

Daten als Ressourcen

Herausforderungen virtueller Kupferstichkabinette

Julia Rössel

Einleitung

In den Anfängen der computergestützten Dokumentation von Sammlungsobjekten stand es im Vordergrund, den Anforderungen lokaler Praktiken, wie etwa der Standortbestimmung eines Objekts in der Sammlung, gerecht zu werden.¹ Die Idee des Museums als Wissensspeicher, das den Zugang zu materiellen Objekten gewährt und das Wissen zu ihnen unter der Nutzung moderner Mediation abrufbar macht, fand in der Datenbank oder genauer dem Museumsmanagement System eine konkrete Form.² Heute ist die Nutzung solcher Systeme üblich. Darüber hinaus haben sich für die Dokumentation von Sammlungsobjekten formale Standards entwickelt, deren Nutzung auch international und institutionell befördert wird. Insbesondere die Durchsetzung des Internet als Massenmedium und die Entwicklungen von Präsentations- und Informationsmöglichkeiten über Webanwendungen haben dazu geführt, dass Sammlungsinstitutionen bereits bei der Erhebung von Daten deren Online-Publikation als Ziel sehen. Das webbasierte Information Retrieval auf eigenen und übergreifenden Plattformen ist eine zentrale Motivation für Museen, ihre Bestände zu digitalisieren. Bilder der sammlungseigenen Objekte sollen zusammen mit verlässlichen Informationen interessierten

-
- 1 Bereits in den späten 1960er Jahren widmet sich Robert G. Chenhall auf Basis der Analyse des Museums als System der Frage nach den Einsatzmöglichkeiten computergestützter Informationssysteme. Vgl. Chenhall (1968).
 - 2 So sprach etwa Hugues de Varine-Bohan, Leiter des ICOM, in einem Interview 1977 von seinem Verständnis des Museums als Datenbank: »Das Museum als Mittel ist gleichsam eine Bank, eine Sachbank. Ich meine dies im Sinne einer Datenbank oder Blutbank [...]. Diese Sachbank hat zum Ziel, eine gewisse Anzahl von Daten in Form von Gegenständen und Dokumenten im zwei- oder dreidimensionalen Raum anzuhäufen. Sie werden gesammelt, klassifiziert, belegt, ins Inventar aufgenommen, beschriftet, konserviert, restauriert... so daß jeder, der sie benötigt, sich jederzeit ihrer bedienen kann: ich sage bedienen, nicht nur einfach sie sehen. Dies betrifft, kurz zusammengefaßt, das Museum als Mittel, als Instrument.« Siehe Rojas/Crespán/Trallero/Vrine-Bohan (1977), S. 12.

Nutzer*innen präsentiert und für sie recherchierbar gemacht werden. Wenn sich GLAM-Institutionen als Datengeber engagieren, so ist damit auch der Gedanke der Authentizität von Daten verbunden. Die institutionelle Herkunft der Daten zu Kulturobjekten soll sowohl die Korrektheit von Bildeindrücken als auch die wissenschaftliche Verlässlichkeit der Metadaten gewährleisten.³

Der Gedanke an die Online-Publikation dominiert und formt allerdings die Erfassung von Objekten im Netz und den institutionellen Umgang mit ihnen. Ein differenzierter Blick zeigt, dass verschiedenste Akteur*innen am Publikationsprozess beteiligt sind, die diverse Anforderungen an die Gestalt des digitalen Objektes stellen.⁴ Dies macht die Einschätzung von Datenqualität, die bedarfsorientiert ausgerichtet ist, zu einem komplexen Evaluationsprozess.⁵ Gerade bei der Nutzung von institutionsübergreifenden Plattformen zeigt sich, wie heterogen die Datenqualität verschiedener Datengeber ist.⁶ Hierdurch stellt sich die Frage, wie ein Qualitätsmanagement für Daten bereits in den ersten Phasen ihres Lebenszyklus, also schon bei der Konzeption des Datenmodells und der Erhebung der Daten durch wissenschaftliche Erschließung und Beschreibung, aussehen kann.⁷ Wie können Kulturinstitutionen selbst dazu beitragen, die Qualität ihrer Daten dauerhaft zu sichern und damit sowohl den eigenen Ansprüchen als auch denen externer Datennutzer*innen gerecht zu werden? Wie können Daten für verschiedene, auch zukünftige oder unbekannte Nutzungskontexte flexibel gehalten werden?

Die folgenden Beobachtungen und Gedanken basieren auf einem Metadaten Assessment, das von 2018 bis 2019 innerhalb des Projektes »Virtuelles Kupferstich-

3 Der Blog »Yellow Milkmaid Syndrom« liefert viele Beispiele für den visuellen Variantenreichtum und damit die visuelle Unzuverlässigkeit von digitalen Bildern von Kunstwerken. U.a. das Rijksmuseum reagierte hierauf mit der Initiative zur offenen Publikation sowohl seiner Bild- als auch Metadaten. Vgl. hierzu Verwayen/Arnoldus/Kaufman (2011).

4 Im Folgenden wird auf den vom Rat für Informationsinfrastrukturen dargestellten Lebenszyklus von Forschungsdaten referenziert, der die Etappen von Daten von ihrer konzeptionellen Entstehung, über die Qualitätskontrolle bis hin zum Teilen und Archivieren darstellt. Vgl. RfII (2019), S. 30.

