

Die besonderen Herausforderungen multimodaler heterogener Daten- und Quellentypen an die Datenverwaltung. Ist eine Forschungsdateninfrastruktur ohne eine Datenbank umsetzbar?

Gerber, Anja

anja.gerber@bbaw.de

Berlin-Brandenburgische Akademie der Wissenschaften, Deutschland

Das Forschungsvorhaben

In einem Kooperationsprojekt zwischen der Potsdamer Arbeitsstelle des Corpus Vitrearum Medii Aevi (CVMA) Deutschland an der Berlin-Brandenburgischen Akademie der Wissenschaften sowie dem Institut für Kunstwissenschaft der Technischen Universität Berlin wird die digitale Erschließung und Repräsentation der Nikolaikirche in Bad Wilsnack sowie der Quellen zur als "Wilsnacksonfahrt" bekannten Wallfahrt umgesetzt. Hierbei handelt es sich um ein interdisziplinäres Forschungsvorhaben aus den Bereichen Kunstgeschichte, Mediävistik, Digital Humanities (DH) und Informationswissenschaften. Verschiedene Forschende haben unterschiedliche Bedarfe an die Infrastruktur sowie die erzeugten digitalen Informationen und Dateien.



Abbildung 1: Wilsnack, Kirche St. Nikolai, Außenansicht, Blick von Westen, Foto: Berenike Rensinghoff.

Die Kirche St. Nikolai in Bad Wilsnack ist auch als Wunderblutkirche bekannt und war seit Auffindung der drei "Wunderbluthostien" im Jahr 1383 ein zentraler Wallfahrtsort (Bednarz et al. 2010, 87–165; Kühne und Ziesack 2005, 9–18). Zahlreiche Quellen verschiedener Art belegen dies, als Stiftungen in Form von Glasmalereien und Ausstattungsstücken des Bauwerks oder als Überlieferungen wie Pilgerzeichen, Testamente und Ablässe. Schriftquellen liegen teilweise nur gedruckt oder nicht erschlossen in Archiven verschiedener nord- und mittel-europäischer Länder vor und benötigen eine systematische Aufarbeitung. Überlieferungen sind zum Teil nicht sicher belegt und nur durch Verweise auf nicht mehr vorhandene Quellen bekannt. Das Bauwerk war aufgrund der politischen und religiösen Geschehnisse im Laufe der Jahrhunderte verschiedenen baulichen Veränderungen ausgesetzt. Neben umfangreichen Restaurierungsmaßnahmen wurden Ausstattungsstücke verändert, bewegt, entfernt oder gar – insbesondere während der Zeit der Reformation und durch verschiedene Eigentümerwechsel – zerstört (Bednarz et al. 2010, 88–92).



Abbildung 2: Wilsnack, Kirche St. Nikolai, Blick in den Chor mit Hochaltarretabel, Foto: Berenike Rensinghoff.

Forschungsdateninfrastruktur

Im Bereich des digitalen Kulturerbes gibt es nur wenige standardisierte Lösungen für digitale Infrastrukturen, auf die zurückgegriffen werden kann. Bekannte virtuelle Forschungsumgebungen sind MonArch und WissKI. MonArch ist jedoch für die Zwecke des Vorhabens nicht ausreichend, da es hier primär um *Building Information Modeling* für die Erfassung und semantische Annotation bauwerksbezogener Inhalte geht (AriInfoware 2022). WissKI ist eine Forschungsumgebung, die über das CMS Drupal Zugriff auf verschiedene Dateitypen bietet und mittels Datenmodell organisiert ist. Eine Repräsentation von Kartendarstellungen und eine Einbindung von 3D-Viewern ist möglich (Germanisches Nationalmuseum o. J.). Im Bereich des Digitalen Kulturerbes findet vor allem das *CIDOC Conceptual Reference Model* (CRM) für die Datenmodellierung Anwendung, da hiermit objektbezogene und ereigniszentrierte Informationen und ihre Eigenschaften modelliert werden können (Bekiaris et al. 2022). So kann eine Nachnutzbarkeit bereits vorhandener aber auch der eigenen, zum Großteil noch zu erstellenden Daten gewährleistet werden. Eine Besonderheit stellt im Bereich der Kunstgeschichte und des Kulturerbes die Vermittlung von Beziehungen zwischen Informationen dar, die getroffene Aussagen über Objekte und Orte in Zusammenhang mit Akteuren und Ereignissen in einem Kontext setzen (Burrichter et al. 2021, 111f).

