

Virtuelle Forschungsumgebungen und Sammlungsräume

Objekte digital modellieren und miteinander vernetzen

Robert Nasarek

Einleitung

Die digitale Repräsentation und informationstechnisch gestützte Analyse von Sammlungsgegenständen sind in den letzten Jahren in etlichen Wissenschafts- und Kultureinrichtungen zu integralen Bestandteilen des Arbeitsfeldes geworden. Umgekehrt ist die Digitalisierung von wissenschaftlichen Sammlungen auch für die Informatik längst kein Neuland mehr. Dennoch stellen sich zu Beginn eines jeden Projekts ähnliche Fragen zum Umgang mit dem Digitalen: Welche Vorteile hat ein digitalisierter Sammlungsbestand, was leistet er, und was nicht? Welche Daten sind wichtig? In welcher Form sollen sie erfasst, gespeichert und repräsentiert werden? So auch in dem vom Bundesministerium für Bildung und Forschung geförderten Verbundprojekt »Objektsprache und Ästhetik«,¹ das im musealen Kontext drei historische Sammlungen von Konchylien (das sind die flachen oder gewundenen Gehäuse von Muscheln und Schnecken) hinsichtlich der Bezeichnung, Beschreibung und Bewertung ihrer Objekte untersucht. Vor dem Hintergrund der in diesem Projekt gewonnenen Erfahrungen und mit gelegentlichen Seitenblicken darauf diskutiert der vorliegende Beitrag die Möglichkeiten und Grenzen eines virtuellen Sammlungsraums und gibt eine kurze Einführung in die damit verbundene Datenmodellierung.

¹ Der vollständige Titel des Projekts lautet »Objektsprache und Ästhetik – Wertdimensionen des Objektbezugs in historischer Perspektive. Das Beispiel Konchylien«, vgl. dazu <https://www.leopoldina.org/ueber-uns/zentrum-fuer-wissenschaftsforschung/projekte/konchylien/> [30.06.2020].

Virtueller Sammlungsraum und virtuelle Forschungsumgebung

Wenn Wissenschafts- und Kultureinrichtungen zur Dokumentation und Präsentation ihrer Sammlungen auf die Mittel und Methoden der Informatik setzen, so schwanken die damit verbundenen Erwartungen oft zwischen den Extremen ›szenografischer Heilsbringer‹ und ›feindliche Übernahme‹.² Auf der einen Seite steht die Überschätzung der Möglichkeiten der Digitalisierung, auf der anderen die Befürchtung, die semantische Formalisierung von Informationen zu Daten würde vor allem zu einem Verlust führen. Indessen beruht diese Sorge häufig auf irrgigen Annahmen, denn auch in digital arbeitenden Forschungsprojekten wird ein nicht unerheblicher Teil der Aufgaben weiterhin konventionell und mit Hilfe von fließtextfokussierten Medien geleistet. Natürliche Sprache ist nach wie vor eines der effizientesten Mittel, um komplexe, mehrdimensionale Informationen (wie sie eigentlich alle Objekte mit sich bringen) zwischen Menschen auszutauschen. Während in natürlicher Sprache Informationen notwendigerweise immer als schon gedeutete und kontextualisierte mitgeteilt werden (als Aussagen über einen bestimmten und auf eine bestimmte Weise wahrgenommenen Ausschnitt der Welt), stellt eine informationstechnische Erfassung und Aufbereitung hingegen Informationen bzw. genauer Daten bereit, die als unstrukturierte und implizite Informationen aus den Quellen selbst gewonnen werden und sodann in digitale Einheiten gegliedert, strukturiert und formalisiert werden. So entstehen Daten in einer formalen Sprache, mittels derer sich die Daten besser vergleichen und reliabler bzw. überhaupt erst computergestützt analysieren lassen.

Solcher Verfahren der Formalisierung bedienen sich auch virtuelle Forschungsumgebungen³. Im konkreten Fall objekt- und sammlungsbezogener Anwendungen sind die hier generierten Daten die re-interpretierbare Darstellung von Bestandsinformationen, geeignet, um von Mensch und Maschine kommuniziert, interpretiert und verarbeitet zu werden. Der Modellierung dieser Daten kommt eine zentrale Aufgabe zu, denn sie bestimmt über die Aussagekraft und Auswertungsmöglichkeiten der Daten, mit Hilfe derer sich Personen gerechtfertigte Überzeugungen von der Wirklichkeit bilden oder auf die Algorithmen zugreifen können, um eigene Analysen durchzuführen oder eigene Datenbestände anzureichern.

Computer dienen hierbei zur Umstrukturierung und Speicherung von Daten, zur Prozessautomatisierung und zur Mustererkennung und bieten über Digitalfunktionen (Suchen, Bearbeiten, Verwalten, Visualisieren etc.) einen Zugang zu größeren Datenbeständen. Diese Funktionen und Prozesse sollen dabei helfen,

2 Krämer/Huber (2018), S. 1.

3 Für eine Definition virtueller Forschungsumgebungen siehe URL: https://www.forschungsdaten.org/index.php/Virtuelle_Forschungsumgebungen [30.06.2020].

zu Erkenntnissen zu gelangen, die dann beispielsweise wiederum als ausführlicher Text in natürlicher Sprache publiziert werden können. Innerhalb virtueller Forschungsumgebungen tritt der Text mit seiner dezidierten Wissensübertragung dennoch in den Hintergrund, während Daten und ihre Visualisierungsformen in den Fokus rücken. Diese Daten sollten für Forscher*innen und Interessent*innen, aber auch für maschinelle Prozesse⁴ nachnutzbar bereitgestellt werden. Indem die digitalisierten Objekte zusammen mit ihren Metadaten und integrierten Anwendungen einem breiteren Publikum zur Verfügung gestellt werden, konstituiert sich ein virtueller Sammlungsraum (Abb. 1). Die virtuelle Forschungsumgebung bildet dabei das Grundgerüst für dessen weitere Ausgestaltung.

