

Abbildung 37 – Der Oberleiserberg in SSO-Ansicht (Von Bwag - Eigenes Werk, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=61101660).

[834] Stuppner 1997, 282; Kern 1987, 3; Lochner 2006, 103 f.

Die Fundstelle ist bereits im 19. Jahrhundert durch M. Much entdeckt worden. Erste detaillierte Untersuchungen wurden von 1925 bis 1931 und 1933 durch H. Mitscha-Märheim und F. Nischer-Falkenhof durchgeführt.[835] Von 1976 bis 1990 wurden weitere Ausgrabungen unter H. Friesinger und ab dem Jahr 1980 unter der Leitung von A. Kern durch das Institut für Ur- und Frühgeschichte[836] durchgeführt.[837] Ab 1996 konnten die Ausgrabungen über das durch den Fond zur Förderung der Wissenschaftlichen Forschung (FWF) finanzierte Projekt „Der Oberleiserberg – eine spätantike Höhensiedlung" zur völkerwanderungszeitlichen Besiedlung durch A. Stuppner wieder aufgenommen werden und wurden bis zum Jahr 2009 fortgesetzt.[838] Um Siedlungsspuren auffinden zu können, wurden auch verschiedene Prospektionenmethoden angewandt. M. Doneus vom Institut für Urgeschichte und Historische Archäologie führte Luftbildaufnahmen durch, wodurch Grundrisse von Gebäuden sowie Gruben und Gräben ausfindig gemacht werden konnten. Geophysikalische Prospektionen durch P. Melichar und W. Neubauer auf dem Plateau ließen erkennen, dass das gesamte Plateau besiedelt gewesen sein muss.[839] 2001 bis 2003 wurden im Zuge des FWF-Projektes „Der Oberleiserberg in der Bronzezeit" der Österreichischen Akademie der Wissenschaften unter der Leitung von M. Lochner die bronzezeitlichen Funde und Befunde aufgearbeitet.[840]

Archäologische Spuren finden sich am Oberleiserberg bereits ab dem späten Neolithikum. Ab der Frühbronzezeit lassen sich Befunde feststellen, welche auf eine Besiedlung der Aunjetitzkultur ab ca. 2300 v. Chr. hinweisen. Nach dieser Siedlungsphase wurde der Oberleiserberg wieder in der späten Bronzezeit, am Ende der älteren Urnenfelderzeit (Ha A1), besiedelt. Die Siedlung wurde wahrscheinlich mit einem Wall aus einer Holz-Erde-Konstruktion gesichert. Die Befunde zeigen Spuren von Pfosten einiger Ständerbauten, eingetiefte Hütten, Siedlungsgruben und Herd- bzw. Ofenplatten.[841] Die gebrannten Lehmplatten und der Kuppelofen zeigen eventuell Reste eines Werkstättenbereichs an, in welchem Keramik hergestellt oder Metall verarbeitet worden sein könnte.[842] Im Umkreis der Ofenrückstände wurden auch Reste von Tongussformen für Ringe gefunden. Die Rekonstruktion einer dieser Gussformen ergab 31 Abdrücke für Ringe.[843] Die Formen, die Ofenreste sowie auch eine Vielzahl an Buntmetallabfällen, wie etwa Metalltropfen, Schlacke und Rohprodukte, deuten auf die Produktion von Metallobjekten zu dieser Zeit hin.

[835] Siehe dazu Mitscha-Märheim – Nischer-Falkenhof 1929.
[836] Heute: Institut für Urgeschichte und Historische Archäologie.
[837] Friesinger 1978, 423 f.; Kern 1982, 513 f.; Kern 1986, 293; Kern 1988, 294; Kern 1990, 230.
[838] Stuppner 1999, 833–836; Stuppner 2001, 698–701; Stuppner 2003, 693–695; Stuppner 2005, 957–961; Stuppner 2007, 718 f.; Stuppner 2009, 604 f.; Stuppner 2010, 461 f.; Stuppner 2011, 296 f.
[839] Stuppner 2006, 9–11.
[840] Kern 2003, 62 f.; Lochner 2006, 103.
[841] Hellerschmid u. a. 2010, 284.
[842] Kern 2013, 105 f.
[843] Lochner 2006, 104–115.

Die Fundobjekte zeigen Keramik, Spinnwirtel, Webgewichte, Reibplatten, Steingeräte und Buntmetallobjekte. Die Fundstücke aus Buntmetall weisen vor allem Nadeln auf, welche typischerweise der Urnenfelderzeit zugeordnet werden, wie Vasenkopfnadeln, Spindelkopfnadeln und Rippenkopfnadeln. Weiters kamen Messer, Pfrieme, Nähnadeln und Werkzeuge zum Vorschein.[844]

Als Hauptphase dieser Besiedlung kann die jüngere Urnenfelderzeit (Ha B1) angesehen werden. Verlassen wurde die Siedlung in etwa am Ende der Stufe Ha B2.[845] Danach wurde die Besiedlung in der Latènezeit in etwa ab der zweiten Hälfte des 3. Jahrhunderts v. Chr. wieder aufgenommen. In der Spätantike wurde ab der zweiten Hälfte des 4. Jahrhunderts n. Chr. eine weitere Siedlung angelegt. Die Siedlung entwickelte sich zu einem germanischen Herrschaftssitz, welche durch die Nähe zum römischen Limes stark romanisiert wurde. Es wurde ein Herrenhof in mehreren Phasen errichtet, welcher in der zweiten Hälfte des 5. Jahrhunderts zerstört wurde.[846] Die weiteren archäologischen Befunde zeigen ein frühmittelalterliches Gräberfeld des 10. und 11. Jahrhunderts sowie eine Siedlung des 12. Jahrhunderts. Ab diesem Zeitpunkt wird der Oberleiserberg als Kirchen- und Wallfahrtsort angesehen.[847]

8.1.2 Datenbasis Oberleiserberg – Datenbank mit Daten zu Kleinfunden

Die Datenbank zu den Kleinfunden enthält Informationen über Fundobjekte der Ausgrabungen der Jahre 1996 bis 2009 durch A. Stuppner sowie zu einigen Altfunden der Ausgrabungen von H. Mitscha-Märheim und E. Nischer-Falkenhof. In der Microsoft Access Datenbank finden sich insgesamt 1.788 Datensätze. Diese setzen sich zusammen aus Daten zu Metallfunden (1.601 Datensätze), Glasfunden (148 Datensätze), Altfunden (30 Datensätze), Einzelfunden (1 Datensatz) und einem Fundprotokoll (8 Datensätze). Die Daten wurden von A. Stuppner in die Datenbank eingegeben. Im Zuge dieser Arbeit wurden 381 Datensätze der Metallfunde zur Datenmodellierung ausgewählt, welche sich dem Zeithorizont der Urnenfelderzeit zuordnen lassen.[848]

8.2 Methodik

Es wurden einige Anleitungen, sogenannte Cookbooks, veröffentlicht, um es Datenherausgeber*innen zu vereinfachen, Linked Data zu produzieren. Sie sollen Schritt für Schritt die Prozesse erklären, welche notwendig sind, um die Daten in RDF-Daten zu konvertieren. Die vom W3C

[844] Hellerschmid u. a. 2010, 284.
[845] Lochner 2006, 103 f.
[846] Stuppner 2013, 48–53.
[847] Stuppner 2006, 40–46.
[848] Vielen herzlichen Dank an dieser Stelle an Ass.-Prof. Mag. Dr. Alois Stuppner, der die Daten für diese Arbeit zur Verfügung gestellt hat.

im Jahr 2011 herausgegebene Anleitung wurde für zu veröffentlichende Regierungs- und Verwaltungsdaten (Open Government Linked Data) kreiert. Sie gibt den Anwender*innen sieben Schritte vor, um zu Linked Data zu gelangen.[849] 2014 wurde diese Anleitung noch um Anweisungen für die Veröffentlichung erweitert.[850]

L. Isaksen stellt in seiner Dissertation ein Cookbook für archäologische Daten vor, welches drei aufeinander aufbauende semantische Ebenen beschreibt. Dies hat den Vorteil, dass der bzw. die Anwender*in bei jeder Ebene stoppen kann. In der ersten Ebene werden die Daten harmonisiert und Standards zur Terminologie angewandt. Die zweite Ebene soll dazu dienen, die Daten mit URIs anzureichern. In der dritten Ebene werden die Daten in ein RDF-Format gebracht.[851]

Im Folgenden wird schrittweise gezeigt, wie die Konvertierung einer Funddatenbank einer Ausgrabung, wie sie in der Archäologie oft vorkommt, in Linked Data durchgeführt werden kann. Dabei wurde sich an den oben genannten Anleitungen angelehnt. Die sieben Schritte, um Linked Data zu produzieren, als Empfehlungen des WC3, sind:[852]

1. *Modellierung der Daten:* Als ersten Schritt werden die Daten ausgewählt, bereinigt und zu einer Ontologie modelliert.

2. *Identifizierung mit URIs:* Zur Identifikation der Objekte soll eine Strategie für die Benennung der Objekte mit URIs soll entwickelt werden. Dabei sollte darauf geachtet werden, stabile HTTP-URIs zu verwenden.

3. *Wiederverwendung von Vokabular und Standards:* Objekte sollen mit zuvor definierten Vokabularen beschrieben werden. Wann immer möglich, sollen Vokabulare, domänenspezifische Standards und Ontologien wiederverwendet werden, bevor neue kreiert werden.

4. *Beschreibung der Daten:* Die Daten sollen möglichst in maschinen- und menschenlesbarer Form selbstbeschreibend sein. Potenzielle Nutzer*innen sollen die Daten ohne weitere Erklärung verstehen können. Das Datenmodell, welches verwendet wurde, sollte angegeben werden.

5. *Konvertierung der Daten in RDF:* Wenn die Daten in einem Modell ausgedrückt wurden, können sie in eine Serialisierung des RDF konvertiert werden. Dabei sollte darauf geachtet werden, Links zu externen Informationen miteinzubinden.

[849] Hyland – Villazón-Terrazas 2011.
[850] Hyland u. a. 2014.
[851] Isaksen 2011, 105–149.
[852] Hyland – Villazón-Terrazas 2011, 4–13.

6. *Spezifizierung einer Nutzerlizenz:* Eine geeignete Lizenz sollte angegeben werden, damit Nutzer*innen wissen, wie die Daten weiterverwendet werden können. Dies erhöht die Wahrscheinlichkeit einer Wiederverwendung der Daten. Lizenzen von Creative Commons[853] eignen sich für Linked Data.

7. *Veröffentlichung der Daten:* Die Daten sollten auf einen öffentlich zugänglichen Server kopiert werden. Im besten Fall sollte ein SPARQL-Endpoint eingerichtet werden, um Nutzer*innen zu ermöglichen, die Daten abzufragen. Eine Möglichkeit, um die Daten vom Server zu laden, ist ebenfalls empfohlen.

8.2.1 Modellierung der Daten

8.2.1.1 Datenbereinigung und Harmonisierung

In diesem ersten Schritt werden die Daten ausgewählt, welche in Linked Data konvertiert werden sollen. Anschließend werden die Begriffe in diesen Daten vereinheitlicht. Dies bereitet die Daten für die weiteren technischen Schritte vor und erleichtert die Konvertierung, da die Daten Konsistenz erhalten und die Struktur standardisiert wird.[854]

Die Harmonisierung ist notwendig, da innerhalb einer Datenbank oft unterschiedliche Terminologien für die gleichen Konzepte verwendet werden. Es werden etwa Ausdrücke für Perioden als Synonyme verwendet, wie beispielsweise „Urnenfelderzeit“ und „UK“. Dies geschieht unter anderem vermehrt, wenn Ausgrabungskampagnen über mehrere Jahre andauern, verschiedene Personen mit der Dokumentation und Aufarbeitung betraut sind oder mehrere Typologien verwendet werden.[855]

Daher sollte in diesem Schritt ein Thesaurus bzw. ein Vokabular ausgewählt und die Daten angepasst werden. Es sollte darauf geachtet werden, ein Konzept mit einem Ausdruck zu beschreiben. Die Struktur, in der die Objekte dokumentiert wurden, sollte standardisiert werden. Dabei ist es am effektivsten, wenn ein Fundobjekt in einer Zeile erfasst wird. Nach diesem Harmonisieren können die Daten einfacher mit anderen Daten, welche auf die gleiche Weise bearbeitet wurden, besser verglichen werden und die Modellierung der Daten wird vereinfacht.[856]

[853] Siehe https://creativecommons.org/licenses/, abgerufen am 21.10.2021.
[854] Isaksen 2011, 110.
[855] Aspöck 2020, 16.
[856] Isaksen 2011, 110.

Die konkreten Arbeitsschritte zur Datenbereinigung orientieren sich an der ersten semantischen Ebene des Cookbooks von L. Isaksen.[857] Aus der vorhandenen Access-Datenbank zu den Fundobjekten wurde eine Excel-Tabelle extrahiert.[858] So kann die Originaldatenbank archiviert und gesichert werden. Die Änderungen sollten auf einer Kopie der ursprünglichen Daten geschehen, um Datenverlust in jedem Fall zu vermeiden. Danach sollte kontrolliert werden, ob und wie alle Spalten benannt sind. Die Benennungen sollten eindeutig sein und keine Abkürzungen, Codierungen (wie Umlaute) oder Leerzeichen enthalten. Als Beispiel für eine Transformation in der Datenbank zum Oberleiserberg können folgende genannt werden (siehe Abbildung 38):

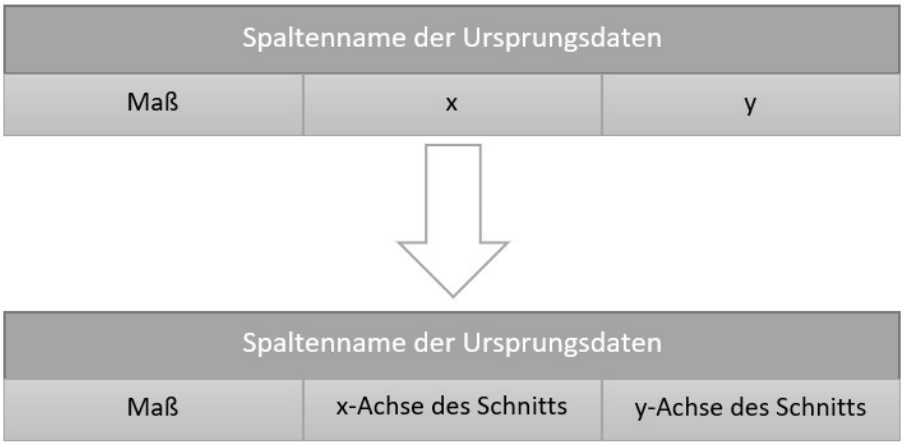


Abbildung 38 – Harmonisierung der Spaltennamen (Grafik: M. Simon).

Danach können die Spalten und Zeilen strukturiert werden. Jeder Fund sollte eine Zeile einnehmen, da dies für die Modellierung vorteilhaft ist. Es ist sinnvoll, neue Spalten hinzuzufügen, wenn die Ansprache spezifiziert werden soll oder sich mehrere Informationen in einer Zelle befinden. Im Folgenden werden Beispiele für die weitere Harmonisierung gegeben.

Die *Ansprache und der Erhaltungszustand*: Oftmals wird der Begriff „-fragment" mit in die Ansprache übernommen. Dieser beschreibt aber eigentlich den Erhaltungszustand eines Fundobjektes. In einer Spalte „Ansprache" könnte beispielsweise „Nadel" stehen. In einer weiteren Spalte kann der Erhaltungszustand festgehalten werden. Handelt es sich um ein Fragment, ist hier der Platz, dies zu dokumentieren, anstatt in der Ansprache den Begriff „Nadelfragment" oder „Nadelbruchstück" zu verwenden (siehe Abbildung 39).

[857] Isaksen 2011, 110–122.
[858] Die Verwendung der freien Software OpenOffice Calc (siehe unter https://www.openoffice.org/product/calc.html, abgerufen am 21.10.2021) kann ebenso verwendet werden.

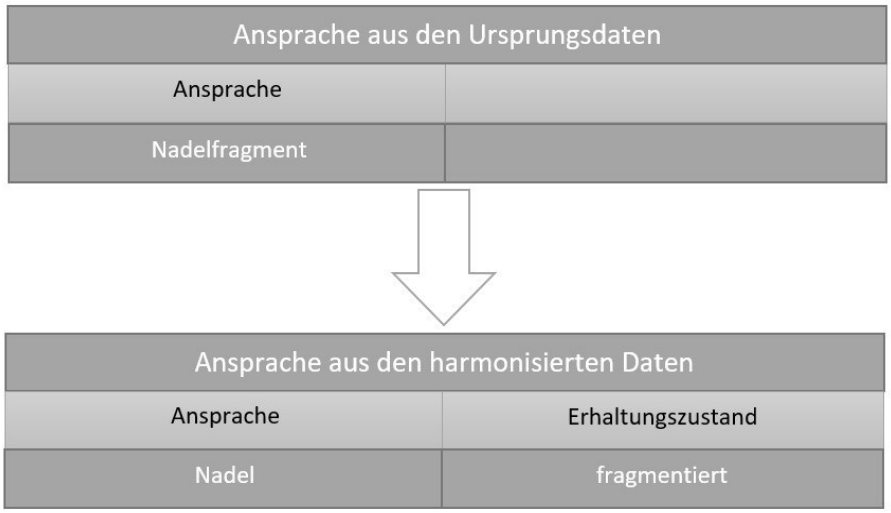


Abbildung 39 - Harmonisierung der Ansprache der Fundobjekte (Grafik: M. Simon).

Abmessungen der Fundobjekte: Um Messdaten analysieren zu können, ist es vorteilhaft, für jeden Wert eine Spalte zu reservieren, welche im Anschluss mit der Ontologie definiert werden kann. Zusätzlich sollte eine Spalte für die Maßeinheit hinzugefügt werden. Anstatt eine Spalte zu haben, wie (siehe Abbildung 40):

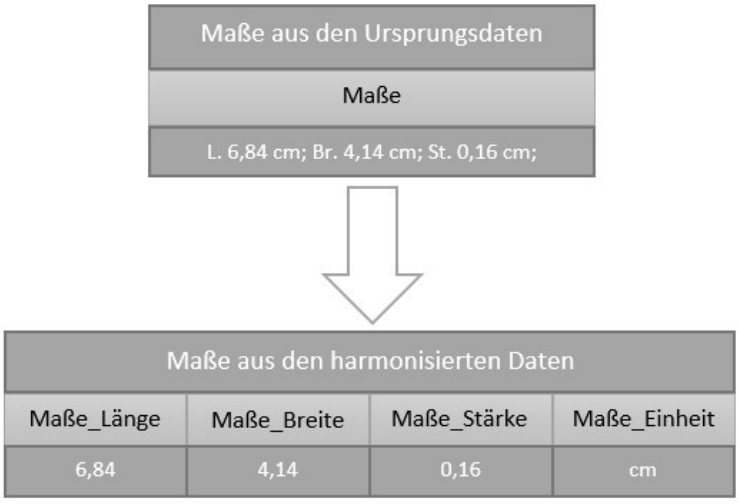
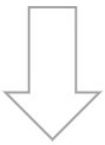


Abbildung 40 - Harmonisierung der Maße der Fundobjekte (Grafik: M. Simon).

Datierungen: Hier eignet sich eine Spalte zur kulturellen Zuordnung, zwei Spalten zur Beschreibung des frühesten und spätesten Jahrhunderts bzw. Periode und zwei Spalten mit absoluten Daten, welche einen zeitlichen Bereich eingrenzen können (siehe Abbildung 41). Dies erhöht die Vergleichbarkeit der Daten und das Erstellen von visuellen Graphen. Die absoluten Daten sollten dabei für Datierungen vor unserer Zeitrechnung mit einem Minus versehen werden, anstatt „v. Chr." oder „BC" anzugeben. Um weitreichende Angaben, wie beispielsweise „8. Jahrhundert v. Chr.", in absolute Zahlen umzuwandeln, wird der weiteste Datierungsrahmen angegeben, um

nicht inakkurat zu werden. Die genaue Ausführung kann projekt- und datenabhängig entschieden und umgesetzt werden.

Datierung aus den Ursprungsdaten	
Kulturelle Zuordnung	Datierung
Urnenfelderzeit	Späte Bronzezeit
	Mittlere – Späte Bronzezeit
Römische Kaiserzeit	Spät. 1. – Spät. 2. Jh.



Datierung aus den harmonisierten Daten				
Kulturelle_Zuordnung	Datierung_von	Datierung_bis	Absolute_Datierung_von	Absolute_Datierung_bis
Urnenfelderzeit	Bz D	Ha B2/3	-1300	-800
	Mittlere Bronzezeit	Späte Bronzezeit	-1650	-800
Römische Kaiserzeit	Spätes 1. Jahrhundert AD	Spätes 2. Jahrhundert AD	60	190

Abbildung 41 - Harmonisierung der Datierung der Fundobjekte (Grafik: M. Simon).

Unsicherheit: Um auszudrücken, dass Einträge nicht gesichert sind, wird oft ein Fragezeichen von Datenbearbeiter*innen eingetragen. Unsicherheiten digital maschinenlesbar auszudrücken, ist nicht einfach zu dokumentieren, da es sehr viele Nuancen der Unsicherheit gibt. Meist handelt es sich um einen unsicheren Wert, wie eine Datierung oder eine Ansprache, oder um eine „Entweder-Oder"-Angabe, wenn zwischen zwei Eingaben geschwankt wird. Wenn es sich um einen unsicheren Wert handelt, kann dieser separat eingegeben werden, damit der Computer dies interpretieren kann. Dafür könnte beispielsweise eine Spalte mit „Unsicherheit" hinzugefügt werden, in der ein „Ja" oder „Nein" eingetragen werden kann. Doch auch dann wäre es nicht klar, bei welchem Wert die Unsicherheit vorhanden ist. Die Fragezeichen sollten jedoch gelöscht werden. Wenn zwischen zwei Werten geschwankt wird, kann entweder ein Wert ausgewählt werden und dann mit einem Unsicherheits-Feld mit „Ja" gekennzeichnet werden, oder den Wert als „unidentifiziert" eintragen. Bei beiden Varianten könnte ein Informationsverlust angezeigt werden, daher könnte zusätzlich ein literales Feld „Anmerkungen" ausgefüllt werden.

Begriffe harmonisieren: Zeichenabfolgen sind wichtig für die Maschinenlesbarkeit und das Abfragen der Daten, daher soll möglichst immer genau die gleiche Bezeichnung für ein Konzept verwendet werden und einheitlich eingetragen werden. Es sollte jedoch auch kontrolliert werden, ob

sich Tippfehler oder Abkürzungen in den Daten befinden. Als besonders hilfreich zur Überprüfung ist hier die Filter-Funktion in Excel. Beispiele für die Harmonisierung sind:

- Groß- und Kleinschreibung („bronzezeit“)
- Zahlen („Ha A2“, „Ha AII“)
- Abkürzungen („BZ“, „UK“)
- Leerzeichen („HaA2“, „SpäteBronzezeit“)
- Umlaute und Accents („Latène“, „Latene“)
- Typographische Fehler („Unenfelderzeit“)
- Zusatzinformationen („Latène (lokal?)“)

Fundort hinzufügen: Da viele Datenbanken von einer einzigen Fundstelle stammen, findet sich diese Information nicht mehr direkt in den Daten wieder. Daher wird eine zusätzliche Spalte mit dem Fundort hinzugefügt, damit dieser zu jedem Fundobjekt zugeordnet werden kann.

Eigentum hinzufügen: Auch der rechtliche Besitz der Fundobjekte sollte in die Daten einfließen. Dies wird ebenfalls mit einer zusätzlichen Spalte festgehalten.

Tabellen trennen: Für die spätere Modellierung der Daten ist es vorteilhaft, wenn die Hauptentität in einer eigenen Tabelle angeführt wird. Bei einer Funddatenbank sind dies die Funde, bei einer Ausgrabungsdatenbank sind es beispielsweise die stratigraphischen Einheiten.

Daten filtern und kontrollieren: Als letzten Schritt dieses Arbeitsprozesses werden die Daten aussortiert und gefiltert. Dies ist nur dann notwendig, wenn beispielsweise nur eine bestimmte Datengruppe, wie die Fundobjekte oder nur eine Fundkategorie, in Linked Data transformiert werden soll. In der Datenbank zu den Funden des Oberleiserbergs werden hier die freigegebenen Metallfunde von Ausgrabungen aus den Jahren 1996 bis 2009 herausgefiltert. Im Anschluss erfolgt eine Kontrolle der transformierten Daten.

8.2.1.2 Datenanreicherung mit URIs

In diesem Schritt liegt der Fokus auf der Zuordnung der URIs zu den vorhandenen Begriffen in den Daten. Dafür kann in der Excel-Datei jeweils eine zusätzliche Spalte für die URIs hinzugefügt werden. Die Auswahl der URIs sollte aus offen zugänglichen Online-Ressourcen erfolgen (siehe Kapitel 6). Der Aufwand dieses Prozesses kann je nach Diversität der Daten gering bis hoch sein, wobei die Kategorie der Ansprache wohl am längsten dauern kann. Die restlichen Kategorien (wie Fundort, Material, etc.) können in der Excel-Tabelle einfach kopiert werden. Im Fall der Daten des Oberleiserbergs wurden folgende URI-Ressourcen ausgewählt:

- *Ansprache*: Hierfür wurde das Vokabular des iDAI-Thesaurus iDAI.vocab[859] und die Getty Vocabularies[860] verwendet.

  Beispiele für die URIs für die Ansprache der Fundobjekte sind:

  - Nadel: <http://archwort.dainst.org/de/term/1910> (im iDAI.vocab als Gewandnadel gelistet, siehe Abbildung 42)
  - Ring: <http://archwort.dainst.org/de/term/7833> (im iDAI.vocab als Ring der Unterkategorie Schmuck gelistet)
  - Niet: <http://archwort.dainst.org/de/term/2491> (im iDAI.vocab als Niet gelistet)
  - Barren: <http://vocab.getty.edu/aat/300137865> (in den Getty Vocabularies als englischer Begriff „ingot" gelistet, welcher Metallbarren definiert)

Abbildung 42 – URI der Ansprache für Gewandnadel (Quelle: iDAI.vocab, unter: https://archwort.dainst.org/de/term/1910).

- *Material*: Zur Identifikation des Materials der Fundobjekte wurde das Vokabular von DBpedia[861] verwendet. Beispiele für URIs von Materialien sind:
  - Buntmetall: <https://dbpedia.org/page/Non-ferrous_metal> (in DBpedia gelistet als Nichteisenmetall)
  - Eisen: <https://dbpedia.org/page/Iron> (in DBpedia als Eisen eingetragen) Hier könnte auch die <URI https://dbpedia.org/page/Ferrous> gewählt werden (welche in DBpedia als eisenhältig beschrieben wird)
- *Erhaltungszustand*: Dieser wurde durch das iDAI.vocab Vokabular und das Wikidata Datenset festgehalten. Als Beispiele dienen hier:

[859] Siehe https://archwort.dainst.org/de/vocab/, abgerufen am 21.10.2021.
[860] Die Suchmaske für die Getty Vocabularies findet sich unter: http://vocab.getty.edu/, abgerufen am 21.10.2021.
[861] Um URIs aus den Daten von DBPedia herauszusuchen, dient die Adresse https://lookup.dbpedia.org/, abgerufen am 21.10.2021.

- Fragmentiert/Fragment: <http://archwort.dainst.org/de/term/1815> (im iDAI.vocab wird hier das Fragment angegeben, weitere Unterkategorien wären hier beispielsweise die Scherbe oder ein Abschlagfragment)
- Bruchstück: <http://archwort.dainst.org/de/term/1595> (das Bruchstück wird im Vokabular dem Fragment als verwandter Begriff zugeordnet)
- Erhalten: <https://www.wikidata.org/wiki/Q56557591> (auf Wikidata wird der Erhaltungszustand „erhalten" definiert)

- *Herstellungstechnologie*: Zur Definition der Technologie, die bestimmt wurde, konnte das Vokabular des iDAI.vocab verwendet werden, da es Begriffe beinhaltet, die in der Archäologie gebräuchlich sind. Beispiele von URIs sind:
  - Gegossen: <http://archwort.dainst.org/de/term/1869> (dieser Begriff ist eine Unterkategorie der Materialeigenschaften im iDAI.vocab)
  - Gehämmert: <http://archwort.dainst.org/de/term/1870> (auch dieser Begriff lässt sich den Materialeigenschaften zuordnen)
  - Getrieben: <http://archwort.dainst.org/de/term/1908> (diese URI ist alleinstehend im iDAI.vocab)
- *Datierung*: Die Zeitperioden wurden über PeriodO[862] definiert. Im PeridO-Client[863] kann das Datenset durchsucht werden nach der geeigneten Datierung, welche mit einem Datierungsrahmen, ortsbezogen und mit Quellenangaben eingegeben ist. Beispiele für die

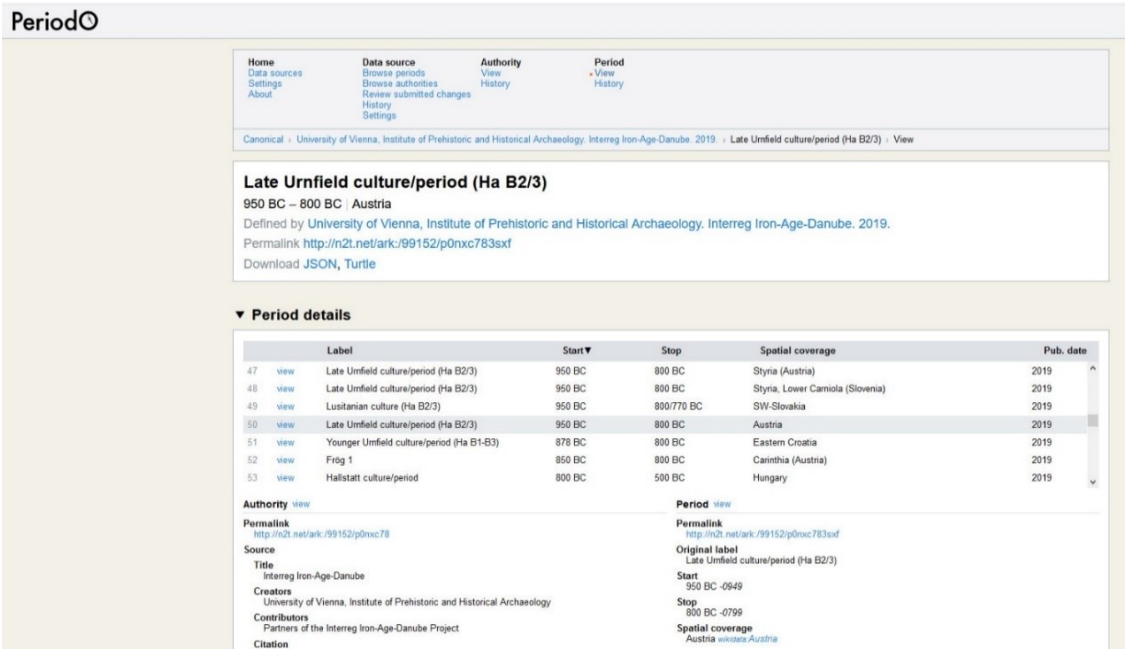


Abbildung 43 – Auszug der Webseite von PeriodO mit der Definition der Späten Urnenfelderzeit in Österreich (Quelle: PeriodO, unter: http://n2t.net/ark:/99152/p0nxc783sxf).

[862] Siehe die Projekthomepage unter: https://perio.do/en/, abgerufen am 21.10.2021.
[863] Siehe den PeriodO-Client unter: https://client.perio.do/?page=open-backend, abgerufen am 21.10.2021.

Bronzezeit in Österreich, welche durch das Projekt Interreg Iron-Age-Danube[864,865] eingepflegt worden sind, sind:

- Late Bronze Age (Späte Bronzezeit): <http://n2t.net/ark:/99152/p0nxc78x87x> (1300 BC – 800 BC)
- Early Urnfield culture/period (Bz D) (Frühe Urnenfelderzeit): <http://n2t.net/ark:/99152/p0nxc789d6p> (1300 BC - 1200 BC)
- Late Urnfield culture/period (Ha B2/3) (Späte Urnenfelderzeit): <http://n2t.net/ark:/99152/p0nxc783sxf> (950 BC - 800 BC) (siehe Abbildung 43)

- *Besitz*: Um festzuhalten, welche Organisation die Besitzrechte des Fundobjektes innehat, wird dies über URIs des VIAF (Virtual International Authority File)[866]definiert. Als Beispiel dient die Definition eines Instituts:
  - Universität Wien, Institut für Urgeschichte und Historische Archäologie: <http://viaf.org/viaf/245829558>
- *Fundstelle und Forschungsaktivität*: Der Fundort wurde mit GeoNames definiert. Die Fundstelle sowie die Ausgrabungsaktivitäten haben für die Ausgrabungen am Oberleiserberg bereits Identifikatoren, da sie durch das Iron-Age-Danube Projekt[867] aufgenommen worden sind.