Eine Speicherung der Dateien erfolgt in einer durch das Projekt vorgegebenen Ordnerstruktur (Kontinent/Land/Bundesland/Stadt/Gebäude/Datentyp). Verschiedene Tools greifen auf die digitalen Ressourcen zu und die Datenschicht ist von der Software getrennt (Gerber 2022, 16–28). Für ihre Erschließung mit Metadaten aber auch für die Erfassung des Datenmodells wird derzeit der bereits in den beiden Arbeitsstellen des CVMA Deutschland verwendete CVMA Digitaler Ressourcen Manager (CVMA DRM) (Gerber und Fischer 2021) weiterentwickelt. In erster Linie wurde er für die Erschließung von digitalen Bilddateien konzipiert. Die Inhalte werden online über ein webbasiertes GUI auf Basis von Javascript aufgerufen und verarbeitet. Der Zugriff auf die Daten erfolgt über einen Webserver. Metadaten werden in einer der jeweiligen digitalen Ressource zugehörigen maschinen- und menschenlesbaren JSON-Datei gespeichert und perspektivisch in GitLab versioniert. Das zu annotierende Datenformat ist unerheblich, da sich die JSON-Dateien mit jedem Datentyp verknüpfen lassen, neben Bilddateien z. B. auch Textdateien, Daten der Punktwolken oder Koordinaten. Eine Zuordnung erfolgt über den Dateinamen und den persistenten Identifikator (PID). Die Metadateninformationen haben denselben Dateinamen wie die jeweils verschlagwortete Ressource, die Dateiendung .meta und sie werden in derselben Ordnerstruktur wie die durch sie beschriebenen Dateien abgelegt. Durch die Speicherung als JSON-Dateien werden zunächst nur die Metainformationen und stark herunter skalierte Bilder bzw. Vorschauansichten der Modelle geladen. Die großen, hochauflösten Daten sind für eine Anzeige explizit auszuwählen, so dass sich die Ladezeiten erheblich verkürzen. Die Konfiguration der Ansicht und des Metadatenschemas erfolgt ebenfalls über JSON-Dateien.

Datenbasis

Die Quellen der Wallfahrt bilden die Basis für die Datenerfassung und eine daraus erfolgende Repräsentation, z. B. auf Karten oder als 3D-Modell des Bauwerks sowie ausgewählter Ausstattung. Die Forschungsdateninfrastruktur, die aktuell entwickelt wird, muss einerseits mit heterogenen Inhalten der Quellen, andererseits auch mit verschiedenen Dateitypen umgehen. Neben Bilddateien als TIFF oder JPG sowie Regesten der Quellen, die in proprietären Office-Formaten erstellt und in ein menschen- und maschinenlesbares Format wie XML oder JSON konvertiert werden, entstehen ebenfalls 3D-Daten (Punktwolken, digitale Rekonstruktionen) in unterschiedlichen Formaten, z. B. OBJ oder PLY, und Koordinaten als GeoJSON für die Kartendarstellungen. Heterogene Dateitypen und verschiedene Informationsschichten in z. T. ungenauem bzw. ungesichertem Überlieferungsstatus werden zu einem multimodalen Datenmodell zusammengeführt. Das Bauwerk betreffende Informationen werden mit Ereignissen in Verbindung gebracht, um zu erfassen und abzufragen, was bspw. Akteure der damaligen Zeit gesehen haben könnten, als sie die Kirche St. Nikolai zu einem bestimmten Zeitpunkt betraten, welche Ausstattungsstücke wann sichtbar waren oder welche Routen Pilger zurückgelegt haben. Die digitalen Ressourcen werden auf einer von der Software getrennten Dateiebene wiederum mit Metadaten angereichert.

