

Die Glaubwürdigkeit des Unbekannten – Modellierung von unvollständigem und dynamischem Wissen in LIDO

Julia Rössel^a | Michelle Weidling^b | Viola Wenz^c

a Abteilung Informationstechnik/Projekt KONDA, Deutsches Dokumentationszentrum für Kunstgeschichte – Bildarchiv Foto Marburg, Philipps-Universität Marburg, Deutschland, roesselj@fotomarburg.de; b Abteilung Forschung und Entwicklung/Gruppe Digital Humanities, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Deutschland, weidling@sub.uni-goettingen.de; c Projekt KONDA, Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, Deutschland, viola.wenz@uni-marburg.de

KURZDARSTELLUNG: Wissen zu Objekten der materiellen Kultur ist von Unsicherheiten, Unschärfe und Dynamik geprägt: Oftmals können Forschende keine zuverlässige Aussage zu einem Sachverhalt treffen oder Fakten, die bisher als gesichert galten, müssen aufgrund neuer Befunde revidiert werden. Dieser Sachverhalt sollte auch Eingang in Forschungsdaten finden, um die Transparenz gegenüber Nutzer*innen dieser Daten zu erhöhen: Erst wenn deutlich gekennzeichnet wird, dass die Aussage in einem Datenfeld nach aktuellem Forschungsdiskurs nicht vollständig gesichert ist, werden Suggestionen vermieden und Glaubwürdigkeit geschaffen. Standardisierte Datenformate wie LIDO bieten bereits jetzt Ansätze, um Unschärfen in der Wissenserschließung zu kennzeichnen; Hierfür liegen jedoch in der Regel keine kontrollierten Vokabulare vor, sodass wiederum semantisch bedingte Unsicherheiten auftreten können. Um in dieser Problematik Abhilfe zu schaffen, werden im Rahmen des Projekts KONDA Empfehlungen erarbeitet, sodass Dynamik und Unschärfe in LIDO ausgedrückt werden können und so ein Beitrag zur Glaubwürdigkeit der Forschungsdaten geleistet wird.

1. EINFÜHRUNG

Gedächtnisinstitutionen wie Museen oder wissenschaftliche Sammlungen produzieren und gestalten aktiv Wissen zu Objekten der materiellen Kultur sowie deren Rezeption, indem sie Informationen zu diesen Objekten erschließen und disseminieren. Damit stehen sie in der Verantwortung, für die Verlässlichkeit und Korrektheit der bereitgestellten Daten zu sorgen.

Aufgrund laufender Forschungsdiskurse und technischer Weiterentwicklungen, die neues Wissen zu einem Objekt generieren bzw. ermöglichen können, sind Daten zu materiellen Sammlungsobjekten stark dynamischer Natur. Vom „einfachen“ Inventarbuch-Eintrag sind viele Datensätze heute weit entfernt; Sie enthalten mitunter detailliertes Wissen, sind von ihren Entstehungskontexten geprägt und unterliegen sich wandelnden Anforderungen und verändertem Wissen zum Gegenstand, den sie repräsentieren.

Auch die Dimension des Nicht-Wissens trägt zur Komplexität der Daten bei: Wissen über

Objekte der materiellen Kultur ist aufgrund ihrer Historizität oftmals unvollständig und unsicher, etwa wenn ein Gemälde einem Künstler zugeschrieben wird, die Urheberschaft aber bisher nicht zweifelsfrei geklärt werden kann. In solchen Fällen müssen Unsicherheiten über das präsentierte Wissen auf Datenebene kenntlich gemacht werden. Dies erfolgt idealerweise in expliziter und standardisierter Form, wird in der Praxis aber häufig in lokalen, nicht standardkonformen Datenmodellen umgesetzt.

