

Datenmodellierung und -visualisierung mit Graphdatenbanken. Konzepte und Erfahrungen anlässlich des Relaunches der Bilddatenbank REALonline

Matschinegg, Ingrid

ingrid.matschinegg@sbg.ac.at
Universität Salzburg, Österreich

Nicka, Isabella

isabella.nicka@sbg.ac.at
Universität Salzburg, Österreich

Mit der Thematik der digitalen Nachhaltigkeit sind Gedächtnisinstitutionen, die sich die umfassende digitale Sicherung und Erschließung des Kulturerbes zum Ziel setzen, gleichermaßen konfrontiert wie Forschungsprojekte, bei denen digitale Daten generiert, computergestützt ausgewertet und in Forschungsdatenbanken zugänglich gemacht werden. Dabei liegt das Hauptaugenmerk nicht nur auf der Langzeitsicherung und Archivierung; die größeren Herausforderungen stellen sich vor allem bei den Aktualisierungen der Datenmodelle, Analysemethoden und Präsentationsformen, um die neuen Werkzeuge der Digital Humanities bestmöglich einsetzen zu können. Die Überlegungen im Vorfeld derartiger Relauncharbeiten bewegen sich erfahrungsgemäß zwischen vorsichtiger Adaptierung und radikalem Umbau der vorliegenden Datenarchitektur. Dass dabei alle vorhandenen Informationen verlustfrei übertragen werden sollen, versteht sich von selbst. Der folgende Beitrag möchte die Transformation von Daten eines Langzeitprojekts in eine neue Datenarchitektur für eine Bilddatenbank vorstellen, bei der eine Graphdatenbank zum Einsatz kommt:

Am Institut für Realienkunde des Mittelalters und der frühen Neuzeit (IMAREAL), einem interdisziplinär ausgerichteten Forschungsinstitut, das Teil der Universität Salzburg ist, wird die materielle Kultur des Mittelalters und der frühen Neuzeit untersucht. Bildquellen bilden dabei neben Schriftquellen und überlieferten Objekten die Grundlagen der Analysen. Mit dem Aufbau der Bilddatenbank REALonline wurde am IMAREAL in den 1970ern auf der Grundlage der von Manfred Thaller speziell für die Anforderungen der historischen Grundwissenschaften entwickelten Datenbanksysteme

begonnen – zunächst *Descriptor* und in weiterer Folge ##### (Thaller 1980 u. 1989). Der Datenbestand von REALonline wurde seither und wird weiterhin kontinuierlich erweitert, damit dargestellte Dinge und ihre Kontexte erforscht werden können. Die Datenbank ist seit 2002 unter <http://tethys.imareal.sbg.ac.at/realonline> online verfügbar (Matschinegg 2004). Anhand der Datenbank ist es möglich, die Bedeutung und Funktion der materiellen Kultur im Bilddiskurs zu untersuchen: Welche Objekte waren zu welchen Zeiten in welchen Gesellschaften und Kontexten als visuelle „Requisiten“ gegenwärtig oder vor- und damit auch darstellbar? Wie wurden Dinge im Bild verhandelt und welche Rolle nehmen sie innerhalb von ins Bild überführten Narrativen ein?

Um diese Fragen beantworten zu können, wurde am IMAREAL entschieden, neben den Metadaten zum Werk bzw. Bildträger systematisch *alle* im Bild dargestellten Elemente auszuzeichnen (Abb. 1). Im Gegensatz zu anderen Bilddatenbanken wird das Dargestellte nicht nur mit einigen wenigen Schlagwörtern erfasst. Diesem Umstand ist es zu verdanken, dass die in REALonline erhobenen Daten sowohl im Rahmen von interdisziplinären Forschungen zur materiellen Kultur ausgewertet werden können, als auch in unterschiedlichen geisteswissenschaftlichen Untersuchungskontexten und für Kulturerbedokumentationen eine wertvolle Ressource darstellen.

