



GOEDOC – Dokumenten- und Publikationsserver der Georg-August-Universität Göttingen

2019

Daten neu verknoten

–
Die Verwendung einer Graphdatenbank für die Bilddatenbank REALonline

Ingrid Matschinegg, Isabella Nicka, Clemens Hafner, Martin Stettner, Stefan Zedlacher

DARIAH-DE Working Papers

Nr. 31

Matschinegg, I.; Nicka, I.; Hafner, C.; Stettner, M.; Zedlacher, St.: Daten neu verknoten : Die Verwendung einer Graphdatenbank für die Bilddatenbank REALonline
Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2019
(DARIAH-DE working papers 31)

Verfügbar:

PURL: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2019-3>
URN: <http://nbn-resolving.org/urn:nbn:de:gbv:7-dariah-2019-3-5>

Dieser Beitrag erscheint unter der Lizenz [Creative Commons Attribution 4.0 \(CC-BY\)](#)



Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <<http://dnb.dnb.de>> abrufbar.

Erschienen in der Reihe
DARIAH-DE working papers

ISSN: 2198-4670

Herausgeber der Reihe
DARIAH-DE, Niedersächsische Staats- und Universitätsbibliothek

Mirjam Blümm, Thomas Kollatz, Stefan Schmunk und Christof Schöch

Abstract: Dieser Beitrag beschreibt den Relaunch von REALonline, der Bilddatenbank des Instituts für Realienkunde des Mittelalters und der frühen Neuzeit in Krems (Universität Salzburg). Viele Projekte, in denen über Jahrzehnte Daten erhoben wurden, müssen sich der Herausforderung stellen, die Daten an die heutigen Möglichkeiten und Standards anzupassen. Der Beitrag skizziert die Überlegungen, die bei der Entwicklung der neuen Datenarchitektur maßgeblich waren und in den Einsatz einer Graphdatenbank mündeten. Ihr Vorteil besteht vor allem in den verbesserten Möglichkeiten der Speicherung und Abfragemöglichkeiten der unterschiedlichen Beziehungen zwischen den erfassten Entitäten, die zudem in der Graphenvisualisierung – sowohl im Frontend als auch im Backend-Editor – einfach nachvollziehbar gemacht werden können. Im neo4j-Browser werden die neuen Abfrage-Möglichkeiten der Graphdatenbank den Nutzer*innen online zugänglich gemacht.

Keywords: Graphdatenbank, Annotation, Datenvisualisierung, Bilddatenbank, Datenarchitektur, Remodellierung

Graphdatabase, Annotation, Data Visualization, Image Database, Data Architecture, Remodelling

Daten neu verknoten

Die Verwendung einer Graphdatenbank für die Bilddatenbank REALonline

Ingrid Matschinegg Isabella Nicka Clemens Hafner
Martin Stettner Stefan Zedlacher

Institut für Realienkunde des Mittelalters und der frühen Neuzeit (IMAREAL)



Ingrid Matschinegg, Isabella Nicka, Clemens Hafner, Martin Stettner, Stefan Zedlacher: „Daten neu verknoten“. [DARIAH-DE Working Papers](#) Nr. 31. Göttingen: DARIAH-DE, 2019.

URN: [urn:nbn:de:gbv:7-dariah-2019-3-5](#).

Dieser Beitrag erscheint unter der
Lizenz [Creative-Commons Attribution 4.0](#) (CC-BY).

Die Dariah-DE Working Papers werden von Mirjam Blümm,
Thomas Kollatz, Stefan Schmunk und Christof Schöch
herausgegeben.

Mitherausgeber dieses Working Papers ist Andreas Kuczera
(ADWL Mainz)



Dieser Beitrag ist im Rahmen der Tagung Graphentechnologien entstanden, die am 19. und 20. Januar 2017 an der Akademie der Wissenschaften und der Literatur Mainz (ADWL Mainz) in Zusammenarbeit mit Dariah-DE und Historical Network Research (HNR) stattfand.

Timo Kissinger (ADWL Mainz) sei gedankt für die überaus konstruktive Unterstützung des Lektorats.

Zusammenfassung

Dieser Beitrag beschreibt den Relaunch von REALonline, der Bilddatenbank des Instituts für Realienkunde des Mittelalters und der frühen Neuzeit in Krems (Universität Salzburg). Viele Projekte, in denen über Jahrzehnte Daten erhoben wurden, müssen sich der Herausforderung stellen, die Daten an die heutigen Möglichkeiten und Standards anzupassen. Der Beitrag skizziert die Überlegungen, die bei der Entwicklung der neuen Datenarchitektur maßgeblich waren und in den Einsatz einer Graphdatenbank mündeten. Ihr Vorteil besteht vor allem in den verbesserten Möglichkeiten der Speicherung und Abfragemöglichkeiten der unterschiedlichen Beziehungen zwischen den erfassten Entitäten, die zudem in der Graphenvisualisierung – sowohl im Frontend als auch im Backend-Editor – einfach nachvollziehbar gemacht werden können. Im neo4j-Browser werden die neuen Abfrage-Möglichkeiten der Graphdatenbank den Nutzer*innen online zugänglich gemacht.

Schlagwörter

Graphdatenbank, Annotation, Datenvisualisierung, Bilddatenbank, Datenarchitektur, Remodellierung

Keywords

Graphdatabase, Annotation, Data Visualization, Image Database, Data Architecture, Remodelling

Inhaltsverzeichnis

1 Eine Datenbank mit Geschichte aktualisieren	4
2 Was haben Bildbeschreibungen und Graphen gemeinsam?	5
3 Ausgangslage	6
3.1 Erschließen, was auf Bildern zu sehen ist	6
3.2 Das bestehende System: KLEIO – ein hierarchisches DBMS	8
4 Graphentechnologien in REALonline	10
4.1 Das konzeptionelle Datenmodell von REALonline	10
4.1.1 Daten- und Graph-Dokumente im konzeptionellen Datenmodell	11
4.1.2 Schemadefinition	13
4.1.3 Definition des Applikations-Datenmodells	14
4.1.4 Speicherung in mongoDB	14
4.2 Die Transformation des hierarchischen Systems in einen Graphen	15
4.3 Graphentechnologien im Backend-Editor von REALonline	17
4.4 Konzeptuelle und funktionale Anforderungen an das Frontend	21
4.4.1 Überlegungen zur Visualisierung mit Graphen	21
4.4.2 Wie kommt der Graph ins REALonline-Frontend?	24
4.5 Die Daten von REALonline via Neo4j abfragen	26
5 Zusammenfassung	32
Literaturverzeichnis	35

1 Eine Datenbank mit Geschichte aktualisieren

Am Institut für Realienkunde des Mittelalters und der frühen Neuzeit (IMAREAL), einem interdisziplinär ausgerichteten Forschungsinstitut der Universität Salzburg, wird die materielle Kultur des Mittelalters und der frühen Neuzeit untersucht. Bildquellen stellen dabei gemeinsam mit Schriftquellen und überlieferten Objekten die wichtigsten Gegenstände von Analysen dar. Bereits in den 1970er Jahren wurde am IMAREAL mit dem Aufbau der Bilddatenbank REAL begonnen, die dann im Jahr 2002 – mit der Erstellung der ersten Internetversion – in REALonline umbenannt wurde. Ziel dieses Langzeitprojektes war und ist es, die visuellen Medien für die Forschung und darüber hinaus für die interessierte Öffentlichkeit zugänglich zu machen. Konkret geht es darum, die auf Bildern dargestellten Dinge und ihre Kontexte systematisch zu erschließen und auszuwerten, sowohl hinsichtlich ihrer Bedeutung und Funktion im Bilddiskurs als auch im Rahmen interdisziplinärer Untersuchungen zur Funktion der Dinge in der alltäglichen Lebensgestaltung.

Mittlerweile ist REALonline eine über vier Jahrzehnte gewachsene Bild- und Textdatenbank, bestehend aus den digitalisierten fotografischen Aufnahmen (bzw. seit einigen Jahren Digitalfotos) sowie den Daten zu den Bildern (im folgenden ‚Werksdaten‘ genannt) und der detaillierten, strukturierten Beschreibungen der Daten auf den Bildern (im folgenden ‚Beschreibungsdaten‘ genannt). Bezuglich der Auswahl werden Bildträger und visuelle Medien so breit wie möglich berücksichtigt: Tafelmalerei, Wandmalerei, illuminierte Handschriften, Druckgrafik, bildtragende materielle Objekte und viele andere mehr. Aufgenommen wurden Bildquellen vom 11. bis zum 18. Jahrhundert, die Mehrzahl der erfassten Werke datiert in den Zeitraum zwischen dem 14. und dem 16. Jahrhundert. Die geografischen Standorte der aufgenommenen Werke liegen in Österreich und angrenzenden Regionen (Bayern, Südtirol, Tschechien, Slowakei, Slowenien, Ungarn) sowie in Siebenbürgen (Rumänien). Insgesamt sind derzeit rund 22.500 Datensätze zu visuellen Medien erfasst. Der Datenbestand wurde und wird kontinuierlich erweitert. Um diese für die interdisziplinäre Forschung so wichtige digitale Ressource auch außerhalb des Instituts benutzen zu können, wurde eine Onlineversion mit einfachen Suchfunktionen entwickelt, die seit 2001 kostenfrei im Internet zugänglich ist (Matschinegg 2004). Komplexe Abfragen konnten aber weiterhin nur von Expert*innen am IMAREAL durchgeführt werden. Die mittlerweile veralteten Funktionalitäten der ersten Webversion bei gleichzeitig rasanten Entwicklungen in der Webtechnologie (Stichwort Web 2.0, 3.0 ...) machten einen grundlegenden Relaunch nötig, der die Erneuerung der gesamten Datenbank umfasste. Im Vorfeld des Relaunchs haben wir die Datenbanken von einigen kommerziellen Museumsanwendungen dahingehend geprüft, ob damit der vollständige Datentransfer und die erwünschten Abfragefunktionen möglich sind. Da dies nur mit Abstrichen realisierbar gewesen wäre, haben wir uns für eine „maßgeschneiderte Lösung“ entschieden. Die Anforderungen für das Backend waren: die Entwicklung eines User-Interface für die Datenerfassung, die Verwaltung der Thesauri und ein Annotationstool zum Taggen von Beschreibungs begriffen direkt am Bild. Dass die neue Datenbank im Frontend mit einer Zoomfunktion für die Bilder ausgestattet sein sollte, zählte ebenso zu den Anforderungen wie die verbesserten Suchfunktionalitäten und die übersichtliche Anzeige der Ergebnisse am Monitor. Sowohl bei der Datenpräsentation im neuen User-Interface als auch bei der Expert*innensuche im Neo4j-Browser kommen Graphentechnologien zur Anwendung, auf die wir im Folgenden noch genauer eingehen möchten.

2 Was haben Bildbeschreibungen und Graphen gemeinsam?

Bilder funktionieren anders als Text. Das wird letztlich auch an der Art ablesbar, wie Bilder von Menschen rezipiert werden:

Wenn wir mehr als den ersten Eindruck über ein Bild herausfinden wollen, wenden wir uns Einzelheiten und Strukturen zu, die wir – je nach Informationsstand, Bildung und Sehroutinen – auf das Gesamt [sic] des Bildes zurückbeziehen. (Klotz 2016, 183)

Auch anhand von historischen Bildbetrachtungen wird ersichtlich, dass Bezüge zwischen semantisch fassbaren Elementen eines Bildes hergestellt wurden. Hier sei etwa auf einen Text aus der Zeit um 1400 verwiesen. Es handelt sich bei der nachstehenden Stelle um die Schilderung eines (fiktiven) Bildes auf dem der Tod, mit Schaufel und Haue bewaffnet, Menschen jeglichen Standes bekämpft:

Doch / wanndu vns' / zu Rome jn einen tempel / an einer want gemalett' / als ein man / auff einem ochssen, / dem die augen verpunden waren' syczend / Derselbe man < führent > ein hawen' / jn seiner rechten hant' / vnde ein schauffel' / jn der linckten. / Domit vacht er' / auf dem ochssen. / Gegen jm sluge' / warff / vnd streyt / ein michel menig volckes, / allerley lewt' / jgliches mensch mit seynes hantwercks gezewge. / Do was auch die nunne / mit deme psalter. Die slugen / vnd wurffen / den man auff dem ochssen. [...]

Neuhochdeutsche Übersetzung:

Indes, / wähntest Du / UNS / zu Rom in einem Tempel / an einer Wand gemalt / als Mann, / auf einem Ochsen, / dem die Augen verbunden waren, / sitzend. / Dieser nämliche Mann führte eine Haue / in seiner rechten Hand / und eine Schaufel / in der linken; / damit focht er auf / dem Ochsen. / Gegen ihn schlug, / warf / und stritt / eine große Menge Volks, / jeder Art Leute, / jegliches Menschenkind / mit seines Handwerks Waffe. / Darunter war auch die Nonne / mit dem Psalter. / Die schlugen / und bewarfen / den Mann auf dem Ochsen.¹

Es werden also in der textlichen Repräsentation Bezüge zwischen den Elementen dieses (fiktiven) Bildes, wie etwa der Konnex zwischen dem Mann und dem Ochsen oder zwischen dem Mann und den Menschen, hergestellt.

Anhand dieses Beispiels lässt sich demonstrieren, worin die Herausforderung bei der Vermittlung dieser Zusammenhänge in Form eines Texts besteht: So kann aufgrund der Satzstellung nicht mit Sicherheit gesagt werden, ob wir uns vorzustellen haben, dass der Mann oder der Ochse mit verbundenen Augen dargestellt wurde – auch wenn aufgrund von Vergleichen mit bildlichen wie literarischen Quellen der Zeit einiges für Ersteres spricht (Bertau 1994, Bd. 2, 317). Im zweiten Satz wird dagegen anhand von zwei Textelementen erkennbar, dass es dem Autor Johannes von Tepl offenbar besonders wichtig war,

¹Johannes von Tepl, Der Ackermann aus Böhmen, Ausschnitt von Kapitel 16, Transkription und Übersetzung zitiert nach der Edition von Bertau (1994, Bd. 1, 53–54 und 187–188).

die Bezüge zwischen den Bildelementen möglichst deutlich zu machen: Hier wird mit den Worten „Derselbe man“ einleitend die Definition geliefert, dass es sich bei den darauffolgenden Erläuterungen nach wie vor um die im ersten Satz beschriebene Figur handelt. Auch die erneut angegebene Position des Mannes, nämlich „auf dem ochssen“, am Ende des Satzes, unterstreicht diesen Umstand.

Gerade aufgrund dieser vielfachen Bezüge und der unter Umständen auch sehr vernetzten Beschaffenheit der Informationen, die in Bildbeschreibungen auftreten können, eignen sich Graphstrukturen besonders gut für die Modellierung dieser Daten. Die Informationen aus dem vorherigen Textzitat könnte man beispielsweise in dieser Form in einen Graphen transformieren (Abb. 1):

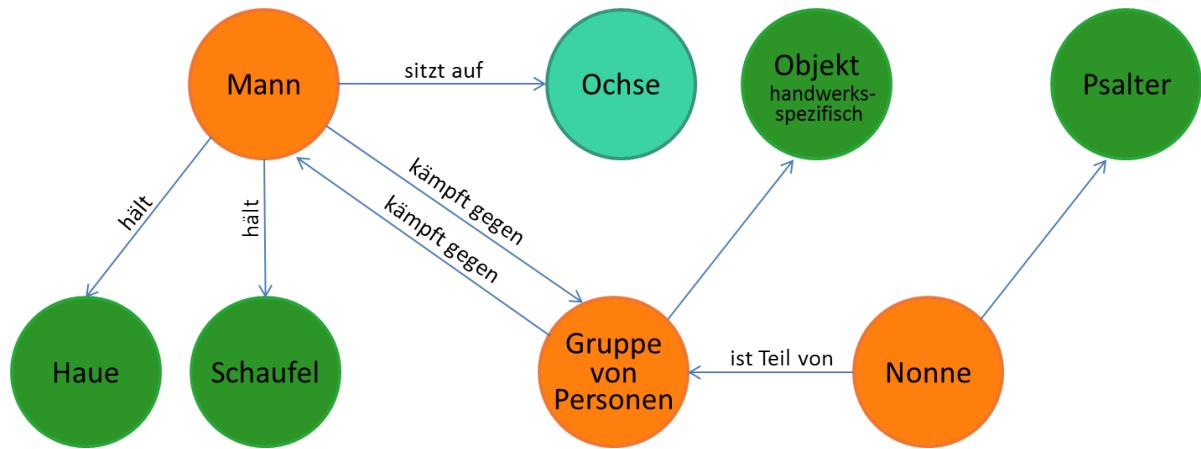


Abbildung 1: Die Informationen der Beschreibung eines fiktiven Bildes aus dem Ackermann aus Böhmen von Johannes Tepl als Graph.

