

Beitrag aus:

Sonderband 4 der ZfdG: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. 2019. DOI: [10.17175/sb004](https://doi.org/10.17175/sb004)

Titel:

Genau, wahrscheinlich, eher nicht: Beziehungsprobleme in einem kunsthistorischen Wissensgraph

Autor/in:

Martin Raspe

Kontakt: raspe@biblhertz.it

Institution: Bibliotheca Hertziana – Max Planck Institut für Kunstgeschichte, Rom
GND: [139144145](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-64861-p0011-9) ORCID: [0000-0003-0861-0412](https://orcid.org/0000-0003-0861-0412)

Autor/in:

Georg Schelbert

Kontakt: georg.schelbert@hu-berlin.de

Institution: Humboldt-Universität zu Berlin, Institut für Kunst- und Bildgeschichte
GND: [133448231](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-64861-p0011-9) ORCID: [0000-0002-7314-8589](https://orcid.org/0000-0002-7314-8589)

DOI des Artikels:

[10.17175/sb004_012](https://doi.org/10.17175/sb004_012)

Nachweis im OPAC der Herzog August Bibliothek:

[1037075307](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-64861-p0011-9)

Erstveröffentlichung:

13.02.2019

Lizenz:

Sofern nicht anders angegeben



Medienlizenzen:

Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:

13.02.2019

GND-Verschlagwortung:

[Datenmodell](#) | [Graphdatenbank](#) | [Konzeptionelle Modellierung](#) | [Kunstwissenschaft](#) | [Wissensrepräsentation](#) |

Zitierweise:

Martin Raspe, Georg Schelbert: Genau, wahrscheinlich, eher nicht: Beziehungsprobleme in einem kunsthistorischen Wissensgraph. In: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. 2019 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 4). PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/sb004_012](https://doi.org/10.17175/sb004_012).

Martin Raspe, Georg Schelbert

Genau, wahrscheinlich, eher nicht: Beziehungsprobleme in einem kunsthistorischen Wissensgraph

Abstracts

Geisteswissenschaftliche Forschungsdaten in einem digitalen Wissensgraph abzubilden, bringt eine doppelte Herausforderung mit sich: Wie werden ungewisse Angaben in Form elektronischer Daten gespeichert und welche Folgen entstehen daraus für Abfrage und Visualisierung? Was drückt Unsicherheit aus und wie beeinflusst sie unser Konzept von Wissen? Das kulturhistorische Informationssystem ZUCCARO der Bibliotheca Hertziana ist ein Beispiel für einen komplexen Wissensgraph. Darin sind drei Fälle von Unsicherheit oder Vagheit zu unterscheiden: Genauigkeit, Plausibilität und negative Aussagen. Der Beitrag untersucht diese drei Aspekte im Hinblick auf die geschichtliche Wirklichkeit, den wissenschaftlichen Forschungsstand und die Benutzbarkeit der Daten für den Wissenschaftler.

To represent research data from the realm of the humanities in a digital knowledge graph presents a double challenge: How is uncertain data to be stored, and what does that imply for retrieval and presentation? Which notions are conveyed through uncertain data, and how do they influence our concept of knowledge? The cultural history information system ZUCCARO created by the Bibliotheca Hertziana is a good example for a complex knowledge graph. Three cases of uncertainty or vagueness are to be distinguished: Precision, plausibility and positively negative assertions. Our contribution studies the three aspects with regard to the historic reality, the state of scientific research and the usability of the content data for scholars.

»Uncertainty is the lack of information«¹

»Uncertainty is the prerequisite to gaining knowledge and frequently the result as well«²

1. Digitalität, Wissensgraph und Unsicherheit

1.1 Wissen

Mit dem Begriff Wissen wird – sowohl in den Natur- als auch in den Geisteswissenschaften – ein von einem größeren Personenkreis geteilter Bestand von begründeten Aussagen bezeichnet.³ Dieser Wissensbestand zeichnet sich im besten Fall durch einen »größtmöglichen Grad an Gewissheit«⁴ aus. So wird erreicht, dass die gemeinsam geteilten Auffassungen als gültig beziehungsweise »wahr« angenommen werden können. Eine absolute Verlässlichkeit kann nicht erreicht werden: Wissen hat prinzipiell und stets den Charakter des Vorläufigen.

¹ Bonneau et al. 2014.

² Hamilton 1936.

³ Einleitend hierzu Wikipedia: [Wissen](#).

⁴ Wikipedia: [Gewissheit](#).

Wissenschaftliche Tätigkeit besteht in der Erzeugung, der Anreicherung und der Vertiefung von Wissen. In der Praxis geschieht dies durch das Aufstellen und das Weitergeben beziehungsweise Mitteilen von Aussagen, die durch Argumente begründet sind. Erst durch die argumentative Begründung kommt die Wissenschaftlichkeit zustande. Die Begründung ist nicht nur deshalb notwendig, weil jeder Aussage über die Wirklichkeit ein Quantum an Distanz, Subjektivität oder Unzuverlässigkeit anhaftet, sondern auch, weil zur Definition der wissenschaftlichen Mitteilung gehört, dass der Adressat überprüfen kann, wie das an ihn weitergegebene Wissen entstanden ist. Nur so ist eine ›Rückkopplung‹ zwischen Empfänger und Sender möglich; nur so wird sichergestellt, dass der Adressat die Aussage des Urhebers für sich übernehmen kann. Er muss die Aussage in ihrem Entstehungskontext verstehen und ihren Grad an Gewissheit beurteilen können. Unsicherheit – oder besser Mangel an Gewissheit – ist also auch in der Praxis geisteswissenschaftlicher Forschung ein Aspekt, der jedweden Wissen innewohnt.

Diese Ungewissheit in einem komplexen digitalen Wissensspeicher abzubilden, bringt eine doppelte Herausforderung mit sich. Einerseits eine praktische: Wie können und sollen ungewisse oder unsichere Daten modelliert und gespeichert werden, und welche Folgen entstehen daraus für die Darstellung und die Abfrage dieser Daten? Andererseits eine theoretische: In welcher Weise wird die in ein Datenformat gebrachte Unsicherheit vom Benutzer inhaltlich verstanden und wie beeinflusst sie unser Konzept von Wissen?

1.2 Das Informationssystem ZUCCARO als Wissensgraph

Bei einem komplexen Wissensgraph wie dem Informationssystem *ZUCCARO* der Bibliotheca Hertziana, das wir als Beispiel heranziehen,⁵ wird kulturgeschichtliches Wissen grundsätzlich in Form von zeitbasierten Relationen zwischen Akteuren beziehungsweise Entitäten dargestellt. Anhand von Einzelfällen wollen wir anschaulich demonstrieren, welche Möglichkeiten es gibt, Unsicherheiten im Rahmen historischer Angaben zu modellieren. *ZUCCARO* ist ein Informationssystem für die historischen Kulturwissenschaften. Es wird an der Bibliotheca Hertziana – Max-Planck-Institut für Kunstgeschichte – in Rom seit 2003 von den Verfassern konzipiert und entwickelt. Seit 2005 ist es in einem relationalen Datenbanksystem prototypisch realisiert und seit 2008 online zugänglich. Das zu Grunde liegende Datenmodell ist so generisch und zugleich erweiterbar ausgelegt, dass es Forschungsprojekte aus unterschiedlichen Bereichen der Kulturwissenschaften unterstützt. Das System kann als universelles Repositorium dienen; prinzipiell ist es an alle gängigen Standardformate in den historischen Disziplinen anpassbar.

Derzeit kommt das System vor allem zur Sammlung von Informationen und Bildmaterial zur italienischen, besonders römischen Kunstgeschichte zum Einsatz. In erster Linie bilden Materialien aus den Forschungsprojekten *Lineamenta* (italienische Architekturzeichnungen des 18. Jahrhunderts) und *ArsRoma* (römische Malerei um 1600 im gesellschaftlichen Umkreis von Caravaggio) die Datenbasis. Das System enthält aber auch zahlreiche weitere Bestände,

⁵ Zugang und Dokumentation: [Zuccaro](#).

speziell zur Topographie der Stadt Rom (Bauwerke, Institutionen, Stadtpläne, Veduten), zu Künstleraufenthalten in Italien im 19. Jahrhundert sowie zu vielen Rara-Digitalisaten aus der Institutsbibliothek. Die Projektdaten sind durch *tags* gekennzeichnet, aber technisch nicht weiter separiert. Der Datenbestand ist also grundsätzlich offen und erweiterbar: Er enthält gemeinsame Stammdaten und zahllose Querverbindungen und kann insofern niemals als ›abgeschlossen‹ gelten. Durch jede neue Eingabe wird das Wissensnetz dichter geknüpft und dadurch nützlicher.