Abbildung 1: Das digitalisierte Objekt ist hier der Scan einer Seite aus einem Manuskript von Friedrich Wilhelm Ernst Wunder (1742-1828) mit Zeichnungen von Konchylien.

Zu den Metadaten gehören etwa die Hauptbenennungen (»Sonnen Strahlen« und »Radii Solis«). Mittels des Mirador Viewers als integrierter Anwendung lässt sich der Scan vergrößern und genauer betrachten. © Archiv Leopoldina

Virtuelle Sammlungsräume bieten einige gewichtige Vorteile. So sind Digitalfunktionen bedeutend effizienter als analoge Prozesse. Sie ermöglichen sekunden-

4 Seit 2016 bieten die FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) Richtlinien an, Daten so zu gestalten, dass Computersysteme sie ohne oder nur mit wenigen menschlichen Eingriffen finden, auf sie zugreifen, mit ihnen interagieren oder sie wiederverwenden können. Vgl. FAIR Principles – GO FAIR, URL: <https://www.go-fair.org/fair-principles/> [30.06.2020]

schnelle Suchen in riesigen Datenbeständen oder das automatisierte Anreichern von Informationen aus externen Datenrepositorien. Mit einem internetfähigen Gerät lässt sich ein virtueller Sammlungsraum von überall aus aufsuchen und besitzt gewissermaßen keine Schließzeiten, was etwa Kooperationen selbst über große Entfernung vereinfacht. Dabei stellen sich allerdings auch neue Fragen nach der Urheberschaft von und den Zugriffsrechten auf Daten⁵ oder der angemessenen Beteiligung verschiedener Projektmitglieder.⁶ Kopien von digitalen Daten sind nahezu kostenfrei anzufertigen, was zusammen mit einer ständigen und unkomplizierten Erreichbarkeit die Nachnutzung der Ergebnisse begünstigt. Hypertext und Linked Data⁷ sind außerdem nur im digitalen Medium wirklich praktikabel umzusetzen und vereinfachen dort die Vernetzung von Wissen und Datenbeständen. Letztlich bieten digitale Methoden und Visualisierungen eine Vielzahl von flexiblen Analyse- und Darstellungsmöglichkeiten.⁸ Neben diesen individuell-technischen Vorteilen stellen virtuelle Sammlungsräume aber auch eine Antwort auf die gegenwärtigen Anforderungen an heutige Museen und Sammlungen dar. So konstatieren etwa Krämer und Huber, kulturelles Erbe überdauere nur, »[...] indem es beständig in die kulturellen Praktiken re-integriert und durch diese wiederbelebt werden kann.«⁹

Nicht selten wird an dieser Stelle der Einwand erhoben, Digitalisate entbehrten etwa der »Aura«, die mit der Anschauung originaler Objekte verbunden sei. Dieser Gedanke beruht sich meist auf Walter Benjamin, der im Jahr 1935 mit Blick auf Fotografie und Film davon gesprochen hat, die technische Reproduzierbarkeit von Kunstwerken entwerte die »Echtheit« und »Einzigartigkeit«, die der Betrachter nur im »Hier und Jetzt« der unmittelbaren Anschauung des Originals erfahren könne. Darüber hinaus verringere eine erweiterte Zugänglichkeit die ursprüngliche Exklusivität und das Besondere des Objektes.¹⁰ So bedenkenswert ein solcher Einwand im Geiste Walter Benjamins sein mag, so kontraproduktiv ist es, das physische Objekt und den realen Raum immer wieder gegen das Digitalisat und die virtuelle Umgebung auszuspielen.

Die Materialität eines Objektes und die soziale Dramaturgie des Ausstellungs- oder Sammlungsraums werden durch digitale Repliken und virtuelle Räume ergänzt – jedoch nicht obsolet. Dabei sollte außerdem beachtet werden, dass die materielle Erfahrbarkeit physischer Sammlungsobjekte für die meisten Menschen, die

5 Vgl. dazu auch die Beiträge von Katja Müller und Oliver Zauzig in diesem Band.

6 Vgl. Deck (2018), S. 5f.

7 Vgl. Homepage W3C – Linked Data, URL: <https://www.w3.org/standards/semanticweb/data> [30.06.2020].

8 Vgl. Cohen/Rosenzweig (2006).

9 Krämer/ Huber (2018), S. 4.

10 Vgl. Benjamin (1980), S. 477.

ein Museum besuchen, ähnlich eingeschränkt ist wie in virtuellen Sammlungsräumen. Im virtuellen Raum *kann* man die Objekte nicht berühren, in Ausstellungen und auch in etlichen Sammlungen *darf* man sie meist nicht berühren. Eine digitale Muschelschale lässt sich in einer virtuellen Umgebung von allen Seiten studieren, während bei einem in einer Vitrine exponierten Objekt häufig nur die eine durch Kuration und Gestaltung vorgegebene Ansicht möglich ist.