Nachdem mehrere Verbindungen von einem Knoten ausgehen können, ist hier beispielsweise auf Anhieb gut zu erkennen, dass es sich um denselben Mann handelt, der einerseits auf einem Ochsen sitzt und andererseits gegen eine Gruppe von Menschen kämpft.

3 Ausgangslage

Ehe im Detail erklärt werden soll, wie wir unsere Daten neu „verknotet“ haben, möchten wir an dieser Stelle noch etwas ausführlicher schildern, welche Datensammlung die Grundlage für diesen Relaunch geboten hat.

3.1 Erschließen, was auf Bildern zu sehen ist

In den vergangenen Jahrzehnten sind beeindruckende Mengen an historischen Bildquellen virtuell verfügbar und – vermehrt auch über kostenfrei zugängliche Plattformen – recherchierbar gemacht worden. Allerdings kann trotz dieser Fülle nur in seltenen Fällen nach ikonografischen Details gesucht werden.

Grund dafür ist hauptsächlich, dass das Erfassen von semantischen Bildinhalten einen sehr zeit- und damit auch kostenintensiven Aspekt der Bildquellenerschließung darstellt (Bry, Schemainda und Schefels 2015, [2]) und gegenwärtig computergenerierte und damit (semi-)automatisierte Verschlagwortungen von historischen visuellen Medien nur in sehr eingeschränktem Maß sinnvolle Ergebnisse liefern (Nicka 2017). Deshalb werden diese Daten – abgesehen vom Bildthema – in vielen Projekten oft gar nicht erhoben oder nur sehr selektiv ein paar Schlagwörter zum Dargestellten verzeichnet. Ist Letzteres der Fall, sind es tendenziell häufiger Subjekte als Objekte² und öfter große oder im Vordergrund sichtbare Bildelemente als kleine oder solche im Hintergrund, die erfasst werden. Das hat zur Folge, dass in vielen Bilddatenbanken außer der über den Titel recherchierbaren Ikonographie nicht (oder nur sehr eingeschränkt) nach dem abgefragt werden kann, was auf den Bildquellen eigentlich zu sehen ist.

In REALonline wird aufgrund des Forschungsinteresses des Instituts für Realienkunde angestrebt, alle semantischen Bildelemente systematisch aufzunehmen und zwar ungeachtet ihrer Größe, ihrer Position, ihrer Relevanz für das Bildthema etc.³

Ein weiteres Spezifikum von REALonline ist zudem, dass die Daten der Bildbeschreibungen nicht als eine Aneinanderreihung von Begriffen gespeichert werden, in der ‚blau‘ gleichwertig neben ‚Apostel‘ oder ‚Holz‘ steht. Hier wird nach Informationstypen (Objekt, Person, Kleidung, Körperteil, etc.) und nach deren untergeordneten Informationsklassen (bei Objekten etwa Bezeichnung, Farbe, Material und Form) differenziert. Diese Kategorien können in weiterer Folge für komplexe Abfragen und Visualisierungen genutzt werden.

Es ist als Alleinstellungsmerkmal der Bilddatenbank REALonline zu werten, dass darüber hinaus auch verschiedene Relationen zwischen einzelnen Informationstypen erhoben und damit in weiterer Folge ebenfalls gezielt auffindbar gemacht werden können. Man kann also etwa dezidiert nach Darstellungen von Katzen suchen, die auf einem Tisch sitzen.

Insgesamt ergibt sich aus den am Forschungsinstitut in Krems erhobenen Daten zu mittelalterlichen und frühneuzeitlichen Bildquellen nicht nur ein reicher Datenschatz für Untersuchungen zur dargestellten materiellen Kultur und deren Kontexte. REALonline bietet darüber hinaus für die Öffentlichkeit wie für unterschiedliche Forschungsrichtungen einen großen Datenpool, auf den bei der Suche und bei dem Vergleich von Bildquellen zugegriffen werden kann. Anhand von Vermittlungsprogrammen in Museen oder Schulen, aber auch im Rahmen von interdisziplinären Forschungen, die historische Bildquellen in den Blick nehmen, hat sich immer wieder gezeigt, wie wichtig es ist, das Dargestellte zu erschließen und damit abfragbar zu machen.

²Eine Ausnahme stellt hierbei ein Projekt am Boijmans Van Beuningen-Museum dar, in dem ausschließlich dargestellte Dinge verschlagwortet werden, um sie besser mit erhaltenen historischen Objekten vergleichen zu können <https://www.boijmans.nl/en/collection/research/almal>, 11.10.2017.

³Erfasst werden auch Merkmale, wie etwa Farbe oder Form eines Bildbestandteils, die mitunter auch für stilistische Auswertungen neben den gewählten Motiven von Relevanz sein können.

3.2 Das bestehende System: KLEIO – ein hierarchisches DBMS

REALonline entstand in den 1970er Jahren auf der Basis der Konzepte der historischen Fachinformatik, die maßgeblich von Manfred Thaller am Max-Planck-Institut für Geschichte in Göttingen entwickelt wurden. Dort wurde mit dem Datenbankprogramm KLEIO auch die Software (kostenfrei) bereitgestellt, die es erlaubt, heterogene, komplexe Daten flexibel abzubilden, zu verwalten und auszuwerten (Thaller 1989). Zahlreiche Projekte aus dem Umfeld der heute den *Digital Humanities* zuzurechnenden Forschungsrichtungen haben ihr Datenmaterial mithilfe des Informationssystems KLEIO verwaltet und wurden von Göttingen aus maßgeblich bei der Entwicklung von geeigneten Datenlösungen unterstützt (Thaller 1980). Da die Software nicht mehr weiterentwickelt wird, musste im Zuge des Relaunchs für REALonline eine Nachfolgelösung gesucht werden. Ein wichtiges Kriterium dafür war die verlustfreie Migration des Altdatenbestandes in das neue Repository. KLEIO verarbeitet Daten in einer XML-konformen hierarchischen Struktur: Informationen werden in Dokumenten abgelegt. Ein Dokument besteht aus einer beliebigen Zahl von (Informations-)Gruppen, die zueinander in einer ganz bestimmten hierarchischen Abhängigkeit stehen. Jede Informationsgruppe kann eine beliebige Zahl von zugehörigen Elementen enthalten, die ihrerseits wiederum in Aspekte unterteilt sein können.

Umgelegt auf die Datenstruktur in REALonline: Für jedes zu beschreibende Werk wird ein Dokument angelegt, bestehend aus Werksdaten und Bildbeschreibungsdaten. Die Werksdaten (Künstler, Standort, Datierung, Herstellungstechnik usw.) umfassen im Wesentlichen die wichtigsten Informationen wie sie in der Kulturerbe- und Museumsdokumentation verwendet werden und können relativ einfach modelliert werden, weshalb dieser Teil hier keiner weiteren Ausführung bedarf. Für die Bildbeschreibung wurde zu Projektbeginn in den 1970ern folgendes Datenmodell entwickelt:

Jedes Bild enthält eine oder mehrere Bildszene mit einer zugehörigen Handlung und einem Ort. Innerhalb jeder Bildszene erfassen wir in REALonline das Dargestellte in Informationsgruppen nach den dargestellten Subjekten und Objekten. Für Personen werden neben dem Subjektnamen die Kategorien Geschlecht, Beruf bzw. Stand, Alter und Gestik/bzw. Haltung erfasst. Bei Objekten, Kleidung und den jeweiligen Teilen werden der Objektname und die Informationen zu Farbe, Material und Form erhoben. Es wird dabei die hierarchische Struktur der Beschreibungsdaten zum am Bild Dargestellten festgehalten und kann damit abfragbar gemacht werden: Direkte Subjekt-Objekt-Beziehungen werden erfasst, um von Figuren gehaltene Objekte zu dokumentieren. Objekt-Objekt-Beziehungen bilden einen Bezug zwischen einzelnen dargestellten Dingen ab. Darüber hinaus können sowohl Körperteile als auch Teile von Objekten abgespeichert werden (Jaritz 1993, 23–43). Zusätzlich verwaltet die Datenbank im Hintergrund Lexika zu Kleidungstermini, Künstlern und Personen sowie zwei große maschinenlesbar gemachte Textbestände zum Alten und Neuen Testament und den Heiligeniten in der Legenda Aurea, sowie Thesauri zu den Bildthemen, den materiellen Objekten und einer Reihe von weiteren Datenfeldern (Vavra 2004, 153).

Welche Beschreibungsdaten zu jedem Bild erfasst werden, kann an einem einfachen, nur aus einer Bildszene bestehenden Bildbeispiel aus einer Handschrift des 15. Jahrhunderts gezeigt werden, die sich in der Österreichischen Nationalbibliothek in Wien befindet: Die kolorierte Federzeichnung stammt aus dem Werk „Die Blumen der Tugend“, einem Tugend- und Lasterspiegel von Hans Vintler vom Beginn des 15. Jahrhunderts (Österreichische Nationalbibliothek, Wien, Codex 13567).⁴ (Abb. 2)

⁴<https://realonline.imareal.sbg.ac.at/detail/?archivnr=007117>, 11.10.2017.

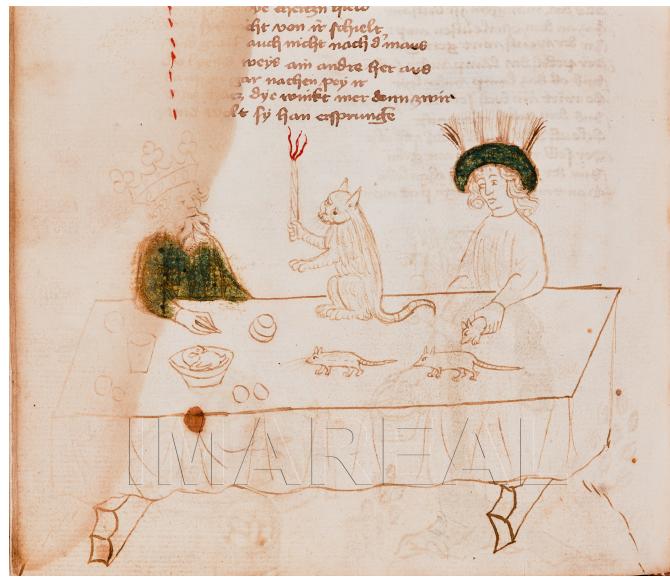


Abbildung 2: Hans Vintler, Blumen der Tugend. Versuchung einer Katze durch Mäuse. Österreichische Nationalbibliothek, Codex 13567, fol. 124v, © Institut für Realienkunde – Univ. Salzburg.

Die dargestellte Szene nimmt Bezug auf eine Stelle im Text, in der die Frage abgehandelt wird, was stärker sei, die Natur oder die Erziehung: König Salomon hat eine Katze so erzogen, dass sie bei Tisch die Kerze hält. Zur Evaluierung der vorgenannten Frage, hat der bei Tisch sitzende Gast drei Mäuse im Ärmel versteckt und lässt eine nach der anderen aus. Anfänglich gehorcht die Katze dem König; bei der dritten Maus schlägt ihre Natur durch, sie lässt die brennende Kerze fallen und versucht die Maus zu fangen (Vonach 2015, 396).

Auf dieser einfachen Zeichnung sind folgende Personen und Objekte dargestellt: König Salomon, ein zweiter Mann bei Tisch sitzend; am Tisch liegend ein Tischtuch, darauf die Katze eine Kerze haltend, Mäuse, ein Trinkbecher, eine Schüssel mit Inhalt (Geflügel), Eier und Brot (Abb. 3).

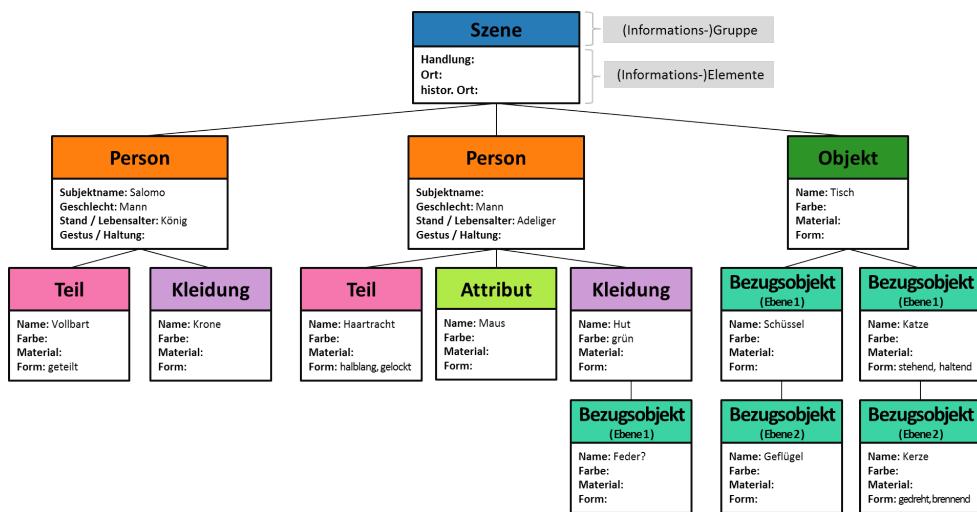


Abbildung 3: Schematische Darstellung ausgewählter Informationen zu Datensatz 007117 im Altsystem.

4 Graphentechnologien in REALonline

Grundsätzlich ergab sich der Einsatz von Graphen als Modellierungskonzept auf natürliche Weise: die Ausgangsdaten in KLEIO wiesen eine hierarchische Struktur auf, die sich gut in einen Graphen überführen ließ, Bildbeschreibungen lassen sich auf intuitive Weise als semantische Graphen abbilden, auch mehrstufige Thesauri bzw. Ontologien werden oft als Graphen repräsentiert und Graphmodelle werden nun auch vermehrt in den digitalen Geisteswissenschaften eingesetzt (Kuczera 2017; Kaufmann und Andrews 2016). Die Graphmodellierung bildet einen integrativen Teil des Gesamtkonzepts von REALonline, auf das im Folgenden näher eingegangen werden soll.

4.1 Das konzeptionelle Datenmodell von REALonline

Für die konkrete Implementierung stellt sich in Folge die Frage, wie diese Daten technisch gespeichert, bzw. verarbeitet und visualisiert werden sollten.