Kulturhistorische Forschung besteht nicht allein im Katalogisieren von materiellen Gegenständen, Artefakten und Bauwerken, sondern bezieht auch den historischen, politisch-gesellschaftlichen und ideell-konzeptionellen Kontext mit ein. Über die Werke hinaus widmet sie sich den Personen, Institutionen, sozialen Gruppen, Berufen und gesellschaftlichen Funktionen. Diese üblicherweise als ›Metadaten‹ bezeichneten Inhalte sind oft selbst wichtige Gegenstände der Forschung. Deshalb dienen sie als Rahmen unseres Informationssystems nicht vorrangig zur Objekterschließung, sondern werden ebenfalls als Datenobjekte erster Ordnung behandelt. Darüber hinaus untersucht die Forschung formale, inhaltliche und topographische Zusammenhänge und berücksichtigt Archivalien, Dokumente und Fachpublikationen. Alle diese Gegenstände werden in ZUCCARO als gleichwertig angesehen und können sowohl mit gezielten Abfragen als auch durch exploratives Browsen studiert werden.

Kulturhistorisches Wissen entsteht durch die Vernetzung von Informationen. Es bildet sich durch die Dokumentation geschichtlicher Ereignisse, an denen Personen, Objekte, Orte und immaterielle Konzepte beteiligt sind und die durch Quellen und Literatur historisch belegt sind. ZUCCARO trägt dieser Struktur des Wissens Rechnung, indem es jeden statischen Gegenstand – sei es ein handelnder ›Akteur‹ oder ein passives ›Objekt‹ – als sogenannte *Entität* behandelt. Die Entität ist lediglich ein digitaler Stellvertreter des Gegenstands. Der Datensatz wird mit Hilfe eines eindeutigen *identifiers* adressiert und kann durch Eigenschaften (*properties*) näher bestimmt werden, zum Beispiel durch Namen oder Bezeichnungen in verschiedenen Sprachen.

Zwischen derartigen abstrakten Entitäten können konkrete *Relationen* angelegt werden. In diesen Beziehungen ist das eigentliche historische Wissen enthalten. Generell sind diese Beziehungen zwischen jeweils zwei Entitäten in der Form von einfachen Aussagesätzen (*semantischen Tripeln*) nach dem Muster Subjekt-Prädikat-Objekt darstellbar, zum Beispiel: Die Nachtwache wurde gemalt von Rembrandt, oder: Der Mann mit dem Goldhelm wurde in der Technik Öl auf Leinwand ausgeführt.⁶

Nimmt man zu solch einer auf das Allerwesentlichste reduzierten Aussage noch den Zeitaspekt hinzu, und zwar in Gestalt einer Datierung, so kommt man zu einer Datensatzform, die man als nicht weiter reduzierbares mikrohistorisches Element ansehen kann, gewissermaßen als das kleinstmögliche geschichtliche Ereignis, wie zum Beispiel:

⁶ Zur Definition Wikipedia: [Semantic triple](#).

›Rembrandt kaufte im Jahre 1639 das Haus in der Jodenbreestraat‹. Dabei sind Rembrandt und das Haus feststehende Entitäten, während das erweiterte Prädikat ›kaufte im Jahre 1639‹ eine historische Beziehung zwischen beiden Entitäten herstellt.

Auf diese Weise steht das historische Ereignis im Zentrum unseres Datenmodells (Abbildung 1). Dieses Konzept, bei dem eine Beziehung zwischen Entitäten durch ein Datum oder einem Zeitraum erweitert wird, bezeichnen wir als ›Ereignis‹ oder *event*. Mit diesem Datenformat können alle inhaltlichen Entitäten auf einfache, generische Weise miteinander verbunden werden. Natürlich muss eine Beziehung nicht zwingend mit einem Datum versehen werden; sie kann außerdem durch zusätzliche Attribute genauer qualifiziert werden. Das kann die Spezifizierung des Beziehungstyps betreffen, zum Beispiel [Rembrandt] – war Schüler von – [Pieter Lastman], oder auch die Angabe einer Quantität, zum Beispiel [das Rembrandthaus] – besitzt eine Anzahl von 4 – [Fensterachsen].

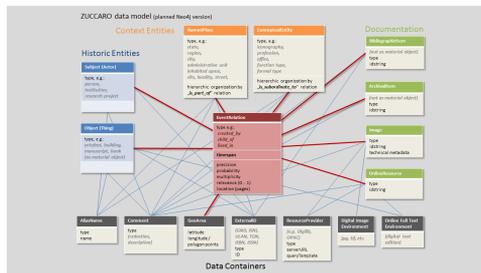


Abb. 1: Schematisches Datenmodell des Informationssystems ZUCCARO (Stand 2013). CC-BY-NC-SA 4.0.

Gewiss sind zahlreiche kulturhistorische Aussagen denkbar, die mit diesem generischen Format nicht ohne weiteres zu erfassen sind. Das Konzept kommt aber der technischen Umsetzung sehr entgegen und ermöglicht eine vielseitige Durchsuchbarkeit der Datenbestände unter beliebigen Aspekten und Fragestellungen. Von entscheidender Bedeutung ist noch ein zweiter Aspekt des Konzepts: Nicht nur Entitäten können miteinander in Beziehung gesetzt werden; Beziehungsdatensätze können ihrerseits durch eigene Beziehungen mit weiteren Datensätzen verbunden werden. Die Beziehungen werden dadurch *reifigiert*,⁷ sie werden also selber als statische Objekte behandelt. Auf diese Weise ist es möglich, mit einer Beziehung zusätzliche Umstände zu verknüpfen wie etwa den Kaufpreis oder den Verkäufer, aber auch den historischen oder fachwissenschaftlichen Beleg, etwa ein Archivstück oder eine Publikation mit Seitenangabe. Insbesondere diese Funktionalität ist von besonderer Wichtigkeit, da durch sie das Element der wissenschaftlichen Begründung abgebildet werden kann. In dem man eine Beziehung mit Belegen verknüpft, werden die historischen Angaben, die die Datenbank macht, nachprüfbar. Erst dann bekommen die Relationen wissenschaftlichen Charakter und unterscheiden sich etwa von einfachen Verknüpfungen im Rahmen von *linked open data*, bei denen nicht deutlich wird, auf welcher Grundlage die Aussage zustande gekommen ist. Beziehungen werden somit ihrerseits zu Objekt-Instanzen,

⁷ Definition Wikipedia: [Reifikation](#).

die in einem Graphen eigentlich durch Knoten repräsentiert werden müssten. Dadurch entsteht eine hybride Struktur, die nicht ohne Weiteres mit der klassischen Graphentheorie kompatibel ist.⁸

Betrachtet man viele solcher ›Mikroereignisse‹ zusammen, so wird klar, wie das Datenmodell komplexere historische Zusammenhänge darstellen kann, etwa die Biographie eines Künstlers anhand der Orte, an denen er sich nacheinander aufgehalten hat, oder seine Karriere anhand der Kontakte zu Förderern und Auftraggebern, oder auch die Abfolge der Personen, die ein Amt ausübten, oder die Kunstwerke, die zu einer historischen Sammlung gehörten, die heute zerstreut ist. Natürlich ist die Voraussetzung dafür, dass zu der gegebenen Fragestellung ein dichter und konsistenter Datenbestand vorliegt.

Wissenschaftliche Anfragen aus derart unterschiedlichen Blickwinkeln sind mit den meisten herkömmlichen Datenbank-Modellen kaum abzubilden. Eine solche ›polyfokale‹ Organisation von Forschungsdaten, die auf unterschiedliche Forschungsfragen reagiert, findet jedoch erfreulicherweise im Bereich der Digital Humanities zunehmend Anklang. Das Datenmodell von ZUCCARO ist angeregt durch die Datenbank zur Antikenrezeption *Census*⁹ und die kulturwissenschaftliche Ontologie *CIDOC-CRM*.¹⁰ Es ist jedoch grundsätzlich generisch konzipiert und legt den Benutzer nicht auf ausgewählte Blickwinkel oder Forschungsgegenstände fest. Im informatischen Sinn stellt das Datenmodell einen sogenannten *property graph* dar. Eine solche netzartige Datenstruktur besteht aus *nodes* (Knoten, bei uns Entitäten) und *edges* (Kanten, bei uns Beziehungen), die beide *properties* (Eigenschaften, also Felder bzw. Attribute) besitzen können.¹¹ Seit einigen Jahren ist ein wachsendes Interesse an derartigen Datenmodellen zu beobachten, was uns hinsichtlich der Richtigkeit des eingeschlagenen Weges bestärkt.