Infolge dessen stehen Gedächtnisinstitutionen vor der Herausforderung, Dynamiken und Unsicherheiten in ihren Objektdaten eindeutig darzustellen, um die historische sowie wissenschaftliche Komplexität ihres Materials zu erfassen und ihrem Anspruch auf Glaubwürdigkeit als Bildungsinstitutionen im digitalen Raum gerecht zu werden. Sie müssen daher (insbesondere aufgrund der vorliegenden Komplexität) sicherstellen, dass die von ihnen öffentlich bereitgestellten Daten regelmäßig in Bezug auf ihre Qualität überprüft und aktualisiert werden. Nur so können sie letztlich auch zuverlässig und glaubwürdig sein.

2. MODELLIERUNG VON DYNAMISCHEM UND UNVOLLSTÄNDIGEM WISSEN IN LIDO

Die Pluralität und Diskursivität, aber auch die Dimension des „Noch-nicht-Wissens“ in Forschungsdaten abzubilden und damit einen glaubwürdigen Ausschnitt der Realität zu repräsentieren, ist eine wesentliche Herausforderung im Umgang mit geistes- und kulturwissenschaftlichen Daten. Dieser Thematik nimmt sich das BMBF-geförderte Verbundprojekt der Universitäten Göttingen und Marburg „KONDA – Kontinuierliches Qualitätsmanagement von dynamischen Forschungsdaten zu Objekten der materiellen Kultur unter Nutzung des LIDO-Standards“ an.

2.1 KONDA

Innerhalb des Projektes wird die Frage der Qualitätssicherung auf drei Ebenen angegangen: Grundlage bilden die Forschungsdaten des Deutschen Dokumentationszentrums für Kunstgeschichte – Foto Marburg (DDK) und der Sammlungen der Universität Göttingen, die teilweise in lokalen Datenformaten vorliegen. Diese Daten werden auf Anforderungen hin untersucht, die spezifische Daten zu Objekten der materiellen Kultur (Architektur, Malerei, Zeichnungen, ethnographische Objekte, ...) bei deren Erschließung mit sich bringen. Die zweite Ebene bildet die Standardisierung und Verbreitung dieser Daten. Ziel ist es, LIDO unter Berücksichtigung der erfassten Anforderungen weiter zu entwickeln. Auf der dritten Ebene werden Anforderungen für generische Qualitätssicherungsprozesse von unsicheren und vagen Daten eruiert und Möglichkeiten ihrer Umsetzung erforscht. Auf diese Weise sollen die für geistes- und kulturwissenschaftliche Forschungsdaten entwickelten Qualitätssicherungsprozesse auch für andere Wissenschaftsdisziplinen nutzbar gemacht werden.

2.2 DYNAMIK

Die Weiterentwicklung von Wissen kann multiple, manchmal auch parallel stattfindende Ursachen haben. So zeigt sich beispielsweise am Fall des Fundes bzw. der Neuzuschreibung von Zeichnungen aus dem Bestand des Kupferstichkabinetts der Kunsthalle in Karlsruhe 2015, wie stark bereits der Kontext der materiellen Sammlung sich auf die Rezeption von Sammlungsobjekten auswirken kann. Die Zeichnungen, die heute als Produkte des italienischen Kupferstechers und

Architekten Giovanni Battista Piranesi (1720–1778) und seiner Werkstatt gelten können, wurden bis zu ihrer Entdeckung in Klebebänden des deutschen Architekten Friedrich Weinbrenner (1766–1826) als Produkte desselben gesehen. Bis dato schien durch den Kontext des Sammelalbums kein Zweifel an der Autorschaft Weinbrenners bestanden zu haben.[1] Provenienz und Sammlungskontext boten also den Hintergrund, der das Wissen über die Zeichnungen zu manifestieren schien. Man vertraut ihnen, doch durch die wiederkehrende Auseinandersetzung mit den Sammlungsgegenständen entstehen immer neue Fragen, die zu solchen Aktualisierungen wie der Zuschreibung an einen anderen Künstler führen können. Die Dynamik des Wissens besteht hier also darin, dass bisher kolportiertes Wissen durch eine neue These in den Diskurskreis der Forschungsgemeinschaft katapultiert wird und, sofern sie mit soliden Argumenten untermauert ist, den Blick auf das Objekt und das Wissen darüber verändert und bisher sicher Geglaubtes in Frage stellt.

