

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

2019

ConedaKOR

–

Von der Tabelle zum Netzwerk

Thorsten Wübbena, Moritz Schepp

DARIAH-DE Working Papers

Nr.30

Wübbena, Thorsten; Schepp, Moritz: ConedaKOR : Von der Tabelle zum Netzwerk
Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2019
(DARIAH-DE working papers 30)

Verfügbar:

PURL: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2019-2>

URN: <http://nbn-resolving.org/urn:nbn:de:gbv:7-dariah-2019-2-9>

Dieser Beitrag erscheint unter der Lizenz [Creative-Commons Attribution 4.0 \(CC-BY\)](https://creativecommons.org/licenses/by/4.0/)



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 Text versteht sich als Essenz des Beitrags, der auf der 2017 durchgeführten Tagung „Graphentechnologien“ (Akademie der Wissenschaften und der Literatur in Mainz) in der Sektion „Bild und Graph“ präsentiert wurde. Im Zentrum der Betrachtung steht hierbei die Applikation ConedaKOR und die damit gewonnene Erfahrung aus der praktischen Arbeit am Deutschen Forum für Kunstgeschichte Paris (Max Weber Stiftung)

Keywords: Graphdatabase; Art-History
Graphdatenbanktechnologie; Kunstgeschichte

ConedaKOR

Von der Tabelle zum Netzwerk

Thorsten Wübbena

Moritz Schepp

Deutsches Forum für Kunstgeschichte Paris / Centre allemand d'histoire de l'art Paris (Max Weber Stiftung)



Thorsten Wübbena, Moritz Schepp: „ConedaKOR“. *DARIAH-DE Working Papers* Nr. 30. Göttingen: DARIAH-DE, 2019. URN: [urn:nbn:de:gbv:7-dariah-2019-2-9](https://nbn-resolving.org/urn:nbn:de:gbv:7-dariah-2019-2-9).

Dieser Beitrag erscheint unter der
Lizenz [Creative-Commons Attribution 4.0](https://creativecommons.org/licenses/by/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 Text versteht sich als Essenz des Beitrags, der auf der 2017 durchgeführten Tagung „Graphentechnologien“ (Akademie der Wissenschaften und der Literatur in Mainz) in der Sektion „Bild und Graph“ präsentiert wurde. Im Zentrum der Betrachtung steht hierbei die Applikation ConedaKOR und die damit gewonnene Erfahrung aus der praktischen Arbeit am Deutschen Forum für Kunstgeschichte Paris (Max Weber Stiftung)

Schlagwörter

Graphdatenbanktechnologie, Kunstgeschichte

Keywords

Graphdatabase, Art-History

Inhaltsverzeichnis

| | | |
|---|-------------------------|---|
| 1 | ConedaKOR | 4 |
| 2 | Beziehungen und Graphen | 4 |
| 3 | Umsetzung | 6 |
| | Literaturverzeichnis | 9 |

1 ConedaKOR

Aber die Arbeit mit Datenbanken bricht eben auch tiefsitzende Überzeugungen geisteswissenschaftlichen Tuns auf, insbesondere dort, wo sie die gerade auch in progressiven Zirkeln verwurzelte Sicherheit unterläuft, dass nur das große künstlerische Individuum substanzielle historische Evidenzerfahrung ermöglicht. (Kohle, 2013, 36)

Wie auch in anderen Disziplinen gilt es in der Kunstgeschichte bei zu planenden Forschungsvorhaben mit Datenbankeinsatz, allgemein die folgenden relevanten Aspekte nacheinander zu betrachten: die reale Welt, deren Abstraktion in Form eines Datenmodells und das Datenbankmodell bzw. das nutzbare Datenbanksystem. Datenbankinformationen geben damit in repräsentativer Form Auskunft über etwas außerhalb der Datenbank Befindliches, andererseits legt das konzeptuelle Modell der Datenbank fest, was überhaupt als Information zählt, d. h. durch welche Informationen die Wirklichkeit beschrieben wird (Konstruktion). Der Blick auf die oberflächliche Präsentation der Datenbank ist somit nie der Blick auf die Datenbank selbst und die Strukturiertheit der inhärenten Informationen, sondern bietet schlicht eine Möglichkeit an.