5 Vgl. RfII (2019), S. 18.

6 So wurde 2013 eine Taskforce zum Thema Datenqualität bei der Europeana gebildet, um die mit einer großen Heterogenität in Daten zusammenhängenden Probleme zu vermitteln und ihren Ursprüngen am Beginn des Datenlebenszyklus entgegenzuwirken. Vgl. Dangerfield/Karlshoven (2015), S. 5.

7 Datenlebenszyklen beschreiben als Modelle die Zustandsphasen von Daten und variieren je nach Domäne. Das zirkuläre Prinzip ist deshalb grundlegend, weil es zeigt, dass Daten immer weiterentwickelt werden und Weiternutzung oder Neukonzeptionen essenziell sind. Das vom RfII für Forschungsdaten entwickelte Modell ist z.T. auch für Daten zu Sammlungsobjekten anwendbar: Der Konzeption und Erhebung von Daten folgt hier u.a. eine Qualitätsprüfung, Beschreibung und später Verknüpfung von Daten. Das Teilen der Daten kann sowohl die Publikation in eigenen Anwendungen als auch die Weitergabe an Dritte bedeuten. Vgl. RfII (2019), S. 30.

kabinett« (VKK) an der Herzog August Bibliothek Wolfenbüttel durchgeführt wurde.⁸

Wie kommen Objekte ins Netz?

Die Erfassung von Daten zu Sammlungsobjekten geschieht vielfach im Kontext drittmittelförderter Forschungs- oder Erschließungsprojekte. Dieser Kontext beeinflusst bereits die Frage nach der Art und Weise der Datenmodellierung und der Vorgehensweise bei der Datenerschließung. So wurde etwa das Virtuelle Kupferstichkabinett als Verbundprojekt der Herzog August Bibliothek Wolfenbüttel und des Herzog Anton Ulrich Museums Braunschweig in mehreren Phasen von verschiedenen Förderern finanziert. Ziel war es, die graphischen Sammlungen der beiden Institutionen über eine Webanwendung gemeinsam recherchierbar zu machen. Als wissenschaftliche Fragestellung lag die Idee einer provenienzgeschichtlich orientierten, virtuellen Rekonstruktion der graphischen Sammlung der Herzöge zu Braunschweig-Wolfenbüttel zugrunde.⁹

Neben dem dezidierten Schwerpunkt auf der Erfassung der Objektgattung Druckgraphik (und später auch Zeichnung), stellte sich für die beteiligten Wissenschaftler*innen bei der Erfassung die Frage nach der Objektgeschichte: Merkmale, die Rückschlüsse auf die Sammlungszugehörigkeit eines Blattes geben, sollten ebenso kommuniziert werden wie technologische Details seiner Herstellung oder materialtechnische Merkmale wie Wasserzeichen. Insbesondere für druckgraphische Werke ist die Verbindung mit anderen Blättern relevant, etwa um den Publikationskontext einer Serie darzustellen. Aus Sicht von Forscher*innen können hierüber wiederum weitere Werke bei einer Recherche aufgefunden werden. Darüber hinaus können natürlich auch auf der Ebene der Darstellung wichtige Fragen der Graphikforschung geklärt werden, weshalb die ikonographische Erschließung ebenso relevant für Druckgraphik im Netz ist. Diese kurzen Einblicke in die Vorüberlegungen zeigen, dass bereits die Forschungsfrage den Blick auf das Objekt formt und dessen Darstellung auf der Ebene der Metadaten zu einem digitalen Bild bezüglich ihrer Ausrichtung und Tiefe steuert. In der Konzeptionsphase kommt daher der Datenmodellierung eine zentrale Rolle zu.

Gerade die Analyse älterer Daten des Virtuellen Kupferstichkabinetts zeigte, von welchen Dynamiken die Datenmodellierung beeinflusst werden kann. In den letzten zehn Jahren haben sich für Technologien und Standards von Forschungsdaten zu Objekten der materiellen Kultur bestimmte Prinzipien etabliert. So ist zum

⁸ Homepage VKK, URL: www.virtuelles-kupferstichkabinett.de/de/ [30.06.2020]. Ich danke David Maus für die Zusammenarbeit und Anregungen für diesen Beitrag.

⁹ Vgl. Pagel (2011), S. 91–104.

Beispiel das Harvestingformat LIDO (Lightweight Information Describing Objects), welches für den Austausch von Daten zu Objekten der materiellen Kultur entwickelt wurde, mittlerweile ein anerkannter Standard, an dem sich bei der Konzeption von Datenmodellen für die Erfassung von Objekten orientiert werden kann.¹⁰ Als Anwendung des ISO-Standards CIDOC-Conceptual Reference Model ist er ereignisbasiert konzipiert, entspricht also der Logik einer Objektgeschichte von der Entstehung bis zur Zerstörung desselben, die als solche immer auch raum-zeitliche Verankerung findet.¹¹ Speziell für die Erfassung von Daten zu Objekten grafischer Sammlungen wurde mit dem *LIDO-Handbuch Graphik* ein Best-Practice-Modell entwickelt, das den LIDO-Standard auf Community-bezogene Anforderungen und Bedarfe hin angepasst vermittelt und für die praktische Umsetzung konkretisiert.¹² Für das Virtuelle Kupferstichkabinett ergab sich durch diese Entwicklung die Anforderung einer Anpassung des lokalen Datenmodells an die Struktur und Prinzipien von LIDO. Um also z.B. der Ereignisorientierung zu entsprechen und auch terminologische Konformität zu schaffen, wurde das lokale Datenmodell erweitert. Im Konkreten bedeutete dies, dass zum Beispiel Personen, die an der Entstehung eines Objektes beteiligt waren, nicht lediglich als »Künstler« eingetragen werden. Vielmehr sollte es möglich sein, ihnen eine spezifischere Rollenbezeichnung zuzuweisen, etwa als Zeichner der Vorlage in Abgrenzung zum Stecher der Platte. Dies ist nur ein Beispiel für die Anpassung des Datenmodells an die Anforderungen bestimmter Forschungsfragen, die eine Modellierung im Zuge der Datenintegration des Teilprojektes »Virtuelles Zeichnungskabinett« nach sich gezogen hat. So ergab sich weiterhin der Bedarf, dass es möglich sein sollte, eine Sammlermarke zu dokumentieren. Hierfür wurde ein Freitextfeld eingefügt. Damit erreichte man zwar eine Vertiefung der Erschließung, aber nicht unbedingt eine Verbesserung der Datenqualität für das Retrieval, da das Element nicht weiter strukturiert und letztlich auch für die Suche nicht indexiert wurde.