Zudem besteht die Dynamik des Wissens zu Objekten der materiellen Kultur auch darin, dass es sich ganz einfach vermehrt. Durch die Arbeit von Forscher*innen zu einem Objekt können neue Aspekte zum Entstehungskontext des Werkes desselben aufgedeckt werden, oder die Geschichte des individuellen Objektes wird durch Erkenntnisse im Bereich der Provenienzforschung erweitert. Eine Anreicherung von Information besteht z.B. bereits dann, wenn die Signatur eines Objektes verändert wird. Da die nunmehr veraltete Information darüber wahrscheinlich noch anderweitig zugänglich ist (z.B. in alten Sammlungskatalogen), wo sie nicht ohne Weiteres ersetzt werden kann, muss sie erhalten bleiben und in den Objektdaten als falsch bzw. veraltet gekennzeichnet werden.

Innerhalb des fachlichen Diskurses verändern sich gerade auch im Moment technologischen Wandels die Interessenlagen zu materiellen Objekten. So sind mittlerweile Fototheken, welche ursprünglich dem Fachpublikum zum Auffinden von Abbildungen kunst- und kulturhistorischer Objekte dienten, gewissermaßen Bilderbibliotheken waren, heute Orte, an denen historische Fotografien lagern und nunmehr musealisiert werden. Die Forschung

nimmt heute nicht mehr nur die Bildinhalte, die dort gefunden werden können, sondern ebenso die Geschichte der Fotos und deren Entstehungskontexte in den Blick.[2]

Auf der Ebene von Forschungsdaten ergibt sich durch die beschriebenen Dynamiken die Anforderung, dass die in Form von Datensätzen vorliegende Information stetig aktualisiert und erweitert werden muss, um den aktuellen Erkenntnisstand abzubilden. Datenmodelle müssen also sowohl der Komplexität der aktuellen Information angemessen, aber auch flexibel gestaltet sein, um die beschriebene Dynamik des Wissens auffangen zu können.

2.3 UNSICHERHEIT

Die oben beschriebene Dynamik des Wissens ist ein Grund für das (nutzerseitige) Misstrauen, welches Forschungsdaten zu Objekten der materiellen Kultur entgegengebracht wird. Einen weiteren Grund können wir in der Unsicherheit des Wissens zu Objekten verorten.

Eine Ebene von Unsicherheit kann als Objekten der materiellen Kultur inhärente Unsicherheit klassifiziert werden. Sie ergibt sich aus der Einzigartigkeit des Objektes und Lücken in der Objektgeschichte, die durch mangelnde Quellen entstehen und seitens der Forschung nicht geklärt werden können. Ein Beispiel hierfür ist der bzw. die Erschaffer*in eines Faustkeils, der bzw. die voraussichtlich für immer unbekannt bleiben wird.

Eine weitere Ebene der Unsicherheiten entsteht durch den Erfassungskontext. Das den Erfasser*innen individuell zugängliche Wissen bildet in diesem Fall einen kleineren Ausschnitt des Forschungsstandes, der mitunter unvollständig ist. Gründe hierfür können z.B. in der Unzugänglichkeit einer wichtigen Publikation oder mangelnde Fremdsprachenkenntnisse des Forschenden liegen, durch die Informationen zu einem Objekt während der Erfassung nicht zur Verfügung stehen. Doch auch die zeitlichen und finanziellen Rahmenbedingungen von Digitalisierungsprojekten können dazu führen, dass nicht alles vorhandene Wissen in einen Datensatz aufgenommen wird.