2 Beziehungen und Graphen

Was macht aber nun eine Graphdatenbank aus? Das ihr zugrunde liegende Prinzip ist leichter zu verstehen, als es bei vielen anderen Modellen der Fall ist. In aller hier gebotenen Kürze und mit dem Hinweis auf die Erläuterungen in den anderen Beiträgen dieser Tagung kann gesagt werden, dass ein Graph aus zwei Elementen besteht: einem Knoten und einer Beziehung (Kante). Jeder Knoten repräsentiert eine Entität (z. B. Person, Ort oder Sache) und jede Beziehung stellt dar, auf welche Art zwei Knoten miteinander verbunden sind. Zum besseren Verständnis hilft hier oftmals ein Vergleich mit dem Aufbau eines einfachen Satzes: Subjekt und Objekt stellen jeweils einen Knoten dar und ein Prädikat steht für die Beziehung. Bei diesem Modell der Graphdatenbanken liegt es auf der Hand, dass, im Gegensatz zu anderen Datenbanksystemen, die Beziehungen Priorität haben. Eine Tatsache, die dem Umstand entgegen kommt, dass die in den Geisteswissenschaften¹ vorkommenden Daten eben nicht alleine wegen ihrer Quantität von Bedeutung sind, sondern ihren Wert aus den Verknüpfungen untereinander beziehen. Die Beziehungen liefern gerichtete, benannte semantisch relevante Verbindungen zwischen zwei Entitäten und da eine Beziehung immer einen Start- und Endknoten hat, kann auch kein Knoten gelöscht werden, ohne nicht auch die zugehörige Beziehung zu löschen. Damit ist sichergestellt, dass eine bestehende Beziehung nie auf einen nicht existierenden Endpunkt zeigt. Nicht zuletzt durch die oben erwähnte Nähe zur Sprache lassen sich auch in einem nicht-technischen Umfeld recht schnell erste Datenmodelle erstellen und wenn zuvor bereits mit einem Objektmodell oder einem Entity-Relationship-Modell gearbeitet wurde, erscheint das Graphmodell sehr vertraut. Ein weiterer Vorteil ist die einfache Erweiterbarkeit des Datenmodells, denn da sich viele Forschungsfragen in geisteswissenschaftlichen Projekten erst im Verlauf des Vorhabens ergeben, muss das Datenbanksystem in der Lage sein, flexibel

¹Ein Einblick in das breite Anwendungsspektrum der Graphentechnologie in den Geisteswissenschaften konnte auf der Tagung „Graphentechnologien. Neue Perspektiven für die Digital Humanities“ (19.–20. Januar 2017, Akademie der Wissenschaften und der Literatur | Mainz) gewonnen werden.