Veränderungen des Datenmodells haben zur Folge, dass ältere Datensätze entsprechend angepasst werden müssen. Gegebenenfalls kann hier eine Information nicht vorliegen, weil sie schlicht nicht erhoben wurde. Die intrinsische, also inhaltsbezogene Vollständigkeit der Daten ist somit nicht mehr gegeben. Für das genannte Beispiel der Sammlermarke muss allerdings ohnehin davon ausgegangen werden, dass viele Objekte der Sammlung gar keine solche aufweisen.¹³ Im

¹⁰ Der Standard wurde durch das ICOM International Committee for Documentation CIDOC entwickelt. Weitere Informationen sowie die Schemadefinitionen finden sich unter: www.lido-schema.org [30.06.2020].

¹¹ Auch das CIDOC CRM wurde in der genannten Arbeitsgruppe entwickelt. Vgl. Homepage CIDOC CRM, URL: www.cidoc-crm.org/ [30.06.2020].

¹² Vgl. Knaus/Stein/Kailus (2019).

¹³ Man könnte fragen, inwiefern es überhaupt nötig sei, diese Information recherchierbar zu machen. Hier kommen sowohl die Anforderungen der Zeichnungsforschung als auch die der

Fall der Rollendefinition für Personen hingegen kann es vorkommen, dass Datensätze ungenau werden, weil den verzeichneten Personen retrospektiv nicht ohne Weiteres konkrete Rollen zugewiesen werden können.

Im Datenlebenszyklus folgt der Phase der Konzeption die der Datenerhebung. Daten zu Objekten der materiellen Kultur werden in der Regel durch Autopsie der Objekte selbst oder anhand digitaler Bilder erhoben. Diese Erschließung der Objekte wird meist durch (kleine) Teams von wissenschaftlichen Hilfskräften umgesetzt. Wie Colum Hourihane in einem Beitrag betont, sollte diese Phase als besonders relevant eingeschätzt werden, da hier die Grundlage für jegliche Weiterverarbeitung der Daten gelegt wird.¹⁴ Hier entscheidet sich, welches Wissen zum Objekt überhaupt kommuniziert und publiziert werden kann. Am Beispiel der Graphik wird deutlich, wie voraussetzungsvoll diese Aufgabe ist: Erfasser*innen benötigen einen geschulten Blick für die Typisierung des Objektes, z.B. im Hinblick auf die Spezifikation der genutzten Drucktechnik: Handelt es sich um einen Kupferstich, eine Radierung, einen Holzschnitt oder eine Lithographie? Außerdem stellt sich die Frage, was das Bild zeigt. Was ist wie dargestellt? Gibt es für diese Darstellung Traditionen oder Vorbilder? Ist im Kupferstich eine Darstellung reproduziert oder handelt es sich um ein eigenständig für das Medium entwickeltes Werk? Für die inhaltliche Beschreibung eines Objektes als Bild bedarf es ikonographischen Wissens und die Beherrschung ikonologischer Methoden.¹⁵ Zudem sind materialkundliches und historisches Wissen hilfreich. Erfasser*innen steht in der Regel ein Apparat von Handbüchern und Nachschlagewerken zur Verfügung, die die Beschreibung auf ein wissenschaftlich gesicherteres Fundament stellen. Dennoch lässt sich hin und wieder die Recherche spezialisierter Literatur nicht vermeiden. Um den Ansprüchen einer wissenschaftlichen Community im Hinblick auf den Informationsgehalt eines Datensatzes entgegenzukommen, bedarf es also bereits bei der Erschließung des Engagements von Personen, die diese Community und deren Prioritäten (etwa durch das Studium) kennengelernt haben.

Diese Ansprüche werden darüber hinaus durch die jeweiligen Vorgesetzten vermittelt, wie etwa Mitarbeiter*innen, die mit der Redaktion der Datensätze betraut sind und meist auch die Aufgaben der Erfassung initiieren, steuern und kontrollieren. Sie sind mit der Prüfung der intrinsischen Datenqualität betraut, das heißt, hier geht es vor allem um die Fehlerfreiheit, Glaubwürdigkeit und gegebenenfalls Objektivität der erfassten Information sowie deren semantische Validität, also z.B.

Provenienzforschung zum Tragen, denen man aufgrund der spezialisierten Ausrichtung der Webanwendung gerecht werden wollte.