Eine dritte Quelle von Unsicherheit bildet die Technologie-Ebene. Hier entsteht Unsicherheit

durch Mängel in der technischen Umsetzung. Durch Konversionen der lokalen Erschließungsformate in Austauschformate wie LIDO kann es durch Fehler in Crosswalks zu Informationsverlusten kommen, sodass ursprünglich genau erfasstes Wissen an Schärfe verliert.

2.4 ABBILDUNG VON DYNAMIK UND UNSICHERHEIT IN LIDO

LIDO (Lightweight Information Describing Objects) ist ein XML-basiertes Schema, das für den Austausch und das Harvesting von Metadaten zu Objekten der materiellen Kultur entwickelt wurde.[3] Es liegt seit 2010 in einer ersten Version (v1.0) vor und wird kontinuierlich unter der Aufsicht der CIDOC LIDO Working Group weiterentwickelt. LIDO hat sich seit seiner Veröffentlichung zu einem internationalen Standard für die Bereitstellung von objektbezogenen Metadaten durch Gedächtnisinstitutionen etabliert und wird z.B. von der Deutschen Digitalen Bibliothek (DDB)[4] und Europeana[5] verwendet.

Als XML-Schema bietet LIDO die Möglichkeit, diverse Informationen zu einem Objekt strukturiert abzubilden. Da LIDO insbesondere in Hinblick auf die Anforderungen aus dem GLAM-Bereich (*galleries, libraries, museums, archives*) entwickelt wurde, ermöglicht es das Schema, Unsicherheiten und Unschärfen abzubilden. So kann z.B. eine Maßangabe wie die Höhe eines Bauwerks als geschätzt und damit als ungenau markiert werden:

```
<lido:objectMeasurements>
  <lido:measurementsSet>
    <lido:measurementType xml:lang="en">
      height
    </lido:measurementType>
    <lido:measurementUnit xml:lang="en">
      m
    </lido:measurementUnit>
    <lido:measurementValue>
      43
    </lido:measurementValue>
  </lido:measurementsSet>
  <lido:qualifierMeasurements
    xml:lang="en">
    approximate
  </lido:qualifierMeasurements>
</lido:objectMeasurements>
```

Ähnlich lassen sich auch nur ungefähr ermittelbare Datumsangaben mit LIDO ausdrücken:

```
<lido:date>
  <lido:earliestDate>
    1840
  </lido:earliestDate>
  <lido:latestDate>
    1860
  </lido:latestDate>
</lido:date>
```

Ein Blick in die Spezifikation von LIDO[6] verdeutlicht jedoch schnell mögliche Probleme bei der Darstellung von Dynamiken und Unsicherheiten in LIDO: Erstere abzubilden ist in der aktuellen LIDO-Version nicht möglich. Der Diskursivität von Wissen kann derzeit also nicht Rechnung getragen werden. Die Modellierung von Unsicherheiten in strukturierter Form beschränkt sich bislang auf die beiden genannten LIDO-Elemente (`lido:qualifierMeasurements` und `lido:earliestDate/lido:latestDate`). Ob ein Bildhauer nur als Erschaffer einer Plastik angenommen wird oder nicht, müsste nach aktuellem Stand im Freitextfeld angegeben werden und ist somit nicht mehr explizit und strukturiert abfragbar. Auch schon bei der Betrachtung von `lido:qualifierMeasurements` fällt ins Auge, dass das Schema zwar Beispielwerte wie “approximate” für den Qualifier angibt. Entscheiden sich aber Institutionen für unterschiedliche Werte wie “ca.”, “circa”, “approx.”, “approximate”, “*”, “?”, ergibt sich durch die Erschließung für Nutzer*innen der Objektdaten noch eine weitere Art der Unsicherheit – die der Semantik, wie Unsicherheiten ausgedrückt werden. Wird mit “circa” in einer Institution eine andere Form der Unschärfe referenziert als mit “approximate”? Meint eine Einrichtung mit der Ortsangabe “Berlin?” das Gleiche wie eine andere Institution mit “(probably) Berlin”? Statt für Transparenz zu sorgen, kann die Angabe von Unsicherheiten ohne ein standardisiertes Vokabular oder explizite Modellierung Klarheit sogar vermindern.