auf diese Anforderungen reagieren zu können. In der praktischen Arbeit wird die Effizienz dann eben auch wesentlich durch das Datenmodell bestimmt, als nur durch die Bedingungen der Datenerhebung und eine Frage in diesem Zusammenhang wäre sicher: „Unter welchen Gegebenheiten können Datensammlungen nun als wissenschaftliche Quellen dienen und Ursprung neuer Erkenntnisse sein?“ In den Digital Humanities dominieren ja bekanntlich *die* Geisteswissenschaften, die mit Texten arbeiten, die bild- und objektorientierte Kunstgeschichte hingegen ist noch immer wenig präsent. Das hängt unmittelbar mit den fachspezifischen Forschungsdaten zusammen, denn wenngleich in der Kunstgeschichte Fachbibliographien und Quellenverzeichnisse erstellt werden – Priorität hat das Bild.² Obschon das Kunstwerk der eigentliche Gegenstand der Forschung ist, hilft in den meisten Fällen ein digitaler Repräsentant, den Untersuchungsgegenstand für die Forscher*innen im gesamten Arbeitsprozess verfügbar zu machen. Für diese Verfügbarmachung betreibt zum Beispiel das Kunstgeschichtliche Institut der Goethe-Universität Frankfurt für Lehre und Forschung seit 2009 das quelloffene Bilddatenbanksystem ConedaKOR, welches zunächst aus den Anforderungen der analogen Diathek hervorgegangen ist, dann aber recht schnell auch konzeptionell rein digital gedacht und entwickelt wurde.³ Die Software wird inzwischen in diversen Hochschulen und Forschungsinstitutionen eingesetzt⁴ und auch als „Software as a Service“-Lösung über DARIAH-DE angeboten⁵. ConedaKOR wurde für die Archivierung, Verwaltung und Recherche von Bild- und Metadaten auf einer gemeinsamen webbasierten Oberfläche entwickelt und ist konzeptuell als Graphdatenbanksystem realisiert worden. Der Datengraph stellt sich – wie oben schon beschrieben – als ein Netzwerk aus virtuellen Entitäten und deren Verknüpfungen zueinander dar und ein Graph besteht somit im Wesentlichen aus Knoten und Beziehungen (Kanten), die jeweils zwei der Knoten miteinander verbinden. Im von ConedaKOR realisierten Graphmodell sind diese Knoten und Kanten typisiert, d. h. übergeordneten Typen zugeordnet, über welche die möglichen Verknüpfungen im Graphen eingeschränkt werden können. Durch weitere freie Attribute können die Knoten zusätzlich näher bestimmt werden.

Exemplarisch ist hier ein winziger Teilgraph aus der Datenbank des Kunstgeschichtlichen Instituts Frankfurt dargestellt. Kanten zwischen Knoten sind immer beidseitig navigierbar und für den hier zu sehenden Graphen bedeutet dies, dass der Bamberger Altar von Veit Stoß ein Teil des südlichen Querhauses des Bamberger Domes ist und entsprechend anders herum gelesen, das Querhaus des Doms als übergeordnetes Werk den Bamberger Altar beinhaltet.

²Diese Kombination aus textueller und visueller Information – mit der hohen Komplexität des Bildes – ist natürlich ein ideales Feld für die Deep Learnig-Forschung.

³Coneda UG, mit Angaben zu ConedaKOR: <https://coneda.net/> bzw. <https://github.com/coneda/kor/> (zuletzt aufgerufen am 01.08.2017).

⁴U. a. in Universitäten in Frankfurt am Main, Freiburg, Saarbrücken und Zürich, sowie im Deutschen Forum für Kunstgeschichte Paris.

⁵Siehe dazu <http://dhd-blog.org/?p=7268> (zuletzt aufgerufen am 01.08.2017).

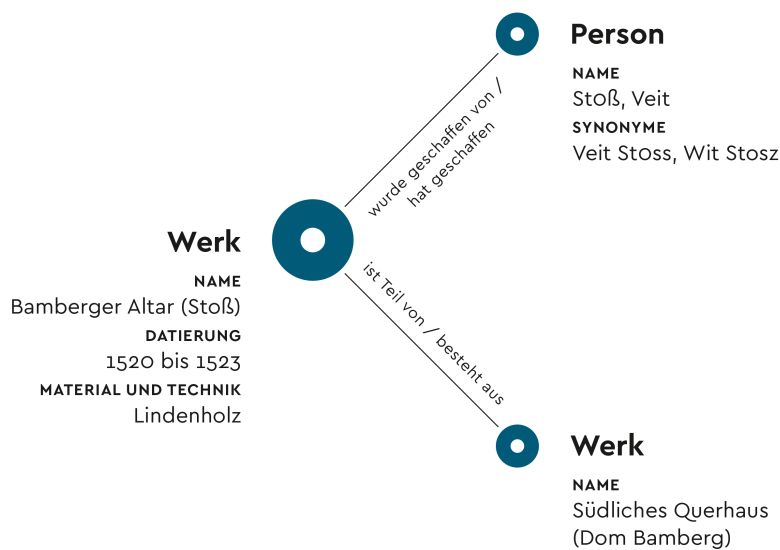


Abbildung 1: Modell eines Ausschnitts aus dem ConedaKOR-Datengraph des Kunstgeschichtlichen Instituts Frankfurt.