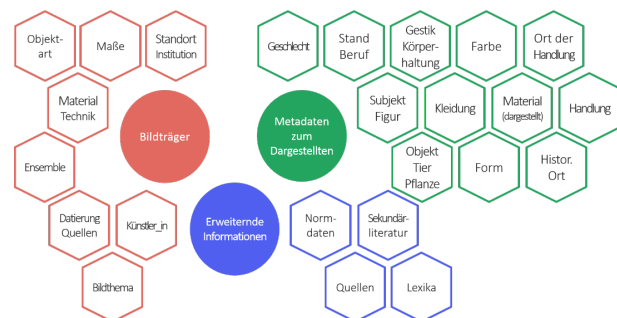


Abb. 1: Erfassungsschema der Metadaten in REALonline

Im Modell für die Erfassung der dargestellten Entitäten im Bild werden folgende Informationen erhoben: Für Subjekte werden neben dem Subjektnamen die Kategorien Geschlecht, Beruf bzw. Stand und Gestik erfasst. Bei Objekten werden der Objektname und die Informationen zu Farbe, Material und Form erhoben. Weiters wird die Struktur dieser Metadaten zu den Bildinhalten festgehalten und kann damit im Rahmen von Analysen abfragbar gemacht werden: Direkte Subjekt-Objekt- bzw. Objekt-Objekt-Relationen werden erfasst, um am Körper getragene bzw. von Figuren gehaltene Objekte zu dokumentieren oder einen Bezug zwischen einzelnen dargestellten Dingen (etwa ein auf einem Tisch stehender Krug) in den Daten abbilden zu können. Darüber hinaus können sowohl Körperteile als auch Teile von Objekten als Metadaten zum Dargestellten gespeichert werden (Jaritz 1993: 23-43).

Graphdatenbanken eignen sich u.a. besonders dafür, die vielfältigen Beziehungen zwischen Personen oder Personen und Gegenständen bzw. Ereignissen sowie auch zwischen den Dingen untereinander möglichst flexibel abzubilden und sind nun auch in der historischen Forschung im Kommen; als Software wird oft Neo4j eingesetzt (Raspe 2014, Kaufmann & Andrews 2015, Kuczera 2015). Die verzweigte Struktur der erfassten Metadaten in REALonline ist einer der Hauptgründe, warum für die neue Datenarchitektur ein property-graph-Modell gewählt wurde. Ein weiterer Leitgedanke war, dass das Beziehungsnetz von Subjekten, Objekten und Handlungen in mittelalterlichen und frühneuzeitlichen Bildern anhand des Modells eines verzweigten Graphen besser veranschaulicht werden kann als in einer langen Liste mit Metadaten. Die Graphdatenbank bietet im Fall von REALonline sowohl bei der Präsentation für die Nutzer_innen im Frontend, für die Abfrage und Darstellung der Abfrageergebnisse als auch für die Eingabe der Daten im Backend (siehe Abb. 2) eine Verbesserung der Usability gegenüber dem zuvor verwendeten hierarchischen Datenbankmodell.

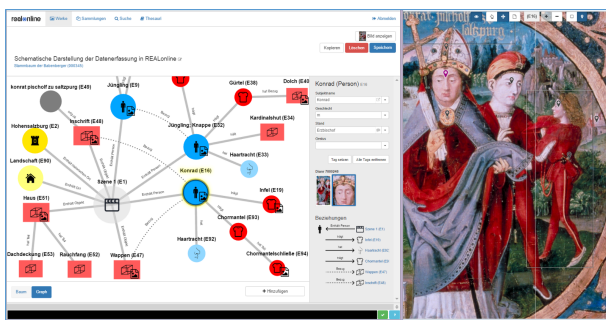


Abb. 2: Screenshot des Graphen zum Dargestellten im Bild (Backend)

In unserem Projekt hat sich die Kombination von Neo4j für die Modellierung und Abfrage der „beziehungsrelevanten“ Daten mit einer NoSQL-Mongo-Dokumentendatenbank angeboten (Abb. 3). Diese Lösung baut einerseits auf dem in der Praxis bereits bewährten softwareseitigen Ineinandergreifen bei der semantischen Transformation der Informationen auf und bietet gleichzeitig die Möglichkeit zur Speicherung und Abfrage von werkgeschichtlich wie auch projektgeschichtlich relevanten Informationen zu den einzelnen Bilddokumenten, die im Verlauf dieses Langzeitprojektes erhoben wurden und laufend erweiterbar bleiben sollen.