3. SCHLUSS

Wenn Dynamiken und unsicheres Wissen in Forschungsdaten abgebildet werden, wird der aktuelle Wissensstand zu einem Objekt präzisiert: Ist die Urheberschaft tatsächlich so sicher, wie ein gesetztes Kernfeld suggeriert, das keine Information darüber enthält, wie

sicher der dargestellte Fakt ist? Hat sich ein Aspekt unseres Wissens durch z.B. technologische Neuerungen radikal verändert und waren Forschende früher ganz anderer Ansicht? Indem Dynamiken und Unsicherheiten in einem Datensatz gekennzeichnet werden, kann das vorhandene Wissen in Relation zu laufenden Diskursen und dem, was aktuell (noch) nicht über ein Objekt ausgesagt werden kann, gesetzt werden. Durch diese Transparenz und Möglichkeit, Informationen zu gewichten, trägt die Einbettung von dynamischem und unsicherem Wissen auch unmittelbar zur Glaubwürdigkeit von Informationen in Forschungsdaten bei und steigert so deren Qualität.

Anhand des aufgeführten Beispiels wird jedoch deutlich, dass LIDO derzeit noch nicht in der Lage ist, Unsicherheiten und Dynamiken in allen Aspekten zu modellieren, in denen sie auftreten können. Zudem können mit einer lokal gewählten und definierten Markierung von Unsicherheiten mitunter schnell Probleme entstehen, insbesondere wenn Institutionen unterschiedliche Bezeichnungen für das gleiche Phänomen bzw. gleiche Codierungen für unterschiedliche Arten von Unsicherheiten und Unschärfen wählen. Konzepte zur Modellierung von Unsicherheit und Dynamik zu entwickeln ist eines der Kernziele von KONDA.

4. DANKSAGUNG

Wir danken an dieser Stelle Arno Kesper und Viola Wenz für ihre Unterstützung und Feedback. Zudem möchten wir dem BMBF für die Förderung von KONDA im Rahmen des Programms „Kurationskriterien und Qualitätsstandards von Forschungsdaten“ danken.

5. LITERATURHINWEIS

[1] Kabierske, Georg: Weinbrenner und Piranesi - Zur Neubewertung von zwei Grafikalben aus dem Besitz Friedrich Weinbrenners in der staatliche Kunsthalle Karlsruhe. In: Baumstark, Brigitte, *Friedrich Weinbrenner 1766-1826 - Architektur und Städtebau des Klassizismus. Ausstellung der Städtischen Galerie Karlsruhe und des Südwestdeutschen Archivs für Architektur und Ingenieurbau am KIT, 27. Juni 2015 bis 4. Oktober 2015*, Michael Imhof, Petersberg, 2015, S. 75-87.

[2] Bärminghausen, Julie at al.: Foto-Objekte, Kerber-Verlag, Berlin, 2018

[3] CIDOC LIDO Working Group: *What is LIDO?* [online], Online im Internet: <http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/> (23.10.2019). (Internetquelle)

[4] Deutsche Digitale Bibliothek: *Welches Metadatenformat müssen die liefernden Einrichtungen bereitstellen?* [online], Online im Internet: <https://www.deutsche-digitale-bibliothek.de/content/ueber-uns/fragen-antworten#5269> (23.10.2019). (Internetquelle)

[5] MUSEU-HUB–Services for museums going digital and aggregating for Europeana: *Share your data* [online], Online im Internet: <http://www.museuhub.eu/en/21/share-your-data> (23.10.2019). (Internetquelle)

[6] Coburn, Erin; Light, Richard; McKenna, Gordon; Stein, Regine; Vitzthum, Axel: *LIDO - Lightweight Information Describing Objects Version 1.0* [online], Online im Internet: <http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html> (23.10.2019). (Internetquelle)