Mit Hilfe eines derartigen Datenmodells, das auf dem Prinzip *property graph* beruht, können sämtliche Daten im Idealfall redundanzfrei als erweiterte Tripelstrukturen dargestellt werden. Zurzeit ist ZUCCARO als Prototyp in einem proprietären, relationalen Datenbanksystem implementiert und mit einem Web-Frontend auf der Basis von XML-Ausgabedaten und Rendering mit Hilfe von XSLT-Templates ausgestattet. Alle Beziehungen müssen durch *joins* zwischen Tabellen abgebildet werden, was sehr komplex werden kann und die Performance belastet. Mit dem Aufkommen der Graphentechnologien steht dem Datenmodell erstmalig eine adäquate Software-Basis gegenüber. Seit einiger Zeit ist die Umsetzung auf ein Graphdatenbanksystem und ein auf modernen Web-Technologien basierendes Framework in Vorbereitung. Vorgesehen ist *Neo4j* als Datenbank-Software und *Mojolicious* oder *Phoenix*¹² als Management Interface. Besonderer Wert wird auf

⁸ Um sich auf ein Datentripel als Ganzes bzw. auf die Instanz einer Relation beziehen zu können, verwendet man in dem Datenformat RDF sogenannte *N-Quads* – um ein viertes Element erweiterte Tripel.

⁹ *Census of Antique Works of Art and Architecture*.

¹⁰ *CIDOC Conceptual Reference Model*.

¹¹ Frisendal 2017.

¹² *Mojolicious*; *Phoenix Framework*.

Anschlussfähigkeit und Schnittstellen im Bereich *semantic web* gelegt. Sie soll durch standardisierte Formate und Schnittstellen erreicht werden, beispielsweise durch eine in GraphQL formulierte API.¹³

1.3 The modelling gap: Unsicherheit durch Abstraktion

Überall dort, wo historische Sachverhalte und Zusammenhänge in ein abstraktes Datenbankformat übertragen werden sollen, ergibt sich das Problem des Übergangs vom analogen Kontinuum zur digitalen Fragmentierung. Die fließende raum-zeitliche Entwicklung der traditionellen, textbasierten Geschichtsdarstellung muss in Ereignisse, Zeitabschnitte, Raumelemente und Sinneinheiten zerlegt werden, um sie in digitaler Form zu speichern, zu vergleichen und je nach Forschungsinteresse neu aggregieren zu können. Entscheidungen zur Zusammenfassung oder Teilung von Dingen werden notwendig: Wo sind natürlich gegebene Grenzen? Wie sehr fragmentieren wir die Welt? Ein Bauwerk bildet beispielsweise eine Einheit – es datentechnisch von seinen Nachbargebäuden zu unterscheiden, leuchtet sofort ein. Andererseits besteht es vielleicht aus Haupt- und Nebengebäuden und hat verschiedene Bauphasen – wie weit geht man bei der Aufteilung?

Nicht alle so entstehende Datenfragmente sind gleichermaßen umfangreich oder wichtig. Im Vergleich zur Wirklichkeit ergibt sich das Problem der Verhältnismäßigkeit der digitalen Granularität: Kleine Dinge erhalten ein überproportionales Gewicht im Datenspeicher, wenn sie einzeln modelliert werden. Der Datensatz kennt keine Dimension: Das Formular für das Fürstenschloss hat gleich viele Felder wie dasjenige für das Gärtnerhaus. Besonders sinnfällig zeigt sich die Diskrepanz im Bereich der CAD-Modelle: Der gesamte Baukörper des aus schlichten, quaderförmigen Elementen bestehenden Bauhaus in Dessau kann beispielsweise mit weitaus weniger Polygonen modelliert werden als ein einziges korinthisches Kapitell mit seinen naturnahen Blattformen und geschwungenen Linien.

Aber es ist nicht nur die Zerlegung und die damit einhergehende Abstraktion, die bei der Übertragung ins Digitale eine Verzerrung der Wirklichkeit mit sich bringen. Im kulturhistorischen Bereich kommt eine weitere Schwierigkeit hinzu, die es in letzter Konsequenz schlichtweg unmöglich macht, im digitalen Raum ein adäquat proportioniertes Abbild der geschichtlichen Wirklichkeit herzustellen – und sei es auch nur eines kleinen Ausschnitts davon. Es ist das Grundproblem jeder historischen Forschung: In den allerseltensten Fällen sind Angaben so vollständig und flächendeckend vorhanden, dass es möglich erscheint, ein ausgewogenes, »statistisch korrektes« Abbild der historischen Lage zu zeichnen. Die Lückenhaftigkeit sowohl der Überlieferung als auch in der historischen Forschung sorgt dafür, dass in aller Regel nur ausschnittshafte, mehr oder minder exemplarische Daten vorliegen – zum Beispiel für einzelne Personen in einer Gruppe, oder für wenige Jahre in einem länger dauernden Vorgang. Natursprachliche Texte können durch eine Vielzahl von Formulierungsweisen die wissenschaftliche Darstellung so nuancieren, dass einerseits die Sachverhalte im Einzelnen zutreffend gewichtet werden und andererseits

¹³ Wikipedia: [GraphQL](#).

ein reflektiertes, angemessenes Gesamtbild entsteht. Datenbanken können dies nicht; sie können gegebenenfalls die Quellen getreu abbilden, eine Wirklichkeit zusammenfassend rekonstruieren oder gar synthetisieren jedoch nicht.

Ein pragmatischer Ausweg aus dem Dilemma kann daher sein, sich bei Entscheidungen in Bezug auf Abstraktion und Granulation von den vorliegenden Quellen leiten zu lassen. Einteilungen von Zeit und Raum werden zum Beispiel bei Melderegister-Einträgen, die nur einmal im Jahr erhoben werden, durch die Quelle vorgegeben. Das gleiche gilt für die Benennung von Bauabschnitten, die aus Bauabrechnungen übernommen werden. Alle Probleme, die sich aus der Lesung und Deutung der Quellen ergeben, finden sich dann in der Datenbank wieder. Die Unsicherheit resultiert in diesem Fall nicht aus der Abstraktion, sondern aus dem Verzicht auf historische Interpretation. Konsequenterweise wäre eine solche Datenbank dann kaum mehr als ein ›kulturhistorisches Informationssystem‹ zu bezeichnen, sondern hätte den Status einer strukturierten Quellenedition.

Ein Informationssystem wie ZUCCARO, das nicht ausschließlich Originalquellen reproduziert, sondern mit impliziten, aus der Sekundärliteratur oder sogar allgemein bekannten Fakten arbeitet, hat also aus systemischen Gründen keine Chance, den *modelling gap*, also den intellektuellen Abstand zwischen Wirklichkeit und Modell, zu überbrücken. Ein Informationssystem hat lediglich Hinweischarakter. Es ist ein Findmittel, das im besten Falle rasch und übersichtlich zum Forschungsstand hinführt, aber keine virtuelle Historie. Dem Missverständnis, Datenbankinhalte könnten die historische Wirklichkeit auch nur näherungsweise darstellen beziehungsweise ›repräsentieren‹, muss stets aufs Neue entgegengetreten werden: Der Wegweiser zeigt den Weg, aber er geht ihn nicht. Ziel eines digitalen Informationssystems ist es, auf bekannte Fakten, Zusammenhänge und Forschungsmeinungen aufmerksam zu machen und dieses Material so transparent und sinnfällig wie möglich aufzubereiten und anzubieten. Der historische Wahrheitsgehalt ist – wie bei jedem wissenschaftlichen Katalog – vom Benutzer selber anhand der angegebenen Quellen zu verifizieren und einzuschätzen. Der adäquaten Wiedergabe von Unschärfen, Wahrscheinlichkeiten und Qualitäten historischer Ereignisse und Sachverhalte sind insofern von vornherein pragmatische Grenzen gesetzt.

2. Genau, wahrscheinlich, eher nicht...

2.1 Genau: Präzision und Trennschärfe

Zunächst ist festzuhalten: In einem graphbasierten Datenmodell sind jeweils zwei Knoten – die Entitäten repräsentieren – durch eine Kante – also eine Beziehung – miteinander verbunden. Unsicherheit kann ausschließlich in der Beziehung ausgedrückt werden. Eine vage oder ungewisse Entität ist zwar vielleicht denkbar (zum Beispiel der mythische Ort Thule oder der unmerklich wirkende »Zeitgeist«), es hat jedoch wenig Sinn, diese in wissenschaftlichen Aussagen zu verwenden – es sei denn, wenn man das gedankliche Konzept meint und nicht den Ort oder Akteur. Auch bei fest eingeführten Fachbegriffen gibt es Randunschärfen, zum

Beispiel: Fällt eine Mischung aus Leinöl, Eigelb und Pigment bereits unter den Begriff Ölfarbe? oder: Kann man diesen Stein als Würfelkapitell bezeichnen, oder ist er nur ein roh behauener Kämpferblock? oder: Wann beginnt, wann endet die Stilepoche des Barock? Trotzdem sind diese Begriffe wohldefiniert. Unsicher oder umstritten ist vielmehr, ob man eine Entität – also einen Forschungsgegenstand – mit vollem Recht einem bestimmten begrifflichen Konzept zuordnen kann. Auch hier liegt also die Ungewissheit in der Relation, nicht in der allzu vagen Definition der Entität.

