

Beitrag aus:

Sonderband 4 der ZfdG: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. 2019. DOI: [10.17175/sb04](https://doi.org/10.17175/sb04)

Titel:

Ambiguität und Unsicherheit: Drei Ebenen eines Datenmodells

Autor/in:

Andreas Wagner

Kontakt:

andreas.wagner@adwmainz.de

Institution:

Akademie der Wissenschaften und der Literatur, Mainz

GND:

136680607

ORCID:

0000-0003-1835-1653

DOI des Artikels:

[10.17175/sb004_007](https://doi.org/10.17175/sb004_007)

Nachweis im OPAC der Herzog August Bibliothek:

[1037073762](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0073-7)

Erstveröffentlichung:

30.01.2019

Lizenz:

Sofern nicht anders angegeben 

Medienlizenzen:

Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:

30.01.2019

GND-Verschlagwortung:

[Digital Humanities](#) | [Datenmodell](#) | [Graphdatenbank](#) | [Kirchenrecht 1567-1681](#) | [Referenzmodell](#) | [Ungewissheit](#) |

Zitierweise:

Andreas Wagner: Ambiguität und Unsicherheit: Drei Ebenen eines Datenmodells. In: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. 2019 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 4). PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/sb004_007](https://doi.org/10.17175/sb004_007).

Andreas Wagner

Ambiguität und Unsicherheit: Drei Ebenen eines Datenmodells

Abstracts

Dieser Beitrag stellt anhand von Forschungen im Rahmen der Max-Planck-Nachwuchsgruppe ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ einen Ansatz vor, Phänomene von Unsicherheit in geisteswissenschaftlichen, genauer: rechtshistorischen Zusammenhängen in einem Datenmodell abzubilden. Die Kernthese besteht im Vorschlag einer Modellierung des Forschungszusammenhangs auf drei Ebenen, welche (a) die historischen Phänomene, (b) die überkommenen Zeugnisse dieser Phänomene und (c) die aktuelle historische Forschung selbst beschreiben. Während Ambiguitäten und Unklarheiten zwar auf allen dreien dieser Ebenen entstehen, werden sie allein auf der dritten Ebene modelliert, denn sie können nicht von der Beobachtung in der Forschung getrennt werden.¹ Im Folgenden wird das Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ im Hinblick insbesondere auf Unsicherheiten in der Datenmodellierung vorgestellt (1.). Im Anschluss daran wird ein vorläufiges, bisweilen noch informelles, Datenmodell beschrieben (2.) und schließlich ein Ausblick vorgenommen, der die weitere Entwicklung des Modells ebenso wie die Herausforderungen im Projektkontext, v.a. in der Dateneingabe und -prozessierung betrachtet (3.).

Using a specific research project, this article presents an approach for displaying phenomena of uncertainty in humanities contexts – here, legal history – in a data model. In the proposal for modeling a research context, the core thesis depends on three levels that describe a) the historical phenomena, b) the traditional evidence for this phenomena, and c) the latest historical research. Although ambiguities and uncertainties will be present on all three levels, they will be modeled on the third level only since they can not be separated from the historical research. The following paper presents the research project ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ with regard to uncertainties in data modeling (1.). The preliminary, at times unofficial, data model will then be described (2.); finally, this article will discuss future prospects for the further development of the model as well as challenges within the context of the project, such as in data entry and processing (3.).

1. Der Projektzusammenhang und unsichere Informationen

1.1 Ein rechtshistorisches Forschungsprojekt

Das Konzil von Trient (1545–1563) gilt als wichtiger Meilenstein in der doktrinären und institutionellen Modernisierung der katholischen Kirche. Mit den Konzilsbeschlüssen reagierte diese auf theologische, juristische und moralische Herausforderungen, die sich unter anderem im Kontext der Reformation, durch ihre eigene Rolle in einer nun globalisierten politischen Landschaft und durch eine virulent gewordene Pluralität und Dynamik sozialer, kultureller

¹ Um genau zu sein, werden auch Informationen, die im Forschungsprozess den ersten beiden Ebenen zugeschrieben werden, zwar eben dort festgehalten, aber mit einer Information auf der dritten Ebene verknüpft, die eben die Unsicherheit der Zuschreibung festhält. Eine unleserliche Abkürzung ist in diesem Sinne zwar ein Phänomen, das in der Quelle selbst verortet ist, aber *unleserlich* ist sie eben immer nur *für eine Leserin* oder *einen Leser*.

und politischer Verhältnisse ergeben hatten. Dies erforderte aber auch handlungsfähige und zugleich flexible, institutionelle Mechanismen, durch die diese Beschlüsse auf die vielfältigen Handlungskontexte und die jeweils anfallenden Probleme anzuwenden oder mit ihnen in Dialog zu bringen wären. Mit dem Auftrag einer autoritativen Interpretation und je zu aktualisierenden Umsetzung wurde 1567 die Konzilskongregation eingerichtet, deren Beratungsprotokolle bis heute im Vatikanischen Geheimarchiv lagern. Dieses Gremium nahm Petitionen und Fragen zur Auslegung der Konzilsbeschlüsse entgegen, die aus kirchlichen Einrichtungen und von Funktionsträgern aus der ganzen Welt nach Rom gesandt wurden. Die Fragen wurden in ggf. mehreren Sitzungsterminen beraten, weitere Informationen wurden eingeholt bzw. Stellungnahmen Dritter wurden eingereicht, bevor eine Entscheidung getroffen und in den Beschluss-Protokollen festgehalten wurde.²

Um die Weise zu untersuchen, in der das Recht als Instrument des beständigen Abgleichs zwischen der universalen Institution der Kirche und den jeweiligen lokalen, konkreten Regulierungsbedarfen eingesetzt wurde oder seine Wirkung entfaltete, wurde 2013 unter der Leitung von Benedetta Albani die Max-Planck-Forschungsgruppe ›**Die Regierung der Universalkirche nach dem Konzil von Trient**‹ am Max-Planck-Institut für europäische Rechtsgeschichte in Frankfurt/Main eingerichtet. Durch verschiedene Einzelfragen erschließen und beforschen die Mitglieder der Gruppe die Akten der Konzilskongregation — eine Datenmenge, die kaum ohne die Hilfe digitaler Methoden intellektuell durchdrungen werden kann:

- Die Forschung betrifft einen Zeitraum von 114 Jahren (1567–1681) und 17 Pontifikaten,
- an den Beratungen waren über 100 Kardinäle beteiligt,
- die Eingaben stammen aus ca. 750 weltweit verteilten Diözesen.
- Die Beratungsprotokolle der Konzilskongregation umfassen 270 Bände Archivmaterialien,
- und sie geben über 33.000 Entscheidungen wieder.