Eine Möglichkeit wäre der Einsatz einer Graphdatenbank wie Neo4j oder auch RDF bzw. ein Triplestore gewesen. Dem steht allerdings eine Reihe von „praktischen“ Argumenten entgegen:

- Nicht alle Anwendungsfälle lassen sich gleich gut mit der Graphmetapher abdecken: beispielsweise erfolgt die Eingabe der Werksdaten im allgemeinen „formularbasiert“, d. h. in Form vorgegebener Eingabefelder, die jeweils bestimmten Kriterien zu genügen haben.
- Bei der administrativen Verwaltung der Daten erfolgt hier auch die Abfrage bzw. Suche eher in der traditionellen Form von Listen und Filtern.
- Die Beschreibungen müssen einer vorgegebenen Struktur folgen, um einerseits Eingabefehler zu vermeiden und andererseits eine Abfrage überhaupt zu ermöglichen. Graphdatenbanken stellen eine sehr allgemeine Struktur zur Verfügung und bieten nur eingeschränkte Möglichkeiten zur Schematisierung bzw. Validierung.
- Werke, Beschreibungen und Thesauri werden in REALonline in unterschiedlichen Arbeitsschritten erfasst bzw. verwaltet: die Werksdaten werden teilweise bereits bei der Aufnahme des Bildes erhoben und gegebenenfalls aus bestehenden Quellen ergänzt, die Beschreibungen werden separat, und oft erst zu einem späteren Zeitpunkt, neu verfasst. Thesauri werden getrennt davon definiert und auch immer wieder geändert.
- Graphdatenbanken bilden im Allgemeinen den gesamten Datenbestand als einen einzigen, uniformen Graphen ab. Die Unterscheidung der einzelnen Teile der Applikation muss also jedenfalls in der Programmlogik erfolgen.

Im Zuge der Datenmodellierung wurde daher im ersten Schritt ein *konzeptionelles Datenmodell* definiert, das die Anforderungen von REALonline abdeckte. Die Umsetzung dieses Datenmodells in einer konkreten Datenbanksoftware erfolgte dann erst in einem zweiten Schritt auf Basis von technischen Überlegungen.

Bei der Datenmodellierung in REALonline kann man drei Ebenen unterscheiden (Abb. 4):

- Die *Applikationsebene* beschäftigt sich mit den konkreten Daten, das heißt mit ‚Werken‘, ‚Bildbeschreibungen‘, ‚Thesauri‘ u. a.
- Das *konzeptionelle Datenmodell* stellt allgemeine ‚Daten-Dokumente‘ sowie darauf aufbauend ‚Graph-Dokumente‘ zu Verfügung. Für diese beiden Dokumenttypen werden außerdem ‚Workflows‘ definiert (Dokument ist ‚gespeichert‘ oder ‚in Bearbeitung‘, Bearbeitungshistorie).
- Die *technische Repräsentation* beschreibt die konkrete Speicherung des konzeptuellen Modells in einer bestehenden Datenbanktechnologie.

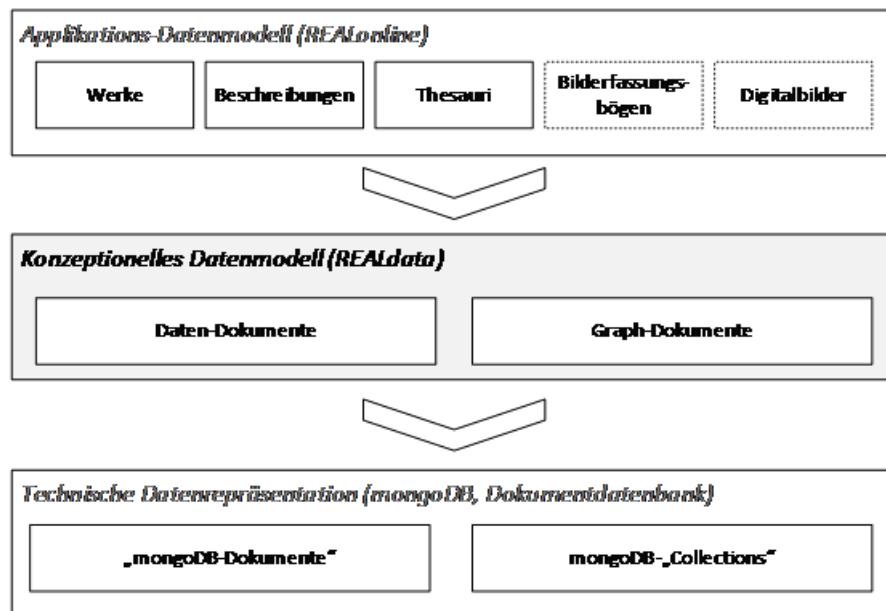


Abbildung 4: Die Datenmodellierung in REALonline.

Die Ebenen greifen zur Beschreibung ihrer Daten dabei jeweils auf Elemente der darunterliegenden Ebene zurück: eine ‚Beschreibung‘ aus dem Applikationsmodell ist definiert als ein ‚Graph-Dokument‘ aus dem konzeptionellen Modell, auf das ein bestimmtes Schema angewendet wird. Ein ‚Graph-Dokument‘ aus dem konzeptionellen Modell wird wiederum auf eine genau definierte Weise in einer dokumentenorientierten Datenbank (mongoDB, siehe Abschnitt 4.1.4) repräsentiert.

4.1.1 Daten- und Graph-Dokumente im konzeptionellen Datenmodell

Das konzeptionelle Datenmodell von REALonline stellt zwei Arten von *Dokumenten* zur Verfügung, die hier zur Unterscheidung als Daten- und Graph-Dokumente bezeichnet werden (Abb. 5).

In beiden Dokumentarten werden eine Reihe von Verwaltungsinformationen gespeichert: jedes Dokument enthält

- einen eindeutigen Bezeichner (ID),
- das Datum der Erstellung bzw. der letzten Bearbeitung,

- den Benutzer*innennamen der Bearbeiterin oder des Bearbeiters,
- den aktuellen Bearbeitungsstatus, die Veröffentlichungsinformationen etc. und
- eine vollständige Bearbeitungshistorie (mit der Möglichkeit, Änderungen nachzuverfolgen bzw. rückgängig zu machen).

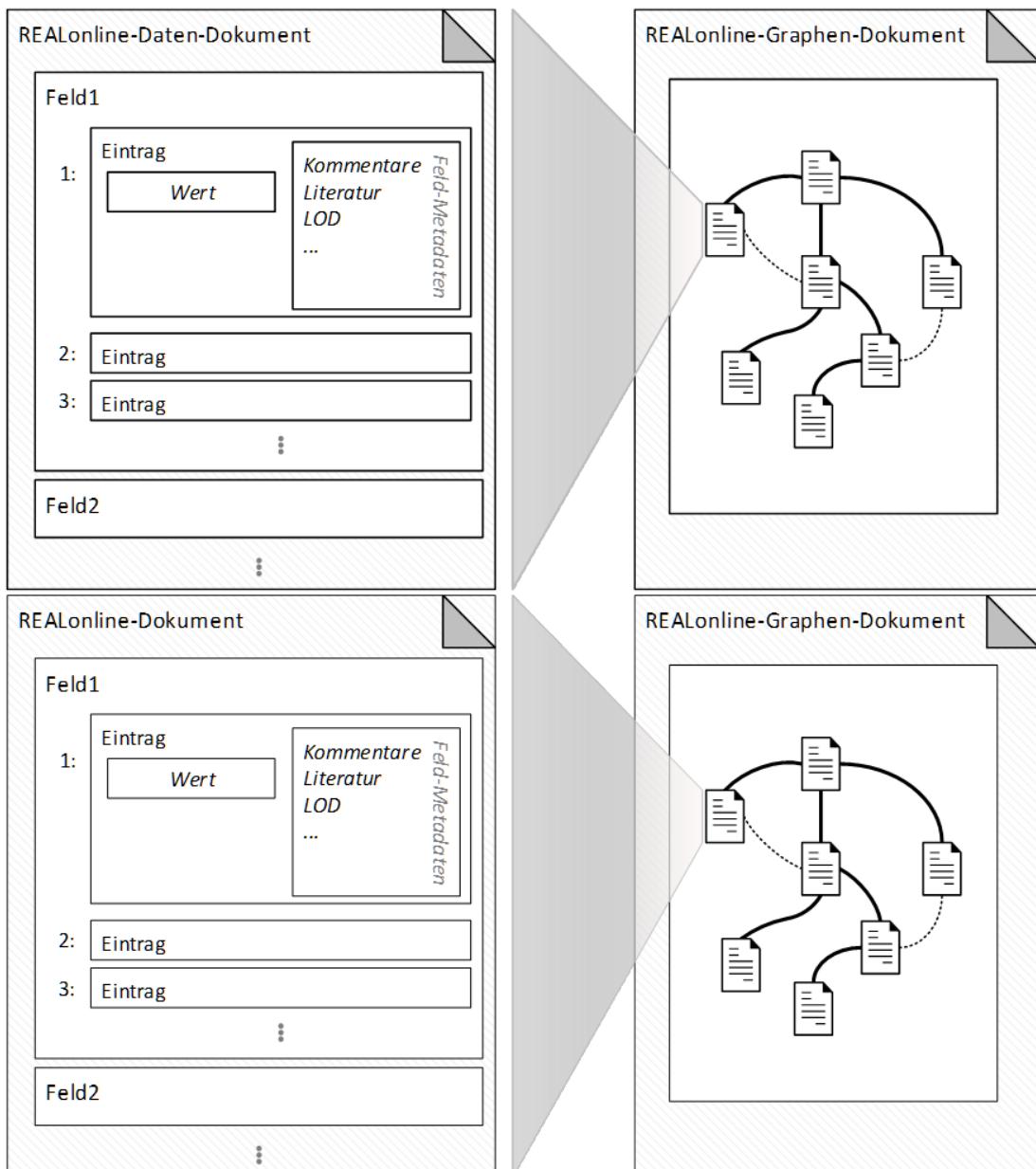


Abbildung 5: Daten- und Graph-Dokumente in REALonline.

Ein *Daten-Dokument* im konzeptionellen Datenmodell von REALonline besteht aus einer Menge von *Feldern*, von denen jedes durch einen eindeutigen Namen bezeichnet wird. Jedes Feld kann eine beliebige Anzahl von *Einträgen* haben, jeder Eintrag besteht aus einem *Wert* (dem „eigentlichen“ Inhalt)

sowie einer Reihe von zusätzlichen Metadaten – wie z. B. Kommentaren, Literaturangaben, Verlinkungen zu Thesaurusbegriffen, Links zu Normdatensätzen und Klassifikationssystemen (ICONCLASS, GND und andere).

Ein *Graph-Dokument* beschreibt einen in sich abgeschlossenen Graphen aus *Entitäten* und zugehörigen *Relationen*. Entitäten werden über einen Bezeichner (ID) referenziert, der im jeweiligen Graph-Dokument eindeutig ist, und erhalten einen frei definierbaren *Entitätstyp*.

Weiters stellt jede Entität eines Graph-Dokuments in sich wieder ein Daten-Dokument mit beliebigen Feldern/Einträgen/Werten bzw. Metadaten dar.

Relationen werden grundsätzlich gerichtet definiert (*von A nach B*) und erhalten als Zusatzinformation einen *Relationstyp*.

Graph-Dokumente unterscheiden sich in folgenden Punkten von Graphen in Graphdatenbanken oder auch RDF/Triplestores:

- Graph-Dokumente stellen in sich abgeschlossene Graphen dar, die unabhängig voneinander angelegt und bearbeitet bzw. auch ex- und importiert werden können. Dadurch ist es z. B. leicht möglich festzulegen, welche Bildbeschreibungen in den Suchindex für die Webseite aufgenommen werden sollen. Im Gegensatz dazu speichern Graphdatenbanken sämtliche Informationen in einem einzigen Graphen.
- Die einzelnen Knoten (Entitäten) der Graph-Dokumente können komplexere Informationen enthalten, als das bei üblichen Graphdatenbanken vorgesehen ist. Insbesondere die Möglichkeit, Mehrfacheinträge mit jeweils zugeordneten Zusatz(Meta-)informationen vorzunehmen, war bei der Modellierung von Bildbeschreibungen ausgesprochen hilfreich.

4.1.2 Schemadefinition

Das konzeptionelle Datenmodell macht keine weiteren Vorgaben zum Inhalt von Daten- oder Graph-Dokumenten. Es wird also nicht definiert, welche Felder bzw. Werte erlaubt sind, oder welche Typen von Entitäten oder Relationen in einem Graphen vorkommen dürfen.

Diese Festlegung wird im Rahmen einer *Schemadefinition* getroffen, wobei – analog zur Definition von Daten- und Graph-Dokumenten – von *Feld-* und *Graph-Schemadefinition* gesprochen wird.

Das *Feldschema* legt fest,

- welche Felder in einem Dokument vorkommen dürfen,
- welche davon verpflichtend und welche optional sind,
- wie viele Einträge in jedem Feld erlaubt sind (Einfach- vs. Mehrfachfelder),
- welche Zusatzinformationen (Feldmetadaten) bei den Einträgen erfasst werden können und
- welche Werte grundsätzlich erlaubt sind (*Datentyp*, z. B. beliebige Zeichenketten, Zahlen, Datierungen, Maßangaben etc.).

Im *Graphschema* wird festgelegt,

- welche Entitätstypen im Graph-Dokument vorkommen können,
- welches Feldschema bei jedem Entitätstyp zur Anwendung kommt und
- welche Relationstypen auftreten und zwischen welchen Entitätstypen die jeweiligen Relationen definiert werden können.

Die Schemadefinitionen sind in der Software als JSON-Dokumente hinterlegt und können jederzeit angepasst bzw. geändert werden.

4.1.3 Definition des Applikations-Datenmodells

Schemadefinitionen werden in der Folge verwendet, um das Applikations-Datenmodell zu definieren: so werden Werke als Daten-Dokumente implementiert, für die ein entsprechendes Feldschema definiert wird. Bildbeschreibungen werden als Graph-Dokumente abgebildet, deren Schema die Elemente (Entitäten) für Bildbeschreibungen in REALonline festlegt.

Die Definition der Schemata erlaubt es, die entsprechenden Eingabemasken bzw. den Grapheditor größtenteils generisch zu implementieren: soll beispielsweise ein neuer Entitätstyp für die Bildbeschreibungen eingeführt werden (wenn man z. B. „Tier“ als eigene Entität im Graphen darstellen wollte) oder wenn neue Felder in der Werksbeschreibung erfasst werden sollen, reicht eine Ergänzung des entsprechenden Schemas, und die Benutzeroberfläche wird automatisch angepasst.

4.1.4 Speicherung in mongoDB

Nach der Definition des konzeptionellen und des Applikationsdatenmodells bleibt noch die Frage nach der physischen Datenablage offen: es muss also definiert werden, in welcher Form bzw. mit welchen Technologien die Daten im konzeptionellen Datenmodell konkret gespeichert werden. Dazu ist grundsätzlich festzuhalten, dass die Technologie prinzipiell frei gewählt werden kann: Datendokumente und Graphen können sowohl in relationalen (SQL-)Datenbanken, als auch in Key-Value-Stores, dokumentenorientierten Datenbanken oder Graphdatenbanken gespeichert werden.

Für REALonline fiel die Wahl schlussendlich auf die dokumentenorientierte Datenbank mongoDB. Die Gründe dafür waren:

- Die Datenbank ist unter einer Open-Source-Lizenz frei verfügbar.
- mongoDB bietet eine ausgezeichnete Performance.
- mongoDB erlaubt einen einfachen Export und Import im JSON-Format. Das machte einerseits Backup-Aufgaben bzw. das Erzeugen von Testdaten einfach, andererseits wurde dadurch auch der Datentransfer an das Web-Frontend sehr vereinfacht.

Durch die Programmarchitektur ist die Wahl allerdings nicht zwangsläufig endgültig. Es ist mit moderatem Aufwand möglich, mongoDB durch eine andere Datenbank zu ersetzen, sollte das in Zukunft sinnvoll oder erforderlich sein.

