

Digitale, formale Methoden und Modelle in den Geschichtswissenschaften.

Am Beispiel digital editierter historischer
Rechnungsbücher.

Masterarbeit

zur Erlangung des akademischen
Grades Master of Arts (MA)

an der Karl-Franzens-Universität Graz

vorgelegt von
Christopher POLLIN, BA MA

am Institut für Geschichte
Begutachterin: Univ.-Prof. Dr. Gabriele Haug-Moritz

Graz, 2020

Inhaltsverzeichnis

1	Einleitung	1
1.1	Aufbau der Arbeit und Literaturübersicht	2
1.2	Related Work	4
1.2.1	Digitale Arbeiten mit historischen Materialien	4
1.2.2	Analoge und digitale Arbeiten mit historischen Rechnungsunterlagen .	5
1.2.3	Digitale Arbeiten mit historischen Materialien unter Einsatz von <i>Web of Data</i> Technologien	7
2	Formale Methoden und Modelle in den Geschichtswissenschaften	8
2.1	Von der erzählenden zur theorieorientierten Geschichtswissenschaft	9
2.1.1	Geschichte und Geschichtswissenschaft: eine Definition	10
2.1.2	Hermeneutik	12
2.1.3	Empirismus	13
2.1.4	Methoden der Geschichtswissenschaft	15
2.2	”Quantifizierer” und ”Traditionalisten”: Entwicklung einer formalen Methodik	18
2.2.1	”Historische Fakten”	19
2.2.2	Interpretation und Theoriebildung	20
2.3	Modellbildung als Kern der Digital History	22
2.3.1	Definition und Terminologie des Modellbegriffs.	22
2.3.2	Modellierung	25
2.3.3	Forschungsdaten in den Geschichtswissenschaften	27
3	Historische Rechnungsbücher als Quelle	31
3.1	Wirtschafts- und Rechnungsbücher	31
3.2	Digital Edition Publishing Cooperative for Historical Accounts	34
3.2.1	The George Washington Financial Papers	34
3.2.2	Cameron Family Papers, Stagville Accounts	36
3.2.3	The Wheaton Accounts	38
4	Web of Data	44
4.1	Geschichte und Vision des <i>Web of Data</i>	44
4.1.1	Kritik am <i>Web of Data</i>	45
4.2	Web of Data Stack	47
4.2.1	Daten als Graph: Resource Description Framework	48
4.2.2	Klassen und Beziehungen: Resource Description Framework Schema (RDFS)	51
4.2.3	Abfragesprache: SPARQL	53
4.3	Formalisierung von Modellen: Ontologien	55

4.3.1	Vom Wissen, über das Semantische Netz zur Ontologie	55
4.3.2	Reasoning	58
5	Anwendung formaler Methoden auf Basis einer Ontologie	59
5.1	Digitale Edition von historischen Rechnungsbüchern	59
5.1.1	Digitale Edition und Web of Data	62
5.2	Die Bookkeeping-Ontology	64
5.2.1	Grundlagen	64
5.2.2	Beschreibung	65
5.2.3	Beispiel	65
5.3	Informationsvisualisierung	70
5.3.1	Voraussetzungen	70
5.3.2	Multidimensionale Rechnungsbücher – Multiple Views und Interaktion	72
5.3.3	Informationsvisualisierung und Web of Data	77
6	Zusammenfassung und Ausblick	78

1 Einleitung

Seitdem Menschen sich an Ereignisse oder Erkenntnisse, die für sie wichtig waren, erinnern, existieren auch Formen dieses Wissen zu repräsentieren und zu speichern. Bereits in den ersten Hochkulturen schrieb man nieder, dass jemand bei jemand anderen in der Schuld stand. Bereits um 3000 v. Chr. lässt sich auf Tontafel finden: *”Alok schuldet dem Sumar 10 Eimer Korn”*. Die Dokumentation ökonomischer Abhängigkeiten zwischen Akteuren stellt eine der ersten Formen dar, wie Wissen verschriftlicht und strukturiert wurde.¹ Für Historikerinnen liefert diese Zeile ein Indiz auf ein vergangenes Ereignis. Wird es in einen bestimmten historischen Kontext gestellt, so lassen sich Forschungsfragen in den Geschichtswissenschaften formulieren, die dabei helfen, ein Ereignis aus der Vergangenheit zu rekonstruieren und auf diese Weise etwas über die Vergangenheit lernen zu können. Dabei können semantische Strukturen helfen, die innerhalb der schriftlichen Überlieferung – der historischen Quelle – stecken. Im Falle von ökonomischen Abhängigkeiten und Transaktionen lässt sich eine ”natürliche” Struktur festmachen, die über Kulturen und Gesellschaften, sowie Zeiten und geographische Räume vergleichbar ist. Dies gilt für eine Tontafel aus Mesopotamien, einer Abgabenliste einer europäischen Stadt im Mittelalter, einem privaten Rechnungsbuch eines Geschäftsmannes aus dem 19. Jahrhundert in den Vereinigten Staaten von Amerika, sowie für einen Rechnungsbeleg aus heutiger Zeit. Akteure schulden oder transferieren wirtschaftliche Güter oder Geldbeträge an andere Akteure. Die Beschreibung einer solchen semantischen Struktur kann durch ein konzeptuelles Modell erfolgen. Die Anwendung formaler Methoden unter Berücksichtigung eines solchen Modells, kann als Grundlage dafür dienen, Interpretationen der Vergangenheit – *”wie es den gewesen sein könnte”*² – mit einer gewissen Datengrundlage zu untermauern. Auf dieser Datengrundlage, die stets offen und nachvollziehbar, sowie in die fachwissenschaftliche Diskussion eingebettet sein muss und deswegen ein konzeptuelles Modell benötigt, können die Interpretationen der Historikerinnen fußen. Die Rekonstruktion, Interpretation und die Einbindung dieser Erkenntnisse in gegenwärtige Diskussion ist die eigentliche Arbeit der Geschichtswissenschaft.

Ziel dieser Arbeit ist eine theoretische und pragmatische Auseinandersetzung mit den Herausforderungen und Möglichkeiten, die formale – und somit auch digitale – Methoden und Modelle in der Geschichtswissenschaft mit sich bringen. Der praktische Anteil dieser Arbeit bezieht sich auf die Anwendung formaler Methoden auf semantisch angereicherte digitale Editionen von historischen Rechnungsbüchern in einem konkreten Projektkontext. Dieser Projektkontext – *Digital Edition Publishing Cooperative for Historical Accounts (DEPCHA)*³ – umfasst eine

¹ROBSON, Keith: Accounting numbers as “inscription”: Action at a distance and the development of accounting. *Accounting, Organizations and Society*, 17 1992 Nr. 7.

²THALLER, Manfred: Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte ’unscharfer’ Systeme [1984]. *Historische Sozialforschung*. Supplement 29, 2017, S.159.

³NHPRC-Mellon Implementation Grants: Digital Edition Publishing Cooperatives, archives.gov/nhprc/announcement/depc. Finanziert durch The Andrew W. Mellon Foundation mellon.org, 05.06.2019.

Kooperation von Partnern aus den USA, koordiniert durch das *Wheaton College Massachusetts*, dem *Center for Digital Editing* der University of Virginia und dem *Zentrum für Informationsmodellierung* (ZIM) der Universität Graz. Es setzt sich zum Ziel eine Publikationsplattform für digitale Editionen historischer Rechnungsbücher zu entwickeln und umzusetzen. Eine Herausforderung in diesem Projekt umfasst die Diskussion und Zusammenführung quantitativer und qualitativer Herangehensweisen von Historikerinnen an Forschungsfragen in den Geschichtswissenschaften. Einen zentralen Stellenwert darin hat die sogenannte *Bookkeeping Ontology*, die als konzeptuelles Modell die semantische Struktur von Rechnungsunterlagen beschreibt und im Sinn des *Web of Data* (aka. Semantic Web) implementiert ist.

Diese Arbeit geht nun folgern Forschungsfrage nach:

Inwieweit können Technologien des *Web of Data* genutzt werden, um geschichtswissenschaftliche Fachdomänen, in diesem Fall die Edition und Analyse von historischen Rechnungsbücher, so zu beschreiben, dass die Anwendung formaler Methoden den geschichtswissenschaftlichen Kriterien gerecht wird und die Zurverfügungstellung als hochstrukturierte Forschungsdaten, sowie Formen der Analyse (z.B. Informationsvisualisierung), die weitere Interpretation von Forschungsfragen verbessert?

1.1 Aufbau der Arbeit und Literaturübersicht

Am Anfang der Arbeit steht neben einer allgemeinen theoretischen Diskussion zu Theorien und Methoden in den Geschichtswissenschaften^{4,5} eine Auseinandersetzung mit der Anwendung formaler Methoden in diesem Fachbereich⁶ unter Einbeziehung von Inhalten aus dem Bereich der historischen Fachinformatik,⁷ der digitalen Geisteswissenschaften⁸ und Digital History.⁹ Ziel ist es, einen Überblick darüber zu verschaffen, welche Ziele die Geschichtswissenschaft verfolgt. Dabei wird besonders Rücksicht auf die "digitale" Komponente^{10,11} in diesem Zusammenhang genommen, um zu zeigen, dass es keine Gegensätze zwischen "analoger" und "digitaler" Geschichtswissenschaft gibt, sondern Synergien.

⁴JORDAN, Stefan: Theorien und Methoden der Geschichtswissenschaft. 4. Auflage. UTB GmbH, 2018.

⁵HARDTWIG, Wolfgang: Geschichtskultur und Wissenschaft. München, 1990.

⁶THALLER: Historische Sozialforschung. Supplement 29 2017.

⁷THALLER, Manfred: Historical Information Science: Is there such a Thing? New Comments on an old Idea [1993]. Historical Social Research. Supplement 29, 2017.

⁸JANNIDIS, Fotis/KOHLE, Hubertus/REHBEIN, Malte: Digital Humanities: Eine Einführung. Springer, 2017.

⁹GRAHAM, Shawn/MILLIGAN, Ian/WEINGART, Scott: Exploring big historical data: The historian's macroscope. London, 2015.

¹⁰FÖHR, Pascal: Historische Quellenkritik im Digitalen Zeitalter. Dissertation, Universität Basel, 2017.

¹¹WINTERGRÜN, Dirk: Netzwerkanalysen und semantische Datenmodellierung als heuristische Instrumente für die historische Forschung. Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2019.

Das nächste Kapitel versteht sich als Quellenstudie und behandelt den Quellentypus historischer Rechnungsbücher¹² und ihre Edition.¹³ In einem zweiten Schritt werden konkrete Editionsprojekte aus DEPCHA angeführt. Diese Auswahl umfasst drei Projekte, die *George Washington Financial Papers*,¹⁴ die *Stagville Accounts*¹⁵ und die *Wheaton Family Papers*.¹⁶

Eine Möglichkeit Modelle^{17,18} zu implementieren, ist mit den Standards des *Web of Data* gegeben. Im vierten Kapitel wird auf den Technologie Stack des *Web of Data* eingegangen und so die Grundlage zur konzeptuellen und daraus abgeleiteten formalen Beschreibung von Information im *World Wide Web* erörtert. Dazu ist eine konzeptionelle,¹⁹ technische²⁰ und kritische²¹ Diskussion dieses Themenbereiches notwendig. Ein zentraler Aspekt dabei sind die Themen Ontologien²² und *Linked Open Data*,²³ die dazu dienen können Modelle und Forschungsdaten²⁴ über das Web auszutauschen.

Der gemeinsame Nenner dieser Projekte umfasst neben der zeitlichen und regionalen Einordnung der Quellen, auch den methodische Zugang zur Erschließung. Diese erfolgt im Sinne einer digitalen Edition.²⁵ Der dabei verwendete Standard ist die *Text Encoding Initiative (TEI)*.²⁶ Dieses Kapitel widmet sich der Fragestellung, wie digitale Editionen von historischen Rechnungsbüchern umgesetzt werden können^{27,28} Die digitale Edition in dieser Arbeit versteht sich als quellen-orientierte und inhaltsbezogene Edition, in der nicht jedes textuelle Phänomen ediert

¹²GLEBA, Gudrun/PETERSEN, Niels: Wirtschafts-und Rechnungsbücher des Mittelalters und der Frühen Neuzeit-Formen und Methoden der Rechnungslegung: Städte, Klöster und Kaufleute. Göttingen, 2015.

¹³VOGELER, Georg: Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert? In BAUM, Constanze/STÄCKER, Thomas (Hrsg.): Grenzen und Möglichkeiten der Digital Humanities. 2015 (URL: www.zfdg.de/sb001_007).

¹⁴The George Washington Financial Papers Project, financial.gwpapers.org, 30.01.2020.

¹⁵Stagville Accounts, fromthepage.com/agbedavies/stagville-accounts, 30.01.2020.

¹⁶Wheaton Family Papers, digitalrepository.wheatoncollege.edu/handle/11040/7928, 30.01.2020.

¹⁷STACHOWIAK, Herbert: Allgemeine Modelltheorie. 1973.

¹⁸MCCARTY, Willard: Modeling: A Study in Words and Meanings. In SUSAN SCHREIBMAN, RAY SIEMENS, John Unsworth (Hrsg.): A Companion to Digital Humanities. John Wiley and Sons, Ltd, 2007. – Kapitel 19.

¹⁹BERNERS-LEE, Tim/HENDLER, James/LASSILA, Ora: The semantic web. Scientific american, 284 2001 Nr. 5.

²⁰BERNSTEIN, Abraham/HENDLER, James/NOY, Natalya: A New Look at the Semantic Web. Commun. ACM, 59 2016 Nr. 9.

²¹SWARTZ, Aaron: A Programmable Web: An Unfinished Work. Synthesis lectures on the semantic web: Theory and Technology, 3 2013 Nr. 2.

²²STUCKENSCHMIDT, Heiner: Ontologien: Konzepte, Technologien und Anwendungen. Springer, 2009.

²³BAUER, Florian/KALTENBÖCK, Martin: Linked open data: The essentials. A Quick Start Guide for Decision Makers. Wien: Edition mono/monochrom, 2011.

²⁴NEHER, Günther/RITSCHER, Bernd: Semantische Vernetzung von Forschungsdaten. In BÜTTNER, Stephan/HOBOHM, Hans-Christoph/MÜLLER, Lars (Hrsg.): Handbuch Forschungsdatenmanagement. Band 6, 2011.

²⁵SAHLE, Patrick: Digitale Editionsformen: Textbegriffe und Recodierung. Band 3, Norderstedt, 2013.

²⁶CUMMINGS, James: The text encoding initiative and the study of literature. A Companion to Digital Literary Studies 2013.

²⁷TOMASEK, Kathryn/BAUMAN, Syd: Encoding financial records for historical research. Journal of the Text Encoding Initiative 2013 Nr. 6.

²⁸VOGELER, Georg et al.: The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies. In SARNOWSKY, Jürgen (Hrsg.): Konzeptionelle Überlegungen zur Edition von Rechnungen und Amtsbüchern des späten Mittelalters. Göttingen, 2016.

werden muss, sondern die semantische Struktur der Quellen im Vordergrund stehen.²⁹ Zur formalen Beschreibung semantischer Strukturen kann das *Web of Data* herangezogen werden. Dies soll am Beispiel der *Bookkeeping Ontology*,³⁰ einem konzeptuellen Modell zur Beschreibung von Transaktionsprozessen in historischen Rechnungsunterlagen, diskutiert und reflektiert werden. Auf diesem Modell und den Forschungsdaten werden weiter formale Methoden zur Analyse von Rechnungsbüchern im Sinne einer Informationsvisualisierung³¹ beschrieben und reflektiert.

1.2 Related Work

Es lässt sich eine Vielzahl von Projekten festmachen, die quantitative, formale und/oder digitale Verfahren zur Bearbeitung und Erschließung von historischen Quellen verwenden und darüber reflektieren. Dabei gibt es auch zahlreiche Beispiele hinsichtlich des Quellentypus der historischen Rechnungsbücher, der digitaler Edition dieser Quellengattung, sowie der Anwendung von *Web of Data* Technologien im Bereich der Geistes- bzw. Geschichtswissenschaften.