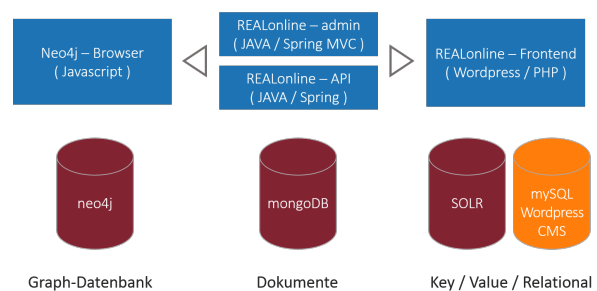


Abb. 3: Datenarchitektur von REALonline

Im Vortrag möchten wir aber auch auf die Herausforderungen hinweisen, denen wir uns im Zuge des Entwurfs der Datenarchitektur von REALonline stellen mussten: So war etwa in der Struktur des hierarchischen Datenmodells die Information zur dargestellten Handlung auf derselben Ebene angesiedelt wie die Entitäten Subjekt und Objekt. Beim Datenexport aus der bis dato verwendeten #####-Datenbank und dem Import in Neo4j konnte – nachdem in diesem Fall keine automatisierten Zuweisungen der Handlung zu Personen bzw. Objekten möglich waren – dieser Umstand nur in das neue Datenmodell mitübernommen werden. Mit der Entscheidung für eine Graphdatenbank ist dennoch gewährleistet, dass in einem weiteren Schritt Informationen, wie jene zur dargestellten Handlung, statt in Knoten in die Kanten des Graphen gelegt werden können und damit die Struktur von RDF-Triples (Subjekt-Prädikat-Objekt) bekommen.

Aufgrund der zeitintensiven Datenerhebung war ein wichtiger Aspekt des Relaunchs, die Dateneingabe so effizient wie möglich zu gestalten. Der Beitrag wird die gefundene Lösung präsentieren. Langfristig gesehen sollte versucht werden, den Zeitaufwand für die Erhebung von Metadaten zu den auf historischen Bildern dargestellten Elementen zu minimieren. Daher möchten wir die in REALonline während mehr als 40 Jahren erhobenen Informationen als Trainingsdaten in transdisziplinäre Projekte zwischen den Geisteswissenschaften und der Computer Vision – insbesondere zur (semi-)automatisierten Bilderkennung – einbringen, so dafür Fördermittel eingeworben werden können.

Mit dem Relaunch von REALonline kann die Menge der erhobenen Metadaten zum im Bild Dargestellten (aktuell sind innerhalb von 23316 Datensätzen 1.165562 Begriffe dazu erfasst) besser zugänglich gemacht werden: Abfragen der Graphdatenbank und Visualisierungen (z.B. mit Software wie *gephi*–The Open Graph Viz Platform oder *yEd graph editor*) dieser Ergebnisse können komplexe Zusammenhänge innerhalb der Bilddetails aufdecken oder Aufschlüsse zu Mustern sowie „Ausreißern“ in unterschiedlichen Samples liefern, die nicht nur als Resultate statistischer Auswertungen verstanden werden sollen, sondern vor allem dazu dienen können, neue Fragen in der (interdisziplinären) Forschung anzustoßen.

Beispielsweise wurde in 154 Datensätzen das Bildthema „Geißelung Christi“ erfasst. Bei der Erschließung dieser Datensätze wurden wiederum 516 Objekte verzeichnet, die von Figuren im Bild in der Hand gehalten werden. Die Visualisierung (Abb. 4) beschränkt sich auf jene Objekte, die nur einmal vorkommen (gelb) und die ihnen übergeordneten Thesauruskategorien (grün, dunkelrot). Während die meisten Objekte dem gängigen Narrativ „Geißelung“ zugeordnet werden können, sind die Objekte *Münze* und *Geldbeutel* nur über einen Konnex zum mittelalterlichen Drama erklärbar (Nicka 2014, 280–282): Die Darstellung einer Bezahlung der Geißler Christi, die nur auf einem Flügelaltar im niederösterreichischen Pöggstall im Bild festgehalten wurde, kennen wir ansonsten nur aus Passionsspielen, wo jüdische Protagonisten negativ gekennzeichnet werden, indem sie den Gerichtsknechten Geld geben, um besonders fest mit den Ruten zuzuschlagen (siehe auch Abb. 2).

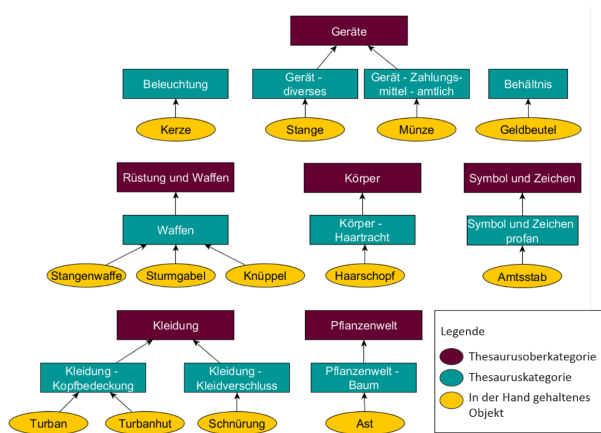


Abb. 4: Visualisierung der dargestellten Objektbegriffe und Körperbezeichnungen aus den Bildbeschreibungen in ihrer Zuordnung zur jeweiligen Thesauruskategorie