¹⁴ Vgl. Hourihane (2002), S. 40.

¹⁵ So basiert etwa der Beschreibungsstandard CDWA (Categories for the Description of Works of Art) auf dieser vom Kunsthistoriker Erwin Panofsky ausgearbeiteten Methode. Vgl. Harping (2002), S. 23.

ob die Angabe im richtigen Feld eingetragen wurde.¹⁶ Mitarbeiter*innen der Redaktion haben meist weitreichende Kenntnisse und Nutzungsrechte innerhalb des Erfassungssystems, die es ihnen erlauben, dessen Such- und Filterfunktionen für Qualitätschecks von Werten einzusetzen und Korrekturen zu steuern. Bei Bedarf wird vielfach an dieser Stelle das Mapping, also die Vorbereitung, Feldauswahl und Übersetzung des lokalen Datenmodells in ein anderes, wie z.B. LIDO, für das Harvesting vorgenommen. Für die darstellungsbezogene Datenqualität, welche die Konsistenz von Daten bezüglich ihrer Formate sowie deren Verständlichkeit und Interpretierbarkeit ihrer Bedeutung beschreibt, ist diese Arbeit ein entscheidender Faktor. Hier entscheidet sich, welche Information wie in andere Formate überführt werden kann. Die Datenredaktion ist zudem meist Anlaufstelle für Feedback von Nutzer*innen der publizierten Daten. Somit hat sie Einblicke in die zugangsbezogene Datenqualität, also z.B. ob zu gewünschten Informationen auch alle vorhandenen Datensätze gefunden werden. Zudem kann über dieses Feedback auch die Erwartungskonformität der Daten für bestimmte Zielgruppen bestimmt werden, also beispielsweise inwiefern die Information zu den erfassten Graphiken korrekt, aktuell und ausreichend ist.

Bei der Publikation von Daten sind häufig das Frontend der Anwendung und die Suchmöglichkeiten vorrangig diskutierte Fragen. Im Fall des Virtuellen Kupferstichkabinetts war es das Ziel der Fachcommunity über den einfachen Suchschlitz hinaus eine detaillierte Suche in allen Feldern zu ermöglichen, die eine sehr zielgerichtete Zusammenstellung der Suchfaktoren, also z.B. Künstler, Inhalt, Technik, Signatur und Provenienz erlaubt. Da die Anwendung auch in den Institutionen selbst von denjenigen Wissenschaftler*innen zur Recherche genutzt wurde, denen die Sammlungen bekannt waren, vermischten sich mit den allgemeinen, weniger granularen Anforderungen an ein Information Retrieval auch Anforderungen an eine Datenbank, die lokale Bedürfnisse zufrieden stellen sollte, wie etwa sehr genaue Abfragen seitens des Bibliotheks- oder Museumspersonals zur Beantwortung von Besucheranfragen. Für das Metadaten Assessment, welches diesem Beitrag zugrunde liegt, wurde die Relevanz der einzelnen Felder für die Suche genauer untersucht und dabei der Zugang zu den Objekten entsprechend der Ziele des Retrieval bewertet, nämlich Objekte zu finden, zu identifizieren und diese auszuwählen, um die zugehörigen Information einzusehen und schließlich auf diese auch zugreifen zu können, also z.B. das digitale Bild oder Metadaten herunterladen zu können.¹⁷ Die Publikation oder Zugänglichkeit von Datensets über Schnittstellen ist besonders im Hinblick auf das Teilen von Daten relevant und wird zudem vielfach als An-

¹⁶ Vgl. Rfill. (2019), S. A-12. Hier wird zwischen intrinsischer, kontextbezogener, darstellungsbezogener und zugangsbezogener Datenqualität unterschieden.

¹⁷ Vgl. Madison/Byrum (2006).

forderung der Mittelgeber vorgeschrieben. Entsprechend sind die Daten des VKK über eine OAI-PMH- Schnittstelle abrufbar.¹⁸

Die vorangegangenen Einblicke in die Erschließung und Ziele von Daten zu Graphik zeigen erstens, dass es verschiedene Interessens- bzw. Nutzer*innengruppen gibt, die jeweils eigene Anforderungen an die Daten stellen. Diese Anforderungen verändern sich im Laufe der Zeit, da sich Technologien, Kontexte oder Akteur*innen weiterentwickeln. Beispielsweise werden Daten aus Sammlungen verstärkt für interdisziplinäre Zusammenhänge genutzt oder sind kommerziell interessant.¹⁹ Ebenso sind, zweitens, die Daten selbst dynamischer Natur, was schon allein ihr wissenschaftlicher Kontext mit sich bringt. Diese Tatsache widerspricht der bislang noch vielfach vorhandenen Einschätzung zur Veröffentlichung von Informationen und Bildern von Sammlungsobjekten, mit dem Erscheinen der Daten auf einer Website sei deren Publikationsprozess abgeschlossen, letzterer sei also vergleichbar mit der Produktion eines gedruckten Katalogs. Drittens schlagen sich Anforderungen, die an Daten gestellt werden, in der Produktion sozio-technischer Artefakte nieder. So musste beispielsweise das Service-Angebot des VKK um persistente Identifikatoren erweitert werden, um die dauerhafte Zitierbarkeit der Daten zu gewährleisten, oder auch um die erwähnte Schnittstelle für das Harvesting. Weniger offensichtlich für externe Nutzer*innen sind Veränderungen des Datenmodells über die Zeit oder Anpassungen der Erfassungspraxis.