Jede einzelne dieser Entscheidungen wird durch Datenfelder wie z.B. das Datum der Sitzung, den Petenten, den infrage stehenden Konzilsbeschluss, die beteiligten Kardinäle und Sekretäre und ggf. den Beschluss der Kongregation beschrieben –nach Erfahrungen der Projektgruppe, die in der einen oder anderen Form bereits seit fünf Jahren an diesem Datenbestand arbeitet, wurden über 100 mögliche Datenfelder identifiziert.³

1.2 Unsicherheiten, Unklarheiten und Ambiguitäten

Die erste Quelle von Zweifeln bezüglich der Informationen, mit denen im Projekt gearbeitet wird, liegt auf der Hand: Denn wie es in solchen Projekten zumeist der Fall ist, ist nicht für jede Entscheidung jede Information in den Protokollen festgehalten. Die formalen Anforderungen

² Vgl. Albani 2009.

³ Diese Zahl beinhaltet allerdings auch in gewisser Weise redundante Felder wie z.B. durch die Projektgruppe zu recherchierende Normdaten-Identifikatoren für die relevanten Personen, eine Standard-Ansetzungsform und die jeweilige Schreibung bzw. Abkürzung ihres Namens im erfassten Protokoll.

sorgen zwar dafür, dass z.B. das Datum der Sitzung in allen Fällen aufgeschrieben wurde, so dass dieses Feld – von Beschädigungen der Archivalien oder unklaren Formulierungen abgesehen – immer befüllt werden kann. Der einschlägige Konzilsbeschluss oder gar formelle wie informelle Einflussnahmen Dritter sind jedoch nicht in allen Fällen dokumentiert und wir haben einen unvollständigen und insofern inkonsistenten Datenbestand. Weitere Probleme ergeben sich durch die Manuskript-Charakteristik der Materials: So liegen häufig unleserliche Schreibungen und – vor allem bei Personennamen – Abkürzungen vor, die z.T. mehrdeutig sind. In anderen Fällen werden Personen eher durch ihre Rolle oder Funktion bezeichnet. Spätestens in diesen Fällen kann eine Differenzierung vorgenommen werden zwischen lückenhafter oder unleserlicher Vorlage auf der einen Seite und unsicherem, bloß wahrscheinlichen Wissen der heutigen Leserinnen auf der anderen. Denn in einem Ausdruck wie »der Domkapitular von Assisi« kann einerseits z.B. der Ortsname unleserlich sein, wodurch nicht klar ist, von welcher Person die Rede ist, es kann aber andererseits auch der Ausdruck ohne weiteres lesbar, jedoch unser Wissen darüber, wer zur betreffenden Zeit diese Rolle ausfüllte, unsicher sein.

Es ist aber noch auf eine dritte Quelle von Unsicherheiten hinzuweisen, die gar nicht direkt mit dem archivalischen Quellenmaterial oder den aktuellen Forschungsprozessen zusammen hängt: Denn schon diesseits ihrer Dokumentation sind in den historischen Phänomenen selbst diffuse Phänomene wichtiger Teil des Praxiszusammenhangs. Schon für einen Teilnehmer der Kongregationsberatungen selbst wäre in vielen Fällen nicht zweifelsfrei entscheidbar, ob z.B. eine dritte Partei Einfluss ausgeübt hat. Und dies nicht erst wegen des unvollkommenen Wissens dieses vorgestellten Teilnehmers, sondern schon wegen der Vagheit der Kategorie der ›Einflussnahme‹. Gewiss ist das Projekt darauf angelegt, formelle und insofern objektivierbare und eindeutige Eigenschaften und Prozesse zu erfassen. Allerdings ist erstens nicht *a priori* überschaubar, ob es sich bei einzelnen zu erfassenden Datenfeldern nicht doch um Informationen handeln könnte, die erst *ex post* vereindeutigt wurden (und sei es durch die Selbst-Interpretation der Teilnehmer) und die in den Dokumenten lediglich präzise scheinen, in der Praxis aber diffuser waren. Und zweitens kann es sich im Zuge der Überlegungen zu einer möglichen Verallgemeinerbarkeit unseres Datenmodells durchaus als produktiv erweisen, über diese Fragen nachzudenken.

Diffuse historische Phänomene, lückenhaftes oder unleserliches Quellenmaterial und unsicheres Kontextwissen der heutigen Forscherinnen sind somit drei Ebenen, auf denen mitunter Zweifel und selbstkritische Vorbehalte bezüglich der Informationen, mit denen im Projekt gearbeitet wird, angebracht sind. Offenkundig würde der wissenschaftliche Prozess ganz wesentlich verarmt, wenn man Informationen, die in einer der genannten Hinsichten fraglich wären, überhaupt nicht mehr berücksichtigen würde. Somit muss es im Rahmen guter wissenschaftlicher Praxis auch in einer solchen Datenbank darum gehen, die entsprechenden Unsicherheiten möglichst transparent und nachvollziehbar zu transportieren und zu kommunizieren, d.h. sie müssen flexibel und genau dokumentiert und eng an die durch sie beschriebenen Daten angelagert werden.⁴

⁴ Hier unterscheiden sich die Anforderungen an Transparenz und systematischer Selbstkritik nicht von denen, die etwa an wissenschaftliche Editionen von Texten gerichtet werden. Vgl. etwa Clement 2011.