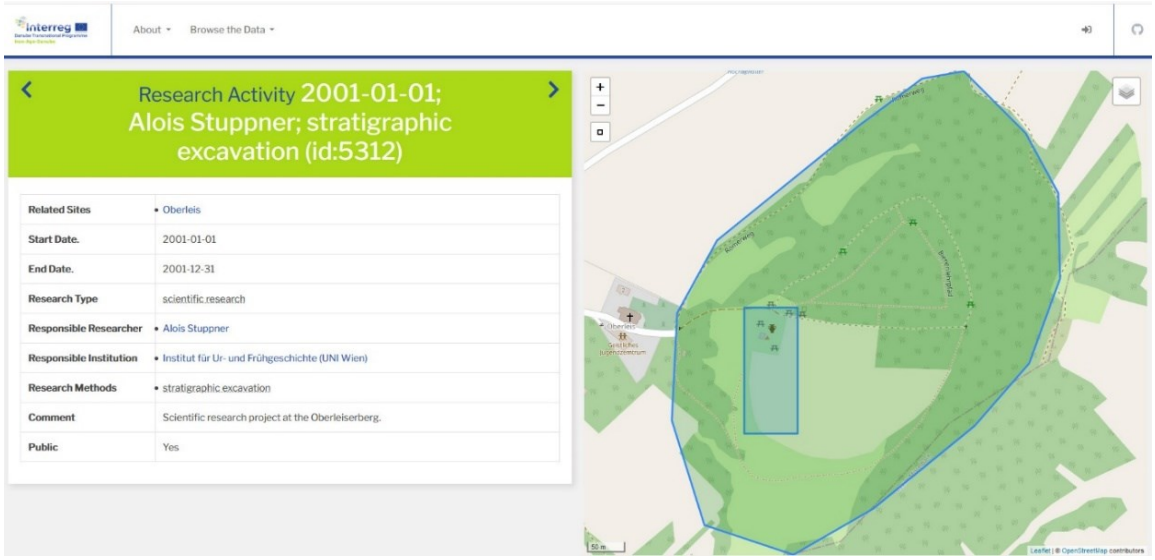


Abbildung 44 – Durch das Interreg-Projekt dokumentierte Forschungsaktivität am Oberleiserberg 2001 (Quelle: https://www.iron-age-danube.eu/archiv/researchevent/detail/5312).

[864] Siehe auch: http://www.interreg-danube.eu/approved-projects/iron-age-danube, abgerufen am 21.10.2021.
[865] Czajlik u. a. 2019.
[866] Siehe http://viaf.org/, abgerufen am 21.10.2021.
[867] Siehe unter: https://www.iron-age-danube.eu/, abgerufen am 21.10.2021.

- Oberleis, Österreich: <https://www.geonames.org/2769960/oberleis.html> (Österreich - Niederösterreich - Korneuburg - Ernstbrunn, Koordinaten: N 48°34'00" E 16°22'00")
- Ausgrabung am Oberleiserberg 2001 durch Alois Stuppner (Institut für Urgeschichte und Historische Archäologie): <https://www.iron-age-danube.eu/archiv/researchevent/detail/5312> (siehe Abbildung 44)

- *Bibliographische Angaben*: Um anzugeben, welche literarischen Quellen für die Datierung der Fundobjekte verwendet wurden, kann in der Datenbank das Zitat angeführt werden. In den Daten zum Oberleiserberg wurden diese Zitate mit der iDAI.bibliography zenon[868] Online-Ressource definiert.
  - Das Zitat „Ríhovský, Jiří. Die Nadeln in Mähren und im Ostalpengebiet (von der mittleren Bronzezeit bis zur älteren Eisenzeit). München: Beck'sche Verlagsbuchhandlung, 1979." kann mit der URI <https://zenon.dainst.org/Record/000896253> definiert werden.

8.2.1.3 Datenmodell mit CIDOC CRM

Nachdem die Daten bereinigt, harmonisiert und mit URIs angereichert wurden, kann das Datenmodell mit CIDOC CRM erstellt werden. Die Definitionen des CIDOC CRM werden in Form von PDF- und Word-Dateien auf der Webseite des CIDOC CRM veröffentlicht.[869] Darin werden die Prinzipien der Modellierung und die Grundkonzepte erklärt. Weiters wird die Hierarchie für alle Klassen und Eigenschaften angegeben. Den Hauptteil bilden die Definitionen für jede einzelne Klasse und Eigenschaft. Es wird definiert, wie jede Klasse und Eigenschaft zu verstehen ist und mit Beispielen illustriert.[870]

Um die Daten zu modellieren, wird für jede Spalte, welche in der Excel-Datei angelegt wurde, eine Klasse in der CIDOC CRM-Ontologie gewählt (siehe Tabelle in Abbildung 45). Es wurde CIDOC CRM in der Version 7.1.1[871], die Erweiterung CRMarchaeo in der Version 1.4.7[872] und die Erweiterung CRMsci der Version 1.2.8[873] verwendet.

[868] Datenbank für Literatur des Deutschen Archäologischen Instituts, zu finden unter: https://zenon.dainst.org/, abgerufen am 21.10.2021.
[869] Siehe unter http://www.cidoc-crm.org/versions-of-the-cidoc-crm, abgerufen am 21.10.2021.
[870] Bekiari u. a. 2021.
[871] Bekiari u. a. 2021.
[872] Doerr u. a. 2018a.
[873] Doerr u. a. 2020b.

Spaltenname in der Exceldatei	*Klasse in CIDOC CRM*
ID_Access_Datenbank	*E42 Identifier*
Fundnummer	*E42 Identifier*
Ansprache	*E55 Type*
Ansprache_uri	*E42 Identifier*
Ansprache_Unsicherheit	*E62 String*
Bibliographische_Angabe	*E31 Document*
Bibliographische_Angabe_uri	*E42 Identifier*
Material	*E57 Material*
Material_uri	*E42 Identifier*
Erhaltungszustand	*E3 Condition State*
Erhaltungszustand_uri	*E42 Identifier*
Maße_Laenge	*E54 Dimension*
Maße_Typ	*E54 Dimension*
Maße_Einheit	*E54 Dimension*
Maße_literal	*E62 String*
Beschreibung_literal	*E62 String*
Verzierung	*E23 Human-Made Feature*
Verzierung_uri	*E42 Identifier*
Technologie_Herstellung	*E55 Type*
Technologie_Herstellung_uri	*E42 Identifier*
Kulturelle_Zuordnung_von	*E4 Period*
Kulturelle_Zuordnung_bis	*E4 Period*
Kulturelle_Zuordnung_von_uri	*E42 Identifier*
Kulturelle_Zuordnung_bis_uri	*E42 Identifier*
Datierung_von	*E4 Period*
Datierung_bis	*E4 Period*
Datierung_von_uri	*E42 Identifier*
Datierung_bis_uri	*E42 Identifier*
Datierung_absolute_Zahlen_von	*E61 Time Primitive*
Datierung_absolute_Zahlen_bis	*E61 Time Primitive*
Datierung_Unsicherheit	*E62 String*
Fundjahr	*E61 Time Primitive*
Forschungsaktivität_uri	*E42 Identifier*
Schnitt-Nummer_Ausgrabungsobjekt	*E26 Physical Feature*

x-Achse	*E59 Primitive Value*
x-Achse_Einheit	*E54 Dimension*
x-Achse_Typ	*E55 Type*
y-Achse	*E59 Primitive Value*
y-Achse_Einheit	*E54 Dimension*
y-Achse_Typ	*E55 Type*
Tiefe_Niveau_literal	*E62 String*
Tiefe_Niveau_Einheit	*E54 Dimension*
Tiefe_Typ	*E55 Type*
Befund	*A2 Stratigraphic Volume Unit*
Aufbewahrungsort	*E53 Place*
Aufbewahrungsort_uri	*E42 Identifier*
Eigentum	*E39 Actor*
Eigentum_uri	*E42 Identifier*
Dateiname_Abbildung	*E31 Document*
Fundort	*E27 Site*
Fundort_uri	*E42 Identifier*
Archaeologische_Maßnahme_Typ	*A9 Archaeological Excavation*
Ausgrabungstechnik	*E29 Design or Procedure*
Kommentar_Freitextfeld_literal	*E62 String*

Abbildung 45 – Zuweisung der Spalten der Excel-Datei zu den Klassen des CIDOC CRM.

Nun können die einzelnen Elemente in einen Zusammenhang gebracht werden. Dafür müssen zu den Klassen die Eigenschaften gewählt werden, um sie zueinander zu verbinden. Für diesen Schritt haben sich Programme für die Erstellung von Mindmaps als hilfreich erwiesen.[874] Die Modellierung sollte von dem zentralen Element der Daten bzw. dieser Tabelle ausgehen. Dies wäre in diesem Fall das Fundobjekt. In Abbildung 46 ist das Ergebnis der wichtigsten Elemente dieser Modellierung zu sehen.

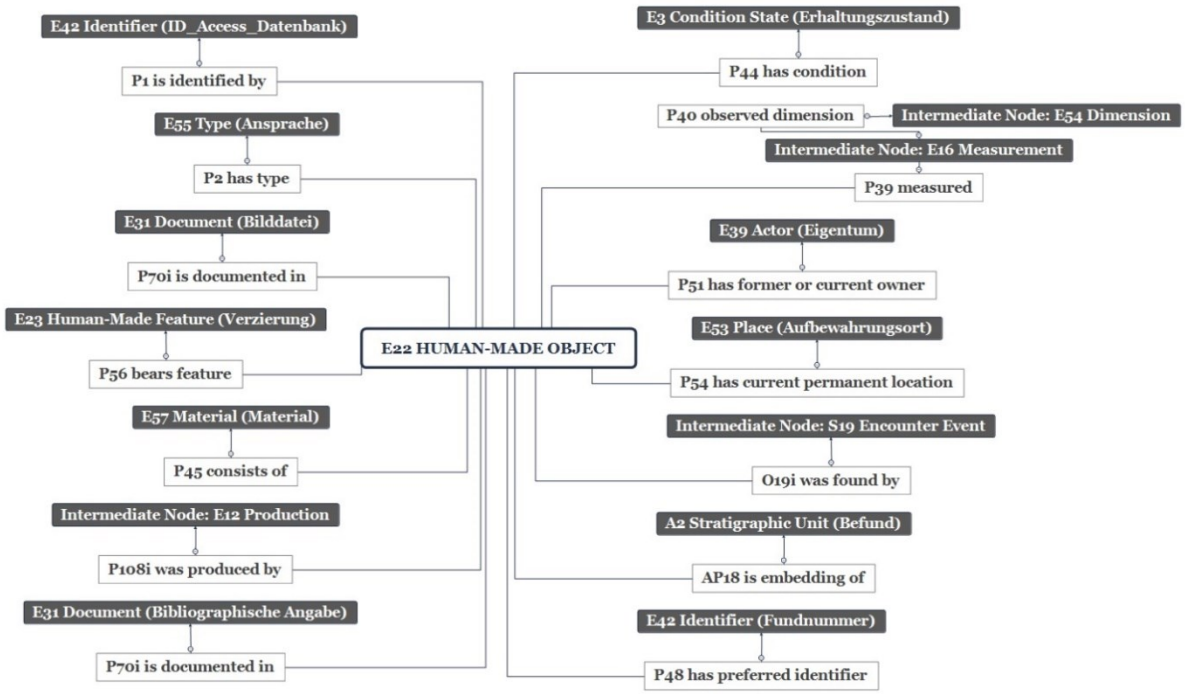


Abbildung 46 - Datenmodell der Funddatenbank des Oberleiserbergs mit CIDOC CRM (Grafik: M. Simon).

8.2.1.3.1 Detaillierte Beschreibung der Modellierung

Im Folgenden sollen die Überlegungen zu den einzelnen Elementen dargelegt werden. Als Zentrum der Modellierung steht das **Fundobjekt** mit *E22 Human-Made Object*. Dem Fundobjekt wird in der Excel-Datei eine URI zugewiesen, welche eindeutig identifizierbar ist und auf das Repositorium hinweisen soll (z. B. <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019> für das Fundobjekt mit der Fundnummer 13019). Dies ist notwendig für die Konvertierung der Daten in ein RDF-Format und wird in Kapitel 8.2.2 näher erläutert, sei an dieser Stelle jedoch für das bessere Verständnis erwähnt. Zu diesem Fundobjekt werden nun alle Informationen in der Datenbank verbunden. Die Verlinkungen zu den URIs der einzelnen Spalten (wie beispielsweise die URI zu der Ansprache) werden in den Abbildungen aus Platzgründen nicht mit dargestellt.

[874] Als Beispiel eines solchen Programms könnte XMind genannt werden.

Das Fundobjekt soll **Identifikatoren** zugewiesen bekommen. In diesem Fall sind dies zwei verschiedene Nummern, da die Daten ursprünglich aus einer Access-Datenbank stammen und so sowohl eine ID dieser Datenbank als auch die bei der Ausgrabung zugewiesene Fundnummer besitzen. Daher wurde die Fundnummer als bevorzugter Identifikator mit der Verbindung *P48 has preferred identifier* und die ID der Access-Datenbank mit *P1 is identified by* verlinkt. Der Fund wird dadurch primär über die Fundnummer identifiziert, die Information aus der Access-Datenbank ist jedoch nicht verloren gegangen und kann so übernommen werden. Die URI stellt einen zusätzlichen Identifikator dar.

Die **Ansprache** des Fundobjekts kann in der Modellierung als *E55 Type* dargestellt werden. Die Verbindung vom Fundobjekt wird mit der Eigenschaft *P2 has Type* vorgenommen. Um **Bilddateien** zum Fundobjekt zu verlinken, kann die Klasse *E31 Document*, welches die Bilddatei darstellt, verwendet werden. Die Dateinamen der Bilddateien sind in der Datenbank hinterlegt (z. B. 13482.tif). Das Fundobjekt wird mit der passiven Variante, welche viele Eigenschaften besitzen, zu der Bilddatei modelliert. Die Eigenschaft *P70 documents* kann auch umgekehrt verwendet werden, dann lautet sie *P70i is documented* in. Das „i" steht dabei für „inverted", also umgekehrt. So wird ausgesagt, dass das Fundobjekt *E22 Human-Made Object* in dieser Bilddatei dokumentiert wird.

Verzierungen können mit der Klasse *E23 Human-Made Feature* beschrieben werden, da sie intentionell vom Menschen kreiert wurden. Die Verlinkung vom *E22 Human-Made Object* kann durch die Eigenschaft *P56 bears feature* durchgeführt werden. Um das **Material** eines Fundobjekts anzugeben, kann es mit der Klasse *E57 Material* modelliert werden. Es wird verbunden mit der Eigenschaft *P45 consists of*.

Die Verlinkung zur **Produktion** wurde als „Intermediate Node" kreiert. Dies ist eine Art Hilfsknoten, welcher in diesem Fall genutzt werden kann, um ein Ereignis zu kreieren. Dieses Ereignis kann in weiterer Folge zu der Datierung des Fundobjekts verlinkt werden, da das CIDOC CRM als ereignis-zentrierte Ontologie Zeitperioden nur zu Ereignissen verlinkt.

Des Weiteren kann die Art der **Herstellungstechnologie**, wie beispielsweise „gegossen" oder „getrieben", durch die Eigenschaft *P2 has type* und die Klasse *E55 Type* verlinkt werden. Die Modellierung der Herstellungstechnologie und der Datierung wird in Abbildung 47 dargestellt. Die *E12 Production* wird mit der Eigenschaft *P108i was produced by* zum Fundobjekt verbunden.

Die **Datierung** wird in den Daten mit drei Kategorien angegeben: absolute Zahlen (mit jeweils einer Spalte „von“ und „bis“), Datierung (z. B. Späte Bronzezeit, jeweils eine Spalte „von“ und „bis“) und kulturelle Zuordnung (wie beispielsweise Urnenfelderzeit, jeweils eine Spalte „von“ und „bis“). Die Modellierung der Kategorien „Datierung“ und „kulturelle Zuordnung“ kann jeweils mit den Eigenschaften *P176i starts after the start of* für die frühere Datierung und *P185 ends before the end of* für den späteren Zeitpunkt der Datierung angegeben werden. Die Verbindung von der Produktion zu den **absoluten Zahlen** führt über die „Intermediate Node“ *E52 Time-Span* mit der Eigenschaft *P4 has time-span*. Als Eigenschaften dienen *P79 beginning is qualified by* für den früheren Zeitpunkt („von“) und *P80 end is qualified by* für den späteren Zeitpunkt („bis“) der Datierung. Die absoluten Zahlen selbst werden mit der Klasse *E61 Time Primitive* angegeben (z. B. -900). Die Kategorie der Datierungsangaben, die in der Spalte „Datierung von“ und „Datierung bis“ eingetragen ist, enthält Angaben wie „Bronzezeit“ oder „Späte Bronzezeit“.

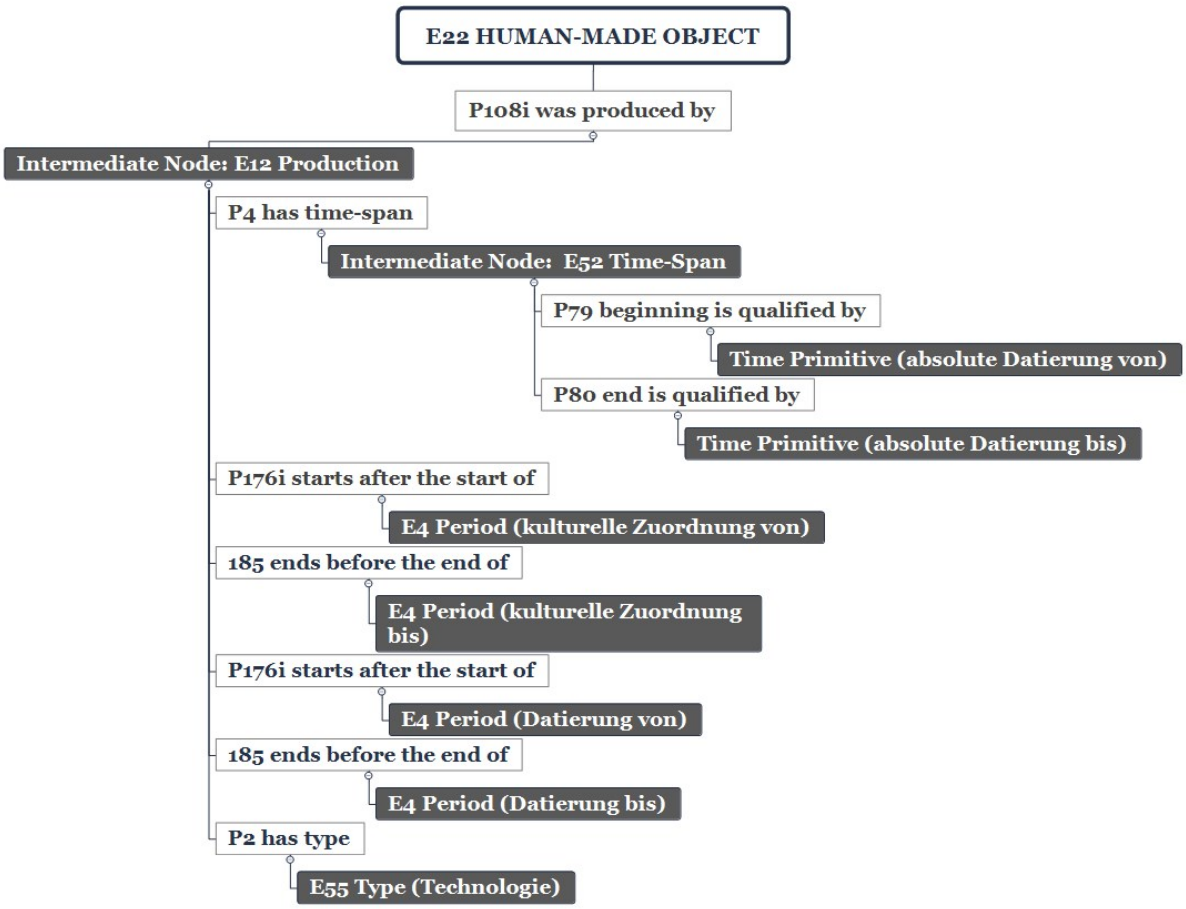


Abbildung 47 - Modellierung der Zeitperioden über das Ereignis der Produktion (Grafik: M. Simon).

In den Daten finden sich auch Informationen über die bei der Bearbeitung der Fundobjekte verwendete Literatur. Diese **bibliographischen Angaben** können als *E31 Document* verwendet werden. Darin befinden sich Informationen zum Typen des Fundobjekts und daher kann es mit *P70i*

is documented in verlinkt werden. Der **Erhaltungszustand** eines Objektes, wie beispielsweise „fragmentiert“ oder „Bruchstück“, kann mit der Klasse *E3 Condition State* angegeben werden. Die Klasse dokumentiert den Zustand des Fundobjektes zum Zeitpunkt der Bearbeitung. Die passende Eigenschaft *P44 has condition* kann das Fundobjekt zum Erhaltungszustand verbinden.

Die **Maßangaben** von Fundobjekten sind aufwändiger zu modellieren. Sie werden in Abbildung 48 detailliert dargestellt. Auch hier muss ein Ereignis durch ein „Intermediate Node“ kreiert werden. Es wird das Messereignis durch die Klasse *E16 Measurement* generiert. In diesem Fall geht die Verbindung zum Fundobjekt vom Knoten *E16 Measurement* aus. Das Messereignis *E16 Measurement* misst (*P39 measured*) das Fundobjekt. In weiterer Folge kann das Messereignis *E16 Measurement* durch die Eigenschaft *P40 observed dimension* zu dem gemessenen Ergebnis verlinkt werden. Das Ergebnis muss ebenfalls als „Intermediate Node“, als *E54 Dimension*, konstruiert werden, da an diesen Knoten die weiteren Punkte zum Messergebnis, wie **Länge**, **Einheit** und **Maße**, hinzugefügt werden sollen. Die Einheit, wie beispielsweise „cm“, kann von der Klasse *E54 Dimension* mit *P91 has unit* zur Einheit mit der Klasse *E58 Measurement Unit* verbunden werden. Die eigentlichen Maßangaben, wie z. B. „1,3“, können als *E59 Primitive Value* modelliert werden. Sie werden über die Eigenschaft *P90 has value* mit der *E54 Dimension* verbunden. Um anzugeben, um welche Art der Maßangabe es sich handelt, wird der Knoten der *E54 Dimension* über die Eigenschaft *P2 has type* zum *E55 Type* verbunden. In der Excel-Datei ist in jeder Zelle dieser Spalte der gleiche Typ, wie beispielsweise „Länge“, einzugeben.

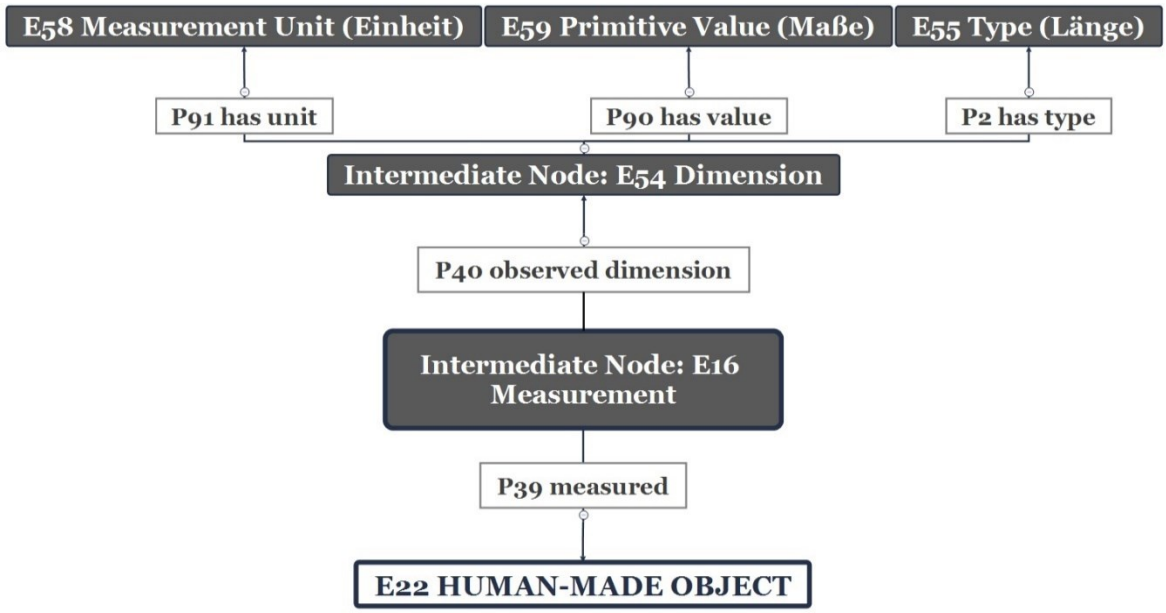


Abbildung 48 - Maßangaben zum Fundobjekt (Grafik: M. Simon).

Um anzugeben, in wessen **Eigentum** sich das Fundobjekt befindet, kann das Fundobjekt mit der Eigenschaft *P51 has former or current owner* zu der Klasse *E39 Actor* verbinden. Diese Eigenschaft

besagt, dass das Eigentum zu einem früheren Zeitpunkt oder zum Zeitpunkt der Modellierung bei dem angegebenen *E39 Actor* liegt. Hier könnte auch die Eigenschaft *P52 has current owner* gewählt werden, die das Fundobjekt zum aktuellen Zeitpunkt als Besitz des *E39 Actors* verbindet.

Der **Aufbewahrungsort** kann als Klasse *E53 Place* angegeben werden. Das Fundobjekt wird mit der Eigenschaft *P54 has current permanent location* verlinkt, um anzugeben, dass dies der zugewiesene Platz für das Fundobjekt ist, es sich jedoch auch an einem anderen Ort, wie beispielsweise einer Ausstellung, befinden kann. Zur Angabe des Aufbewahrungsortes könnten auch die Eigenschaften *P53 has former or current location*, welche besagt, dass das Fundobjekt in früherer Zeit oder zum jetzigen Zeitpunkt diesen Aufbewahrungsort hat bzw. hatte, oder *P55 has current location*, welche eine Spezifizierung der Eigenschaft *P53 has former or current location* darstellt, verwendet werden.

Die Informationen rund um die **Ausgrabung** (siehe Abbildung 49) können so modelliert werden, dass das Fundobjekt *E22 Human-Made Object* bei einem Auffindungsevent durch den Hilfsknoten *S19 Encounter Event* gefunden wurde (*O19i object was found by*). Die archäologische Ausgrabung kann als Klasse *A9 Archaeological Excavation* modelliert werden. Sie besteht aus (*P9 consists of*) den einzelnen Auffindungsevents (*S19 Encounter Event*), welche auf der Ausgrabung stattfinden. Die Informationen zu der Ausgrabung, wie das Jahr, der Fundort und die **Ausgrabungsmethode** werden von der Klasse *A9 Archaeological Excavation* verbunden. Die Ausgrabungsmethode, wie beispielsweise „Planumsgrabung“, wird mit der Eigenschaft *P32 used general technique* zu der Klasse *E29 Design or Procedure* modelliert. Der **Fundort** mit der Klasse *E27 Site* wird über die Eigenschaft *P7*

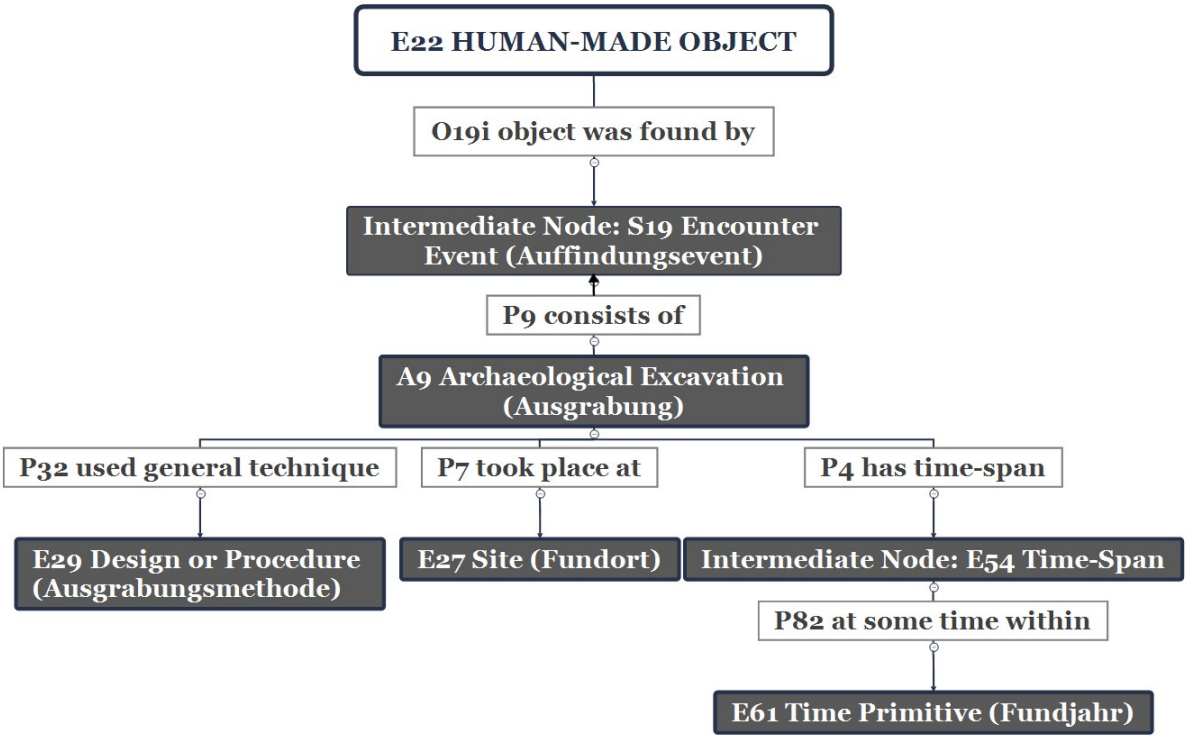


Abbildung 49 - Modellierung der archäologischen Ausgrabung (Grafik: M. Simon).

took place at verlinkt. Das **Fundjahr** kann mit einem Hilfsknoten einer Zeitspanne *E52 Time-Span* verbunden werden. Die Ausgrabung hat eine Zeitspanne (*P4 has time-span*), welche an einem Zeitpunkt innerhalb der angegebenen Zeit (*P82 at some time within*) angesetzt wird. Das Jahr wird mit der Klasse *E61 Time Primitive* angegeben. Diese Klasse kann für Zeitangaben verwendet werden, die in weiterer Folge weiterverarbeitet werden sollen, wie beispielsweise in Zeitdarstellungen.

Die Fundobjekte wurden bei den Ausgrabungen am Oberleiserberg mittels Schnitts, **x- und y-Achse**, **Tiefenniveau** und wenn gegeben, mit der Zugehörigkeit zu einem **Befund**, wie einer Grube, dokumentiert. In der Modellierung (siehe Abbildung 50) wird angegeben, dass das Fundobjekt in einem „Intermediate Node" *A7 Embedding* eingebettet war mittels der Eigenschaft *AP18 is embedding of*. Diese Klasse *A7 Embedding* kann nun weiter konkretisiert werden. Es wurde gefunden bei einem Auffindungsevent während der Ausgrabung, welches als Klasse „Intermediate Node" *S19 Encounter Event* mit der Eigenschaft *AP17 is found by* verlinkt wird. Um anzugeben, in welchem Befund das Fundobjekt gefunden wurde, bzw. in welchem Befund das Fundobjekt eingebettet war, kann das *A7 Embedding* des Fundobjektes mit der Eigenschaft *AP19 ist embedding in* mit dem Befund der Klasse *A2 Stratigraphic Unit* verbunden werden. Die Eigenschaft *AP19 is embedding in* beschreibt die *A2 Stratigraphic Unit*, in welcher das Fundobjekt eingebettet ist.