Daten als Ressourcen

Werfen wir einen vergleichenden Blick auf andere Ressourcen, die in medienhistorischer Perspektive die Vorgänger zu Daten und Softwaresystemen in Museen bilden. Sammlungen werden in Museen hergestellt, indem Objekte unterschiedlichster Herkunft und Bedeutung nicht nur räumlich dort versammelt werden. Sie werden in Objektklassen unterteilt, nach Materialität sortiert und schließlich in bestimmte Ordnungs- und Wissensstrukturen eingebunden. Ihnen werden darüber hinaus andere materielle Dinge hinzugefügt, beispielsweise Passepartouts.²⁰ Die

¹⁸ Um den Austausch von Metadaten zu gewährleisten, wurde von der Open Archives Initiative ein Harvesting Protokoll entwickelt. OAI-PMH steht für Open Archives Initiative Protocol for Metadata Harvesting. Vgl. Homepage Open Archives Initiative, URL: <https://www.openarchives.org/pmh/> [30.06.2020].

¹⁹ Vgl. RfII (2019), S. 11. Auf neue Anforderungen, die insbesondere die Digital Humanities mit sich bringen, wird auch im Paper zum Stand der Kulturgutdigitalisierung hingewiesen: Klaffki/Schmunk/Stäcker (2018), S. 26.

²⁰ Verschiedene Studien thematisieren die materielle Kultur des Ansammelns und Aufbewahrens, die allerdings vielfach in den unzugänglichen Archivräumen der Institutionen verborgen bleibt, vgl. te Heesen/Spary (2001) und te Heesen/Michels (2007).

Sammlungsobjekte durchlaufen dabei einen Wandel von Nutzungskontexten. Eine Lithographie, die als Bildpublikation gedruckt wurde, wird zum historisch und künstlerisch relevanten Gegenstand, indem sie dem Archivraum der graphischen Sammlung zugefügt wird. Der Archivierungsprozess produziert also ein museales Objekt. André Malreaux beschreibt diesen Transformationsprozess als Entstehung eines Mediums: »[...] dafür hat es [das Museum] nur Abbilder von Dingen, die etwas anderes als die Dinge selbst sind und gerade auf diese spezifische Unterschiedenheit ihre Daseinsberechtigung gründen.«²¹ Ein Besuch des Studiensaals einer Graphischen Sammlung beginnt oft mit dem Griff in einen Handapparat, wo sich meist neben Handbüchern und Nachschlagewerken auch die Publikationen des Hauses sowie Sammlungs- und Ausstellungskataloge befinden. Ihnen können erste Informationen zur Sammlung und zu den Objekten entnommen werden. Beim Durchblättern stößt man gegebenenfalls auf weitere interessante Objekte. Mitarbeiter*innen stehen als direkte Ansprechpartner *innen zur Verfügung und liefern die gewünschten Originale aus dem Depot. Sie nutzen Zettelkataloge oder Inventare zur Standortermittlung. Die auf Basis der eigenen Sammlung produzierten Produkte werden also in funktionale Abläufe alltäglicher musealer Tätigkeiten eingebunden. Als Sammlungsinstitutionen mit Bildungsauftrag haben Museen ein Interesse an der Aufbewahrung ihrer eigenen Produkte, weil sie sowohl repräsentative als auch praktische Funktionen erfüllen. Die genannten Medien sind somit auf die Ziele des Findens, Identifizierens, Auswählens und des Zugänglichmachens ausgerichtet.

Wenn Museen in den digitalen Raum erweitert werden, gilt es bisherige museale Strategien in Bezug auf den Umgang mit den eigenen Produkten für die dortigen Umstände anzupassen. Für digitale Medien, wie Daten zu Sammlungsobjekten oder Anwendungen, bedeutet dies, dass sie eines institutionellen strategischen Rahmens bedürfen, der ihren Lebenszyklus und ihre dynamische Natur berücksichtigt. Die Qualität von Daten bemisst sich am Kontext und den Anforderungen ihrer aktuellen und auch zukünftigen Nutzung. Daten zu Sammlungsobjekten können als eine Ressource ähnlich der Museumssammlung begriffen werden, die innerhalb eines »virtuellen Museums« die Möglichkeit immer neuer Interpretationen bieten und als Basis für weitere Produkte dienen können. Diese Ressource will verwaltet, gepflegt und genutzt werden. Weder für Daten noch für die genutzten Systeme haben sich bislang ausreichende qualitätssichernde Maßnahmen und Prozesse etabliert.

²¹ Siehe Malreaux (1956), S. 10.