4.2 Die Transformation des hierarchischen Systems in einen Graphen

Das Datenmodell der Bildbeschreibungen in REALonline, wie es für den Gebrauch in KLEIO entworfen wurde, hatte aufgrund der bereits genannten Charakteristika von Bildern und der Art, wie Menschen deren semantische Inhalte erfassen, bereits viele Elemente, die eine Transformation in ein graph-orientiertes Datenbankschema sinnvoll gemacht haben: So wurden etwa die Beziehungen zwischen Subjekten und Kleidungen, zwischen Subjekten und Attributen sowie zwischen Objekten und ihren Bezugsobjekten bereits über die Kategorien im hierarchischen Datenbanksystem als solche modelliert.

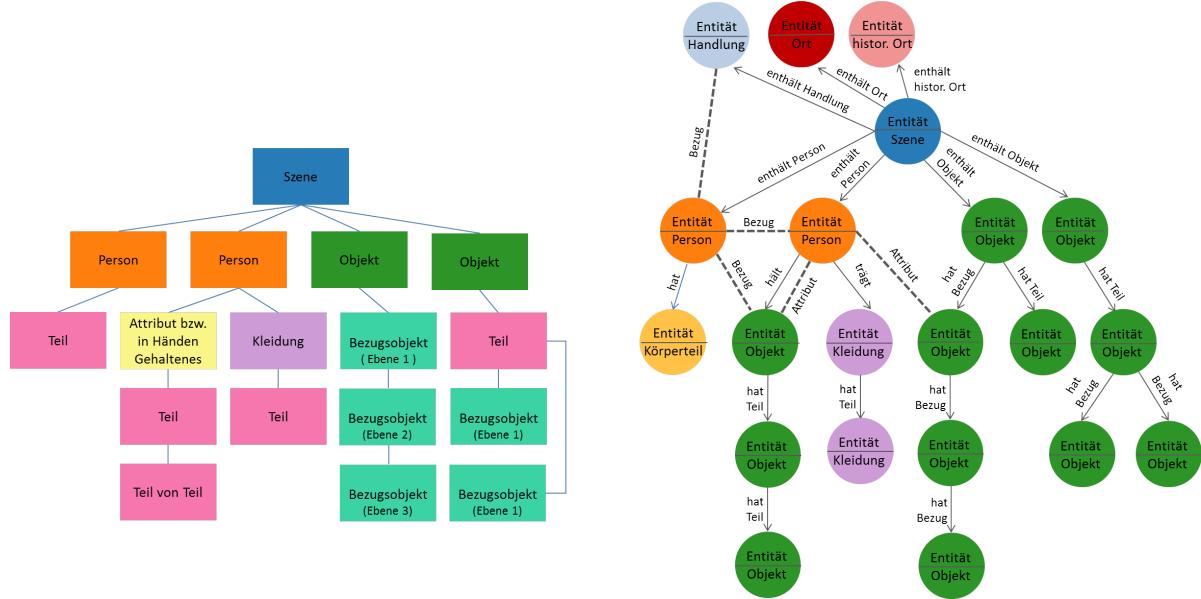


Abbildung 6: Die Transformation der Daten: Der linke Teil zeigt einen schematisch dargestellten Datensatz, wie er nach dem System für REALonline in KLEIO erfasst wird. Rechts ist ein Beispiel für einen Datensatz im neuen graph-orientierten Erfassungsschema in REALonline zu sehen.

Diese Bezüge wurden in KLEIO auf Ebene der Informationstypen und nicht auf der Ebene der Relationen zwischen den einzelnen erhobenen Bildelementen festgehalten. Dieser Umstand ist anhand von Abbildung 6 auf einen Blick daran zu erkennen, dass in der schematischen Darstellung des KLEIO-Datensatzes den Verbindungen selbst keine Information zugeordnet ist.

Um beispielsweise eine konkrete Beziehung zwischen Personen und Objekten herzustellen, wurde im KLEIO-Schema der Informationstyp „In Händen Gehaltenes“ festgelegt. Nachdem die Attribute der Heiligen sehr oft in Händen gehalten dargestellt wurden, konnte dieser Informationstyp bei Bedarf auch dafür verwendet werden.⁵

Im neuen graph-orientierten Schema der Beschreibungen wird die Information ‚hält‘ als Verbindung modelliert. Damit entspricht der Teilgraph einem einfachen Satz aus Subjekt, Prädikat und Objekt: „Eine Person hält ein Objekt“. Diese Nähe zu unseren sprachlichen Gewohnheiten erleichtert in weiterer Folge die Abfragen der Datenbank.

⁵Die Differenzierung erfolgte in diesem Fall über den zusätzlichen Wert ‚H‘ (für Heiligenattribut) in der Bezeichnung des Elements.

Mit der Transferierung der Information ‚Heiligenattribut‘ in die Verbindung können zudem auch Bildelemente, die nicht in der Hand gehalten werden, als Heiligenattribute definiert werden. Die Verbindungen ‚Heiligenattribut‘ und ‚Bezug‘ – in Abbildung 6 strichiert dargestellt – definieren einen alternativen Verbindungstyp zwischen einzelnen Elementen. Mit der Verbindung ‚Bezug‘ ist es im neuen Schema auch möglich Beziehungen herzustellen, die bis dato aufgrund des hierarchischen Systems (abgesehen von Verweisen in Kommentaren) nicht erhoben werden konnten. Damit können nun auch folgende Bezüge ausgedrückt werden:

- jene zwischen Personen, also etwa wenn Maria das Jesuskind hält, oder
- wenn über ein Objekt eine Beziehung zu einer anderen Person hergestellt wird, wenn beispielsweise jemand einen Brief an eine andere Person überreicht.

Ein Teil der Überarbeitung des Datenmodells der Beschreibungen hat daher auch darin bestanden, die im KLEIO-Schema dem Informationstyp ‚Szene‘ untergeordneten Informationsklassen ‚Handlung‘, ‚historischer Ort‘ und ‚Ort‘ zu eigenen Informationstypen umzuwandeln, damit in Zukunft gewährleistet ist, dass z. B. auch Bezüge zwischen einzelnen Personen oder Objekten und einer Handlung hergestellt werden können.

Was die Informationstypen ‚Teil‘ und ‚Bezugsobjekt‘ des KLEIO-Schemas angeht, konnte bei der Überführung in das graph-orientierte Schema in mehrfacher Hinsicht eine nötige Weiterentwicklung geleistet werden. Hier wurde ebenfalls die Information ‚hat Teil‘ oder ‚hat Bezug‘ in die Verbindung der Knoten gelegt. Im KLEIO-Datenschema war die Anzahl bei Teilen auf zwei und bei Bezugsobjekten auf drei hierarchische Ebenen beschränkt. Diese Vorgaben wurden in der neuen Version von REALonline aufgehoben. Die Modellierung der Beschreibung als Graph und die Speicherung der Information ‚hat Teil‘ oder ‚hat Bezug‘ als Kante statt als eigenen Informationstypen ermöglicht nun dort, wo es angebracht ist (wenn etwa ein Tisch ein Tischbein als Teil und eine Vase als Bezugselement hat), Teile und Bezugselemente auf der gleichen Ebene anzulegen.

Die Daten zu Bezügen zwischen den einzelnen Elementen als eine Art der Verbindung zu speichern und nicht mehr als eigenen Informationstypen, hat das Modell der Beschreibungen einheitlicher gemacht. So wird z. B. ein Pilgerzeichen nun immer unter der Kategorie ‚Objekt‘ geführt, egal, ob es in der Hand gehalten wird, ob es an einem Kleidungsstück angebracht ist oder auf dem Boden liegt, während im alten Schema dafür die Informationstypen Attribut, Bezugsobjekt oder Teil verwendet worden sind. Dies ist auch bei der Suche hilfreich, wenn man etwa alle Objekte in REALonline abfragen möchte, die auf Bildquellen von 1330 bis 1350 dargestellt sind. Dazu muss nicht im Zuge der Bildung der Query berücksichtigt werden, dass es sich neben dem Informationstyp ‚Objekt‘ auch um ‚Attribut‘, ‚Bezugsobjekt‘ oder ‚Teil‘ handeln könnte.⁶

Umgekehrt war es in den Fällen, wo Teile von Personen erhoben worden sind, sinnvoll diese Daten nicht unter dem Informationstyp ‚Person‘ zu speichern – weil bei Personen Daten zu anderen Informationsklassen (Subjektname, Geschlecht, Stand, Gestus/Haltung) aufgenommen werden –, sondern

⁶In KLEIO war diesbezüglich von Vorteil, dass außer bei Szene und bei Person dieselben Informationselemente (Name, Farbe, Material, Form) verwendet wurden und damit auch für die genannte Query nach allen Datensätzen, die zwischen 1330 und 1350 datieren und weiters nach dem Informationselement ‚Name‘ eingeschränkt werden konnte. In dieser Aufstellung waren dann allerdings auch alle erfassten Körperteile und Kleidungsstücke enthalten.

dafür den neuen Informationstyp ‚Körperteil‘ zu schaffen. Grundsätzlich wird in der neuen Version des Beschreibungsschemas zu allen Knoten auch der „allgemeine“ Informationstyp ‚Entität‘ gespeichert, wodurch informationstypunabhängige Abfragen ermöglicht werden.

Insgesamt konnten bei der Überführung der Daten in das neue Schema nur die bisher im Rahmen der Informationstypen definierten Verbindungen generiert werden. Eine Ausweitung der Arten von Verbindungen ist aber prinzipiell möglich, sodass etwa in Zukunft Zusammenhänge wie „eine Person reitet auf einem Pferd“ oder „eine Person sitzt auf einem Schemel“ durch neue Werte in den Verbindungen wie ‚reitet auf‘ oder ‚sitzt auf‘ modelliert werden können.

Einen entscheidenden Vorteil der graph-orientierten Speicherung der Beschreibungsdaten (etwa gegenüber einer relationalen Datenbank) sehen wir in der potentiellen Erweiterbarkeit des Schemas: Zum Zeitpunkt der Erstellung der Datenbank muss damit noch nicht Anzahl und Art aller möglichen Knoten und Verbindungen feststehen, sondern das System kann hier mit einem vertretbaren Aufwand angepasst werden, wenn etwa eine später hinzukommende Bildquellengattung mit abweichenden Charakteristika oder innovative Forschungsansätze dies notwendig machen.

4.3 Graphentechnologien im Backend-Editor von REALonline

Aufgrund der vielen Daten, die in REALonline zum Bild Dargestellten eingegeben werden, und dem damit einhergehenden großen Zeitaufwand, war es eine Anforderung an das neue System, die Eingabe so effizient wie möglich zu gestalten. Bisher wurden die Informationen (sowohl für die Werks- als auch die Beschreibungsdaten) als Text gespeichert, der in einem Texteditor auf der Basis des Datenbankschemas geschrieben wurde (Abb. 7).

```
N$007117//7008943/3011083/7008943
G$Federzeichnung koloriert/Illustrationszyklus Dichtung/Illustration//Hans Vintler. Blumen der Tugend #
Versuchung einer Katze durch Mäuse/Tirol/Vintler:06755-06783
D$1411/1411/Wien/Österreich/Wien/Österreichische Nationalbibliothek/cod. 13567/fol. 124v
C$/Hans Vintler. Blumen der Tugend
B$
S$M/Salomo/König
T$Haartracht//halblang;gelockt
T$Vollbart//geteilt
K$Krone
K$Obergewand/grün
S$M//Adeliger
T$Haartracht//halblang;gelockt
A$Maus
K$Hut/grün
X$Feder?
K$Obergewand
O$Tisch
X$Tischdecke
X$Katze//stehend;haltend # Kerze
Y$Kerze//gedreht;brennend
X$Brot?//rund;klein
X$Ei?
X$Schüssel
Y$Geflügel
X$Maus
```

Abbildung 7: Datensatz 007117, wie er als Textdatei für die Weiterverarbeitung in KLEIO gespeichert wurde.

Diese Textdateien wurden in weiterer Folge in KLEIO verarbeitet. Dabei handelt es sich um eine Form der Eingabe, bei der – wenn die Bearbeiter*innen das Datenbankschema bereits gut kennen – relativ schnell viele Daten erhoben werden können. Das Fehlen einer auf REALonline abgestimmten grafischen Benutzungsoberfläche (GUI) wurde als Desiderat empfunden, weil unter anderem in den mitunter sehr viele Textzeilen enthaltenden Datensätzen der unmittelbare anschauliche Überblick verlorenging, kein direkter Bezug zwischen Eingabe und den in anderen Listen vorgehaltenen Thesauri bestand und Überarbeitungen der Daten nur von gut eingearbeiteten Mitarbeiter*innen und mit einem sehr hohen Arbeitsaufwand durchgeführt werden konnten.

Im neuen System wurde für die Erfassung der Werksdaten eine „klassische“ Eingabemaske gewählt. Über ein Kontextmenü können u. a. weitere Metadaten (wie Kommentare oder Verknüpfungen zu Normdatenrepositorien) hinzugefügt oder die Verlinkungen mit dem Thesaurus bearbeitet werden (Abb. 8).

Abbildung 8: Die Erfassung der Werksdaten im neuen System für Datensatz 007117.

Das GUI für die Beschreibungsdaten im Backend des neuen Systems bietet unterschiedliche Möglichkeiten der Eingabe und integriert verschiedene Visualisierungen, um die Zusammenhänge der vielen im Eingabeprozess erhobenen Daten adäquat zu veranschaulichen und gleichzeitig zu gewährleisten, dass ein besonders effizientes Erfassen der Daten erlaubt wird.

Die Daten können im Rahmen von Befehlen eingegeben werden, die über eine integrierte Kommandozeile abgesetzt werden (Abb. 9). Diese Befehle stellen eine umfassende Weiterentwicklung der Vorgangsweise dar, wie die Textdateien für KLEIO geschrieben wurden. Einerseits wurden die Eingabekontrollen überarbeitet und optimiert und andererseits werden nun nach der Ausführung eines Befehls unterschiedliche Prüfungen durchgeführt, sodass insgesamt einerseits die Eingabegeschwindigkeit erhalten und gleichzeitig die zu erwartende Konsistenz und Richtigkeit der Daten verbessert werden konnten.⁷

The screenshot shows the REALonline application interface. At the top, there is a navigation bar with links for Werke, Sammlungen, Suche, Thesauri, and user information (Angemeldet als Isabella, Abmelden). Below the navigation is a title 'Beschreibung für Verkündigung an Maria' with a status message: 'Verkündigung an Maria (000168) Aktualisiert am: 16.8.2017 18:09 Status: Bestätigt von Isabella Öffentlich'. To the right are buttons for Kopieren, Löschen, and Speichern.

The main area contains a tree view of objects on the left and a detailed editing mask on the right. The tree view lists various object types such as Streifenmuster Objekt, Lesepult Objekt, Buch Objekt, Bucheinband Objekt, Buchseite Objekt, Gewölbe Objekt, Gewölberippe Objekt, Gewölbekappe Objekt, Wand Objekt, Steinquader Objekt, Rundbogenöffnung Objekt, Dienst Objekt, Schlussstein Objekt, Fliesenboden Objekt, Fliese Objekt, Balkendecke Objekt, Querbalken Objekt, Wand Objekt, Rundbogenöffnung Objekt, Kreuzstockfenster Objekt, Tür Objekt, Beschlag Objekt, Türschloss Objekt, Schatten Objekt, and Innenraum Ort. The 'Buch Objekt' node is selected.

The editing mask for 'Buch (Objekt)' includes fields for Name (Buch), Farbe (dropdown), Material (dropdown), Form (geöffnet), and a 'Beziehungen' section showing relationships like 'hat Bezug' to 'Lesepult (E37)', 'hat Teil' to 'Bucheinband (E39)', and 'hat Teil' to 'Buchseite (E40)'. There are also thumbnail images of the book and its pages.