Der Schnitt, in welchem das Fundobjekt gefunden wurde, kann ebenfalls durch die Eigenschaft *AP19 is embedding* in angegeben werden. Der Schnitt wird dabei als Klasse *E26 Physical Feature*

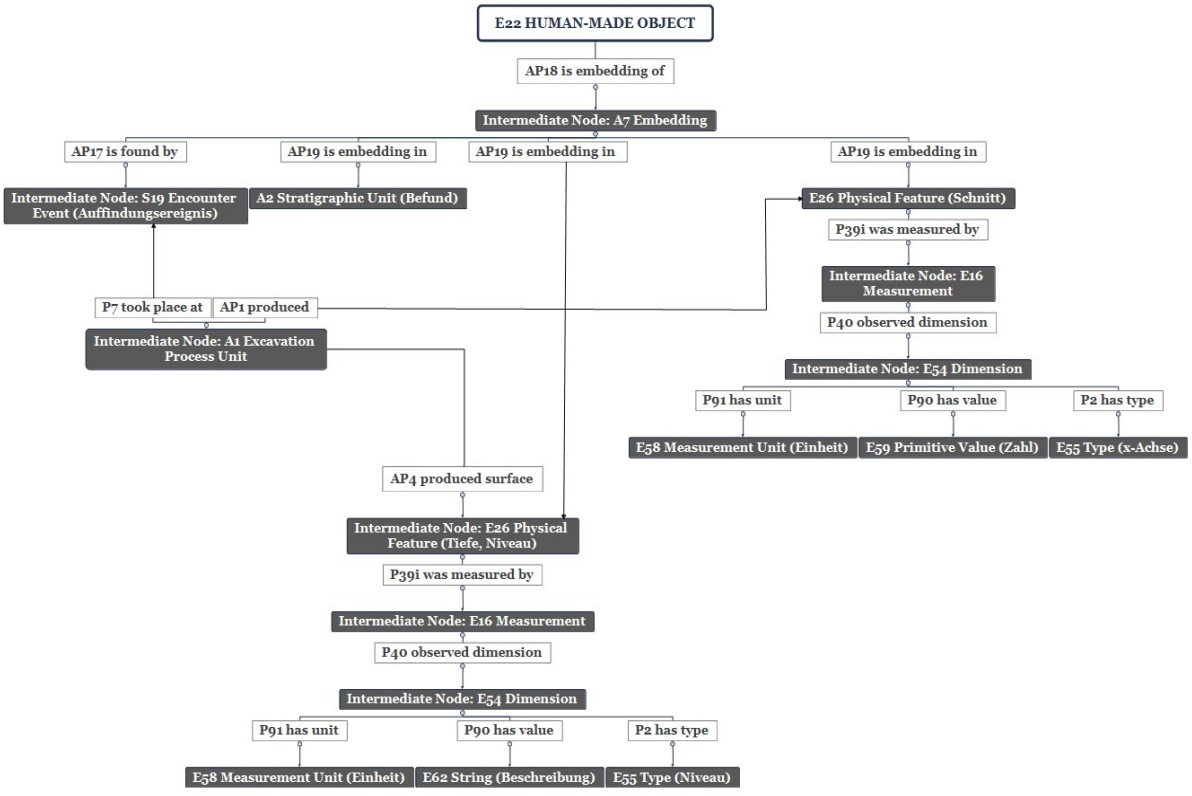


Abbildung 50 - Modellierung des Befundes, der x-Achse und des Tiefenniveaus (Grafik: M. Simon).

modelliert. In Abbildung 50 werden exemplarisch die Informationen zur x-Achse modelliert. Sie können ebenfalls auf die y-Achse übertragen werden. Die Vorgehensweise dabei ist ähnlich wie bei der Modellierung der Maßangaben. Es wird ein „Intermediate Node" *E16 Measurement* kreiert, um das Ereignis der Messung darzustellen. Das *E16 Measurement* kann durch die Eigenschaft *P40 observed dimension* zum nächsten Hilfsknoten, *E54 Dimension*, verlinkt werden. Zu dieser Klasse können die Informationen zur x-Achse verbunden werden, die Einheit, wie beispielsweise „m" (*E58 Measurement Unit*), der Wert bzw. die Position auf der x-Achse (z. B. 6,3) als *E59 Primitive Value* und der Typ der *E54 Dimension*, die „x-Achse". In CRMarchaeo können die Arbeitseinheiten der Ausgrabung als *A1 Excavation Process Unit* modelliert werden. So können Angaben zu den Arbeitsprozessen dokumentiert werden. In diesem Modell wird ein Arbeitsprozess mit einem Hilfsknoten mit der Eigenschaft *P7 took place at* zum Auffindungsereignis *S19 Encounter Event* verlinkt. Ein Arbeitsprozess, welcher bei der Ausgrabung stattgefunden hat, produzierte sowohl den Schnitt (*E26 Physical Feature*) als auch die Oberfläche des Niveaus (*E26 Physical Feature*). Die Klasse *A1 Excavation Process Unit* wird mit der Eigenschaft *AP1 produced* zum Schnitt und mit der Eigenschaft *AP4 produced surface* zum Niveau verbunden, da es sich hierbei um eine Oberfläche handelt. Die Informationen zum Niveau werden ebenfalls, wie die Maßangaben und die Informationen zum Schnitt, mittels Hilfsknoten *E16 Measurement* und *E54 Dimension* modelliert.

8.2.2 Konvertierung des Datenmodells in ein RDF-Format

Wenn die Datenmodellierung mit CIDOC CRM abgeschlossen ist, ist es ratsam, sich mit einem oder einer Expert*in der Materie auszutauschen und zu kontrollieren, ob die Modellierung korrekt ist. Wenn dies der Fall ist, können die Daten in ein RDF-Format gebracht werden. Im Zuge des Arbeitsprozesses zeigte sich, dass dies der späteste Zeitpunkt ist, um sich mit einem oder einer Semantic Web-Expert*in auszutauschen bzw. zusammenzuarbeiten, um die Konvertierung korrekt ausführen zu können, da jedes Datenset einzigartig ist und Fehlerquellen bei einem ersten Versuch so vermieden werden können. Für die Konvertierung sind mehrere Schritte notwendig, welche im Folgenden genauer beschrieben werden. Der Arbeitsprozess richtete sich nach Ergebnissen aus dem Projekt „Open Research Data for Prehistoric Mining Archaeology"[875,876] der Universität Innsbruck unter der Leitung von Gerald Hiebel[877].

[875] Zum Projekt siehe https://www.uibk.ac.at/projects/ord4mining-archaeo/, abgerufen am 21.10.2021.
[876] Siehe zum Projekt u. a. Hiebel u. a. 2021a; Hiebel u. a. 2018.
[877] Vielen herzlichen Dank an dieser Stelle an Dr. Gerald Hiebel für die Informationen über den Workflow und die Arbeitsprozesse zur RDF-Konvertierung und Modellierung und die persönliche Hilfestellung bei der Umsetzung.

8.2.2.1 Einbauen der Intermediate Nodes

Wie im vorangehenden Kapitel 8.2.1.3 beschrieben, werden für die Modellierung an manchen Stellen Hilfsknoten verwendet. Diese finden sich in den Daten aber noch nicht als eigene Spalte, da sie keinen Inhalt besitzen. Deshalb muss für jeden Hilfsknoten („Intermediate Node") eine Spalte in der Excel-Datei erstellt werden, um sie ins Datenmodell einbauen zu können. In den Spalten werden URIs für die Hilfsknoten generiert, welche eindeutig die Institution und das Projekt erkennen lassen sollten. Die URI für das Fundobjekt selbst, welches in der Modellierung als *E22 Human-Made Object* angegeben wurde, zählt ebenso zu diesen Hilfsknoten. Die Tabelle in Abbildung 51 zeigt Beispiele, wie diese generierten URIs für die Fundnummer 13019 aussehen könnten. Als Basis für die URIs wurde < http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/> gewählt, dann eine Kurzbezeichnung für den jeweiligen Hilfsknoten hinzugefügt (wie beispielsweise „E22-object/") und die Fundnummer des jeweiligen Fundes angehängt.

Hilfsknoten	*URI*
E22 Human-Made Object	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019
E54 Dimension (Maße)	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Abmessung/13019
E12 Production	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019
E4 Period	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E4/13019
E52 Time-Span (Datierung)	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E52-Datierung/13019
E52 Time-Span (Ausgrabung)	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E52-Ausgrabung/13019
A7 Embedding	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/A7/13019
A1 Excavation Process Unit	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/A1/13019
S19 Encounter Event	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/S19/13019
E53 Place (für x-y-Achsen)	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E16/13019

E54 Dimension (für x-y-Achsen)	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Achse/13019

Abbildung 51 – Generierung von URIs für die Hilfsknoten der Modellierung.

In der Excel-Datei wird für jede Fundnummer und jeden Hilfsknoten eine URI in dieser Zeile generiert. Auf diese Weise erhält jedes Fundobjekt eine eigene URI für jeden Hilfsknoten. In Abbildung 52 wird exemplarisch die Liste in der Excel-Datei der URIs für den Hilfskoten *E22 Human-Made Object* dargestellt. Die URI des Hilfsknotens der Fundnummer 13019 ist demnach <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019> und die URI des Hilfskotens der Fundnummer 13007 ist <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13007>.

	B	BB
1	Fundnummer	E22_Human-made_object_uri
44	12718	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/12718
45	12797	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/12797
46	12827	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/12827
47	12897	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/12897
48	13007	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13007
49	13019	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019
50	13044	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13044
51	13074	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13074
52	13107	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13107

Abbildung 52 - Generierung von URIs für die Hilfsknoten der Modellierung in der Excel-Datei.

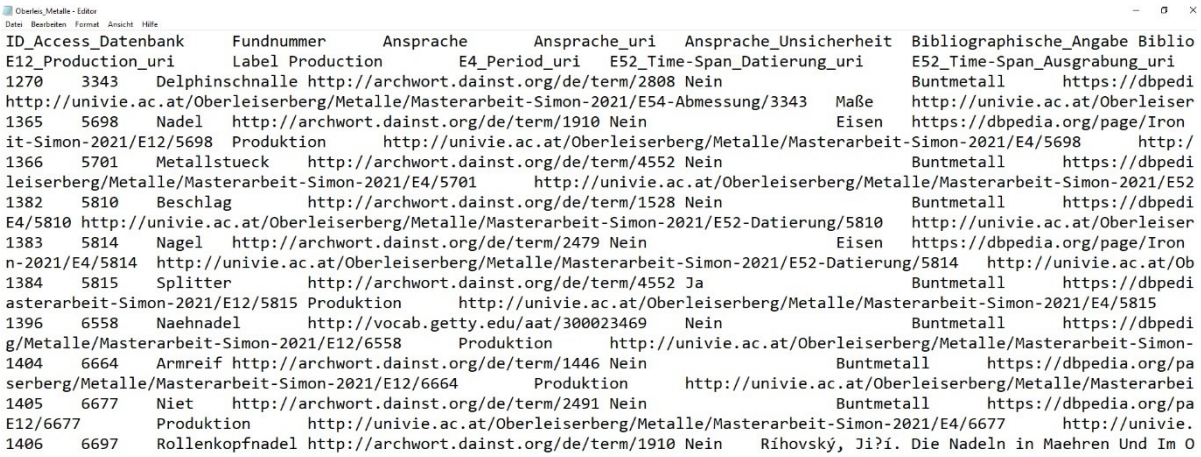
8.2.2.2 Namen für die Hilfsknoten vergeben

Um die Visualisierung in einem späteren Schritt zu verbessern, können zusätzlich Spalten eingefügt werden, welche die jeweilige URI mit einem Namen versehen (siehe Abbildung 53). Dieser Zwischenschritt ist optional. Diese können in der Modellierung als „Label" der jeweiligen URIs hinzugefügt werden. Im Fall der zu bearbeitenden Daten waren dies:

Hilfsknoten	*Label*
E54 Dimension	Maße
E12 Production	Produktion
S19 Encounter Event	Auffindungsevent
E16 Measurement	Messung
E52 Time-Span	Zeitspanne

Abbildung 53 – Namen der Hilfsknoten.

8.2.2.3 Anwendung der Modellierung


Oberleis_Metalle - Editor
Datei Bearbeiten Format Ansicht Hilfe

```
ID_Access_Datenbank     Fundnummer      Ansprache       Ansprache_uri   Ansprache_Unsicherheit  Bibliographische_Angabe Biblio
E12_Production_uri      Label Production        E4_Period_uri   E52_Time-Span_Datierung_uri     E52_Time-Span_Ausgrabung_uri
1270    3343    Delphinschnalle http://archwort.dainst.org/de/term/2808 Nein                    Buntmetall      https://dbpedi
http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Abmessung/3343   Maße    http://univie.ac.at/Oberleiser
1365    5698    Nadel   http://archwort.dainst.org/de/term/1910 Nein                    Eisen   https://dbpedia.org/page/Iron
it-Simon-2021/E12/5698  Produktion      http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E4/5698      http:/
1366    5701    Metallstueck    http://archwort.dainst.org/de/term/4552 Nein                    Buntmetall      https://dbpedi
leiserberg/Metalle/Masterarbeit-Simon-2021/E4/5701      http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E52
1382    5810    Beschlag        http://archwort.dainst.org/de/term/1528 Nein                    Buntmetall      https://dbpedi
E4/5810 http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E52-Datierung/5810   http://univie.ac.at/Oberleiser
1383    5814    Nagel   http://archwort.dainst.org/de/term/2479 Nein                    Eisen   https://dbpedia.org/page/Iron
n-2021/E4/5814  http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E52-Datierung/5814   http://univie.ac.at/Ob
1384    5815    Splitter        http://archwort.dainst.org/de/term/4552 Ja                      Buntmetall      https://dbpedi
asterarbeit-Simon-2021/E12/5815 Produktion      http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E4/5815
1396    6558    Naehnadel       http://vocab.getty.edu/aat/300023469    Nein                    Buntmetall      https://dbpedi
g/Metalle/Masterarbeit-Simon-2021/E12/6558      Produktion      http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-
1404    6664    Armreif http://archwort.dainst.org/de/term/1446 Nein                    Buntmetall      https://dbpedia.org/pa
serberg/Metalle/Masterarbeit-Simon-2021/E12/6664        Produktion      http://univie.ac.at/Oberleiserberg/Metalle/Masterarbei
1405    6677    Niet    http://archwort.dainst.org/de/term/2491 Nein                    Buntmetall      https://dbpedia.org/pa
E12/6677        Produktion      http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E4/6677      http://univie.
1406    6697    Rollenkopfnadel http://archwort.dainst.org/de/term/1910 Nein    Říhovský, Ji?í. Die Nadeln in Maehren Und Im O
```

Abbildung 54 - Daten in einer UTF-8 Textdatei mit Tabulatortrennung.

In einem nächsten Schritt werden die Daten mithilfe des Online-Modellierungswerkzeugs Karma[878,879] modelliert. Hierfür werden die Daten aus der bearbeiteten Excel-Datei in Form einer Text-Datei benötigt, welche die Codierung UTF-8 besitzt und mit Trennzeichen („Tabstop“) getrennt ist. Dieses Dateiformat kann durch Exportieren der Excel-Datei durch den Dateityp „Text (Tabstopp-getrennt) Textformat mit Tabulatortrennung“ mit der Einstellung „Code: UTF-8“ generiert werden. Ein Ausschnitt, wie diese Datei in etwa aussehen kann, wird in Abbildung 54 gezeigt.

Diese Datei soll im Anschluss in Karma geladen werden. Wenn Karma gestartet wurde, wird zunächst ein neues Repositorium angelegt. Hierfür kann in der Menüleiste der Punkt „OpenRDF“ ausgewählt werden. Danach öffnet sich die „OpenRDF-Workbench“, in welcher unter dem Menüpunkt „Repositories“ ein neues Repositorium angelegt werden kann. Im Anschluss wird die ID und der Titel des Repositoriums angegeben. Dieser darf keine Umlaute, Sonderzeichen oder Bindestriche enthalten (siehe Abbildung 55). Mit „Create“ wird das Repositorium angelegt und das Fenster kann wieder geschlossen werden.

[878] Siehe unter https://usc-isi-i2.github.io/karma/, abgerufen am 21.10.2021.
[879] Installationsanweisungen zu Karma finden sich unter https://github.com/usc-isi-i2/Web-Karma/wiki/Installation%3A-One-Click-Install, abgerufen am 21.10.2021.

Nun soll der Daten-Endpunkt (Data Endpoint) festgelegt werden. Dies kann in Karma im unteren rechten Bildschirmrand eingegeben werden, wenn auf die Adresse nach „Data Endpoint:“ geklickt wird. Der letzte Teil der URL „/karma_data“ wird mit dem Titel des Repositoriums, welcher zuvor eingegeben wurde, ausgetauscht. In diesem Fall würde dies „/Oberleiserberg_ID“ sein (siehe Abbildung 55).

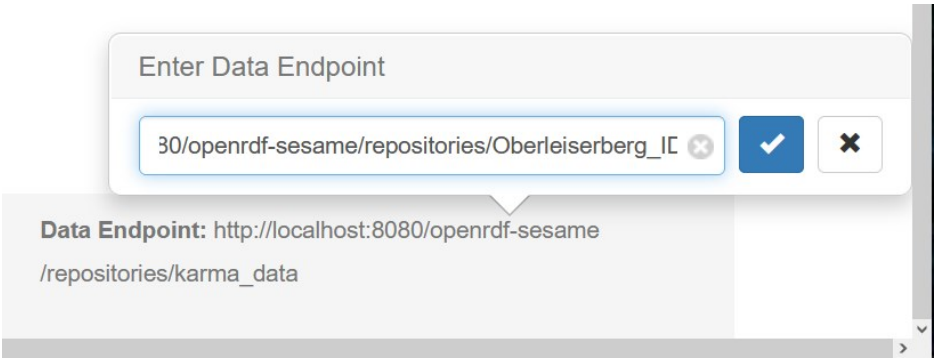


Abbildung 55 – Eingabe des Daten-Endpunkts in Karma.

Um in Karma Daten nach einer Ontologie zu modellieren, müssen diese in Form von Daten hinterlegt werden.[880] Dies kann in der unteren Leiste unter „Karma Home“ durchgeführt werden. Der Pfad, welcher an dieser Stelle steht, sollte die entsprechenden Dateien in dem Karma-Ordner „preloaded-ontologies“ enthalten (z. B. [Pfad bei Karma Home]\preloaded-ontologies).

[880] Die RDF-Repräsentationen der CIDOC CRM Ontologie finden sich auf der Webseite unter http://www.cidoc-crm.org/versions-of-the-cidoc-crm (CIDOC CRM), http://www.cidoc-crm.org/crmsci/fm_releases (CRMsci) und http://www.cidoc-crm.org/crmarchaeo/fm_releases (CRMarchaeo), abgerufen am 21.10.2021. Andere Ontologien, wie RDF-Schema, SKOS oder dcterms finden sich unter der Tutoriums-Seite von Kama https://github.com/szeke/karma-tcdl-tutorial/wiki, abgerufen am 21.10.2021.

Nun kann die Textdatei mit den Daten, welche in einem vorigen Schritt als UTF-8 Textdatei erzeugt wurde, in Karma importiert werden. Dafür kann in Karma in der oben angebrachten Menüleiste „Import“ und in weiterer Folge „From File“ ausgewählt werden und die UTF-8 Datei hineingeladen werden. Im erscheinenden Fenster kann dann die Option „CSV Text File“ angewählt werden. Daraufhin erscheint ein Fenster, welches die Einstellungen für den Import regelt. Hier sollte unter „Column delimiter“ „tab“ für Tabulatortrennung, bei „Text qualifier“ die Anführungszeichen ("), bei „Header start index“ „1“, bei „Data Start index“ „2“, bei „Encoding“ „Unicode UTF-8“ und bei „Rows to import“ „10.000“ eingetragen werden (siehe Abbildung 56). Die Einstellungen geben an, dass die Textdatei mit Tabulatortrennung im Unicode UTF-8 Format gelesen werden kann. Im unteren Bereich des Fensters ist eine Vorschau für die importierten Daten sichtbar. Wenn diese richtig erscheint, können die Daten mit „Import“ hineingeladen werden.[881]

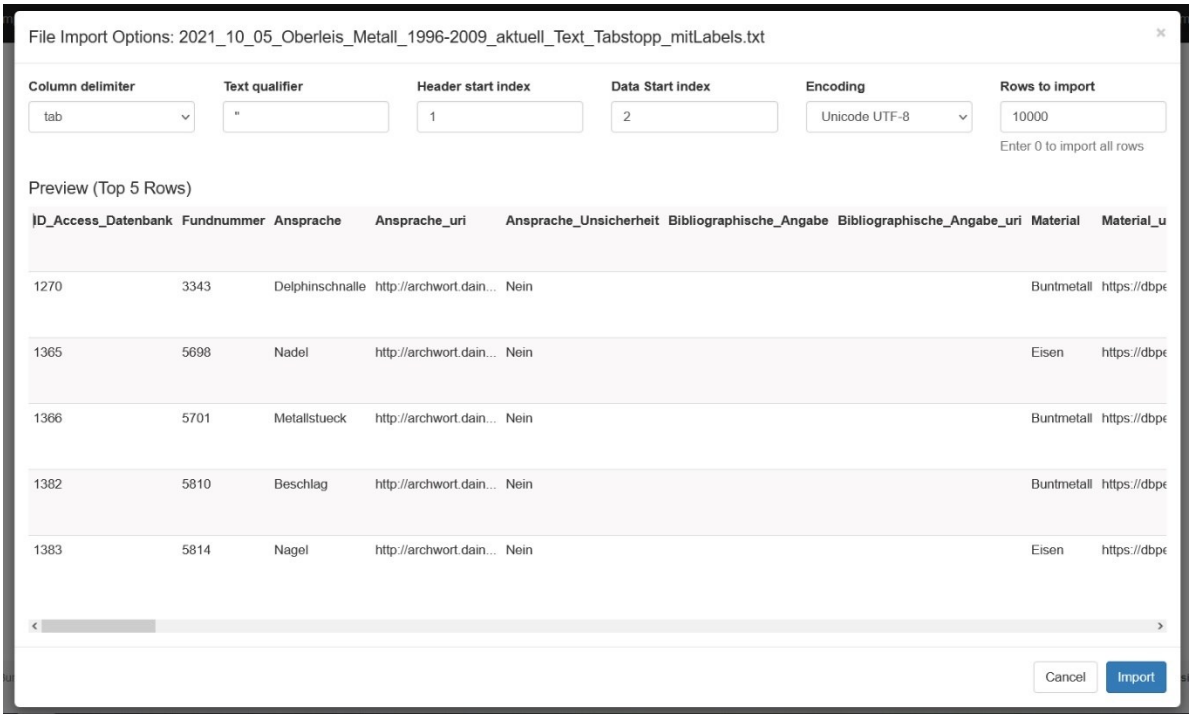
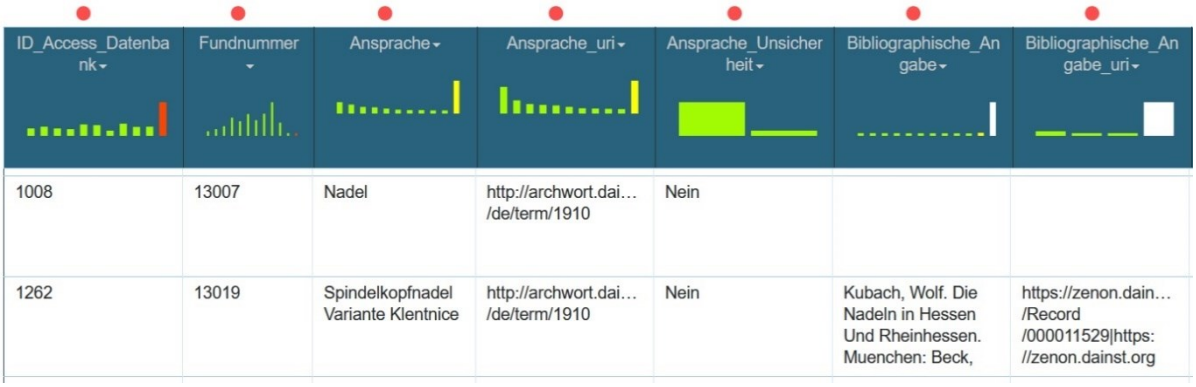


Abbildung 56 - Import der Textdatei in Karma.

[881] Wenn der Import nicht problemlos funktioniert, kann die UTF-8 Textdatei zuerst in den Karma Home Ordner, in welchem sich auch die Ontologien befinden, in den Unterordner „user-uploaded-files“ kopiert werden.

Die Daten erscheinen nun in einer Tabelle, welche über jeder Spalte einen roten Punkt aufweist (siehe Abbildung 57). Diese Knoten stellen die semantischen Verlinkungen zwischen den Spalten dar und können nun miteinander nach dem vorhin erstellten Datenmodell verbunden werden.



ID_Access_Datenbank	Fundnummer	Ansprache	Ansprache_uri	Ansprache_Unsicherheit	Bibliographische_Angabe	Bibliographische_Angabe_uri
1008	13007	Nadel	http://archwort.dai… /de/term/1910	Nein		
1262	13019	Spindelkopfnadel Variante Klentnice	http://archwort.dai… /de/term/1910	Nein	Kubach, Wolf. Die Nadeln in Hessen Und Rheinhessen. Muenchen: Beck,	https://zenon.dain… /Record /000011529|https: //zenon.dainst.org

Abbildung 57 - Datenimport in der Tabelle in Karma.

Zu Beginn kann der semantische Typ der erstellten URI für das Fundobjekt festgelegt werden. Dazu wird in der Tabelle die Spalte „E22_Human-made_object_uri“ des Hilfsknotens gesucht und das Dropdownmenü geöffnet. Hier kann der semantische Typ über den Punkt „Set Semantic Type“ ausgewählt werden (siehe Abbildung 58).

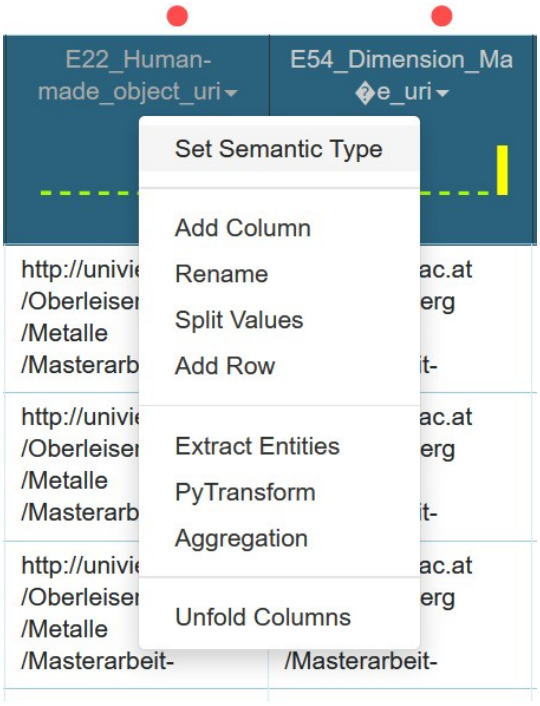


Abbildung 58 – Festsetzen des semantischen Typs der URI des Fundobjekts in Karma.

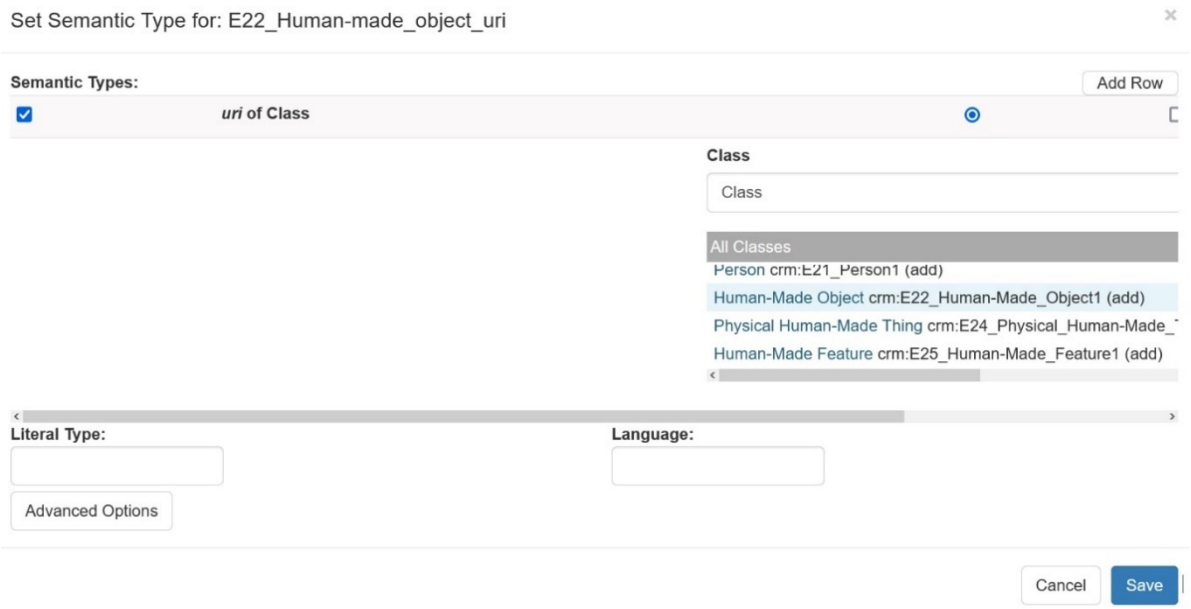


Abbildung 59 - Auswahl der Klasse aus der Ontologie für eine Spalte in der Tabelle in Karma.

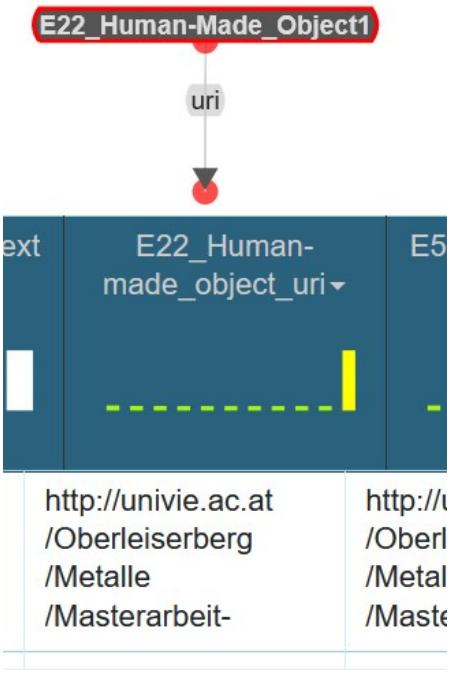


Abbildung 60 - Semantischer Typ der Spalte „E22_Human-made_object_uri" in Karma.

Im Fenster, welches sich daraufhin öffnet, kann der Punkt „uri of Class" angehakt werden und unter dem Punkt „edit" spezifiziert werden. Hier kann die die Klasse *E22 Human-Made Object* aus der Ontologie CIDOC CRM ausgewählt und zum Modell hinzugefügt werden (siehe Abbildung 59).

Nun ist der erste semantische Typ im Datenmodell festgesetzt und kann mit weiteren Spalten verbunden werden (siehe Abbildung 60). Die URI des Fundobjekts bildet den Ausgangspunkt für die weitere Modellierung. Alle weiteren Verbindungen, wie sie im CIDOC CRM-Datenmodell dargestellt sind (siehe Kapitel 8.2.1.3), wie beispielsweise die Produktion, das Material oder der Erhaltungszustand, können von diesem Typ aus verbunden werden.

Dafür kann auf den schwarzen Knoten der Klasse „E22_Human-Made_Object1“ geklickt werden und eine von dieser Klasse ausgehende Verbindung über den Punkt „Outgoing Link“ hinzugefügt werden. Hier wird die Produktion des Fundobjekts im Modell hinzugefügt. In einem Fenster kann die Klasse und Eigenschaft der Verbindung ausgewählt werden. In diesem Fall wird die Klasse *E12 Production* und die Eigenschaft *P108i was produced by* verwendet (siehe Abbildung 61). In Karma wird der Knoten „E12_Production1“ hinzugefügt. Das Resultat wird im Datenmodell anschließend über Verbindungen visualisiert und die weitere Modellierung kann fortgesetzt werden (siehe Abbildung 62).

Abbildung 61 – Verbindungen von Klasse und Eigenschaft im Datenmodell in Karma herstellen.

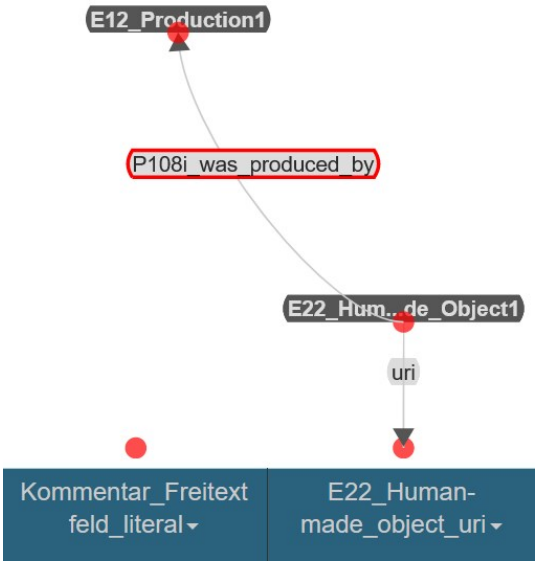


Abbildung 62 - Darstellung der Verbindungen im Datenmodell in Karma.

Als Beispiel für die Modellierung in Karma soll gezeigt werden, wie die Produktion und die Datierung der Fundobjekte verbunden sind. Die Produktion wurde im Datenmodell des CIDOC CRM (siehe Kapitel 8.2.1.3, Abbildung 47) mit der Datierung des Fundobjekts verbunden. Dies kann in Karma über die Verbindung des Knotens „E12_Production1" zu mehreren Knoten der Klasse *E4 Period* und zu einer Klasse *E52 Time-Span* für die absoluten Zahlen modelliert werden. Zunächst wird eine ausgehende Verbindung des Knotens „E12_Production1" über die Eigenschaft *P176i starts after the start of* zu einem neuen Knoten „E4_Period1" hergestellt. Die Nummerierung der Knoten ermöglicht es, mehrere Knoten der Klasse *E4 Period* hinzuzufügen. Es werden insgesamt vier Knoten der Klasse *E4 Period* benötigt, jeweils einer für die Spalten „Datierung von" und „Datierung bis" und die Spalten „Kulturelle Zuordnung von" und „Kulturelle Zuordnung bis". Der neue Knoten der „E4_Period1" kann nun zu den Spalten „Datierung_von" und „Datierung_von_uri" verlinkt werden. Hierbei ist es für spätere Visualisierungen vorteilhaft, die Spalte „Datierung_von" mit der Eigenschaft *rdfs:label*[882,883] zu versehen. Diese stammt aus dem RDF Schema, welches ein Vokabular für RDF-Daten darstellt und ermöglicht es, die Inhalte der Spalte mit einem menschenlesbaren Titel zu versehen.[884] In der Visualisierung ist hier beispielsweise „Bronzezeit" zu lesen. Um die Spalte mit der Eigenschaft zu versehen, wird der semantische Typ der Spalte „Datierung_von" mit „Set Semantic Type" gesetzt. Im erscheinenden Fenster wird

[882] Das RDF Schema-Vokabular muss ebenfalls in den Ordner „preloaded-ontologies" geladen werden. Der Pfad dorthin wird im unteren linken Bildschirmrand angezeigt. Das Vokabular findet sich auf der Webseite des Karma-Tutorials unter https://github.com/szeke/karma-tcdl-tutorial/wiki, abgerufen am 21.10.2021, in den „Tutorial Files".
[883] Rdfs:label ist die Abkürzung durch einen Namensraum, welcher in der Ontologie des RDF Schemas festgelegt ist. Die Abkürzung steht für <http://www.w3.org/2000/01/rdf-schema#label>.
[884] Brickley u. a. 2014, Kapitel 3.6.