Deins, Meins, Unser

Ein zentraler Aspekt für (digitale) Produkte ist die Frage danach, wer sie nutzt oder nutzen soll. Wenn wir für Kulturinstitutionen wie Museen oder Bibliotheken die Anforderung erhoben haben, dass ihre Produkte sowohl von ihnen selbst als auch von anderen nutzbar sein sollen, können wir von »UsWare« sprechen.²² In Abgrenzung dazu, gibt es an vielen Institutionen Software, die nur für interne Zwecke und ein einziges Ziel genutzt wird. Darunter fallen beispielsweise Filemaker-Datenbanken oder Excel-Tabellen, deren Nutzung sich innerhalb einer bestimmten Praxis etabliert hat, die aber nur für Mitarbeiter*innen zugänglich sind und deren Kategorien nur im Kontext der Sammlung verständlich sind. Solche »MeWare« muss und soll für andere gar keinen Sinn ergeben. Lediglich die Entwickler*innen müssen damit umgehen können. In zahlreichen Fällen ist proprietäre Museumsmanagement-Software wiederum Beispiel für »ThemWare«. Diese wurde von Entwickler*innen für eine bestimmte Zielgruppe geschaffen, zu der sie selbst nicht gehören. Zentral für unsere Fragestellung ist, dass mit der Kategorie der »UsWare« idealerweise eine Qualitätssicherung des Produkts einhergeht: Wenn es für alle Seiten nützlich ist, haben alle ein Interesse daran, seine Qualität zu sichern oder zu verbessern. Damit ändert sich bestenfalls das Verhältnis der einzelnen Person zum Produkt als etwas Eigenes. In der Diskussion um die Strategie des Dogfooding unter Software-Entwickler*innen,²³ also die Nutzung des eigenen Produkts mit dem Ziel der Qualitätssicherung, finden sich einige kritische Gegenstimmen. Zu den negativen Mechanismen gehöre z.B., dass man sich mit dem eigenen Produkt zufriedengeben würde, weil man keine Lust oder auch Zeit habe daran noch weiter zu arbeiten. Die Konfrontation mit dem eigenen Produkt befördere nicht gerade kritisches Denken, weil die Auseinandersetzung mit Fehlern im System nicht zuletzt neue Arbeit generiere.²⁴ Zudem könne es dazu führen, Standards zu missachten oder einfach eigene zu entwickeln. Außerdem spielt der Faktor Zeit eine nicht zu unterschätzende Rolle, wenn man parallel zu neuen Produkten auch noch bisherige optimieren solle.²⁵

²² Referenziert wird hier auf einen Blogbeitrag des Entwicklers Eric Sink, der darin Software in drei verschiedene Kategorien unterteilt. Vgl. Sink (2006).

²³ Der Begriff wurzelt in einer Überzeugungsstrategie des Marketings, nach der das Produkt so gut ist, dass der Hersteller es selbst konsumiert. Vgl. Wikipedia, Eating your own dog food, URL: https://en.wikipedia.org/wiki/Eating_your_own_dog_food [30.06.2020]. In der Software-Branche wurde der Begriff durch Microsoft auf Test- und Qualitätsmanagement-Verfahren für eigene Software übertragen. Vgl. Homepage Codinghorror, URL: <https://blog.codinghorror.com/the-difficulty-of-dogfooding/> [30.06.2020].

²⁴ Vgl. Harrison (2006).

²⁵ Vgl. Dvorak (2007).

Vorteilhaft daran sein eigenes Produkt zu nutzen ist, dass Dogfooding ein zirkuläres Prinzip ist, an dem verschiedene Nutzer*innen-Gruppen teilhaben. Sie vertreten unterschiedliche Interessen und formen dadurch die Gestalt digitaler Objekte. Durch Feedback von allen Seiten kann eine Software oder können Daten auch für alle optimiert werden. Durch die fortwährende Nutzung entsteht der Druck, eventuelle Qualitätsmängel tatsächlich zu beheben. Das macht es zu einem komplexen Prozess, der allerdings der Komplexität digitaler Produkte entspricht. Software ist nicht linear, kein Teil gleicht dem anderen, was dazu führt, dass das »Verstellen eines Reglers« immer Auswirkungen auf das gesamte System hat. Dogfooding ist nicht nur eine plane Feedbackschleife, sondern erlaubt mehrere Perspektiven auf einen vieldimensionalen Gegenstand. Eine Software oder Daten müssen in vielen Interfaces funktionieren, die für sehr verschiedene Nutzer*innen gedacht sind. Dogfooding bezieht noch eine weitere essentielle Eigenschaft von Software ein: ihre Wandelbarkeit. »All successful software gets changed.«²⁶ Wenn eine Software sowohl von ihren Entwickler*innen als auch von anderen Zielgruppen genutzt werden kann, ist es nicht abwegig, dass sich über die Zeit noch weitere Nutzungsbereiche für die Funktionen der Software ergeben. Diese Nutzbarkeit macht sie nachhaltig. Durch Dogfooding können Entwickler*innen zudem auf Neuentwicklungen schneller reagieren und so Problemen durch eventuelle Umstellungsprozesse im Vorfeld entgegenwirken.

Datenqualität – aber wie?

Wenngleich die Welt der Museen und Bibliotheken in Bezug auf die Nutzung von Software und die Produktion qualitätvoller Daten von einem solchen Idealzustand aus vielerlei Gründen weit entfernt ist, so lohnt es sich, die geschilderten Prinzipien als ideelle Basis zu sehen. Zwar ist ein Großteil der Institutionen auf Service-Leistungen Dritter (ThemWare) angewiesen, allerdings können zumindest für die eigenen Produkte (Daten) Anforderungen an UsWare geltend gemacht werden.

Ein zentraler Aspekt der Qualitätssicherung ist konsequente Standardisierung. Für Daten zu Sammlungsobjekten wie Graphik und Zeichnungen bedeutet dies z.B. den Einsatz normierter Vokabulare und Thesauri schon bei der Erfassung. Dies ist heute an den meisten Institutionen üblich. Zur Anpassung der Qualität älterer Datensätze empfiehlt sich die wiederkehrende automatisierte Aktualisierung lokaler Vokabulare, die über Schnittstellen oder Tools möglich ist.²⁷ In der Phase der Konzeption von Daten gilt es zudem ebenfalls auf Standards der Modellierung

²⁶ Siehe Brooks (1995), S. 4.