Ein Grundproblem bei der Modellierung von Unschärfen und Unsicherheiten besteht darin, dass das übliche Datenmodell eines Graphen, das sich von der mathematischen Graphentheorie herleitet, überhaupt keine unterschiedlich gewichteten Kanten (in unserem Fall Wissensrelationen) vorsieht. Um eine solche Gewichtung vorzunehmen zu können, muss man einen *property graph* verwenden, bei dem die Relationen mit zusätzlichen Angaben angereichert, also *reifizi*ert werden können. In einem oder mehreren dieser Felder können Werte gespeichert werden, die den Grad der Unsicherheit beziehungsweise Vagheit abbilden.

Der einfachste Fall ist hierbei die *Präzision*: Die Relation enthält eine Angabe über den Grad der Genauigkeit, mit dem sie zutrifft. Hierbei steht in der Regel nicht die Relation selber in Frage, sie wird als gegeben betrachtet; vielmehr ist eines ihrer Attribute oder ein Parameter mehr oder weniger zutreffend. Das kann zum Beispiel die Datierung eines Ereignisses, die Zuordnung einer Entität zu einer Kategorie oder auch die Lokalisierung eines Ortes betreffen.

Ein Datenmodell kann hierzu festlegen, wie etwa zeitliche Ungenauigkeit oder inhaltliche Unklarheit kodiert werden sollen. Ein ungewisses Datum kann in einen gewissen Zeitraum fallen. Man kann also einen *terminus post quem* und einen *terminus ante quem* angeben, ohne genauer festzulegen, welches Datum als das wahrscheinlichste gelten kann; man kann aber auch nur ein Datum angeben und dazu die mögliche Streuung festhalten.¹⁴

Im geisteswissenschaftlichen Kontext tritt häufig der Fall auf, dass es für ein bestimmtes Ereignis unterschiedliche, einander ausschließende Datierungsvorschläge gibt. In diesem Fall erscheint es sinnvoll, mehrere Ereignis-Beziehungen anzulegen und mit unterschiedlichen Zeitspannen zu versehen. Die jeweilige Begründung kann dann mit der entsprechenden Beziehung verknüpft werden. Es ist auch möglich, zusätzlich eine allgemeinere Ereignis-Beziehung anzulegen, die den gesamten vorgeschlagenen Zeitraum umfasst. Wählt man die Lösung mit mehreren, alternativ datierten Ereignissen, so schließt sich sofort die Frage an, wie diese dargestellt werden, in welcher Reihenfolge oder in welcher Auswahl. Je genauer man es mit der Präzision nimmt, umso problematischer werden die Visualisierung und die intuitive Verständlichkeit der Daten.

An dieser Stelle wird deutlich, dass es keine allgemeingültige Lösung für derartige Präzisionsprobleme geben kann, die jeden Einzelfall abdeckt. Die Modellierung von Datierungen kann sich nicht allein danach richten, wie die objektive Wirklichkeit ausgesehen

¹⁴ Vgl. dazu die Definition der Entität [E52 Time-Span](#) im CIDOC-CRM.

hat, da diese in vielen Fällen unbekannt oder nicht verfügbar ist. Vielmehr wird man sich oft auf vermutetes, aber verschiedenartig begründetes Wissen stützen. Die Frage, wie ein ungenau bekannter Sachverhalt in einer Datenbank technisch umgesetzt und damit inhaltlich dargestellt werden soll, kann also nicht allein anhand dessen entschieden werden, wie es sich in der historischen Wirklichkeit verhielt. Entscheidend ist vielmehr die Überlegung, wie der Sachverhalt dem Benutzer der Datenbank im Suchergebnis dargestellt werden soll. Modellierungsfragen können sich also nicht primär an den Charakteristika der Inhalte orientieren, sondern müssen die Bedürfnisse der Benutzer beziehungsweise den Verwendungszweck der Datenbank zu Grunde legen.

2.2 Wahrscheinlich: Plausibilität als Gewichtungskriterium

Aus dem geschilderten Beispiel, bei dem mehrere Datierungsvorschläge miteinander konkurrieren, ergibt sich ein zweiter, prinzipiell anders gelagerter Fall von Ungewissheit in einer kulturhistorischen Datenbank. Hier geht es nicht mehr allein darum, die mangelnde Präzision einer Aussage festzuhalten, sondern ihre *Plausibilität* im Vergleich zu anderen, widersprechenden Aussagen, oder ihre Wahrscheinlichkeit im Vergleich zu ihrem Gegenteil abzubilden.

In den wenigsten Fällen ist historisches Wissen mit absoluter Sicherheit belegbar. Es ist also durchaus verständlich und legitim, einen Grad an Wahrscheinlichkeit angeben zu wollen, ob eine im Datenbestand angelegte Beziehung zwischen zwei Entitäten tatsächlich historisch bestanden hat. Hier kann man in der Tat von »verschieden stark« ausgeprägten oder gewichteten Graph-Kanten sprechen.

Es leuchtet unmittelbar ein, dass dieser Fall zwei Probleme aufwirft, die im Fall der mangelnden Genauigkeit nicht auftreten. Zum einen betrifft die fehlende Sicherheit hier die Existenz der Beziehung als solcher. Wenn wir die genaue Datierung eines Ereignisses nicht kennen, so ist dadurch noch nicht in Frage gestellt, dass es das Ereignis gab. Wenn aber unklar ist, ob das Ereignis in dem angegebenen Zeitraum überhaupt stattgefunden hat, dann kann der Fall eintreten, dass in der Datenbank ein Sachverhalt gespeichert wird, der möglicherweise der historischen Wahrheit widerspricht.

Das alleine wäre noch kein grundsätzliches Problem, denn auch in der analogen Geisteswissenschaft werden regelmäßig hypothetisch mögliche Sachverhalte vorgeschlagen, sofern ihnen eine gewisse Wahrscheinlichkeit zu eigen ist. Ausschlaggebend ist hier erneut die mit der Aussage verknüpfte, durch Argumentation untermauerte wissenschaftliche Begründung, die dem Adressaten die Möglichkeit gibt, das Wissensfaktum und seine Plausibilität selber zu beurteilen.

Das zweite Problem betrifft den Begriff der Wahrscheinlichkeit, der mit dieser Art von Unsicherheit verbunden ist. Gemeint ist hier nicht die objektivistische, mathematisch berechenbare Wahrscheinlichkeit,¹⁵ die üblicherweise aufgrund von empirisch beobachteten Vorgängen auf das mögliche Eintreten zukünftiger Ereignisse schließt. Auf unseren Fall ist der subjektive Wahrscheinlichkeitsbegriff¹⁶ anzuwenden, bei dem die persönliche Einschätzung als Maß für die Sicherheit eines Sachverhalts dient. Da es sich hier um gemeinschaftliches Wissen handelt, ist es nicht die Einschätzung des Einzelnen, die zählt, sondern die der wissenschaftlichen Community. Daher ist der Begriff ›Plausibilität‹ zu bevorzugen, um den Aspekt der Zustimmung zu betonen.

Zwar kann Plausibilität als Zahlenwert, und zwar als Quotenverhältnis¹⁷ dargestellt werden, das Problem liegt jedoch darin, dass die Zahlenwerte sich aus dem Verhältnis der vorhandenen Alternativen zueinander ergeben. Plausibilität ist ein relativer Wert, er kann nicht absolut angegeben werden: Eine Vermutung verliert an Wahrscheinlichkeit, wenn eine plausiblere hinzukommt. Wenn mehrere alternative Sachverhalte mit ihrer jeweiligen Plausibilität festgehalten werden sollen, dann steht jeder Zahlenwert in einem festgelegten Verhältnis zu jedem anderen und zur Gesamtsumme. Dies hat zur Folge, dass man nicht einen Plausibilitätswert verändern, hinzufügen oder löschen kann, ohne dass alle anderen angepasst werden müssten. Andernfalls würde sich das Gesamtverhältnis verschieben, und die Datenbank geriete in einen inkonsistenten Zustand – es sei denn, in solchen Fällen würde ein automatischer Korrekturmechanismus greifen.