„Property of Class“ ausgewählt und weiter editiert. Im Editierungsfenster kann unter „Property“ *rdfs:label* und unter „Class“ der bereits angelegte Knoten „E4_Period1“ ausgewählt werden (siehe Abbildung 63). Um die URI für die Datierung des Knotens „E4_Period1“ hinzuzufügen, wird von der Spalte „Datierung_von_uri“ aus der semantische Typ mit „uri of Class“ ausgewählt und die Klasse „E4_Period1“ eingegeben (siehe Abbildung 64). Die Spalte „Datierung_von_uri“ stellt demnach die URI der Periode „E4_Period1“ dar. Das Zwischenergebnis der Modellierung ist in Abbildung 65 abgebildet.

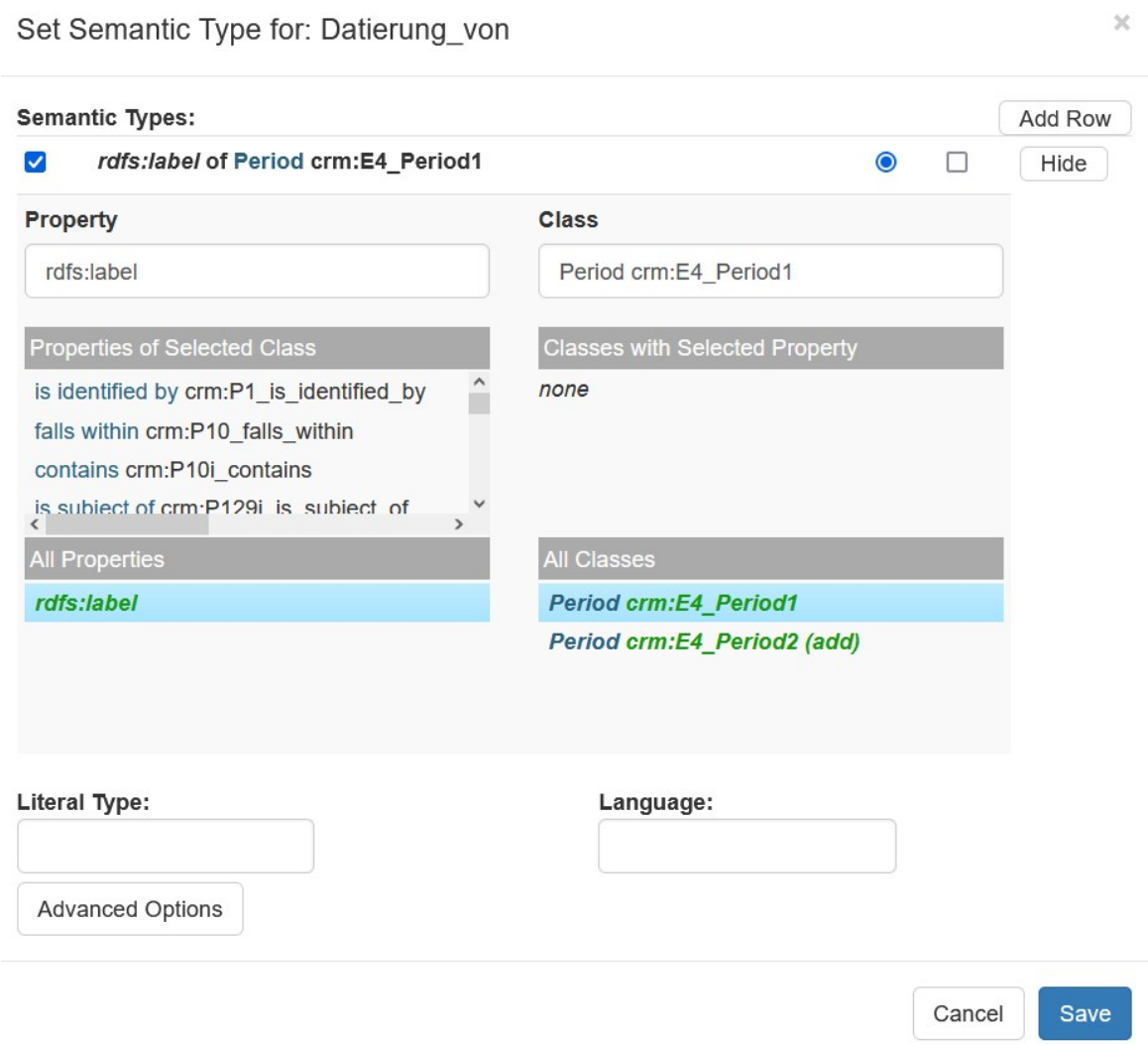


Abbildung 63 – Titel für die Datierung des Fundobjekts über die Eigenschaft rdfs:label in Karma hinzufügen.

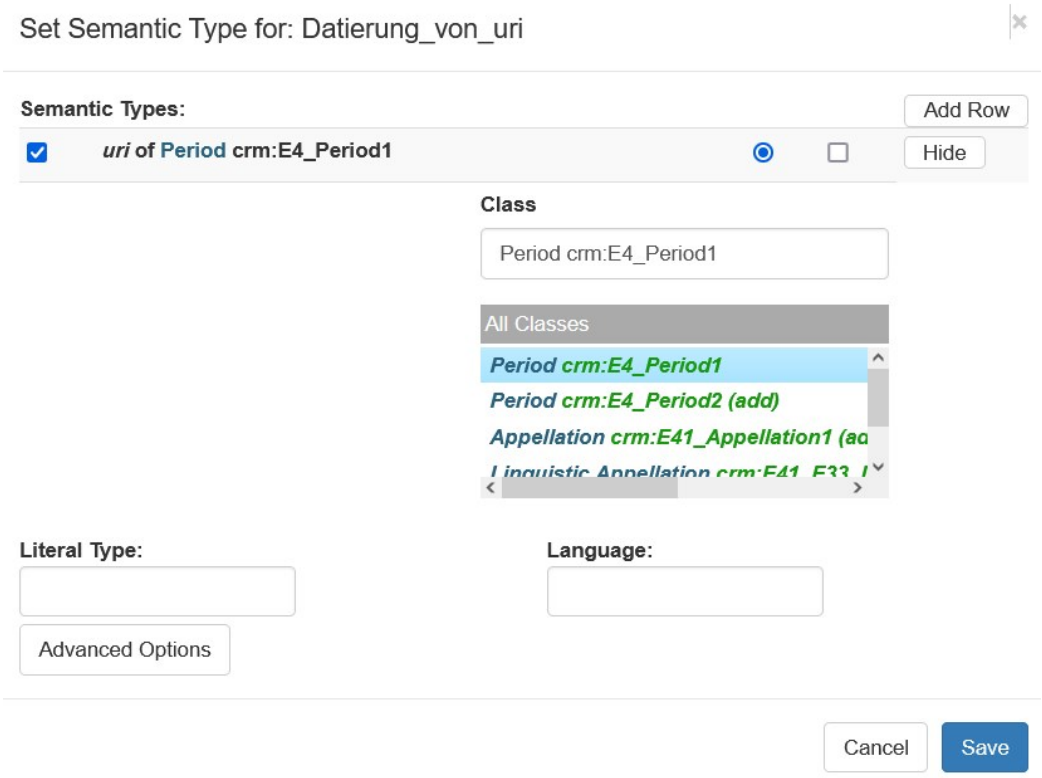


Abbildung 64 - Festlegen der URI der Klasse „E4_Period1" in Karma.

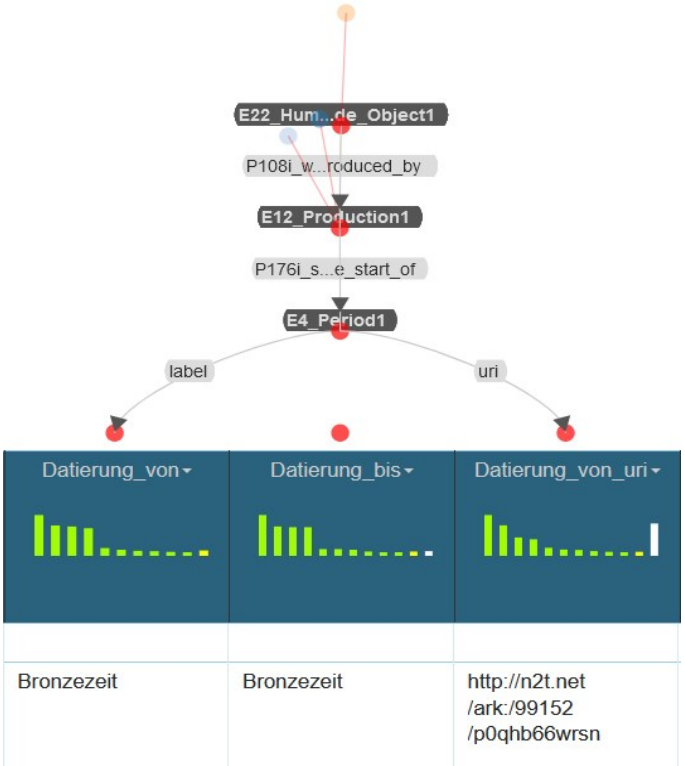


Abbildung 65 - Modellierung der Periode „E4 Period1" in Karma.

Die gleiche Vorgehensweise kann nun bei der „Datierung_bis“ angewandt werden. Hierfür wird ein neuer Knoten der „E4_Period2“ angelegt, welcher zu den Spalten „Datierung_bis“ mit rdfs:label und „Datierung_bis_uri“ als URI verbunden wird (siehe Abbildung 66).

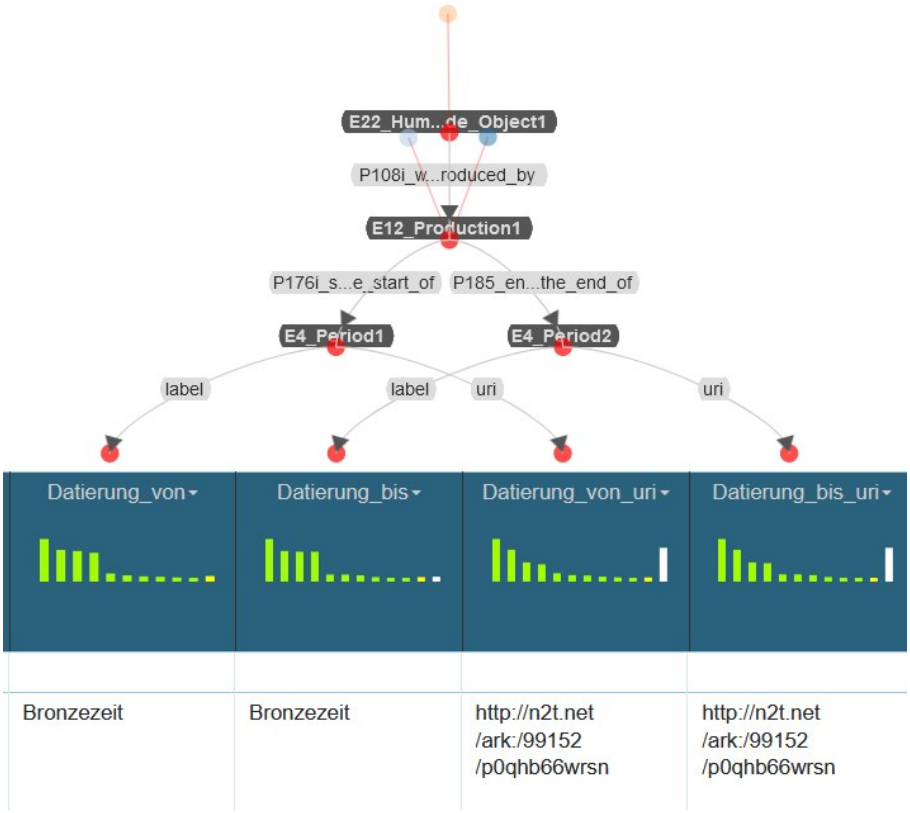


Abbildung 66 - Modellierung der Perioden „E4 Period1“ und „E4 Period2“ in Karma.

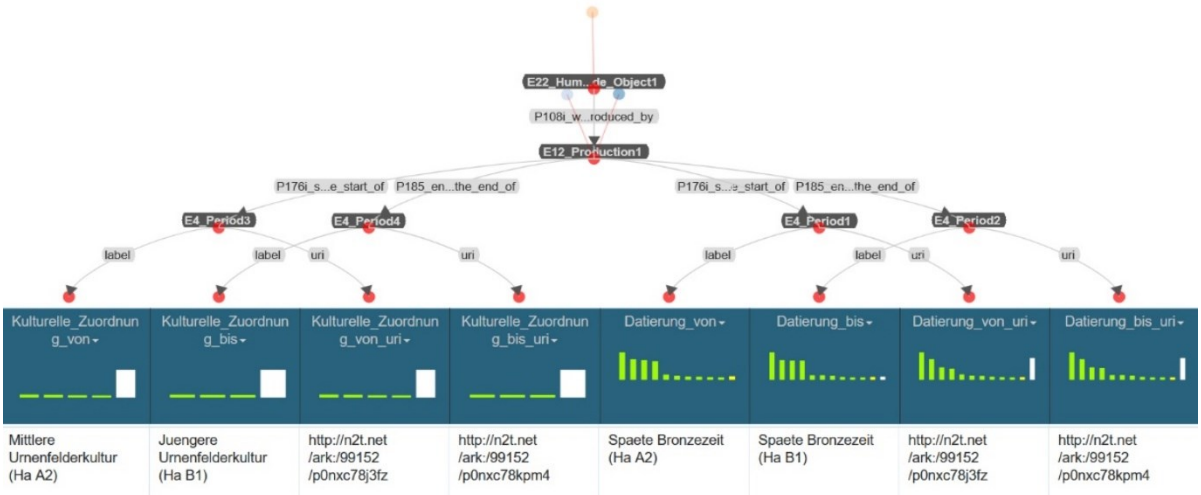


Abbildung 67 - Modellierung aller vier Perioden „E4 Period“ 1 bis 4 in Karma.

Die gleichen Schritte können nun für die Spalten „Kulturelle Zuordnung von“ und „Kulturelle Zuordnung bis“ angewandt werden. Auch hier werden jeweils die URI und die Label hinzugefügt, diesmal zu Periode 3 und Periode 4 (siehe Abbildung 67).

Um die absoluten Zahlen in die Modellierung miteinzubinden, können die beiden Spalten „Datierung_absolute_Zahlen_von“ und „Datierung_absolute_Zahlen_bis“ zu dem „Intermediate Node“ „*E52 Time-Span*“ verlinkt werden. Dieser muss zuerst mit einer URI versehen werden, die für die „Intermediate Node“ *E52 Time-Span* erstellt wurde. Danach wird die *E52 Time-Span* zu der Produktion verbunden. Dies kann mit einem „Outgoing Link“ von dem Knoten „E12 Production1“ mit der Eigenschaft *P4 has time-span* durchgeführt werden. Die Verlinkung zu den Spalten mit den absoluten Zahlen wird mit den Eigenschaften *P79 beginning is qualified by* für die Spalte „Datierung_absolute_Zahlen_von“ und *P80 end is qualified by* für die Spalte „Datierung_absolute_Zahlen_von“ vorgenommen. Die Modellierung der Datierung des Fundobjekts über den Knoten der Produktion ist in Abbildung 68 dargestellt. Auf diese Weise kann in weiterer Folge die Modellierung fortgesetzt werden.

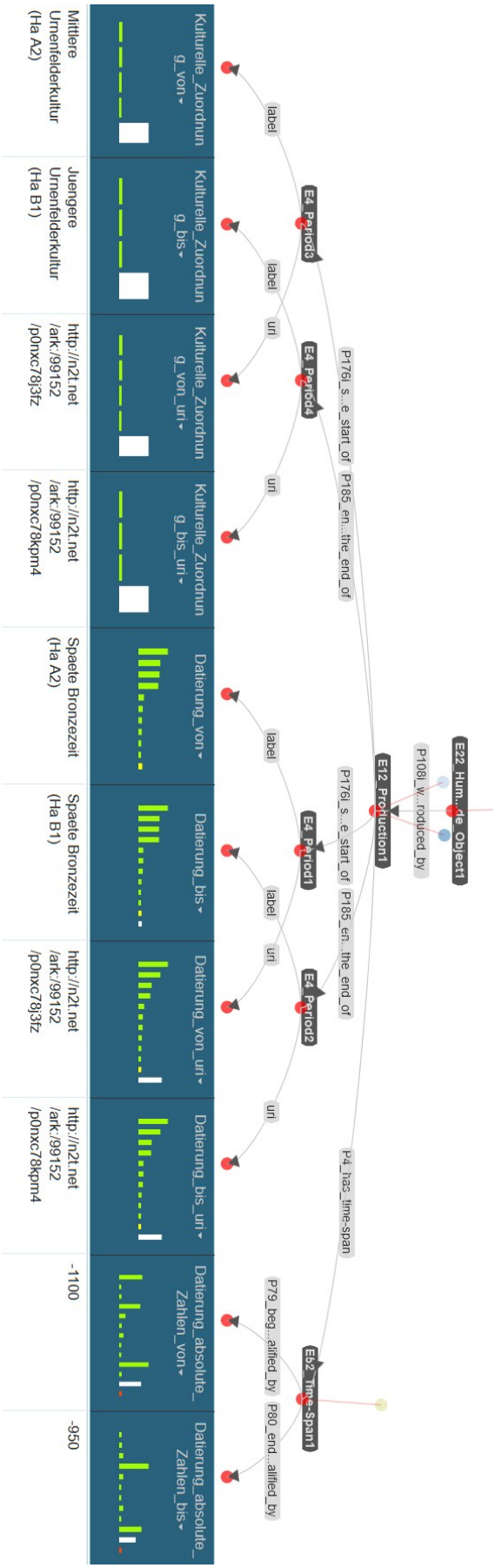


Abbildung 68 – Modellierung der Datierung des Fundobjekt in Karma.

8.2.2.4 Export in RDF

Wenn das Modell fertiggestellt ist, können die Daten in einem RDF-Format aus Karma exportiert werden. Hierfür muss das Modell zuerst publiziert werden. Um dies durchzuführen, kann das Menü in der Leiste über dem Datenmodell, in welcher der Dateiname steht, angeklickt werden und die Menüpunkte „Publish“ und folgend „RDF“ ausgewählt werden (siehe Abbildung 69).

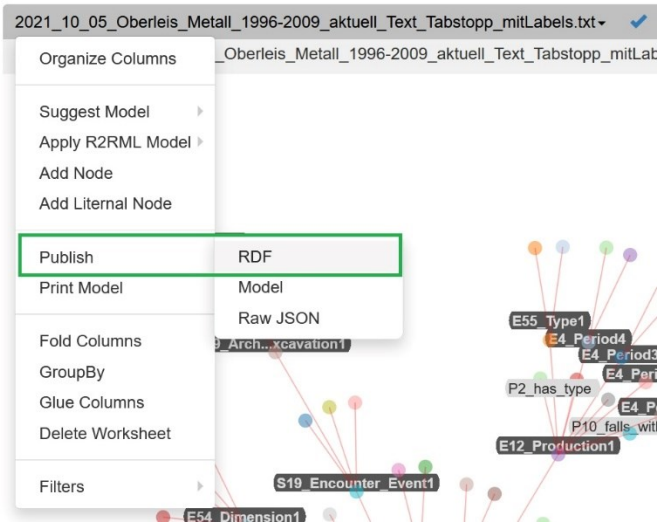


Abbildung 69 – Das Datenmodell in Karma publizieren.

Im neu erscheinenden Fenster werden die Einstellungen „Create New Context“ und „Replace Existing Data“ ausgewählt. Der Graph, welcher erstellt wird, wird bestmöglich eindeutig der Institution, dem Projekt und der Tabelle zuordenbar in der Zeile „Create New Graph“ benannt (siehe Abbildung 70). Im Anschluss kann in Karma auf der rechten oberen Bildschirmecke zur „OpenRDF-Workbench“ gewechselt werden und dort unter „Repositories“ das aktuelle Repositorium ausgewählt werden. Im Menüpunkt „Explore“ kann der Unterpunkt „Contexts“ ausgewählt werden. Hier befindet sich der RDF-Graph (siehe Abbildung 71).

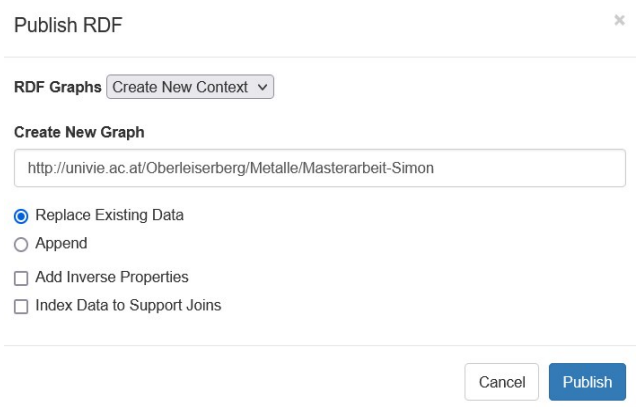


Abbildung 70 – Einstellungen zum Erstellen des RDF-Graphen in Karma.

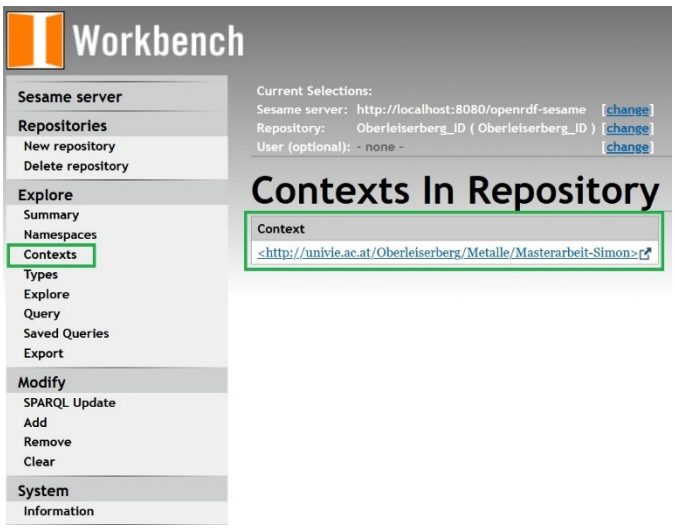


Abbildung 71 – RDF-Kontext in OpenRDF.

Um den Graphen zu exportieren, kann in den Menüpunkt „Export“ gewechselt werden und das Repositorium im gewünschten Format heruntergeladen werden (siehe Abbildung 72). Es können un-

ter anderen auch die Formate N-Triples, RDF/XML, Turtle, RDF/JSON und N-Quads ausgewählt werden. Wenn mehr als ein Datenset weiterverarbeitet wird, ist es ratsam, N-Quads zu wählen, da hier immer der Kontext, aus welchem Graphen die Daten stammen, hinzugefügt wird.

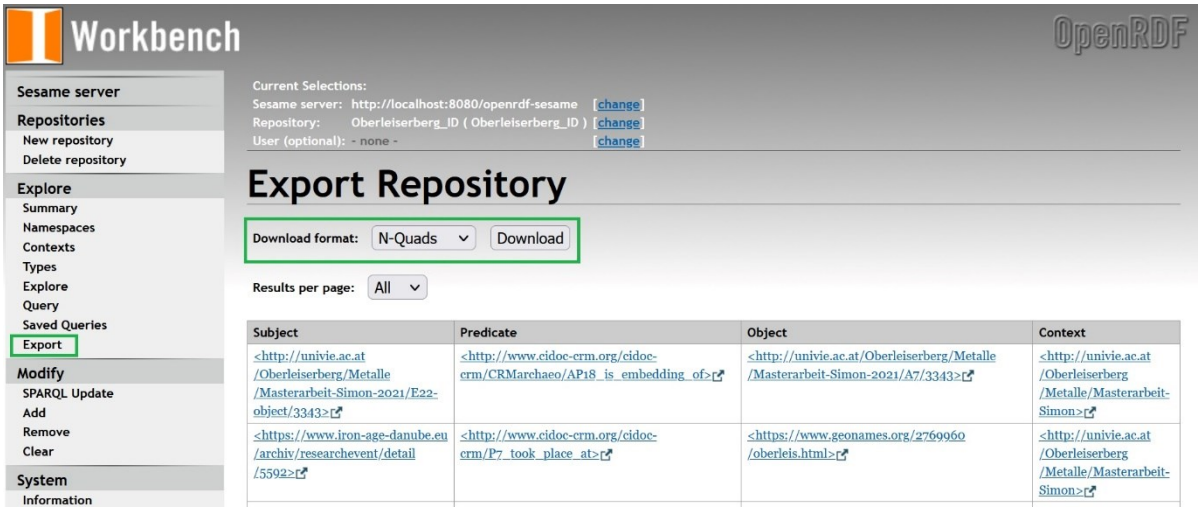


Abbildung 72 – Export des RDF-Graphen aus OpenRDF.

Die Daten des Oberleiserbergs wurden im Format N-Quads als NQ-Datei mit der Endung „.nq“ heruntergeladen. Zur Schreibweise der N-Quads Serialisierung siehe Kapitel 4.3.2.2.3. Wenn die Datei im Texteditor geöffnet wird, können die Triples mit Kontext gelesen werden. Im Folgenden ist ein Auszug der Daten zu sehen:

```
1       <http://univie.ac.at/Oberleiserberg/Metalle/
        Masterarbeit-Simon-2021/E22-object/3343> <http://
        www.cidoc-crm.org/cidoc-crm/P45_consists_of>
        <https://dbpedia.org/page/Non-ferrous_metal>
        <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon> .
2       <http://univie.ac.at/Oberleiserberg/Metalle/
        Masterarbeit-Simon-2021/E12/3343> <http://www.cidoc-crm.org/
        cidoc-crm/P2_has_type> <http://archwort.dainst.org/de/term/1869>
        <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon> .
3       <http://univie.ac.at/Oberleiserberg/Metalle/
        Masterarbeit-Simon-2021/E12/3343> <http://www.cidoc-crm.org/
        cidoc-crm/P79_beginning_is_qualified_by> "-2200"
        <http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon> .
```

Die Daten sind in Triples mit vollständigen URIs angeordnet, wobei der Graph jeweils nachgestellt ist (hier fett gedruckt). Das erste Triple beschreibt, dass die URI des Fundobjekts der Fundnummer 3343 aus Buntmetall besteht. Das Triple in Zeile zwei beschreibt den Typ der Produktion (*E12 Production*) der Fundnummer 3343 mit der URI <http://archwort.dainst.org/de/term/1869>. Diese ist im Vokabular des Deutschen Archäologischen Instituts hinterlegt als „gegossen“. Das dritte Triple beschreibt den Beginn der Produktion (*E12 Production*) des Fundobjekts der Fundnummer 3343 mit der Zahl „400“.

8.2.3 Visualisierung des Repositoriums

Für die Visualisierung kann beispielsweise die Online-Graphendatenbank GraphDB[885] verwendet werden. Die Graphendatenbank unterstützt RDF-Graphen und kann SPARQL-Abfragen mit den geladenen Daten durchführen. Um den RDF-Graphen hineinzuladen, soll zunächst ein neues Repositorium angelegt werden. Dafür muss im Menüpunkt „Setup“ der Unterpunkt „Repositories“ ausgewählt und rechts ein neues Repositorium unter „Create new repository“ angelegt werden (siehe Abbildung 73). Im nächsten Schritt kann die Repository ID eingetragen und die Voreinstellungen übernommen werden. Um mit dem Repositorium arbeiten zu können, muss es im rechten oberen Bildschirmeck im Dropdown-Menü ausgewählt werden.

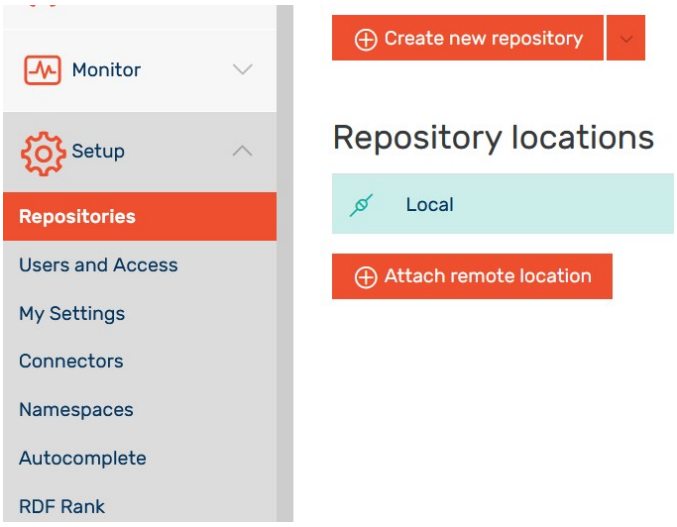


Abbildung 73 – Ein neues Repositorium anlegen in GraphDB.

Nun kann die RDF-Datei importiert werden. Dabei ist darauf zu achten, dass sich keine Sonderzeichen (z. B. |) in der Datei befinden. Im linken Menü kann unter dem Punkt „Import“ – „RDF“ die Kategorie „Upload RDF files“ ausgewählt werden (siehe Abbildung 74). Nun erscheint die Datei in einer Liste, kann ausgewählt und importiert werden. Die Voreinstellungen können übernommen werden. Im Menüpunkt „Explore“ und „Graphs overview“ kann der Graph ausgewählt und angesehen werden. Er wird in einer Liste mit Subjekt, Prädikat und Objekt angegeben. Wenn ein Link in der Spalte des Subjekts ausgewählt wird, erscheinen alle Triples, die mit ihm in Verbindung stehen. Im Folgenden sollen Beispiele für die visuelle Darstellung der Graphen gezeigt werden.

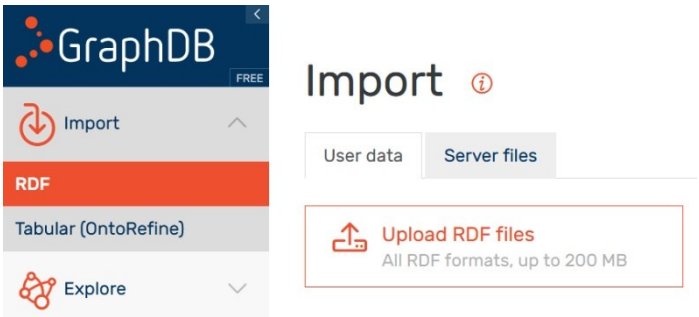


Abbildung 74 – Import der RDF-Datei in GraphDB.

[885] Zum Download unter https://www.ontotext.com/products/graphdb/, abgerufen am 21.10.2021. Informationen unter https://graphdb.ontotext.com/free/index.html, abgerufen am 21.10.2021.

8.2.3.1 Produktion

Die Abbildung 75 zeigt alle Triples, welche mit der Produktion des Fundobjekts der Fundnummer 13019 zusammenhängen. In Abbildung 76 wird der visuelle Graph gezeigt, welcher erscheint, wenn auf „Visual Graph“ geklickt wird.

Produktion

Source: http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019

subject | predicate | object | context | all — Explicit only — Show

	subject	predicate	object
1	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019	http://www.cidoc-crm.org/cidoc-crm/P176i_starts_after_the_start_of	http://n2t.net/ark:/99152/p0nxc78j3fz
2	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019	http://www.cidoc-crm.org/cidoc-crm/P185_ends_before_the_end_of	http://n2t.net/ark:/99152/p0nxc78kpm4
3	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019	http://www.cidoc-crm.org/cidoc-crm/P4_has_time-span	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E52-Datierung/13019
4	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019	rdf:type	http://www.cidoc-crm.org/cidoc-crm/E12_Production
5	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019	rdfs:label	"Produktion"@de
6	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.cidoc-crm.org/cidoc-crm/P108i_was_produced_by	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019

Abbildung 75 - Triples der Produktion des Fundobjekts Nr. 13019 in GraphDB.

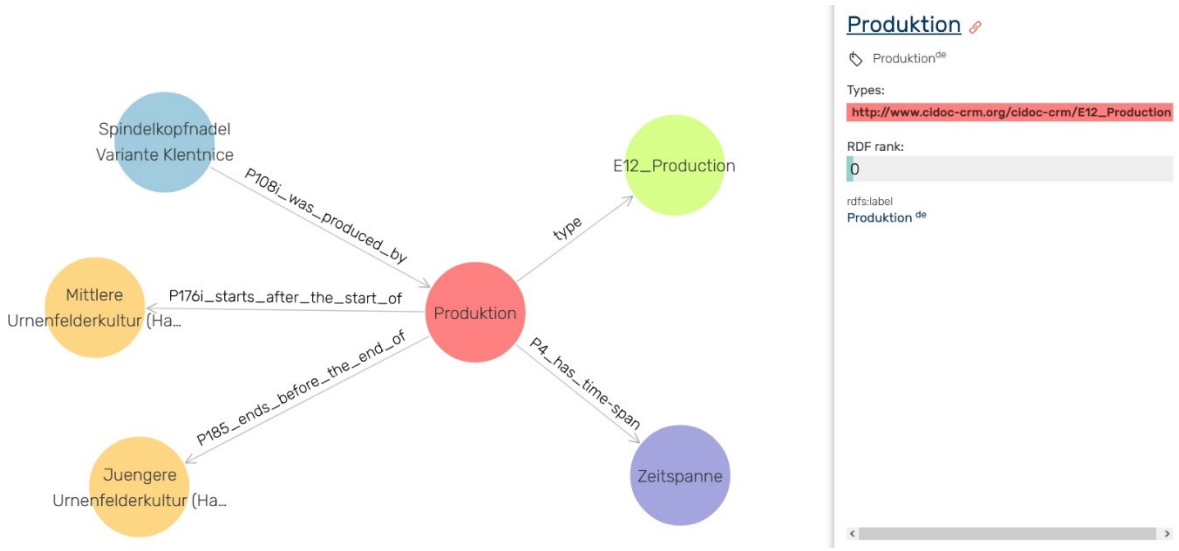


Abbildung 76 - Visueller Graph der Triples zur Produktion des Fundobjekts Nr. 13019 in GraphDB.