Abbildung 3: Ansicht Graphical User Interface (GUI) des CVMA Digitaler Ressourcen Manager für die Erschließung der Forschungsobjekte, Anja Gerber 2022.

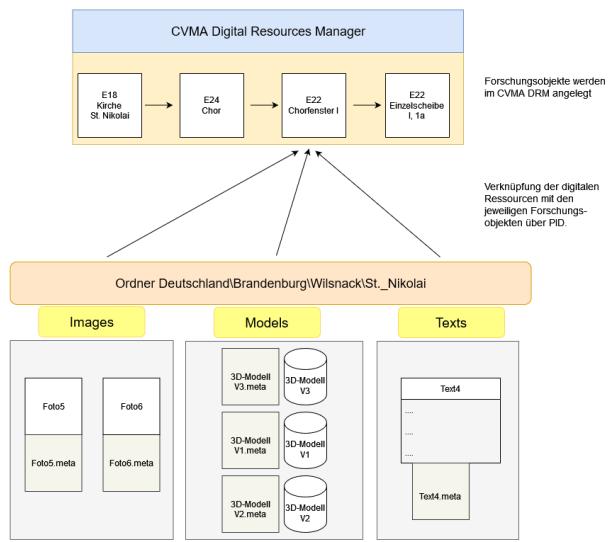


Abbildung 4: Verknüpfung der Forschungsobjekte mit Ressourcen und Metadatendateien. Anja Corber 2022

Die erschlossenen Inhalte der Quellen werden durch das Datenmodell organisiert und mittels verschiedener Viewer, u. a. Kompakkt für 3D-Modelle (Universität zu Köln 2018–2022), Digilib für Bilder (digilib Community 2001–2022) und das Chronotopische Tool für Kartendarstellungen (Fischer und Thomas 2021), repräsentiert. In diese Viewer können zudem weitere digitale Ressourcen, z. B. Bilder oder beschreibende Texte, eingebunden und mit den dort dargestellten Daten angezeigt, sowie die Inhalte mit zusätzlichen Informationen außerhalb der Metadaten annotiert werden. Ein einzelnes System müsste aufwendig angepasst oder es muss auf verschiedene Datenbanken zurückgegriffen werden, um die verschiedenen Dateitypen speichern zu können. Daher ist die Entwicklung einer Weboberfläche geplant, über die der Ressourcenmanager zur Erfassung und Recherche, aber auch die Tools zur Anzeige der Präsentationsschicht eingebunden und angezeigt werden. Bisher erfolgt ein Aufruf der jeweiligen Tools über verschiedene Links. Sie befinden sich bereits auf demselben Webserver und greifen auf denselben Datenbestand zu. Für die Requesten und

Texte der Quellen muss zum Zeitpunkt des Abstracts noch eine Lösung gefunden werden.

Datenmodellierung

Über das Datenmodell werden die sehr heterogenen Informationen und Dateien inhaltlich organisiert und mittels automatisiert vergebener PID miteinander verknüpft. Dieser Prozess erfolgt manuell. Aus den Quellentexten entnommene Angaben zu Akteuren, Datierungen, Ereignissen, Objekten und Orten erhalten konkrete Bezeichnungen und bilden die Gruppe der Forschungsobjekte. Diese werden dann verschiedenen Kategorien des Datenmodells zugewiesen, z. B. St. Nikolai als Bauwerk der Klasse *E18 Physical Object*, Wilsack als Stadt der Klasse *E53 Place* oder Bischöfe und Pilger als Personen der Klasse *E39 Actor*. Die Quelle liegt als digitale Ressource vor und wird mit Metadaten angereichert sowie den entsprechenden Forschungsobjekten zugewiesen. Der Quellentyp, wie Ablass, Testament, Urkunde, wird zusammen mit ihrem Fundort in den Metadaten erfasst. Es ist nicht ausreichend, das Datenmodell nur am Bauwerk oder an geographischen Informationen auszurichten, da heterogene Daten miteinander in Verbindung gebracht werden müssen und Informationen zu Bauwerk und Ereignissen beinhalten.