To the right of the mask is a context menu with options like Hinzufügen, Entfernen, Thesaurus, Kommentare, Literatur, LOD-Einträge, Bemerkenswert, and Unsicherheit. At the bottom, a command line shows the input 'o#tintenf_schwarz' followed by a warning: 'Kein passender Begriff für „tintenf“ im Thesaurus „Gegenstände“ gefunden, mögliche Begriffe sind: „Tintenfass“, „Tintenfassdeckel“, „Tintenfisch“, „Tintenfläschchen“'.

Abbildung 9: GUI von REALonline – Listenansicht. Kommandozeile (unten), Bearbeitungsmaske mit Kontextmenü (rechts), Übersicht zu vorhanden Tags und bestehenden Beziehungen der ausgewählten Entität zu anderen Entitäten (rechts unten).

⁷Der eingegebene Befehl wird nicht ausgeführt, wenn die korrekte Anzahl der möglichen Informationsklassen bei einem konkreten Informationstyp überschritten wurde oder wenn die Eingabe zu einer vom Schema nicht vorgesehenen Abfolge von Informationstypen führen würde. Weiters werden die erfassten Begriffe mit den bei der jeweiligen Informationsklasse hinterlegten Thesauri verglichen und, wenn notwendig, die Groß- bzw. Kleinschreibung korrigiert. Stimmt die eingegebene Zeichenfolge nicht überein, wird der Befehl ausgeführt, aber es erscheint ein Hinweis, dass dieser Begriff nicht im Thesaurus enthalten war (was wiederum impliziert, dass dieser auch nicht mit dem Thesaurus verlinkt werden konnte). Diesem Hinweis wird, wenn möglich, ein Vorschlag basierend auf Textähnlichkeit hinzugefügt, wie beispielsweise: „Kein passender Begriff für „kreuzripp“ im Thesaurus „Gegenstände“ gefunden, mögliche Begriffe sind: „Kreuzrippe“, „Kreuzrippengewölbe“.“ Weitere Erleichterungen betreffen unter anderem die Verwendung von Standardannahmen (sowohl hinsichtlich der Erstellung von Knoten als auch für die Festlegung der Verbindungen) sowie die einfache Ansprache von übergeordneten Elementen oder die Wiederholbarkeit von Befehlen bzw. Befehlsabfolgen.

Einer der wichtigsten Aspekte des neuen REALonline-GUI ist die Visualisierung der eingegebenen Daten. Einerseits kann hierzu eine Anzeige gewählt werden, die die Knoten und ihre Informationstypen als Liste ausgibt (Abb. 9). Die weiteren, zu einem Knoten gespeicherten Werte in den Informationsklassen werden in dem hellgrauen Bereich rechts wiedergegeben. Symbole neben den Einträgen zeigen an, wenn zusätzlich Kommentare, Verknüpfungen zu Normdaten⁸ oder bibliografische Quellen vorhanden sind; ein Frage- bzw. ein Ausrufezeichen markiert einen Wert als fraglich oder besonders hervorhebenswert. Unterhalb dieses Informationsblocks vermittelt eine Grafik, welche direkten Verbindungen zu anderen Knoten bestehen.

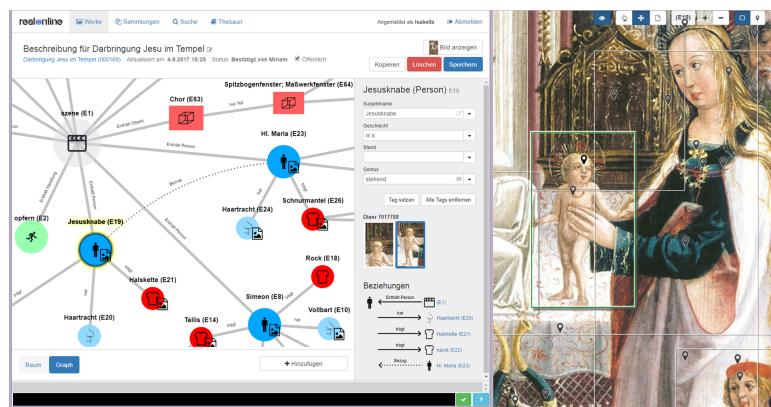


Abbildung 10: GUI von REALonline – Graph-Ansicht, Tagging-Interface (rechts).

Noch deutlicher führt die Visualisierung der Daten als Graph die Zusammenhänge zwischen den vorhandenen Informationen zu den Bildelementen vor Augen (Abb. 10). Die über die Kommandozeile erstellten Knoten werden gemäß ihrem Informationstyp in der Visualisierung in unterschiedlichen Farben und Formen (rund oder rechteckig) und mit unterschiedlichen Symbolen ausgezeichnet; die beiden Verbindungstypen werden ebenfalls durch unterschiedliche Strichformen angezeigt. Zusätzlich sind alle Knoten mit ihrer Entitätsnummer und einer Bezeichnung versehen und auch die Art der Verbindungen ist als Text ausgewiesen. Die Visualisierung als Graph bringt auch mit sich, dass die Verbindungen ‚Bezug‘ und ‚Heiligenattribut‘, die in der Listenansicht zwischen mitunter räumlich sehr weit voneinander entfernten Elementen bestehen, einfach hergestellt werden können. Dabei ist essentiell, dass die Knoten der Graphen im GUI von den Bearbeiter*innen nicht nur verschoben, sondern auch fixiert werden können, sodass deren Position gespeichert und damit auch bei späteren Aufrufen der Beschreibung im Backend wieder angezeigt wird. Die zoombare Graphvisualisierung ermöglicht es, sich eine Übersicht über die Zusammensetzung und die Struktur des Graphen zu machen oder spezielle Teilgraphen vergrößert zu betrachten. Damit hier möglichst viel Raum des Bildschirms für die Anzeige der Daten genutzt werden kann, kann die Datei mit dem zu beschreibenden Bild, in einem zweiten Fenster oder einem zweiten Monitor geöffnet werden.

Neben der Visualisierung der Daten kann die Graph-Ansicht aber auch als Eingabeeditor genutzt werden. Ein Rechtsklick auf einen Knoten oder auf eine Verbindung öffnet ein Kontextmenü, das angibt, welche

⁸Gegenwärtig wird vor allem auf ICONCLASS- und GND-Notationen verwiesen (vgl. <http://www.iconclass.nl/home>, 11.10.2017 und http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html, 11.10.2017) Eine Ausweitung dieser Bezüge, etwa durch das Anreichern der Daten mit AAT-Referenzen (<http://www.getty.edu/research/tools/vocabularies/aat/about.html>, 11.10.2017), ist geplant.

Knoten an dieser Stelle hinzugefügt oder entfernt werden können, bzw. welche Bezüge hergestellt oder geändert werden können.⁹ Diese Art der Eingabe ermöglicht es, dass geringfügige datensatzspezifische Erweiterungen oder Überarbeitungen der gespeicherten Informationen auch von Anwender*innen vorgenommen werden können, die grundsätzlich mit der Art, Daten als Befehle über die Kommandozeile einzugeben, oder mit dem Datenschema nicht ausreichend vertraut sind.

Eine Anforderung an die neue Version von REALonline war es, Koordinaten zu den erfassten Bildelementen speichern zu können, sodass hier ein deutlich nachvollziehbarer Bezug zwischen textlicher und bildlicher Information hergestellt werden kann.¹⁰ Dies kann entweder nachträglich geschehen, oder die Auszeichnung der Bildbereiche, zu denen Informationen erhoben werden sollen, kann auch am Beginn der Erfassung stehen: Im dafür vorgesehenen Modus werden durch das Taggen des Bildes die primären Knoten und Kanten des Graphen generiert.

4.4 Konzeptuelle und funktionale Anforderungen an das Frontend

Die Realisierung des Frontends für die neue Version von REALonline erfolgte im Rahmen einer eigenständigen Entwicklung. Wichtig erschien es uns hierbei unter anderem, die Vorteile der Graphansicht auch für die User im Internet zur Verfügung zu stellen (Matschinegg und Nicka 2018).

4.4.1 Überlegungen zur Visualisierung mit Graphen

Warum ein Graph? Diese Frage ist wohl eine der meist gestellten und kann aus dem Blickwinkel der Informationstechnologie allein nicht einfach beantwortet werden. Ab einer gewissen Anzahl von Knoten und Kanten (Verbindungen) wird jeder Graph unübersichtlich und schwer zu handhaben. Aber er bietet eben auch, wie in Abbildung 11 zu sehen ist, für ganz unterschiedlich dichte Werke eine einheitliche Möglichkeit schnell die Struktur (nicht die Daten!) des Werkes zu erkennen. Diese entspricht dem Gesetz der Einfachheit (Maeda 2006) genauso wie dem Ansatz, über spielerische Momente Daten bzw. deren Aussagen verständlicher zu machen¹¹. Diese beiden Konzepte wurden beim Entwurf des Frontends neben dem Einsatz klassischer Webdesign-Elemente zur verständlichen Vermittlung umfangreicher Daten herangezogen (Abb. 12). Letztlich wurde – und das darf durchaus als „work in progress“ angesehen werden – bei der Umsetzung des Graphen dessen Darstellung um diverse Funktionen wie Zoom und Pan erweitert. Im visuellen Graph können Nutzer*innen zudem über die Knoten und Kanten navigieren, sodass per mouse-over die Detailinformationen in einer Infobox angezeigt werden (siehe Abbildung 12).

⁹Die Eingabe der Werte in den Informationsklassen erfolgt in den Feldern im grauen Bereich rechts und wird durch angezeigte Vorschläge aus dem jeweiligen, bei der Informationsklasse hinterlegten Thesaurus erleichtert. Aktionen, wie etwa einen Kommentar hinzufügen, sind über ein Kontextmenü abrufbar.

¹⁰Für eine optimale Handhabung wurde die Anwendung so programmiert, dass sie im Zweisichtsbetrieb läuft. Aktuell können hier Koordinaten von Punkten oder Rechtecken gespeichert werden. Eine Ausweitung auf andere Polygone bedürfte unseres Erachtens die Integration von computergestützter Bildsegmentierung und ist gegenwärtig aufgrund des hohen Zeitaufwands bei einem ausschließlich manuellen Verfahren nicht vorgesehen.

¹¹<http://visualjournalism.unibz.it/play-data>, 11.10.2017.

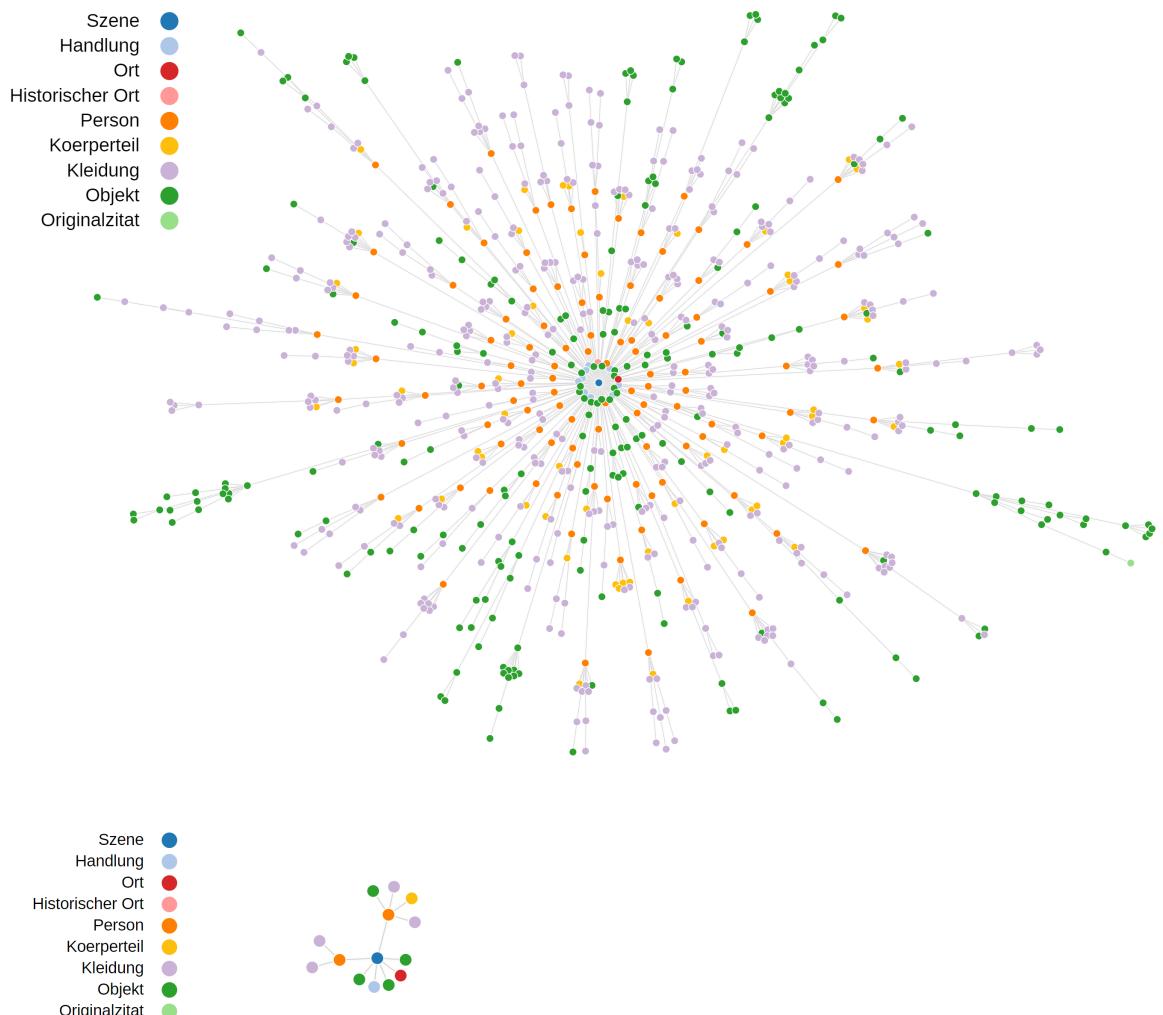


Abbildung 11: Einfache und umfangreiche Graphen in REALonline.

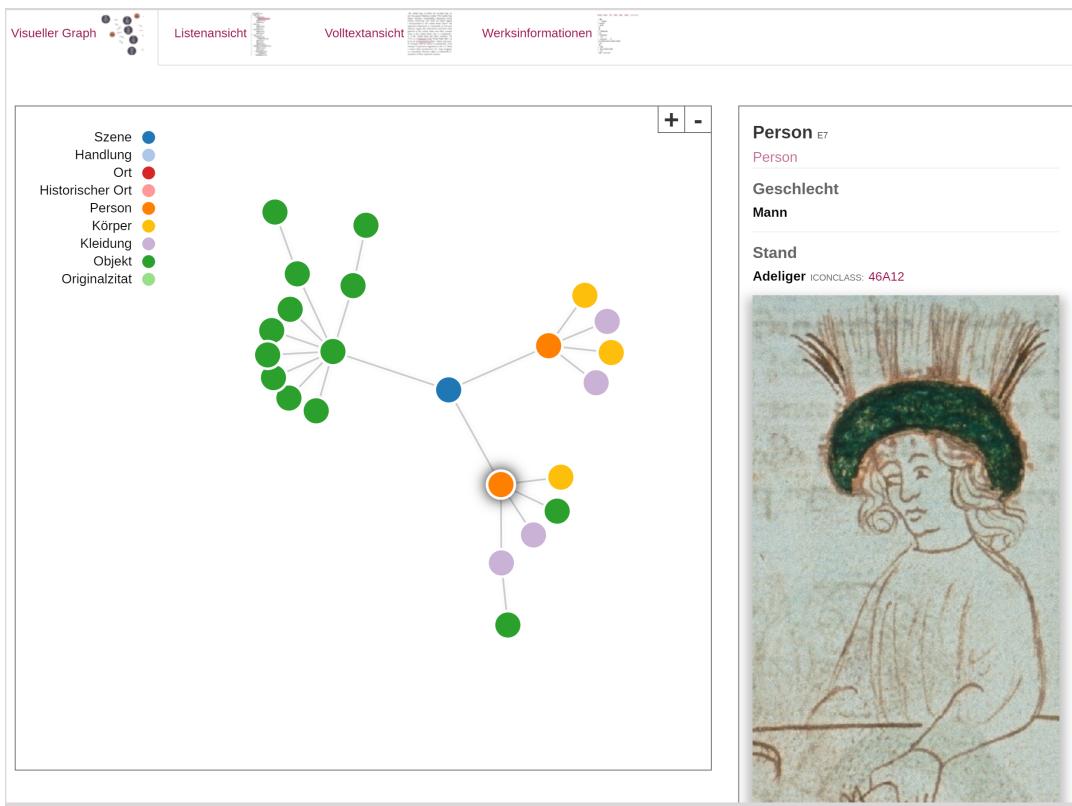


Abbildung 12: Anzeige der zu einem Knoten gespeicherten Informationen und getaggten Bildausschnitte im Frontend.