In der bisherigen Projektarbeit erfassen die Forscherinnen Informationen aus den Archivmaterialien in einer Office-Tabellenkalkulation und markieren Ergänzungen und Vermutungen wie üblich typographisch, z.B. durch Kursivierungen, geschweifte oder eckige Klammern (>[...]<) usw. (Dazu gibt es ein im Projekt erstelltes Richtliniendokument, in dem aufgeschlüsselt wird, welche typographische Hervorhebung in welchem Datenfeld was zum Ausdruck bringt.). Über die Notiz hinaus, dass eine bestimmte Information durch die heutigen Forscherinnen in das Material getragen wurde, befassen wir uns mit Möglichkeiten, die Gewissheit oder Zuversicht der Forscherin bezüglich der Richtigkeit dieser Eintragung festzuhalten, eventuell neben einer Bemerkung dazu, auf welche (>externen<) Quellen sie sich dabei stützt. Das hier vorgestellte Datenmodell räumt diese Möglichkeiten ein, es wird aber auch kritisch zu überprüfen sein, ob sie in praktikabler Weise in den Arbeitsprozess integriert werden können.

2. Die drei Ebenen des Datenmodells

Aktuell wird das Datenmodell in einem iterativen Verfahren weiterentwickelt, das sich einerseits (kurz- bzw. mittelfristig) an einer flexiblen und explorativen Beforschung des Materials und andererseits (langfristig) an einer interoperablen Veröffentlichung der Daten orientiert. Konkret bedeutet das eine konzeptuelle Orientierung am CIDOC-CRM-Standard⁵ und zugleich eine umgehende Implementierung und kontinuierliche Anpassung in einer Property-Graph-Datenbank (*Neo4j*). Die konsequente Konversion des gesamten Datenbestandes in ein CIDOC-CRM-kompatibles Format und ihre Publikation als RDF / Linked Open Data werden zu einem späteren Zeitpunkt erfolgen, während bereits jetzt projektintern mit den Werkzeugen der Graphdatenbank gearbeitet wird. So können, auch während kontinuierlich noch weiter Daten erfasst werden, bereits im aktuellen Stand der Datenbank z.B. über einfache Graphanalysen Dubletten identifiziert oder Hypothesen über Problemkonjunkturen oder Personennetzwerke gebildet werden. Eine weitere Herausforderung im Prozess der Datenmodellierung besteht darin, den Besonderheiten spezifisch juridischer Phänomene durch die Integration von Einsichten aus historischen Projekten (z.B. *Trials of the Late Roman Republic*⁶) oder von Erfahrungen und etablierten Standards in der digitalen Erfassung geltenden Rechts und juridischer Deliberationen (z.B. *LKIF-core*⁷) Rechnung zu tragen.

2.1. Historische Vorgänge: Der Fall

In den Akten sind die Aufzeichnungen zuerst nach Beratungsterminen organisiert und werden in den ersten Schritten der Datenerfassung weiter aufgegliedert, so dass in der vorläufigen Datenerfassung durch die Forscherinnen in einer Office-Tabellenkalkulation einzelne Zeilen einzelnen Gesprächs->Runden< an diesen Terminen entsprechen (vgl. Abbildung 1).

⁵ vgl. Ore et al. 2017.

⁶ vgl. Sperberg-McQueen 2016.

⁷ vgl. Hoekstra et al. 2008.

Abb. 1: Datenerfassung im Spreadsheet. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.

Allerdings beziehen sich Erkenntnisinteresse und wissenschaftliche Fragestellung in erster Linie nicht auf diese Archivalien oder ihre Organisation selbst, sondern auf die in ihnen dokumentierten historischen Prozesse. Van Ruymbeke et al. schlagen eine Differenzierung zwischen der ›historischen‹ oder ›phänomenalen‹ und der ›informationellen‹ Dimension vor, also zwischen den beobachtbaren Phänomenen und dem Diskurs darüber.⁸ In ähnlicher Weise konzentrieren wir uns auf dieser ersten Ebene zunächst auf die historischen Phänomene: auf den oder die Petenten (in der Abbildung orange hervorgehoben), auf die Diözese, aus der die Eingabe erfolgt ist oder die von dem Problem betroffen ist (gelb), auf die in dieser Angelegenheit interpretationsbedürftigen Beschlüsse des Trienter Konzils (grün), sowie auf die an der Beratung beteiligten Kardinäle (dunkelblau), und schließlich auf das Datum (hellblau) und die Kurzfassung ihres Beschlusses (rot).

So ist der hinter der Erfassung und eben auch schon hinter dem Archiveintrag stehende historische (und genuin juristische) Zusammenhang eines ›Falls‹ die zentrale Entität der Modellierung des historischen Phänomens, ihm sind die Besprechungstermine dann erst zugeordnet. Diese Intuition wird dadurch bestätigt, dass die Protokolle durchaus innerhalb der Beratungssitzungen eine Gliederung nach den verschiedenen am jeweiligen Termin behandelten Fällen aufweisen, und sie wird weiter bekräftigt durch andere Projekte und Forschungen, die vergleichbare gerichtliche oder gerichts-analoge Vorgänge modellieren und den ›Fall‹ ebenso ins Zentrum stellen.⁹ An den Fall werden nun weitere Informationen angelagert, teils als Eigenschaften des Falls, teils als mit ihm verbundene, weitere Entitäten: Ein Fall wird eröffnet dadurch, dass eine Person eine Eingabe an die Kongregation macht und die Auslegung eines Beschlusses des Trienter Konzils erbittet. Dies wird dadurch wiedergegeben, dass Personen als eigenständige Entitäten modelliert und dann mit einer qualifizierten n:m-Relation (»Petent_in«) mit dem Fall verknüpft werden. So wird die Unterscheidung zwischen der *Rolle* der Petentin und der physischen Person in der getrennten Notation von Personen- und Relations-Eigenschaften aufgezeichnet. Eine gegebene Person kann in mehreren Fällen als Petent und in anderen Fällen in einer anderen Rolle auftreten, während umgekehrt ein Fall mehrere Petenten haben kann.¹⁰ Ebenso wird der angesprochene Konzilsbeschluss

⁸ Vgl. Van Ruymbeke et al. 2017, *passim*. Dabei gewinnen sie ihre Definitionen in Auseinandersetzung mit der CIDOC-CRM-Erweiterung CRMsci (Doerr et al. 2018) und behalten leider ein (hier allerdings nicht weiter zu diskutierendes) physikalistisches Verständnis der ›Realität‹ bei.

⁹ Z.B. Wyner / Hoekstra 2012, Sperberg-McQueen 2016.