1.2.1 Digitale Arbeiten mit historischen Materialien

VASOLD hat in seiner Arbeit in den 1990er Jahren mit der computergestützten Verarbeitung der Itinerar³² des Erzbischofs Konrad IV. von Salzburg die Möglichkeiten und den Mehrwert formaler Methoden in den Geschichtswissenschaften gezeigt. Er hat auf Basis eines mittelgroßen Quellenbestandes aus Urkunden und historiographischen Belegen, die Reisetätigkeiten eines Erzbischofes aus dem Mittelalter rekonstruiert, in dem er ein Modell zur formalen Beschreibung der gesammelten Belege entwickelt und die Daten daraus im quellenorientierten Datenbanksystem Kleio verarbeitet hat. Kleio ist ein Datenverarbeitungsprogramm, aus den 80iger und 90iger Jahren, das die Arbeit mit komplexen (historischen) Daten ermöglicht und dadurch die Forschungstätigkeit von Historikerinnen unterstützt.³³ Als Vorteil dieser Herangehensweise führt VASOLD an, dass durch die Ein- und Ausblendung bestimmter Strukturen und Quellen, sich ein stabiles Bild des Weges einer Person skizzieren lässt.³⁴

²⁹VOGELER, Georg: The 'assertive edition'. On the consequences of digital methods in scholarly editing for historians. *International Journal of Digital Humanities*, 1 2019.

³⁰POLLIN, Christopher: Digital Edition Publishing Cooperative for Historical Accounts and the Bookkeeping Ontology. In *Proceedings of the Doctoral Symposium on Research on Online Databases in History*. 2019.

³¹PREIM, Bernhard/DACHSELT, Raimund: *Interaktive Systeme: Band 1: Grundlagen, Graphical User Interfaces, Informationsvisualisierung*. 2. Auflage. Berlin / Heidelberg, 2010.

³²Als Itinerar versteht man nicht nur Reiseberichte, sondern auch die Methode der Rekonstruktion der getätigten Wege einer Person oder Gruppe.

³³HOFFMANN, Barbara/HORN, Klaus: Ehepaare, Eheverläufe und Lebenslauf in Leipzig 1580-1730 KLEIO-Anwendung in einem historischen Forschungsprojekt. *Historische Sozialforschung*, 15 1990 Nr. 3 (55, S.172-174).

³⁴VASOLD, Gunter: *Das Itinerar Erzbischof Konrads IV. von Salzburg: 1291-1312; computergestützte Itinerarerstellung und Itineraranalyse*. Selbstverl. d. Inst. für Geschichte d. Karl-Franzens-Univ. Graz, 1996.

BURGHARTZ, CALVI und VOGELER haben eine TEI/XML basierte digitale Edition des Basler Urfehdebuchs X, das Urfehdeeeinträge aus den Jahren 1563 bis 1569 beinhaltet, im Web veröffentlicht. Als Urfehde versteht man ein Rechtsmittel in der Neuzeit, um einen Konflikt zweier Parteien zu schlichten. Die digitale Edition liefert eine transkribierte Quelle, die nach einem Modell semantisch beschrieben wurde. Täter, Tat, Opfer, Strafen und personenbezogene Information werden so maschinenlesbar strukturiert und können weiter ausgewertet werden.³⁵

Ein weiteres Beispiel für die Zurverfügungstellung eines administrativen, historischen Quellenkorpus über das Web ist das *Itinera Nova* Projekt. Im Verlauf dieses Projekts hat das Stadtarchiv Leuven in Kooperation mit dem *Cologne Center for eHumanities* und dem *Data Center for the Humanities* der Universität Köln 950.000 Seiten aus Schöffenenakten aus der Zeit zwischen 1362-1795 digitalisiert und im Volltext indiziert.³⁶

1.2.2 Analoge und digitale Arbeiten mit historischen Rechnungsunterlagen

Da Rechnungsbüchern eine relativ einheitliche und "natürliche" Struktur aufweisen und ihr Inhalt sich für die Anwendung quantitativer und computergestützter Methoden eignet, lassen sich eine Vielzahl von Projekten festmachen, die Rechnungsunterlagen formal ausgewertet haben. Hier sind nur einige wenige angeführt:

Im Projekt "*Rechnungsbücher von Bistritz aus den Jahren 1461–1520 ediert*", werden Steuererzeichnisse erschlossen, die einen Einblick in die wirtschaftliche und soziale Struktur der drittgrößten Stadt Siebenbürgens zu dieser Zeit liefern.³⁷

NEUWÖHNER beschreibt ausführlich eine Geschichte der Stadt Paderborn im 17. Jahrhundert, die auf den Einnahmen und Ausgaben der Stadtrechnungen basiert. Nicht die Erfassung und Edition der historischen Quellen liegt in dieser Arbeit im Vordergrund, sondern veranschaulicht sie, wie reichhaltig die geschichtswissenschaftliche Interpretation des Quellentypus Stadtrechnung sein kann, um die finanzielle, wirtschaftliche und soziale Dimension einer Stadt zu erforschen.³⁸

Die Arbeit von ZIEGLER, zeigt, dass die Überlegungen zur Anwendung quantitativer Methoden auf historische Rechnungsbücher dazu führen kann, komplexere Sachverhalte zu bearbeiten, wie etwa den Staatshaushalt Bayerns in der zweiten Hälfte des 15. Jahrhunderts. In seiner

³⁵POLLIN, Christopher/VOGELER, Georg: Semantically Enriched Historical Data. Drawing on the Example of the Digital Edition of the "Urfehdebücher der Stadt Basel". In WHiSe, ISWC. 2017, S.27-29.

³⁶Itinera Nova - Redaktions- und Präsentationssystem für die mittelalterlichen Schöffenenakten der Stadt Leuven, itineranova.be, 30.01.2020.

³⁷Rechnungsbücher von Bistritz aus den Jahren 1461–1520 ediert, tinyurl.com/rgxw27q, 15.07.2019

³⁸NEUWÖHNER, Andreas: Haushalt und Finanzen im Spiegel kommunaler Rechnungen: die Stadt Paderborn im 17. Jahrhundert. In PÄTZOLD, Stefan/STUMPF, Marcus (Hrsg.): Mittelalterliche und frühneuzeitliche Rechnungen als Quellen der landesgeschichtlichen Forschung. Münster, 2016, S.11-50.

Zusammenfassung ist sich ZIEGLER bewusst, dass auf Grund der unscharfen und lückenhaften Datengrundlage es sich nur um eine historische Annäherung handeln kann, die aber dennoch für die Geschichtswissenschaften relevant sind.³⁹

In einer anderen Kooperation haben BURGHARTZ und VOGELER die *”Jahrrechnungen der Stadt Basel”* in den Jahren 1535 bis 1610 digital ediert und im Web veröffentlicht. Jedes Rechnungsbuch beschreibt Einkünfte und Ausgaben der Stadt Basel, wie beispielsweise Einnahmen durch das *Weinungeld*, einer Abgabe an die Stadt, die durch den Ausschank von Wein fällig wurde. Die digitale Edition verfügt über Funktionalitäten einzelne Einträge zu sammeln und Berechnungen damit anzustellen.⁴⁰

”Die digitale Edition der Augsburger Baumeisterbücher”,⁴¹ erlaubt den Zugang zu XML-basierten Transkriptionen der Quellen. Sowohl Digitalisate, wie auch Indices zu Personennamen, Orten und Schlagwörtern sind im Web aufrufbar und durchsuchbar. Eine Normalisierung von Namensformen wird über die GND⁴² sichergestellt.⁴³

Unter Verwendung der *”dynamisch integrierten computergestützten Edition”* hat PERSTLING das Steirische Marchfutterurbar aus den Jahren 1412 bis 1426 so erschlossen, dass unterschiedliche Bearbeitungsschichten des Dokuments als digitale Edition über das Web nachvollziehbar werden.⁴⁴ So wird das Fundament für künftige Erforschungen von beispielsweise Preisen, Münzen oder prosopographischen,⁴⁵ wie auch onomastischen⁴⁶ Studien gelegt. Einen vergleichbaren Zugang verfolgt auch REHBEIN.⁴⁷

³⁹ZIEGLER, Walter: Studien zum Staatshaushalt Bayerns in der zweiten Hälfte des 15. Jahrhunderts. Die regulären Kammereinkünfte Niederbayerns 1450-1500. München, 1981, S.221-230.

⁴⁰VOGELER et al.: The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies, S.11-13, gams.uni-graz.at/srbas.

⁴¹Die Augsburger Baumeisterbücher. Digitale Edition der mittelalterlichen Stadtrechnungen von 1320 bis 1466 augsbuerger-baumeisterbuecher.de

⁴²Gemeinsame Normdatei, dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html

⁴³WÜRZ, Simone: Methoden der Digital Humanities in der Bearbeitung und Erforschung mittelalterlicher Rechnungsbücher. Möglichkeiten und Grenzen am Beispiel der digitalen Edition der Augsburger Stadtrechnungsbücher. In PÄTZOLD, Stefan/STUMPF, Marcus (Hrsg.): Mittelalterliche und frühneuzeitliche Rechnungen als Quellen der landesgeschichtlichen Forschung. Münster, 2016, S.109-113.

⁴⁴PERSTLING, Matthias: Multimediale Dokumentation und Edition mehrschichtiger Texte : das steirisch-landesfürstliche Marchfutterurbar von 1414/1426. Dissertation, Universität Graz, 2013.

⁴⁵Bezeichnet in der Geschichtswissenschaft die systematische Erforschung eines bestimmten Personenkreises.

⁴⁶Bezeichnet die Namensforschung, die die Verbreitung, Bedeutung und Herkunft von Eigennamen erforscht.

⁴⁷REHBEIN, M.: Reconstructing the textual evolution of a medieval manuscript. Literary and Linguistic Computing, 24 08 2009.

1.2.3 Digitale Arbeiten mit historischen Materialien unter Einsatz von *Web of Data* Technologien

Drei geschichtswissenschaftliche Projekte, die einen *Linked Open Data* (LOD) Zugang verfolgen sind *BiographyNet*, *CEDAR* und das *Dutch Ships and Sailors*-Projekt.⁴⁸

Wie bestehende LOD-Vokabulare wiederverwendet werden können, veranschaulicht *BiographyNet*.⁴⁹ In diesem Projekt werden Daten angereichert, ohne die eigentliche Integrität der ursprünglichen Datenquellen zu beeinträchtigen. Die Wiederverwendung bestehender Vokabulare für Aufgaben wie Provenienz- und Objektmodellierung sparte nicht nur Zeit bei der Erstellung des BiographyNet-Schemas, sondern führte auch zu einer besseren Kompatibilität der Daten mit anderen Quellen. Darüber hinaus ermöglichte die Verwendung von LOD die schrittweise Erweiterung des Datenkorpus.

Das CEDAR-Projekt⁵⁰ zeigt, wie eine große Datenmengen über lange Zeitbereiche, von 1795 bis 1971, aufbereitet werden können. Die Verwendung von LOD-Vokabularen bietet ein einheitliches konzeptionelles Modell, um statistische Beobachtungen, auch historischer Art, zu beschreiben. Zur Darstellung historischer Konzepte werden diese Beobachtungen mit bekannten Klassifikationssystemen der Sozialgeschichte verknüpft, um beispielsweise Berufsbezeichnungen zu normalisieren. Als LOD-Datensatz hat die niederländische Volkszählung ein großes Potenzial, andere Datensätze mit Wohnungs-, Demographie-, und Arbeitsdaten anzureichern.

Zentral bei *Ships and Sailors*⁵¹ ist die Zusammenführung von verschiedene Datensätze zu einer LOD-Sammlung. Die einzelnen Datensätze werden in Zusammenarbeit mit den Geschichtswissenschaftlerinnen separat modelliert, so dass eigenständige Datensätze entstehen. Durch Verknüpfungen zu einem gemeinsamen Metadatenschema sowie direkte datensatzübergreifende Verknüpfungen kann ein gewisses Maß an Integration erreicht werden. Die so entstandenen Datensammlung kann abgefragt und somit erforscht und analysiert werden.

⁴⁸BOER, Víctor de/MERONO PENUELA, A/OCKELOEN, CJ: Linked data for digital history: lessons learned from three case studies. *Anejos de la Revista de Historiografía*, 2016 Nr. 4, S.5-22.

⁴⁹BiographyNet, <http://www.biographynet.nl/>, 28.08.2020

⁵⁰CEDAR | Dutch historical census data on the Semantic Web, <https://www.cedar-project.nl/>, 28.08.2020

⁵¹Dutch Ships and Sailors, <https://2017.semantics.cc/dutch-ships-and-sailors>, 28.08.2020

2 Formale Methoden und Modelle in den Geschichtswissenschaften

Am ersten August des Jahres 1808 hat ein gewisser *James Haley* 1/4 Pfund Pulver, 1 Pfund Munition und 1 Pfund Zucker zum Preis von 2 Schilling und 6 Pence im Laden der *Stagville Plantage* in *North Carolina* käuflich erworben. Diese Art von Information über die Vergangenheit lässt sich in historischen Rechnungsbüchern finden. Dabei ist nicht der Einzeleintrag von großer Bedeutung für Fragestellungen in den Geschichtswissenschaften, sondern die Aggregation vieler Einzelinformationen, um eine Datengrundlage zu schaffen, auf der eine weitere fachwissenschaftliche Auseinandersetzung – eine Interpretation der Vergangenheit – durch Historikerinnen fußt.⁵²

Damit diese Datengrundlage ausreichend groß ist und Aussagen nicht nur auf Ausschnitte eines Quellenkorpus reduziert sind, ist es notwendig, formale Verfahren zur Erschließung, Veröffentlichung und Analyse zu verwenden. Auf historische Rechnungsbücher bezogen, können so Berechnungen angestellt werden, die zeigen, wie sich etwa der Preis einer Ware über eine bestimmte Zeitspanne entwickelt hat, oder welche Waren von welchen Akteuren gekauft und verkauft wurden. Die Menge an unterschiedlichen Fragestellungen ist groß und wird in Kapitel 3 weiter veranschaulicht. Setzt man diese "historischen Fakten" aus der Quelle in einen größeren historischen Zusammenhang und als Diskussionsbeitrag in die fachwissenschaftliche Diskussion, beginnt die "klassische" Arbeit von Historikerinnen.

Im Mittelpunkt dieses Kapitels steht eine Diskussion über **Theorien und Methoden der Geschichtswissenschaften** und einer darauf aufbauenden Auseinandersetzung mit der Entwicklung der Anwendung formaler Verfahren innerhalb des Faches. Es wird der Frage nachgegangen, welche theoretischen und methodischen Herangehensweisen sich in den Geschichtswissenschaften etabliert haben. So stellt sich beispielsweise die Frage was man als "historische Fakten" verstehen kann und wie man sie nachvollziehbar, ohne die Quellen zu verzerren, algorithmisch verarbeiten kann. Bereits in den 1980er Jahren hat man sich im Fachbereich *Historische Fachinformatik* mit diesen Fragen beschäftigt.⁵³

Getrieben von einer analytischen Geschichtswissenschaft, in der die Theoriebildung ein zentrales Moment ist, wird die Anwendung formaler und computergestützter Verfahren zur Verarbeitung von Quellen und eine methodische Diskussion darüber immer wichtiger. Aus den Sozialwissenschaften kommend, in der **Historischen Fachinformatik** im Selbstverständnis einer Hilfswissenschaft weiterentwickelt, erleben formale Methoden eine neue Renaissance. Die **Digital Humanities** und die **Digital History** sind zwei Begriffe die dafür stehen. Ein Kernaspekt

⁵²VOGELER: Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert?, 06.06.2019.

⁵³THALLER: Historical Social Research. Supplement 29 2017, S.260-286.

dieser fachlichen Ausrichtungen ist die Erarbeitung und Anwendung von Modellen, um **historische Forschungsdaten** sinnvoll in eine nutzbare Struktur überführen zu können. So setzt sich dieses Kapitel auch mit einer grundlegenden Erörterung der **Modelltheorie** und Modellbildung in den Geisteswissenschaften bzw. Geschichtswissenschaften auseinander.

2.1 Von der erzählenden zur theorieorientierten Geschichtswissenschaft

*”Ich glaube an diese ganze Theoriebedürftigkeit der Geschichte nicht. Die Historie ist eine Kunst, die auf Kenntnissen beruht, und weiter ist sie gar nichts”*⁵⁴

Führte MANN im Jahr 1979 an, um auszudrücken, dass durch die Einführung damals neuartiger, theoretischer Ansätze die Hauptaufgabe der Geschichtswissenschaften – das Erzählen von Geschichten – aufgeweicht wird. Bereits in den 1960er Jahren forderten Historiker wie MOMMSEN, KOCKA oder WEHLER eine *”Reform der Geschichtswissenschaft jenseits des Historismus”* hin zu einer analytischen und theorieorientierten Geschichtswissenschaft. Ihre Kritik richtet sich an den individualistischen Ansatz des Historismus, der Geschichte als Produkt menschlichen Handelns und nicht als rationales Ergebnis eines gesellschaftlichen Prozesses versteht. Die Kritik richtet sich explizit an dieses zentrale Element des Historismus. Die Errungenschaften der historisch-kritischen Methode, der Quellennähe und der Verwissenschaftlichung des Faches, die sich als *”historische Betrachtung”* als allgemeines Paradigma der Geistes- bzw. Kulturwissenschaften durchsetzte,⁵⁵ werden stets positiv hervorgehoben. Sie fordern eine stärkere Theoriebildung in der Geschichtswissenschaft und orientieren sich mehr und mehr an der Methodik der benachbarten Sozialwissenschaften, wie etwa der Soziologie. Die reine Nacherzählung der Vergangenheit, wie sie MANN noch einforderte, ist im heutigen Verständnis der Geschichtswissenschaft, ohne Anbindung an jegliche theoretische Fragestellungen, nicht mehr denkbar.