Im Zentrum steht die URI der Produktion, welche zu Beginn generiert wurde (<http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019>). Durch die Modellierung mit der Eigenschaft *rdfs:label* in Karma ist im zentralen Knoten „Produktion“ zu lesen. Die kulturelle Zuordnung wird in den linken beiden Knoten angezeigt. Die Produktion des Fundobjekts beginnt (*P176i starts after the start of*) in der Mittleren Urnenfelderkultur (Ha A2) und endet (*P185 ends before the end of*) in der Jüngeren Urnenfelderkultur (Ha B1). Auf der rechten Bildschirmseite werden weitere Informationen angezeigt. Die absoluten Zahlen vom Beginn und dem Ende der Zeitspanne der Produktion werden angezeigt, wenn auf den Knoten „Zeitspanne“ geklickt wird. Die „Produktion“ hat den Typ (*P2 has type*) *E12 Production* aus der Ontologie CIDOC CRM. Dies ist auch in der Grafik ersichtlich. Links oben ist der Knoten mit der Ansprache des Fundobjektes „Spindelkopfnadel Variante Klentnice“ zu sehen. Diese wurde produziert (*P108i was produced by*) durch die „Produktion“.

8.2.3.2 Fundobjekt

Abbildung 77 zeigt alle Triples, welche mit dem Knoten *E22 Human-Made Object* des Fundobjekts 13019 zusammenhängen. Der visuelle Graph ist in Abbildung 78 dargestellt. Er zeigt alle direkten Verbindungen zu dem Knoten *E22 Human-Made Object*, welcher mit *rdfs:label* den Titel „Spindelkopfnadel Variante Klentnice“ zugewiesen bekommen hat.

Spindelkopfnadel Variante Klentnice

Source: http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019

subject | predicate | object | context | all — Explicit only

	subject	predicate	object
1	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.cidoc-crm.org/cidoc-crm/CRMarchaeo/AP18_is_embedding_of	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/A7/13019
2	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.cidoc-crm.org/cidoc-crm/P108i_was_produced_by	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E12/13019
3	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.cidoc-crm.org/cidoc-crm/P2_has_type	http://archwort.dainst.org/de/term/1910
4	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.cidoc-crm.org/cidoc-crm/P3_has_note	"Runder Schaft mit rillenverziertem Spindelkopf; abgebrochen."@de
5	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.cidoc-crm.org/cidoc-crm/P44_has_condition	http://archwort.dainst.org/de/term/1595
6	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.cidoc-crm.org/cidoc-crm/P45_consists_of	https://dbpedia.org/page/Non-ferrous_metal
7	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.cidoc-crm.org/cidoc-crm/P51_has_former_or_current_owner	http://viaf.org/viaf/245829558
8	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.cidoc-crm.org/cidoc-crm/P54_has_current_permanent_location	http://viaf.org/viaf/245829558
9	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	http://www.ics.forth.gr/isl/CRMsci/O19i_was_object_found_by	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/S19/13019
10	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	rdf:type	http://www.cidoc-crm.org/cidoc-crm/E22_Human-Made_Object
11	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019	rdfs:label	"Spindelkopfnadel Variante Klentnice"@de

Abbildung 77 - Triples der Klasse E22 Human-Made Object des Fundobjekts Nr. 13019 in GraphDB.

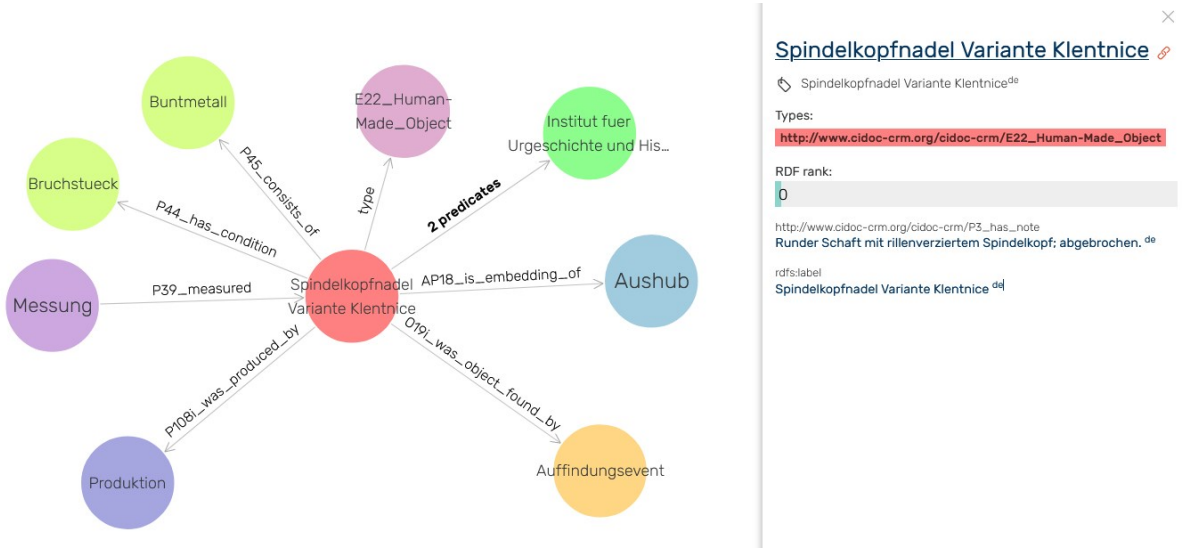


Abbildung 78 - Visueller Graph zur Klasse E22 Human-Made Object des Fundobjekts Nr. 13019 in GraphDB.

Die „Intermediate Nodes" des CIDOC CRM-Datenmodells Auffindungsevent (*S19 Encounter Event*), Produktion (*E12 Production*) und Messung (*E16 Measurement*) sind auf der unteren Hälfte des Graphen sichtbar. Sie können noch erweitert werden, um die weiteren verbundenen Informationen anzuzeigen. Das Fundobjekt wurde bei einem Auffindungsevent gefunden, beim Ereignis der Produktion produziert und bei der Messung abgemessen. Der Erhaltungszustand der Nadel ist als „Bruchstück" mit der Verbindung *P44 has condition* angezeigt. Der Knoten „Institut für Urgeschichte und Historische Archäologie" hat zwei Prädikate, welche beim Anklicken sichtbar werden. Das Fundobjekt wird über die Eigenschaft *P51 has former or current owner* und *P54 has current permanent location* mit dem Institut verbunden. Dies bedeutet, das Fundobjekt ist im Besitz des Instituts und hat auch den ihm permanent zugewiesenen Platz dort. Weiters ist zu sehen, aus welchem Material das Fundobjekt besteht. Das Material „Buntmetall" wird mit der Eigenschaft *P45 consists of* zum Objekt verbunden. Die Nadel wurde eingebettet in einer Klasse *A7 Embedding*. Diese wurde in Karma zur Spalte des „Befunds" verlinkt, deshalb ist hier „Aushub" zu lesen. Im Informationsfeld auf der rechten Seite ist die wörtliche Beschreibung des Fundobjekts zu lesen. Mit der Eigenschaft *P3 has note* können literale Notizen angehängt werden.

8.2.3.3 Auffindungsevent

Das Auffindungsevent wird im CIDOC CRM-Datenmodell mit der Klasse S19 Encounter Event modelliert. Es beschreibt das Ereignis, in welchem das Fundobjekt gefunden wurde. Abbildung 79 zeigt alle Triples im Graphen, welche zu dieser Klasse Verbindungen aufweisen.

Auffindungsevent

Source: http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/S19/13019

subject | predicate | object | context | all — Explicit only — Show Blan

	subject	predicate	object
1	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/S19/13019	http://www.cidoc-crm.org/cidoc-crm/P82_at_some_time_within	"1999"
2	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/S19/13019	rdf:type	http://www.ics.forth.gr/isl/CRMsci/S19_Encounter_Event
3	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/S19/13019	rdfs:label	"Auffindungsevent"@de

Abbildung 79 - Triples zur Klasse S19 Encounter Event (Auffindungsevent) des Fundobjekts der Nr. 13019 in GraphDB.

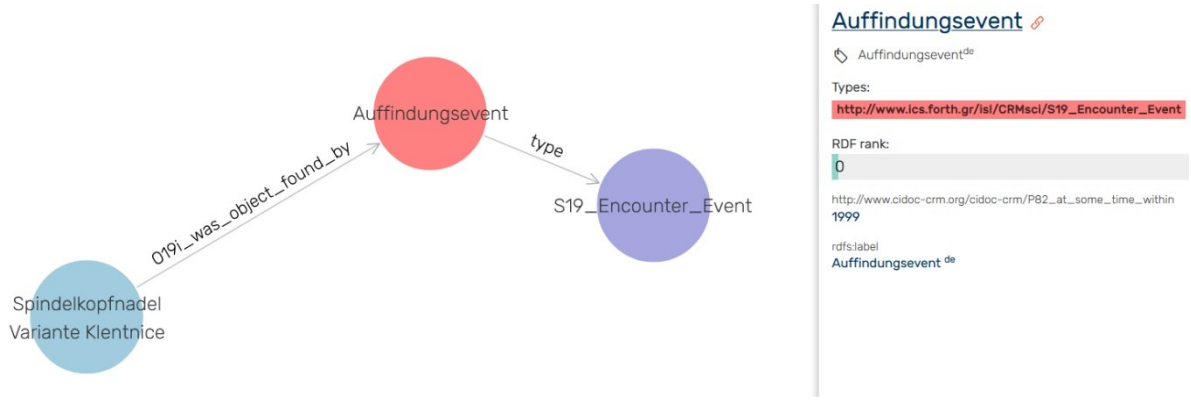


Abbildung 80 - Visueller Graph der Klasse S19 Encounter Event (Auffindungsevent) des Fundobjekts der Nr. 13019 in GraphDB.

Der visuelle Graph des Auffindungsevents ist in Abbildung 80 dargestellt. Er wurde als Typ mit der Klasse *S19 Encounter Event* versehen. Die Nadel wurde bei diesem Event gefunden. Weitere Informationen zu der Ausgrabung können hier im Modell angehängt werden. In der Informationsbox auf der rechten Seite ist zu sehen, dass dieses Ereignis im Jahr 1999 stattgefunden hat.

8.2.3.4 Messung

Als Messung wird im Datenmodell das Ereignis der Abmessung bezeichnet (*E16 Measurement*). In Abbildung 81 sind alle zugehörigen Triples im Graphen aufgelistet. Der visuelle Graph in Abbildung 82 zeigt, dass das Ereignis der Messung die Maße des Fundobjekts durch die Eigenschaft *P40 observed dimension* beobachtet hat und dass die „Spindelkopfnadel Variante Klentnice“ über das Ereignis mit der Eigenschaft *P39 measured* vermessen wurde.

Messung

Source: http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E16/13019

subject | predicate | object | context | all — Explicit only — Show

	subject	predicate	object
1	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E16/13019	http://www.cidoc-crm.org/cidoc-crm/P39_measured	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E22-object/13019
2	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E16/13019	http://www.cidoc-crm.org/cidoc-crm/P40_observed_dimension	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Abmessung/13019
3	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E16/13019	rdf:type	http://www.cidoc-crm.org/cidoc-crm/E16_Measurement
4	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E16/13019	rdfs:label	"Messung"@de

Abbildung 81 - Triples zur Klasse E16 Measurement des Fundobjekts der Nr. 13019 in GraphDB.

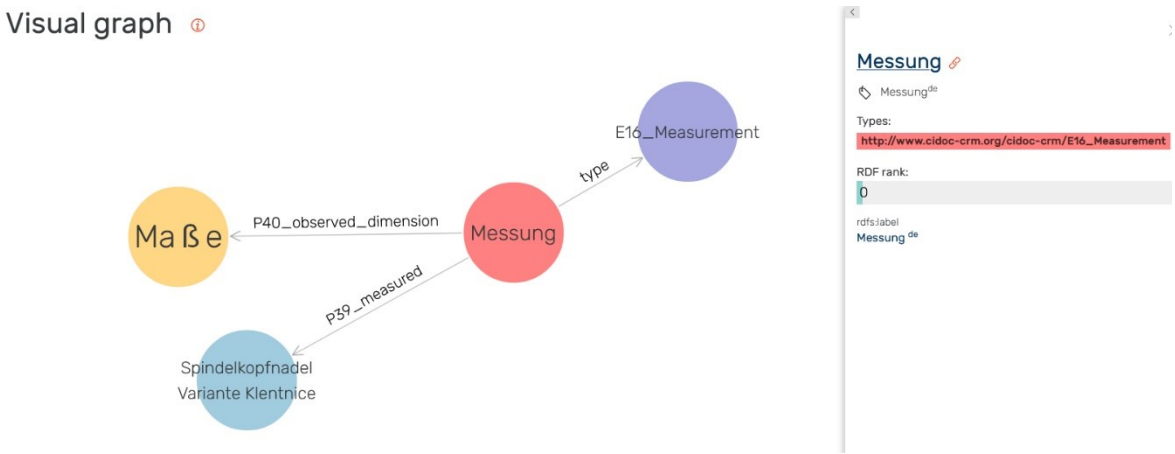


Abbildung 82 - Visueller Graph der Klasse E16 Measurement des Fundobjekts der Nr. 13019 in GraphDB.

8.2.3.5 Maße

Die Maße wurden mit der Klasse *E54 Dimension* modelliert und sind das Ergebnis des Messereignisses *E16 Measurement*. In Abbildung 83 sind alle Triples, die mit der Klasse *E54 Dimension* des Fundobjekts 13019 verbunden sind, aufgelistet. Die Abbildung 84 zeigt den dazugehörigen visuellen Graph. Die aus der Modellierung stammenden Ergebnisse sind auf der rechten Seite ausgewiesen. Dabei wurde angegeben, dass die Klasse *E54 Dimension* einen Typ hat (*P2 has type*). Der Typ „Länge" beschreibt, dass dieses Maß die Länge des Fundobjekts darstellt. Der Wert der Messung ist mit der Eigenschaft *P90 has value* angegeben. In diesem Fall ist der Wert der Messung „3,41". Die Einheit wird separiert mit der Eigenschaft *P91 has unit* angegeben („cm").

Maße

Source: http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Abmessung/13019

subject | predicate | object | context | all — Explicit only — Show

	subject	predicate	object
1	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Abmessung/13019	http://www.cidoc-crm.org/cidoc-crm/P2_has_type	"Länge"
2	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Abmessung/13019	http://www.cidoc-crm.org/cidoc-crm/P90_has_value	"3,41"
3	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Abmessung/13019	http://www.cidoc-crm.org/cidoc-crm/P91_has_unit	"cm"
4	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Abmessung/13019	rdf:type	http://www.cidoc-crm.org/cidoc-crm/E54_Dimension
5	http://univie.ac.at/Oberleiserberg/Metalle/Masterarbeit-Simon-2021/E54-Abmessung/13019	rdfs:label	"Maße"@de

Abbildung 83 - Triples der Klasse E54 Dimension des Fundobjekts der Nr. 13019 in GraphDB.

Abbildung 84 - Visueller Graph der Klasse E54 Dimension des Fundobjekts der Nr. 13019 in GraphDB.

In Abbildung 85 wird nun der gesamte Graph mit allen Verbindungen einer Rippenkopfnadel angezeigt. Es wird der Eigentümer und der Auffindungsort (Institut für Urgeschichte und Historische Archäologie), der Befund (Objekt 138), das Auffindungsereignis, das Material (Buntmetall), die Produktion mit der kulturellen Zuordnung und der Produktionsart und die Maße angezeigt.

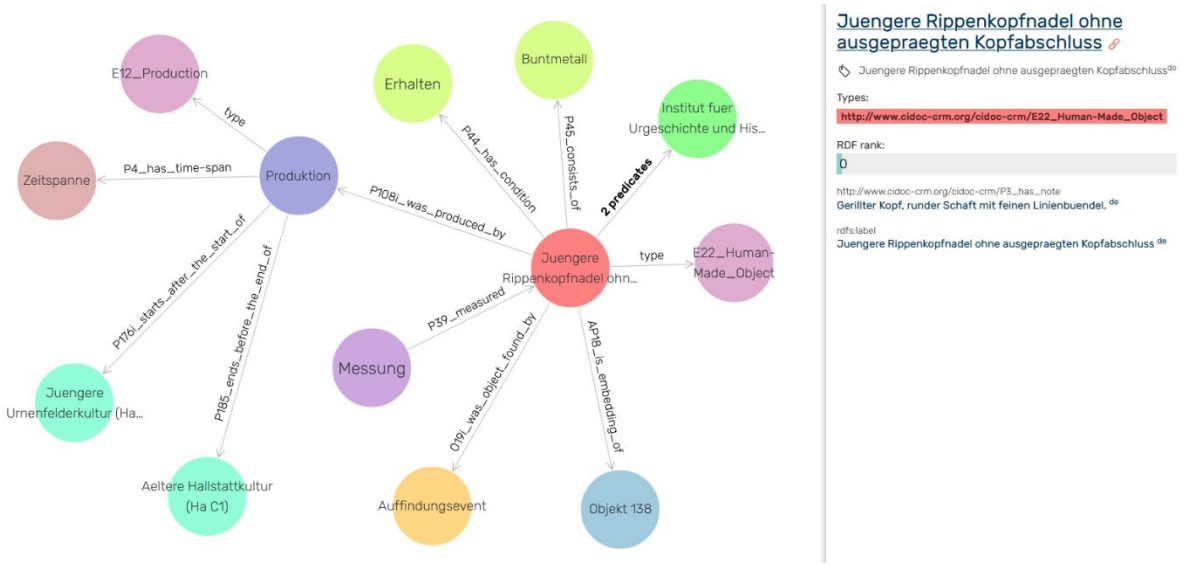


Abbildung 85 - Gesamter visueller Graph einer Rippenkopfnadel (Fundnummer 18308) in GraphDB.

8.2.4 Abfrage in SPARQL

SPARQL-Abfragen können durchgeführt werden, wenn der Graph in einen Triplestore geladen wurde (siehe Kapitel 4.5.2). Wenn die RDF-Daten in einem Repositorium in GraphDB hinterlegt sind, wie oben beschrieben, können auch hier SPARQL-Abfragen durchgeführt werden.[886] Hier-

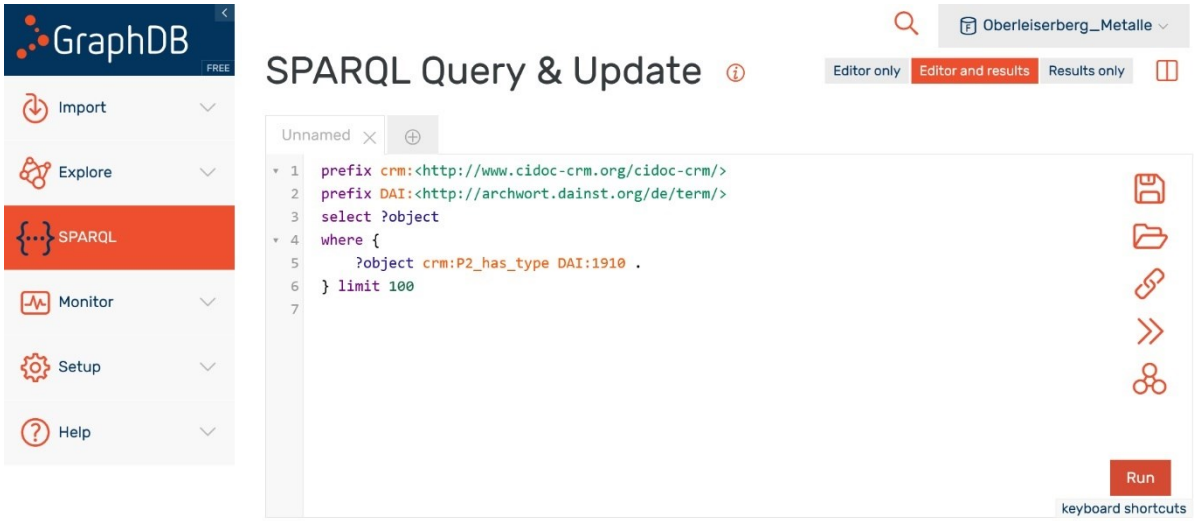


Abbildung 86 - Eingabemaske der SPARQL-Abfrage in GraphDB.

[886] Siehe hierfür https://graphdb.ontotext.com/free/quick-start-guide.html#query-your-data, abgerufen am 21.10.2021.

für kann, wenn das Repositorium an rechten oberen Bildschirmrand ausgewählt ist, in der Menüleiste unter „SPARQL“ die Eingabemaske aufgerufen werden. Im Anschluss können die Daten über die Eingabemaske via SPARQL-Abfragen durchsucht und aktualisiert werden. Abbildung 86 zeigt die Eingabemaske in GraphDB.

Abbildung 86 zeigt ein Beispiel einer SPARQL-Abfrage der Daten des Oberleiserbergs. Darin wird nach Objekten gefragt, welche als „Gewandnadel“ angesprochen werden. Die ersten beiden Zeilen zeigen die Präfixe an. In diesem Fall steht „crm:“ für <http://www.cidoc-crm.org/cidoc-crm/> und „DAI:“ für <http://archwort.dainst.org/de/term/>. Es können in der Abfrage demnach Abkürzungen für CIDOC CRM und dem Vokabular des Deutschen Archäologischen Instituts verwendet werden. Mit diesem wurde der Hauptanteil der Fundobjekte mit URIs angereichert. Der URI <http://archwort.dainst.org/de/term/1910> führt zur Webseite des Vokabulars, in dem die Ansprache „Gewandnadel“ angezeigt wird. In Zeile drei wird mit „select“ die Variable bestimmt, nach der in der Abfrage gesucht werden soll. Die Wahl des Wortes hat hier keinerlei Einfluss auf die Abfrage, es erscheint im Resultat als Überschrift über den Ergebnissen. Die nächste Zeile leitet mit dem Schlüsselwort „where“ die Abfrage ein. Hier wird ein Triple angegeben, welches die Variable beinhalten kann, nach der gesucht wird. In diesem Fall wird nach der Variablen als Subjekt gesucht. Das Prädikat ist „P2 has type“ nach dem CIDOC CRM-Vokabular und das Objekt der Aussage ist die Gewandnadel als URI. Es sollen demnach alle Triples ausgegeben werden, die den Typ Gewandnadel aufweisen können. Das Resultat der Abfrage erscheint unter der Eingabemaske (siehe Abbildung 87). In diesem Fall wurden 89 Ergebnisse gefunden, welche mit dieser Aussage übereinstimmen. Es handelt sich hier um die URIs

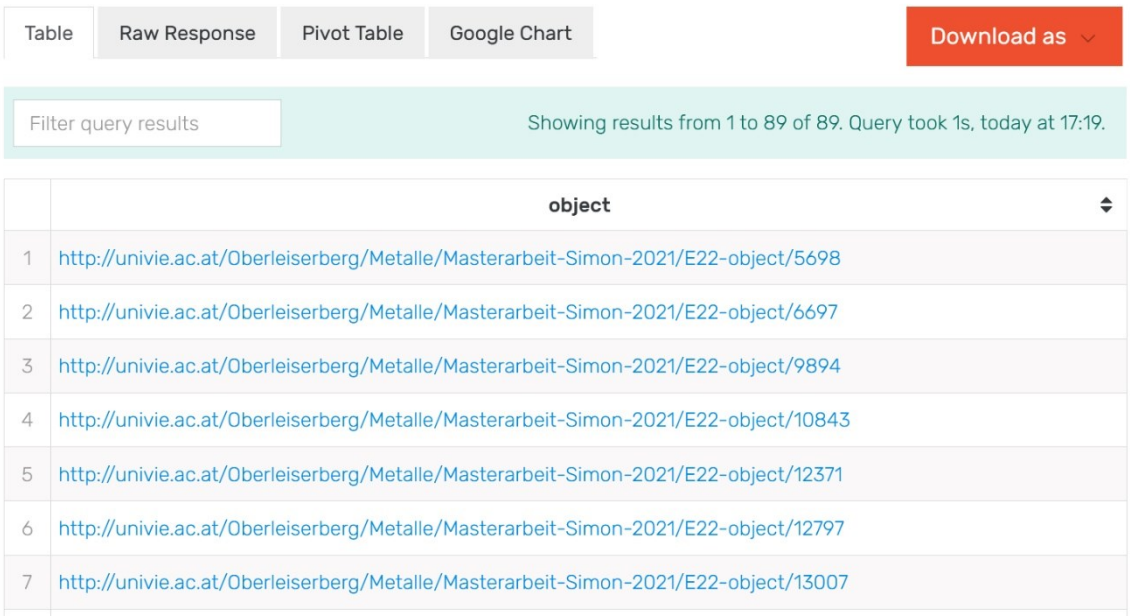


Abbildung 87 – Ergebnisse der SPARQL-Abfrage.

der Fundobjekte. Das Ergebnis kann in verschiedenen Formaten (z. B. JSON, XML, CSV) heruntergeladen werden.

8.3 Ergebnis

In der Case Study wurden Daten des Oberleiserbergs bearbeitet, welche in Form einer Access-Datenbank übernommen wurden. Durch den Export in ein Excel-Format konnten die Daten weiterverarbeitet werden. Ziel war, die Daten in ein RDF-Format zu konvertieren, um sie zukünftig für Linked Open Data verwenden zu können.

Der erste Schritt, um zu einer Modellierung der Daten zu kommen, die Datenbereinigung und Harmonisierung, konnte innerhalb von ca. vier Tagen bzw. 32 Stunden abgeschlossen werden. Die Bereinigung und Harmonisierung von Tippfehlern, uneinheitlichen Ansprachen von Fundkategorien, Datierungsangaben oder Maßangaben und das Hinzufügen einiger nur indirekt vorhandener Kategorien, wie z. B. dem Fundort, schuf eine geeignete Datenbasis für die Weiterverarbeitung. Dieser Schritt konnte bereits die weitere Arbeit mit den Daten verbessern.

Als nächsten Schritt wurden die Daten mit URIs aus den in Kapitel 6 vorgestellten Standards angereichert. Dies fördert die Eindeutigkeit der Daten und bringt den Nutzer bzw. die Nutzerin dazu, sich tiefgehend mit den Daten zu beschäftigen. Dieser Schritt konnte in ca. zwei Tagen bzw. 16 Stunden abgeschlossen werden.

Um mit diesen Daten in einem nächsten Schritt ein Datenmodell in der Ontologie CIDOC CRM anzufertigen, war sehr viel aufwändiger. Zunächst musste sich sehr eingehend mit der Methodik, den Regeln und Beispielen der Ontologie vertraut gemacht und danach diese auf die Daten angewandt werden. Die Schwierigkeit lag darin, die Richtigkeit des Datenmodells zu verifizieren. Die Beispiele, welche in der Literatur von archäologischen Forschungsprojekten, von Publikationen oder auf der CIDOC CRM-Webseite zu finden sind, konnten nicht unverändert übernommen werden. Denn jedes Datenset weist andere Anforderungen an ein Datenmodell auf. Aufgrund der Zeit, die in diesem Schritt anfänglich aufgewendet werden musste, um sich mit der Ontologie auseinanderzusetzen, dauerte er in etwa elf Tage bzw. 88 Stunden.

Nachdem das Datenmodell von einem CIDOC CRM-Experten begutachtet wurde, konnte der nächste Schritt begonnen werden. Die Konvertierung der Daten anhand des erstellten Datenmodells konnte nicht gänzlich ohne Hilfestellung eines Experten durchgeführt werden. Modellierungsdetails für RDF-Daten, wie der Umgang mit Labels oder URIs, können oft nur in der Anwendung erlernt werden. Um Fehler zu vermeiden und die Arbeitsprozesse kennenzulernen, ist es ratsam, die anfängliche Modellierung in Karma gemeinsam mit einem oder einer Expert*in

durchzuführen. Es wurde nicht das komplette Datenmodell umgesetzt, da dies in den meisten Fällen sehr aufwändig und nicht zweckmäßig ist. Vor allem bei der Aufarbeitung von Altdaten muss entschieden werden, welche Daten in ein RDF-Format gebracht werden sollen. Als Beispiel kann hier die Beschreibung der Auffindungssituation der Fundobjekte gebracht werden. Oft werden bei Ausgrabungen die stratigraphischen Einheiten dokumentiert und finden sich in den Daten über das Fundobjekt wieder. Die Dokumentation im Fall des Oberleiserbergs wurde mittels x- und y- Achsen, Niveautiefe und Objekten durchgeführt. Dies bedeutet einen erheblichen Mehraufwand in der Modellierung eines Datenmodells. Nachdem die wichtigsten Komponenten des Datenmodells gewählt wurden, wurde das Modell mit dem Online-Integrationstool Karma in das RDF-Format N-Quads konvertiert und als Repositorium in die Online-Graphendatenbank GraphDB überspielt. Für diesen Schritt wurden zusammen in etwa vier Tage bzw. 32 Stunden aufgebracht.

Die Visualisierung und die Abfrage der Daten in GraphDB sind nützliche Werkzeuge, um die Daten weiter zu verarbeiten. Dabei zeigte sich, dass die Visualisierung hilfreich ist, die Daten zum einen darstellen zu können und weiters auch das Datenmodell überprüfen zu können. Die Abfragen in SPARQL sind technisch herausfordernd. Komplizierte Abfragen müssen von Experten durchgeführt werden. Es dauert ca. eine halbe Stunde, um sich in der Online-Datenbank zurechtzufinden.

Zusammenfassend lässt sich feststellen, dass der Schritt der Modellierung, also die Bereinigung der Daten, die Anreicherung der Daten mit URIs und die Erstellung des Datenmodells mit CIDOC CRM von Archäolog*innen ohne besondere IT-Kenntnisse ohne große Probleme durchgeführt werden kann. Die Konvertierung der Daten in ein RDF-Format verlangt jedoch Hilfe eines Experten bzw. eine Einführung für die erste Modellierung. Die Lernkurve ist jedoch sehr steil und nach anfänglicher Schwierigkeit ist das Modellieren in Karma durchaus für Personen ohne Programmierkenntnisse zu bewältigen. Der Umgang mit der Online-Graphendatenbank GraphDB ist benutzerfreundlich und ein guter Kontrollmechanismus der Modellierung in Karma. Um SPARQL-Abfragen durchführen zu können, ist das eingehende Beschäftigen mit der Materie vorausgesetzt. Für komplexe Abfragen müssen Anwender*innen Zeit aufwenden, um die Abfragesprache zu erlernen oder Expert*innen hinzuziehen. Die Modellierung der Daten konnte in etwa innerhalb von ca. drei Wochen bzw. 168 Stunden durchgeführt werden. Dabei sollte jedoch bedacht werden, dass zuvor bereits Recherchen für die Datenbereinigung, passende URIs und CIDOC CRM durchgeführt worden sind.

9 Diskussion

Ein Großteil archäologischer Daten ist heterogen gestaltet, da unterschiedliche Standards und Vokabulare verwendet werden. Aus diesem Grund ist ihre Wiederverwendung für neue Forschungen sowie die Integration und Interoperabilität mit anderen Daten nur schwer möglich. Semantische Technologien können Lösungen für dieses Problem archäologischer Daten bieten. In dieser Arbeit wurde untersucht, wie Archäolog*innen ohne besondere IT-Kenntnisse archäologische Daten mit Standards versehen, zu einer Ontologie modellieren und in ein Linked Open Data Format konvertieren können und ab welchem Zeitpunkt hierbei eine Unterstützung durch Expert*innen notwendig ist. Weiters wurden die Vor- und Nachteile semantischer Technologien im Bereich der Archäologie untersucht.

Dafür wurden die grundlegenden Methoden erläutert, Anwendungsbeispiele archäologischer Forschungsprojekte vorgestellt und die Durchführung der Konvertierung eines Datensets in ein Linked Data-Format anhand der Case Study schrittweise dargestellt. Die Umsetzung der Konvertierung beinhaltete die Bereinigung der Daten, die Anreicherung der Inhalte mit breit anerkannten Standards und Vokabularen, die Modellierung des Datensets mit der Ontologie CIDOC CRM und die Konvertierung in das RDF-Format N-Quads über das Integrationstool Karma. Im Anschluss wurde das Datenset mithilfe der Online-Graphdatenbank GraphDB visualisiert und mit der Abfragesprache SPARQL abgefragt. Der gesamte Zeitaufwand der Konvertierung von 381 Fundobjekten der Funddatenbank des Oberleiserbergs kann in etwa auf drei Wochen, bzw. 168 Stunden, angegeben werden. In dieser Schätzung ist die Zeit enthalten, die es benötigt, sich die Ontologie CIDOC CRM in Grundzügen anzueignen, nicht jedoch die Zeit, um grundlegende Konzepte semantischer Technologien zu verinnerlichen, Daten inhaltlich zu bearbeiten (wie z.B. die Datierungen) oder passende Vokabulare oder Ressourcen von URIs zu recherchieren. Dabei ist zu bedenken, dass sich diese Schätzung auf den ersten Versuch einer Modellierung mit CIDOC CRM und den Umgang mit Konvertierungswerkzeugen bezieht.