Aber auch, wenn es diesen gäbe, wäre es schwierig, das Verhältnis von Plausibilitäten verschiedener Sachverhalte zueinander mit einem Algorithmus zu ermitteln. Wonach soll geurteilt werden? Nach einem ähnlich vagen *credibility factor* der ForscherInnen, welche die jeweilige Aussage getätigt haben? Nach der Zahl der angegebenen Quellen? Nach einer Abstimmung durch die wissenschaftlichen Benutzer? Die Absurdität und der fragwürdige wissenschaftliche Nutzen einer solchen Berechnung liegen auf der Hand.

Kulturhistorische Aussagen bestehen außerdem oft aus einer Vielzahl von abhängigen Sachverhalten, die sich kaum separat beurteilen lassen. Wenn ein Kunstwerk einer Person zugeschrieben wird, impliziert dies Überlegungen zur Datierung, zur Malweise, zum Entstehungskontext, zur Schaffensphase und so weiter. In der Praxis wird man nicht die einzelnen Parameter separat hinsichtlich ihrer Stichhaltigkeit bestimmen und daraus die wahrscheinlichste mutmaßliche ›Realität‹ berechnen, sondern mehrere Relationen bilden, die unterschiedliche wissenschaftliche Meinungen darstellen. Ein Beispiel ist der ZUCCARO-Datensatz zu einer Zeichnung in der Berliner Kunstbibliothek (Abbildung 2). Zu dem Blatt, das wohl in Rom entstanden ist und vermutlich einen Entwurf für eine Villa zeigt, gibt es voneinander abweichende wissenschaftliche Meinungen zur Bestimmung und zur Autorschaft. Diese sind in den zugehörigen Relationen in Plausibilitätswerten, Kommentaren und zusätzlichen bibliographischen und archivalischen Belegen ausgedrückt.¹⁸

¹⁵ Wikipedia: [Objektivistischer Wahrscheinlichkeitsbegriff](#).

¹⁶ Wikipedia: [Subjektiver Wahrscheinlichkeitsbegriff](#).

¹⁷ Wikipedia: [Chancenverhältnis](#).

¹⁸ Vgl. Kieven / Schelbert 2014.

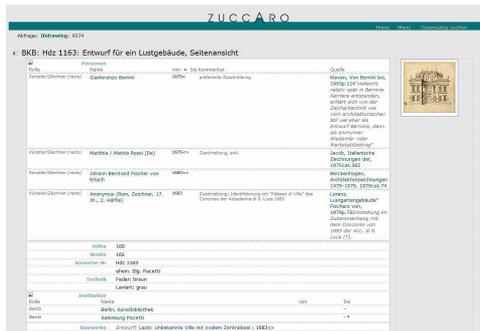


Abb. 2: Unbekannter Künstler (Gianlorenzo Bernini?): Entwurf für ein Lustgebäude, vermutlich eine Villa in Rom. Datensatz in ZUCCARO. CC-BY-NC-SA 4.0.

Erneut stellt sich also die Frage, wie der Unsicherheitsfaktor in einer Datenbank eingebaut werden kann, und erneut kann die abgebildete Wirklichkeit nicht als Kriterium dienen. Wieder müssen wir festhalten: Es kommt darauf an, wozu wir den Aspekt der Plausibilität eigentlich benötigen, was der Datenbank-Benutzer daraus ersehen kann beziehungsweise welchen Zweck wir mit der Datenbank anstreben.

2.3 Eher nicht: Negative Aussagen als Basis kulturwissenschaftlichen Wissens

Kulturhistorisches Wissen, insbesondere die dazu notwendige Argumentation, basiert nicht selten auf dem Ausschlussprinzip. Häufig wissen wir lediglich positiv, dass eine bestimmte historische Beziehung mit Sicherheit niemals bestanden hat. Mit treffendem Witz hat der Karikaturist Freimut Woessner den Vorgang beispielhaft in einem Cartoon dargestellt, der 1991 anlässlich der Ausstellung »Rembrandt – Der Meister und seine Werkstatt« im Berliner Stadtmagazin »zitty« erschienen ist (Abbildung 3). Die in der Karikatur erzählte Geschichte bezieht sich auf das Faktum, dass die Mitglieder des Rembrandt Research Project 1986 das Gemälde mit dem Titel *Der Mann mit dem Goldhelm*, bis dahin eine Zimelie der Berliner Museen, aus dem eigenhändigen Oeuvre des Meisters ausgeschieden hatten. Diese neue wissenschaftliche Erkenntnis erregte großes Aufsehen und wurde mit Erstaunen, gelegentlich auch mit Empörung aufgenommen. Die Berliner Ausstellung von 1991 stellte die Ergebnisse des Projekts einem größeren Publikum vor Augen.

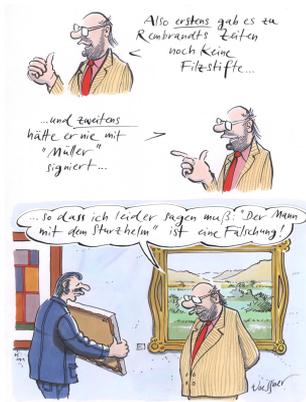


Abb. 3: Freimut Woessner: Der Mann mit dem Sturzhelm, 1991, Zeichnung, Archiv des Künstlers. Mit freundlicher Genehmigung des Künstlers.

In Woessners Bildergeschichte wird der kunsthistorische Argumentationsprozess auf die Schippe genommen. Angesichts eines ihm zu Prüfung vorgelegten Gemäldes bringt der Experte zwei Einwände gegen eine Zuschreibung an Rembrandt, um dann schlussfolgernd darzulegen, dass es sich um eine Fälschung handeln müsse. Die Komik besteht darin, dass abgesehen davon, dass die vorgeführte Filzstiftzeichnung anscheinend einen Mann im Helm zeigt, offenkundig überhaupt kein Bezug zu dem niederländischen Meister des 17. Jahrhunderts besteht und insofern die Bezeichnung als ›Fälschung‹ völlig absurd ist.

Kulturhistorisches Wissen besteht – und das ist charakteristisch für unsere Disziplinen – nicht selten in solchen negativen Statements. Der Experte kann keine positive Zuschreibung vornehmen, sondern lediglich aussagen, dass zwischen der Filzstiftzeichnung und Rembrandt keine Beziehung besteht. Wenn wir diesen Sachverhalt in unser Datenmodell übertragen wollen, so tritt der paradoxe Fall ein, dass wir eine Beziehung zwischen den zwei Entitäten ›Rembrandt‹ und ›Filzstiftzeichnung‹ anlegen müssen (also eine Kante zwischen zwei Knoten), um damit auszusagen, dass gerade *keine* Beziehung besteht.

Man könnte sich hier aus der Verlegenheit helfen, indem man kulturwissenschaftlich festgestellte Nicht-Beziehungen definiert als normale Beziehungen mit einer Plausibilität von Null. In unserem Fall gibt daneben keine weitere Beziehung zwischen Werk und Künstler, denn der Urheber ist ja unbekannt. Die Gleichbehandlung von Beziehungen und Nicht-Beziehungen führt also zu einer weiteren Paradoxie, nämlich dass sich beide systematisch nicht mehr unterscheiden lassen. Praktisch bedeutet das, dass in einer Auflistung der Werke Rembrandts auch Werke erscheinen, die überhaupt nichts mit dem Meister zu tun haben – und je mehr solche negativen Statements wir in die Datenbank aufnehmen, umso deutlicher treten diese in Erscheinung. Aus einem Wissensgraphen würde ein Graph des Nichtwissens. »Die Welt ist alles, was nicht *nicht der Fall* ist«: Eine Datenbank, die für sämtliche Verbindungen, die mit Gewissheit auszuschließen sind, Beziehungsdatensätze anlegt, würde sich selber ad absurdum führen.