Heutzutage ist die Orientierung an Theorien in den Geschichtswissenschaften wichtiger Bestandteil. KOCKA versteht darunter einen expliziten Gebrauch von Modellen, Theorien und Begriffen zur Strukturierung des Gegenstandes.⁵⁶ Je stärker man sich beispielsweise mit historischen Fragestellungen, die soziale oder wirtschaftliche Phänomene untersuchen, wie die Beziehungen zwischen sozialen Klassen oder ökonomisches Wachstum, steigt auch die Notwendigkeit mehr über die Fragestellungen und Theorien der Sozial- und Wirtschaftswissenschaft zu wissen.⁵⁷ So ist ein grundlegendes Verständnis über Modelle der Konjunkturtheorie oder klas-

⁵⁴MANN, Golo: Plädoyer für die historische Erzählung. In KOCKA, Jürgen/NIPPERDEY, Thomas (Hrsg.): Theorie und Erzählung in der Geschichte. München, 1979, S.53.

⁵⁵HARDTWIG: Geschichtskultur und Wissenschaft, S.28.

⁵⁶MAGERSKI, Christine et al.: Wie schreibt man Geschichte? Zagreber Germanistische Beiträge, 2009 Nr. 18, S.2.

⁵⁷KOCKA, Jürgen: Theorien in der Geschichtswissenschaft. In Theoriedebatte und Geschichtsunterricht: Sozialgeschichte, Paradigmawechsel und Geschichtsdidaktik in der aktuellen Diskussion. Vorträge auf der Jahrestagung des Landesverbandes nordrhein-westfälischer Geschichtslehrer am 23. November 1981 in Essen Paderborn:

sentheoretische Konzepte der Sozialgeschichte unabdingbar.⁵⁸ Der vermehrte Einsatz formaler Verfahren zur Erschließung, Auswertung und Darstellung von Information in historischen Quellen, befördert diese Notwendigkeit. Gleichzeitig ist aber auch ein naiver Positivismus, der davon ausgeht absolute Wahrheiten aus den Quellen alleine abzuleiten oder gar "auszurechnen", ein Weg, den eine "sinnvolle" Geschichtswissenschaft ablehnt.

2.1.1 Geschichte und Geschichtswissenschaft: eine Definition

*"Die Geschichte ist das Wissen der Menschen von sich, ihre Selbstgewißheit"*⁵⁹

Dieser Satz aus "Grundriss der Historik" von DROYSEN bringt die Aufgabe von Geschichte auf den Punkt: so wie jedes Individuum sich selbst gewahr ist, so ist der Gegenstandsbereich der **Geschichte** das Handeln der Menschen in der Vergangenheit. Ein Handeln, das jeden einzelnen und die Gemeinschaft erst zu dem gemacht hat was sie sind. Dieses Handeln ist aber für niemanden mehr erlebbar, sondern muss über Belege der Vergangenheit, den Quellen, ob als Überrest oder Tradition, rekonstruiert werden. Die Unterscheidung erfolgt nach DROYSEN zwischen "unabsichtlich" erzeugten Überresten, wie etwa ein Kleidungsstück, das bei einer archäologischen Ausgrabung gefunden wurde, und absichtlich überlieferten Quellen, wie etwa ein Denkmal. Letztere werden als Tradition verstanden.⁶⁰ Bei dieser Einteilung handelt es sich aber um ein Ideal, da für eine genaue Zuordnung das Wissen über die Intention der Schöpferinnen notwendig wäre.

KIRN führt in seinem Einführungswerk in die Geschichtswissenschaft mehrere Definitionen des Begriffes Geschichte an: Wo HELLPACH über *"die bewußte Gestaltung menschlichen Gemeinschaftslebens aus schöpferischem Willen"* spricht und die menschliche Handlung in den Vordergrund stellt, ist es für HUIZINGA, aus dem Verständnis von Geschichtswissenschaft als Kulturgeschichte, *"die geistige Form, in der sich eine Kultur über ihre Vergangenheit Rechenschaft gibt"*. Zentral in dieser Menge an Definitionen und Zugängen zur Geschichte ist, dass es sich um die wissenschaftliche Darstellung des Vergangenen handelt.⁶¹ Zweiter zentraler Aspekt ist, dass der Mensch als gesellschaftliches und kulturelles Wesen etwas über sich selbst, seine vergangenen Gesellschaft- und Kulturformen wissen möchte. Man könnte es als einen Akt der kollektiven Selbstreflexion verstehen. Genau das meint das eingangs angeführte Zitat von DROYSEN.

Schöningh, 1982, S.6-8.

⁵⁸SCHMIEDER, Felicitas/SOKOLL, Thomas: Theorien und Methoden der Geschichtswissenschaft., S.1.

⁵⁹HARDTWIG, Wolfgang: Über das Studium der Geschichte. Deutscher Taschenbuch Verlag, 1990, S.111.

⁶⁰OPGENOORTH, Ernst/SCHULZ, Günther: Einführung in das Studium der Neueren Geschichte. Paderborn / München / Wien / Zürich, 2010, 49–55.

⁶¹KIRN, Paul: Einführung in die Geschichtswissenschaft. Berlin, Boston, 2015, S.7-12.

Jedenfalls ist der Begriff mehrdeutig; Geschichte kann als **Abfolge von realen Tatsachen und Ereignissen, Geschichtsforschung** (also der Darstellung dieser Realgeschichte), oder **Geschichtsschreibung** (die wissenschaftliche Disziplin), verstanden werden.⁶² Geschichtsschreibung als Produkt der Geschichtsforschung liegt üblicherweise als Narrativ über vergangene Ereignisse vor. Sie ist jedoch nicht auf die Textform beschränkt, sondern kann auch in anderer Form vorliegen. Dies umfasst mündliche, wie bildliche Überlieferungen oder sämtliche Schöpfungen, die von Menschen erbracht wurden, um in einen Kommunikationsprozess einzutreten.⁶³

Bei DEMANDT finden sich weitere Ausführungen zu den grundlegenden Überlegungen des Begriffes *Geschichte*. Neben dem Mythos, der in vorschriftlichen Zeiten und in erzählerischer Form Vergangenes von einer Generation zu anderen weitergibt, war die Chronistik eine frühe Form der verschriftlichten Erinnerung. Chroniken berichten über Ereignisse, ohne die Texte literarisch aufzuarbeiten. Die **Historie**, die *Erkundung der Vergangenheit*, geht noch einen Schritt weiter und versucht Ereigniszusammenhänge einer "wahren" Vergangenheit darzustellen. Sie etabliert sich als autonome Disziplin, die mehr den Wandel der Geschichte und das Geschehene an sich versucht zu erklären, denn moralisch und belehrend auf Rezipierende zu wirken, wie es die Geschichtsüberlieferung davor getan hat.⁶⁴

Ein Zitat von SCHILLER fasst dieses neue Paradigma der Geschichtswissenschaft zusammen, welches bis heute gilt: "*Das Verhältnis eines historischen Datums zu der heutigen Weltverfassung ist es also, worauf gesehen werden muss*". Geschichtswissenschaft, ist erst dann vollkommen, wenn sie mit der Gegenwart in Beziehung gesetzt wird und umgekehrt.⁶⁵

Das moderne Verständnis des Wortes Geschichte umfasst sowohl die Gesamtheit von realen Geschehnissen, als auch den Bericht und die Reflexion darüber. Zentral in der **modernen Geschichtswissenschaft**, also der Erforschung der Vergangenheit nach wissenschaftlichen Kriterien, ist die Auseinandersetzung mit historischen Quellen, die mittels kritischer Methode untersucht werden. In diesem Prozess müssen die angewandten Methoden, Ideen und Ergebnisse diskutierbar und nachprüfbar sein.⁶⁶ Jede und jeder soll in der Lage sein, feststellen zu können, ob eine Interpretation eines historischen Sachverhalts durch eine angeführten Quelle belegbar ist. Forschungsgegenstand und Forschungsfrage sind in den Geschichtswissenschaften nicht klar abgegrenzt und die Interessen von Historikerinnen sind interdisziplinär, weswegen auch die angewandte Methodik vielfältig ist.⁶⁷ Auf die Methoden in den Geschichtswissenschaften wird in Kapitel 2.1.4 detaillierter eingegangen.

⁶²GOTTFRIED, Gabriel: Fakten oder Fiktion? Zum Erkenntniswert der Geschichte. Oldenbourg, S.1.

⁶³FRANK, Ingo: Visualisierungswerkzeuge zur Erklärung historischer Ereignisse: geschichtstheoretische Anforderungsanalyse und zeichentheoretisches Rahmenwerk. INF-DH-2018 2018, S.5-7.

⁶⁴HARDTWIG: Geschichtskultur und Wissenschaft, S.24.

⁶⁵ebd., S.25.

⁶⁶DEMANDT, Alexander: Philosophie der Geschichte von der Antike zur Gegenwart. 2011, S.13-32.

⁶⁷REICHE, Ruth/BECKER, Rainer/BENDER, Michael: Verfahren der Digital Humanities in den Geistes- und Kulturwissenschaften. Niedersächsische Staats- und Universitätsbibliothek Göttingen, 2014, S.13.

Zwei grundlegende Strömungen in der Erkenntnis- und Wissenschaftstheorie sind jedoch wichtig und werden im Folgenden weiter ausgeführt: die Hermeneutik und der Empirismus.

2.1.2 Hermeneutik

*”Hermeneutik ist das Nachdenken darüber, wie Wissen zustande kommt und wie es plausibel vermittelt werden kann”*⁶⁸

Bei der Hermeneutik handelt es sich um eine Theorie der Interpretation und des Verstehens von Texten, die nicht nur auf die Geschichtswissenschaft angewendet wird, sondern auf viele geisteswissenschaftliche oder rechtswissenschaftliche Fächer. DILTHEY betont, *”dass das Verstehen [...] die spezifische Operation des Geisteswissen darstellt.”*⁶⁹

Im großen Unterschied zu naturwissenschaftlichen Fragestellungen, ist die Vergangenheit nicht in einem Experiment beweis- bzw. widerlegbar. Geprägt wird die Hermeneutik, als Theorie für die Geschichtswissenschaften, unter anderen, von den Arbeiten von SCHLEIERMACHER, der sie zu einer *universellen Theorie des Verstehens* weiterentwickelte, in der jedes Verstehen einen individuellen Prozess darstellt. Für SCHLEIERMACHER ist gerade die Sprache, als Ausdruck der Gedanken von Individuen, wichtig.⁷⁰

Aus dem neuen Geschichtsverständnis des 19. Jahrhunderts entwickelte DILTHEY die Hermeneutik zu einer historischen Methodik. Bei ihm wird Verstehen zur essentiellen Methode in den Geisteswissenschaften. Wo Naturwissenschaften die Welt von außen betrachten und auf Gesetzmäßigkeiten schließen, beispielsweise durch Messverfahren, so ist die Aufgabe der Geisteswissenschaften, die Welt von innen zu verstehen.⁷¹

Unter GADAMER wird die Hermeneutik zur Philosophie weiterentwickelt. Es bedarf immer einem Vorverständnis, um in einen Prozess des Verstehens eintauchen zu können, an dessen Ende nur eine Annäherung an eine Wahrheit möglich ist.⁷² Voraussetzung dafür ist die Frage nach der Anwendung und Konkretisierung der Fragestellung, die GADAMER als grundsätzliche Verwandtschaft von juristischer und hermeneutischer Rationalität beschreibt.⁷³

Nicht nur das Wissen über historische Quellen, sondern alles Wissen beruht auf einem Verstehensprozess und ist in diesem Sinne zirkelförmig. Jemand tritt mit einem gewissen Vorwissen an eine Fragestellung heran. Es werden in der Auseinandersetzung mit der Materie Erfahrun-

⁶⁸Digitale Hermeneutik - Ein Update für das kritische Denken in den Geisteswissenschaften, <https://tinyurl.com/y3opg5bs>, 27.08.2020

⁶⁹FICARA, Elena: Texte zur Hermeneutik. Von Plato bis heute. Reclam, 2015, S.82.

⁷⁰ebd., S.72-81.

⁷¹ebd., S.82-83.

⁷²OPGENOORTH/SCHULZ: Einführung in das Studium der Neueren Geschichte, S.19-30.

⁷³FICARA: Texte zur Hermeneutik. Von Plato bis heute, S.176-178.

gen gesammelt und ein tieferes Verständnis der Thematik erarbeitet. Mit diesem gefestigten Verständnis bewegt man sich im sogenannten *hermeneutischen Zirkel*, einem iterativen Prozess gleich, weiter fort und erhält auf diese Weise ein gefestigtes Verständnis zu einem Thema. Eine Problemstellung muss zuerst gefunden, erkannt und verstanden werden, damit sie weiter verarbeitet werden kann.⁷⁴

In folgendem Beispiel sind die Wörter „*vihe*“, „*verwürckt*“ oder „*fewer*“ ohne jeglichen Kontext unverständlich. Werden sie aber in ihrem Kontext (der historischen Quelle, dem Wissen über die Quelle und über die Zeit in der die Quelle verortet wird) gelesen, wird die Bedeutung des Textes offensichtlich:

*"Item so eyn mensch mit eynem vihe, mann mit mann, weib mit weib, vnkeusch treiben, die haben auch das leinen verwürckt, vnd man soll sie der gemeynen gewonheyt nach mit dem fewer vom leben zum todt richten"*⁷⁵

Zentrale Begriffe lassen sich häufig erst aus dem Zusammenhang erschließen. Für das vollständige Verstehen des Gesamttextes müssen alle Einzelheiten nachvollziehbar sein. Die hermeneutische Spirale beinhaltet, dass Teilstücke ganzheitlich verstanden oder erweitert werden und sich das Ganze von den Teilaspekten her definieren lässt. Gerade in den Geschichtswissenschaften ist die Berücksichtigung und das Verstehen des Kontextes, in dem eine Aussage über die Vergangenheit steht, essentiell.

Werden hermeneutischer Verfahren auf historisches Quellenmaterial angewendet, spricht man von einem historischen Verstehen, mit dem Ziel Geschichte und Vergangenheit in Relation zu setzen. Es ist der reflexive Prozess sich in eine Schöpferinnen eines geistigen Produktes, etwa einem historischen Rechnungsbuch, hineinzusetzen. So sollen Erfahrungen, Denkweisen oder gar Gefühle reproduziert werden, um Handlungen in der Geschichte verständlich erscheinen zu lassen.⁷⁶ Für die Zusammenführung und Interpretation historischer Quellen sind diese Schritte unerlässlich.

2.1.3 Empirismus

Wissenschaften, die Sachverhalte und Objekte der Welt durch Experimente im Labor, Beobachtung oder Befragung im Feld untersuchen, werden als empirische Wissenschaften zusammengefasst. Als sogenannte Erfahrungswissenschaft wird Erkenntnis aus der Erfahrung der Welt abgeleitet. Die klassischen Vertreter sind etwa die Disziplinen der Naturwissenschaften, wie

⁷⁴MANTZAVINOS, C., "Hermeneutics", The Stanford Encyclopedia of Philosophy (Spring 2020 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2020/entries/hermeneutics/>.

⁷⁵Beispiel aus "Peinliche Halsgerichtsordnung Kaiser Karls V." von 1532, <http://ra.smixx.de/media/files/Constitutio-Criminalis-Carolina-1532.pdf>, 15.10.2020.