Digitalisate dienen weiterhin dem Schutz und Erhalt der Originale, deren Abbild sie sind. Nicht für jede forschende oder lehrende Befassung braucht es gleich das Original. Digitalisate sind zudem meist leichter verfügbar und in mancher Hinsicht auch einfacher zu nutzen. Mit den weiter voranschreitenden Möglichkeiten der Modellierung und digitalen Repräsentation entwickeln sich darüber hinaus erweiterte, teils eigene Formen der Darstellung, der Analyse und Vermittlung von Objekten. Dass es für etliche Forschungsfragen und ebenso im Kontext der Lehre weiterhin und wesentlich die Arbeit am Original braucht, ist davon unbenommen. Nur sind Zugang und Zugriff auf die originalen Objekte leider oft genug erschwert. Virtuelle Sammlungsräume können hier dazu beitragen, quantitative (Zugangs-)Beschränkungen zu minimieren, was die Sichtbarkeit vor allem von kleineren oder peripheren Sammlungen und Institutionen erhöht. Gleichwohl stellt die digitale Dokumentation der Bestände für nahezu alle sammlungstragenden Einrichtungen eine nicht unerhebliche Herausforderung dar.¹¹ Das gleiche gilt für die Verfügbarmachung der dabei entstehenden Daten: Bildet der auf dem Bildschirm bislang am häufigsten anzutreffende Suchschlitz die einzige Verbindung zur digitalen Sammlung auf einer Website, so verengt sich auch hier der Zugang zu den Objekten und wird auf die Eingabe einiger Schlagwörter reduziert. Hier sind stattdessen weitere (und weitergehende) Möglichkeiten zur Erkundung der Daten gefragt, die auch eine vertiefte Nutzung der digitalen Bestände erlauben.¹² Solche erweiterten Funktionalitäten, Visualisierungen und Analysemöglichkeiten des digitalen Sammlungsraums sind jeweils eng verbunden mit der benutzten Software bzw. virtuellen Forschungsumgebung und insbesondere dem verwendeten Datenmodell. Die Datenmodellierung der Sammlungsbestände ist daher eine fundamentale Aufgabe bei der Entwicklung virtueller Sammlungsräume.¹³

11 Vgl. hierzu etwa auch den Beitrag von Stefan Przigoda in diesem Band, der der Frage nach »Menge und Qualität« der digitalen Dokumentation nachgeht, oder den Aufsatz von Julia Rössel zu den Herausforderungen virtueller Kupferstichkabinette.

12 Hier sei auf die Projekte des Urban Complexity Lab hingewiesen, welche innovative Ansätze zur kreativen und kritischen Visualisierung von Datensätzen entwerfen. Siehe Projects – UCLA, URL: <https://uclab.fh-potsdam.de/projects/> [30.06.2020].

13 Vgl. Elgammal/Kang/Leeuw (2017).

Daten und Datenmodellierung

Datenmodelle sind Formalisierungen und Abstraktionen, die die inhaltliche Struktur von Beständen, sowohl im Allgemeinen als auch unter Berücksichtigung spezifischer Fragestellungen und Funktionen, repräsentieren.¹⁴ Die Qualität eines Datenmodells bemisst sich an einer Vielzahl von Anforderungen wie Vollständigkeit, Redundanzfreiheit, Gegenstandstreue, Wiederverwendbarkeit, ein ausgewogenes Verhältnis von Stabilität und Flexibilität, flüssige Integration in bestehende Systeme, gute Kommunizierbarkeit bei größtmöglicher Einfachheit.¹⁵ Kein Datenmodell wird all diesen Anforderungen vollständig gerecht werden können, doch können gründliche Vorüberlegungen zu möglichen Funktionen, zur Organisationsform der Daten und zum Modellierungsvorgang die Qualität des jeweiligen Datenmodells deutlich erhöhen.

Datenmodelle sind Konstrukte, abhängig von den Anforderungen, Fragestellungen, Perspektiven und Voraussetzungen ihrer Entwickler*innen. Sie stellen keine außerhalb jeder subjektiven Wahrnehmung existierende, »objektive« Wirklichkeit dar. Dennoch erfordert ein gutes Datenmodell strenge und präzise Definitionen der verwendeten Begriffe, eine logische und formalisierte Organisation der Informationen und damit einen hohen Anteil an expliziten Informationen. Somit sind Datenmodelle einer strengen strukturierten Heuristik unterworfen. Datenmodelle (vor allem gut dokumentierte) erhöhen potenziell die Nachvollziehbarkeit, Intersubjektivität und Reproduzierbarkeit von Wissen und ermöglichen es, die Validität von Ergebnissen besser einzuschätzen.¹⁶ Die Bedingung, sich dabei auf explizite, möglichst intersubjektive und raum-zeitlich bestimmbare Entitäten festzulegen, mag manchen Forscher*innen als Informationsverkümmерung im Vergleich zu einem interpretationsreichen Fließtext vorkommen, doch gerade die analytische Beschränkung auf möglichst präzise Begriffe und Zusammenhänge begünstigt umgekehrt eine ergänzende Form der Hermeneutik und Reflexion über den Gegenstand: »Reflections like these may lead to a more precise quantitative analysis of argumentative strategies, but would certainly not have come into view by just reading the texts.«¹⁷