¹⁰ Vgl. erneut Hoekstra et al. 2008. Ob eine Person in diesem Kontext in ein- und demselben Fall in verschiedenen Rollen auftreten kann, ist erst noch wissenschaftlich zu diskutieren und kann ggf. später in einem Schema oder anderen konsistenzgewährenden Mechanismen sichergestellt werden. Zu thematischen Rollen allgemeiner vgl. Goy et al. 2018. Dort wird auch diskutiert, dass, anders als für soziale Rollen (z.B.

als ›Materie‹ mit den üblichen Indikatoren Sitzung und Dekret bzw. Canon festgehalten. Analog verhält es sich mit den anderen an dem Verfahren beteiligten Personen, mit den getroffenen Beschlüssen, den anderen Fällen, auf die die Beratungen mitunter verweisen und auch den einzelnen Anhörungsterminen. **Abbildung 2** gibt eine vorläufige Konzeption dieser Zusammenhänge wieder.

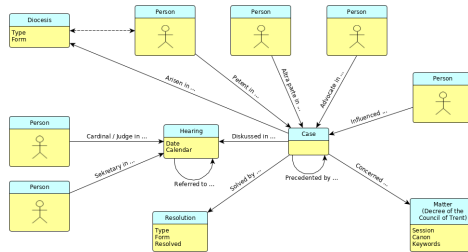


Abb. 2: Datenmodell der Ebene der historischen Phänomene. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.

Dabei wird das Datenmodell, werden also die Entitätstypen, ihre Attribute und Relationen (sowie einstweilen auch Attribute von Relationen) im Ausgang von der über Jahre ausdifferenzierten und bewährten Datenerfassung zunächst intuitiv formuliert und dann in der schemafreien Graphdatenbank *Neo4j* implementiert. Eine Abbildung auf stärker standardisierte Modelle wie CIDOC CRM bzw. RDF wird später evaluiert.¹¹ Zwar ist diese Abbildung als Exportschnittstelle in jedem Fall vorgesehen, allerdings ist es aktuell durchaus auch vorstellbar, dass der Datenbestand im Projekt allein im Rahmen der aktuell aufgebauten (und ggf. weiterentwickelten) Graphdatenbank beforscht und im Datenexport z.B. auf eine SPARQL-Schnittstelle verzichtet wird.

2.2. Schriftliche Quellen: Positiones

Die historischen Fälle wurden, wie oben beschrieben, in Berichten, sog. *Positiones* protokolliert, die sich in den Archiven finden lassen. In diesem Zusammenhang können z.B. der schriftführende Sekretär und Informationen über Siegel- und Pergament-Beschaffenheiten erfasst werden. Vor allem aber ist auf dieser Ebene der protokollarische Bericht (eine ideelle Entität, die sich auf die Phänomene der ersten Ebene bezieht) mit einem schriftlichen Dokument (einem symbolischen Ausdruck) und dieses wiederum mit einem materiellen

Äbtissin, Kanzler), thematische Rollen eine Modellierung als eigenständige Entität i.d.R. nicht erfordern und es nahe liegt, sie ausschließlich in Verbindung mit dem jeweiligen Fall zu modellieren — im vorliegenden Projekt eben als Typ der Relation zwischen Person und Fall.

¹¹ Vor allem auf dieser Ebene konzentriert sich im Übrigen die rechtliche Besonderheit des Forschungsgegenstands und -projekts. Die im Weiteren beschriebenen zweite und dritte Ebene sind auf historische Forschung bzw. auf geisteswissenschaftliche Forschung als solche zugeschnitten und weisen jenen spezifischen juristischen Charakter nicht mehr auf. Das legt auch nahe, auf dieser ersten Ebene sorgfältig zu überlegen, ob neben dem aus dem Bedarf von Kulturerbe-Institutionen stammenden CIDOC CRM Standard nicht auch Taxonomien, Vokabularien oder Ontologien aus dem Bereich des geltenden Rechts und der Rechtsinformatik integriert werden können. Zu denken wäre etwa an *LKIF-Core* (Hoekstra et al. 2008), *UNDO* (Peroni et al. 2017) oder die verschiedenen Elemente aus dem *European Legal Taxonomy Syllabus* (Ajani et al. 2016).

Träger verknüpft, der seinerseits in bestimmten archivalischen Fundzusammenhängen (Medieneinheit, Regal, Fundus, Gebäude, Adresse) eingeordnet ist. Damit ist auf dieser Ebene auch eine wichtige Möglichkeit angesiedelt, einige der angesprochenen Unsicherheiten in der Datenbank abzubilden. Denn neben der allgemeinen Relation zwischen Fall und Bericht werden auch die speziellen Einträge von Personen und deren Relationen auf der Ebene der historischen Vorgänge den Namen bzw. den Symbolen und Zeichenfolgen in den *Positiones* abgelesen. Wir möchten also idealerweise zweierlei Dinge explizit festhalten: Erstens, dass sich z.B. unter den Zeichenfolgen im Dokument eine befindet, die wir als eine Namensnennung identifiziert haben; zweitens, dass diese Namensnennung unserer Ansicht nach eine bestimmte historische Person bezeichnet. Streng genommen sind die historischen Personen also nur über Instanzen von Namensnennungen in den Dokumenten mit den Fällen assoziiert.

Informationen dieser Art sind in der aktuellen Datenerfassung nur sehr implizit und grundsätzlich enthalten, so dass diese uns für die Modellierung auf dieser Ebene keine ausreichende Orientierung bietet. Die oben in Anschlag gebrachte Begrifflichkeit von ideeller Entität, symbolischem Ausdruck und materiellem Träger legt allerdings nahe, sich an den CIDOC CRM Standard (und seine Erweiterungen FRBROo bzw. CRMtex)¹² anzulehnen. In einem ersten Anlauf verwenden wir also auf dieser Ebene Namen für Entitäts- und Relationstypen, die aus jenem Kontext stammen, wie in *Abbildung 3* ersichtlich.¹³ Die *Positio* ist dem Wesen nach ein CIDOC CRM »E31_Document«, welches den Fall dokumentiert (»P70_documents«). Im Besonderen enthält das Dokument (»P165_incorporates«) einen symbolischen Ausdruck, genauer: die Namensnennung eines Akteurs (»E82_Actor_Appellation«, eine Unterklasse von »E41_Appellation« bzw. »E90_Symbolic_Object«). Diese Namensnennung wird in der Quelle in einer Kette von Glyphen (»TX7_Written_Text_Fragment«) physisch abgebildet, und sie identifiziert (»P131_identifies«) einen Akteur (»E39_Actor«, entweder eine einzelne »E21_Person« oder eine ganze »E74_Group«) auf der historischen Ebene 1.