⁷⁶DETEL, Wolfgang: Geist und Verstehen. Historische Grundlagen einer modernen Hermeneutik. Philosophische Abhandlungen Band 104, 2011, S.14-20.

die Physik oder Biologie. Die Mathematik, die Philosophie oder auch die Informatik wiederum sind keine empirischen Wissenschaften. In diesen Disziplinen findet keine direkte Beobachtung statt, sondern Aussagen basieren auf einer logischen bzw. formalen Herangehensweise.⁷⁷ Abbildung 1 zeigt die zentralen Begriffe der Empirie und ihre Wechselwirkung. Als Empirie wird die methodisch-systematische Sammlung von Daten verstanden, die zu einer Hypothese oder deren Widerlegung führt. Die **Deduktion** versteht die Schlussfolgerung gegebener Prämissen auf die logisch zwingenden Konsequenzen und die **Induktion** beschreibt den abstrahierenden Schluss aus beobachteten Phänomenen auf eine allgemeinere Erkenntnis.⁷⁸ Oder anders formuliert: Schlussfolgerung vom Allgemeinen auf das Besondere und umgekehrt. Für die Geschichtswissenschaften bedeutet das ausgehend von einer Forschungsfrage verschaffen sich so Historikerinnen einen kritischen Überblick über vorhandenes Quellenmaterial, entwickeln eine These und überprüfen diese anhand des Materials.⁷⁹

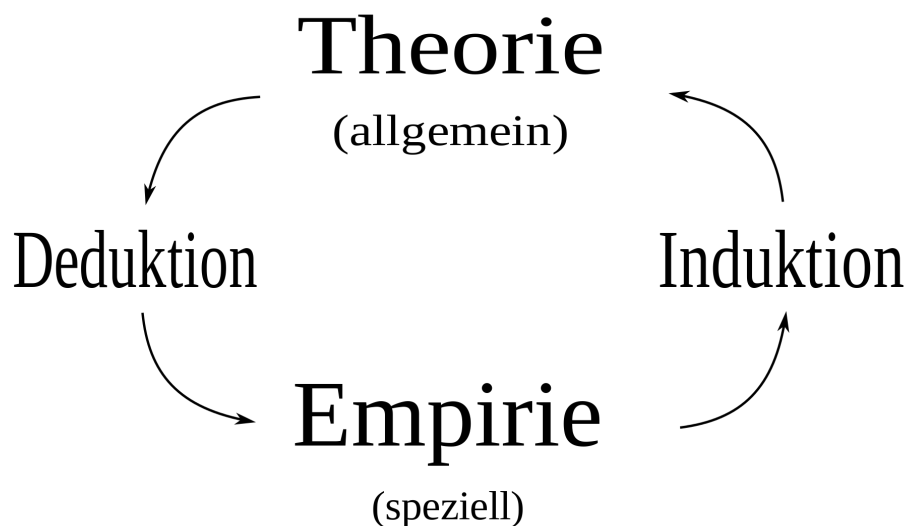


Abbildung 1: Zusammenhang zwischen Induktion und Deduktion in empirischen Wissenschaften

In der Geschichtswissenschaft werden Textquellen mittels hermeneutischer Methoden interpretiert. Jedoch kann die **Geschichtswissenschaft auch als empirische Wissenschaft verstanden werden**. Nämlich dann, wenn empirische Daten, aus den überlieferten und erschlossenen (edierten) Quellen, im Rahmen einer Hypothese hinsichtlich einer historischen Rekonstruktion der Vergangenheit herangezogen werden. Wichtig ist zu berücksichtigen, dass es keine absolute Reproduktion der Ergebnisse geben kann, keine Experimente möglich sind und es sich immer um einmalige Untersuchungsgegenstände handelt und damit eine Generalisierung der Art "immer

⁷⁷FROMMER, Sören: Grundbegriffe aus Theorie, Methode und Wissenschaftsgeschichte. (URL: www.historische-archaeologie.de/downloads/grundbegriffe.pdf), 08.11.2019.

⁷⁸Siehe CHALMERS, Alan F: Wege der Wissenschaft: Einführung in die Wissenschaftstheorie. Springer, 2007.

⁷⁹JORDAN: Theorien und Methoden der Geschichtswissenschaft, S.45-46.

wenn A eintritt, dann folgt B" unmöglich ist. Vertreter einer einheitswissenschaftlichen Position wie HEMPEL aber definieren die Geschichtswissenschaft als Gesetzeswissenschaft,⁸⁰ wohingegen DILTHEY und WRIGHT den besonderen Charakter hermeneutisch vorgehender Wissenschaften hervorheben. Sie setzen sich gegen eine "blinde" Übernahme naturwissenschaftlicher Erkenntnismodelle auf die Geistes- und Sozialwissenschaften ein. Das Verhältnis dieser beiden wissenschaftlichen Ausrichtungen ist bis heute kontrovers.⁸¹ Aber gerade im Zusammenspiel zwischen empirischer Methoden und der Herangehensweise an Forschungsfragen im Sinne der Hermeneutik kann ein Mehrwert für die Bildung von Hypothesen und die Interpretation eines Themas entstehen. Im Zuge der empirischen Überprüfung von Sachverhalten spielt die Hermeneutik eine wesentliche Rolle bei der Quantifizierung von qualitativen Aussagen. Auch die Interpretation von empirischen Resultaten ist ein hermeneutischer Vorgang.

Die Hermeneutik und die Empirie sind zwei unterschiedliche Herangehensweisen wie Erkenntnis gewonnen werden kann. Bei ersterem steht das Verstehen und bei zweiterem das Erklären im Vordergrund. Oft wird empirische Erkenntnis als "objektiver" und hermeneutische Erkenntnis als "subjektiver" dargestellt. DANIEL⁸² diskutiert in diesem Zusammenhang die Begriffe "Objektivität" und "Subjektivität" in der Kulturgeschichte. Der Anspruch an Objektivität ist unsinnig, also "*wie es wirklich gewesen sei*", da es unmöglich ist die historische Wirklichkeit zu erfüllen. Ganz explizit schreibt THALLER auch, in Bezug auf die Umsetzung von Informationssystemen in den Geschichtswissenschaften, "*zu zeigen, 'mit welcher Wahrscheinlichkeit es unter welchen Voraussetzungen wie gewesen sein könnte'*".⁸³ Vielmehr entsteht "objektives" historisches Wissens aus dem Diskutieren und Nachdenken über subjektive Werturteile und Motive und fließt in die historische Interpretation ein.

2.1.4 Methoden der Geschichtswissenschaft

Ein kontrolliertes Verfahren zur Erreichung eines Forschungsziel bzw. zur Gewinnung von Erkenntnis wird als Methode bezeichnet. Die Methodologie beschäftigt sich damit, ob eine gewählte Methode für ein bestimmtes Ziel geeignet ist, oder nicht.⁸⁴ Sie sind wichtiger Bestandteil der wissenschaftlichen Arbeit und somit auch ein nicht wegzudenkendes Werkzeug in den Geschichtswissenschaften, wie ein Blick auf die Vielfalt **historischer Hilfswissenschaften**, deren Aufgabe die Erschließung von bestimmten Quellentypen ist, zeigt. Die Hilfswissenschaften bzw. Grundwissenschaften, wie sie auch genannt werden – "*als Abzweigungen der allgemei-*

⁸⁰NÜSSEL, Friederike et al.: Offenbarung als Geschichte: Implikationen und Konsequenzen eines theologischen Programms. Vandenhoeck & Ruprecht, 2018, S.79-80.

⁸¹WELLMER, Albrecht: Georg Henrik von Wright über "Erklären" und "Verstehen". 1979 (URL: <https://www.jstor.org/stable/42571419>).

⁸²UTE, Daniel: Kompendium Kulturgeschichte. Theorien, Praxis, Schlüsselwörter. 5. Auflage. Frankfurt am Main, S.390-400.

⁸³THALLER: Historische Sozialforschung. Supplement 29 2017, S.159.

⁸⁴BLAKSTAD Oskar: Die wissenschaftliche Methode, explorable.com/de/die-wissenschaftliche-methode, 28.12.2019.

nen historischen Arbeitsform“⁸⁵ – haben das Ziel die historischen Quellen aufzubereiten. Sie erstrecken sich von der Paläografie, Kodikologie über die Diplomatik, Insignienkunde und Aktenkunde bis hin zur Genealogie, um nur einige Vertreter zu nennen. Charakterisiert sind die historischen Hilfswissenschaften durch ihren Bezug zur Quelle. Die äußeren Merkmale einer Quelle (Tinte, Papier, Schrift etc.) stehen den inneren Merkmalen gegenüber und es ist gemeinsames Ziel einer jeden Hilfswissenschaft zu überprüfen, ob eine Quelle authentisch ist, oder nicht. Gerade diese Frage nach der Authentizität der Quelle bzw. der Integrität der Daten ist essentieller Bestandteil der historischen Forschung.⁸⁶

Die **Historische Fachinformatik**, sozusagen als Vorgänger der Digital History, versteht sich auch als Hilfswissenschaft innerhalb der Geschichtswissenschaften und beschäftigt sich mit der Anwendung und Reflexion von formalen Verfahren im Fach. Sie ist auf der Suche nach in Algorithmen ausdrückbaren, sprich formalisierbaren, Verfahren der Geschichtswissenschaft im Umgang mit historischen Quellen.⁸⁷

Neben Methoden der Erschließung historischer Quellen in ihrer Materialisierung als historische Hilfswissenschaften ist die **Quellenkritik** immanente Methode der Geschichtswissenschaft. Die Quellenkritik entwickelte sich aus dem Dreischritt von Heuristik, Kritik und Interpretation. Historikerinnen formulieren eine Fragestellung, als Ergebnis eines spezifischen Forschungsstandes, um eigene bzw. neue Erkenntnisse über eine Forschungsfrage zu gewinnen. Relevante Quellen werden gesammelt, erschlossen und einer kritischen Prüfung auf Vollständigkeit, Glaubwürdigkeit und Echtheit unterzogen. Die Ergebnisse aus dem Quellenstudium werden mit den Erkenntnissen von Fachkollegeninnen verglichen und in übergeordnete Kontexte eingeordnet.⁸⁸ Für DROYSEN mündet die Reflexion auf die historische Methode in eine Theorie der historischen Erkenntnis: der Hermeneutik. Aus diesem Verständnis heraus entstand auch die Anwendung der Hermeneutik als Methode der Geschichtswissenschaft.

Unterschieden werden qualitative und quantitative Methoden. **Qualitative Methoden** verfolgen im Gegensatz zu quantitativen Methoden einen offeneren und flexibleren Zugang zum Forschungsgegenstand. Es gibt Platz, dass neue Phänomene entdeckt werden können. Der Forschungsprozess wird als dynamisch aufgefasst und es werden Methoden eingesetzt wie Interviews, Inhaltsanalysen und Beobachtungen. Die **quantitativen Methoden** sind statisch und folgen einem festgelegten Muster. Zu Beginn des Forschungsprozesses müssen Theorien und Modelle über den Gegenstand der Forschung vorliegen. Es werden Hypothesen abgeleitet, die

⁸⁵VON BRANDT, Ahasver: Werkzeug des Historikers: eine Einführung in die historischen Hilfswissenschaften. Band 33, W. Kohlhammer Verlag, 2007, S.10.

⁸⁶FICKERS, Andreas: Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik? Zeithistorische Forschungen, 17 2020 Nr. 1 (URL: <http://hdl.handle.net/10993/44086>), S.7.

⁸⁷Was ist HFI?, Kurzbeschreibung des Faches und seiner Positionierung hfi.uni-graz.at/basisinformation/was-ist-hfi, 23.05.2019.

⁸⁸JORDAN: Theorien und Methoden der Geschichtswissenschaft, S.43-45.

im Forschungsprozess überprüft werden. Hierzu werden Überlegungen angestellt, welche Indikatoren für Forschungsfragen sinnvollerweise messbar sind und die Daten über statistische Verfahren ausgewertet. Die so "berechneten" Erkenntnisse werden abschließend wieder auf das theoretische Modell bezogen und interpretiert.⁸⁹ Die Geschichtswissenschaft setzt allgemeine Hypothesen zur Erklärung spezifischer Entwicklungen ein und die Sozialwissenschaften nutzen Daten zur Formulierung von allgemeinen Gesetzmäßigkeiten. Der Zweck quantitativer und qualitativer Methoden in der Geschichtswissenschaft, die stets als Mittel zum Zweck zu betrachten sind, ist es, Erkenntnis über die Vergangenheit zu gewinnen.⁹⁰

⁸⁹WOLF, Willi; KÖNIG, E./ZEDLER, P. (Hrsg.): Qualitative versus quantitative Forschung. Band Bd. I: Grundlagen qualitativer Forschung, Beltz, 1995, S.309–329.

⁹⁰JARAUSCH, Konrad Hugo/ARMINGER, Gerhard/THALLER, Manfred: Quantitative Methoden in der Geschichtswissenschaft. Wissenschaftliche Buchgesellschaft, 1985, S.203-206.

2.2 "Quantifizierer" und "Traditionalisten": Entwicklung einer formalen Methodik

*"Aber alle Autoren halten daran fest, dass sich die Quantifizierung als ein gewichtiges Werkzeug historischer Analyse erwiesen habe."*⁹¹

In den 1980iger Jahren standen sich in der Diskussion über den methodischen Zugang zu historischen Quellen zwei Gruppen gegenüber, die sich als "Traditionalisten" und "Quantifizierer" festmachen lassen können. Im Gegensatz zu den "Traditionalisten", die einen hermeneutischen Zugang wählten, um historische Quellen zu verstehen, fokussierten die "Quantifizierer" formale Methoden, wie etwa statistische Verfahren aus den Sozialwissenschaften, um sie auf Quellenkorpora anzuwenden und die daraus gewonnenen empirischen Fakten für ihre Interpretation zu nutzen.⁹² SCHRÖDER definiert die historischen Sozialwissenschaften als *"theoretisch und methodisch reflektierte, empirische, besonders auch quantitativ gestützte Erforschung sozialer Strukturen und Prozesse in der Geschichte."*⁹³ Die Forschung in der historischen Sozialwissenschaft setzt sich aus folgenden Schritten zusammen:⁹⁴

- Problemauswahl und -formulierung.
- Theorie und Hypothesenbildung
- Methoden und Quellenauswahl (Heuristik)
- Datenerhebung, Verarbeitung, Analyse und Repräsentation

Ergebnis dieser Auseinandersetzung war es, dass im Zusammenhang mit der Anwendung formaler Methoden besonders auf die Nachvollziehbarkeit geachtet werden muss, damit nicht Dinge, die nicht empirisch beweisbar sind, auch nicht so missverstanden werden können. Aus diesem Grund muss die Quelle ohne jegliche Vorannahmen zur Verfügung gestellt werden, einsehbar sein und das angewandte formale Modell bzw. die Methode in ihrer Gänze offen gelegt werden. Dies gilt grundsätzlich für jede Art wissenschaftlicher Arbeit.

Darum wird empfohlen die grundlegenden Annahmen und Definitionen, die in der Verarbeitung und anschließenden Interpretation von Quellen verwendet werden in einer, und so haben es bereits Vertreter der Historischen Fachinformatik gefordert, gemeinsamen *knowledge domain* zu formalisieren.⁹⁵ Diese *knowledge domain* beschreibt die interne Komplexität einer Fragestellung.

⁹¹JARAUSCH/ARMINGER/THALLER: Quantitative Methoden in der Geschichtswissenschaft, S.191-206.

⁹²ebd.

⁹³SCHRÖDER, Wilhelm Heinz: Historische Sozialforschung: Forschungsstrategie Infrastruktur Auswahlbibliographie. Historical Social Research 1988 Nr. 1, S.5.

⁹⁴ebd., S.5-8.

⁹⁵THALLER: Historical Social Research. Supplement 29 2017, S.263.

Die Konzepte, Standards und Technologien des *Web of Data* (aka. Semantic Web) und der *Linked Open Data* Paradigmen haben zum Ziel Wissensdomänen als konzeptuelle, maschinenlesbare Modelle zu formalisieren und auf diese Weise strukturierte Daten über das Web zur Verfügung zu stellen. Diese können dazu dienen, die von THALLER geforderten *knowledge domain* für historische Fragestellungen in die Realität umzusetzen. Auf diese Aspekte wird später in Kapitel 4 und 5 weiter eingegangen. Jedenfalls können diese Technologien ein Weg sein, um Information in einer Quelle so zu beschreiben, dass sie als "historische Fakten" in einem Informationssystem weiter verarbeitet werden können.

2.2.1 "Historische Fakten"

Algorithmen verarbeiten nur einen Input nach einem vordefinierten Programm. Die Logik dieser Verarbeitung innerhalb der binären Logik von Rechenmaschinen ist "wahr" oder "falsch" (0 oder 1). Die Information, die in historischen Quellen zu finden ist, kann aber nicht so einfach einem absoluten Wahrheitswert zugeordnet werden.