Mit Routine in Datenmodellierung mit CIDOC CRM und Konvertierung in Karma kann diese Zeit erheblich verkürzt werden, vermutlich auf ca. ein Drittel der Zeit, eine Woche bzw. 56 Stunden. Die Zeitersparnis ergibt sich dabei vor allem durch die Modellierung mit CIDOC CRM, da hier alle benötigten Klassen und Eigenschaften bereits bewusst wären und durch die Konvertierung in Karma, da die Funktionen des Werkzeugs und die gängigen Modellierungspraktiken bekannt wären. In der Case Study wurden 381 Fundobjekte aus Buntmetall bearbeitet, welche von einer Fundstelle stammen. Viele Kriterien in den Daten, wie die Fundstelle, der Besitz oder das Material sind ident. Dies führt zu einer schnelleren Harmonisierung und Zuordnung von URIs in diesen Kategorien als bei anderen Datenbanken. Bei Datenbanken mit mehr Einträgen würde

sich der Aufwand der Modellierung in ein Datenmodell und die Konvertierung in ein RDF-Format nicht verändern, da die Anzahl der Spalten gleichbleibt. Die Zuordnung der URIs für mehr Fundobjekte würde jedoch einen Mehraufwand bedeuten. Ein noch höherer Aufwand ist bei Daten zu erwarten, die sich in ihrer Ansprache, dem Material, den Fundstellen oder der Datierung stark unterscheiden, denn dann muss eine höhere Anzahl an unterschiedlichen URIs aus den Online-Ressourcen herausgesucht werden.

Die Anwendung der Case Study zeigte, dass die Schritte von der Bereinigung der Daten bis zur Erstellung des Datenmodells ohne große Probleme durchführbar waren. Die Evaluierung des Datenmodells, die erste Anwendung der Konvertierung in ein RDF-Format und vor allem die Abfrage der Daten mittels SPARQL bedurfte jedoch der Unterstützung eines Fachexperten. Mithilfe dieser Methode konnten die heterogenen Daten mit Standards und Vokabularen versehen und in ein Format konvertiert werden, welches den Linked Data Empfehlungen (siehe Kapitel 3.2) entspricht und wiederverwendbar und integrierbar mit anderen konvertierten Daten gemacht werden. Nach Vollendung dieser Arbeit wird ein Repositorium auf Zenodo[887] angelegt werden, um die Daten gemeinsam mit dieser Arbeit zu veröffentlichen und ihnen einen Identifikator in Form einer DOI zuzuweisen. So können die Daten im Rahmen anderer Forschungsfragen weiterverwendet werden. Im Folgenden sollen die Erkenntnisse über die Vorteile und Grenzen semantischer Technologien für die Archäologie dargelegt und im Rahmen der Case Study diskutiert werden.

Semantische Technologien sind mit Sicherheit im Stande, einige Herausforderungen und Probleme im Bereich des Datenmanagements zu lösen. Es wurde jedoch niemals als Allheilmittel verkauft und hat in breiten Feldern noch Verbesserungsbedarf.[888] Anhand der Case Study zeigten sich einige Vorteile. Die Daten wurden in ein maschinenlesbares Format konvertiert und können so auch von Maschinen gelesen und weiterverarbeitet werden. Schon T. Berners-Lee zeigte die Vorteile maschinenlesbarer semantischer Daten auf. Dokumente werden in einem gemeinsamen Rahmen wie RDF erschaffen und so wird es Computern ermöglicht, miteinander zu kommunizieren und zu arbeiten.[889,890]

Durch die Generierung von HTTP-URIs und deren Verwendung als Standards und Vokabulare können in den Daten interaktiv neue Konzepte und Informationen entdeckt werden. Nutzer*in-

[887] Siehe unter https://zenodo.org/, abgerufen am 21.10.2021.
[888] Isaksen 2011, 6.
[889] Berners-Lee – Fischetti 1999, 189–192.
[890] Berners-Lee u. a. 2001, 34.

nen können Informationen über die Daten nachsehen und finden dabei immer neue Informationen, Links, Konzepte und Datensätze, die nützlich sein können. Dieser Prozess bzw. Effekt wird auch als „Follow your nose"-Prinzip beschrieben.[891,892]

Die Daten sind in diesem Format integrierbar und interoperabel mit anderen Daten. Dies und die Veröffentlichung als Repositorium können Datensilos und heterogene Daten verhindern. Syntaktische und semantische Interoperabilität durch die Anwendung von RDF und der Ontologie CIDOC CRM ermöglicht es nun, die Daten mit anderen Daten der Bronzezeit zu verbinden und es können neue Erkenntnisse gewonnen werden. Auch T. Berners-Lee betont den Gewinn, der entsteht, wenn heterogene und gestreute Daten zusammengeführt werden.[893] Syntaktische und semantische Interoperabilität werden ebenfalls ermöglicht. Wenn die Daten in einer RDF-Repräsentation zur Verfügung stehen, können neue Datensets oder Eigenschaften einfacher hinzugefügt und unterschiedliche Daten integriert werden. RDF-Strukturen und Prädikate ermöglichen es ebenso, Synonyme zu Typen oder Konzepten hinzuzufügen. M. K. Bergmann sieht hier eine der wichtigsten Fähigkeiten des Semantic Web, welche für die Archäologie zweifellos von Bedeutung ist.[894]

Im Semantic Web kann mittels Inferenz Wissen erlangt werden, welches durch logische Schlussfolgerungen gewonnen wurde. Mit Hilfe von definierten Eigenschaften der Daten, ihrer Beziehungen und Ontologien können neue Aussagen aus den Daten geschöpft werden, die in dieser Form gar nicht eingegeben wurden, sondern nur auf den Schlussfolgerungen beruhen.[895,896] Als Beispiel für Inferenz in den Daten der Case Study kann die Datierung dienen. Durch das Modellieren der Produktion zu den drei Datierungsangaben in den Daten „Kulturelle Zuordnung", „Datierung" und „Absolute Datierung" (z. B. Mittlere Urnenfelderkultur bis Jüngere Urnenfelderkultur, Späte Bronzezeit und „-1100" bis „-950") können Informationen, wie dass die Mittlere bis Jüngere Urnenfelderkultur in die Späte Bronzezeit fällt und diese wiederum jedenfalls im Zeitraum zwischen „-1100" und „-950" anhält, indirekt herausgelesen werden. Diese Information mag Archäolog*innen unwichtig erscheinen, jedoch könnten aus größeren Datenmengen interessante Schlussfolgerungen zutage kommen, die ansonsten verborgen bleiben. Es könnten beispielsweise Untersuchungen zu bestimmten Nadeltypen, ihren Datierungen und ihrem Vorkommen in Grä-

[891] W3C World Wide Web Consortium 2011, siehe "Follow your nose".
[892] Berners-Lee 2006, 1.
[893] Berners-Lee u. a. 2001, 37.
[894] Bergman 2009, 4 f.
[895] Allemang – Hendler 2008, 114.
[896] Berners-Lee u. a. 2001, 35.

bern mit anthropologischen Untersuchungen zu den Leichenbränden und weiteren Untersuchungen mit archäometallurgischen Ergebnissen zu verschmolzenen Nadeln verlinkt werden und dadurch neue Erkenntnisse gewonnen werden.

Die kreativen Möglichkeiten an Fragestellungen sind grenzenlos. Die Integrität der Daten wird erhöht durch die Bereinigung und den Einsatz von Standards und Vokabularen. Dabei werden Inkonsistenzen und Fehlerquellen, wie verschiedene Begriffe für die Ansprache oder die Datierung der Daten aufgedeckt. Durch die Modellierung mit CIDOC CRM erhöht sich die Beschäftigung mit den Daten enorm und fehlende Daten, wie beispielsweise die explizite Erwähnung der Fundstelle des Oberleiserbergs in den Daten, können ergänzt werden.[897] Kontraindiktionen können leichter aufgedeckt werden und neue Forschungsfragen entstehen.[898]

Die Daten sind durch die Veröffentlichung als Repositorium auf Zenodo offen zugänglich und können von anderen wiederverwendet werden. Die Verwendung geschieht oft in unvorhersehbarer Weise und eröffnet neue Möglichkeiten. Meist haben andere Personen kreative Ideen und eine neue Perspektive, wie die Daten noch verwendet werden können. Das berühmte Zitat von Rufus Pollock drückt dies folgendermaßen aus: „*The Best Thing to Do With Your Data Will be Thought of By Someone Else.*“[899,900]

Ein weiterer positiver Aspekt der Anwendung der breit anerkannten Ontologie CIDOC CRM kann für die Langzeitarchivierung von Daten gesehen werden. Die Ontologie ermöglicht es, die Bedeutung der Datensätze mit in die Daten einzubauen. Dies bedeutet, dass die Daten sich selbst beschreiben können und selbst in einigen Jahrzehnten von unbeteiligten Personen so interpretiert werden können, wie es Datenurheber*innen beabsichtigen.

Die Kontextualisierung, die durch zufällige Verlinkungen zu den Daten geschehen kann, sowie die Dokumentenzerlegung, bei welcher nur Teile des Datensets gefunden bzw. wiederverwendet werden, wären in diesem Fall besser möglich, wenn die Daten Teil der Linked Open Data Cloud wären. Dies ist jedoch für ein Datenset in dieser Größenordnung nicht möglich. Die auf der CAA 2021 gestartete Special Interest Group (SIG) „Semantics and LOUD in Archaeology“[901] setzt sich daher und aufgrund der vielen Repositorien, die immer schwieriger zu überblicken sind, für den

[897] Tao 2010, 330.
[898] Geser 2016, 16.
[899] Pollock 2012, Folie 26.
[900] Siehe auch einen Twitter-Eintrag von Rufus Pollock vom 20.04.2017: https://twitter.com/rufuspollock/status/855054231153369090, abgerufen am 21.10.2021.
[901] Siehe S2 der CAA (Computer Applications & Quantitative Methods in Archaeology) 2021. Hic sunt dracones – Improving knowledge exchange in the Semantic Web with Linked Open and FAIR data (Florian Thiery, Martina Trognitz und Ethan Gruber) unter https://2021.caaconference.org/sessions/, abgerufen am 21.10.2021.

Aufbau einer Archaeological Linked Data Cloud ein.[902] Dies könnte ein großer Meilenstein für archäologische Datensets sein, welche mit semantischen Technologien aufbereitet werden.

Die größten Hürden bei der Anwendung semantischer Technologien bilden im Allgemeinen die Probleme bei der breiten Adaptierung, fehlende Kosten-Nutzen-Analysen, die geringen Möglichkeiten, seine Daten zu anderen Daten zu verlinken und die hohe Einstiegshürde. Diese erkannten S. Haustein und J. Pleumann bereits 2002[903], L. Isaksen 2011[904] und G. Geser 2016[905]. Im Folgenden wird auf die Schwierigkeiten und Hürden eingegangen, welche bei der Konvertierung des Datensets zu Daten des Oberleiserbergs beobachtet wurden. Zu Beginn soll die erfahrene Einstiegsbarriere evaluiert werden. Die Autorin dieser Arbeit kann nach den persönlichen Erfahrungen die Annahmen aus der Literatur, dass die Einstiegsbarriere hoch ist, bestätigen.[906] Zum einen ist der Großteil der Literatur für Forschende verfasst, die bereits als Experten auf diesem Gebiet angesehen werden können. Dies erschwert den Einstieg für Interessierte. Zum anderen sind die Anwendungsbeispiele, welche sich in der Literatur finden, meist nur ein Ausschnitt der Anwendung und entwickelte Werkzeuge für die Konvertierung in ein RDF-Format, welche den Einstieg für Nicht-Experten erleichtern sollen, sind meist wenig benutzerfreundlich gestaltet und erfordern jedenfalls eine Einführung durch Expert*innen. Generell sind die grundlegenden semantischen Technologien nicht einfach anzuwenden für Nicht-Expert*innen. Dies erschwert die breite Adaptierung von Linked Open Data. Als Vergleich zur Adaptierung kann das World Wide Web genommen werden. Würden alle Nutzer*innen nach wie vor Webseitenprogrammierkenntnisse aufweisen müssen, um Informationen aus dem Web gewinnen zu können, wäre der große Erfolg wohl ausgeblieben.

Schwierigkeiten bei der Modellierung mit der Ontologie CIDOC CRM werden vor allem bei der Evaluierung des eigenen Modells erfahren. Es gibt wenige Möglichkeiten, dessen ontologische Korrektheit zu überprüfen, außer sich an Expert*innen zu wenden. Das Modell anhand vorhandener Beispiele zu überprüfen, ist aufgrund der Unterschiedlichkeit der Daten meist eher problematisch. Als besonders herausfordernd wird die Modellierung der Datierungen empfunden. In der Case Study wird die Datierung des Fundobjekts mit drei Komponenten in den Daten beschrieben. Der Einsatz der Klassen *E4 Period* und *E52 Time-Span* an der richtigen Stelle sowie die Unterscheidung der Eigenschaften, welche zur Datierung dienen[907], ist nicht trivial und die Definitionen des CIDOC CRM müssen sehr genau analysiert werden.

[902] Thiery 2019.
[903] Haustein – Pleumann 2002, 448 f.
[904] Isaksen 2011, 5 f.
[905] Geser 2016, 43.
[906] Siehe z. B. Stuart 2012, 176; Binding u. a. 2015, 1 f.; Isaksen 2011, 66.
[907] Siehe beispielsweise die Eigenschaften P79-82, P86, P89, P173-176 und P182-185.

Eine weitere Schwierigkeit, welche in der Bearbeitung der Daten beobachtet wurde, wird aktuell ebenfalls in Fachkreisen adressiert.[908] Archäologische Daten weisen oft eine Unsicherheit auf in Bezug auf die Ansprache, Datierung oder Lokalisierung von Fundobjekten sowie Unsicherheiten der Stratigraphie oder Festlegung auf die Ansprache von Befunden auf. Dies wird in den Daten meist mithilfe einer Beschreibung oder mit einem Fragezeichen gekennzeichnet. Die Unsicherheit in der Datenmodellierung auszudrücken ist zurzeit noch nicht ausreichend möglich. Dieser Umstand kann Probleme bei der Auswertung der Daten verursachen. Um transparente Daten zu produzieren und zu veröffentlichen, ist eine Lösung für die Eingabe der Unsicherheit besonders wichtig. In der Case Study wird versucht, dies mit jeweils einer eigenen Spalte wie „Datierung Unsicherheit" oder „Ansprache Unsicherheit" und der booleschen Auswahl „Ja" oder „Nein" auszudrücken. Es konnte jedoch keine zufriedenstellende Modellierung hierfür gefunden werden.

Die Modellierung der absoluten Zahlen der Datierungen der Fundobjekte mit positiven und negativen Zahlenangaben (z. B. -1100) sind mit der Klasse *E61 Time Primitive* vorgenommen worden. Um die Zahlen besser für Visualisierungen, Statistiken und verschiedenen Rechenoperationen verwenden zu können, wurden die Angaben „BC" bzw. „AD" weggelassen. Dies könnte jedoch eventuell die Eindeutigkeit der Zahlen bezüglich der verwendeten Kalenderart beeinträchtigen.

Bei der Aufarbeitung von Daten, welche von Altgrabungen stammen, wie die Daten des Oberleiserbergs, treten oft Unklarheiten mit der Modellierung der CIDOC CRM-Erweiterung CRMarchaeo auf. Das CRMarchaeo ist geeignet, um stratigraphische Einheiten zu modellieren. Im Fall der Case Study wurden die Fundobjekte nicht nach stratigraphischen Einheiten dokumentiert, sondern nach einem Achsensystem. Die Funde wurden nach x- und y-Koordinaten des Grabungsschnittes sowie der Tiefe von der Humusoberfläche aus eingemessen und dokumentiert. Dieser Dokumentationsweise konnte sich in der Modellierung nur angenähert werden (siehe Kapitel 8.2.1.3). Bei einer weiteren Bearbeitung der Daten könnten diese Koordinaten in ein Landeskoordinatensystem umgerechnet werden, dies war jedoch im Rahmen dieser Arbeit nicht vorgesehen.

Neben der Anwendung der Ontologie sind die Serialisierungen von RDF für Nicht-Expert*innen schwierig zu verstehen und anzuwenden. Selbst wenn Werkzeuge zur Erstellung von RDF-Formaten verwendet werden, muss ein gewisses Grundverständnis hierfür vorhanden sein, um mit den Daten weiterarbeiten zu können. Die Serialisierungen N-Triples und N-Quads sind am geeignetsten für Einsteiger*innen, da sie weder Präfixe noch eine Hierarchie verwenden. SPARQL-

[908] Siehe die S2 der CAA 2021 (Computer Applications & Quantitative Methods in Archaeology). Hic sunt dracones – Improving knowledge exchange in the Semantic Web with Linked Open and FAIR data (Florian Thiery, Martina Trognitz und Ethan Gruber) unter https://2021.caaconference.org/sessions/, abgerufen am 21.10.2021.

Abfragen müssen jedoch in der Turtle-Serialisierung formuliert werden. Dies birgt ein großes Hindernis für Nicht-Expert*innen, um Daten abfragen zu können. Datenabfragen sind jedoch ein wichtiger Teil der Linked Data Anwendung. Um komplexe Abfragen durchführen zu können, muss eine lange Lernzeit in Kauf genommen werden.

Im Zuge der Case Study wurde das Integrationstool Karma verwendet. Die Durchführung der Modellierung der einzelnen Spalten anhand des erstellten Datenmodells kann nach einer gewissen Eingewöhnungszeit bewältigt werden. In diesem Schritt kommen jedoch die in Expertenkreisen als „gute Praxis" verstandenen Modellierungseinzelheiten hinzu, welche in der Literatur oder Tutorien nicht ausreichend adressiert werden. Dies ist beispielsweise die Praxis, Konzepte wie die Ansprache eines Fundobjekts, mit einem *rdfs:label* (aus dem RDF-Schema) und einer URI zu versehen. Die Schritte, um die Daten zuerst zu importieren und nach der Modellierung als RDF-Format zu exportieren, sind ohne ausführliche Anleitung nicht ersichtlich. Es müssen einige Zwischenschritte durchgeführt werden, welche in Karma selbst nicht angeführt sind und nur über eine Einführung durch Expert*innen bekannt werden.

Diese Arbeit behandelt vorrangig die Aspekte von semantischen Technologien auf den Bereich der Archäologie und deren schrittweise Anwendung auf ein Datenset, deshalb können keine Ergebnisse ausgegeben werden, die auf der Auswertung von mehreren Datensets und deren Integration beruhen. Die Vorteile, die sich mit der Verbindung von großen Datenmengen ergeben würden und großflächige Abfragen von Daten können daher nicht aufgezeigt werden und sind Themen für weiterführende Forschungsfragen. Diese Arbeit richtet sich in erster Linie an Datenherausgeber*innen, nicht jedoch an die weitere Verarbeitung von Linked (Open) Data.

10 Fazit und Ausblick

Die Ergebnisse der Case Study zeigen, dass zurzeit ein Zusammenspiel von archäologischen und semantischen Expert*innen notwendig ist, um semantische Technologien in der Archäologie anwenden zu können. Der Mehrwert von semantisch aufgearbeiteten archäologischen Daten ergibt sich in der Möglichkeit, sie mit anderen Daten integrieren und vergleichen, für neue Forschungsfragen wiederverwenden und sie offen zugänglich in einem Repositorium in einem maschinenlesbaren Format zur Verfügung stellen zu können. Der Mehraufwand gestaltet sich bei einer ersten Anwendung semantischer Technologien hoch, da eine Vielzahl an neuen Kenntnissen, wie die Anwendung einer Ontologie und die grundlegenden Konzepte von Linked Data, gewonnen werden müssen. Bei weiteren Anwendungen können diese Kenntnisse bereits angewandt werden und der zeitliche Aufwand reduziert sich. Die weitere Verwendung von Linked Data und die Vorteile sind jedoch noch nicht ausreichend in der Praxis ersichtlich. Diese Arbeit soll Archäolog*innen die grundlegenden Konzepte semantischer Technologien, ihre Vorteile und die schrittweise Konvertierung von archäologischen Daten in Linked Data näherbringen und so einen Teil zur weiteren Verbreitung dieser Technologien beitragen.

Das Datenset zu Funden des Oberleiserbergs, welches im Zuge dieser Arbeit in ein Linked Data-Format konvertiert wurde, kann für weitergehende Forschung im Bereich der Bronzezeit verwendet werden. Das Forschungsprojekt Open Research Data for Prehistoric Mining Archaeology (ORD4Mining-Archaeo)[909] veröffentlicht Daten aus dem DACH-Projekt Prehistoric copper production in the eastern and central Alps[910] als FAIRe Daten. Diese Daten zeigen Forschungsergebnisse zur Minenaktivität, Technologietransfer und Handelsbeziehungen in der Bronzezeit und frühen Eisenzeit in den Alpen und beinhalten Daten zu Fundstellen prähistorischer Minen- und Metallurgieaktivitäten. Sie wurden mithilfe der Ontologie CIDOC CRM modelliert und beinhalten Informationen zu Materialuntersuchungen, Typologien, Landschaftsarchäologie, naturwissenschaftlicher Analyseergebnisse und Archäometallurgie. Am Oberleiserberg wurde durch Funde von Gussformen, Buntmetallkügelchen, Schlacke und weitere Metallverarbeitungsrückstände indiziert, dass dort jedenfalls Metallverarbeitung betrieben wurde. Eine Verbindung der Daten zu den in dieser Arbeit konvertierten Daten zu urnenfelderzeitlichen Metallfunden und Daten aus dem ORD4Mining-Archaeo-Projekt könnte weitere Erkenntnisse zur Metallverarbeitung und Handelsbeziehungen der späten Bronzezeit am Oberleiserberg erbringen.

[909] Siehe unter https://www.uibk.ac.at/projects/ord4mining-archaeo/, abgerufen am 21.10.2021.
[910] Siehe unter https://www.uibk.ac.at/himat/projekte/dach/index.html.de, abgerufen am 21.10.2021.

Eine Empfehlung für weitere Forschung zu semantischen Technologien sind nützliche und vor allem benutzerfreundliche Werkzeuge für Archäolog*innen zu entwickeln, welche in den Alltag integriert werden können und die Aneignung von semantischen Technologien auf ein mögliches Minimum reduzieren können, wenn nicht komplett auszuklammern. Denn der große Durchbruch von Linked Open Data kann nur dann erreicht werden, wenn möglichst viele Nutzer*innen einen Beitrag dazu leisten können und die Daten für die Öffentlichkeit auch benutzerfreundlich nutzbar sind. Eine Möglichkeit ist die Implementierung von semantischen Technologien und einer Ontologie im Hintergrund von Software, die im Feld bei Ausgrabungen und bei der Aufarbeitung der Dokumentation und der Fundobjekte verwendet werden kann. Würde eine solche Software für die Dokumentation von Ausgrabungen verwendet werden und beim Bundesdenkmalamt mit Vorschreibungen zur Langzeitarchivierung und Wiederverwendung der Daten zusammenkommen können, würde dies unendlich viele Optionen für zukünftige Forschungsfragen ermöglichen.

11 Zusammenfassung/Abstract

11.1 Zusammenfassung

Archäologische Dokumentation wird vorwiegend digital angefertigt und generiert jedes Jahr eine große Menge an Daten. Durch die Verwendung unterschiedlicher Standards und Vokabulare sind archäologische Daten höchst heterogen. Dies verhindert nachhaltiges Wiederverwenden der Daten und die Ausschöpfung des enormen Potentials dieser Datenmengen.

Das Ziel der vorliegenden Arbeit ist es zu beantworten, wie archäologische Daten durch semantische Technologien integrierbar, interoperabel und wiederverwendbar werden können, welche Schritte dafür notwendig sind, ab wann Unterstützung durch semantische Expert*innen angeraten wird und welche Vorteile sie der Archäologie bieten können.

Um dies beantworten zu können, wurden Datenmanagementpläne sowie semantische Technologien und ihre Anwendung in der Archäologie untersucht. Zur Darstellung der schrittweisen Konvertierung in ein Linked (Open) Data-Format wurde eine Datenbank zu urnenfelderzeitlichen Funden des Oberleiserbergs verwendet. Spezifisch wurden die Daten bereinigt, mit Standards und Vokabularen angereichert, ein Datenmodell mithilfe der im kulturhistorischen Bereich beliebten Ontologie CIDOC CRM erstellt und die Daten in ein Linked (Open) Data-Format umgewandelt.

Die Anwendung zeigt die einzelnen Schritte im Detail und soll helfen, archäologische Daten wiederverwendbar zur Verfügung stellen zu können. Eine sinnvolle Vorgehensweise hierbei ist die Zusammenarbeit archäologischer und semantischer Expert*innen.

Die weitere Verwendung von Linked (Open) Data birgt nach wie vor Herausforderungen für archäologische Daten, doch je mehr offene, integrierbare archäologische Daten veröffentlicht werden, desto größer kann das Netz an Daten in der Archäologie wachsen und neue ungeahnte Erkenntnisse eröffnen. Diese Arbeit soll einen Beitrag dazu leisten, Bewusstsein für die Herausforderungen des Datenmanagement zu bilden und Linked (Open) Data in der Archäologie zu etablieren.

11.2 Abstract

Archaeological documentation is mostly produced digitally and generates a large amount of data. Due to the use of different standards and vocabularies, archaeological data are extremely heterogeneous. This prevents sustainable reuse of the data and the exhaustion of the enormous potential of this data.

The aim of this thesis is to answer how archaeological data can be integrated and made interoperable and reusable through semantic technologies, which steps are necessary for this, at which point of the process support from semantic experts is recommended and which benefits they can offer archaeological data.

To be able to answer this, data management as well as semantic technologies and their application in archaeology were examined. To illustrate the step-by-step conversion into a linked (open) data format, a database of finds from the Urnfield period found at the Oberleiserberg was used. Specifically, the data was cleaned up, enriched with standards, a data model was created using the CIDOC CRM ontology and the data was converted into a linked (open) data format.

The application shows the individual steps and is intended to help make archaeological data reusable. A sensible approach here is the collaboration of archaeological and semantic experts.

The further use of linked (open) data still harbours challenges, but the more open, integrable archaeological data is published, the larger the network of data in archaeology can grow and open undreamt-of knowledge. This work is intended to help raise awareness of the challenges of data management and to establish linked (open) data in archaeology.

12 Literaturverzeichnis

ALCTS Association for Library Collections and Technical Services 2000

ALCTS Association for Library Collections and Technical Services. Committee on Cataloging: Description and Access. CC:DA, Task Force on Metadata - Final Report June 16, 2000, <https://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html> (abgerufen am 22.10.2021).

Allemang – Hendler 2008

D. Allemang – J. Hendler, Semantic Web for the Working Ontologist. Effective Modeling in RDFS and OWL (Amsterdam 2008).

Allemang – Hendler 2012

D. Allemang – J. Hendler, Semantic Web for the Working Ontologist. Effective Modeling in RDFS and OWL. Second Edition (Waltham 2012).

Allison 2008

P. Allison, Dealing with Legacy Data - an introduction, Internet Archaeology 24, 2008, <https://doi.org/10.11141/ia.24.8> (abgerufen am 22.10.2021).

Antoniou u. a. 2012

G. Antoniou – P. Groth – F. van Harmelen – R. Hoekstra, A Semantic Web Primer. Third edition (Cambridge Massachusetts 2012).

ARIADNE Plus 2019

ARIADNE Plus, ARIADNE Plus - Advanced Research Infrastructure for Archaeological Dataset Networking in Europe, 2019, <https://ariadne-infrastructure.eu/wp-content/uploads/2019/02/ARIADNEPlus_synopsis-short.pdf> (abgerufen am 22.10.2021).

Aspöck u. a. 2015

E. Aspöck – K. Kopetzky – B. Horejs – M. Bietak – M. Kucera – W. Neubauer, A Puzzle in 4D Digital Preservation and Reconstruction of an Egyptian Palace, in: G. Guidi – J. C. Torres – R. Scopigno – H. Graf – F. Remondino – P. Brunet – J. Barceló – L. Duranti – S. Hazan (Hrsg.), 2015 Digital Heritage. Conference 28. Sept.-02. Oct. 2015, Granada, Spain. Proceedings of the 2015 Digital Heritage International Congress, Volume 2 (Granada 2015) 675–678.

Aspöck 2019

E. Aspöck, ARIADNE and ARIADNEplus in Austria, in: J. D. Richards – F. Niccolucci (Hrsg.), The ARIADNE Impact (Budapest 2019) 51–62.

Aspöck u. a. 2020

E. Aspöck – G. Hiebel – K. Kopetzky – M. Durco, A Puzzle in 4D: Archiving Digital and

Analogue Resources of the Austrian Excavations at Tell el-Daba, Egypt, in: E. Aspöck – S. Štuhec – K. Kopetzky – M. Kucera (Hrsg.), Old Excavation Data - What Can We Do? Proceedings of the Workshop held at the 10th ICAANE in Vienna, April 2016, Oriental and European Archaeology 16 (Wien 2020) 79–100.

ASPÖCK 2020

E. Aspöck, Old Excavation Data – What Can We Do? An Introduction, in: E. Aspöck – S. Štuhec – K. Kopetzky – M. Kucera (Hrsg.), Old Excavation Data - What Can We Do? Proceedings of the Workshop held at the 10th ICAANE in Vienna, April 2016, Oriental and European Archaeology 16 (Wien 2020) 11–23.

AUER U. A. 2007

S. Auer – C. Bizer – G. Kobilarov – J. Lehmann – R. Cyganiak – Z. Ives, DBpedia: A Nucleus for a Web of Open Data, in: K. Aberer – K.-S. Choi – N. Noy – D. Allemang – K.-I. Lee – L. Nixon – J. Golbeck – P. Mika – D. Maynard – R. Mizoguchi – G. Schreiber – P. Cudré-Mauroux (Hrsg.), The Semantic Web. 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007. Proceedings, Lecture Notes in Computer Science 4825 (Berlin - Heidelberg 2007) 722–735.

AUER U. A. 2009

S. Auer – S. Dietzold – J. Lehmann – S. Hellmann – D. Aumueller, Triplify – Light-Weight Linked Data Publication from Relational Databases, in: J. Quemada – G. León (Hrsg.), WWW '09: Proceedings of the 18th international conference on World Wide Web, Madrid, Spain, April 2009 (New York 2009) 621–630.

AUER 2014

S. Auer, Introduction to LOD2, in: S. Auer – V. Bryl – S. Tramp (Hrsg.), Linked Open Data - Creating Knowledge Out of Interlinked Data. Results of the LOD2 Project, Lecture Notes in Computer Science 8661 (Cham 2014) 1–17.

BAINES – BROPHY 2006

A. Baines – K. Brophy, What's another word for Thesaurus? Data standards and classifying the past, in: P. Daly – T. L. Evans (Hrsg.), Digital Archaeology. Bridging Method and Theory (London 2006) 210-222.

BALL 2014

A. Ball, How to License Research Data, DCC How-to Guides. Edinburgh: Digital Curation

Centre, 2014, <https://www.dcc.ac.uk/guidance/how-guides/license-research-data> (abgerufen am 22.10.2021).

BAUER U. A. 2015

B. Bauer – A. Ferus – J. Gorraiz – V. Gründhammer – C. Gumpenberger – N. Maly – J. M. Mühlegger – J. L. Preza – B. Sánchez Solís – N. Schmidt – C. Steineder, Forschende und Ihre Daten. Ergebnisse einer österreichweiten Befragung – Report 2015 - Executive Summary und Empfehlungen, Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 68, 3-4, 2015, 566-580, <https://doi.org/10.31263/voebm.v68i3.1298> (abgerufen am 22.10.2021).

BAUER – FERUS 2018

B. Bauer – A. Ferus, Österreichische Repositorien in OpenDOAR und re3data.org: Die Entwicklung der Infrastruktur für den Grünen Weg des Open Access, und mehr, Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 71, 1, 2018, 70–86, <https://doi.org/10.31263/voebm.v71i1.2037> (abgerufen am 22.10.2021).

BAUER – KALTENBÖCK 2012

F. Bauer – M. Kaltenböck, Linked Open Data: The Essentials. A Quick Start Guide for Decision Makers (Wien 2012).

BAUER – KALTENBÖCK 2016

F. Bauer – M. Kaltenböck, Linked Open Data: The Essentials. The Climate Knowledge Brokering Edition. Second Edition (Wien 2016).

BECKETT 2014

D. Beckett, RDF 1.1 N-Triples. A line-based syntax for an RDF graph. W3C Recommendation 25 February 2014, <https://www.w3.org/TR/2014/REC-n-triples-20140225/> (abgerufen am 22.10.2021).

BECKETT U. A. 2014

D. Beckett – T. Berners-Lee – E. Prud'hommeaux – G. Carothers, RDF 1.1 Turtle. Terse RDF Triple Language. W3C Recommendation 25 February 2014, <https://www.w3.org/TR/2014/REC-turtle-20140225/> (abgerufen am 22.10.2021).

BEER U. A. 2014

N. Beer – K. Herold – W. Kolbmann – T. Kollatz – M. Romanello – S. Rose – N. O. Walkowski, Interdisciplinary Interoperability, DARIAH-DE Working Papers 3, Göttingen: DARIAH-DE 2014, <https://d-nb.info/1048629198/34> (abgerufen am 22.10.2021).

BEKIARI U. A. 2021

C. Bekiari – G. Bruseker – M. Doerr – C.-E. Ore – S. Stead – A. Velios, Definition of the CIDOC Conceptual Reference Model. Version 7.1.1., April 2021, <http://www.cidoc-crm.org/sites/default/files/cidoc_crm_v.7.1.1_0.pdf> (abgerufen am 22.10.2021).

BELSHE U. A. 2015

M. Belshe – R. Peon – E. M. Thomson, Hypertext Transfer Protocol Version 2 (HTTP/2). Internet Engineering Task Force, RFC 7540, 2015, <https://tools.ietf.org/rfc/rfc7540.txt> (abgerufen am 22.10.2021).