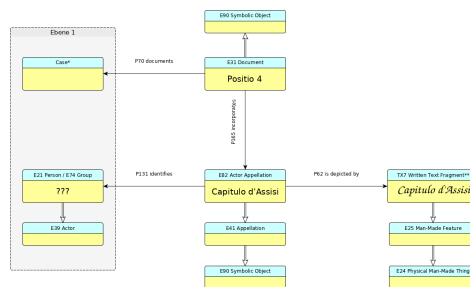


Abb. 3: Datenmodell der Ebene der archivalischen Quellen. ©Forschungsprojekt »Die Regierung der Universalkirche nach dem Konzil von Trient: 2018.

¹² vgl. Bekiari et al. 2017 und Murano / Felicetti 2017.

¹³ An dieser Stelle sei noch einmal darauf hingewiesen, dass es sich um einen aktuell noch andauernden, iterativen Entwicklungsprozess handelt und die hier vorgestellte Konzeption sich erst noch in Implementierung und praktischer Datenerfassung und -analyse bewähren muss.

Im aktuellen Entwicklungsstand ist noch offen, in welcher Weise diese Information mit den Aspekten der verschiedenen Rollen auf der Ebene 1 der historischen Vorgänge und Entitäten integriert wird. Denn es geht einerseits nicht allein um die Nennung einer Person im Zusammenhang mit einem Fall, sondern auch um die Zuweisung einer bestimmten Rolle innerhalb des Falles an diese Person, und andererseits wird die Person oft nicht mit persönlichem Namen, sondern mit ihrer sozialen Rolle (z.B. »der Domkapitular von Assisi«) genannt, die Namensnennung (»E82_Actor_Appellation«) identifiziert also zunächst gar keine konkrete Person. Ein erster Ansatz unter Rückgriff auf den Vorschlag von Martin Doerr¹⁴ versucht, dies über Attribute von Attributen (»P14.1_in_the_role_of« als Attribut von »P14_performed«) abzubilden. Dies würde uns aber zwingen, eine Aktivität in das Datenmodell hinein zu modellieren, über welche die genannten Attribute mit der Person verbunden werden könnten. Zudem ist zweifelhaft, ob diese Konstruktion sich in der Implementierung und im späteren Export in RDF leicht umsetzen lässt. Eine andere Möglichkeit wäre, die Rollen von Petenten, beratenden Kardinälen usw. doch zu eigenständigen Entitäten zu machen. So lassen sie sich zu Endpunkten weiterer Relationen, etwa mit den Namensnennungen, machen. Aber auch diese Strategie ist nicht unproblematisch: Wir würden — entgegen der ursprünglichen, oben dargestellten Intuition und entgegen der Empfehlung in¹⁵ — die Entitäten in uneleganter Weise vervielfachen, was nicht zuletzt die Abfragen komplizierter macht; die Namensnennung (oder die Nennung der sozialen Rolle, z.B. des »Domkapitular von Assisi«) ist ja selbst gar nicht auf die Rolle »Petent« bezogen; schließlich wird womöglich die spätere Integration mit CIDOC CRM erschwert. Kurzum: die Überlegungen hierzu sind zum gegenwärtigen Zeitpunkt noch nicht abgeschlossen.

2.3. Forschungsprozesse: Assignments

Auf der und durch die dritte Ebene wird es schließlich konkret möglich, die im Forschungszusammenhang auftretenden Zweifel festzuhalten. Dabei führt kein Weg daran vorbei, die Forscherinnen selbst ins Datenmodell mit einzubauen. Denn sie sind es, die Unklarheiten des Quellenmaterials feststellen oder selbst fallible Hypothesen aufstellen. Die Attribute und Relationen auf den ersten beiden Ebenen gehen zurück auf Interpretationen der Forscherinnen und müssen ggf. mit einer Art »Zweifels-Koeffizienten« versehen werden. Dies abzubilden bringt jedoch beträchtliche Komplikationen mit sich: In der Implementierung in der Graphdatenbank ist es zwar möglich, Relationen mit Eigenschaften zu versehen, nicht aber Attribute. So könnte man zwar z.B. eine ungewisse Einmischung des spanischen Königs als eine Relation zwischen König und Fall modellieren, an die man die Eigenschaft »ungewiss« und eventuell noch weitere Eigenschaften wie Quellen notiert, auf die sich die Annahme stützt oder solche, die sie umgekehrt gerade in Zweifel ziehen. Es ist aber schon nicht mehr einfach möglich, die Form eines Beschlusses, wenn sie als Attribut des Beschlusses notiert ist, zum Gegenstand einer vergleichbaren Qualifikation zu machen. Auf der Seite des uns in vielen Fragen zur Orientierung dienenden CIDOC CRM Ökosystems werden die entsprechenden Probleme von Alexiev¹⁶ beschrieben.

¹⁴ vgl. Martin Doerr 2015, *passim*.

¹⁵ vgl. Goy et al. 2018.

¹⁶ vgl. Alexiev 2012.

Wir erproben in diesem Zusammenhang die Umsetzung eines Vorschlages von Niccolucci und Hermon: In zweifelhaften Fällen wird die Feststellung eines Wertes selbst als ein Ereignis (»E13_Attribute_Assignment« oder eine seiner Unterklassen) modelliert, das selbst zum Gegenstand einer Einschätzung der Zuverlässigkeit werden kann (vgl. Abbildung 4).¹⁷ Auf diese Weise ist es möglich, konkurrierende Interpretationen verschiedener Forscherinnen hinsichtlich bestimmter Attribute oder Relationen vollständig zu erfassen. Ebenso können verschiedene Forscherinnen einer jeden solchen Interpretation unterschiedliche Gewissheits- und Zuverlässigkeits-Werte (Abschnitt 2.3) zuschreiben.