Abschließend bleibt zu erwähnen, dass sich mit der Notwendigkeit, eine gut eingeführte, aber in ihren technischen Funktionalitäten nicht mehr zeitgemäße Bilddatenbank zu modernisieren, auch die Chance zur besseren Nutzung der umfangreichen Datenbestände verbinden lässt. Im Zuge der Relaunchvorarbeiten haben wir die Konzepte und Lösungsansätze aufgegriffen, die gegenwärtig in den Digital Humanities diskutiert und getestet werden. Wir haben die Umsetzung in enger Zusammenarbeit mit den Grazer Entwicklerfirmen complement.at und zedlacher.net realisiert. Der Beitrag gibt einen knappen Überblick über die wichtigsten Entscheidungsfindungsprozesse sowie die Schwierigkeiten und Potentiale, die bei der Überführung in die neue Datenarchitektur und die gewählte Frontend-Lösung entstanden sind. Der Aspekt der Nachhaltigkeit hat dabei von Anfang an eine große Rolle gespielt; sowohl bei der Erhaltung aller vorhandenen Informationen als auch bei der nachhaltigen Nutzbarkeit der erhobenen Daten.

So ist die Zitierbarkeit der Daten über einen PID (persistent identifier) mit einem handle gewährleistet und die Metadaten werden mit einer *Creative Commons by-nc-sa* 4.0-Lizenz zur Verfügung gestellt. Die neue Online-Version von REALonline wird gegenwärtig getestet und optimiert und 2017 freigeschaltet.

Bibliographie

Jaritz, Gerhard (1993): *Images: A Primer of Computer-Supported Research with ##### IAS*. Halbgraue Reihe zur historischen Fachinformatik A 22. St. Katharinen: Scripta Mercaturae Verlag.

Kaufmann, Sascha / Andrews, Tara Lee (2016): „Bearbeitung und Annotation historischer Texte mittels Graph-Datenbanken am Beispiel der Chronik des Matthias von Edessa“, in: *DHd 2016: Modellierung - Vernetzung - Visualisierung* 176–178 <http://dhd2016.de/boa.pdf> [letzter Zugriff 20. August 2016].

Kuczera, Andreas (2015): „Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi“, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120) [letzter Zugriff 20. August 2016].

Matschinneg, Ingrid (2004): „REALonline – IMAREAL's Digital Image-Server“, in: *[Enter the Past]. The E-way into the Four Dimensions of Cultural Heritage. CAA 2003 | Computer Applications and Quantitative Methods in Archaeology | Proceedings of the 31st Conference*, Vienna, Austria, April 2003 (BAR International Series 1227). Oxford: archaeopress, 214–216.

Nicka, Isabella (2014): „Interfaces. Berührungszonen von Transzendenz und Immanenz im spätmittelalterlichen Sakralraum“, in: Meyer, Marion / Klimburg-Salter, Deborah (eds.): *Visualisierungen von Kult.* Wien / Köln / Weimar: Böhlau, 260–293, Abb. auf 438–444.

Nicka, Isabella (im Erscheinen): „REALonline–Explore and Find Out. Wohin führt das Digitale die Kunstgeschichte?“, Beitrag zum Tagungsband der vom 6.–8. Nov. 2015 in Wien abgehaltenen Konferenz „Newest Art History“. *Wohin geht die jüngste Kunstgeschichte?*

Raspe, Martin (2014): *Zuccaro. Ein modernes, konfigurierbares Informationssystem für die Geisteswissenschaften*. <http://zuccaro.biblherz.it/dokumentation/zuccaro> [letzter Zugriff 20. August 2016].

Thaller, Manfred (1980): „Descriptor: Probleme der Entwicklung eines Programmsystems zur computerunterstützten Auswertung mittelalterlicher Bildquellen“, in: *Europäische Sachkultur des Mittelalters: Gedenkschrift aus Anlaß des 10jährigen Bestehens des Instituts für mittelalterliche Realienkunde Österreichs* (Veröffentlichungen des Instituts für mittelalterliche Realienkunde Österreichs 4 / Sitzungsberichte der Akademie der Wissenschaften, Phil.-Hist. Klasse 374). Wien: Verlag der Österreichischen Akademie der Wissenschaften, 167–194.

Thaller, Manfred (1989): #####. Ein Datenbanksystem.
Halbgraue Reihe zur historischen Fachinformatik B1. St.
Katharinen: Scripta Mercaturae Verlag.