Derzeit orientiert sich das Modell an der Basisontologie von CIDOC CRM. Im Laufe der folgenden Projektphase wird in verschiedenen Unterontologien geprüft, welche Klassen und Eigenschaften zutreffend sein könnten, um bestimmte Angaben zu präzisieren. Hier erscheinen insbesondere CIDOC CRMba für archäologische Bauwerke (Ronzino et al. 2016), CRMdig für Herkunftsmetadaten (Doerr, Stead und Theodoridou 2016) sowie CRMgeo für räumlich-zeitliche Daten (Hiebel et al. 2015) als geeignet. Die Kategorien für die Erfassung der Forschungsobjekte entsprechen zu diesem Zeitpunkt hauptsächlich den *High Level Classes*, für die Hierarchisierung des Bauwerks werden bereits untergeordnete Klassen verwendet. Die Modellierung der Beziehungen zwischen den Forschungsobjekten dient deren Kontextualisierung und erfolgt unter Verwendung der den jeweiligen Klassen zugewiesenen Eigenschaften. Aus Gründen der Datenlogik lässt sich nicht jede Klasse mit jeder Eigenschaft verbinden. Eine feinere Granulierung erfolgt während des Fortschreitens des Projekts, da die Entwicklung des Datenmodells im Dialog mit den Erkenntnissen der Forschung erfolgt. Eine zu grobe oder kleinteilige Erfassung führt zu ungenauen oder zu geringen Informationen, die miteinander in Verbindung gebracht werden. Anpassungen werden im Projekt dokumentiert. Es besteht ebenfalls die Möglichkeit, ab Herbst 2022 das in der Weiterentwicklung befindliche Datenmodell für die semantische Annotation von 3D Artefakten der NFDI4Culture (Rossenova 2021) zu testen und für das Wilsnackprojekt anzupassen.

Die sehr heterogenen Forschungsobjekte erhalten bereits während des Erfassungsprozesses konkrete Bezeichnungen wie "Kirche St. Nikolai", "Chor", "Chorfenster I", "Hostienwunder", "1383", "Hostienverbrennung", "Johann Ellefeldt", "1552", u. s. w., so dass eine Zuordnung eindeutig ist. Nicht jedes Ausstattungsstück oder jeder Akteur der Zeit werden erfasst. Gemeinsam in der Projektgruppe erfolgt eine Priorisierung. Gleich benannte

Forschungsobjekte, wie Säulen, nicht näher bezeichnete Altäre oder die Wunderbluthostien, werden nummeriert. Eine Dokumentation erfolgt in den Metadaten der zugehörigen digitalen Ressourcen und im Datenmodell. Das Bauwerk wird durch die Kategorien Bauwerk – Gebäudeteile – immobile und mobile Ausstattung hierarchisiert. Den Forschungsobjekten werden alle zugehörigen Informationen und Dateien – egal welchen Typs – zugewiesen, z. B. werden alle Fotografien, Bauberichte und entstehenden Modelle des Chors mit dem Forschungsobjekt "Chor" verknüpft. Jede erfasste Information, egal ob Forschungsobjekt oder Datei, erhält einen PID, so dass eine Eindeutigkeit gewährleistet ist. Die Zuweisung erfolgt händisch und kann korrigiert werden. Einzelscheiben der Montagen oder des Gesamtfensters können gruppiert werden, wenn sie dieselben Bewegungen im Raum durchlaufen. Eine Erfassung erfolgt jedoch für jedes Forschungsobjekt einzeln, da sie verschiedene Ereignisse durchlaufen haben und noch können.