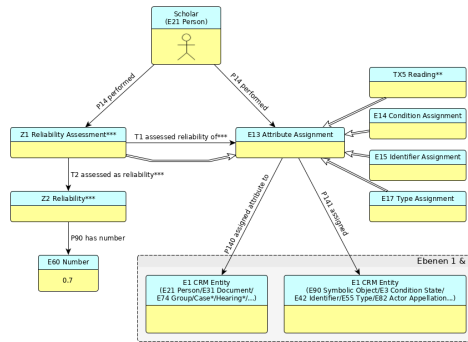


Abb. 4: Datenmodell der Ebene des Forschungsprozesses. ©Forschungsprojekt »Die Regierung der Universalkirche nach dem Konzil von Trient: 2018.
18

Als Alternative zu diesem Vorschlag wäre die explizite Modellierung von ganzen Interpretationszusammenhängen als »Überzeugungssystemen« denkbar, wie sie die CRMinf Erweiterung¹⁹ definiert, die mit neuen Entitäten (»I1_Argumentation«, »I2_Belief«, »I4_Proposition_Set« u.ä.) den Prozess des wissenschaftlichen Interpretierens und Schlussfolgerns insgesamt abbildet.²⁰ Das würde einerseits eine deutlichere und weitergehende Modellierung der wechselseitigen Stützung und Erklärung einzelner Aspekte jenes Wissens, idealerweise auch ausführliches Schlussfolgern über hier erst noch implizit enthaltenes Wissen erlauben. Aus Gründen der pragmatischen Implementierung und Anwendung folgen wir jedoch dem einfacheren Ansatz von Niccolucci und Hermon, der von diesen als mit CRMinf zwar kompatibel, jedoch mit den genannten Vorzügen qualifiziert beschrieben wird.²¹

¹⁷ vgl. Niccolucci / Hermon 2016.

¹⁸ Bezeichnungen der Entitäten und Relationen nach CIDOC CRM v6.2. Doppelte Pfeile geben Unterklassen-Relationen an. (*) Für die Entitäten »Case« und »Hearing« ist noch keine CIDOC-CRM-kompatible Definition gefunden. (**) Die Entität »TX5 Reading« ist definiert in der CIDOC CRM Erweiterung CRMttx; vgl. Murano / Felicetti 2017. (***) Die Entitäten »Z1 Reliability Assessment« und »Z2 Reliability« sowie die Relationen »T1 assessed reliability of« und »T2 assessed as reliability« sind definiert in Niccolucci / Hermon 2016.

¹⁹ vgl. Stead et al. 2015.

²⁰ vgl. dazu auch erneut Van Ruymbeke et al. 2017.

²¹ vgl. Niccolucci / Hermon 2016, S. 286. Im Speziellen für die Fragen der graphologischen Unsicherheiten vgl. Murano / Felicetti 2017.

Die Zuverlässigkeitsschätzung nach Niccolucci und Hermon²² erlaubt es, neben einer quantitativen Bestimmung der Zuverlässigkeit auch Faktoren festzuhalten, die die Einschätzung beeinflussen (über »P15_was_influenced_by« oder »P33_used_specific_technique«, die etwa die Klassifikation über einen definierten Kriterienkatalog aufnehmen kann) und auf weitere Dokumentation, z.B. einen Aufsatz, in dem die Forscherin zur Frage Stellung bezieht, hinzuweisen (über »P70_is_documented_in«). Gründe wie die Tatsache, dass einer der möglichen Werte der historischen Realität wahrhaft entspricht, dass diese vergangene historische »Realität« aber eben nicht mehr objektiv überprüfbar ist, führen Niccolucci und Hermon dazu, die Zuverlässigkeitswerte als numerische Werte zwischen 0 und 1 im Rahmen eines »Fuzzy« Kalküls zu verstehen. Damit können die Alternativen nicht nur mit Vorbehalten versehen und verglichen werden, sondern es können im Prinzip auch weitergehende Berechnungen nach einem formalen Kalkül angestellt werden.²³

3. Ausblick

In diesem Ausblick werden einige Herausforderungen genannt, die in der Modellierung und im praktischen Projektkontext noch offen sind oder sich als Herausforderungen durch die bisherigen Schritte erst deutlich herauskristallisiert haben.

Erstens gilt es, die oben genannten Vorläufigkeiten aufzusuchen, d.h. Punkte wie die Verankerungen der Relationen zwischen den Ebenen auf der jeweils tieferen Ebene zu klären und das Datenmodell konsequenter an CIDOC CRM auszurichten. Entitäts- und Relationstypen sollten ebenso wie Attribut-Bezeichnungen den durch CIDOC CRM definierten Bezeichnungen entsprechen. Gegebenenfalls kann auf offizielle Erweiterungen wie CRMinf zurückgegriffen werden. Um die Komplexität in der Eingabe und Bearbeitung der Daten zu beschränken oder zu maskieren, könnten auch eigene Bezeichnungen gewählt werden, solange diese möglichst eindeutig auf CIDOC CRM (und Erweiterungen) gemappt werden können. Sollte dies schließlich aus Gründen der ökonomischen und intuitiven Arbeit mit dem Datenbestand nötig werden, könnte es sich anbieten, einzelne Phänomene nach einer eigenen Konzeption zu modellieren und sie erst in einem letzten Export-Schritt durch eine wohldefinierte Transformation in RDF bzw. CIDOC CRM zu überführen. Dabei ist vor allem an die Möglichkeiten zu denken, bei der projektinternen Erfassung und Beforschung der Daten mit einem *Property Graph* zu arbeiten, also den Relationen ihrerseits Attribute zuzuordnen zu können. Dies lässt sich nicht direkt in RDF Tripel abbilden, sondern nur über die (automatische) Erzeugung zusätzlicher Entitäten anstelle der Relationen, die dann einerseits mit den relationierten Entitäten verknüpft werden und andererseits die Relations-Attribute aufnehmen können.

²² vgl. Niccolucci / Hermon 2016.

²³ Der Begriff der unscharfen, »fuzzy« Mengen, Systeme und Methoden wurde von Zadeh eingeführt, vgl. Zadeh 1965. Zu seiner Verwendung im geisteswissenschaftlichen Zusammenhang vgl. Termini 2012 und Thaller 2018. In unserem Fall werden beim Import aus dem Spreadsheet Daten, die als Interpretation gekennzeichnet sind, in einem ersten Schritt mit einem Standard-Gewissheitswert von 0,95 versehen. Durch eine Abfrage von solchermaßen qualifizierten *Assignments* können sie in einem zweiten Durchgang in Fällen, wo es sich um eine tatsächlich ungesicherte Interpretation handelt, nach unten korrigiert werden. Welche Workflows und Werte sich hier optimal verwenden lassen, muss sich aber erst noch in der Praxis ergeben.