²⁷ Zur Anreicherung von Daten wird beispielsweise vielfach Open Refine genutzt, welches über Schnittstellen zu vielen Normdatenrepositorien (z.B. zur CND) verfügt.

zurückzugreifen. Für Daten zu Kulturobjekten ist seit vielen Jahren das bereits erwähnte CIDOC Conceptual Reference Model ein wichtiger Standard. Hierauf basiert auch das Harvestingformat LIDO, welches für den Austausch von Daten des Virtuellen Kupferstichkabinetts genutzt wird. Die dortigen Erfahrungen haben gezeigt, dass es sinnvoll ist, die Phase der Konzeption regelmäßig zu wiederholen, da auch Standards sich weiterentwickeln. Wenn sie zu Beginn eines Digitalisierungsprojektes noch nicht etabliert waren, können sie sich nach fünf oder zehn Jahren bereits durchgesetzt haben. Die Datenmodellierung muss entsprechend überdacht und ggf. angepasst werden, um die Offenheit der Daten für Anwendungen und damit verbunden auch ihre Qualität weiterhin zu gewährleisten.

Hier zeigt sich zudem die Relevanz von digitalen Strategien als Rahmenbedingung für qualitätvolle Daten und Services, die immer wieder aktualisiert werden sollten. Wenn darin wiederkehrende Prozesse der Qualitätssicherung in allen Phasen des Datenlebenszyklus berücksichtigt werden, bedeutet dies gegebenenfalls die Anpassung von Aufgabenfeldern und technischer wie personeller Ressourcen. Nicht nur für die Konzeptionsphase in Drittmittelprojekten, sondern auch für eine digitale Strategie ist die Orientierung an Datenmanagementplänen sinnvoll. Hier gestellte Ausgangsfragen sind z.B., welche Daten wie gesammelt werden, was und wie dokumentiert wird, welche rechtlichen Rahmenbestimmungen gelten sollen, wie Daten geteilt werden und nicht zuletzt, wer für das Datenmanagement zuständig ist.²⁸

Ausgehend vom Beispiel des Virtuellen Kupferstichkabinetts kann an dieser Stelle ein Szenario umrissen werden, das die angesprochenen Problematiken und Anforderungen berücksichtigt: Da eine auf Graphik spezialisierte Datenbank in einer Bibliothek oder im Museum nur eines von mehreren Webangeboten sein kann, könnte zunächst der Zugriff auf objektbezogene Daten vereinheitlicht werden. Jegliche an der jeweiligen Institution produzierten Objektdaten könnten über ein Datenbanksystem zugänglich sein. Als weiteres Prinzip basiert dieses Szenario auf einer Funktionstrennung von basalen Daten und Forschungsdaten. Das Datenmodell für Basisdaten bezieht lediglich die zur Identifikation und zum Auffinden des Objektes notwendigen Informationen ein, enthielte also z.B. eine Objektklassifikation, einen Titel oder inhaltliche Beschreibung sowie eine Signatur. Obligatorisch wäre hier die Nutzung normierten Vokabulars für die Beschreibung. Damit würde Information, die essentiell ist, um ein Objekt zu finden, und deren Faktizität sich kaum oder nur sehr langsam verändert, gebündelt werden. An einer Bibliothek könnten so Daten zu druckgraphischen Objekten ebenso wie Bücher und andere Medien über ein Katalogisierungssystem erschlossen, verwaltet und aufgefunden

28 Vgl. Homepage des Digital Curation Centre, URL: www.dcc.ac.uk/dmponline [30.06.2020]; vgl. RfII (2019), S. 43.

werden. Für eine Repräsentation in einer anderen Webanwendung bzw. eine forschungsorientierte Erschließung könnten solche Basisdaten über eine Schnittstelle abgerufen und weiterentwickelt werden. Diese Forschungsdaten bedeuten eine vertiefte Erschließung des Gegenstands im Kontext eines Projekts mit einer forschungsspezifischen Fragestellung. Das repräsentierte Wissen weist eine höhere Dynamik auf, weil es weniger gefestigt ist. Für eine Zeichnung würde der Forschungsstand etwa durch Fragen der Zuschreibung des Werks, die Provenienz des Objektes oder die Spezifikation der Technik repräsentiert. Zudem würden bei der Modellierung die Darstellbarkeit der Beziehung des Objektes zu anderen berücksichtigt. Für Druckgraphik ist beispielsweise die Zugehörigkeit zu Serien oder Vierlagen relevant. Für die Struktur solcher Forschungsdaten empfiehlt sich die Orientierung am CIDOC CRM oder an LIDO, um die Interoperabilität und Weitergabe zu befördern. Zudem müsste hier der Grad der Aktualität der Angaben angezeigt werden.