In Forschungsvorhaben, an denen mehrere Einrichtungen beteiligt sind, arbeiten die Projektpartner*innen üblicherweise ebenso an einer übergeordneten Fragestellung wie an spezifischen Aufgabenstellungen. Oftmals fließt zum Schluss trotzdem ein Großteil der Daten in einem Datenmodell zusammen und wird gemeinsam digital dargestellt. Im bereits erwähnten Verbundprojekt »Objektsprache

14 Vgl. Jannidis/Kohle/Rehbein (2017), S. 100.

15 Vgl. Simsion/Witt (2009), S. 32.

16 Vgl. Heßbrüggen-Walter (2015), S. 8.

17 Ebd., S. 10.

und Ästhetik« benutzt jedes Teilprojekt daher zunächst sein eigenes Datenmodell, um relevante Daten zu erfassen. Dies soll zunächst den Freiraum geben, bedeutsame und bemerkenswerte Informationen aus eigener Perspektive zu beschreiben, ohne in die Pfadabhängigkeiten eines schon vorher definierten Schemas zu geraten. Die Teilprojekte tauschen sich im Verlauf des Projekts in ständiger Absprache über Entitäts-, Attributs- und Beziehungsdefinitionen aus und gleichen die semantischen Definitionen mit Standards ab, um so schriftweise ein gemeinsames Modell zu erstellen. Für die Verknüpfung der Objekte haben sich die beteiligten Institutionen auf ein gemeinsames Attribut geeinigt: die moderne taxonomische Bezeichnung der Konchylie. Über den Weg der modernen binominalen Klassifikation können so die Objekte aus allen drei Sammlungen jenseits der individuellen historischen Begrifflichkeiten in Beziehung gesetzt werden. Dieser Teil der Zusammenarbeit erfordert ein nicht zu unterschätzendes Bewusstsein für die Fragestellungen und Herangehensweisen der jeweils anderen Fachrichtungen, damit gemeinsame Konzepte auch verstanden werden können. Regelmäßige Absprachen mit Projektpartner*innen und Nutzer*innen mit Blick auf die Daten und den zu modellierenden virtuellen Sammlungsraum erhöhen die Wahrscheinlichkeit, dass das Endergebnis zu den Erwartungen und Zielen passt und Konsens über ein solides und kommunizierbares Datenmodell erzielt werden kann.

Die Organisationsform des Datenmodells wird als »Metamodell« bezeichnet.¹⁸ Die Wahl des Metamodells steht am Anfang der Datenmodellierung, da es die Form der Datenaufnahme bestimmt bzw. festlegt, in welcher Form die Daten letztlich gespeichert werden sollen. Es werden hauptsächlich vier Arten von Metamodellen im Bereich der digitalen Geisteswissenschaften angewendet:

Tabellen bieten sich als ein relationales Modell für stärker strukturierte Informationen an. Es ist derzeit das am häufigsten verwendete Metamodell und bereits mit geringen Vorkenntnissen in einem Tabellenkalkulationsprogramm umzusetzen. Tabellen konzentrieren sich auf singuläre Begriffe und Daten, die normiert und kategorisiert werden, wobei als Produkt meist nur noch die Interpretation der Quellinformationen in der Datenbankinstanz verbleibt.

Ein anderes Metamodell sind digitale Editionen; diese sind oft in der Annotationssprache XML verfasst, bei der Struktur, Metadaten und Interpretation durch Annotationen auf die Quellinformationen »aufgesetzt« werden. So bleiben unstrukturierte Informationen – in den meisten Fällen handelt es sich dabei um [Fließ-]Text – in ihrer ursprünglichen Form erhalten. Dies ermöglicht einen Rückbezug auf den nativen Inhalt und die Struktur des Quellenmaterials, erfordert jedoch meist einen höheren Aufwand als eine Erfassung in Tabellenform. XML-Annotationen behalten Richtung und zeitliche Abfolge des Informationsflusses

18 Vgl. Flanders/Jannidis (2019), S. 33.

bei, werden aber auch schnell unübersichtlich und sind in ihrer Gesamtheit ohne einen Transformations- oder Visualisierungsschritt *de facto* nicht erschließbar.

Eine dritte Variante sind graphbasierte Metamodelle; diese eignen sich besonders zur Modellierung von Netzwerken, Taxonomien und Ontologien, weil sie die Beziehungen der einzelnen Dinge in den Fokus rücken und es einfacher gestalten, semantische Pfade innerhalb eines Bedeutungsgeflechts zu verfolgen. Ontologien¹⁹ werden meist im Resource Description Framework (RDF) und hauptsächlich in der Web Ontology Language (OWL) umgesetzt. Auch wenn relationale Datenbanken mit verschachtelten Mengen (Nested Sets) ebenfalls über einen performanten Weg zur Abfrage innerhalb von Bäumen und Netzwerken verfügen,²⁰ ist der Graph die native Form von Informationsgeflechten mit unterschiedlichen Abstraktionsniveaus.²¹ Die durch die Graphenstruktur obligatorische Dreiteilung der Informationen in Ausgangsknoten – Beziehung – Eingangsknoten erzwingt bei komplexen Objekteigenschaften meist eine aufwendige Reifikation²² von Informationen. Ohne eine passende Eingabehilfe (z.B. über das Erfassungsinterface der virtuellen Forschungsumgebung WissKI²³) oder ein automatisiertes Importverfahren gestaltet sich die Datenaufnahme in Graphenstrukturen äußerst aufwendig.