Zweitens liegt auf der Hand, dass mit der Entwicklung des Datenmodells nur der allererste (oder jedenfalls nur ein sehr früher) Schritt getan ist, dass es also in einer Datenbank (a) implementiert und mit den bereits vorliegenden und weiter angelieferten Daten (b) befüllt werden muss. Dabei gehen wir im Projekt parallel vor: eine auf der Basis der bereits vorliegenden Erfassungs-Tabellen intuitiv erzeugte und befüllte Graphdatenbank (vgl. *Abbildung 5*) wird zunehmend und durch systematische Transformationen an das reflektierte Datenmodell angepasst, während umgekehrt das Datenmodell nicht zuletzt durch die Arbeit mit der bereits bestehenden Graphdatenbank verfeinert wird. Durch die systematische Definition und Redefinition der Datenbank kann diese jederzeit mit unterschiedlichen Datenmodellen aus den ursprünglichen (CSV-)Tabellen neu erzeugt werden.

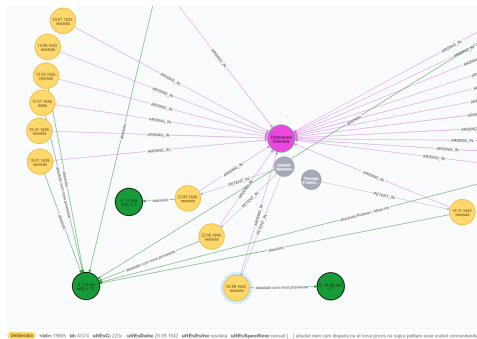


Abb. 5: Ausschnitt der Graphdatenbank. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.

Daran muss drittens die Entwicklung von Oberflächen-Elementen anschließen, die zum einen auch technisch weniger versierten Projektbeteiligten die Erfassung, Bearbeitung und Abfrage von Daten erlaubt. Beispielsweise sollte die Komplexität der Einschätzung von Zuverlässigkeiten aus arbeitsökonomischen Gründen durch technische Vorkehrungen wie z.B. ein- und ausblendbare Eingabefelder, Default-Werte, und möglicherweise grafische Elemente wie Repräsentation und Auswahl der möglichen Werte auf einer Farbskala (anstelle von numerischen Werten und Komma- oder Prozentzahlen) reduziert bzw. ausgeblendet und nur bei Bedarf erweitert und expliziert werden. Auch für die Dokumentation der Forschung und als Angebot an andere an dem Datensatz interessierte Besucher kann darüber nachgedacht werden, bestimmte Ansichten und Analysen der Daten aufzubereiten und über den Datenbank-Server frei zugänglich zu machen. Dazu eignen sich die sogenannten *Browser Guides* in Neo4j hervorragend (vgl. *Abbildung 6*).²⁴

²⁴ In den *Neo4j Browser Guides* sind der Mechanismus und das gleichnamige Format dokumentiert.

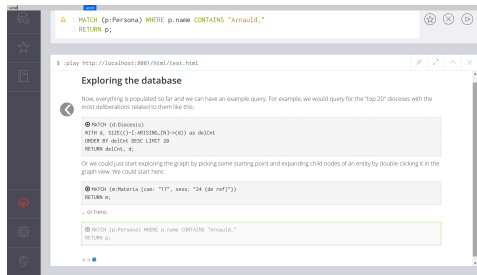


Abb. 6: Beispiel eines *Browser Guides*. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.

Viertens und letztens schließlich muss geklärt werden, in welcher Weise die hier im Fokus stehenden Informationen zu Ungewissheiten und Zweifeln eigentlich ausgewertet und in den Suchen und Analysen berücksichtigt werden sollen. Unter welchen Bedingungen und in welcher Form sollen z.B. Diskussionsteilnehmer, deren Identität unsicher ist, in Abfragen angezeigt werden? Nur, wenn die Abfragenden dies in irgendeiner Weise explizit anfordern oder immer? Sollen sie chronologisch bzw. alphabetisch in die Ergebnismenge einsortiert werden oder anhand der Gewissheit, mit der sie zur korrekten Ergebnismenge gehören? Gibt es einen praktikablen Weg, den ›Zuverlässigkeitswert‹ eines Feldes in der grafischen Anzeige zu repräsentieren? Auch den einer Relation? Wie kann abgebildet werden, dass mit einer Bezeichnung zwei Personen im Datenbestand gemeint sein könnten? Und noch vor jeder Überlegung zu Fragen der Darstellung: Wie verändert sich die Gewissheit, wenn nach zwei Feldern gefragt ist und beide ungewiss sind, wie wenn ein Feld sehr ungewiss ist?²⁵

Die weitere Projektarbeit wird zeigen, ob sich auf alle diese Fragen und für alle genannten Herausforderungen befriedigende, praktikable und zugleich wissenschaftlich fortschrittliche Lösungen und Antworten finden lassen.

²⁵ Zu Ansätzen, Kalküle aus der *fuzzy logic* in historischer Forschung oder allgemeiner zur Integration verschiedener Expertenmeinungen heranzuziehen vgl. Thaller 2018 bzw. Yager et al. 2012.