Aktuell enthält die Datenbank des Kunstgeschichtlichen Instituts Frankfurt ca. 285.000 Entitäten (Knoten) und 600.000 Beziehungen (Kanten) im Datengraph⁶. Die wichtigsten Entitätstypen repräsentieren Medien, Werke, Personen, Literatur, Institutionen und Orte, die mit einer Vielzahl von Relationen miteinander verbunden werden können.

3 Umsetzung

Bei einem genaueren Blick „unter die Haube“ von ConedaKOR ist festzustellen, dass es sich hier nicht um ein natives Graphdatenbanksystem handelt. Als Storage-Backend dient MySQL und die Ablage der Medien erfolgt im Dateisystem – eine wenig revolutionäre, dafür umso robustere Lösung, was sich in der Nutzung für System- und DatenkuratorInnen als sehr erfreulich darstellt. Die Web-Applikation läuft auf der Basis von Ruby on Rails und Riot.js, so dass insgesamt nur Standard-Komponenten gewartet werden müssen.⁷ Bei Applikationen dieser Art fallen zudem weitere Aufgaben, wie die Verwaltung von Benutzer*innen, Rechten und Sammlungen, an, die mit einem relationalen Datenbankmanagementsystem (RDBMS) im Backend leichter integriert werden konnten. Das Konzept „Graph“ wird in ConedaKOR in erster Linie als ontologisches Ordnungssystem verstanden und fungiert als Benutzerschnittstelle, die nicht nur Daten abfragebasiert anzeigt, sondern den Nutzer*innen auch ermächtigt, direkt in die Modellierung einzugreifen. Die Erstellung der Graphstruktur erfolgt über ein einfach zu bedienendes Web-Interface, welches auch die Eingabe und die Anzeige umsetzt, so dass für die einzelnen Prozesse die Umgebung konstant und kohärent bleibt.

⁶Stand August 2017.

⁷Ein großer Vorteil im universitären Kontext, da in den Hochschulrechenzentren „exotische“ Systemkomponenten ja eher ungern gesehen werden.

Die Erfahrungen stellen mittlerweile das monolithische Konzept einer „Virtuellen Forschungsumgebung“ in Frage⁸ und stattdessen ist eine modulare Systemarchitektur inzwischen State-of-the-Art. Ganz in diesem Sinne ist auch der Ansatz von ConedaKOR zu sehen, welches weitere Komponenten nutzen oder Daten in diverse Umgebungen einspielen kann. So können beispielsweise Teile der Daten in Instanzen von Elasticsearch⁹ und/oder Neo4j¹⁰ migriert und langfristig aktualisiert werden, sodass weitere Funktionalität zur Verfügung steht. Die Rails-Applikation stellt Schnittstellen zur Verfügung, welche dann durch verschiedene „Verbraucher“ genutzt werden können. Denn jenseits des reinen Einsatzes als Datenbanksystem mit eigener Benutzeroberfläche ergeben sich auch Anwendungsfälle, die ConedaKOR als schnittstellenstarkes Werkzeug im Backend in Erscheinung treten lassen. Sei es mit spezifisch gestaltetem Frontend oder als Repositorium, welches Daten über ein Javascript-Widget in bestehende Umgebungen einfließen lässt. Dabei ist es von Vorteil, Inhalte u. a. in statische HTML-Seiten, WordPress-Installationen oder auch z. B. Drupal-Systeme integrieren zu können, ohne die jeweilige Plattform anpassen zu müssen. Über jede Instanz von ConedaKOR kann eine wachsende Bibliothek von Widgets abgerufen werden, welche die direkte Integration von Inhalten in andere Produkte ermöglichen. Nach vorherigem Einfügen der Bibliothek auf der Zielseite lassen sich dort Platzhalter verwenden, die beim Aufruf durch den Browser durch verschiedene Inhalte ersetzt werden. Da hierbei an das Zielsystem keinerlei technische Anforderungen gestellt werden, können hierdurch auch dann Inhalte eingebunden werden, wenn das Zielsystem nicht hinreichend veränderbar ist oder auf einer andersartigen Technologie basiert.¹¹ Auch bietet ConedaKOR insgesamt vier Endpunkte für das Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) an, welche jeweils die Inhalte von Entitätstypen, Entitäten, Verknüpfungstypen und Verknüpfungen abrufbar machen. Die Wiedergabe der Daten erfolgt hier paginiert und auf Wunsch inkrementell, sodass keine zentrierte, hohe Last entsteht. Insgesamt wird hierdurch der gesamte Nutzinhalt der Instanz geordnet abgefragt und dieser Ansatz eignet sich vor allem dann, wenn die Reproduktion oder Zusammenführung der Daten im Vordergrund stehen, also etwa für Backups, Hub-Instanzen oder z. B. auch für die Weitergabe der Daten an Harvester – ein in der Kunstgeschichte prominentes Beispiel: prometheus – Das verteilte digitale Bildarchiv für Forschung und Lehre¹². Das Gegenstück zu diesem OAI-PMH-Repository stellt die Harvester-Komponente von ConedaKOR dar, die im Speziellen dafür Sorge trägt, dass die Abfragen idempotent sind und sich von Instanz zu Instanz transitiv verhalten.