Objektorientierte Metamodelle schließlich bieten sich an, wenn Daten mittels einer Programmiersprache über einen Webservice abgefragt und ad hoc weiterverarbeitet werden sollen. Zusätzlich zu den zuvor dargestellten Metamodellen speichern und vererben objektorientierte Metamodelle nicht nur die Eigenschaften eines Gegenstandes, sie bieten auch die Möglichkeit, Methoden zu definieren, die festlegen, wie beim Aufruf der Methode mit bestimmten Eigenschaften umgegangen werden soll.

Da innerhalb eines Projektes verschiedene Quellengattungen mit unterschiedlichem Strukturgrad (Fließtext, Register, Tabellen) auftreten können und externe Datenrepositorien für die automatisierte Abfrage ihrer Daten eine Programmiersprache voraussetzen, ist es meist nicht sinnvoll oder auch nur möglich, sich auf

19 Unter einer Ontologie wird hier eine zwischen Menschen geteilte formalisierte Sprache zur Beschreibung von Begriffen und deren Beziehungen zueinander verstanden, die maschinenlesbar umgesetzt wurde. Vgl. Guarino/Oberle/Staab (2009).

20 Vgl. Celko (2015), S. 547-579.

21 Vgl. Robinson/Webber/Eifrem (2015), S. 8-10.

22 Der Dreischritt Ausgangsknoten/Beziehung/Eingangsknoten erlaubt lediglich Aussagen nach dem Schema: Subjekt–Prädikat–Objekt. Komplexe Informationen von Aussagen über Aussagen müssen daher in mehrere »Sätze« aufgeteilt werden. Mehr dazu bei Stevens/Lord (2010).

23 Die Wissenschaftliche Kommunikationsinfrastruktur (WissKI) ist eine virtuelle Forschungsumgebung, die insbesondere die wissenschaftliche Kommunikation und Dokumentation im Bereich des kulturellen Erbes unterstützen soll und eine Ontologie-gestützte Wissensverarbeitung informationstechnisch umsetzt. Vgl. Homepage WissKI, URL: <http://wiss-ki.eu/> [30.06.2020]. Siehe auch den Beitrag von Udo Andraschke und Sarah Wagner in diesem Band.

nur ein einzelnes Metamodell festzulegen. In arbeitsteiligen Projekten ist das Fertigkeitsniveau im Umgang mit den verschiedenen Metamodellen zudem oft unterschiedlich. Daher werden entweder die Mitarbeiter*innen im Umgang mit den jeweiligen Metamodellen geschult oder die Metamodelle werden nach den Fertigkeiten der Mitarbeiter*innen ausgewählt und ein*e Datenkurator*in fügt die Teile in einem Metamodell zusammen.

Im Projekt »Objektsprache und Ästhetik« wird in einem Teilprojekt eine digitale Edition des Sammlungskatalogs in XML erarbeitet. Hingegen ist es für die beiden anderen Teilprojekte aufgrund der stark strukturierten Quellen zielführender, die Informationen in Tabellen aufzunehmen. Um die Daten mit Hilfe verschiedener Webservices mit weiteren Informationen anreichern zu können, werden sie über das vereinfachte Austauschformat der »Comma Separated Values« (CSV) in ein objektorientiertes Metamodell einer Python-Umgebung übertragen. CSV-Dateien besitzen aufgrund ihres simplen Aufbaus zwar die Stärke einer universellen Importier- und Exportierbarkeit, jedoch mangelt es ihnen dafür an nahezu allen anderen Funktionen eines Datenbankmanagementsystems (Datentypenbeschränkungen, Primär- und Fremdschlüssel, Abfragen etc.). Aus diesem Grund laufen die Daten in einer relationalen Datenbank wieder zusammen, um letztlich in WissKI importiert zu werden, die jedoch auf einem graphenbasierten Metamodell aufbaut. Damit finden sich im Projekt »Objektsprache und Ästhetik« alle hier vorgestellten Metamodelle wieder, zum Teil sogar in verschiedenen Abstufungen der Komplexität. Sie wurden nach dem Strukturierungsgrad der Quelle, den Stärken ihrer Features, softwarebedingten Pfadabhängigkeiten und den Fertigkeiten der Projektpartner*innen ausgewählt.

Datenmodelle in Forschungsvorhaben richten sich nach den Fragestellungen und Aufgaben des Projekts und der Projektpartner*innen. Begrenzt werden diese Anforderungen durch die Möglichkeiten, die das vorhandene Quellenmaterial bietet. Andererseits ergeben sich Fragestellungen und Erkenntnisse oft erst aus der Arbeit mit dem Material, die dann im Datenmodell implementiert werden. Außerdem muss beachtet werden, wie das Datenmodell und dessen Inhalte später dargestellt werden sollen und über welche Funktionen sie erreicht werden können. Entscheidend hierfür sind auch die Interessen der nachnutzenden Nutzer*innen.