Technisch wurden die Graphen mit den Datenvisualisierungsbibliotheken des D3JS-Frameworks¹² (Bostock, Ogievetsky, Heer 2011) aufgebaut.

Die Farben des Graphen basieren auf der Arbeit von Cynthia Brewer (Brewer, Hatchard und Harrower 2003) und sind mittlerweile in der Version 2.0 online¹³ verfügbar. Die Verwendung dieser Zuordnung hatte zwei Vorteile gegenüber einer individuellen bzw. einer anderen Farbklassifizierung: Es konnten verschiedene Farbschemata ohne großen Aufwand getestet werden und die Skalierung des ColorBrewer-Schemas ist so angelegt, dass auf eine unterschiedliche Anzahl von Kategorien leicht Rücksicht genommen werden kann. Schließlich war auch die langjährige Erfahrung und Erprobung der Schemata speziell für web-basierte (GIS-)Anwendungen ein Auswahlkriterium und es besteht weiterhin die Option, bestimmten Nutzer*innenkreisen bzw. für eine barrierefreie Version der Seite eigene und individuell einstellbare Schemata zur Verfügung zu stellen.

Um die Frage beantworten zu können, wie die komplexen Informationen und Beziehungen aus dem Backend, die dort von einer sehr versierten und spezialisierten Nutzer*innengruppe eingegeben und verwendet werden, an das Frontend geliefert werden, müssen wir etwas ausholen.

Die Nutzung des Frontends beruht, im Gegensatz zum Backend, auf „Freiwilligkeit“. Mit anderen Worten: es besteht die Aufgabe für Entwickler*innen also darin, die Fragen der Nutzer*innen schnell,¹⁴ zufrieden-

¹²<https://d3js.org>, 11.10.2017.

¹³<http://colorbrewer2.org>, 11.10.2017.

¹⁴<https://www.nngroup.com/articles/how-long-do-users-stay-on-web-pages>, 11.10.2017.

stellend und möglichst eindeutig zu beantworten. Ansonsten wird die Nutzerin oder der Nutzer die Seite verlassen und nicht wiederkommen. Darüber hinaus sind die User und ihre Intentionen überwiegend unbekannt. Das Frontend muss daher

- eine möglichst einfache Ausgabe bereitstellen, die dennoch die Tiefe der Informationen abbildet und
- eine möglichst intuitive Interaktion bieten.

Aufgrund des in REALonline vorliegenden Datenmaterials lag es nahe, eine textbasierte Suche, wie sie von anderen Datenbanken und von Google bekannt ist, zu nutzen. Der nächste Schritt war, zu überlegen, wie eine solche Suche dann technisch am besten zu realisieren sei. Für die Verarbeitung und Darstellung der Daten aus dem Backend wird ein Apache Solr-Suchserver¹⁵ verwendet. Der Solr-Index der ca. 22.500 in REALonline enthaltenen Datensätze ist die Basis für die gesamte Anzeige im Frontend. Dieses Ergebnis wird als komprimiertes JSON-Dokument¹⁶ von Solr geliefert. Damit wird die Verbindung zwischen dem OpenSeadragon¹⁷ als Image-Server und textbasierter Indexabfragen hergestellt. Als Suchergebnis werden standardmäßig Vorschaubilder angezeigt.

4.4.2 Wie kommt der Graph ins REALonline-Frontend?

Es lag nahe, die vielfältigen Relationen der Beschreibungen auch in einer JSON-Notation abzulegen. Dafür standen wiederum zwei Möglichkeiten zur Auswahl: Einerseits hätte man das JSON-Format dazu verwenden können, um die Hierarchie des Graphen abzubilden und so schon jedes Element im JSON-Baum an die „richtige“ Stellen zu setzen.¹⁸ Da dies aber nur für den Graphen, nicht aber für die anderen Darstellungsmethoden der Detailansicht günstig war, entschieden wir uns für die andere Methode, nämlich die Elemente von den Beziehungen zu trennen. So werden die Beziehungen als eigene Zuordnung im Detailaufruf der API mitgegeben (Abb. 13). Der Graph im Frontend wird also aus einer JSON-formatierten Liste aus Elementen (mit IDs) und Relationen (parent/child-Verhältnis von jeweils zwei Objekten) erzeugt.

¹⁵ <http://lucene.apache.org/solr>, 11.10.2017.

¹⁶ <http://www.json.org/json-de.html>, 11.10.2017.

¹⁷ <https://openseadragon.github.io>, 11.10.2017.

¹⁸ <https://lakshminp.com/building-nested-hierarchy-json-relational-db>, 11.10.2017.

```

    ],
    {
        "id": "E8",
        "type": "Objekt",
        "name": [
            {
                "value": "Besen",
                "thesaurus_id": "gegenstand:besen"
            }
        ],
        "farbe": [
            {
                "value": "braun",
                "thesaurus_id": "farbe:braun"
            }
        ],
        "form": [
            {
                "value": "geschnürt",
                "thesaurus_id": "form:geschnürt"
            }
        ]
    },
    {
        "id": "E9",
        "type": "Kleidung",
        "name": [
            {
                "value": "Schleier",
                "thesaurus_id": "gegenstand:schleier",
                "thesaurus_lexicon": true
            }
        ],
        "farbe": [
            {
                "value": "weiß",
                "thesaurus_id": "farbe:weiß"
            }
        ]
    },
    {
        "id": "E10",
        "type": "Kleidung",
        "name": [
            {
                "value": "Rise",
                "thesaurus_id": "gegenstand:rise",
                "thesaurus_lexicon": true
            }
        ],
    }
]

```

Abbildung 13: JSON-Ergebnis der API aus dem Backend.

Ein Detailaufruf an die API (Application Programming Interface), aus dem wir den Graphen aufbauen, sieht so aus: <https://realonline.imareal.sbg.ac.at/api/data/work?apiKey=&detail=true&archivnr=005179A>, 11.10.2017.

So werden also nicht nur die Suchergebnisse („Suche“) sondern auch die Detailinformationen als Anfrage an die „API des Backends“ generiert. Diese „API des Backends“ ist genauer gesagt ein Interface der Daten des Backend-Servers am Frontend. Dies ist aus Gründen der Datenkonsistenz (Eingabe und Auslieferung können unterschiedlich sein) und Sicherheit notwendig. Lediglich die Startseite wird nicht über den Solr-Index generiert.

Ein Wort noch zu AngularJS: Wichtig an dieser Stelle ist, dass die Applikation als *single page application* um Angular 2¹⁹ wegen der dynamischen Inhalte erweitert wurde. Dadurch wird es möglich, eine Performance zu gewährleisten, wie sie vielen Nutzerinnen und Nutzern von anderen Services (wie etwa Amazon) bekannt ist. Dies erfordert aber den Einbau bzw. Umbau in jeder weiteren Komponente. Im Unterschied zum Backend (Java auf einem Tomcat Server) baut das Frontend auf Javascript (Angular, Wordpress CMS, Twitter Bootstrap, JQuery) PHP und Solr auf.

4.5 Die Daten von REALonline via Neo4j abfragen

Nachdem, wie oben bereits ausgeführt, die Relationen zwischen dargestellten Elementen grundlegend für das Verständnis der mittelalterlichen und frühneuzeitlichen Bildquellen sind, ist es nicht nur wichtig diese zu erheben, sondern sie auch in weiterer Folge abfragen zu können.

Graphdatenbanken bieten sich an, um zusammenhängende Daten in Form eines Graphen effizient und abfragbar zu speichern. Sie sind daher – im Gegensatz zu herkömmlichen SQL-Datenbanken – besonders für heterogen vernetzte Informationen geeignet (Angles und Gutierrez 2008, 5f.; Yoon, Kim und Kim 2017). Für das REALonline-Projekt fiel die Wahl auf Neo4j, eine weit verbreitete Graphdatenbank mit hervorragenden Anbindungsmöglichkeiten (Kumar Kaliyar 2015, 787). Neo4j ist unter einer Open-Source-Lizenz veröffentlicht und eignet sich daher umso mehr für die Integration in eine zu entwickelnde Projektinfrastruktur.²⁰

Neben dem im Zuge des Relaunchs entwickelten Frontend, das unterschiedlichen Nutzer*innengruppen – von interessierten Laien bis zu Forscher*innen – ermöglichen soll, das Material in REALonline einfach zu durchsuchen, sollen alle erhobenen Daten darüber hinaus in einer Graphdatenbank komplex abfragbar sein, ohne zu präjudizieren, welche Fragen dies sein könnten. Damit können die Daten nun online in diversen Analysekontexten geistes- und kulturwissenschaftlicher Forschung oder für kunst- und kulturvermittelnden Bestrebungen als Grundlage dienen. Im bisherigen System war eine solche Nutzung nur vor Ort am Institut in Krems möglich. Im Zuge des Relaunchs von REALonline war es uns ein besonderes Anliegen, diesen Service unabhängig von Platzkapazitäten am Institut oder von Reisekostenbudgets anbieten zu können. Hierfür ist die so genannte Expert*innensuche vorgesehen, die vom REALonline-Frontend direkt zu den Daten im Neo4j-Browser führt. Dieser kann mit jedem gängigen Webbrowser verwendet werden und ermöglicht es, Abfragen an die Datenbank zu stellen und das Ergebnis auf unterschiedliche Weise zu betrachten.

¹⁹ <https://angular.org>, 11.10.2017.

²⁰ <https://neo4j.com/product>, 11.10.2017.

Die gesamten Daten in REALonline – alle erhobenen Informationen zum Werk und zur Bildbeschreibung, alle Thesauri und alle zusätzlichen Angaben, wie Kommentare oder Normdaten – werden in dieser online abrufbaren Graphdatenbank vorgehalten.

Um die Daten aus der Mongo-Dokumentendatenbank in einer Graphdatenbank zu exportieren, müssen diese entsprechend aufbereitet werden. Diese Transformation besteht im Wesentlichen aus drei Schritten.

Einerseits werden die Informationen zum Kunstwerk bzw. Bildträger exportiert. Diese resultieren in Knoten für die Werke (*node label ,Work'* in der Neo4j-Datenbank) und eigene Knoten (*node label ,Entry'*) für die Eigenschaften (*properties*) der Werke. Des Weiteren werden die Beschreibungen (*node label ,Description'*) und deren Entitäten (*node label ,Entity'*) mit eigenen Knoten (*node label ,Entry'*) für ihre Eigenschaften exportiert. Die Lösung, Werks- und Entitätseigenschaften nicht nur als *properties* in den Knoten, sondern auch als eigene Knoten abzubilden, wurde gewählt, da viele Eigenschaften Mehrfacheinträge umfassen können. Jeder einzelne dieser Einträge kann wiederum multiple Kommentare oder Informationen zu Normdaten, bibliografische Referenzen sowie Metadaten beinhalten, die auf eine Unsicherheit oder auf eine Besonderheit im Datenmaterial hinweisen (Abb. 14). Auch nach diesen Knoten, Eigenschaften und Zusatzinformationen kann somit einfach abgefragt werden.

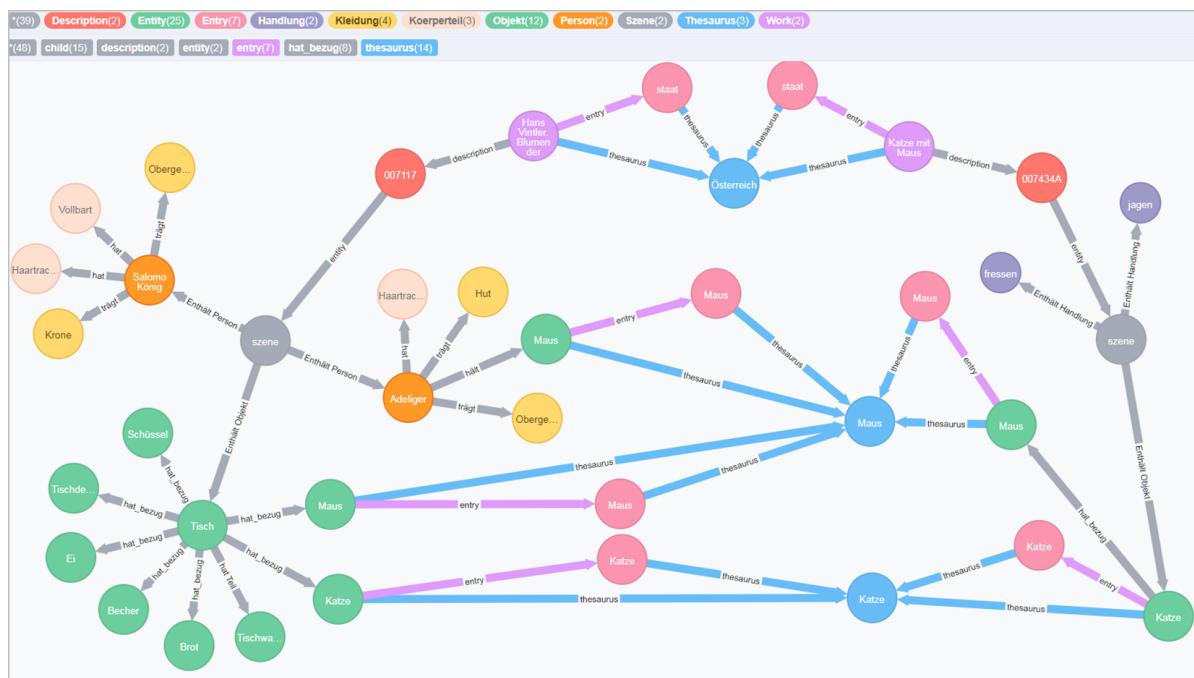


Abbildung 14: Graph von den Datensätzen 007117 und 007434A. Gezeigt werden ausgewählte Knoten und deren Verbindungen. Zu sehen ist auch, wie die beiden Datensätze über die Thesaurusknoten (hellblau) verbunden sind.

Die Beziehungen zwischen den Entitäten einer Beschreibung sind bereits in der Dokumentendatenbank gespeichert. Die Verknüpfung verschiedener Werke und deren Entitäten werden schließlich über die gemeinsamen Knoten der Thesauri realisiert. Dadurch ist es möglich, Entitäten eines Werkes mit jenen eines anderen Werkes in Beziehung zu setzen und mögliche Gemeinsamkeiten abzufragen.