⁸So wiesen dann z. B. Stefan Buddenbohm, Harry Enke, Matthias Hofmann, Jochen Klar, Heike Neuroth und Uwe Schwiigelshohn in *Erfolgskriterien für den Aufbau und nachhaltigen Betrieb Virtueller Forschungsumgebungen* auch zurecht auf folgenden Umstand hin: „Ein werkzeugorientierter Ansatz hat gegenüber einem dem gesamten Forschungsprozess abdeckenden Ansatz Vorteile. Er konzentriert sich auf bestimmte Abschnitte des Forschungsprozesses und ist in der Lage, besser zwischen Community-„Nachfrage“ und Infrastruktur-„Angebot“ zu vermitteln. Er wird auch der Tatsache besser gerecht, dass generische Werkzeuge (bspw. Wiki, Clouddienste) in der Arbeitsumgebung der Forscher in der Regel bereits vorhanden sind, gegen die generische Werkzeuge einer umfassenden VRE vielleicht nur schwer bestehen können.“, Göttingen 2014, S. 30, <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2014-7.pdf> (zuletzt aufgerufen am 02.08.2017).

⁹Suchmaschine (Java) auf Basis von Apache Lucene.

¹⁰Eine in Java implementierte native Graphdatenbank.

¹¹Mittels JSON API können Inhalte in Desktop- und Mobile-Browsern angezeigt werden und über diesen Weg ist die Funktionalität der Anwendung vollständig abgebildet. Bei Vorhaben mit hohem und/oder transparentem Integrationsbedarf, bietet sich die Schnittstelle durch flexible Konfigurationsmöglichkeiten der Cross-Origin Resource Sharing-Header (CORS-Header) an. Als ein Beispiel sei an dieser Stelle nur das Projekt „Travellog“ von Philip Ursprung genannt, <http://travellog.gta.arch.ethz.ch/> (zuletzt aufgerufen am 01.02.2017).

¹²<http://prometheus-bildarchiv.de/> (zuletzt aufgerufen am 01.08.2017).