Aus dem Dilemma gibt es keinen prinzipiellen Ausweg, denn man kann derartige affirmativ-negative Aussagen auch nicht dadurch darstellen, dass man sie einfach weglässt. Es besteht nämlich grundsätzlich keine Unterscheidungsmöglichkeit zwischen einer Angabe, die prinzipiell unbekannt ist, und einer Angabe, die lediglich nicht in die Datenbank eingetragen wurde. Wenn zu einem Kunstwerk kein Künstler angegeben ist, heißt dies nicht, dass es keinen Urheber gab, es heißt auch nicht, dass der Urheber unbekannt ist. Es bedeutet lediglich »keine Information eingetragen«. Es wäre absurd, die Datenbank überall dort, wo Angaben tatsächlich nicht bekannt oder verfügbar sind, mit Aussagen über eben diesen Sachverhalt zu füllen. Trotzdem ist es in der Kulturgeschichte durchaus üblich, auch Unbekanntes positiv festzuhalten, etwa bei anonymen, aber künstlerisch profilierten Meistern Notnamen zu vergeben, oder mit allgemeinen Bezeichnungen wie »Niederländischer Maler des 15. Jahrhunderts« zu arbeiten. Insbesondere im zweiten Fall wird deutlich, dass zu einer näheren Bestimmung der Person durchaus weitere Kriterien vorhanden sind. Man könnte statt der Künstlerperson eine abstrakte Personengruppe anlegen, deren Merkmale sind, dass sie ungefähr im 15. Jahrhundert bestanden hat und mit der kulturellen Region Niederlande in Beziehung stand. Dadurch wird vermieden, dass man eine Unzahl anonymer Künstler anlegen muss, die jeweils einzeln näher zu spezifizieren sind. Hier wird erneut deutlich, dass eine Datenbank kein Abbild der geschichtlichen Realität sein kann: Eine solche Personengruppe hat historisch nicht existiert. Sie ist lediglich ein Platzhalter, der zur Organisation des Wissens dient.

Zu einem ähnlichen Schluss wird man kommen, wenn man das Problem der negativen Aussagen dadurch löst, dass man es auf solche Fälle beschränkt, wo die Forschung auf vergleichbare Weise im Ausschlussverfahren gearbeitet und sich vorläufig auf negative Aussagen zurückgezogen hat. Wo es einen Anlass gab, eine Urheberschaft Rembrandts zu vermuten, da kann man auch positiv festhalten, wenn sich herausstellt, dass sich diese Beziehung nicht bestätigt hat. Diesen Fall zeigt auch die Karikatur, wenn sie als Beleg zwei weitere Negativaussagen anführt. Ob es allerdings sinnvoll ist, Selbstverständlichkeiten digital festzuhalten wie »Rembrandt hat nicht mit Müller signiert«, steht dahin. Nicht jedes Argument einer kunsthistorischen Erörterung muss in einem Informationssystem Aufnahme finden. Andererseits sind Argumentationsketten wie »Rembrandt hat keine Filzstifte verwendet, weil es sie zu seiner Zeit noch nicht gab« durchaus nicht trivial. In bestimmten Fällen kann es durchaus sinnvoll sein, sie in einem Informationssystem zu dokumentieren. Sie in Beziehungen zu zerlegen und im Rahmen eines *property graph* zu modellieren, macht die Sache allerdings beträchtlich komplizierter.

Zwei Dinge werden daran deutlich. Erstens: Negative Aussagen sind nicht dazu geeignet, die historische Wirklichkeit zu repräsentieren. Dennoch haben sie in einem kulturhistorischen Informationssystem ihre Berechtigung. Zweitens: Am Beispiel der negativen Aussagen wird deutlich, dass eine Datenbank nicht die reale Vergangenheit, sondern nur das durch Forschung erworbene Wissen darüber widerspiegeln und verwalten kann. Dieses wird immer lückenhaft sein. Die Unsicherheit ist eine Funktion der systembedingten Unvollständigkeit. Auch eine Forschungsdatenbank kann uns nichts wirklich Neues lehren, sie kann nicht selber Forschung treiben und automatisch Wissenslücken füllen. Sie lässt uns lediglich bereits Bekanntes rascher

wiederfinden als zuvor. Darüber hinaus ermöglicht sie uns, durch intelligente Strukturierung anhand eines Datenmodells das scheinbar altbekannte Material nahezu mühelos neu anzuordnen und aus anderen Blickwinkeln zu betrachten.

3. Letztlich alles eine Frage der Darstellung ...

3.1 Das Problem des ranking in der Wissensrepräsentation

Der Fall der negativen Beziehungen verweist auf ein weiteres, grundlegendes Problem kulturwissenschaftlicher Informationssysteme. Ausschlaggebend für die Vermittlung des gespeicherten Wissens ist nämlich nicht allein das Datenmodell, durch welches die Informationen in kleinste Einheiten zerlegt werden, sondern in mindestens gleichem Umfang die Frage nach der Anordnung der Relationen bei der Ergebnisausgabe im Rahmen einer Datenbankabfrage. Nach welchen Kriterien werden unscharfe beziehungsweise unsichere Beziehungen sortiert? Vom *ranking* der Suchergebnisse nach wissenschaftlicher Relevanz hängt die Verständlichkeit des Materials und Vertrauenswürdigkeit der im System gespeicherten Informationen in hohem Maße ab. Lässt sich die unterschiedliche inhaltliche Gewichtung im Hinblick auf Genauigkeit oder Plausibilität so allgemeingültig quantifizieren, dass damit ein Sortier-Algorithmus umgehen kann? Welche Rolle spielen dabei der Abfragekontext, das Erkenntnisinteresse und der subjektive Blickwinkel des Benutzers? Ist es überhaupt zulässig, das Ranking intelligenten Algorithmen zu überlassen, oder ist es Bestandteil der wissenschaftlichen Aussage und muss daher separat modelliert werden? Diese Fragen sollen hier nur anhand einiger praktischer Beispiele aufgeworfen, aber nicht abschließend beantwortet werden.

Die Kehrseite des Umgangs mit Ungenauigkeit und Plausibilität, nämlich das Problem der Anordnung und Visualisierung, wurde im Rahmen der internen Evaluation von *ZUCCARO* immer wieder deutlich. Nicht selten gab es beim wissenschaftlichen Publikum Reaktionen wie ›da ist ja alles durcheinander‹, ›und was ist nun wichtig?‹. Einen charakteristischen Fall bildeten die Einträge zu der heutzutage nicht mehr existierenden *Villa del Pigneto Sacchetti* von Pietro da Cortona.¹⁹ Um diese aufzurufen, wurde über das Abfrageformular nach einem Bauwerk gesucht, dessen Bezeichnung die Zeichenkette ›Villa Sacchetti‹ enthält.

Zwar fand das System alle zu der Villa gehörenden Daten, das Ergebnis befriedigte die Erwartungen dennoch keineswegs: In der Ausgabetablelle wurde die Trefferliste alphabetisch angeordnet. Dadurch erschienen sämtliche Teilgebäude der Villa, das Casino, die Grotte, das Nymphäum zuoberst. Der Haupteintrag der Villa erschien erst ganz unten, weil bei ihm die Bezeichnung mit dem Buchstaben ›V‹ anfängt.

¹⁹ Vgl. hierzu auch Kieven 2011.

Selbst die semantische Hierarchisierung der Bestandteile untereinander durch Beziehungen wie ›ist enthalten in‹ oder ›bildet Einheit mit‹ reichte nicht aus, um für den Benutzer Klarheit zu schaffen. Ursache dafür ist letztlich erneut die Diskrepanz zwischen dem Zwang zur Festlegung und der daraus resultierenden Trennschärfe des Digitalen im Angesicht der fließenden Gegebenheiten in der Realität. Der Ausgabe-Algorithmus war nicht darauf vorbereitet, in der Trefferliste die hierarchische Struktur zu erkennen und diese in der Anordnung zu berücksichtigen. Inzwischen wurde die Ausgabelogik für diesen Fall geändert (Abbildung 4), aber natürlich müssten noch viele andere Ausnahmen berücksichtigt werden. Der *property graph* allein reicht nicht aus, einen vernünftigen Überblick zu gewährleisten, damit der Zusammenhang des Wissens nicht durch die Fülle an Detailinformationen verloren geht.

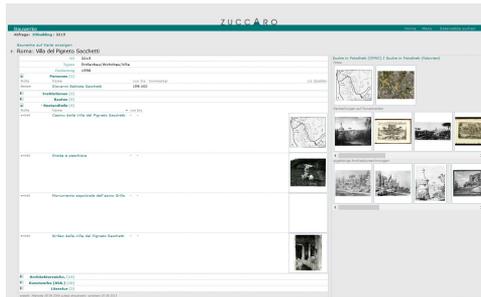


Abb. 4: Rom, Villa del Pigneto Sacchetti. Datensatz in ZUCCARO. CC-BY-NC-SA 4.0.

Es wird deutlich, dass ohne die zusätzliche Berücksichtigung semantischer Zusammenhänge die visuelle Repräsentation des Wissens in einem Informationssystem so unverständlich ausfallen kann, dass der Benutzer eher verwirrt als informiert wird. Dies trifft besonders dann zu, wenn er nicht weiß, mit welchen Ordnungskriterien der Ausgabe-Algorithmus arbeitet. Die Anordnung der Daten nach ihrer Relevanz für den Benutzer erfordert eine tiefere semantische Analyse sowohl der Abfragesituation als auch der im System enthaltenen Ergebnismenge – und dies, obwohl mit der Suche nach einem Bauwerk bereits ein inhaltlicher Kontext vorgegeben war.