Im Bereich des kulturellen Erbes ist das Informationsspektrum der untersuchten Gegenstände zumeist groß (wie viele Informationsarten – z.B. Autor*in, Fundort Herstellungsdatum – existieren insgesamt?) und die Informationsdichte oft sehr unterschiedlich (wie viele Informationsarten sind für einen bestimmten Gegenstand verfügbar?). Hinzu kommt, dass in den Geistes- und Kulturwissenschaften die Zusammenhänge der Dinge, über die etwas ausgesagt werden soll (die Entitäten), oftmals sehr komplex sind. Hier dominiert ein heuristisches Vorgehen

zur Erkenntnisgewinnung: Mit begrenztem Wissen wird die Datenbankinstanz²⁴ anhand des bearbeiteten Materials Schritt für Schritt erweitert. Je komplexer die Zusammenhänge innerhalb des Datenmodells jedoch sind und je häufiger Änderungen vorgenommen werden, desto aufwendiger wird es, die Datenquantität und -qualität konsistent zu halten. Gerade zu Beginn der Modellierung verändert sich das Datenmodell häufig, Kategorien entstehen neu oder werden wieder verworfen, Schwerpunkte verlagern sich und Zusammenhänge werden erst nach und nach erkannt. Das Metamodell muss daher eine möglichst flexible Gestaltung, eine Entwicklung mit dem Erkenntnisprozess zulassen und nah am Quellenmaterial bleiben. Markup-Sprachen wie XML oder Software zur qualitativen Datenanalyse lassen die Ursprungsdaten intakt, bei der Nutzung von Tabellenformaten genügen oft exakte Quellenverweise. So kann schnell zu kritischen oder unklaren Stellen zurückgesprungen werden, um sie zu bearbeiten, wenn sich neue Erkenntnisse eingestellt haben. Außerdem bilden sie eine Brücke zwischen den Herausforderungen der inhaltlichen Erschließung und der Struktur der Quelle bzw. des Gegenstands. Dabei ist zu beachten, dass Softwareumgebungen die Arbeit mit dem Material tendenziell erleichtern können, ihr Datenausgabeformat jedoch die spätere Integration in andere Umgebungen oder nachträgliche Transformation erschweren kann. Metamodelle können tendenziell frei gewählt werden, so lange sie zum Material passen und sie von den Bearbeiter*innen adäquat verwendet werden können. Jedes Metamodell lässt sich im Prinzip in jedes andere Metamodell transformieren.

Auch die verwendete Methodik sollte einem heuristischen Vorgehen gerecht werden. Hier bietet sich eine adaptierte Form der qualitativen Inhalts- und Objektnalyse²⁵ an. Um eine gleichbleibende Datenqualität zu gewährleisten, sollte iterativ vorgegangen werden: Zu Beginn wird ein repräsentativer Teil der Daten ausgewählt (10-50 %), mit dem der ganze Workflow durchgearbeitet wird. Zu diesem Modell werden dann größere Teile oder der Rest der Daten hinzugefügt. Ergeben sich hierbei gravierende Änderungen, wird das gesamte bisherige Material erneut überarbeitet.²⁶ So wächst das Datenmodell mit der wachsenden Erkenntnis und die Daten werden einheitlich aufgenommen. Eine Automatisierung innerhalb dieser repetitiven Phasen bringt den zusätzlichen Vorteil mit sich, dass die – bei größeren Datenbeständen nur schwer zu identifizierenden – händischen Zufallsfehler vermieden werden können.

Die Datenmodellierung bewegt sich im Spannungsfeld zwischen forschungsorientierten bzw. kurativen Aufgaben und einer konzeptbezogenen bzw. objekt-

²⁴ Das ist der Datenbankzustand zu einer bestimmten Zeit, d.h. das gewählte Datenmodell und die aktuell eingetragenen Daten.

²⁵ Vgl. Wikipedia, Qualitative Inhaltsanalyse, URL: https://de.wikipedia.org/wiki/Qualitative_Inhaltsanalyse [30.06.2019].

²⁶ Vgl. Kuckartz (2016), S. 86.

bezogenen Modellierung. Eine forschungsorientierte Datenmodellierung versucht projektspezifische Fragestellungen zu beantworten und Funktionen zu erfüllen; eine kurative Datenmodellierung soll die Interessen von nachnutzenden Personen abdecken.²⁷ Eine Möglichkeit, diesen beiden Aufgabenstellungen gerecht zu werden, ist die Nutzung von Standards und Normen. An den Richtlinien der Text Encoding Initiative (TEI)²⁸ oder dem Referenzmodell des Internationalen Komitees für Dokumentation (CIDOC CRM)²⁹ orientieren sich viele Projekte, die kulturelles Erbe modellieren. Die TEI besteht seit über 40 Jahren und bietet umfassende Richtlinien zur Kodierung von Texten und gilt inzwischen als De-facto-Standard. Das CIDOC CRM, eine zertifizierte ISO-Norm, ist ein Beschreibungsmodell für die Dokumentation von kulturellem Erbe, das ein gemeinsames Verständnis von Begriffen in Verbindung mit Sammlungsobjekten erleichtern soll. Es bietet eine formale Sprache für die Beschreibung und Strukturierung dieser Konzepte an und unterbreitet dazu Modellierungsvorschläge. Das CIDOC CRM beinhaltet Namenskonventionen und Definitionen auf einer relativ abstrakten Ebene und ist offen für Erweiterungen eigener Konzepte. Das macht dieses Modell zu einer sogenannten Top-Level-Ontologie. Es wird stetig weiterentwickelt und von der *CIDOC CRM Special Interest Group (SIG)* gewartet.