BERGMAN 2009

M. K. Bergman, Advantages and Myths of RDF. A 10th Birthday Salute to RDF's Role in Powering Data Interoperability, AI3 Adaptive Information, Adaptive Innovation, Adaptive Infrastructure, 2009, <https://www.mkbergman.com/483/advantages-and-myths-of-rdf/> (abgerufen am 22.10.2021).

BERNERS-LEE 1998A

T. Berners-Lee, Semantic Web Road map. W3C Design Issue, 1998, <https://www.w3.org/DesignIssues/Semantic.html> (abgerufen am 22.10.2021).

BERNERS-LEE 1998B

T. Berners-Lee, What the Semantic Web is not. What the Semantic Web can represent. W3C Design Issue, 1998, <https://www.w3.org/DesignIssues/RDFnot.html> (abgerufen am 22.10.2021).

BERNERS-LEE 1999

T. Berners-Lee, Web Architecture from 50,000 feet. W3C Design Issue, 1999 <https://www.w3.org/DesignIssues/Architecture.html> (abgerufen am 22.10.2021).

BERNERS-LEE U. A. 2001

T. Berners-Lee – J. Hendler – O. Lassila, The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American 5, 284, 2001, 34–38, <https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf> (abgerufen am 22.10.2021).

BERNERS-LEE U. A. 2005

T. Berners-Lee – R. Fielding – L. Masinter, Uniform Resource Identifier (URI): Generic Syntax. Network Working Group, RFC 3986, 2005, <https://www.ietf.org/rfc/rfc3986.txt> (abgerufen am 22.10.2021).

BERNERS-LEE 2006

T. Berners-Lee, Linked Data. W3C Design Issue, 2006, <https://www.w3.org/DesignIssues/LinkedData.html> (abgerufen am 22.10.2021).

BERNERS-LEE – FISCHETTI 1999

T. Berners-Lee – M. Fischetti, Weaving the Web. The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor (New York 1999).

BINDING U. A. 2008

C. Binding – K. May – D. Tudhope, Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction Via the CIDOC CRM, in: B. Christensen-Dalsgaard – D. Castelli – B. A. Jurik – J. Lippincott (Hrsg.), Research and Advanced Technology for Digital Libraries. 12th European conference, ECDL 2008, Aarhus, Denmark, September 14-19, 2008. Proceedings, Lecture Notes in Computer Science 5173 (Berlin - Heidelberg 2008) 280–290.

BINDING 2010

C. Binding, Implementing Archaeological Time Periods Using CIDOC CRM and SKOS, in: L. Aroyo – G. Antoniou – E. Hyvönen – A. t. Teje – H. Stuckenschmidt – L. Cabral – T. Tudorache (Hrsg.), The Semantic Web: Research and Applications. 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 - June 3, 2010. Proceedings, Part 1, Lecture Notes in Computer Science 6088 (Berlin - Heidelberg 2010) 273–287.

BINDING 2011

C. Binding, STELLAR Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources, Description of applications and templates (STELLAR manual), 2011, <https://hypermedia.research.southwales.ac.uk/kos/stellar/stellar-applications/> (abgerufen am 22.10.2021).

BINDING U. A. 2015

C. Binding – M. Charno – S. Jeffrey – K. May – D. Tudhope, Template Based Semantic Integration: From Legacy Archaeological Datasets to Linked Data, International Journal on Semantic Web and Information Systems 11, 1, 2015, 1–29, <https://doi.org/10.4018/IJSWIS.2015010101> (abgerufen am 22.10.2021).

BIZER U. A. 2009

C. Bizer – T. Heath – T. Berners-Lee, Linked Data - The Story So Far, International Journal on Semantic Web and Information Systems 5, 3, 2009, 1–22, <https://doi.org/10.4018/jswis.2009081901> (abgerufen am 22.10.2021).

BLUMESBERGER 2018

S. Blumesberger, Neue Anforderungen – viele offene Fragen. Zu den vielfältigen Rollen von Repositorien am Beispiel der UB Wien, Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 71, 1, 2018, 149–161, <https://doi.org/10.31263/voebm.v71i1.2003 > (abgerufen am 22.10.2021).

BLUMESBERGER 2020

S. Blumesberger, Repositorien als Tools für ein umfassendes Forschungsdatenmanagement: Am Beispiel von PHAIDRA an der Universitätsbibliothek Wien, Bibliothek Forschung und Praxis 44, 3, 2020, 503–511, <https://doi.org/10.1515/bfp-2020-2026> (abgerufen am 22.10.2021).

BORST 1997

W. N. Borst, Construction of Engineering Ontologies for Knowledge Sharing and Reuse (PhD Thesis University of Twente 1997).

BRAY U. A. 2008

T. Bray – J. Paoli – C. M. Spielberg-McQueen – E. Maler – F. Yergeau, Extensible Markup Language (XML) 1.0 (Fifth Edition). W3C Recommendation 26 November 2008, <https://www.w3.org/TR/2008/REC-xml-20081126/> (abgerufen am 22.10.2021).

BRAY U. A. 2009

T. Bray – D. Hollander – A. Layman – R. Tobin – H. S. Thompson, Namespaces in XML 1.0 (Third Edition). W3C Recommendation 8 December 2009, <https://www.w3.org/TR/2009/REC-xml-names-20091208/> (abgerufen am 22.10.2021).

BRICKLEY U. A. 2014

D. Brickley – R. V. Guha – B. McBride, RDF Schema 1.1. W3C Recommendation 25 February 2014, <https://www.w3.org/TR/2014/REC-rdf-schema-20140225/> (abgerufen am 22.10.2021).

BUNDESDENKMALAMT 2018

Bundesdenkmalamt, Richtlinien für archäologische Maßnahmen. 5. Fassung - 1. Jänner 2018, Wien 2018, <https://bda.gv.at/fileadmin/Medien/bda.gv.at/

SERVICE_RECHT_DOWNLOAD/Richtlinien2020_Download.pdf> (abgerufen am 22.10.2021).

CANNY 2015

N. Canny, Opening Access to Archaeology, Archäologische Informationen 38, 2015, 21–29, <https://doi.org/10.11588/ai.2015.1.26109> (abgerufen am 22.10.2021).

CAROTHERS 2014

G. Carothers, RDF 1.1 N-Quads. A line-based syntax for RDF datasets. W3C Recommendation 25 February 2014, <https://www.w3.org/TR/2014/REC-n-quads-20140225/> (abgerufen am 22.10.2021).

CHEN U. A. 2008

D. Chen – G. Doumeingts – F. Vernadat, Architectures for enterprise integration and interoperability: Past, present and future, Computers in Industry 59, 7, 2008, 647–659, <https://doi.org/10.1016/j.compind.2007.12.016 > (abgerufen am 22.10.2021).

COX – LITTLE 2020

S. Cox – C. Little, Time Ontology in OWL. W3C Candidate Recommendation 26 March 2020, <https://www.w3.org/TR/2020/CR-owl-time-20200326/> (abgerufen am 22.10.2021).

CRIPPS U. A. 2004

P. Cripps – A. Greenhalgh – D. Fellows – K. May – D. Robinson, Ontological Modelling of the work of the Centre for Archaeology. CIDOC CRM Technical Paper, 2004, <http://old.cidoc-crm.org/docs/Ontological_Modelling_Project_Report_%20Sep2004.pdf > (abgerufen am 22.10.2021).

CROFTS U. A. 2011

N. Crofts – M. Doerr – T. Gill – S. Stead – M. Stiff, Definition of the CIDOC Conceptual Reference Model. Version 5.0.4., November 2011, <http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_5.0.4.pdf> (abgerufen am 22.10.2021).

CYGANIAK U. A. 2014

R. Cyganiak – D. Wood – M. Lanthaler, RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014, <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/> (abgerufen am 22.10.2021).

CZAJLIK U. A. (HRSG.) 2019

Z. Czajlik – M. Črešnar – M. Doneus – M. Fera – A. Hellmuth Kramberger – M. Mele (Hrsg.), Researching Archaeological Landscapes Across Borders. Strategies, Methods and Decisions for the 21st Century (Graz 2019).

DEICKE 2016

A. Deicke, CIDOC CRM-based modeling of archaeological catalogue data, in: E. W. de Luca – P. Bianchini (Hrsg.), DHC 2016. Digital Humanities and Digital Curation. Proceedings of the First Workshop on Digital Humanities and Digital Curation, Goettingen, Germany, November 22, 2016. CEUR Workshop Proceedings 1764, 2016, Session 2 Digital Curation, Article Number 1, 1-6, <http://ceur-ws.org/Vol-1764/4.pdf > (abgerufen am 22.10.2021).

DENARD 2009

H. Denard, Die Londoner Charta. Für die computergestützte Visualisierung von kulturellem Erbe, Entwurf 2.1, 2009, <https://www.londoncharter.org/fileadmin/templates/main/docs/london_charter_2_1_de.pdf> (abgerufen am 22.10.2021).

DEUTZ U. A. 2020

D. B. Deutz – M. C. H. Buss – J. S. Hansen – K. K. Hansen – K. G. Kjelmann – A. V. Larsen – E. Vlachos – K. F. Holmstrand, How to FAIR: a website to guide researchers on making research data more FAIR, Zenodo, 2020, <https://doi.org/10.5281/zenodo.3712065> (abgerufen am 22.10.2021).

DMSG

DMSG, Denkmalschutzgesetz - DMSG. Bundesgesetz betreffend den Schutz von Denkmalen wegen ihrer geschichtlichen, künstlerischen oder sonstigen kulturellen Bedeutung, StF: BGBl. Nr. 533/1923, 2021, <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009184> (abgerufen am 22.10.2021).

DOERR 2003

M. Doerr, The CIDOC Conceptual Reference Module: An ontological approach to semantic interoperability of metadata, AI Magazine, 24, 3, 2003, 75–92.

DOERR 2009

M. Doerr, Ontologies for Cultural Heritage, in: S. Staab – R. Studer (Hrsg.), Handbook on Ontologies. Second Edition, International Handbooks on Information Systems (Berlin - Heidelberg 2009) 463–486.

DOERR U. A. 2016

M. Doerr – M. Theodoridou – E. Aspöck – A. Masur, Mapping Archaeological Databases to CIDOC CRM, in: S. Campana – R. Scopigno – G. Carpentiero – M. Cirillo (Hrsg.), CAA2015. Keep the Revolution going: Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology, 30 March - 02 April 2015, Siena, Italy (Oxford 2016) 443–451.

DOERR U. A. 2018A

M. Doerr – A. Felicetti – S. Hermon – G. Hiebel – A. Kritsotaki – A. Masur – K. May – P. Ronzino – W. Schmidle – M. Theodoridou – D. Tsiafaki – E. a. o. Christaki, Definition of the CRMarchaeo. An Extension of CIDOC CRM to support the archaeological excavation process. Version 1.4.7, November 2018, <http://www.cidoc-crm.org/crmarchaeo/sites/default/files/CRMarchaeo_v1.4.7.pdf> (abgerufen am 22.10.2021).

DOERR U. A. 2018B

M. Doerr – A. Felicetti – S. Hermon – G. Hiebel – A. Kritsotaki – A. Masur – K. May – P. Ronzino – W. Schmidle – M. Theodoridou – D. Tsiafaki – E. Christaki, Definition of the CRMarchaeo. An Extension of CIDOC CRM to support the archaeological excavation process. Version 1.4.4, January 2018, <http://www.cidoc-crm.org/crmarchaeo/sites/default/files/CRMarchaeo_v1.4.4.pdf> (abgerufen am 22.10.2021).

DOERR U. A. 2020A

M. Doerr – G. Bruseker – C. Bekiari – C. E. Ore – T. Velios – S. Stead, Definition of the CIDOC Conceptual Reference Model. Version 6.2.9., April 2020, <http://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM_v6.2.9%2030-4-2020%20.pdf> (abgerufen am 22.10.2021).

DOERR U. A. 2020B

M. Doerr – A. Kritsotaki – Y. Rousakis – G. Hiebel – M. Theodoridou, Definition of the CRMsci. An Extension of CIDOC-CRM to support scientific observation. Version 1.2.8, February 2020, <http://www.cidoc-crm.org/crmsci/sites/default/files/CRMsci%20v.1.2.8.pdf> (abgerufen am 22.10.2021).

DOERR U. A. 2020C

M. Doerr – R. Light – G. Hiebel, Implementing the CIDOC Conceptual Reference Model in RDF. Version 1.1, October 2020, <http://www.cidoc-crm.org/sites/default/files/issue%20443%20-%20Implementing%20CIDOC%20CRM%20in%20RDF%20v1.1.pdf> (abgerufen am 22.10.2021).

DOERR – IORIZZO 2008

M. Doerr – D. Iorizzo, The dream of a global knowledge network - A new approach, Journal on Computing and Cultural Heritage 1, 1, 2008, Article Number 5, 1-23, <https://doi.org/10.1145/1367080.1367085 > (abgerufen am 22.10.2021).

DÜRST – SUIGNARD 2005

M. Dürst – M. Suignard, Internationalized Resource Identifiers (IRIs). Network Working Group, RFC 3987, 2005, <https://www.ietf.org/rfc/rfc3987.txt> (abgerufen am 22.10.2021).

EFFINGER – BÜTTNER 2015

M. Effinger – A. Büttner, Open Access – Open Archaeology. Wissenschaft und Bibliothek als Dream-Team?, Archäologische Informationen 38, 2015, 73–82, <https://doi.org/10.11588/ai.2015.1.26114> (abgerufen am 22.10.2021).

EGGERT 2005

M. K. H. Eggert, Prähistorische Archäologie. Konzepte und Methoden. 2. Auflage (Tübingen 2005).

EIBL U. A. 2019

G. Eibl – B. Lutz – P. Parycek – S. Pawel, Rahmenbedingungen für Open Government Data Plattformen, 2019, <https://neu.ref.wien.gv.at/at.gv.wien.ref-live/documents/20189/68315/Rahmenbedingungen_f%C3%BCr_Open_Government_Data_Portale_1.3_fin.pdf/51921bf6-cc0a-49c0-bc96-20fc23dcdfba> (abgerufen am 22.10.2021).

EICHERT 2021

S. Eichert, Digital Mapping of Medieval Cemeteries: Case Studies from Austria and Czechia, Journal on Computing and Cultural Heritage 14, 1, 2021, Article 3, 1–15, <https://doi.org/10.1145/3406535 > (abgerufen am 22.10.2021).

EOSC 2017

EOSC, EOSC Declaration. European Open Science Cloud. New Research & Innovation Opportunities. Brussels, 26 October 2017, <https://eosc-portal.eu/sites/default/files/eosc_declaration.pdf> (abgerufen am 22.10.2021).

EUROPEAN COMMISSION 2017

European Commission, H2020 Programme, Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2, 2017,

<https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf> (abgerufen am 22.10.2021).

FANIEL U. A. 2013

I. Faniel – E. Kansa – S. Whitcher Kansa – J. Barrera-Gomez – E. Yakel, The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse, in: J. S. Downie – R. H. McDonald – T. W. Cole – R. Sanderson – F. Shipman (Hrsg.), JCDL '13. Proceedings of the 13th ACM/IEEE-CS joint Conference on Digital Libraries, Indianapolis, Indiana, USA, 22.07.2013 - 26.07.2013 (New York 2013) 295–304.

FELICETTI U. A. 2015

A. Felicetti – P. Gerth – C. Meghini – M. Theodoridou, Integrating Heterogeneous Coin Datasets in the Context of Archaeological Research, in: P. Ronzino (Hrsg.), EMF-CRM 2015. Extending, Mapping and Focusing the CRM 2015. Proceedings of the Workshop on Extending, Mapping and Focusing the CRM, Poznań, Poland, September 17, 2015, CEUR Workshop Proceedings 1656, 2015, 13–27, <http://ceur-ws.org/Vol-1656/paper2.pdf > (abgerufen am 22.10.2021).

FERNIE 2013

K. Fernie, CARARE: Connecting Archaeology and Architecture in Europeana, Uncommon Culture, Interviews and Projects, 2013, 85-92, <https://uncommonculture.org/ojs/index.php/UC/article/view/4753/3691> (abgerufen am 22.10.2021).

FIELDING – RESCHKE 2014

R. Fielding – J. Reschke, Hypertext Transfer Protocol (HTTP/1.1). Semantics and Content. Internet Engineering Task Force, RFC 7231, 2014, <https://www.rfc-editor.org/rfc/pdfrfc/rfc7231.txt.pdf> (abgerufen am 22.10.2021).

FILZWIESER – EICHERT 2020

R. Filzwieser – S. Eichert, Towards an Online Database for Archaeological Landscapes. Using the Web Based, Open Source Software OpenAtlas for the Acquisition, Analysis and Dissemination of Archaeological and Historical Data on a Landscape Basis, Heritage 3, 4, 2020, 1385–1401, <https://doi.org/10.3390/heritage3040077> (abgerufen am 22.10.2021).

FRIESINGER 1978

H. Friesinger, Oberleis, Fundberichte aus Österreich 16 (1977), 1978, 423–424.

GANDON – SCHREIBER 2014

F. Gandon – G. Schreiber, RDF 1.1 XML Syntax. W3C Recommendation 25 February 2014,

<https://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/> (abgerufen am 22.10.2021).

GEORGIS 2018

C. Georgis, CIDOC-CRM URIs namespace mechanism (303-redirect), 2018, <http://www.cidoc-crm.org/Resources/cidoc-crm-uris-namespace-mechanism-using-303-redirect-mechanism> (abgerufen am 22.10.2021).

GERTH 2017

P. Gerth, Spatial Cultural Heritage Data in the Linked Open Data Cloud: Datenintegration von räumlichen, musealen und archäologischen Daten anhand von Linked Open Data (LOD) Technologien (Masterarbeit Hochschule Anhalt, Fachbereich Architektur, Facility Management und Geoinformation 2017).

GESER 2014

G. Geser, Open Access to cultural heritage content. Results of the CreativeCH workshop on "Open Access, IPR and management of rights". Results of the CreativeCH workshop held within the Museum and the Web Florence 2014 pre-conference programme "Horizon2020 and Creative Europe vs Digital Heritage – A European Projects Crossover", Florence, 18th of February 2014, 2014,<https://creativech-toolkit.salzburgresearch.at/open-access-to-cultural-heritage-content/> (abgerufen am 22.10.2021).

GESER 2016

G. Geser, ARIADNE WP15 Study: Towards a Web of Archaeological Linked Open Data, Version 1.0, 2016 <http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/ARIADNE_archaeological_LOD_study_10-2016-1.pdf> (abgerufen am 22.10.2021).

GESER 2019

G. Geser, ARIADNEplus and community data repositories. Innovative solutions for sharing open archaeological data. Abstracts of the 19th International Conference on Cultural Heritage and New Technologies 2019 (CHNT 24, 2019), Vienna, Austria Nov. 4-6. 2019, 2019, <https://www.chnt.at/wp-content/uploads/ARIADNEplus-and-community-data-repositories.pdf> (abgerufen am 22.10.2021).

GEWIN 2016

V. Gewin, Data sharing: An open mind on open data, Nature 529, 7584, 2016, 117–119, <https://doi.org/10.1038/NJ7584-117A > (abgerufen am 22.10.2021).

GOLDENBERG U. A. (HRSG.) 2011

G. Goldenberg – U. Töchterle – K. Oeggl – A. Krenn-Leeb (Hrsg.), Forschungsprogramm HiMAT - Neues zur Bergbaugeschichte der Ostalpen, Archäologie Österreichs Spezial 4 (Wien 2011).

GRUBER 1993

T. R. Gruber, A translation approach to portable ontology specifications, Knowledge Acquisition 5, 2, 1993, 199–220, <https://doi.org/10.1006/knac.1993.1008> (abgerufen am 22.10.2021).

GUARINO U. A. 2009

N. Guarino – D. Oberle – S. Staab, What is an Ontology?, in: S. Staab – R. Studer (Hrsg.), Handbook on Ontologies. Second Edition, International Handbooks on Information Systems (Berlin - Heidelberg 2009) 1–17.

HAGMANN 2018

D. Hagmann, Überlegungen zur Nutzung von PHAIDRA als Repositorium für digitale archäologische Daten, Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 71, 1, 2018, 53–69, <https://doi.org/10.31263/voebm.v71i1.1974> (abgerufen am 22.10.2021).

HANNEMANN – KETT 2010

J. Hannemann – J. Kett, Linked Data for Libraries. World Library and Information Congress: 76TH IFLA General Conference and Assembly. 10-15 August 2010, Gothenburg, Sweden, 2010, <https://www.ifla.org/past-wlic/2010/149-hannemann-en.pdf> (abgerufen am 22.10.2021).

HARRIS – SEABORNE 2013

S. Harris – A. Seaborne, SPARQL 1.1 Query Language. W3C Recommendation 21 March 2013, <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/> (abgerufen am 22.10.2021).

HASAPIS U. A. 2015

P. Hasapis – D. Vergeti – A. Liapis – A. Ramfos – G. Flouris – K. Stefanidis – I. Chrysakis – Y. Roussakis – M. Meimaris – G. Papastefanatos – Y. Stavrakas – C. Pateritsas – T. Galani – P. Buneman – J. Cheney – S. Staworko – S. D. Viglas – J. Debattista – N. Friesen – L. Petit – S. Jupp – T. Burdett – R. Isele – K. Möller, A Toolset for Supporting Evolution and Preservation of Linked Data: the DIACHRON approach, Semantic Web Journal 1019, 2015, 1–9,

<http://www.semantic-web-journal.net/system/files/swj1019.pdf> (abgerufen am 22.10.2021).

HASLHOFER – KLAS 2010

B. Haslhofer – W. Klas, A survey of techniques for achieving metadata interoperability, ACM Computing Surveys 42, 2, 2010, Article Number 7, 1–37, <https://doi.org/10.1145/1667062.1667064> (abgerufen am 22.10.2021).

HAUSTEIN – PLEUMANN 2002

S. Haustein – J. Pleumann, Is Participation in the Semantic Web Too Difficult?, in: I. Horrocks – J. A. Hendler (Hrsg.), The Semantic Web - ISWC 2002. First International Semantic Web Conference, Sardinia, Italy, June 9-12, 2002, Proceedings, Lecture Notes in Computer Science 2342 (Berlin - New York 2002) 448–453.

HEATH – BIZER 2011

T. Heath – C. Bizer, Linked Data: Evolving the Web into a Global Data Space, Synthesis Lectures on the Semantic Web: Theory and Technology 1:1 (San Rafael 2011).

HEINRICH U. A. 2014

M. Heinrich – S. Jahn – F. Schäfer, Stakeholderanalyse zu Forschungsdaten in den Altertumswissenschaften, IANUS Forschungsdatenzentrum, 2014, <https://doi.org/10.13149/000.jah37w-q> (abgerufen am 22.10.2021).

HELLERSCHMID U. A. 2010

I. Hellerschmid – D. Kern – M. Lochner, Oberleiserberg – Stillfried – Thunau. Drei Höhensiedlungen der mitteldonauländischen Urnenfelderkultur im Vergleich, in: B. Gediga – W. Piotroski (Hrsg.), Rola głównych centrów kulturowych w kształtowaniu oblicza kulturowego Europy Środkowej we wczesnych okresach epoki żelaza (Rolle der wichtigen Kulturzentren in der Gestaltung des Kulturbildes Mitteleuropas in den frühen Perioden der Eisenzeit). Symposium Biskupin 23.-25.06.2008, Archäologisches Museum in Biskupin, Biskupiner Archäologische Arbeiten Nr. 8, Polnische Akademie der Wissenschaften, Abteilung Wroclaw, Arbeiten der Archäologischen Kommission Nr. 18 (Biskupin-Wrocław 2010) 283–297.

HIEBEL U. A. 2010

G. Hiebel – K. Hanke – I. Hayek, GIS Supported Implementation of Ontology Based Data Management for Multidisciplinary Research, in: P. Anreiter – G. Goldenberg – K. Hanke – R. Krause – W. Leitner – F. Mathis – K. Nicolussi – K. Oeggl – E. Pernicka – M. Prast – J. Schibler – I. Schneider – H. Stadler – T. Stöllner – G. Tompedi – P. Tropper (Hrsg.), Mining in European History and its Impact on Environment and Human Societies. Proceedings for

the 1st Mining in European History-Conference of the SFB-HIMAT, 12.–15. November 2009, Innsbruck (Innsbruck 2010) 405–409.

HIEBEL U. A. 2013

G. Hiebel – K. Hanke – I. Hayek, Methodology for CIDOC CRM based data integration with spatial data (corrected version), in: F. Contrearas – M. Farjas – F. J. Melero (Hrsg.), CAA 2010: Fusion of Cultures: Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain, April 2010, BAR International Series 2494 (Oxford 2013) 547-555.

HIEBEL U. A. 2014

G. Hiebel – M. Doerr – K. Hanke – A. Masur, How to Put Archaeological Geometric Data into Context? Representing Mining History Research with CIDOC CRM and Extensions, International Journal of Heritage in the Digital Era 3, 3, 2014, 557–577, <https://doi.org/10.1260/2047-4970.3.3.557> (abgerufen am 22.10.2021).

HIEBEL U. A. 2015A

G. Hiebel – M. Doerr – Eide, Øyvind, Theodoridou, Maria, CRMgeo: a Spatiotemporal Model. An Extension of CIDOC-CRM to link the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement. Version 1.2, September 2015, <http://www.cidoc-crm.org/crmgeo/sites/default/files/CRMgeo1_2.pdf> (abgerufen am 22.10.2021).

HIEBEL U. A. 2015B

G. Hiebel – M. Doerr – Ø. Eide, Integration of CIDOC CRM with OGC Standards to Model Spatial Information, in: A. Traviglia (Hrsg.), CAA 2013 - Across Space and Time. Papers from the 41st Conference on Computer Applications and Quantitative Methods in Archaeology. Perth, 25-28 March 2013 (Amsterdam 2015) 303–310.

HIEBEL U. A. 2017

G. Hiebel – K. Hanke – G. Goldenberg – M. Staudt – C. Grutsch, Information Integration in a Mining Landscape, Studies in Digital Heritage 1, 2, 2017, 692–699, <https://doi.org/10.14434/sdh.v1i2.23231> (abgerufen am 22.10.2021).

HIEBEL U. A. 2018

G. Hiebel – K. Hanke – G. Goldenberg – M. Staudt – M. Scherer-Windisch, Interdisziplinäre Analyse von Geländemodellen zur Auffindung von (prä)historischen Bergbauspuren. VGI, Österreichische Zeitschrift für Vermessung und Geoinformation 106, 3, 2018, 183–194.

HIEBEL U. A. 2021A

G. Hiebel – G. Goldenberg – C. Grutsch – K. Hanke – M. Staudt, FAIR data for prehistoric

mining archaeology, International Journal on Digital Libraries 22, 2020, 267-277, <https://doi.org/10.1007/s00799-020-00282-8> (abgerufen am 22.10.2021).

HIEBEL U. A. 2021B

G. Hiebel – E. Aspöck – K. Kopetzky, Ontological Modeling for Excavation Documentation and Virtual Reconstruction of an Ancient Egyptian Site, Journal on Computing and Cultural Heritage 14, 3, 2021, Article 32, 1–14, <https://doi.org/10.1145/3439735> (abgerufen am 22.10.2021).

HIEBEL – HANKE 2009

G. Hiebel – K. Hanke, Datenmodellierung und Systemarchitektur in einem GIS-gestützten multidisziplinären Forschungsprojekt, in: J. Strobl – T. Blaschke – G. Griesebner (Hrsg.), Angewandte Geoinformatik 2009. Beiträge zum 21. AGIT-Symposium Salzburg (Heidelberg 2009) 653–662.

HIEBEL – HANKE 2017

G. Hiebel – K. Hanke, Datenintegration zur Lokalisierung und Identifikation von Bergbauspuren, in: K. Hanke – T. Weinold (Hrsg.), 19. Internationale Geodätische Woche Obergurgl 2017 (Berlin 2017) 260–264.

HORTON – DCC 2016

L. Horton – DCC, Overview of UK Institution RDM Policies, Version 6, August 2016, Digital Curation Centre, <https://www.dcc.ac.uk/guidance/policy/institutional-data-policies> (abgerufen am 22.10.2021).

HYLAND U. A. 2014

B. Hyland – G. Atemezing – B. Villazón-Terrazas, Best Practices for Publishing Linked Data. W3C Working Group Note 09 January 2014, <https://www.w3.org/TR/2014/NOTE-ld-bp-20140109/> (abgerufen am 22.10.2021).

HYLAND – VILLAZÓN-TERRAZAS 2011

B. Hyland – B. Villazón-Terrazas, Linked Data Cookbook. Cookbook for Open Government Linked Data, W3C, 2011, <https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook> (abgerufen am 22.10.2021).

ICOMOS 2017

ICOMOS, The Seville Principles. International Principles of Virtual Archaeology. Ratified by the 19th ICOMOS General Assembly in New Dehli, December 2017, <http://sevilleprinciples.com/> (abgerufen am 22.10.2021).

ISAKSEN 2011

L. Isaksen, Archaeology and the Semantic Web (Thesis for the degree of Doctor of Philosophy University of Southampton 2011).

ISO 2010

ISO, International Organization for Standardization (2010), Information technology - Universal Coded Character Sets (UCS). ISO/IEC 10646:2010, ISO/IEC International Standard, Final Committee Draft, <http://unicode.org/L2/L2010/10038-fcd10646-main.pdf> (abgerufen am 22.10.2021).

ISO 2012

ISO, International Organization for Standardization (2012), Information and documentation - Digital object identifier system. ISO 26324:2012, <https://www.iso.org/standard/43506.html> (abgerufen am 22.10.2021).

ISO 2014

ISO, International Organization for Standardization (2014), Information and documentation - A reference ontology for the interchange of cultural heritage information. ISO 21127:2014, <https://www.iso.org/standard/57832.html>, (abgerufen am 22.10.2021).

ISO 2017A

ISO, International Organization for Standardization (2017), Information and documentation - International Standard Book Number (ISBN). ISO 2108:2017, <https://www.iso.org/standard/65483.html> (abgerufen am 22.10.2021).

ISO 2017B

ISO, International Organization for Standardization (2017), Information and documentation - The Dublin Core metadata element set - Part 1: Core elements. ISO 15836-1:2017, <https://www.iso.org/standard/71339.html> (abgerufen am 22.10.2021).

ISO 2019

ISO, International Organization for Standardization (2019), Information and documentation - The Dublin Core metadata element set - Part 2: DCMI Properties and classes. ISO 15836-2:2019 <https://www.iso.org/standard/71341.html> (abgerufen am 22.10.2021).

ISO 2020

ISO, International Organization for Standardization (2020), Information and documentation - International standard serial number (ISSN). ISO 3297:2020, <https://www.iso.org/standard/73846.html> (abgerufen am 22.10.2021).

JACOBS – HOLLAND 2007

P. Jacobs – C. Holland, Sharing Archaeological Data: The Distributed Archive Method, Near Eastern Archaeology 70, 4, 2007, 197–200, <https://doi.org/10.1086/NEA20361334> (abgerufen am 22.10.2021).

JEFFREY U. A. 2009

S. Jeffrey – J. Richards – F. Ciravegna – S. Waller – S. Chapman – Z. Zhang, The Archaeotools project: faceted classification and natural language processing in an archaeological context, Philosophical transactions of the Royal Society of London. Series A, Mathematical, physical and engineering sciences 367, 1897, 2009, 2507–2519, <https://doi.org/10.1098/rsta.2009.0038> (abgerufen am 22.10.2021).

JEFFREY 2012

S. Jeffrey, A new Digital Dark Age? Collaborative web tools, social media and long-term preservation, World Archaeology 44, 4, 2012, 553–570, <https://doi.org/10.1080/00438243.2012.737579> (abgerufen am 22.10.2021).

KAIER – EDIG 2020

C. Kaier – X. van Edig, Publizieren in wissenschaftlichen Zeitschriften, in: K. Lackner – L. Schilhan – C. Kaier (Hrsg.), Publikationsberatung an Universitäten. Ein Praxisleitfaden zum Aufbau publikationsunterstützender Services (Bielefeld 2020) 53–78.

KALTENBÖCK 2013

M. Kaltenböck, Linked Open Data Pilotprojekt Österreich (LOD PILOT AT), netidee Förderungen, Förderjahr 2013, Projekt Call 8, ProjectID 484, 2013, <https://www.netidee.at/sites/default/files/2017-08/LOD-Zusammenfassung_1.pdf> (abgerufen am 22.10.2021).

KÄNDLER 2020

U. Kändler, Open-Access-Finanzierung, in: K. Lackner – L. Schilhan – C. Kaier (Hrsg.), Publikationsberatung an Universitäten. Ein Praxisleitfaden zum Aufbau publikationsunterstützender Services (Bielefeld 2020) 181–202.

KANSA 2016

E. C. Kansa, Click Here to Save the Past, in: E. W. Averett – J. M. Gordon – D. B. Counts (Hrsg.), Mobilizing the Past for a Digital Future. The Potential of Digital Archaeology (Grand Forks 2016) 443–472.

KANSA – KANSA 2018

S. W. Kansa – E. C. Kansa, Data Beyond the Archive in Digital Archaeology. An Introduction to the Special Section, Advances in Archaeological Practice 6, 2, 2018, 89–92, <https://doi.org/10.1017/aap.2018.7> (abgerufen am 22.10.2021).

KANSA – WHITCHER KANSA 2013

E. C. Kansa – S. Whitcher Kansa, We All Know That a 14 Is a Sheep: Data Publication and Professionalism in Archaeological Communication, Journal of Eastern Mediterranean Archaeology and Heritage Studies 1, 1, 2013, 88–97.