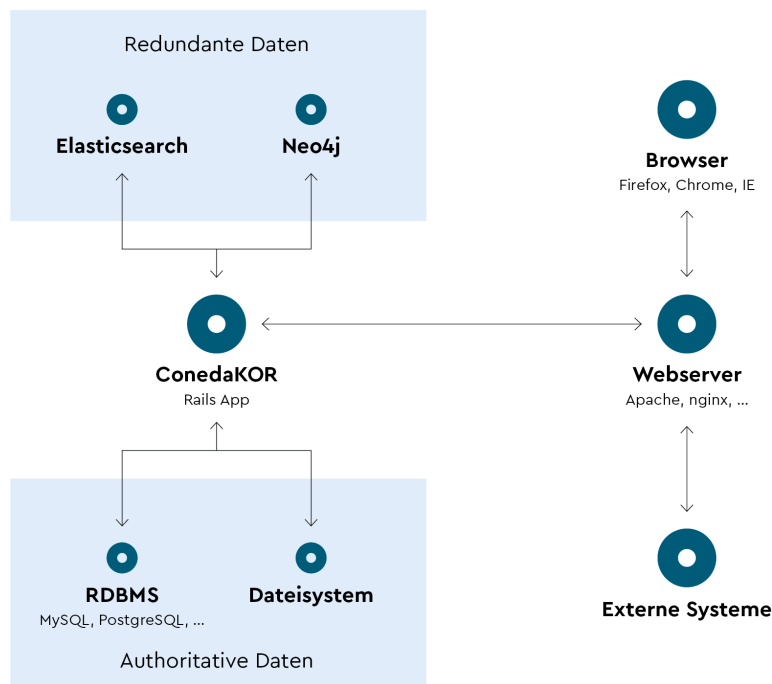


Abbildung 2: Schematische Darstellung der ConedaKOR-Architektur

Im Zuge der Anpassung und Weiterentwicklung von ConedaKOR wurde in jüngster Zeit besonderes Augenmerk auf die Möglichkeiten der Vernetzung der Instanzen unter Berücksichtigung des CIDOC Conceptual Reference Model (CIDOC CRM) gelegt¹³. Hier wurde eine lokale und verteilte semantische Datenintegration mittels des CIDOC CRM als Upper-Level-Ontologie entwickelt, was in der Praxis bedeutet, dass jedes Vorhaben mit einer projektspezifischen Konzeption eigener Daten samt Terminologie die Datenbank nutzen kann und sich damit nicht an ein erst einmal zu erlernendes, globales Datenschema halten muss. Mit den Konzepten des CRM lässt sich eine (in der Regel abstraktere) Entsprechung für die spezifischen Modellierungen der jeweiligen Datenbankprojekte finden, womit die heterogenen Daten integriert werden können. Damit die konkretere Semantik der Elemente der spezifischen Datenbankmodelle nicht auf jene des abstrakten CRM reduziert werden muss, wird das CRM über den vorgesehenen Vererbungsmechanismus um die Elemente des Schemagraphen der spezifischen Datenbankmodelle erweitert. Legt man dann das CRM als standardmäßig integriertes Basisdatenmodell zugrunde, so lässt sich die semantische Datenintegration von der lokalen Datenbank auf ein verteiltes System von Datenbanken erweitern. Auf technischer Ebene wird diese Vernetzung durch die oben bereits erwähnte Implementierung des OAI-PMH realisiert. Das CRM-spezifische Verfahren ist letztlich aber losgelöst von der dahinter stehenden Softwarelösung, die ein Mapping auch Nicht-CRM-basiert durchführen kann.

¹³Siehe hierzu auch Sven Peter, *Abbildung relationaler Daten auf die Ontologie des CIDOC CRM*, 2015, <http://archiv.ub.uni-heidelberg.de/artdok/3454/> (zuletzt aufgerufen am 01.08.2017).

Literaturverzeichnis

Stefan Buddenbohm, Harry Enke, Matthias Hofmann, Jochen Klar, Heike Neuroth, Uwe Schwiegelshohn, *Erfolgskriterien für den Aufbau und nachhaltigen Betrieb Virtueller Forschungsumgebungen*, Göttingen 2014, GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, DARIAH-DE working papers 7, <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2014-7.pdf> (zuletzt aufgerufen am 01.08.2017).

Hubertus Kohle, *Digitale Bildwissenschaft*, Glückstadt 2013.

Sven Peter, *Abbildung relationaler Daten auf die Ontologie des CIDOC CRM*, 2015, <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bsz:16-artdok-34540> (zuletzt aufgerufen am 01.08.2017).