Der Begriff "historische Fakten" wird in der Fachliteratur ausführlich aus vielfältigen Blickwinkeln diskutiert. Für manche Historikerinnen sind sie gegeben, wenn zur Rekonstruktion der Vergangenheit Sachverhalte in Quellen herangezogen werden, die ohne eigenen soziokulturellen Hintergrund interpretiert werden. Information in den Quellen wird über den geschichtswissenschaftlichen Diskurs zu einem "historischen Faktum" transformiert. Ein gegensätzlicher Standpunkt ist, dass ein Faktum unabhängig von der Arbeit der Historikerinnen existiert und die historische Wirklichkeit nicht von diesen überprüft oder geschaffen werden muss, sie sondern nur "gefunden" werden muss.⁹⁶

Als "**Vetorecht der Quellen**"⁹⁷ wird ein geschichtstheoretische Denkmodell bezeichnet, "*nach der quellenkritischen Deutung historischer Quellen die Funktion zukommt, historisch unwahre Aussagen als solche kenntlich werden zu lassen*". Der Begriff stammt von KOSELLECK und hat sich in der Auseinandersetzung zwischen Verfechterinnen einer objektiven Geschichtswissenschaft und historischen Materialistinnen entwickelt. Diese Herangehensweise folgt dem Falsifizierungsprinzip von POPPER und beschreibt, dass eine historische Quelle, sofern sie die Kriterien erfüllt (Quellenkritik), eine historische Interpretation widerlegen kann. Das "Vetorecht der Quelle" ist die Kritik an einem naiven Objektivismus, nach dem historische Fakten ohne die Perspektivität der Historikerinnen und den soziokulturellen Kontexten in denen sie stehen, verstanden und dargestellt werden können.

⁹⁶EVANS, Richard J: Fakten und Fiktionen: über die Grundlagen historischer Erkenntnis. Campus-Verlag, 1998.

⁹⁷JORDAN, Stefan: Vetorecht der Quellen. 2010 (URL: <http://dx.doi.org/10.14765/zzf.dok.2.570.v1>).

Bei DANIEL wird die historische Herkunft der Begriffe *Tatsache*, *Objekt* und *Wahrheit* erörtert.⁹⁸ **Tatsachen** bestehen immer aus einer handelten und zeitlichen Dimension und bezeichnen ausdrücklich, im Kontext „*historisches Faktum*“, einen geschichtlichen Sachverhalt. Als Tatsachen versteht sie (vorläufige!) Resultate der Forschung. Unter **Objekten** wird verstanden, was einer Betrachterin gegenübersteht. Also ein Buch das angefasst werden kann. Unter **Wahrheit** wird aufgefasst, was als wissenswert, richtig und begründungsfähig gehalten wird und wofür jemand eine gewisse Verantwortung übernimmt. Darin steckt kein absoluter, positivistischer Wahrheitsanspruch.

2.2.2 Interpretation und Theoriebildung

Die Interpretation von neu gewonnenen Ergebnissen ist das eigentliche Ziel der quantitativen Methode. Quantitative Methoden ermöglichen nicht nur einen punktuellen Einblick in einen Quellenbestand, sondern versetzen die Historikerinnen in die Lage größere Mengen von Belegen zu verarbeiten. Genau darin besteht der Mehrwert dieser Methodik aber auch ihre Herausforderung. Es besteht die Gefahr, so JARAUSCH et. al., dass die eigentliche Arbeit, die Interpretation der Quellen, im reinen quantitativen Verarbeiten und Verwalten liegen bleibt. Der Aussagewert der Ergebnisse aus diesen Methoden ist abhängig von einer ganzen Reihe von qualitativen Entscheidungen. Denkt man beispielsweise an die Einordnung von historischen Berufen in einzelne Gruppen, um diese besser gegenüber stellen zu können, so bedarf es einer ausführlichen Argumentation, wie sich diese Gruppen zusammenstellen. Fehlerhafte Vorbedingungen führen zu fehlerhaften Ergebnissen, die wiederum verzerrte Interpretationen mit sich führen können. Gehört ein Kaufmann in die Gruppe des Besitzbürgertums, oder zum alten Mittelstand? Es ist notwendig, dass das Erkenntnisziel eines Verfahrens im Vorhinein definiert wird und das Verhältnis der quantitativen Daten zu den qualitativen Quellen geklärt und diskutiert wird.

Gerade Information aus historischen Quellen ist stark kontextabhängig.⁹⁹ Bereits 1988 veranschaulichte THALLER das in seinem sogenannten „**Preußenbeispiel**“. Der politische bzw. geographische Begriff „Preußen“ bezieht sich im Jahre 1670 auf deutlich andere Koordinaten, als im Jahre 1770. Wer als „Preuße“ zu kategorisieren ist, ist abhängig zu welcher Zeit und an welchem Ort eine Person lebte bzw. geboren wurde.¹⁰⁰ Abgehoben von dieser Frage ist dann, ob jemand sich auch als „Preuße“ verstanden hat. Historische Information ist unscharf, unvollständig, heterogen und mehrdeutig. Die korrekte Interpretation solcher Daten hängt stark von den sich verändernden Parametern ab. Ein Grundsatz ist, dass nicht „alles mit allem“ verarbeitet wird, in der Hoffnung so auf noch ungeahnte Erkenntnisse zu stoßen. Es müssen bereits im Vorhinein Überlegungen angestellt werden, inwieweit die innere Logik eines Verfahrens auch

⁹⁸UTE: Kompendium Kulturgeschichte. Theorien, Praxis, Schlüsselwörter, S.381-398.

⁹⁹JARAUSCH/ARMINGER/THALLER: Quantitative Methoden in der Geschichtswissenschaft, S.182-193.

¹⁰⁰THALLER: Historical Social Research. Supplement 29 2017, S.264-266.

geeignet ist um passende Resultate zu erzeugen, in denen Verarbeitungsfehler bzw. falsche Annahmen ausgeschlossen werden können. Ein statistischer Zusammenhang heißt noch keine Kausalbeziehung: das Aussterben der Störche kann statistisch einen Einfluss auf den Niedergang der Geburtenrate suggerieren (Korrelation), einen kausalen Zusammenhang gibt es in diesem Beispiel aber nicht (Kausalität). Statistik kann nichts erklären, sondern nur Erklärungsmodelle auf ihre Übereinstimmung mit den Daten überprüfen. Es ist wichtig aus der statistischen Auswertung aufzutauchen und deren Befunde mit der Fachliteratur zu diskutieren. Nur so kann man den aktuellen Wissensstand mit dem alten vergleichen und neue Thesen aufstellen.¹⁰¹ Im praktischen Teil dieser Arbeit wird nicht explizit auf Statistik eingegangen, sondern auf die Informationsvisualisierung als Produkt bzw. eine Repräsentationsform von mathematischen und statistischen Verfahren.

Aus diesem Grund ist auch die **Theoriebildung für die Anwendung quantitativer Methodik** in den Geschichtswissenschaften relevant. KOCKA definiert die Theorie als einen expliziten und konsistenten Satz von verwandten Begriffen, die benützt werden können, historische Zusammenhänge, die nicht aus dem Studium der Quellen alleine abgeleitet werden können, zu strukturieren und zu erklären. So unterscheidet KOCKA sechs Funktionen von Theorie:¹⁰²

- Eine Theorie hilft bei forschungsstrategischen Entscheidungen und grenzt eine Forschungsfrage von anderen ab. Dies erleichtert die Diskussion über einen Forschungsprozess.
- Theorien helfen dabei, besonders für wirtschaftliche, politische und soziale Daten, Hypothesen aufzustellen, die dann überprüft werden können.
- Theorien sollen bei der empirischen Auswertung der verfügbaren Quellen helfen, um eine Hypothese zu überprüfen.
- Theorien helfen dabei den begrifflichen Rahmen für Vergleiche zwischen Gesellschaften und Epochen zu konstruieren.
- Theorien können als Rahmen zur Definition von Kriterien dienen, um die historische Periodisierung zu erleichtern.
- Theorien helfen dabei Vergangenheit zu gegenwärtigen Fragen und Kontroversen in Beziehung zu setzen

¹⁰¹JARAUSCH/ARMINGER/THALLER: Quantitative Methoden in der Geschichtswissenschaft, S.182-191.

¹⁰²SCHRÖDER: Historical Social Research, Nr. 1, 1988, S.10-14.

2.3 Modellbildung als Kern der Digital History

*”Im wissenschaftlichen wie außerwissenschaftlichen Sprachgebrauch hat gegenwärtig der Modellbegriff zunehmend Relevanz erlangt. Bei zahlreichen passenden – leider auch unpassenden – Gelegenheiten ist von ’Modellen’ die Rede.”*¹⁰³

Der Kern der digitalen Geisteswissenschaften, also der Verwendung digitaler Werkzeuge zu Bearbeitung geisteswissenschaftlicher Fragestellungen, ist für PIOTROWSKI zweigeteilt. Zum einen umfasst es die **theoretischen digitalen Geisteswissenschaften**, die sich mit der Erforschung und Entwicklung von Methoden, die für die Erstellung von formalen Modellen in den Geisteswissenschaften nötig sind, zum anderen die **angewandten digitalen Geisteswissenschaften**, die die Anwendung dieser Mittel und Methoden zur Erstellung konkreter formaler Modelle in den geisteswissenschaftlichen Disziplinen befördert.¹⁰⁴ Zentral in dieser Definition, die gleichrangig neben vielen anderen Definitionen der digitalen Geisteswissenschaften steht, ist der Modellbegriff.

Dieses Kapitel widmet sich der terminologischen Klärung des Modellbegriffs, mit welchen Werkzeugen modelliert werden kann und welche Rolle sie in Hinblick auf (Forschungs)Daten in den Geschichtswissenschaften einnehmen.

2.3.1 Definition und Terminologie des Modellbegriffs.

Bereits 1973 beschreibt STACHOWIAK in seiner Einleitung seiner ausführlichen Abhandlung zur Allgemeinen Modelltheorie, die unscharfe Verwendung des Modellbegriffes. STRACHOWIAK definiert: *”Ein Modell ist eine verkürzte, zweckorientierte Abbildung von der Wirklichkeit”*. Weiter führt er drei Hauptaugenmerkmale von Modellen an:¹⁰⁵

- **Abbildung:** jedes Modell ist Abbild von einem Teilaspekt der Wirklichkeit
- **Verkürzung:** jedes Modell ist eine Abstraktion eines Teilaspekts der Wirklichkeit
- **Pragmatismus:** jedes Modell wird deswegen geschaffen, um für jemanden eine Problemstellung innerhalb eines bestimmten Zeitbereiches zu bearbeiten.

Auch KOBLEER führt an, dass der Modellbegriff in fast jeder wissenschaftlichen Disziplin auftritt, wobei eine einheitliche Definition des Begriffes nicht vorzufinden ist. Kritik richtet sich oft daran, dass für die Definition des Modellbegriffes andere Konzepte wie *Abstraktion*, *Entität* oder *System* verwendet werden, die wiederum terminologische Unschärfe mit sich bringen. Im deutschen Sprachgebrauch lässt sich weiters eine Doppelbedeutung des Begriffs festmachen:

¹⁰³ STACHOWIAK: Allgemeine Modelltheorie, S.1.

¹⁰⁴ PIOTROWSKI, Michael: Digital Humanities. Computational Linguistics, and Natural Language Processing. Lectures on Language Technology and History 2016.

¹⁰⁵ STACHOWIAK: Allgemeine Modelltheorie, S.131–133.

als Abbild von Etwas, sowie Vorbild für Etwas (jemand steht Modell beim Malen).¹⁰⁶ Die Notwendigkeit der ersten Bedeutung geht daraus hervor, dass die Erfahrung und das Verstehen der Welt komplex ist. In ihrer Gesamtheit übersteigen sie die kognitiven Fähigkeiten des Menschen, weswegen eine Reduktion von Information notwendig ist. Gerade diese Fertigkeit der Abstraktion und Generalisierung von Abbildern der Realität (von Dingen in der Welt) stellt die Grundlage menschlicher Kultur. Der Fokus in dieser Arbeit betrachtet nun nur externalisierte, niedergeschriebene Modelle, die sich theoretisch in Datenmodelle überführen lassen, da diese Art von Modellen, wie anfangs erwähnt, für die digitalen Geisteswissenschaften und somit für die Digital History relevant sind.

In den unterschiedlichen Disziplinen haben Modelle eine abweichende Funktion. In den Ingenieurwissenschaften arbeitet man mit formalen Modellen die einen Sachverhalt, an den man sich so nah wie möglich annähern will, beschreiben und erklären. In den Wirtschaftswissenschaften haben Modelle oft beschreibende Funktion. In der Wissenschaftstheorie werden Theorien in Form von mathematischen Modellen dargestellt. In der Wirtschaftsinformatik, wiederum werden sie für vielfältige andere Aufgaben eingesetzt, sodass man zusammenfassend eine große Menge an Modelltypen festmachen kann.¹⁰⁷

Wie für Methoden lassen sich auch qualitative von quantitativen Modellen unterscheiden. **Qualitative Modelle** beschreiben die wesentlichen Bestandteile und Beziehungen eines Systems. Zentral ist die Diskussion von Mustern, Wechselwirkungen und Strukturen. Es beschreibt beispielsweise welche Sequenz von Ursachen ein Phänomen bedingen, welche Objekte ihm zugeordnet werden können, oder welche kausalen Zusammenhänge es gibt. **Quantitative Modelle** geben das Regelwerk vor, in welchem Verhältnis sich Zahlen oder empirische Messwerte zueinander verhalten: wie lange dauert ein Prozess an, wieviele Objekte gibt es oder wie hoch ist die Wahrscheinlichkeit von X in Abhängigkeit von Y. Dabei bedarf es einer Konkretisierung qualitativer Begriffe auf der empirischen Ebene. Allerdings bedeutet die schlichte Verwendung von Zahlen noch nicht, dass es sich um ein quantitatives Modell handelt.¹⁰⁸

Eine weitere Unterscheidung von Modellen erfolgt durch statische und dynamische Modelle. In einem **statischen Modell** werden Komponenten und deren Beziehungen zwischen den Komponenten zu einem festen Zeitpunkt beschrieben. In einem **dynamischen Modell** werden prozesshafte Strukturen definiert, sprich Zustandsübergänge und Zustände, wie sie etwa durch ein Zustandsübergangsdiagramm beschrieben werden können.

Weiters gibt es eine Einteilung in nicht-formale (informelle), semiformale und formale Modelle. Existiert keine eindeutige Beschreibungssyntax für ein Modell, es also mittels natürlicher Sprache beschrieben wird, so spricht man von einem **informellen Modell**. Ein vorzufinden ist, kann

¹⁰⁶ STACHOWIAK: Allgemeine Modelltheorie, S.129.

¹⁰⁷ KOBLE, Maximilian: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung. Logos Verlag Berlin GmbH, 2010, S.41-44.

¹⁰⁸ WOLF: Qualitative versus quantitative Forschung, S.309–329.

als ein solches Modell betrachtet werden. Oft ist dies aber begründet durch die Verwendung natürlicher Sprache nicht immer eindeutig und Eindeutigkeit ist für eine formale Verarbeitung unerlässlich. Betrachtet man beispielsweise den Satz *”Ich sah den Mann auf dem Berg mit dem Fernrohr.”*, so lassen sich fünf Bedeutungen ableiten, je nachdem welche Person, wen, wo mit dem Fernrohr betrachtet. Abbildung 2 veranschaulicht dies.



Abbildung 2: Nichteindeutigkeit nicht-formaler Modelle

Ein **semiformales Modell** ist teilweise exakt. Es verfügt über eine definierte Syntax, spezifiziert aber nicht die ganze Domäne und verfügt über keine semantischen Konstruktionsregeln. Umsetzen lassen sich solche Modelle beispielsweise mit dem Entity-Relationship-Modell (ER-Diagramm), das anschließend ausführlicher beschrieben wird.

Formale Modelle verfügen über eine konkrete Syntax und Semantik. Das erlaubt die formale Definition von Modellen beispielsweise auf Basis deskriptiver Logiken.¹⁰⁹ Für formale Modelle lassen sich Algorithmen formulieren, die es erlauben Daten, die durch ein solches Modell beschrieben sind, automatisiert zu validieren. So beschriebene Modelle lassen sich in Datenmodelle überführen. Oder anders formuliert: formale Modelle können eindeutig mathematisch beschrieben werden.

Im Gegensatz zum Computer kann der Mensch auch mit informellen Modellen arbeiten. Man kann stark strukturierte und formalisierte Modelle, die algorithmisch genutzt werden können, offenen Modellen gegenüberstellen, in die sich ein Mensch hineindenken kann.