Für interpretierbare Datentypen wurden eigene *properties* angelegt, die eine einfachere Abfragbarkeit der Einzelheiten dieser Datentypen ermöglichen. So werden z. B. für Maße – so verfügbar – eigene Properties für Höhe, Breite, Tiefe und Sondermaße angelegt. Um eine möglichst einfache Abfragbarkeit zu gewährleisten wurde bei den Beschreibungsdaten die Option von Neo4j genutzt, mehrere Bezeichner (*labels*) zu den Knoten zu speichern (Hunger 2014, 37f.). So kann etwa ein Kleidungsstück sowohl unter dem allgemeinen *label* „Entity“ als auch anhand des zusätzlich vergebenen *labels* „Kleidung“ in Querys angesprochen werden. REALonline umfasst mit über 2,5 Millionen Knoten und mehr als 5 Millionen Kanten eine Anzahl an Elementen, die weder auf einem Bildschirm sinnvoll angezeigt, noch in anderer Form auf eine Weise dargestellt werden kann, die es einer Benutzerin oder einem Benutzer ermöglicht, Rückschlüsse direkt aus der Übersicht über die Daten zu ziehen. Wie bei allen umfangreichen Datensammlungen ist es daher essentiell, diese präzise abfragen zu können. Neo4j bietet mit *cyphe*²¹ eine mächtige Sprache, die es ermöglicht, spezifische Abfragen auch in sehr großen Datenbeständen durchzuführen. Dadurch können als interessant erachtete Details genau analysiert und für die gegenständliche Untersuchung irrelevante Informationen verborgen werden. Gerade für Expert*innen ist es oft auch wichtig, gewisse Aspekte eines Graphen zu untersuchen, während andere Elemente nicht immer von Belang sein müssen.

Dabei bietet *cyphe* einerseits eine niedrige Lernschwelle und eignet sich somit gut, um schnell zu interessanten Ergebnissen zu kommen (Robinson, Webber und Eifrem 2015, 28), ist andererseits jedoch mächtig genug, um auch komplexe Abfragen stellen zu können.

Je nachdem, wie man die Parameter in der Ausgabeanforderung definiert, kann als Ergebnis entweder eine Visualisierung des Teilgraphen oder eine Tabelle ausgegeben werden. Bei ersterem kann das angezeigte Suchresultat von den Nutzerinnen und Nutzern interaktiv erweitert werden, um darüberhinausgehende Kontexte des Teilgraphen zu erschließen. Die Darstellung der Knoten und Kanten kann mit Grass (*Graph Style Sheet*), einer einfachen an CSS angelehnten Style-Sheet-Sprache, an die jeweiligen Bedürfnisse angepasst werden, wodurch eine übersichtlichere Darstellung ermöglicht wird. Zusätzlich können die Ergebnisse aber auch in den Formaten JSON und CSV exportiert werden. Für die Verwendung der Daten gelten die in REALonline festgelegten Nutzungsbedingungen.²²

Die *cyphe*-Abfragen im Neo4j-Browser bieten damit über die Möglichkeiten von Volltext-, Facetten-, Thesaurus- und Geo-Browser-Suche im REALonline-Frontend hinausgehende, komplexe und spezifische Fragen an die in REALonline erfassten Daten zu stellen.²³

So kann etwa konkret nach Darstellungen von Personen gesucht werden, die über den Thesaurus „Stand/Lebensalter“ dem im Haushalt tätigen Personal zugeordnet wurden und die Objekte in Händen halten. Bei diesem Beispiel haben wir die Objekte (grüne Knoten) und zusätzlich die Thesauruskategorien (blaue Knoten), denen sie zugeordnet sind, ausgegeben, damit Muster oder Besonderheiten im Datenmaterial einfacher detektierbar sind (Abb. 15).

²¹<https://neo4j.com/docs/developer-manual/current/cypher>, 11.10.2017.

²²<https://realonline.imareal.sbg.ac.at/nutzung>, 25.6.2018.

²³Die nachstehenden Beispiel-Querys sind auf den aktuellen Stand der Graphdatenbank abgestimmt (25.6.2018). Aufgrund von Weiterentwicklungen kann es künftig notwendig sein, geringfügige Adaptionen vorzunehmen, um zu Ergebnissen zu kommen. Bei Fragen wenden Sie sich bitte an sekretariat.imareal@sbg.ac.at.

Abfrage 1

```
match (a:Person)-[:thesaurus|broader *1..2]->(:Thesaurus {LongName: 'Haushalt'})
```

with a

```
match (a)-[:child {RelationLabel: 'hält'}]->(b:Entity)-[c:thesaurus|:broader*1..5]->(d:Thesaurus {ThesaurusName: 'Haushalt'})
```

return b,c,d

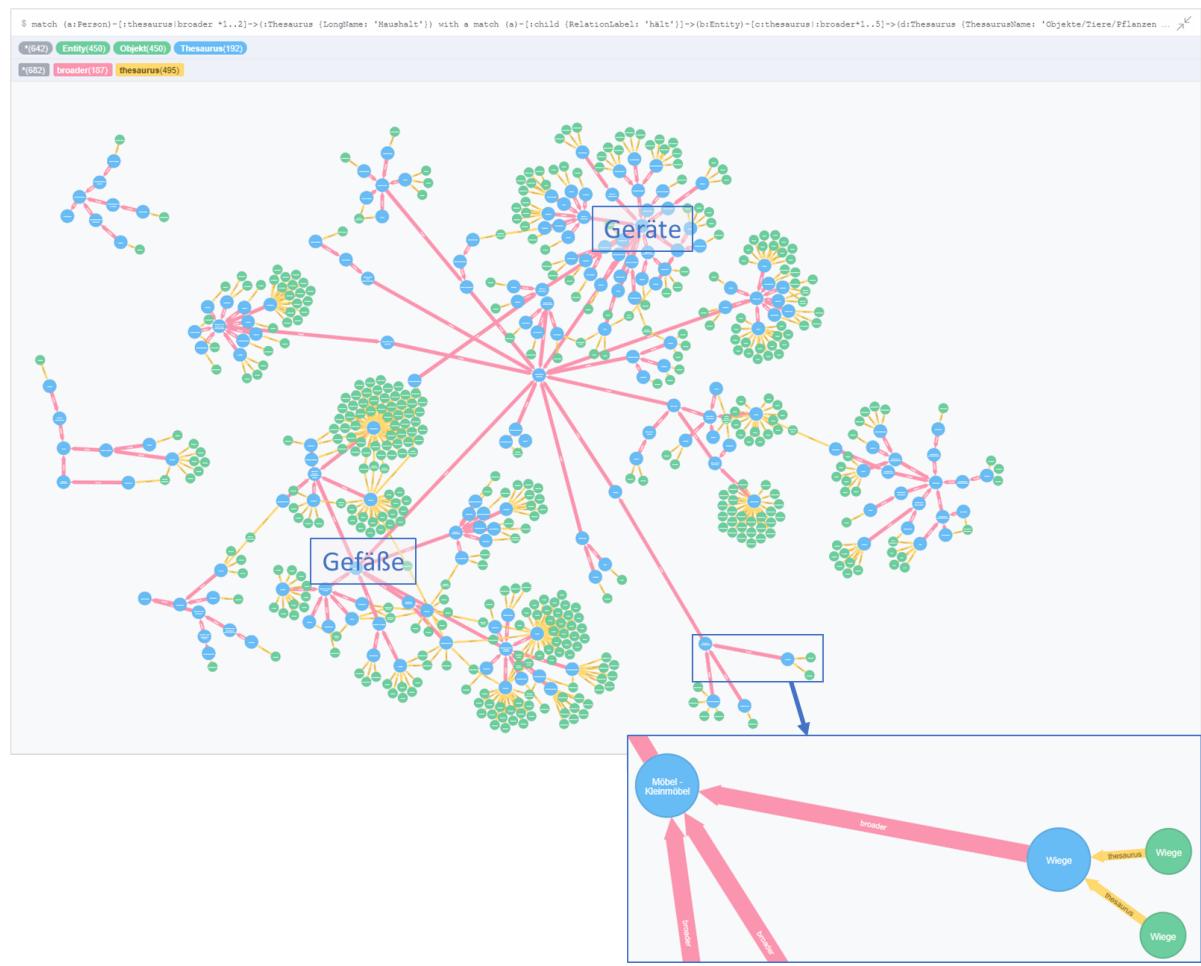


Abbildung 15: Screenshot der Ausgabe von Abfrage 1 im Neo4j-Browser; hinzugefügt wurden in der Grafik ausgewählte Begriffe und die vergrößerte Darstellung eines Teilgraphen. Verbindungen zwischen Thesaurusknoten sind rosa Pfeile, während Relationen zwischen Objekten und Thesaurusknoten in grau dargestellt sind.

Die größte Anzahl an Objekten kann der Thesaurusgruppe Gefäß zugeordnet werden (insgesamt 205 von 450), auch unterschiedliche Geräte werden (erwartungsgemäß) oft von diesen Personen in Händen gehalten. Überraschender ist der Befund, dass Kleimöbel nur in wenigen Fällen getragen dargestellt

wurden. Die von einer Magd unter den Arm geklemmte Wiege auf einem Holzschnitt aus Albrecht Dürers Marienleben²⁴ und eine Darstellung auf einem Flügelaltar von 1579²⁵, die sich eindeutig auf die Dürer'sche Druckgrafik bezieht, sind die einzigen beiden Beispiele im gesamten Datenbestand von REALonline, wo dieses Kleinmöbel von einer Person getragen gezeigt wird.²⁶ Es handelt sich hier also wohl um eine Erfindung Dürers, der mit dieser im Bild hergestellten Verbindung zwischen Magd und Wiege ein Moment der Unmittelbarkeit in der Szene vermittelt: Die Geburt der Gottesmutter hat eben erst stattgefunden; das kleine Bett für Maria steht noch nicht bereit, sondern wird erst gerade an seinen Bestimmungsort gebracht. Die Abfrage der Daten in REALonline macht es hier also sowohl möglich, motivische Zusammenhänge in über 70 Jahren auseinanderliegenden Kunstwerken zu entdecken als auch auf einen konkreten Wandel in den Darstellungsmöglichkeiten und Vermittlungsstrategien visueller Medien aufmerksam zu werden.²⁷

In einem weiteren Beispiel für eine Abfrage der Daten über den Neo4j-Browser möchten wir herausfinden, auf welchen Bildelementen hebräische (bzw. hebraisierende) Inschriften oder Schriftzeichen dargestellt wurden.

Abfrage 2

```
match (a:Thesaurus {ThesaurusName:'Formen'})->-[m]-(b:Entity)->-[n]-(c:Entity)
where 'hebräisch' in a.LongName
with a,b,c,m,n
match (c)-[e:thesaurus|broader*]->(f:Thesaurus{ThesaurusName: 'Objekte/Tiere/Pflanzen etc.'})
return a,b,c,e,f,m,n
```

²⁴<https://realonline.imareal.sbg.ac.at/detail/?archivnr=014301>, 11.10.2017.

²⁵<https://realonline.imareal.sbg.ac.at/detail/?archivnr=015601>, 11.10.2017.

²⁶Vgl. <https://realonline.imareal.sbg.ac.at/2017/09/01/suchbild-mit-wiege/#more-2325>, 11.10.2017. Die Frage, warum v. a. Kleinmöbel auf Bildern nur in ganz bestimmten Zusammenhängen von einer Person getragen werden, wird in einem Kapitel der Dissertation von Isabella Nicka, „Prozessierte Objekte? Dargestellte Möbel in den visuellen Medien des Mittelalters“, erörtert. (Publikation in Vorbereitung)

²⁷In der interdisziplinären Forschungsperspektive *object links – objects link* des IMAREAL (<https://realonline.imareal.sbg.ac.at/home/forschung/object-links/>, 11.10.2017.) ist das Erkenntnisinteresse, welche Funktionen und Bedeutungen die Verbindungen zwischen Objekten und Personen bzw. zwischen Objekten und Objekten haben. Ein geplantes Teilprojekt soll sich den dargestellten Objekt-Objekt- bzw. Person-Objekt-Relationen widmen und auf dem Datenmaterial in REALonline aufbauen.

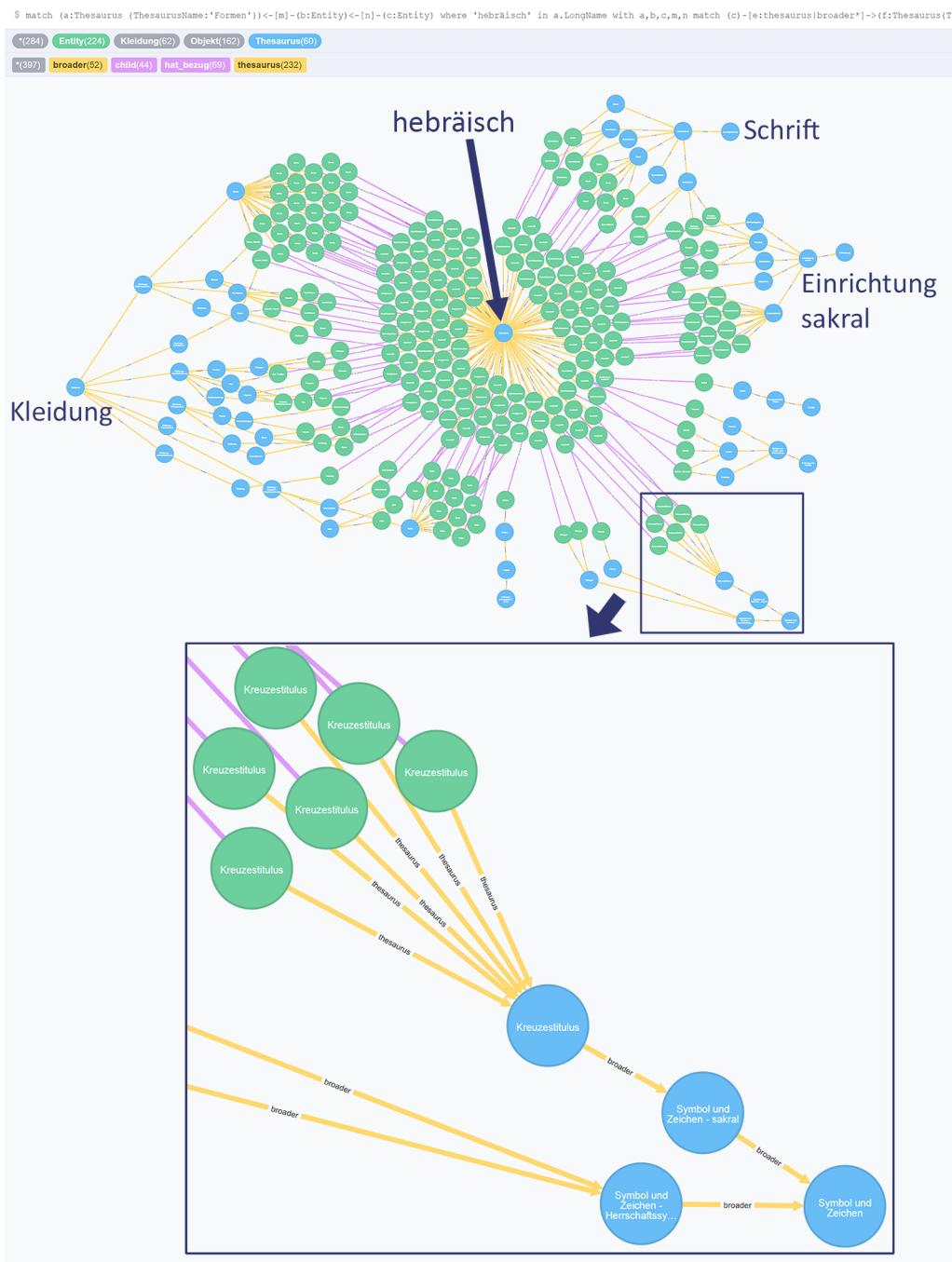


Abbildung 16: Screenshot der Ausgabe von Abfrage 2 im Neo4j-Browser; hinzugefügt wurden in der Grafik ausgewählte Begriffe und die vergrößerte Darstellung eines Teilgraphen. Verbindungen zwischen Thesaurusknoten sind gelbe Pfeile, während Relationen zwischen Objekten und Thesaurusknoten in lila dargestellt sind.