Bei dieser sich auch in der Softwarearchitektur der Institution manifestierenden Funktionstrennung bleibt allerdings die Frage offen, wie Validität und Aktualität der Daten beider Ausrichtungen zu synchronisieren wären. Vorteilhaft an dieser Form der Funktionstrennung wäre, dass die Basisdaten den Mindestanforderungen für die Nutzung innerhalb der Institution, nämlich das Auffinden des Objektes in der Sammlung, ebenso entspreche, wie sie offen für verschiedene Anwendungsszenarien bliebe. Durch die konsequente Nutzung des eigenen Basisdatenbestandes im Sinne des Dogfooding könnten Qualitätsmängel schneller ausgemacht und behoben werden. Datenbestände könnten darüber hinaus systematisch entwickelt werden. Eine konsequente Qualitätssicherung wirkt sich auf Arbeitsprozesse und Personalentscheidungen aus. Es ergeben sich Aufgaben, wie die Analyse und Kuration von Daten, die kaum nebenbei erledigt werden können. Zentral für dieses Szenario ist daher die Rolle eines/r Datenkurator*in. Diese Person vermittelt zwischen verschiedenen Akteur*innen und Anforderungen und hat aktuelle technologische Entwicklungen ebenso im Blick wie Bedarfe von Forscher*innen oder Bibliothekar*innen. Sie berücksichtigt Anforderungen von Nutzer*innen ebenso wie die Möglichkeiten der technologischen Infrastruktur.

Resümierend bleibt festzuhalten, dass die Produktion und Sorge um die eigenen digitalen Objekte für Kulturinstitutionen eine langfristig zu integrierende Aufgabe darstellt. Das Bewusstsein um die dynamische Natur digitaler Objekte und die nutzungsorientierte Bewertung ihrer Qualität erfordert die stärkere Einbindung computergestützter Qualitätsmanagementprozesse in allen Phasen des Datenlebenszyklus.²⁹ Dabei müssen solche Prozesse teils noch entwickelt und ihre

²⁹ Der Erforschung von Möglichkeiten und Erprobung von Techniken für einen kontinuierlichen Qualitätsmanagementprozess, der alle Bereiche des Datenlebenszyklus berücksichtigt, wid-

Praktikabilität im Alltag überhaupt erst evaluiert werden. Die geschilderten Prinzipien und Ideen können daher nur Ausgangspunkte für eine Diskussion bilden, die es zukünftig weiterzuführen gilt. Es zeigt sich anhand unseres Beispiels, dass die computergestützte Erfassung und Verwaltung von Sammlungen langfristig auch Konsequenzen für die gesamte Institution mit sich bringt.

Literatur

- Brooks, Frederick P.: No Silver Bullet – Essence and Accident in Software Engineering, in: The Mythical Man-Month – Essays on Software Engineering, Crawfordsville 1995, S. 1-16; verfügbar unter: <https://doi.org/10.13225/j.cnki.jccs.2017.1345> [30.06.2020].
- Chenhall, Robert G.: Information Retrieval and Systems Analysis, in: The Metropolitan Museum of Art (Hg.): Computers and Their Potential Applications in Museums, New York 1968, S. 59-79.
- Dvorak, John C.: The Problem with eating your own dogfood, in: Homepage PCMag.Com, 11.05.2007; verfügbar unter: <https://web.archive.org/web/20090628161803/www.pcmag.com/article2/0,2817,2217007,00.asp> [30.06.2020].
- Harping, Patricia: The Language of Images: Enhancing Access to Images by Applying Metadata Schemas and Structured Vocabularies, in: Introduction to Art Image Access, 2002, S. 20-39.
- Harrison, Warren: Eating your own dogfood, in: IEEE Software, May/June 2006, Vol. 23, S. 5-7.
- Hourihane, Colum: It Begins with the Cataloguer: Subject Access to Images and the Cataloguers Perspective, in: Introduction to Art Image Access, Los Angeles 2002, S. 40-58.
- Klaffki, Lisa/Schmunk, Stefan/Stäcker, Thomas: Stand der Kulturgutdigitalisierung in Deutschland – Eine Analyse und Handlungsvorschläge des Dariah-DE Stakeholdergremiums »Wissenschaftliche Sammlungen«, Dariah-DE Working Papers, 2018.
- Madison, Olivia/Byrum, John: Funktionelle Anforderungen an Bibliografische Datensätze – Abschlussbericht IFLA Study Group, 2006; verfügbar unter: http://infoserv.inist.fr/wwsympa.fcgi/d_read/frbr/FRBR_bibliography.rtf [30.06.2020].
- Malreaux, André: Das Imaginäre Museum, in: ders.: Stimmen der Stille, München 1956.

met sich seit 2019 das Verbundprojekt KONDA am DDK-Foto Marburg, der Universität Marburg sowie der SUB Göttingen.

- Pagel, Christiane: Virtuelles Kupferstichkabinett – Druckgraphik der Frühen Neuzeit Online, in: Wolfenbütteler Notizen zur Buchgeschichte, 26 (2011), S. 91-104.
- RfII, Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel, Göttingen 2019.
- Rojas, Roberto/Crespán, José Luis/Trallero, Manuel/Vrine-Bohan, Hugues: Museen der Welt – Vom Musentempel zum Aktionsraum, Hamburg 1977.
- Sink, Eric: Yours, Mine and Ours (2006); verfügbar unter: https://Ericsink.Com/Articles/Yours_Mine_Ours.Html [30.06.2020].
- te Heesen, Anke/Michels, Anette: Auf/Zu. Der Schrank in den Wissenschaften, Berlin 2007.
- te Heesen, Anke/Spary, Emma (Hg.): Sammeln als Wissen – Das Sammeln und seine wissenschaftsgeschichtliche Bedeutung, Göttingen 2001.
- Verwayen, Harry/Arnoldus, Martijn/Kaufman, Peter B.: The Problem of the Yellow Milkmaid. A Business Model Perspective on Open Metadata, 2011.