KERN 1982

A. Kern, Oberleis, Fundberichte aus Österreich 20 (1981), 1982, 513–514.

KERN 1986

A. Kern, Oberleis, Fundberichte aus Österreich 23 (1984), 1986, 293.

KERN 1987

A. Kern, Die Urgeschichtlichen Funde vom Oberleiserberg, MG. Ernstbrunn. Die unstratifizierten Bestände aus Privatsammlungen, Bundes- und Landes- und Heimatmuseen, 1: Text, Katalog, (Dissertation Universität Wien 1987).

KERN 1988

A. Kern, Oberleis, Fundberichte aus Österreich 24/25 (1985/86), 1988, 294.

KERN 1990

A. Kern, Oberleis, Fundberichte aus Österreich 28 (1989), 1990, 230.

KERN 2003

D. Kern, Der Oberleiserberg in der Bronzezeit, in: M. Lochner (Hrsg.), Die Urnenfelderkultur in Österreich - Standort und Ausblick. Symposium. Eine Veranstaltung der Prähistorischen Kommission der Österreichischen Akademie der Wissenschaften, Wien 24.-25. April 2003, Broschüre zum Symposium (Wien 2003) 62–63.

KERN 2013

D. Kern, Ausgewählte Siedlungsbefunde aus der urnenfelderzeitlichen Siedlung auf dem Oberleiserberg bei Ernstbrunn, VB Korneuburg, NÖ, Archaeologia Austriaca 95/2011, 2013, 101–124, <https://doi.org/10.1553/archaeologia95s101> (abgerufen am 22.10.2021).

KINTIGH 2006

K. Kintigh, The Promise and Challenge of Archaeological Data Integration, American Antiquity 71, 3, 2006, 567–578, <https://doi.org/10.2307/40035365> (abgerufen am 22.10.2021).

KINTIGH 2009

K. W. Kintigh, The Challenge of Archaeological Data Integration, in: A. Velho – H. Kamermans (Hrsg.), Technology and Methodology for Archaeological Practice. Practical applications for the past reconstruction. Proceedings of the XV World Congress of the International Union for Prehistoric and Protohistoric Sciences, Lisbon, 4.-9. September 2006, BAR International Series 2029 (Oxford 2009) 81–86.

KINTIGH U. A. 2018

K. W. Kintigh – K. A. Spielmann – A. Brin – K. S. Candan – T. C. Clark – M. Peeples, Data Integration in the Service of Synthetic Research, Advances in archaeological practice 6, 1, 2018, 30–41, <https://doi.org/10.1017/aap.2017.33> (abgerufen am 22.10.2021).

KNOCHE 2017

M. Knoche, Die Idee der Bibliothek und ihre Zukunft (Göttingen 2017).

KROMP U. A. 2019

B. Kromp – M. Seissl – T. Zarka, Austrian Transistion to Open Access (AT2OA) – ein Überblick, Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 72, 1, 2019, 28–34, <https://doi.org/10.31263/voebm.v72i1.2274> (abgerufen am 22.10.2021).

KURT 2018

S. Kurt, Why do authors publish in predatory journals?, Learned Publishing 31, 2, 2018, 141–147, <https://doi.org/10.1002/leap.1150> (abgerufen am 22.10.2021).

LASSILA – SWICK 1999

O. Lassila – R. R. Swick, Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation 22 February 1999, <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/> (abgerufen am 22.10.2021).

LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES 2013

League of European Research Universities, LERU Roadmap for research data. LERU Research Data Working Group. Advice Paper No. 14, December 2013, <https://www.leru.org/files/LERU-Roadmap-for-Research-Data-Full-paper.pdf> (abgerufen am 22.10.2021).

LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES 2018

League of European Research Universities, Open Science and its role in universities: A roadmap for cultural change. Advice Paper No. 24, May 2018, <https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf> (abgerufen am 22.10.2021).

LEHMANN U. A. 2015

J. Lehmann – R. Isele – M. Jakob – A. Jentzsch – D. Kontokostas – P. N. Mendes – S. Hellmann – M. Morsey – P. van Kleef – S. Auer – C. Bizer, DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia, Semantic Web 6, 2, 2015, 167–195, <https://doi.org/10.3233/SW-140134> (abgerufen am 22.10.2021).

LOCHNER 2006

M. Lochner, Tongussformen für Ringe aus urnenfelderzeitlichen Siedlungen Niederösterreich, Archaeologia Austriaca 88/2004, 2006, 103–120, <https://doi.org/10.1553/archaeologia88s103> (abgerufen am 22.10.2021).

MACHADO U. A. 2019

L. M. O. Machado – R. R. Souza – M. Graça Simões, Semantic Web or Web of Data? A Diachronic Study (1999 to 2017) of the Publications of Tim Berners-Lee and the World Wide Web Consortium, Journal of the Association for Information Science and Technology 70, 7, 2019, 701–714, <https://doi.org/10.1002/asi.24111> (abgerufen am 22.10.2021).

MASUR U. A. 2014

A. Masur – K. May – G. Hiebel – E. Aspöck, Comparing and mapping archaeological excavation data from different recording systems for integration using ontologies, in: Museen der Stadt Wien - Stadtarchäologie (Hrsg.), Proceedings of the 18th International Conference on Cultural Heritage and New Technologies 2013 (CHNT 18, 2013). Vienna, Austria 11. - 13. November 2013 (Wien 2014) Chapter Infrastructures and services for sharing of archaeological documentation, Article Number 4, 1-11.

MCMANUS U. A. 2012

R. McManus – E. C. Nemec – D. S. Ferer – K. F. Gumpper, Suggested definitions for informatics terms: Interfacing, integration, and interoperability, American Journal of Health-System Pharmacy 69, 13, 2012, 1163–1165, <https://doi.org/10.2146/ajhp110612> (abgerufen am 22.10.2021).

MCMURRY U. A. 2017

J. A. McMurry – N. Juty – N. Blomberg – T. Burdett – T. Conlin – N. Conte – M. Courtot – J. Deck – M. Dumontier – D. K. Fellows – A. Gonzalez-Beltran – P. Gormanns – J. Grethe – J. Hastings – H. Hermjakob – J.-K. Hériché – J. C. Ison – R. C. Jimenez – S. Jupp – J. Kunze – C. Laibe – N. Le Novère – J. Malone – M. J. Martin – J. R. McEntyre – C. Morris – J. Muilu – W. Müller – P. Rocca-Serra – S.-A. Sansone – M. Sariyar – J. L. Snoep – N. J. Stanford – S. Soiland-Reyes – N. Swainston – N. Washington – A. R. Williams – S. Wimalaratne – L. Winfree – K. Wolstencroft – C. Goble – C. J. Mungall – M. A. Haendel – H. Parkinson,

Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data, PLoS biology 15, 6, 2017, e2001414: 1-18, <https://doi.org/10.1371/journal.pbio.2001414> (abgerufen am 22.10.2021).

MEGHINI U. A. 2017

C. Meghini – R. Scopigno – J. Richards – H. Wright – G. Geser – S. Cuy – J. Fihn – B. Fanini – H. Hollander – F. Niccolucci – A. Felicetti – P. Ronzino – F. Nurra – C. Papatheodorou – D. Gavrilis – M. Theodoridou – M. Doerr – D. Tudhope – C. Binding – A. Vlachidis, ARIADNE: A Research Infrastructure for Archaeology, Journal on Computing and Cultural Heritage 10, 3, 2017, Article 18, 1–27, <https://doi.org/10.1145/3064527> (abgerufen am 22.10.2021).

MEIMARIS U. A. 2015

M. Meimaris – G. Papastefanatos – C. Pateritsas, An Archiving System for Managing Evolution in the Data Web, in: J. Debattista – M. D'Aquin – C. Lange (Hrsg.), DIACHRON 2015 - Managing the Evolution and Preservation of the Data Web. Proceedings of the First DIACHRON Workshop on Managing the Evolution and Preservation of the Data Web, Portorož, Slovenia, May 31, 2015, CEUR Workshop Proceedings 1377, 2015, 16–21, <http://ceur-ws.org/Vol-1377/paper2.pdf> (abgerufen am 22.10.2021).

MITSCHA-MÄRHEIM – NISCHER-FALKENHOF 1929

H. Mitscha-Märheim – E. Nischer-Falkenhof, Der Oberleiserberg - Ein Zentrum vor- und frühgeschichtlicher Besiedlung. Bericht über die in den Jahren 1925 bis 1928 mit Unterstützung der Akademie der Wissenschaften in Wien durchgeführten Arbeiten, Mittelungen der Prähistorischen Kommission der Österreichischen Akademie der Wissenschaften 2, 5 (Wien 1929) 392-438.

MONTELIUS 1903

O. Montelius, Die älteren Kulturperioden im Orient und in Europa, 1: Die Methode (Stockholm 1903).

MOORE – RICHARDS 2015

R. Moore – J. D. Richards, Here Today, Gone Tomorrow: Open Access, Open Data and Digital Preservation, in: A. T. Wilson – B. Edwards (Hrsg.), Open Source Archaeology. Ethics and Practice (Warsaw - Berlin 2015) 30–43.

MORGAN – EVE 2012

C. Morgan – S. Eve, DIY and digital archaeology: what are you doing to participate?, World

Archaeology 44, 4, 2012, 521–537, <https://doi.org/10.1080/00438243.2012.741810> (abgerufen am 22.10.2021).

NICCOLUCCI – RICHARDS 2013

F. Niccolucci – J. D. Richards, ARIADNE: Advanced Research Infrastructures for Archaeological Dataset Networking in Europe, International Journal of Humanities and Arts Computing 7, 1-2, 2013, 70–88, <https://doi.org/10.3366/ijhac.2013.0082> (abgerufen am 22.10.2021).

NIKOLOVA 2015

L. Nikolova, What was Published is as Important as How it was Published, in: A. T. Wilson – B. Edwards (Hrsg.), Open Source Archaeology. Ethics and Practice (Warsaw - Berlin 2015) 83–91.

NOVOTNY – SEYFFERTITZ 2018

G. Novotny – T. Seyffertitz, Institutionelle Repositorien – Traum und Wirklichkeit, Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 71, 1, 2018, 87–106, <https://doi.org/10.31263/voebm.v71i1.2002> (abgerufen am 22.10.2021).

NOWACK 2009

B. Nowack, The Semantic Web Technology Stack (not a piece of cake…), Abbildung, 2009, <http://bnode.org/media/2009/07/08/semantic_web_technology_stack.png> (abgerufen am 22.10.2021).

OBERLÄNDER-TÂRNOVEANU 2005

I. Oberländer-Târnoveanu, Multilingual Access to Cultural Heritage Resources, Internet Archaeology 2005, 18, <https://doi.org/10.11141/ia.18.7> (abgerufen am 22.10.2021).

OLDMAN U. A. 2010

D. Oldman – M. Theodoridou – G. Samaritakis, Using Mapping Memory Manager (3M) with CIDOC CRM. Version 4g, 2010, <http://83.212.168.219/DariahCrete/sites/default/files/mapping_manual_version_4g.pdf> (abgerufen am 22.10.2021).

PAN 2009

J. Z. Pan, Resource Description Framework, in: S. Staab – R. Studer (Hrsg.), Handbook on Ontologies. Second Edition, International Handbooks on Information Systems (Berlin - Heidelberg 2009) 71–90.

PARAPONTIS 2012

I. Parapontis, Semantic Web Technology Stack and Linked Data Technology Stack, 2012,

<https://parapontis.files.wordpress.com/2012/12/
semantic-web-ld-stack3.jpg?w=564&h=378> (abgerufen am 22.10.2021).

PERRY – HERRING 2012

M. Perry – J. Herring, OGC GeoSPARQL - A Geographic Query Language for RDF Data, Open Geospatial Consortium Standard, Version 1.0, 2012, <http://www.opengis.net/doc/IS/geosparql/1.0> (abgerufen am 22.10.2021).

POLLOCK 2012

R. Pollock, Open Knowledge Foundation & Open Data. Opening Up Meeting Presentation, Open Data Conference, Cambridge, September 2012, <https://de.slideshare.net/OpeningUp/open-data-conference-rufus-pollock> (abgerufen am 22.10.2021).

PRINTZ 2014

S. Printz, Ein Beitrag zu Normen und Standards der Geoinformatik in der archäologischen Feldforschung (Dissertation Technische Universität Dortmund 2014).

RICHARDS 2002

J. D. Richards, Digital Preservation and Access, European Journal of Archaeology 5, 3, 2002, 343–366, <https://doi.org/10.1179/eja.2002.5.3.343> (abgerufen am 22.10.2021).

RICHARDS 2006

J. D. Richards, Archaeology, e-publication and the Semantic Web, Antiquity 80, 310, 2006, 970–979, <https://doi.org/10.1017/S0003598X00094552> (abgerufen am 22.10.2021).

RICHARDS 2009

J. D. Richards, From anarchy to good practice: the evolution of standards in archaeological computing, Archaeologia e Calcolatori 20, 2009, 27–35.

RICHARDS 2015

J. D. Richards, Ahead of the curve: adventures in e-publishing in Internet Archaeology, Archäologische Informationen 38, 2015, 63–71, <https://doi.org/10.11588/ai.2015.1.26113> (abgerufen am 22.10.2021).

RICHARDS U. A. 2021

J. D. Richards – U. Jakobsson – D. Novák – B. Štular – H. Wright, Digital Archiving in Archaeology: The State of the Art. Introduction, Internet Archaeology 58, 2021, <https://doi.org/10.11141/ia.58.23> (abgerufen am 22.10.2021).

RICHARDS – HARDMAN 2008

J. D. Richards – C. S. Hardman, Stepping back from the trench edge: an archaeological perspective on the devleopment of standards for recording and publication, in: M. Greengass – L. Hughes (Hrsg.), The Virtual Representation of the Past (Farnham - Surrey - Burlington 2008) 167–188.

RILEY 2017

J. Riley, Understanding metadata. What is metadata, and what is it for?, NISO Primer series, National Information Standards Organization (Baltimore 2017), <https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf> (abgerufen am 22.10.2021).

SALISBURY 2017

R. B. Salisbury, Open Access, archaeology, and the future, The European Archaeologist 51, Winter 2017, 12–14.

SÁNCHEZ SOLÍS U. A. 2020

B. Sánchez Solís – P. Marín-Arraiza – C. Stork – M. Andrae, Forschungsdatenmanagement und Publizieren von Forschungsdaten – Aufbau von Services am Beispiel der TU Wien, in: K. Lackner – L. Schilhan – C. Kaier (Hrsg.), Publikationsberatung an Universitäten. Ein Praxisleitfaden zum Aufbau publikationsunterstützender Services (Bielefeld 2020) 101–122.

SAUERMANN – CYGANIAK 2008

L. Sauermann – R. Cyganiak, Cool URIs for the Semantic Web. W3C Interest Group Note 03 December 2008, <https://www.w3.org/TR/2008/NOTE-cooluris-20081203/> (abgerufen am 22.10.2021).

SCHÄFER U. A. 2015

F. Schäfer – M. Heinrich – A. Sieverling – M. Trognitz – R. Förtsch – O. Dally – F. Fless, Forschungsrohdaten für die Altertumswissenschaften – eine kurze Bilanz der aktuellen Situation von Open Data in Deutschland, Archäologische Informationen 38, 2015, 125–136, <https://doi.org/10.11588/ai.2015.1.26156> (abgerufen am 22.10.2021).

SCHELOSKE 2012

M. Scheloske, Das Dilemma alternativer wissenschaftlicher Publikationsformate. Weshalb Open Access und Wissenschaftsblogs auch weiter Akzeptanzprobleme haben werden, Blogbeitrag, 2012, <https://www.wissenswerkstatt.net/2012/das-dilemma-alternativer-wissenschaftlicher-publikationsformate/> (abgerufen am 22.10.2021).

SCHILHAN – LACKNER 2020

L. Schilhan – K. Lackner, Qualitätssicherung und Predatory Publishing in der Publikationsberatung, in: K. Lackner – L. Schilhan – C. Kaier (Hrsg.), Publikationsberatung an Universitäten. Ein Praxisleitfaden zum Aufbau publikationsunterstützender Services (Bielefeld 2020) 163–180.

SCHOLL U. A. 2011

H. J. Scholl – H. Kubicek – R. Cimander, Interoperability, Enterprise Architectures, and IT Governance in Government, in: M. Janssen – H. J. Scholl – M. A. Wimmer – Y. Tan (Hrsg.), Electronic Government. 10th IFIP WG 8.5 International Conference, EGOV 2011 Delft, The Netherlands, August 28 - September 2, 2011. Proceedings, Lecture Notes in Computer Science 6846 (Berlin - Heidelberg 2011) 345–354.

SCHREIBER – RAIMOND 2014

G. Schreiber – Y. Raimond, RDF 1.1 Primer. W3C Working Group Note 24 June 2014, <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/> (abgerufen am 22.10.2021).

SCHRÖTER 2017

M. Schröter, Erfolgreich recherchieren – Altertumswissenschaften und Archäologie (Berlin - Boston 2017).

SHADBOLT – BERNERS-LEE 2006

N. Shadbolt – T. Berners-Lee, The Semantic Web Revisited, IEEE Intelligent Systems 21, 3, 2006, 96–101, <https://doi.org/10.1109/MIS.2006.62> (abgerufen am 22.10.2021).

SHAMASH 2016

K. Shamash, Article processing charges (APCs) and subscriptions. Monitoring open access costs, Jisc Report, 2016, <https://www.jisc.ac.uk/sites/default/files/apc-and-subscriptions-report.pdf> (abgerufen am 22.10.2021).

SIEGMUND 2013

F. Siegmund, Schnell, weltweit frei zugänglich und mit zusätzlichen Daten: Die Zeitschrift Archäologische Informationen erscheint im Open Access mit Early Views, Archäologische Informationen 36, 2013, 81–99, <https://doi.org/10.11588/ai.2013.0.15204> (abgerufen am 22.10.2021).

SIEGMUND – SCHERZLER 2015

F. Siegmund – D. Scherzler, Open Access und Open Data in der Ur- und Frühgeschichte:

Bestandsaufnahme und Ausblick, Archäologische Informationen 38, 2015, 11–19, <https://doi.org/10.11588/ai.2015.1.26108> (abgerufen am 22.10.2021).

SIEGMUND – SCHERZLER 2016

F. Siegmund – D. Scherzler, Frank Siegmund und Diane Scherzler zur Initiative "Open-Access 2020", "Die Archäologie muss sich einmischen". Die Initiative "Open Access 2020" und der Open-Access-Rummel im Frühling 2016. Ein Kommentar von PD Dr. Frank Siegmund und Diane Scherzler M. A., 2016, <https://www.dguf.de/ngo/kommentare/418-frank-siegmund-und-diane-scherzler-zur-initiative-open-access-2020> (abgerufen am 22.10.2021).

SIMON U. A. 2016

R. Simon – L. Isaksen – E. Barker – P. d. S. Cañamares, Peripleo: A Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time, Code4Lib Journal 31, 2016, <https://journal.code4lib.org/articles/11144> (abgerufen am 22.10.2021).

SIMON 2017

R. Simon, Pelagios Commons. Vortrag am Seminar von BAM-Austria (Bibliotheken, Archive, Museen) am 11.05.2017 in Wien: „Metadatennormierung heute für morgen: Open Access und Linked Open Data“, 2017, <https://www.univie.ac.at/voeb/fileadmin/Dateien/Bibliothekswesen/BAM/2017_Simon_Pelagios-Common_BAM_Wien.pdf> (abgerufen am 22.10.2021).

SPORNY U. A. 2020

M. Sporny – D. Longley – G. Kellogg – M. Lanthaler – P.-A. Champin – N. Lindstöm, JSON-LD 1.1. A JSON-based Serialization for Linked Data. W3C Recommendation 16 July 2020, <https://www.w3.org/TR/2020/REC-json-ld11-20200716/> (abgerufen am 22.10.2021).

STADLER – KAIER 2020

E. Stadler – C. Kaier, Publizieren von wissenschaftlichen Büchern, in: K. Lackner – L. Schilhan – C. Kaier (Hrsg.), Publikationsberatung an Universitäten. Ein Praxisleitfaden zum Aufbau publikationsunterstützender Services (Bielefeld 2020) 79–100.

STEAD 2007

S. Stead, AHRC ICT Methods Network Workshop. Space and Time: Methods in Geospatial Computing for Mapping the Past. Welsh E-Science Centre, Cardiff University, 10-11 October 2007, Session 3: Standards and Metadata, Standards and Metadata - Rapporteur's Report., 22-

24, 2007, <http://www.methodsnetwork.ac.uk/redist/pdf/act24report.pdf> (abgerufen am 22.10.2021).

STELLAR 2011

STELLAR, STELLAR Mapping and Extraction Guidelines (STELLAR Introduction), 2011, <https://hypermedia.research.southwales.ac.uk/kos/stellar/stellar-applications/> (abgerufen am 22.10.2021).

STIGLER – STEINER 2018

J. H. Stigler – E. Steiner, GAMS – An infrastructure for the long-term preservation and publication of research data from the Humanities, Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 71, 1, 2018, 207–216, <https://doi.org/10.31263/voebm.v71i1.1992> (abgerufen am 22.10.2021).

STUART 2012

D. Stuart, FOAF Within UK Academic Web Space: A Webometric Analysis of the Semantic Web, in: G. Widén – K. Holmberg, Social Information Research, Library and Information Science 5 (Bingley 2012) 173–191.

STUDER U. A. 1998

R. Studer – V. R. Benjamins – D. Fensel, Knowledge Engineering: Principles and methods, Data & Knowledge Engineering 25, 1, 1998, 161–197, <https://doi.org/10.1016/S0169-023X(97)00056-6> (abgerufen am 22.10.2021).

STUPPNER 1997

A. Stuppner, Der Oberleiserberg bei Ernstbrunn. Vorrömische Siedlung - germanischer Fürstensitz, in: H. Friesinger – F. Krinzinger (Hrsg.), Der römische Limes in Österreich. Führer zu den archäologischen Denkmälern (Wien 1997) 282–287.

STUPPNER 1999

A. Stuppner, Oberleis, Fundberichte aus Österreich 37 (1998), 1999, 833–836.

STUPPNER 2001

A. Stuppner, Oberleis, Fundberichte aus Österreich 39 (2000), 2001, 698–701.

STUPPNER 2003

A. Stuppner, Oberleis, Fundberichte aus Österreich 41 (2002), 2003, 693–695.

STUPPNER 2005

A. Stuppner, Oberleis, Fundberichte aus Österreich 43 (2004), 2005, 957–961.

STUPPNER 2006

A. Stuppner, Rund um den Oberleiserberg - Von der Urzeit bis zum Mittelalter. Archäologische Denkmale der Gemeinden Ernstbrunn und Niederleis (Ernstbrunn 2006).

STUPPNER 2007

A. Stuppner, Oberleis, Fundberichte aus Österreich 45 (2006), 2007, 718–719.

STUPPNER 2009

A. Stuppner, Oberleis, Fundberichte aus Österreich 47 (2008), 2009, 604–605.

STUPPNER 2010

A. Stuppner, Oberleis, Fundberichte aus Österreich 48 (2009), 2010, 461–462.

STUPPNER 2011

A. Stuppner, Oberleis, Fundberichte aus Österreich 49 (2010), 2011, 296–297.

STUPPNER 2013

A. Stuppner, Die ländliche Besiedlung im mittleren Donauraum von der Spätantike bis zum Frühmittelalter, Antiquité Tardive 21, 2013, 47–62, <https://doi.org/10.1484/J.AT.5.101403> (abgerufen am 22.10.2021).

TAO 2010

J. Tao, Adding Integrity Constraints to the Semantic Web for Instance Data Evaluation, in: P. F. Patel-Schneider – Y. Pan – P. Hitzler – P. Mika – L. Zhang – J. Z. Pan – I. Horrocks – B. Glimm (Hrsg.), The Semantic Web - ISWC 2010. 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010. Revised Selected Papers, Part II, Lecture Notes in Computer Science 6497 (Berlin 2010) 330–338.

TAUBERT 2016

N. Taubert, Open Access und digitale Publikation aus der Perspektive von Wissenschaftsverlagen, in: P. Weingart – N. Taubert (Hrsg.), Wissenschaftliches Publizieren. Zwischen Digitalisierung, Leistungsmessung, Ökonomisierung und medialer Beobachtung, Interdisziplinäre Arbeitsgruppen Forschungsberichte 38 (Berlin - Boston 2016) 75–102.

THIERY 2013

F. Thiery, Semantic Web und Linked Data: Generierung von Interoperabilität in archäologischen Fachdaten am Beispiel römischer Töpferstempel (Masterarbeit Fachhochschule Mainz 2013).

THIERY 2019

F. Thiery, Hic sunt dracones! Der digitale Wandel zur Archäologie 4.0 am Beispiel römischer

Münzen, 9. Workshop der AG CAA (AG CAA), Wilhelmshaven, Germany, Zenodo, 2019, <https://doi.org/10.5281/zenodo.3403058> (abgerufen am 22.10.2021).

THIERY U. A. 2019

F. Thiery – M. Trognitz – E. Gruber – D. Wigg-Wolf, Hic sunt dracones! The modern unknown Data Dragons, Research Squirrel Engineers - Squirrel Papers, Zenodo, 2019, <https://doi.org/10.5281/zenodo.3345715> (abgerufen am 22.10.2021).

THIERY – MEES 2017

F. Thiery – A. Mees, Das Labeling System: Erstellung kontrollierter Linked Open Data Vokabulare als Metadaten-Hub für archäologische Fachdatenbanken, 8. Workshop der AG CAA, Heidelberg, Germany, Zenodo, 2017, <https://doi.org/10.5281/zenodo.292554> (abgerufen am 22.10.2021).

TONTO U. A. 2015

Y. Tonto – G. Dogan – U. Al – O. Madran, Open Access Policies of Research Funders: The Case Study of the Austrian Science Fund (FWF), Zenodo, 2015, <https://doi.org/10.5281/zenodo.35616> (abgerufen am 22.10.2021).

TORGGLER – ANDRAE 2018

A. Torggler – M. Andrae, Aus dem Leben einer/s Repman – ein Bericht aus dem österreichischen „Netzwerk für RepositorienmanagerInnen“, Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 71, 1, 2018, 107–124, <https://doi.org/10.31263/voebm.v71i1.1986> (abgerufen am 22.10.2021).

TROGNITZ 2021

M. Trognitz, Saving Us from the Digital Dark Age: the Austrian perspective, Internet Archaeology 58, 2021, <https://doi.org/10.11141/ia.58.2> (abgerufen am 22.10.2021).

TROGNITZ – ĎURČO 2018

M. Trognitz – M. Ďurčo, One Schema to Rule them All. The Inner Workings of the Digital Archive ARCHE, Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 71, 1, 2018, 217–231, <https://doi.org/10.31263/voebm.v71i1.1979> (abgerufen am 22.10.2021).

TUDHOPE U. A. 2011

D. Tudhope – C. Binding – S. Jeffrey – K. May – A. Vlachidis, A STELLAR Role for Knowledge Organization Systems in Digital Archaeology, Bulletin of the American Society for Information Science and Technology 37, 4, 2011, 15–18, <https://doi.org/10.1002/bult.2011.1720370405 > (abgerufen am 22.10.2021).

TUDHOPE U. A. 2013

D. Tudhope – C. Binding – K. May – M. Charno, Pattern based mapping and extraction via CIDOC CRM, in: V. Alexiev – V. Ivanov – M. Grinberg (Hrsg.), CRMEX 2013. Practical Experiences with CIDOC CRM and its Extensions. Proceedings of the Workshop Practical Experiences with CIDOC CRM and its Extensions, Valetta, Malta, September 26, 2013, CEUR Workshop Proceedings 1117, 2013, 23–36, <http://ceur-ws.org/Vol-1117/paper3.pdf > (abgerufen am 22.10.2021).

ULEBERG U. A. 2020

E. Uleberg – M. Matsumoto – C. E. Ore – J. Kile-Vesik, Sustainable archaeological data in Norway. EAA 2020 Virtual, Abstract 1665, 2020, <https://submissions.e-a-a.org/eaa2020/repository/preview.php?Abstract=1665> (abgerufen am 22.10.2021).

VAKIL 2019

C. Vakil, Predatory journals. Authors and readers beware, Canadian Family Physician 65, 2, 2019, 92–94, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6515480/> (abgerufen am 22.10.2021).

W3C OWL WORKING GROUP 2012

W3C OWL Working Group, OWL 2 Web Ontology Language, Document Overview (Second Edition). W3C Recommendation 11 December 2012, <https://www.w3.org/TR/2012/REC-owl2-overview-20121211/> (abgerufen am 22.10.2021).

W3C SPARQL WORKING GROUP 2013

W3C SPARQL Working Group, SPARQL 1.1 Overview. W3C Recommendation 21 March 2013, <http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/> (abgerufen am 22.10.2021).

W3C WORLD WIDE WEB CONSORTIUM 2008

W3C World Wide Web Consortium, W3C Opens Data on the Web with SPARQL. Powerful Technology for Querying Distributed and Diverse Data. Press Release, 2008, <https://www.w3.org/2007/12/sparql-pressrelease> (abgerufen am 22.10.2021).

W3C WORLD WIDE WEB CONSORTIUM 2011

W3C World Wide Web Consortium, Linking patterns, 2011, <https://www.w3.org/2001/sw/wiki/Linking_patterns> (abgerufen am 22.10.2021).

WAIT 2017

G. Wait, Das „Chartered Institute for Archaeologists“: Der systematische Aufbau von Professionalität, Macht und Einfluss in Archäologie und Denkmalpflege, Archäologische Informationen 40, 2017, 121–130, <https://doi.org/10.11588/ai.2017.1.42472> (abgerufen am 22.10.2021).

WEINGART 2016

P. Weingart, Vertrauen, Qualitätssicherung und Open Access – Predatory Journals und die Zukunft des wissenschaftlichen Publikationssystems, in: P. Weingart – N. Taubert (Hrsg.), Wissenschaftliches Publizieren. Zwischen Digitalisierung, Leistungsmessung, Ökonomisierung und medialer Beobachtung, Interdisziplinäre Arbeitsgruppen Forschungsberichte 38 (Berlin - Boston 2016) 283–290.

WIGG-WOLF – DUYRAT 2017

D. Wigg-Wolf – F. Duyrat, La révolution des Linked Open Data en numismatique: Les exemples de nomisma.org et Online Greek Coinage (The Linked Open Data Revolution in Numismatics: The examples of nomisma.org and Online Greek Coinage), Archéologies numériques 17, 1, 1, 2017, 1–12, <https://doi.org/10.21494/ISTE.OP.2017.0171> (abgerufen am 22.10.2021).

WILKINSON U. A. 2016

M. D. Wilkinson – M. Dumontier – I. J. J. Aalbersberg – G. Appleton – M. Axton – A. Baak – N. Blomberg – J.-W. Boiten – L. B. da Silva Santos – P. E. Bourne – J. Bouwman – A. J. Brookes – T. Clark – M. Crosas – I. Dillo – O. Dumon – S. Edmunds – C. T. Evelo – R. Finkers – A. Gonzalez-Beltran – A. J. G. Gray – P. Groth – C. Goble – J. S. Grethe – J. Heringa – P. A. C. 't Hoen – R. Hooft – T. Kuhn – R. Kok – J. Kok – S. J. Lusher – M. E. Martone – A. Mons – A. L. Packer – B. Persson – P. Rocca-Serra – M. Roos – R. van Schaik – S.-A. Sansone – E. Schultes – T. Sengstag – T. Slater – G. Strawn – M. A. Swertz – M. Thompson – J. van der Lei – E. van Mulligen – J. Velterop – A. Waagmeester – P. Wittenburg – K. Wolstencroft – J. Zhao – B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Scientific data 3, 2016, 160018: 1–9, <https://doi.org/10.1038/sdata.2016.18> (abgerufen am 22.10.2021).

WILKINSON U. A. 2019

M. D. Wilkinson – M. Dumontier – I. J. J. Aalbersberg – G. Appleton – M. Axton – A. Baak – N. Blomberg – J.-W. Boiten – L. B. da Silva Santos – P. E. Bourne – J. Bouwman – A. J. Brookes – T. Clark – M. Crosas – I. Dillo – O. Dumon – S. Edmunds – C. T. Evelo – R.

Finkers – A. Gonzalez-Beltran – A. J. G. Gray – P. Groth – C. Goble – J. S. Grethe – J. Heringa – P. A. C. 't Hoen – R. Hooft – T. Kuhn – R. Kok – J. Kok – S. J. Lusher – M. E. Martone – A. Mons – A. L. Packer – B. Persson – P. Rocca-Serra – M. Roos – R. van Schaik – S.-A. Sansone – E. Schultes – T. Sengstag – T. Slater – G. Strawn – M. A. Swertz – M. Thompson – J. van der Lei – E. van Mulligen – J. Velterop – A. Waagmeester – P. Wittenburg – K. Wolstencroft – J. Zhao – B. Mons, Addendum: The FAIR Guiding Principles for scientific data management and stewardship, Scientific data 6, 1, 2019, 6: 1-2, <https://doi.org/10.1038/s41597-019-0009-6> (abgerufen am 22.10.2021).

WIMMER 2017

U. Wimmer, When Global Meets Personal – Überlegungen zur Einführung von RDA aus Sicht der Standardisierungsforschung, in: K. Umlauf – K. U. Werner – A. Kaufmann (Hrsg.), Strategien für die Bibliothek als Ort. Festschrift für Petra Hauke zum 70. Geburtstag (Berlin - Boston 2017) 281–296.

WISE – MILLER 1997

A. Wise – P. Miller, Why metadata matters in archaeology, Internet Archaeology 2, 1997, <https://doi.org/10.11141/ia.2.5> (abgerufen am 22.10.2021).

ZOZUS 2017

M. Zozus, The Data Book. Collection and Management of Research Data (Boca Raton 2017).

13 Abbildungsnachweise