Bibliographische Angaben

- Gianmaria Ajani / Guido Boella / Luigi di Caro / Livio Robaldo / Llio Humphreys / Sabrina Praduroux / Piercarlo Rossi / Andrea Violato: The European Legal Taxonomy Syllabus: A Multi-Lingual, Multi-Level Ontology Framework to Untangle the Web of European Legal Terminology. In: *Applied Ontology* 11 (2016), H. 4, S. 325–375. [[Nachweis im GBV](#)]
- Benedetta Albani: In universo christiano orbe: la Sacra Congregazione del Concilio e l'amministrazione dei sacramenti nel Nuovo Mondo (secoli XVI–XVII). DOI: [10.3406/mefr.2009.10577](#) In: *Mélanges de l'École française de Rome / Italie et Méditerranée* 121 (2009), H. 1, S. 63–73. [[online](#)] [[Nachweis im GBV](#)]
- Vladimir Alexiev: Types and Annotations for CIDOC CRM Properties. Report at Digital Presentation and Preservation of Cultural and Scientific Heritage (DiPP2012, Veliko Tarnovo 18.–21.09.2012). In: *ontotext.com*. Documents. Publications. Veliko Tarnovo 18.09.2012. PDF. [[online](#)]
- Chrysoula Bekiari / Martin Doerr / Patrick Le Bœuf / Pat Riva: FRBRoo: object-oriented definition and mapping from FRBRER, FRAD and FRASD. In: *cidoc-crm.org*. Frbrroo. Version 3.0. von September 2017. PDF. [[online](#)]
- Tanya Clement: Knowledge Representation and Digital Scholarly Editions in Theory and Practice. DOI: [10.4000/jtei.203](#) In: *Journal of the Text Encoding Initiative* 1 (2011). DOI: [10.4000/jtei.125](#)
- Martin Doerr: How to model Roles in the CIDOC CRM RDF encoding. In: *cidoc-crm.org*. Frequently Asked Questions. Beitrag vom 12.02.2015. [[online](#)]
- Martin Doerr / Athina Kritsotaki / Yannis Rousakis / Gerald Hiebel / Maria Theodoridou et al.: Definition of the CRMsci. An Extension of CIDOC-CRM to support scientific observation. In: *cidoc-crm.org*. CRMsci. Version 1.2.5. von Mai 2018. PDF. [[online](#)]
- Anna Goy / Diego Magro / Marco Rovera: On the role of thematic roles in a historical event ontology. In: *Applied Ontology* 13 (2018), H. 1, S. 19–39. [[Nachweis im GBV](#)]
- Rinke Hoekstra / Joost Breuker / Marcello Di Bello / Alexander Boer: The LKIF Core Ontology of Basic Legal Concepts. PDF. [[online](#)] In: *Proceedings of LOAIT 07: II Workshop on Legal Ontologies and Artificial Intelligence Techniques*. Hg. von Pompeu Casanovas / Maria Angela Biasiotti / Enrico Francesconi / Maria Teresa Sagri. (LOAIT: 2, Stanford, CA, 04.06.2007) Aachen 2008, S. 43–63. URN: [urn:nbn:de:0074-321-1](#)
- Francesca Murano / Achille Felicetti: Definition of the CRMtex. An Extension of CIDOC CRM to Model Ancient Textual Entities. In: *cidoc-crm.org*. CRMtex. Version 0.8. von Januar 2017. PDF. [[online](#)]
- Franco Niccolucci / Sorin Hermon: Expressing Reliability with CIDOC CRM. In: *International Journal on Digital Libraries* 18 (2016), H. 4, S. 281–87. Artikel vom 07.10.2016. [[Nachweis im GBV](#)]
- Definition of the CIDOC Conceptual Reference Model. Hg. von Christian Emil Ore / Martin Doerr / Patrick LeBœuf / Stephen Stead. In: *cidoc-crm.org*. Version 6.2.2. von September 2017. PDF. [[online](#)]
- Silvio Peroni / Monica Palmirani / Fabio Vitali: UNDO: The United Nations System Document Ontology. In: *International Semantic Web Conference 2017*. Hg. von Claudia D'Amato. 2 Bde. (ISWC 2017, Wien, 21.–25.10.2017) Cham 2017. Bd. 2, S. 175–183. [[Nachweis im GBV](#)]
- Muriel Van Ruymbekke / Pierre Hallot / Roland Billen: Enhancing CIDOC-CRM and Compatible Models with the Concept of Multiple Interpretation. DOI: [10.5194/isprs-annals-IV-2-W2-287-2017](#) In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences 2017 IV-2/W2*. Hg. von J. Hayes / C. Oumet / M. Santana Quintero / S. Fai / L. Smith. (Ottawa, Canada 28.08.–01.09.2017) Red Hook 2017, S. 287–94. [[online](#)] [[Nachweis im GBV](#)]
- Michael Sperberg-McQueen: Trials of the Late Roman Republic: Providing XML infrastructure on a shoe-string for a distributed academic project. In: *Proceedings of Balisage: The Markup Conference 2016*. (Balisage: 17, Washington, DC, 02-05.-08.2016) Rockville, MD 2016. (= Balisage Series on Markup Technologies, 17) [[online](#)]
- Stephen Stead / Martin Doerr et al.: CRMinf: the Argumentation Model. An Extension of CIDOC-CRM to support argumentation. In: *cidoc-crm.org*. CRMinf. Version 0.7. von Februar 2015. [[online](#)]
- Settimo Termini: On some «Family Resemblances» of Fuzzy Set Theory and Human Sciences. In: *Soft Computing in Humanities and Social Sciences*. Hg. von Rudolf Seising / Veronica Sanz. Berlin u.a. 2012, S. 39–54 (= *Studies in Fuzziness and Soft Computing*, 273) [[Nachweis im GBV](#)]
- Manfred Thaller: On Information in Historical Sources. In: *A Digital Ivory Tower. Prolegomena for a computer science for historical studies*. Blogbeitrag vom 24. April 2018. [[online](#)]
- Adam Wyner / Rinke Hoekstra: A Legal Case OWL Ontology with an Instantiation of Popov v. Hayashi. In: *Artificial Intelligence and Law* 20 (2012), H. 1, S. 83–107. [[Nachweis im GBV](#)]
- Ronald Yager / Henri Prade / Didier Dubois: Merging Fuzzy Information. In: *Fuzzy Sets in Approximate Reasoning and Information Systems*. Boston, MA u.a. 2012. (= *The Handbook of Fuzzy Sets Series*, 5) [[Nachweis im GBV](#)]
- Lotfi Zadeh: Fuzzy Sets. In: *Information and Control* 8 (1965), H. 3, S. 338–353. [[Nachweis im GBV](#)]

Abbildungslegenden- und nachweise

Abb. 1: Datenerfassung im Spreadsheet. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.

Abb. 2: Datenmodell der Ebene der historischen Phänomene. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.

Abb. 3: Datenmodell der Ebene der archivalischen Quellen. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.

Abb. 4: Datenmodell der Ebene des Forschungsprozesses. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.

Abb. 5: Ausschnitt der Graphdatenbank. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.

Abb. 6: Beispiel eines *Browser Guides*. ©Forschungsprojekt ›Die Regierung der Universalkirche nach dem Konzil von Trient‹ 2018.