Abbildung 5: Wilsnack, Kirche St. Nikolai, Nördliches Querhaus, Fenster n VIII, CVMA Deutschland Potsdam/Berlin-Brandenburgische Akademie der Wissenschaften, Foto: Holger Kupfer.

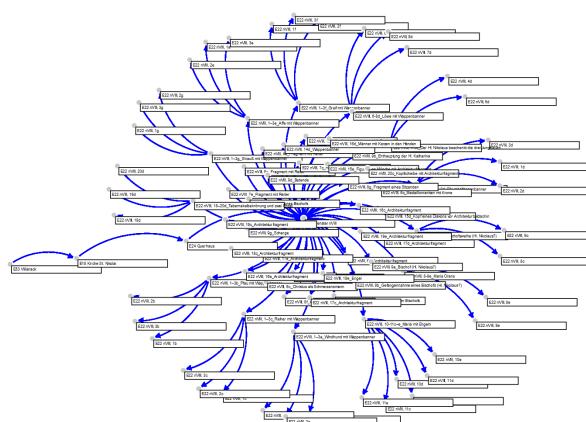


Abbildung 6: Wilsnack, Kirche St. Nikolai, Modellierung des Fenster n VIII anhand CIDOC CRM, Visualisierung: Gordon Fischer / Daten: Anja Gerber 2022.

Erschließung der Daten

Eine Annotation von digitalen Ressourcen mit Metadaten erfolgt auf der Dateiebene mit einer projekteigenen und sich in der Entwicklung befindlichen Spezifikation, die sich aus Gründen der Nachnutzbarkeit an bekannten und etablierten Standards orientiert. Der Erstentwurf verwendet das Austauschschemata LIDO 1.1 (ICOM-CIDOC LIDO Working Group 2021), welches sich u. a. an CIDOC CRM, CDWA Lite (J. Paul Getty Trust 2006) und Spectrum (Collections Trust 2017) orientiert, sowie Elemente aus Dublin Core (DCMI 1995–2022, DCMI 2020). Bei LIDO handelt es sich um einen etablierten Standard für die (Meta-)Datenerfassung im Bereich des Digitalen Kulturerbes (Knaus, Stein und Kailus 2019, 12–15). Dieses Schema ist neben der Erfassung von Grafiken und Fotografien auch für Ausstattungsstücke geeignet (Knaus, Kailus und Stein 2022, 18–21). Die Spezifikation des Projekts wird im Hinblick auf die Annotation der zu erstellenden 3D-Modelle um für diese erforderliche Angaben erweitert. Hier erscheint der Standard CARARE 2.0 als geeignet, da technische Angaben (digitale Herkunftsmetadata) und beschreibende Informationen für Gebäude, Artefakte und born-digital Objekte möglich sind (Fernie, Gavrilis und Angelis o. J.). Mit LIDO können Metadaten desselben Typs unter Zuweisung verschiedener Kategorien und Rollen erfasst werden, z. B. ob jemand an der Erstellung einer Ressource beteiligt war oder lediglich auf dieser abgebildet ist. Administrative, beschreibende und technische Metadaten werden abgedeckt. So können die Datei selbst aber auch ihr Erstellungsprozess beschrieben werden. Die Konfiguration des Metadatenschemas erfolgt ebenfalls über eine JSON-Datei. Anpassungen werden dokumentiert und versioniert. Um eine eindeutige Zuordnung zu Akteuren, Ereignissen, Orten oder auch eine zweifelsfreie Klassifizierung der Objekte zu ermöglichen, werden bereits etablierte Normdatenvokabulare, wie GND, Wikidata, Iconclass, Getty Art & Architecture Thesaurus oder Geonames, genutzt und die Identifier erfasst. Zudem gibt es Überlegungen, Normdaten im Rahmen der NFDI4Culture über Wikibase (Rosse-

nova u. a. 2021) zu erfassen und als Linked Open Data in den Wissensgraphen (NFDI4Culture o. J.) einzuspeisen.