Ort	Datum	Dauer
Berlin	1829	1830
Worms	1830	1831
Worms	1831	1832
Worms	1832	1833
Worms	1833	1834
Worms	1834	1835
Worms	1835	1836
Worms	1836	1837
Worms	1837	1838
Worms	1838	1839
Worms	1839	1840
Worms	1840	1841
Worms	1841	1842
Worms	1842	1843
Worms	1843	1844
Worms	1844	1845
Worms	1845	1846
Worms	1846	1847
Worms	1847	1848
Worms	1848	1849
Worms	1849	1850
Worms	1850	1851
Worms	1851	1852
Worms	1852	1853
Worms	1853	1854
Worms	1854	1855
Worms	1855	1856
Worms	1856	1857
Worms	1857	1858
Worms	1858	1859
Worms	1859	1860
Worms	1860	1861
Worms	1861	1862
Worms	1862	1863
Worms	1863	1864
Worms	1864	1865
Worms	1865	1866
Worms	1866	1867
Worms	1867	1868
Worms	1868	1869
Worms	1869	1870
Worms	1870	1871
Worms	1871	1872
Worms	1872	1873
Worms	1873	1874
Worms	1874	1875
Worms	1875	1876
Worms	1876	1877
Worms	1877	1878
Worms	1878	1879
Worms	1879	1880
Worms	1880	1881
Worms	1881	1882
Worms	1882	1883
Worms	1883	1884
Worms	1884	1885
Worms	1885	1886
Worms	1886	1887
Worms	1887	1888
Worms	1888	1889
Worms	1889	1890
Worms	1890	1891
Worms	1891	1892
Worms	1892	1893
Worms	1893	1894
Worms	1894	1895
Worms	1895	1896
Worms	1896	1897
Worms	1897	1898
Worms	1898	1899
Worms	1899	1900

Abb. 5: Aufenthalte und Reisen des sächsischen Architekten Robert Wimmer. Datensatz in ZUCCARO. CC-BY-NC-SA 4.0.

Die Anordnung von Suchergebnissen nach Relevanz, Bedeutung oder Umfang eines Sachverhalts wird außerdem erschwert durch die oben beschriebenen Unverhältnismäßigkeiten, die aus der digitalen Segmentierung des Materials resultieren. Lebensstätten, Reisestationen, Studien- und Wirkungsorte eines Künstlers lassen sich in einem Graphenmodell hervorragend und – im Gegensatz zum Karteiformular – in beliebiger Tiefe abbilden. In ZUCCARO wird jeder Aufenthalt als eine zeitlich und modal definierte Beziehung zu einer Lokalität ausgedrückt. Als Beispiel diene hier die – soweit bislang bekannte – Biographie des sächsischen Architekten Robert Wimmer (1829–1907), dessen in der Kunstbibliothek Berlin aufbewahrte Skizzen einer Italienreise kürzlich ausgewertet werden konnten (Abbildung 5). In der Standardansicht erscheint jeder dieser Aufenthalte gleichwertig, ob es sich um einen mehrjährigen Studienaufenthalt handelt oder nur um eine kurze, aber dokumentierte Reisestation. Immerhin enthält der Datensatz die Zeitdauer, so dass die Relevanz mit einem geeigneten Visualisierungsinstrument, etwa dem Geo-Browser von DARIAH-DE,²⁰ anschaulich besser gewürdigt werden kann (Abbildung 6).



Abb. 6: Visualisierung eines Teils der Reisestationen des Architekten Robert Wimmer im Geo-Browser von DARIAH-DE. CC-BY-NC-SA 4.0.

²⁰ Beschreibung des Geo-Browsers und der Benutzeroberfläche.

3.2 Wissensrepräsentation als fragestellungsabhängige Visualisierung

Im Umgang mit Unsicherheiten und Unschärfen spielt die Anordnung des Wissens eine vergleichbar wichtige Rolle. Eine Standardlösung gibt es auch hier nicht. Dazu genügt es, sich die Probleme bei der Anordnung von sowohl präzise als auch unscharf datierten Datensätzen vor Augen zu führen. Hat man einen Datensatz, wo als Datierung der Zeitraum 1408–1415 eingetragen ist, und einen anderen, der präzise auf 1411 datiert ist – welcher wird zuerst ausgegeben? Und was geschieht mit einem weiteren Datensatz, der den Zeitraum 1405–1420 umfasst?

Bei der Anordnung des Wissens spielt der jeweilige Fragekontext eine wichtige Rolle. Nicht in jedem Fall kann das gleiche Sortierkriterium angewandt werden. Wenn man einen Künstler betrachtet, dann möchte man seine Werke vermutlich in der Reihenfolge ihres Entstehens aufgeführt haben; vielleicht möchte man außerdem zunächst nur die sicher zugeschriebenen Werke sehen und die unsicheren später auf Wunsch einblenden. Betrachtet man die Werke aus der Sicht einer Sammlung, dann möchte man sie vielleicht lieber nach Ankaufsdatum oder nach dem Saal, in dem sie hängen, geordnet sehen. Was geschieht in diesen Fällen mit Datensätzen, die zu wenige oder ungenaue Informationen enthalten? Hierzu sind vermutlich für jeden Einzelfall eigene Überlegungen anzustellen. Festzuhalten bleibt, dass *Relevanz* in jedem Betrachtungskontext anders definiert ist. Ob ein befriedigender generischer Algorithmus für das *ranking* überhaupt existiert, muss offenbleiben.

Es ist darüber hinaus grundsätzlich fraglich, ob ein Ergebnis, das durch *opaque algorithms*, also undurchsichtige Berechnungen, zustande kommt, überhaupt als Repräsentation von Wissen aufgefasst werden kann. Zur Wissenschaft gehört die Überprüfbarkeit, und die ist in diesem Fall nicht gegeben. Demzufolge wäre es nur konsequent, das jeweilige Ranking, also die Gewichtung der Datensätze bei der Ausgabe, bei der Datenredaktion explizit festzulegen. Dem stehen jedoch große Schwierigkeiten entgegen. Die eine liegt darin, dass bei einem *property graph* jeder Knoten in den Mittelpunkt der Betrachtung rücken kann. Daher müssten explizite Angaben zum Ranking der zugehörigen Beziehungen im Knoten selber gespeichert und dort gegebenenfalls auch modifiziert werden. Es ist fraglich, ob sich der damit verbundene *overhead* nicht negativ auf die Komplexität des Systems und seine Performanz auswirken würde. Die andere Schwierigkeit entsteht dadurch, dass nicht immer ein einzelner Knoten betrachtet wird, sondern oft eine ganze Auswahl, wie in dem genannten Beispiel der Villa Sacchetti. In diesem Fall gibt es keinen Speicherort für Ranking-Angaben im Datenbestand des Wissensgraphen, sondern das System selbst müsste darüber Buch führen. Wie dies im Datenmodell aussehen könnte, wäre gesondert zu überlegen.

Die Visualisierung von Unsicherheiten und Unschärfen ist ein eigenes Gebiet, das von der Forschung auch in anderen Fächern längst thematisiert worden ist.²¹ In ZUCCARO haben wir zum Beispiel damit experimentiert, bei der Zuschreibung einer Architekturzeichnung an einen historischen Zeichner den jeweiligen Grad der Gewissheit mit unterschiedlichen Farben zu kodieren (Abbildung 7).

The screenshot shows a table with columns: PERSON, INSTITUTION, OFFICE, BUILDING, WORK OF ART, DRAWING, PRINT, LOCATION, SELO, DOCUMENT, BINDER. The table lists several entries, each with a colored square indicating a certainty level: yellow for 'probable', red for 'likely', and green for 'certain'. The entries include details about the drawing's origin, date, and location.

Abb. 7: Kodierung der Zuschreibungswahrscheinlichkeit mit Farben in ZUCCARO (Desktopansicht 2007). CC-BY-NC-SA 4.0.