Standards bieten den entscheidenden Vorteil, auf der Erfahrung einer wissenschaftlichen Community aufbauen zu können; auch führt ihre breite Akzeptanz zu einer besseren Nachnutzbarkeit und Vergleichbarkeit der Ergebnisse.³⁰

Bei der Modellierung von Sammlungsobjekten sind zudem die Unterschiede zwischen objektbezogenem und konzeptbezogenem Modellieren und die aus diesen Unterschieden folgenden Konsequenzen zu beachten: Werden konkrete individuelle Objekte modelliert, die das Wissen oder die Perspektive in einem zeitlich statischen, historischen Kontext zeigen sollen, handelt es sich um objektbezogenes Modellieren. Es wird ein konkretes Objekt mit individuell bezogenen Objektdaten modelliert.

Werden hingegen normative Konzepte modelliert, die auf mehrere Objekte passen und nach gegenwärtigen Ansprüchen und gegenwärtigem Wissen Richtigkeit und Vollständigkeit anstreben – d.h. wird eine standardisierte Idee von einem Objekt entwickelt –, so handelt es sich um konzeptbezogenes Modellieren.

Auch wenn das primäre Erkenntnisinteresse eines Projekts wie »Objektsprache und Ästhetik« ein (wissenschafts)historisches ist, nämlich die Untersuchung

27 Vgl. Jannidis/Kohle/Rehbein (2017), S. 99.

28 Vgl. Homepage Text Encoding Initiative, URL: <https://tei-c.org/> [30.06.2020].

29 CIDOC steht für Comité International pour la Documentation, das Komitee des Internationalen Museumsverbandes; CRM für Conceptual Reference Model. Vgl. Homepage CIDOC CRM, URL: www.cidoc-crm.org/ [30.06.2020].

30 Vgl. Jannidis (2016).

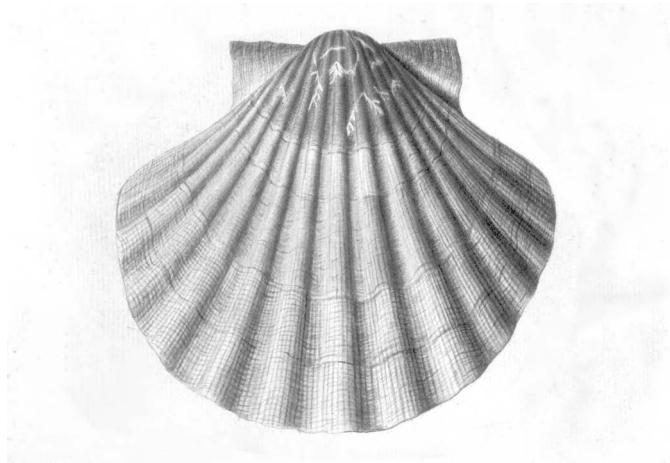
des Wandels der Terminologie im Sammlungskontext, so liefern die im Projekt genutzten modernen Datenbanken zugleich auch zahlreiche weitere Informationen, etwa zur Taxonomie der untersuchten Meereslebewesen, zu ihrem Lebensraum und ihrer Verbreitung. Und sie tun dies mit dem Anspruch, möglichst umfassende, korrekte und aktuelle Daten zu liefern. Um diesen Umstand im Sinne eines möglichen Erkenntnismehrwerts fruchtbar zu machen, empfiehlt sich die Verbindung eines objektbezogenen Ansatzes mit einem konzeptbezogenen (ohne dabei das Prinzip der historisch-individuellen Ebene infrage zu stellen).

Resümee: Ceci n'est pas une coquille

Virtuelle Forschungsumgebungen und Sammlungsräume bieten erhebliche Potenziale für die objekt- und sammlungsbezogene Forschung und die digitale Präsentation physischer Sammlungen. Sie unterstützen fächer- und ortsübergreifende Kooperation und Kollaboration, sie erhöhen die Sichtbarkeit der Bestände und sie ermöglichen innovative Darstellungen der Objekte und Objektinformationen. Virtuelle Forschungsumgebungen und Sammlungsräume ersetzen dabei nicht bestehende Infrastrukturen oder gar die Originale, sondern ergänzen sie und passen sie der heutigen Zeit bzw. dem heutigen Methodenspektrum an. Was datenproduzierende und -getriebene Projekte von früheren oder konventionellen Erschließungsvorhaben unterscheidet und ihre Qualität potenziell erhöht, ist demnach nicht die grundlegende Art des wissenschaftlichen Arbeitens mit den Objekten, sondern der zweckbezogene Grad der Formalisierung und Explizitheit von Informationen sowie die weitere »Zielgruppe« der Maschinen und Algorithmen. Eine hierfür geeignete Auswahl des Metamodells und eine gelungene Datenmodellierung unterstützt und ergänzt das klassische objekt- und sammlungsbezogene wissenschaftliche Arbeiten durch eine digitale Hermeneutik. Die gewonnenen Daten erweitern die potenziellen Nutzungsmöglichkeiten ebenso wie den Nutzer*innenkreis und tragen außerdem dazu bei, die Objektivität und Nachprüfbarkeit der gewonnenen Ergebnisse zu erhöhen.