¹⁰⁹Es gibt unterschiedliche Logiken, die sich in ihrer Ausdrucksmächtigkeit und Entscheidbarkeit unterscheiden.

Konzeptuelle oder auch konzeptionelle Modelle (**conceptual model**) bzw. Informationsmodelle (information model) sind beispielsweise in der Wirtschaftsinformatik von zentraler Bedeutung. Sie sind ein Werkzeug das die Kommunikation zwischen Informatik, Wirtschaftsinformatik und Wirtschaft erleichtert.¹¹⁰ Genauso können solche Modelle in der interdisziplinären Arbeit zwischen Informatik, digitale Geisteswissenschaft und Geschichtswissenschaft fungieren.

Als **Datenmodelle** versteht man formale Modelle, die von Computern verarbeitbar sind und aus diesem Grund in expliziter und eindeutiger Form vorliegen. Datenmodelle dienen zur Repräsentation von konkreten oder abstrakten Objekten. JANNIDIS führt folgende Funktionen von Datenmodellen an:¹¹¹

- Maschinelle Verarbeitung
- Kommunikation in der Entwicklung von Algorithmen bzw. Software
- Definition der Struktur der Daten für ein Programm
- Standardisierung im Austausch von Daten

Zusammenfassend lässt sich sagen, dass der Modellbegriff, je nach Disziplin unterschiedlich verwendet wird. Gemeinsam jedoch ist stets die Funktion eine Domäne zu abstrahieren, zu vereinfachen und so zu beschreiben, dass dadurch Kommunikations- und Verstehensprozesse unterstützt werden. Zentral ist auch wie ein Modell beschrieben wird und welche Formen des Modellierens es gibt. Der aktive Prozess eine Domäne und ihre Daten in ein Modell zu überführen nennt man Modellieren.

2.3.2 Modellierung

Für McCARTY stellt die Modellerstellung, sprich das Modellieren, einen heuristischen Prozess dar, der die Manipulation und Konstruktion von Modellen umfasst. So, ganz im Sinne von STACHOWIAK, betrachtet er ein Modell als Repräsentation von etwas, dessen Zwecke der Entwurf zur Implementierung von etwas Neuen, oder der Analys einer Domäne dienlich ist.¹¹²

Drei Besonderheiten hebt JANNIDIS in der Modellierung von Daten in den Geisteswissenschaften hervor: die Historizität und Kontextabhängigkeit von Objekten, die für die geisteswissenschaftliche Analyse essentiell sind, die Unschärfe von Information und *”das systembedingte Primat des Individuellen vor dem Allgemeinen”*. Historizität meint, dass nicht nur jedes Objekt eine Geschichte aufweist, sondern auch jede Untersuchung eines Objektes. Als unscharf

¹¹⁰KOBLER: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung, S.44-47.

¹¹¹JANNIDIS, Fotis: Grundlagen der Datenmodellierung. In Digital Humanities Springer, 2017, S.99-100.

¹¹²MCCARTY: Modeling: A Study in Words and Meanings, S.255.

versteht man, dass es keine definitiven und fixen Werte gibt, die aber für eine maschinelle Verarbeitung notwendig sind. Und Letzteres meint, dass in den Geisteswissenschaften oft der Blick auf ein einzelnes Objekt, z.B. ein Roman, tendenziell interessanter ist als auf eine Vielzahl.¹¹³

In der Modellierung, auf einer konzeptuellen Ebene, ist es notwendig die im Bezug auf den Verwendungszweck wichtigsten Entitäten und ihre Attribute, sowie Relationen zu identifizieren.¹¹⁴ Das **Entity-Relationship-Modell** (ER-Diagramm) ist eine Notationsformen, die genau auf diesen drei Begriffen beruht: Relationen, Attributen und Entitäten (Klassen). Die Entitäten – alle potenziellen Untersuchungsobjekte – verfügen über Eigenschaften, sogenannte Attribute. Weiters stehen sie zu anderen Entitäten in unterschiedlichen Beziehungen, sogenannten "Relationen".¹¹⁵ Eine gängige Art ER-Diagramme umzusetzen, besonders in der Softwareentwicklung, ist die *Unified Modeling Language* (UML). Dabei handelt es sich um eine Sprache zur Modellierung, Dokumentation und Spezifikation von Systemen und Softwarekomponenten. Als ISO genormte Sprache definiert UML Bezeichner für wichtige Begriffe einer Modellierung und legt mögliche Beziehungen zwischen diesen Begriffen fest. UML definiert weiter grafische Notationen für diese Begriffe und für Modelle statischer Strukturen und dynamischer Abläufe, die man mit diesen Begriffen formulieren kann.¹¹⁶ Folgende Abbildung veranschaulicht die Bestandteile von UML:

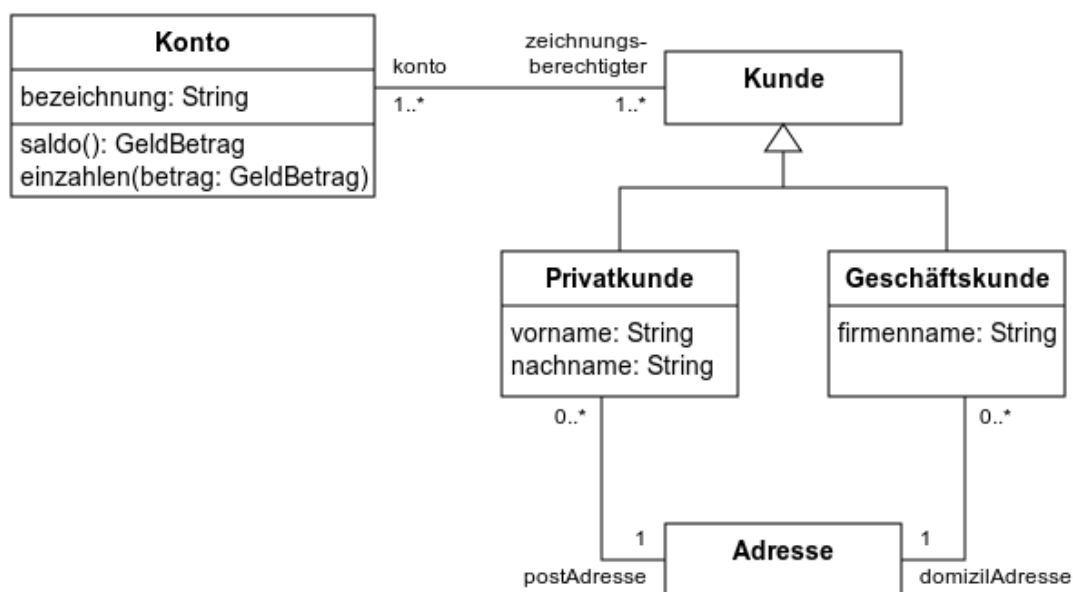


Abbildung 3: Beispiel für ein Klassendiagramm nach UML

¹¹³ JANNIDIS: Grundlagen der Datenmodellierung, S.106-108.

¹¹⁴ ebd., S.102-104.

¹¹⁵ CHEN, Peter Pin-Shan: The entity-relationship model—toward a unified view of data. ACM Transactions on Database Systems (TODS), 1 1976 Nr. 1.

¹¹⁶ RUMBAUGH, James/JACOBSON, Ivar/BOOCH, Grady: Unified modeling language reference manual, the. Pearson Higher Education, 2004.

Es existiert eine Klasse *Kunden*, die wiederum zwei Unterklassen *Privatkunde* und *Geschäftskunde* besitzt. Die beiden Unterklassen (weißer Pfeil definiert 'hat Superklasse') unterscheiden sich in ihren unterschiedlichen Attributen. Wo die *Privatkunde*-Klasse aus *vorname* und *nachname* vom Datentyp *string* zusammengestellt wird, verfügt der *Geschäftskunde* über das Attribut *firmenname*. Die Linien zwischen den Klassen definieren Relationen und die daneben stehenden Zahlwerte und Symbole die Kardinalität dieser Beziehungen. So hat eine Instanz aus der Klasse *Privatkunde* eine Relation *postAdresse* zu einer Klasse *Adresse*, aber eine Adresse kann mit mehreren Privatkunden verknüpft sein.¹¹⁷

Man kann nun aus der Klasse *Privatkunde* beliebig viele Instanzen erzeugen. Instanzen entsprechen den eigentlichen Daten, die beispielsweise in einer Datenbank abgelegt werden können. So könne man das oben angeführte konzeptuelle Modell folgendermaßen instanziiieren, sodass ein Privatkunde "Christopher Pollin" eine Adresse und ein Konto hat:

Privatkunde → *vorname* "Christopher" *nachname* "Pollin"

"Christopher Pollin" *postAdresse* "Neubaugasse 1"

"Christopher Pollin" *konto* "Konto 1"

Die Beschreibung von historischen Daten und eines Modells, das diese Daten definiert, kann durch ER bzw. UML umgesetzt werden und hilft bei der Überführung in Datenmodelle.

2.3.3 Forschungsdaten in den Geschichtswissenschaften

Die Grundlage jeder formalen Verarbeitung sind Daten. Die Definition von Daten, im Sinne der Informationswissenschaft, lässt sich am besten in der Kontextualisierung mit drei anderen Begriffen fassen: Signal, Information und Wissen. Wo **Signale** eine physikalische Entität darstellen, beispielsweise eine Lichtwelle, versteht man unter **Daten** Zeichen zur Verarbeitung und Repräsentation von Information. Noch spezifischer wird der Begriff in der Informatik verwendet. In diesem Fach versteht man unter Daten Informationseinheiten, die für Maschinen lesbar und bearbeitbar sind. Von **Information** spricht man, wenn Daten in einem bestimmten Kontext interpretiert werden. WERSIG, aus einem kybernetischen Verständnis, spricht bei Information von einer "*Reduktion von Ungewissheit aufgrund von Kommunikationsprozessen*".¹¹⁸ Noch eine Hierarchiestufe höher steht der Wissensbegriff. **Wissen**, so FAVRE-BULLE, ist gegeben, wenn ein kognitiver Agent, beispielsweise ein Mensch, auf Basis von Information eine Handlung in der Welt setzt.¹¹⁹

¹¹⁷JANNIDIS/KOHLE/REHBEIN: Digital Humanities: Eine Einführung, S.99-108.

¹¹⁸WERSIG, Gernot: Information-Kommunikation-Dokumentation. Band 5, Verlag Dokumentation, 1971, S.74.

¹¹⁹FAVRE-BULLE, Bernard: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001, S.93-97.

Folgende Darstellung zeigt die Abhängigkeit dieser vier Begriffe zueinander, mit einem anwachsen der Komplexität der Struktur, sowie der Anbindung an einen kognitiven Agenten.¹²⁰

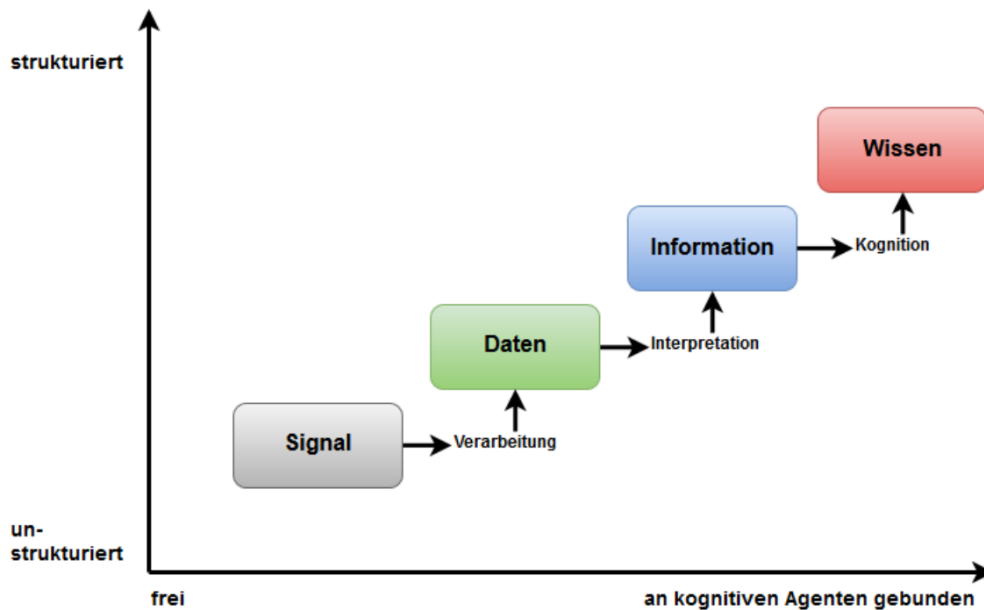


Abbildung 4: Signal, Daten, Information und Wissen

Entstehen Daten während eines, oder als Ergebnis eines Forschungsprozesses, so spricht man von **Forschungsdaten**. Damit werden sowohl Daten aus den Naturwissenschaften (Messdaten aus einem Experiment), Sozialwissenschaften (Interviews) oder den Kulturwissenschaften (dokumentierte Beobachtung) zusammengefasst.¹²¹

ANDORFER spricht sich für einen nicht *”inflationären Gebrauchs dieses Begriffes”* und einer *”präzisere[n] Terminologie”* im Zusammenhang des Begriffes geisteswissenschaftliche Forschungsdaten aus. Es soll klar definiert sein, was man als Forschungsdaten auffasst und was davon ausgeschlossen ist. Weiters soll es bei den für Fachwissenschaftlerinnen vertrauten Begrifflichkeiten bleiben.¹²² Diese Datenpyramide, in Abbildung 5, beschreibt, dass aus den Kulturerbeeinrichtungen (Archiv, Bibliothek und Museum), sowie den digitalen Repositorien, die Quellen für die Fachwissenschaftlerinnen zur Verfügung gestellt werden. Beispielweise liegen die originalen und analogen historischen Rechnugsbücher in einem Stadtarchiv. Wird auf Basis dieser Quellen – und man sieht bereits, dass die Pyramide schmaler wird und eine Selektion aller vorhandenen Objekte geschieht – eine weitere Leistung produziert, wie etwa eine Transkription

¹²⁰Eine ausführlichere Auseinandersetzung mit den Begriffen Daten, Information und Wissen findet sich in meiner ersten Abschlussarbeit. POLLIN, Christopher: Vom Suchen, Stöbern und Finden : Information Retrieval am Beispiel der Digitalen Sammlung des Hans Gross Kriminalmuseums. Diplomarbeit, Universität Graz, 2017, Masterarbeit Graz, S.20-28.

¹²¹KINDLING, Maxi et al.: Forschungsdatenmanagement an Hochschulen: Das Beispiel der Humboldt-Universität zu Berlin. LIBREAS. Library Ideas, 2013 Nr. 23 (URL: libreas.eu/ausgabe23/07kindling), 09.06.2019.

¹²²ANDORFER, Peter: Forschungsdaten in den (digitalen) Geisteswissenschaften: Versuch einer Konkretisierung. DARIAH-DE, Working Papers 14 2015 (URL: webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-14.pdf).

des Rechnungsbuch. In diesem Fall kann man von Arbeitsdaten sprechen. Solche Arbeitsdaten sind Forschungsdaten und können in einem geeigneten Repository gesichert und verfügbar gemacht werden. Ein solches Repository ist beispielsweise GAMS.¹²³ Eine weitere Auswahl führt von Arbeitsdaten zu veröffentlichten Publikationen, also ein Aufsatz oder eine digitale Edition, dessen Grundlage die Transkription eines Rechnungsbuch war.