Im Zentrum der Visualisierung findet sich der blaue Thesaurusknoten ‚hebräisch‘, der mit 112 Objekten (grüne Knoten in der Mitte) verbunden ist (Abb. 16). Der größte Teil der Bildelemente, die mit hebräischen Schriftzeichen versehen sind, können den Kategorien Kleidung oder Textilien zugeordnet werden. Objekte, die mit Schriftlichkeit in Zusammenhang stehen (wie Bücher oder Schriftrollen) sind ebenso wie der sakralen Einrichtung zuzuordnende Elemente (wie dargestellte Retabel oder aufgestellte Gesetzestafeln) oft mit hebräischer Schrift versehen.

Im Datenmaterial von REALonline finden sich auch sechs von insgesamt 293 erfassten Kreuzestituli, die eine hebräische oder hebraisierende Inschrift erkennen lassen. Einer davon ist auf einer Tafel mit der Beweinung zu sehen, die Teil des ehemaligen Retabels am Hochaltar der Schottenkirche in Wien war.²⁸ Margaretha Boockmann (2013, 225–264) versammelt in ihrer Publikation zu den dargestellten hebräischen und hebraisierenden Inschriften auf Gemälden vor 1540 47 Beispiele für die bildliche Umsetzung einer konkreten neutestamentlichen Stelle: Dieser zufolge ist ein Schild am Kreuz Christi befestigt worden, worauf in den drei Sprachen Hebräisch, Lateinisch und Griechisch „Jesus von Nazareth, der König der Juden“ geschrieben war.²⁹ Am genannten Beispiel in REALonline, das in der Jubiläumsfestschrift des Instituts für Realienkunde noch ausführlicher behandelt wird,³⁰ fällt auf, dass es sich nicht um einen mehrsprachigen Titulus handelt, sondern hier offenbar eine additive Rezeptionsstrategie von den Betrachter*innen eingefordert wurde: Die lateinische Variante INRI findet sich nämlich an der ein Bildfeld links vor der Beweinung angebrachten Kreuzigung.³¹ Komplexe Abfragen der in REALonline erfassten Bilddetails und ihrer Beziehungen zu anderen Bildelementen können so also auf Besonderheiten, wie diese Art des Erzählens der Passionsgeschichte am Wiener Schottenretabel, aufmerksam machen und damit neue Forschungsfragen anstoßen.³²

5 Zusammenfassung

Für viele datenbankgestützte Langzeitprojekte stellt sich irgendwann die Frage, ob sie auf der Basis der eingesetzten Software und der zur Anwendung kommenden Datenmodelle weiterbetrieben werden können oder sollen. Zudem erfordert die mit dem Medienwandel einhergehende Öffnung der Forschung, die oft in Jahrzehntelanger Arbeit in Datenbanken eingepflegten Forschungsdaten über das Internet zugänglich und abfragbar zu machen, wodurch „in die Jahre gekommene“ Datenbanken an die Grenzen des Machbaren stoßen können.

Im Zuge des Relaunchs der Bilddatenbank REALonline haben wir uns folgende Fragen gestellt, die exemplarisch mitunter auch für andere Projekte relevant sein können:

²⁸<https://realonline.imareal.sbg.ac.at/detail/?archivnr=000707>, 11.10.2017. Margaretha Boockmann (2013, 245, Kat.Nr. 330) hält für den Kreuzestitulus fest, dass manche der vier Zeichen hebräischen Schriftzeichen ähneln, aber nicht deutlich ist, was konkret damit gemeint sein soll.

²⁹Johannes 19, 19–22 und ähnlich in den synoptischen Evangelien.

³⁰Der Beitrag „Object links in/zu Bildern mit REALonline analysieren“ von Isabella Nicka erscheint in der Festschrift, die anlässlich des 50-jährigen Bestehens des Instituts für Realienkunde 2019 erscheinen wird.

³¹<https://realonline.imareal.sbg.ac.at/detail/?archivnr=000301>, 11.10.2017.

³²Weitere Beispiele dafür werden bei Nicka (2017, 229–231) und bei Matschinegg und Nicka (2018, 25–29) angeführt.

- Wie können wir die vorliegenden (mitunter komplex vernetzten) Daten verlustfrei und vollständig in eine neue Systemumgebung migrieren?
- Welche Technologien stehen zur Verfügung?
- Können wir vorkonfektionierte Systeme verwenden bzw. adaptieren oder sollen wir eine „maßgeschneiderte Lösung“ neu entwickeln?
- Ist das neue System offen für Modifikationen, die sich im Laufe des Datenbankumbaus abzeichnen?

Die Datenbank, anhand der diese und viele andere Problemstellungen in der Praxis umgesetzt wurden, ist die über mehr als 40 Jahre angewachsene Bilddatenbank des Instituts für Realienkunde des Mittelalters und der frühen Neuzeit, in der visuelle Medien systematisch erschlossen werden, um sie für die Erforschung der materiellen Kultur und für bildwissenschaftliche Analysen bereit zu stellen. Jedes in REALonline angelegte Dokument enthält die in der Kulturerbedokumentation standardmäßig erfassten Informationen zu den Werksdaten (Künstler, Standort, Material, Technik, ...). Das Alleinstellungsmerkmal dieser Datenbank ist jedoch die detaillierte Beschreibung der Darstellungen auf den Bildquellen, bei der alle am Bild erkennbaren Personen einschließlich der von ihnen getragenen Kleidung, alle Objekte wie Objektteile, Pflanzen und Tiere sowie die Beziehungen zwischen den Personen und den Objekten erfasst werden. In der Ausgangssituation lagen diese Daten zu ca. 22.500 Werken in einer hierarchischen Strukturierung vor, mit einer Gesamtzahl von ca. 2,3 Millionen an textuellen Elementen und zusätzlichen Daten zu Thesauri und lexikalischen Informationen im Umfang von ca. 195.000 Informationseinheiten, die es neu zu modellieren galt.

Bei der Realisierung des Relaunchs haben wir uns für die Neukonzeption der Datenbank mitsamt einer effizienten Datenerfassung im Backend, sowie einer grundlegend neuen Verfügbarmachung der digitalen Bilder und der textuellen Daten im Internet entschieden. Dabei kommen auf mehreren Ebenen Graphentechnologien zum Einsatz, mit denen die Repräsentation komplexer und vernetzter Daten der Bildbeschreibungen optimal unterstützt wird.

Auf der Ebene der Datenmodellierung konnten die vorliegenden Daten zu den Bildbeschreibungen – mit einigen Modifikationen – in Knoten (also Personen, Objekte etc.), Kanten (Beziehungen zwischen den Knoten) und Eigenschaften (Werte, die Knoten und Kanten besitzen können) transformiert werden. Für das Backend wurde, basierend auf der Graphenstruktur des Datenmodells der Beschreibungen, ein GUI mit einer Graphvisualisierung entwickelt, das die Vernetzung der Daten ermöglicht und zudem eine im Vergleich zu Eingabemasken oder -listen bessere Sicht auf die Daten gewährleistet. Auf der Ebene der Beschreibungen von Einzelbildern bietet die Graphvisualisierung eine gute Übersicht auf die Struktur der Beschreibungsdaten, die wir auch für die User im Frontend als Standardansicht bereitstellen. Diese anspruchsvolle Form der Informationsvisualisierung im Frontend erfolgt auf der Grundlage von forschungsmäßig ausgetesteten Farbschematas und kann neben vielen anderen Verbesserungen des User-Interfaces als eine innovative Komponente der neuen Internetversion von REALonline angesehen werden. User, die sich bereits mit Graphdatenbanken auskennen oder gewillt sind, dies zu lernen, erhalten via Expert*innensuche einen read-only Zugang zur Graphdatenbank wo sie im Neo4j-Browser mithilfe der Abfragesprache *cypher* mit dem gesamten Datenbestand forschen können. Zu diesem Zweck haben wir die Werks- und Beschreibungsdaten, alle verfügbaren Thesauri und die Anreicherungen

mit Referenzen zu Normdaten in einem Graphen exportiert. Neo4j verfügt über einen interaktiven Grapheditor der durch den Klick auf einen Knoten diesen öffnet und Beziehungen zu den weiteren Knoten zueinander klickbar macht und auch individuelle Formatierungen ermöglicht.

In diesem Beitrag haben wir schwerpunktmäßig den Einsatz von Graphentechnologien beim Relaunch von REALonline beschrieben. Darüber hinaus haben wir versucht, die Schemagrundlagen für die Datenstrukturierung mit JSON und die Entwicklung der Dokumentendatenbank zu skizzieren, wofür wir mit mongoDB eine nosql-Lösung als die am besten Geeignete erachteten. Daraus werden die aktuellen Daten an den SOLR-Index im Frontend geliefert wie auch der Graph für die Neo4j-Datenbank generiert. Alle in REALonline eingesetzten Komponenten, darunter auch der OpenSeadragon als Imageserver und Wordpress als WebCMS, bilden eine *Open-Source*-Entwicklungsumgebung.

Literaturverzeichnis

- Angles, Renzo; Gutierrez, Claudio. 2008. Survey of Graph Database Models. In: *ACM Computing Surveys* 40/1, Article 1, February 2008, DOI = 10.1145/1322432.1322433, <http://doi.acm.org/10.1145/1322432.1322433>, 11.10.2017.
- Bertau, Karl. 1994. *Epistola cum Libello ackerman und das buichlein ackerman*. Nach der Freiburger Hs. 163 und nach der Stuttgarter Hs. Hb X 23. Herausgegeben und übersetzt von Karl Bertau, 2 Bde. Berlin: Walter De Gruyter.
- Boockmann, Margaretha. 2013. *Schrift als Stigma*. Hebräische und hebraisierende Inschriften auf Gemälden der Spätgotik. Schriften der Hochschule für jüdische Studien Heidelberg 16. Heidelberg: Universitätsverlag Winter.
- Bostock, Michael; Ogievetsky, Vadim; Heer, Jeffrey. 2011. D3: Data-Driven Documents. In: *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12), 2301–2309, DOI: 10.1109/TVCG.2011.185.
- Brewer, Cynthia A.; Hatchard, Geoffrey W.; Harrower, Mark A. 2003. ColorBrewer in Print: A Catalog of Color Schemes for Maps. In: *Cartography and Geographic Information Science*, 30(1), 5–32, DOI: 10.1559/152304003100010929.
- Bry, François; Schemainda, Corina; Schefels, Clemens. 2015. *A Gaming Ecosystem Crowdsourcing Deep Semantic Annotations*. Research Report. <http://www.en.pms.ifi.lmu.de/publications/PMS-FB/PMS-FB-2015-1/PMS-FB-2015-1-paper.pdf>, 11.10.2017.
- Hunger, Michael. 2014. *Neo4j 2.0: eine Graphdatenbank für alle*. Frankfurt am Main: Entwickler.press.
- Jaritz, Gerhard. 1993. *Images. A Primer of Computer-Supported Research with KLEIO IAS*. Halbgraue Reihe zur historischen Fachinformatik A 22. St. Katharinen: Scripta Mercatura Verlag.
- Kaufmann, Sascha; Andrews, Tara Lee. 2016. Bearbeitung und Annotation historischer Texte mittels Graph-Datenbanken am Beispiel der Chronik des Matthias von Edessa. In: *Konferenzabstracts der DHd 2016, Univ. Leipzig*: Modellierung, Vernetzung, Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma. Publikationskoordination Elisabeth Burr, 176–178. <http://dhd2016.de/boa.pdf>, 11.10.2017.
- Kuczera, Andreas. 2017. Graphentechnologien in den Digitalen Geisteswissenschaften. In: *AbI Technik. Zeitschrift für Automation, Bau und Technik im Archiv-, Bibliotheks- und Informationswesen*, 37/3. <https://www.degruyter.com/downloadpdf/j/abitech.2017.37.issue-3/abitech-2017-0042/abitech-2017-0042.pdf>, 11.10.2017.
- Kumar Kaliyar, Rohit. 2015. Graph databases: A survey. In: *Computing, Communication & Automation (ICCA)*, 2015 International Conference on IEEE, 785–790.
- Klotz, Peter. 2016. „Ekphrasis und Kunstbeschreibung“. In: *Handbuch Sprache in der Kunstkommunikation*. Herausgegeben von Heiko Hausendorf und Marcus Müller, 176–197. Handbücher Sprachwissen 16. Berlin: Walter De Gruyter.
- Maeda, John. 2006. *Simplicity*. Design, Technology, Business, Life. Cambridge (Mass.) u.a.: MIT Press.

Matschinegg, 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.

Matschinegg, Ingrid; Nicka, Isabella. 2018. REALonline Enhanced. Die neuen Funktionalitäten und Features der Forschungsbilddatenbank des IMAREAL. In: MEMO 2 (2018): Digital Humanities & Materielle Kultur, 10–32. Pdf-Format, <http://memo.imareal.sbg.ac.at/warticle/memo/2018-matschinegg-nicka-realonline-enhanced/> doi: 10.25536/20180202.

Nicka, Isabella. 2014. Interfaces. Berührungszonen von Transzendenz und Immanenz im spätmittelalterlichen Sakralraum. In: *Visualisierungen von Kult*. Herausgegeben von Marion Meyer und Deborah Klimburg-Salter. Wien–Köln–Weimar: Böhlau, 260–293, Abb. auf 438–444.

Nicka, Isabella. 2017. REALonline—Explore and Find Out. Wohin führt das ‚Digitale‘ die Kunstgeschichte? In: *Newest Art History. Wohin geht die jüngste Kunstgeschichte?* Tagungsband zur 18. Tagung des Verbandes österreichischer Kunsthistorikerinnen und Kunsthistoriker. Herausgegeben vom Verband österreichischer Kunsthistorikerinnen und Kunsthistoriker (VöKK). Wien, Online-Publikation, 223–235. <http://www.kunsthistoriker-in.at/?nav=&gruppe=&artikel=10822>, 11.10.2017.

Robinson, Ian; Webber, Jim; Eifrem, Emil. 2015. *Graph Databases*. New Opportunities for Connected Data. 2. Aufl. Sebastopol: O’Reilly.

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 Bd. 4 / Sitzungsberichte der Akademie der Wissenschaften, Phil.-Hist. Klasse 374. Wien: Verlag der Österreichischen Akademie der Wissenschaften, 167–194.

Thaller, Manfred. 1989. *KLEIO. Ein Datenbanksystem*. Halbgraue Reihe zur historischen Fachinformatik B1. St. Katharinen: Scripta Mercaturae Verlag.

Vavra, Elisabeth; Harant, Patricia; Blaschitz, Gertrud. 2004. Die Bild- und Textdatenbank am Institut für Realienkunde des Mittelalters und der frühen Neuzeit. In: *Mediävistik und Neue Medien*. Herausgegeben von Klaus van Eickels, Ruth Weichselbaumer Ruth und Ingrid Bennewitz. Ostfildern: Thorbecke, 149–164.

Vonach, Andreas. 2015. Biblische Zitate und Reminiszenzen in Hans Vintlers „Die Blumen der Tugend“. In: *Hans Vintler: Die Blumen der Tugend (1411)*. Symposium nach 600 Jahren (Bozen, 28.–30. September 2011). Aktenband. Innsbruck: Universitätsverlag Wagner (= Schlerm-Schriften, 362), 385–397.

Yoon, Byoung-Ha; Kim, Seon-Kyu; Kim, Seon-Young. 2017. Use of Graph Database for the Integration of Heterogeneous Biological Data. In: *Genomics and Informatics* 15/1, 19–27.