Bibliographie

- AriInfoWare GmbH.** "Aktuelle Informationen – MonArch". In *MonArch*. 2022. Zugegriffen 28. Juli 2022. <https://openmonarch.org/informationen/>.
- Bednarz, Ute, Eva Fitz, Peter Knüvener, Frank Martin, Markus Mock, Götz J. Pfeiffer und Martina Voigt.** "Die mittelalterlichen Glasmalereien in Berlin und Brandenburg." *Corpus Vitrearum Medii Aevi Deutschland XXII*. Berlin: Akademieverlag, 2010.
- Bekari, Chryssoula, George Bruseker, Erin Canning, Martin Doerr, Philippe Michon, Christian-Emil Ore, Stephen Stead, und Athanasios Velios.** "Volume A: Definition of the CIDOC Conceptual Reference Model". CIDOC CRM Special Interest Group. Juni 2022. <https://cidoc-crm.org/versions-of-the-cidoc-crm>.
- Burrichter, Brigitte, Björn Gebert, Christoph Mackert, und Gabriel Viehhäuser.** "Digitale Mediävistik". *Das Mittelalter. Perspektiven mediävistischer Forschung* 26, Nr. 1, (2021): 101-117. <https://doi.org/10.17885/heiup.mial.2021.1.24312>.
- Collections Trust.** "Spectrum 5.0. All Procedures." 2017. Zugegriffen 28. Juli 2022. <https://collectionstrust.org.uk/spectrum/procedures/>.
- digilib Community.** "digilib - The Digital Image Library". 2001-2022. Zugegriffen 28. Juli 2022. <https://robcast.git-hub.io/digilib/>.
- Doerr, Martin, Stephen Stead und Maria Theodoridou.** "Definition of the CRMdig. An Extension of CIDOC-CRM to support provenance metadata." Proposal for approval by CIDOC CRM-SIG. Version 3.2.1. April 2016. https://cidoc-crm.org/crmdig/sites/default/files/CRMdig_v3.2.1.pdf.
- Dublin Core Metadata Initiative (DCMI).** "Specifications." 1995–2022. Zugegriffen 28. Juli 2022. <http://dublincore.org/specifications/dublin-core/>.
- Dublin Core Metadata Initiative (DCMI).** "DCMI Metadata Terms." 2020. Zugegriffen 28. Juli 2022. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>.
- Fernie, Kate, Dimitris Gavrilis, und Stavros Angelis.** "The CARARE Metadata Schema, v.2.0". o. J. Zugegriffen 28. Juli 2022. <http://3dicons-project.eu/wp-content/uploads/2018/08/The-CARARE-metadata-schema2.pdf>.
- Fischer, Gordon und Christian Thomas.** "Alexander von Humboldt auf Reisen: Chronotopische Zugänge zur edition humboldt digital." *vDhd2021 – Experimente*. 2021. Zugegriffen 28. Juli 2022. <https://vdhd2021.hypotheses.org/292>.
- Germanisches Nationalmuseum.** „Features I wiss-ki.eu“. *WissKI - a scientific communication infrastructure*. o. J. Zugegriffen 28. Juli 2022. <https://wiss-ki.eu/features>.
- Gerber, Anja.** "Forschungsdateninfrastruktur für multimodale digitale Daten- und Quellentypen am Beispiel des Standortes Wilsnack, St. Nikolai." Masterarbeit, Fachhochschule Potsdam und Humboldt Universität zu Berlin, 2022.
- Gerber, Anja und Gordon Fischer. "CVMA Foto Manager - ein Open-Source-Metadateneditor für die Erschließung von Bilddaten." *NFDI4Culture Community Plenary*. 17.-19.11.2021. Zugegriffen 28. Juli 2022. <https://nfdi4culture.de/de/aktuelles/nachrichten/second-culture-community-plenary-abstracts-of-lightning-talks-and-presentations.html>.