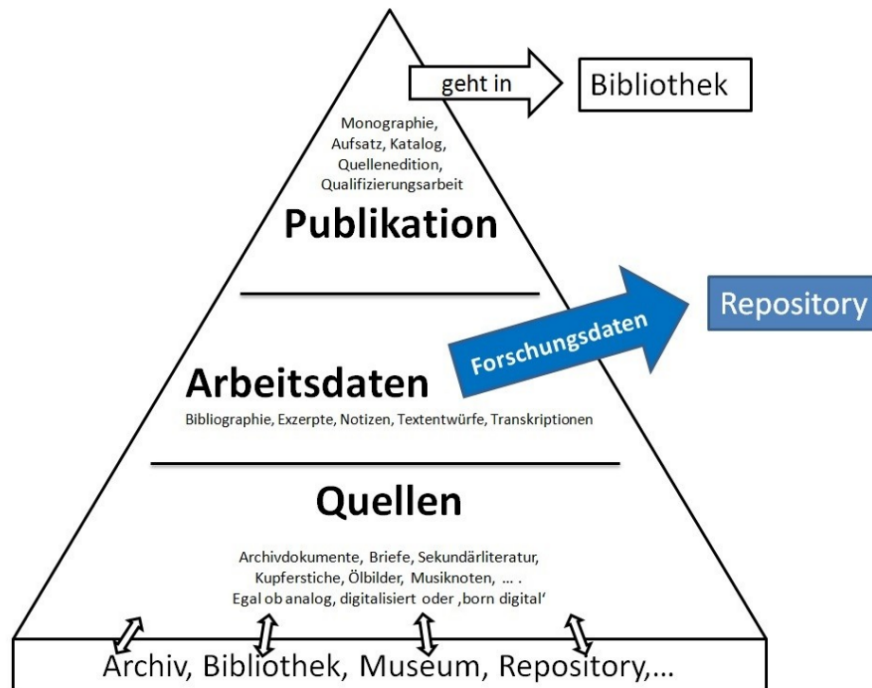


Abbildung 5: Datenpyramide geisteswissenschaftlicher Forschungsdaten im institutionellen Kontext nach ANDORFER

Für die Geisteswissenschaften und im speziellen für die Geschichtswissenschaften, in denen viel stärker hermeneutischer Prozess im Vordergrund stehen, ist der Begriff der Forschungsdaten nicht eindeutig. HILTMANN sieht einen engeren und einen weiteren Begriff von Forschungsdaten in den Geschichtswissenschaften. Im engeren Sinne versteht man Metadaten als Annotationen von historischen Quellen in all ihren Ausprägungen. Als Beispiel ist die archivarische Erschließung eines Quellenkorpus angeführt, damit die Verwahrung und Auffindbarkeit für eine potentielle, wissenschaftliche Auseinandersetzungen ermöglicht wird. Im weiteren Verständnis kann alles, das digital vorhanden ist, zur historischen Quelle und somit zu relevanten Daten für die Geschichtswissenschaft werden.¹²⁴

Damit Forschungsdaten nachnutzbar und nachvollziehbar sind bedarf es eines mitgelieferten Modells. Diese Modelle wiederum müssen auf Standards beruhen. Es ist zeitintensiv die Datenmodelle andere Fachwissenschaftlerinnen zu verstehen und oft ist es leichter sein eigenes

¹²³Geisteswissenschaftliches Asset Management System, <https://gams.uni-graz.at>, 06.05.2020.

¹²⁴HILTMANN, Torsten: Forschungsdaten in der (digitalen) Geschichtswissenschaft. Warum sie wichtig sind und wir gemeinsame Standards brauchen. 2018 (URL: dig.w.hypotheses.org/2622), 09.06.2019..

Modell zu entwickeln. Ein generelles Problem mit Informationssystem und Forschungsdaten ist, dass sie oft nur für eine bestimmte Fragestellung bzw. ein bestimmtes Projekt entwickelt werden. Werden bei der Generation der Forschungsdaten aber bereits Konzepte und Standardisierungen verfolgt, so fällt es leichter die "Datensilos" zu verlassen und die Nachnutzung zu fördern.

Eine erste Form der Standardisierung in den (digitalen) Geisteswissenschaften ist die *Text Encoding Initiative* (TEI). Die TEI entwickelt seit Ende der 1980er Jahre einheitliche Verfahren für die strukturelle und semantische Auszeichnung verschiedener Text- und Dokumenttypen. XML/TEI ermöglicht die korrekte Erfassung von Zeichenfolgen in jeder Art von Text, die Integration von Textstrukturen sowie von Phänomenen und Kontextinformationen. Wenn auch die Umsetzungen der TEI nicht einem 'klassischen' Standard entspricht, so sind die Überlegungen, wie man Text strukturieren und annotieren kann, sehr wohl standardisiert.

Die TEI-Richtlinien decken jedoch ein breites Spektrum unterschiedlicher Textarten ab und die Mehrzahl der Merkmale werden nicht für ein einzelnes Projekt benötigt. Darüber hinaus erlaubt die TEI oft mehrere Auszeichnungsvarianten für dasselbe Textphänomen.¹²⁵

Es ist eine Notwendigkeit, dass Standards auf der einen Seite ausdrucksstark sein müssen, um die unterschiedlichen Problemstellungen fassen zu können, andererseits es sich aber daraus formale Modelle bzw. Datenmodelle ableiten lassen müssen. Standards zur Beschreibung von Modellen sollen vielmehr einen *beschreibenden* statt *vorschreibenden* Charakter haben.¹²⁶

¹²⁵SAHLE, Patrick: Digitale Edition. Digital Humanities. Eine Einführung, 2017, S.234-249.

¹²⁶THALLER, Manfred: The Need for Standards: Data Modelling and Exchange [1991]. Historische Sozialforschung. Supplement, 2017, S.204.

3 Historische Rechnungsbücher als Quelle

In diesem Kapitel wird der Quellentypus Rechnungsbuch als historische Quelle dargestellt. Dazu werden inhaltliche, wie auch äußerliche Eigenschaften von Repräsentanten dieser Quellengattung beschrieben und die dazugehörige Fachliteratur diskutiert. Anfangs wird der Quellentypus an sich beschrieben, gefolgt von einer Beschreibung konkreter digitale Editionsprojekte von Rechnungsbüchern, in denen ein Blick auf die textuelle und semantische Strukturen der Quellen geworfen wird. Weiters wird beispielhaft erörtert, welche geschichtswissenschaftlichen Forschungsfragen sich damit beantworten lassen.

3.1 Wirtschafts- und Rechnungsbücher

*”Wirtschafts- und Rechnungsbücher kennen keine Grenzen – keine räumlichen, wenig ständische, keine institutionellen Grenzen – und auch keine disziplinären.”*¹²⁷

Bei Rechnungsbüchern handelt es sich um pragmatisches, oft seriell, zum Teil über Jahrzehnte hinweg erhaltenes, aus dem Alltag der Menschen stammendes Schriftgut, das zur Dokumentation und als Gedächtnisstütze für den Austausch von Waren, Dienstleistungen und Rechten diente. Zu meist auf Papier, aber im Laufe der Zeit mittels unterschiedlichsten Schreib- und Beschreibstoffen und ohne übergeordnete Standards verfasst, folgen sie keinem repräsentativem Zweck. Menschen mussten zumindest über Schreib- und Lesekenntnisse verfügen, die von einfachen Rechenfertigkeiten bis hin zur Erstellung vollständiger Kalkulationen reichten. Das gemeinsame Element dieses differenzierten Quellentypus ist die Ausrichtung auf wirtschaftliche Aktivitäten in Form von Transaktionen. Daraus ergibt sich eine ”natürliche” – oft auch textuelle – semantische Struktur, die zeigt, dass es sich um ein *”Werkzeug des Alltages”* gehandelt hat, das man an unterschiedlichen Orten und Zeiten der Menschheitsgeschichte finden kann.¹²⁸

Eine Transkription allein reicht nicht aus, um die linguistische/textuelle, quantifizierbare und die semantische Dimension einer solchen Quelle abzudecken. Vielmehr ist es die Edition der Quellen, die dabei helfen kann, Forschungsinteresse zu befriedigen.¹²⁹ So wird die Vielfalt der ableitbaren Inhalte aus Rechnungsbüchern, sowie die Möglichkeiten der *”interdisziplinären Erschließung der Quelle Rechnungsbuch”* hervorgehoben. Die Herausforderungen der Erschlie-

¹²⁷GLEBA, Gudrun: Rechnen. Wirtschaften. Aufschreiben. Vernetzte Schriftlichkeit – Wirtschafts- und Rechnungsbücher als Quellen klösterlicher Alltagsgeschichte. In PÄTZOLD, Stefan/STUMPF, Marcus (Hrsg.): Mittelalterliche und frühneuzeitliche Rechnungen als Quellen der landesgeschichtlichen Forschung. Münster, 2016, S.51.

¹²⁸Eine Übersicht zu historischen Rechnungsbüchern und Handelspraktiken findet sich auch bei: DENZEL, Markus A/HOCQUET, Jean Claude/WITTHÖFT, Harald: Kaufmannsbücher und Handelspraktiken vom Spätmittelalter bis zum beginnenden 20. Jahrhundert. Franz Steiner Verlag, 2002, 163.

¹²⁹VOGELER: Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert?.

ßung der Quelle sind vielfältig: Maß-, Größen- und Gewichtsordnungen und Information zu Geldwerten sind abhängig von Ort und Zeit. Dennoch lassen sich viele Forschungsfragen aus unterschiedlichen Disziplinen neben den Geschichtswissenschaften, wie etwa der Historischen Linguistik, der Editions- oder Medienwissenschaft, mit solchen Quellen bearbeiten: wie haben sich Güter, Dienstleistungen und Geldgeschäfte über die Zeit hinweg verändert, welche wirtschaftlichen, aber besonders auch welche kultur- und sozialgeschichtliche Relationen gibt es? Rechnungsbücher gehen nach KLAPP *”weit über wirtschafts- und verwaltungsgeschichtliche Aspekte hinausgehen”*¹³⁰ und so erlauben Fragestellungen dieser Art weitere soziale, gesellschaftliche und wirtschaftliche Implikationen, die uns ein Verständnis darüber geben können, wie das Leben der Menschen in der Vergangenheit ausgesehen haben kann.¹³¹

GLEBA unterscheidet vier Kategorien¹³² in diesem Zusammenhang:

- **Urbare** sind Besitzverzeichnisse, in denen Besitzrechte einer Grundherrschaft und der damit verbundenen Abgaben aufgelistet werden. Hier finden sich meist auch Angaben zur Lokalisierung der Liegenschaften.
- Sogenannte **Heberegister** beinhalten die Abgabepflichten an wirtschaftlichen Gütern, wie Arbeit oder Naturalien.
- In **Wechselbüchern** wird der Wechsel von Menschen in einer anderen Grundherrschaft dokumentiert.
- **Rechnungsbücher** führen vor allem Transaktionen, sprich monetäre Einnahmen und Ausgaben an. Sind aber nicht nur auf diese beschränkt.

BRUCH sieht Rechnungsbücher als geeignete Quelle, aus denen Daten zur Beantwortung der oben angeführten Forschungsfragen extrahiert werden können. Diese können anschließend mittels quantitativen Methoden weiterverarbeitet werden. Am Quellenbestand des *Zisterzienserinnenkloster Pielenhofen bei Regensburg*, der beispielhaft für die Arbeit in diesem Bereich steht, wird diese Vorgehensweise veranschaulicht. Nach einer notwendigen Kontextualisierung der Quelle, datiert auf 1224-1348, wird die wirtschaftliche Situation des Klosters über diesen Zeitbereich hinweg, beschrieben. BRUCH fasst zusammen:¹³³

¹³⁰KLAPP, Sabine: Die “Äbtissinnenrechnungen” des Klosters St. Klara auf dem Werth. Alltag und Festtag einer geistlichen Frauengemeinschaft Straßburgs am Ausgang des Mittelalters. Zeitschrift für die Geschichte des Oberrheins Bd. 159, 2011, S.14.

¹³¹GLEBA, Gudrun/PETERSEN, Niels: Zur Einleitung. In GLEBA, Gudrun/PETERSEN, Niels (Hrsg.): Wirtschafts- und Rechnungsbücher des Mittelalters und der Frühen Neuzeit-Formen und Methoden der Rechnungslegung: Städte, Klöster und Kaufleute. Universitätsverlag Göttingen, 2015, S.7-10.

¹³²GLEBA: Rechnen. Wirtschaften. Aufschreiben. Vernetzte Schriftlichkeit – Wirtschafts- und Rechnungsbücher als Quellen klösterlicher Alltagsgeschichte, S.51-54.

¹³³BRUCH, Julia: Die Kunst, Daten in Informationen umzuwandeln. Zur Auswertung eines zisterziensischen Rechnungsbuchs aus dem 13. und 14. Jahrhundert und den Herausforderungen in der Analyse serieller Wirtschaftsquellen. Wirtschafts- und Rechnungsbücher des Mittelalters und der Frühen Neuzeit, 2015, S.13-37.

*”Allerdings zeigt sich auch, dass die für ein Jahr gewonnenen Daten miteinander in Beziehung gesetzt werden können: Einnahmen gegenüber Ausgaben, Getreide gegenüber Geld, Schulden gegenüber Einnahmen und Ausgaben, Klosterinsassen oder der Viehbestand gegenüber den Einnahmen bzw. Ausgaben. [...] So wandeln sich die reinen Daten in Informationen, die zur Beantwortung konkreter Fragestellungen herangezogen werden können.”*¹³⁴

BRUCH kommt weiters zum Schluss, dass die Abrechnung durch das Mutterkloster nicht auf einem Rechnungsjahr basiert und so die Bezugszeiträume für die Zahlen nicht zur Gänze nachvollziehbar sind. Die lückenhafte Überlieferung ist eine Momentaufnahme und somit eine mathematisch-statistische bzw. quantitative Analyse nur in Ansätzen möglich. Nichtsdestotrotz lassen sich die Daten aus dem Rechnungsbuch in Informationen überführen. Folgendes lässt sich Gegenüberstellen und auf dieser Informationsbasis weiter untersuchen:¹³⁵

- Die Haushaltsführung, sprich Einnahmen und Ausgaben eines Jahres
- Getreide- und Geldeinnahmen bzw. Ausgaben, um die wirtschaftlichen Schwerpunkte des Klosters zu eruieren.
- Schulden können gegen die Einnahmen aufgerechnet werden
- Die Zahl der Klosterinsassen oder der Viehbestand können mit Einnahmen bzw. Ausgaben in Verbindung gebracht werden. Hinzu kommt, dass man, wenn man die unterschiedlichen Visitationszeitpunkte beachtet, Tendenzen ausmachen kann, ob die Klöster mit strukturellen Defiziten zu recht kommen musste, oder prosperierten.

¹³⁴BRUCH: Wirtschafts- und Rechnungsbücher des Mittelalters und der Frühen Neuzeit 2015, S.37.

¹³⁵ebd., S.37-44.

3.2 Digital Edition Publishing Cooperative for Historical Accounts

*”Thus, the information that is available about the past is grounded in verifiable facts based on evidence that we make available to everyone. We’re demonstrating the importance of evidence and facts in understanding our world.”*¹³⁶

Daten - aus unterschiedlichen Formaten - sollen auf einer gemeinsamen Plattform zusammengeführt werden und adäquate Formen des Retrievals, der Discovery und der Visualisierung eröffnen, um die Arbeit mit den Quellen zu erleichtern. Die Überführung in das Resource Description Framework (RDF) auf Basis der im Projektkontext entwickelten *Bookkeeping-Ontologie*, die Transferprozesse historischer Rechnungsbücher formalisiert, erlaubt die Interoperabilität, Verlinkung und das Zusammenführen von Informationen im Sinne des *Web of Data* und *Linked Open Data*. Diese Begriffe werden im Kapitel 4 ausführlicher erörtert.

Drei ausgewählte Rechnungsbücher, die im Zuge von DEPCHA bearbeitet und veröffentlicht werden, werden nun ausführlicher behandelt. Dabei wird sowohl die Struktur und der Kontext der Quelle beschrieben und herausgearbeitet welche Forschungsinteressen mit diesen Quellen bedient werden können.

3.2.1 The George Washington Financial Papers

Die digitale Edition der *George Washington Financial Papers* (1748-1799) haben zum Ziel die Geschäfts- und Haushaltsakten des US-Präsidenten George Washington über das Web zugänglich zu machen. Basierend auf einer technischen Open Source Lösung, dem Content Management System Drupal,¹³⁷ sind Benutzerinnen in der Lage die Transkriptionen der drei Rechnungsbücher (*Ledgers A, B, and C*) von Washington zu lesen, Suchanfragen an die aus den Editionen extrahierten Daten zu stellen, sowie die Daten zu exportieren und lokal weiter zu verarbeiten. Auf diese Weise wird ein Einblick in das Leben der Person George Washington, sowie anderer Themen, wie etwa die materielle Kultur, Sozialgeschichte, Gewerbe und Landwirtschaft dieser Zeit ermöglicht.¹³⁸ Beispielhafte Forschungsfragen, die für Historikerinnen im Zuge dieses Editionsprojektes von Interesse sind, lauten wie folgt:

- Wie viel Geld hat Washington jährlich für bestimmte Rohstoffe ausgegeben?
- Welche Rolle spielt der Sklavenhandel in Washingtons wirtschaftlicher Tätigkeit?
- Wie hat sich der Preis bestimmter Rohstoffe über diesen Zeitraum hinweg verändert?
- Wie sieht das Netzwerk der Geschäftspartner rund um Washington aus?