Am Beispiel des digitalen Konchylienkabinetts wollte dieser Beitrag zeigen, wie sich mehrere Sammlungen auf Grundlage einer virtuellen Forschungsumgebung virtuell vereinen, sich unter verschiedenen fachübergreifenden Aspekten erforschen, sich mit heutigem Wissen verknüpfen und angemessen präsentieren lassen. Essenziell für dieses Vorhaben war und ist eine regelmäßige und klare Kommunikation zwischen den Projektpartner*innen über die Inhalte und Funktionalitäten des Datenmodells sowie über die »Ausstattung« des virtuellen Sammlungsraums. Als digitales Abbild der Sammlung kann dieser nur das präsentieren und zugänglich machen, was durch die an diesem Projekt beteiligten Mitarbeiter*in-

Abbildung 2: Das ist keine Muschelschale. Digitalisierte Objekte und ihre Daten sind Abbilder von Originalen und als solche aus einer bestimmten Perspektive angefertigt, um spezifische Funktionen zu erfüllen. © Archiv Leopoldina.



nen digitalisiert, erforscht und als bedeutsam festgelegt wurde – und somit nicht länger zwischen zwei Muschelschalen verschlossen blieb.

Literatur

- Benjamin, Walter: Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit, in: Tiedemann, R./Schweppenhäuser, H. (Hg.): *Walter Benjamin – Gesammelte Schriften*. Frankfurt a.M. 1980, S. 471-508.
- Celko, Joe: *Joe Celko's SQL for Smarties. Advanced SQL Programming* (5., überarb. Aufl.), Burlington, MA 2015.
- Cohen, Daniel J./Rosenzweig, Roy: *Digital History. A Guide to Gathering, Preserving, and Presenting the Past on the Web*, Philadelphia, PA 2006; verfügbar unter: <http://chnm.gmu.edu/digitalhistory/book.php> [30.06.2020].
- Deck, Klaus-Georg: *Digital Humanities – Eine Herausforderung an die Informatik und an die Geisteswissenschaften*, in: Martin Huber/Sibylle Krämer (Hg.): *Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden*. Sonderband der Zeitschrift für digitale Geisteswissenschaften (Bd. 3.), Wolfenbüttel 2018; verfügbar unter: http://zfdg.de/sb003_002 [30.06.2020].
- Elgammal, Ahmed/Kang, Yan/Leeuw, Milko: *Picasso, Matisse, or a Fake? Automated Analysis of Drawings at the Stroke Level for Attribution and Authentication*, 2017; verfügbar unter: <https://arxiv.org/pdf/1711.03536.pdf> [30.06.2020].
- Flanders, Julia/Jannidis, Fotis: *A Gentle Introduction to Data Modeling*, in: dies. (Hg.): *The Shape of Data in Digital Humanities. Modeling Texts and Text-based Resources*, London 2019, S. 26-95.
- Guarino, Nicola/Oberle, Daniel/Staab, Steffen: *What Is an Ontology?*, in: Steffen Staab/Rudi Studer (Hg.): *Handbook on Ontologies*, Dordrecht u.a. 2009, S. 1-17.
- Heßbrüggen-Walter, Stephan: *What People Said: The Theoretical Foundations of a Minimal Doxographical Ontology and Its Use in the History of Philosophy*, in: Constanze Baum/Thomas Stäcker (Hg.): *Grenzen und Möglichkeiten der Digital Humanities*. Sonderband der Zeitschrift für digitale Geisteswissenschaften (Bd. 1.), Wolfenbüttel 2015; verfügbar unter: http://zfdg.de/sb001_001 [30.06.2020].
- Jannidis, Fotis: *Digital Humanities*, Berlin/Heidelberg 2016; verfügbar unter: <https://gi.de/informatiklexikon/digital-humanities/> [30.06.2020].
- Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hg.): *Digital Humanities. Eine Einführung*, Stuttgart 2017.
- Krämer, Sybille/Huber, Martin: *Dimensionen Digitaler Geisteswissenschaften. Einführung*, in: dies. (Hg.): *Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden*. Sonderband der Zeitschrift für digitale Geisteswissenschaften (Bd. 3.), Wolfenbüttel 2018; verfügbar unter: http://zfdg.de/sb003_013 [30.06.2020].

- Kuckartz, Udo: Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung, Weinheim/Basel 2016.
- Müller, Katja: Digitale Objekte – subjektive Materie. Zur Materialität digitalisierter Objekte in Museum und Archiv, in: Hans Peter Hahn/Friedemann Neumann (Hg.): Dinge als Herausforderung: Kontexte, Umgangsweisen und Umwertungen von Objekten, Bielefeld 2018, S. 49-66.
- Robinson, Ian/Webber, Jim/Eifrem, Emil: Graph Databases. New Opportunities for Connected Data, Sebastopol, CA 2015.
- Simsion, Graeme C./Witt, Graham C.: Data Modeling Essentials. San Francisco, CA 2009.
- Stevens, Robert/Lord, Phillip: Reification of properties in an ontology, in: Ontogenesis (2010); verfügbar unter: <http://ontogenesis.knowledgeblog.org/993/> [30.06.2020].
- Weindl, Roman: Die »Aura« des Originals im Museum. Über den Zusammenhang von Authentizität und Besucherinteresse, Bielefeld 2019.