- ICOM-CIDOC LIDO Working Group.** "LIDO - Lightweight Information Describing Objects Version 1.1." *LIDO v1.1 Documentation*. 20. Dezember 2021. <https://lido-schema.org/schema/v1.1/lido-v1.1.html>.
- Institut für Museumsforschung der Staatlichen Museen zu Berlin – Preußischer Kulturbesitz.** "SPEC-TRUM 3.1. The UK Museum Documentation Standard, deutsche erweiterte Fassung." *Materialien aus dem Institut für Museumsforschung – Sonderheft 5*. Berlin, 2013. https://www.smb.museum/fileadmin/website/Institute/Institut_fuer_Museumsforschung/Publikationen/Materialien/Sonderhefte/materialien/Sonderheft_5-SPECTRUM_3_1.pdf.
- J. Paul Getty Trust.** "CDWA Lite. Specification for an XML Schema for Contributing Records via the OAI Harvesting Protocol 1.1". 2006. Zugegriffen 28. Juli 2022. https://www.getty.edu/research/publications/electronic_publications/cdwa/cdwalite.pdf.
- Hiebel, Gerald, Martin Doerr, Øyvind Eide, und Maria Theodoridou.** "CRMgeo: a Spatiotemporal Model. An Extension of CIDOC-CRM to link the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement." Proposal for approval by CIDOC CRM-SIG. Version 1.2. September 2015. https://www.cidoc-crm.org/crmgeo/sites/default/files/CRMgeo1_2.pdf.
- Knaus, Gudrun, Regine Stein und Angela Kailus.** "LIDO-Handbuch für die Erfassung und Publikation von Metadaten zu kulturellen Objekten. Band 1: Graphik." Heidelberg: arthistoricum.net, 2019. <https://doi.org/10.11588/arthistoricum.382.544>.
- Knaus, Gudrun, Angela Kailus und Regine Stein.** "LIDO-Handbuch für die Erfassung und Publikation von Metadaten zu kulturellen Objekten. Band 2: Malerei und Skulptur". Heidelberg: arthistoricum.net, 2022. <https://doi.org/10.11588/arthistoricum.1026>.
- Kühne, Hartmut und Anne-Katrin Ziesak.** "Wunder – Wallfahrt – Widersacher. Die Wilsnackfahrt". Regensburg: Pustet Verlag, 2005.
- NFDI4Culture.** "Aufgabenbereich 5. Technologien, Recht & Ethik. Übergreifende technische, ethische und rechtliche Aktivitäten". o. J. Zugegriffen 28. Juli 2022. <https://nfdi4culture.de/de/aufgaben/aufgabenbereiche/aufgabenbereich-5.html>.
- Ronzino, Paola, Franco Niccolucci, Achille Felicetti, und Martin Doerr.** "Definition of the CRMba. An extension of CIDOC CRM to support buildings archaeology documentation." Proposal for approval by CIDOC CRM-SIG. Editorial Status: Under Revision since 3.12.2016. Version 1.4. Dezember 2016. https://cidoc-crm.org/crmba/sites/default/files/2016-12-3%23CRMba_v1.4.1_LUR.pdf.
- Rossenova, Lozana.** "Semantic Annotation for 3D Cultural Artefacts: MVP." 29. Oktober 2021. <https://doi.org/10.5281/zenodo.5628847>.
- Rossenova, Lozana, Zoe Schubert, Richard Vock, und Ina Blümel.** "Beyond the render silo - Semanti-

cally annotating 3D data within an integrated knowledge graph and 3D-rendering toolchain". Potsdam, 7. März 2022. <https://doi.org/10.5281/zenodo.6328155>.

Universität zu Köln, Department of Digital Humanities. "Kompakkt - 'cause the world is multidimensional." 2018-2022. Zugegriffen 28. Juli 2022. <https://kompakkt.de/home>.