Besonders augenfällig wird das Thema im Bereich der digitalen räumlichen Modellierung. Ein CAD-Modell kann einen Mauerzug prinzipiell nur in einer ganz konkreten Ausdehnung und Position wiedergeben. Verglichen mit den tatsächlichen historischen Kenntnissen sind die im CAD-System gespeicherten Maßangaben demzufolge oft viel zu präzise und spiegeln dadurch eine Genauigkeit des Wissens vor, die keine fundierte Grundlage hat. Die virtuelle Präzision ist die digitale Kehrseite der *fuzziness* des Wissens. Daher haben sich auf diesem Gebiet bereits verschiedene Strategien im Umgang mit Unsicherheit und Unschärfe herausgebildet.²²

Unsicherheit, Unvollständigkeit und Granularität der Daten werfen weitere Fragen auf. Zwar legt Datenhaltung in einem *property graph* die Möglichkeit nahe, darauf Methoden des automatisierten *reasoning* anzuwenden, also mit Hilfe von Netzwerkalgorithmen implizites Wissen in explizites umzuwandeln und dadurch den Datenbestand zu konsolidieren. Wie weit solche Verfahren durch die genannten Probleme behindert würden, wäre zu prüfen.

Ein automatisiertes Ranking beispielsweise, das sich auf die Anzahl der verknüpften Elemente stützt, oder auf die Häufigkeit von Anfragen, oder Benutzerbewertungen, dürfte stets fehlerbehaftet bleiben. Ebenso problematisch erscheint es, aus den vorhandenen Inhalten automatisiert Schlussfolgerungen zu ziehen oder statistische Auswertungen vorzunehmen. Zwar können derartige Verfahrensweisen den Charakter umfangreicher Simulationen

²¹ Vgl. Bonneau et al. 2014; eher praktisch Yau 2018.
²² Standards für die Kenntlichmachung der hypothetischen Bestandteile eines Modells wurden durch die *London Charter* am 19.09.2017 definiert. Vgl. speziell hierzu auch – aus eher gestalterischer als modeltheoretischer Sicht: Lengyel / Toulouse 2011 und Lengyel / Toulouse 2015.

annehmen, um nicht ausreichende Informationen zu interpolieren und damit wieder in den Datenbestand zurückzuwirken, doch die Zulässigkeit ist fraglich. Derartige Versuche und Überlegungen werden im Projekt *Venice Time Machine* und im Projektverbund *Time Machine* angestellt (Abbildung 8).²³

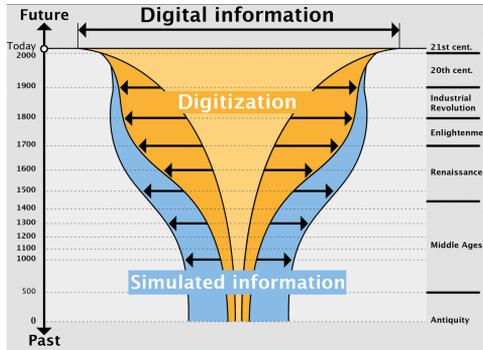


Abb. 8: Frédéric Kaplan: Überbrückung von fehlenden Quellen durch Simulation. Kaplan 2013, fig. 3.

3.3 Fazit

Unsere Erfahrungen bei der zwölfjährigen Arbeit mit dem Informationssystem ZUCCARO sind bescheidener. Als Quintessenz der hier angestellten Überlegungen lässt sich die Erkenntnis formulieren, dass ein kunsthistorisches Informationssystem der Forschung in erster Linie der inhaltlichen Erschließung des historischen Materials dienen sollte. Die Ausarbeitung des Datenmodells und die Algorithmen des Datenbanksystems sollten stärker darauf Rücksicht nehmen, wie man die Inhalte dem Benutzer möglichst übersichtlich und durchschaubar präsentieren kann. Das Datenmodell extrem elaboriert und detailliert zu strukturieren, damit es der historischen Wirklichkeit möglichst nahe kommt, ist demgegenüber zweitrangig und in manchen Fällen sogar kontraproduktiv. Eine möglichst getreue Simulation einer geschichtlichen Wirklichkeit anzustreben, sollte nicht das Ziel eines *knowledge graph* sein. Unsicherheiten im Datenbestand zu dokumentieren und für den Benutzer kenntlich zu machen, ist sinnvoll – sie automatisiert auszuwerten jedoch nicht.

²³ Vgl. Kaplan 2013 und Kaplan 2017.

Bibliographische Angaben

Georges-Pierre Bonneau / Hans-Christian Hege / Chris R. Johnson / Manuel M. Oliveira / Kristin Potter / Penny Rheingans / Thomas Schultz: Overview and State-of-the-Art of Uncertainty Visualization. In: Scientific visualization: uncertainty, multifield, biomedical, and scalable visualization. Hg. von Charles D. Hansen / Min Chen / Christopher R. Johnson / Arie E. Kaufman / Hans HageLondon u.a. 2014, S. 3-27. (= Mathematics and Visualization, 37) [[Nachweis im GBV](#)]

Thomas Frisendal: Property Graphs: The Swiss Army Knife of Data Modeling. In: dataversity.net. Big Data Blogs. Blogbeitrag vom 22.09.2017. [[online](#)]

Edith Hamilton: Spokesmen for God. The Great Teachers of the Old Testament. New York 1936. [[Nachweis im GBV](#)]

Frédéric Kaplan: Lancement de la »Venice Time Machine«. In: fkaplan.wordpress.com. Frederic Kaplan. Blogbeitrag vom 14.03.2013. [[online](#)]

Frédéric Kaplan / Isabella di Lenardo: Big Data of the Past. In: Frontiers in Digital Humanities (2017). Artikel vom 29.05.2017. DOI : [10.3389/fdigh.2017.00012](https://doi.org/10.3389/fdigh.2017.00012)

Elisabeth Kieven / Georg Schelbert: Architekturzeichnungen, Architektur und digitale Repräsentationen. Das Projekt LINEAMENTA. In: kunsttexte.de 4 (2014). DOI: [10.18452/6832](https://doi.org/10.18452/6832)

Elisabeth Kieven: Research Infrastructures for Historic Artefacts: Knowledge Networks. In: Research Infrastructures in the Digital Humanities. Hg. von der European Science Foundation. (ESF Science Policy Briefing: 42, Straßburg, 09.2011) Straßburg 2011, S. 13-15. [[online](#)]

Dominik Lengyel / Catherine Toulouse: Darstellung von unscharfem Wissen in der Rekonstruktion historischer Bauten. In: Von Handaufmaß bis High Tech III. 3D in der historischen Bauforschung. Hg. von Katja Heine. Darmstadt u.a. 2011, S. 182-186. [[Nachweis im GBV](#)]

Dominik Lengyel / Catherine Toulouse: Die Bedeutung architektonischer Gestaltung in der Vermittlung von Unschärfe. In: gams.uni-graz.at. Präsentation vom 25.02.2015. (DHd 2015, Graz, 23.-27.02.2015) Graz 2015. [[online](#)]

Alan M. MacEachren: Visualizing Uncertain Information. DOI: [10.14714/CP13.1000](https://doi.org/10.14714/CP13.1000) In: Cartographic Perspectives 13 (1992), S. 10-19. [[online](#)] [[Nachweis im GBV](#)]

Martin Raspe / Georg Schelbert: ZUCCARO. Ein Informationssystem für die historischen Wissenschaften. In: IT Information Technology 51 (2009), H.4, S. 207-215. DOI: [10.11588/artdok.00005812](https://doi.org/10.11588/artdok.00005812) [[Nachweis im GBV](#)]

Nathan Yau: Visualizing the Uncertainty in Data. In: flowingdata.com. Guides. Beitrag vom 08.01.2018. [[online](#)]

Abbildungslegenden und -nachweise

Abb. 1: Schematisches Datenmodell des Informationssystems ZUCCARO (Stand 2013). CC-BY-NC-SA 4.0.

Abb. 2: Unbekannter Künstler (Gianlorenzo Bernini?): Entwurf für ein Lustgebäude, vermutlich eine Villa in Rom. Datensatz in ZUCCARO. CC-BY-NC-SA 4.0.

Abb. 3: Freimut Woessner: Der Mann mit dem Sturzhelm, 1991, Zeichnung, Archiv des Künstlers. Mit freundlicher Genehmigung des Künstlers.

Abb. 4: Rom, Villa del Pigneto Sacchetti. Datensatz in ZUCCARO. CC-BY-NC-SA 4.0.

Abb. 5: Aufenthalte und Reisen des sächsischen Architekten Robert Wimmer. Datensatz in ZUCCARO. CC-BY-NC-SA 4.0.

Abb. 6: Visualisierung eines Teils der Reisestationen des Architekten Robert Wimmer im Geo-Browser von DARIAH-DE. CC-BY-NC-SA 4.0.

Abb. 7: Kodierung der Zuschreibungswahrscheinlichkeit mit Farben in ZUCCARO (Desktopansicht 2007). CC-BY-NC-SA 4.0.

Abb. 8: Frédéric Kaplan: Überbrückung von fehlenden Quellen durch Simulation. Kaplan 2013, fig. 3.