¹³⁶TOMASEK Kathryn: Accounting for the past, <https://wheatoncollege.edu/news/accounting-for-the-past>, 30.01.2020.

¹³⁷Drupal, drupal.org, 12.10.2019

¹³⁸The George Washington Financial Papers Project, financial.gwpapers.org, 08.10.2019.

- Wie wurde der Wert der Ware Tabak in verschiedenen Währungen berechnet und bewertet?

Im Laufe seines Lebens trug Washington Tausende von Finanzdokumenten zusammen, in denen er jede Transaktion genau nach modernster Finanzlehre aufzeichnete. Dies umfasste Ausgaben für Lebensmittel und Textilien oder seinen Landsitz in Mount Vernon, sowie Kosten für die Ausbildung seiner Kinder oder die medizinischen Behandlungen für Sklaven. Weiters dokumentiert Washington an wen er Geld verliehen hat, seine Anteile an Firmen oder eingehobene Mieten. Aus den Rechnungsbüchern schließt STERTZER, dass Washington ein versierter Unternehmer war, der Innovationen suchte, kalkulierte Risiken einging und aufgrund seiner methodischen Buchführung ein breites Verständnis für die amerikanische Wirtschaft hatte.¹³⁹

An einem Beispiel soll Veranschaulicht werden, wie Einträge und deren Struktur in diesem Rechnungsbuch aussehen. Abbildung 6 zeigt die erste Seite aus dem Rechnungsbuch *Ledger A* von Washington. Die Überschrift definiert die Debit-Seite der Transaktionen, gekennzeichnet durch die Abkürzung *Dr*, mit der Person *George Fairfax* der nun folgenden Einträge. In der ersten Spalte wird das Datum der Transaktion festgehalten, wie *Novr 6* für den 06.11.1750, gefolgt von der Verschriftlichung der Transaktion in der mittleren Spalte. Diese definiert zu meist den Status, Art, Wert und die beteiligten Personen der Transaktion. *”To Ditto”* meint, dass es sich um die zuvor angeführte Person handelt. Für eine formale Verarbeitung müssen diese Phrasen normalisiert werden. Die letzten 3 Spalten stehen für die Währungseinheiten Pfund, Schilling und Pence. Zum Teil werden diese Spalten auch genutzt um eine (Zwischen)Summe zu berechnen.¹⁴⁰

George Fairfax Esqr. Dr				
1750 April 2	To John Welton's Order on you	4	10	
Octr 7	To Cash lent	2	3	
Novr 8	To Ditto paid Robt Worthington & Michl Sweim Chr Govr	17	3	

Transkription von drei Einträge von Seite 1 des Ledger Book A

Inhaltlich zusammengefasst lassen sich 3 Geldzahlungen von George Washington an *George Fairfax* festmachen: eine Zahlung in Zusammenhang einer weiteren Person *John Welton's Order*, ein Geldbetrag von 2 Pfund und 3 Schilling der verliehen wurde und die Bezahlung für *Robt Worthington & Michl Sweim*.

¹³⁹STERTZER, Jennifer: Working with the Financial Records of George Washington: Document vs. Data. Digital Studies/le Champ Numérique, 2014 (URL: doi.org/10.16995/dscn.57).

¹⁴⁰Ledger A, 1750 - 1772: pg.1, financial.gwpapers.org/?q=content/ledger-1750-1772-pg1

I

George Fairfax Esq^r Dr

1750		
April 2	To John Wellton's Order on you.....	4..10..
Oct ^r 7	To Cash lent.....	2 3..
Nov ^r 8	To Ditto paid Rob ^t Worthington & Mich ^l Swaim At Car ^r	17. 3
	To Surveying a piece of Land for Mich ^l Swaim p ^r Ord ^r	2..3..
	To Cash reciev ^d of Benj ^m Rutherford.....	2..3..
		<u>211..16 3</u>
Jan ^y 5	To Cash paid Doct ^r Lynn for Hartshorn's Havings.....	0..
750	To Ditto paid W ^m Yates for 2 p ^r Mens Shoes.....	1..1..
	To Ditto for a Saddle cloth of Rich ^d Lewis.....	7 6
Feb ^y 6	To Ditto for Ditto of W ^m Yates.....	7 6
	To Ditto for 600 brags Nails.....	6..
	To 2000 4 ^o Ditto of M ^r Hunter.....	7..
	To 2000 6 ^o Ditto of Ditto.....	11..

Abbildung 6: 1. Seite von George Washington, 1750-72, *Ledger Book A*, Library of Congress, Manuscript Division.

3.2.2 Cameron Family Papers, Stagville Accounts

Das historische Stagville ist eine Tabakplantage nördlich von Durham (North Carolina, USA), die seit dem späten 18. Jahrhundert im Besitz der Familie Bennehan und Cameron ist. Ihr Eigentum, vor Beginn des Amerikanischen Bürgerkriegs, umfasste 30.000 Hektar Land und um die 900 versklavten Menschen, die auf dieser Plantage arbeiten mussten.¹⁴¹

Die *Cameron Family Papers* (1767-1892) beinhalten umfangreiche Geschäftsunterlagen im Zusammenhang mit einem Wirtschaftsbetrieb rund um eine Tabakplantage, die Transaktionen zwischen dem Geschäftsladen und den Mitgliedern, sowie unfreien Arbeiterinnen der Gemeinschaft dokumentieren. Beispielsweise findet sich darin ein sogenanntes "Slave-Ledger". Dabei handelt es sich um ein separat geführtes Rechnungsbuch, das Transaktionen zwischen dem Geschäft auf der Tabakplantage und versklavten Kundeninnen dokumentiert. AGBE-DAVIS und BRUMFIELD stellen bei der Untersuchung des "Slave-Ledger" fest, dass die Unterlagen nicht isoliert erforscht werden sollten, sondern im Kontext aller Dokumente in den *Cameron Family Papers* betrachtet werden müssen. Forschungsfragen in diesem Projektkontext sind:

- Wie bezahlen unfreie Arbeiterinnen Güter, die sie im Laden der Plantage erwerben konnten?

¹⁴¹ BRUMFIELD, Ben/AGBE-DAVIES, Anna: Encoding Account Books Relating to Slavery in the U.S. Modeling semantically Enriched Digital Edition of Accounts, 2015 (URL: medea.hypotheses.org/182).

- Welche wirtschaftlichen Abhängigkeiten (Verschuldungen) bestehen zwischen den unterschiedlichen Akteuren?
- Welche Waren werden gemeinsam gekauft?
- Welche soziale Stellung haben sie und inwieweit beeinflusst das ihre Einkäufe.

In folgender Abbildung ist die dritte Seite des Rechnungsbuches aus Stagville dargestellt.¹⁴² Der erste Eintrag beschreibt den Kauf eines Schuhmessers zum Preis von 1 Schilling und 6 Pence. Im "Slave-Ledger" verkauft Cameron, aber ein Schuhmesser für 2 Schilling. Ob die Änderung des Preises durch den zeitlichen Unterschied gegeben ist, oder eben durch den sozialen Status der kaufenden Person, ist eine relevante geschichtswissenschaftliche Fragestellung.¹⁴³

Stagville August 1st 1808				
69	James Haley Self Dr.			
	1/4 lb Powder 2/6	1 lb Shot 2/6	1 lb Sugar 2/6	7. 6
	1 Shoe Knife 1/6			1. 6
				9 -

Abbildung 7: Seite 3 des Objektes 00133_sv0061 der Stagville Accounts

Stagville August 1st 1808

69 James Haley Self Dr.

1/4 lb Powder 2/6 1 lb Shot 2/6 1 lb Sugar 2/6 7 6

1 Shoe Knife 1/6 1 6

9 -

Transkription der zweier Einträge auf Seite 3 der Stagville Accounts.

Zur Veranschaulichung dieser Quelle zeigt Abbildung 7 wieder eine ähnliche Struktur, die vergleichbar ist mit derer der *George Washington Financial Paper*. In der Überschrift wird Ort und Datum, der 1. August 1808, angegeben und alle nun folgenden Einträge beziehen sich darauf. Die tabellarische Struktur verfügt über 5 Spalten. Die erste Spalte enthält eine Referenz auf eine Seite in einem anderen Dokument, in dem die *Credit* gegen geschrieben werden. Die zweite

¹⁴²Stagville Accounts, From the Page, fromthepage.com/agbedavies/stagville-accounts/00133-sv0061/display/7884, 08.10.2019.

¹⁴³BRUMFIELD, Ben: The Stagville Accounts in DEPCHA: Plantation Financial Records as Linked Data. From the Page, 2019 (URL: content.fromthepage.com/stagville-accounts-in-depcha/).

Spalte beinhaltet den Fließtext der Einträge, wobei die Nennung der Person in der ersten Zeile jede weitere Zeile als einen eigenen Eintrag mit Geldbeträgen in der 4. und 5. Zeile deklariert. Inhaltlich zusammen gefasst hat "James Haley" einen 1/4 Pfund Pulver, einen Pfund Munition und einen Pfund Zucker zum Preis von je 2 Schilling und 6 Pence, sowie ein Schuhmesser zum Preis von 1 Schilling und 6 Pence käuflich erworben. Die Summe aller Einträge ist somit 9 Schilling, da 1 Schilling 12 Pence entspricht.

3.2.3 The Wheaton Accounts

Die *Wheaton Family Papers* umfassen mehrere Bücher im Zeitraum von 1828 bis 1859, die unter anderem die finanziellen Geschäfte von *Laban Morey Wheaton*, einem Geschäftsbesitzer eines *dry goods store* aus Norton, Massachusetts in den Vereinigten Staaten von Amerika, dokumentieren. Unter *dry goods* versteht man einen aus dem 17. Jahrhundert stammenden historischen Begriff, der je nach Region zum Teil unterschiedliche Güter zusammenfasst. Darunter werden im Allgemeinen Produkte verstanden, die "trocken" sind, beispielsweise Textilien, Konfektionskleidung, Tabak oder bestimmte Lebensmittel, wie etwa Kartoffel.¹⁴⁴

Die Stadt Norton ist eine charakteristische Kleinstadt im Nordosten der USA, die von landwirtschaftlich Betrieben und Manufakturen geprägt ist. Sie befindet sich im Hinterland der regional bedeutenden Häfen Boston, New Bedford, Newport und Providence. In einer Geschichte der Stadt Norton von CLARK wird *Laban Morey Wheaton* namentlich genannt. Er lebte von 1796 bis 1865, hat Rechtswissenschaften an der *Brown University* studiert und jahrelang als *Postmaster of Norton* gearbeitet. Er war politisch aktiv und war Vertreter im *Massachusetts General Court*, sowie Mitglied im *Massachusetts Governor's Council*. Weiters war er verheiratet und Angehöriger des Kongregationalismus, einer Form der christlichen Gemeindeverfassung. Die Wheaton Familie besaß eine Milchviehherde und Manufakturen zur Produktion von Baumwollwatte, die *Wheaton Manufacturing Company*, und eine zur Produktion von Strohhüten.¹⁴⁵ So gehört Laban Wheaton, von ihm ist bei CLARK ein Portrait überliefert, zu einem der wohlhabendsten Männer der Stadt.¹⁴⁶

Aus dieser Überlieferung über *Laban Morey Wheaton* kann man die These aufstellen, dass er ein mächtiger Mann war. Auf der einen Seite politisch aktiv und politischer Vertreter, auf der anderen Seite vermögend und Besitzer mehrere Immobilien und eines Geschäftes. Gerade der im *Daybook* dokumentierte Verkauf von Gütern, in dem Menschen Lebensmittel im Austausch für Arbeit erwerben, zeigt eine gewisse Machtposition. Denn die, nicht vollständig überlieferten

¹⁴⁴Definition von *Dry Goods*, chestofbooks.com/reference/Dictionary-of-Dry-Goods/Dry-Goods.html, 23.05.2019, COLE, George S: A complete Dictionary of Dry Goods and History of Silk, Linen, Wool and other Fibrous Substances. 1892. 2015.

¹⁴⁵TOMASEK/BAUMAN: Journal of the Text Encoding Initiative, Nr. 6, 2013, S.6.

¹⁴⁶CLARK, George Faber: A History of the Town of Norton, Bristol County, Massachusetts, from 1669-1859. Crosby, Nichols, and Company, and author at Norton, 1859, S.496.

Rechnungsbücher von Wheaton, zeigen neben Angaben zu Gütern, Dienstleistungen und Geldbeträgen in den einzelnen Transaktionen, auch eine Vielzahl von Individuen und ihrer Familien, die in Beschäftigung oder anderen Verhältnissen zu Wheaton stehen, oder an die er Objekte, Tiere und Gebäude vermietete. TOMASEK führt an, dass diese Rechnungsbücher detaillierte Informationen über das tägliche Leben der Einwohnerinnen in Norton, New England geben. Ein städtisches Leben, das durch eine gemischte Land-, sowie Industriewirtschaft im zweiten Quartal des 19. Jahrhunderts, geprägt ist.¹⁴⁷



Abbildung 8: Portrait von Laban Morey Wheaton

Das *Wheaton College Massachusetts* stellt die *Wheaton Family Papers* Dokumente in ihrem digitalen Repositorium online zur Verfügung.¹⁴⁸ Die wissenschaftliche Auseinandersetzung erfolgt durch Kathryn Tomasek, Professorin für Geschichte am *Wheaton College Massachusetts*. Im Zuge des Projektes *The Wheaton College Digital History Project* wurden Teile der Wheaton Papers, darunter beispielsweise das oben angeführte *Daybook* gemeinsam mit Studierenden transkribiert und mit dem Standard der *Text Encoding Initiative* (TEI) ausgezeichnet und im Sinne einer Vorarbeit einer digitalen Edition ediert. Der Fokus dieser Arbeiten liegt auf der Erschließung der einzelnen Transaktionen und der damit verbundenen Personen, Güter und Dienstleistungen.¹⁴⁹ Teile davon wurden im Zuge des DEPCHA Projektes für die Umsetzung eines Prototyps zur Publikation von semantisch angereicherten, digitalen Edition von Rechnungsbüchern verwendet.¹⁵⁰

¹⁴⁷ CLIFFORD, Anderson et al.: *Modeling semantically Enhanced Digital Edition of Accounts as Historical Method*. 2016, S.5.

¹⁴⁸ Wheaton Family Papers, digitalrepository.wheatoncollege.edu/handle/11040/7928, 19.05.2019.

¹⁴⁹ TOMASEK Kathryn: *The Wheaton College Digital History Project: Undergraduate Research in a Local Collection*, writinghistory.trincoll.edu/teach/wheaton-college-digital-history-project-tomasek, 23.05.2019. ALEXANDER, Bryan/DAVIS, Rebecca Frost: *Should liberal arts campuses do digital humanities? Process and products in the small college world. Debates in the digital humanities*, 2012, S.379.

¹⁵⁰ LMW Day Book in DEPCHA, <https://gams.uni-graz.at/o:depcha.wheaton.1>, 29.12.2019.

Im Alltag dieser Zeit in einer Kleinstadt wie Norton trafen sich Menschen regelmäßig und man kannte sich persönlich. Die systematische Buchführung bot Laban Morey Wheaton eine Möglichkeit, seine Geschäfte von seinen persönlichen Interaktionen zu trennen. Zur Dokumentation dieser Geschäfte führte er mehrere Bücher nach dem Prinzip der Doppelten Buchführung und trennte HABEN und SOLL auf zwei Bücher. Transaktionen zwischen lokalen Geschäftspartnerinnen und seinen Kundinnen wurden chronologisch im *Daybook* erfasst, jeweils mit einem Verweis in der linken Spalte auf der Seite im Hauptbuch, auf der der Geschäftsmann die laufende Schuld der Kundin und die Zeiten, zu denen die Schuld beglichen wurde, erfasst hat.¹⁵¹

Abbildung 9 zeigt Seite 100 des *Daybooks*. An diesem Beispiel soll nun die grundlegende Struktur dieser Quelle beschrieben werden. Die linke Seite beginnt mit einer Überschrift "*Norton Tuesday Sept 6 1831*". Die ersten Einträge in dieser tabellarischen Struktur beziehen sich nun auf den 6. September des Jahres 1831. An diesem Tag hat "*Wheaton Wheeler*" einen halben "*Burell Corn*", also ein halbes Büschel Mais, für den Geldbetrag von 50 Cent erworben.

Die Zahl "393" an der ersten Spalte referenziert auf das Hauptbuch und verweist auf eine Seite, auf der Information zum Status der Transaktion zu finden ist und zeigt somit die Doppelte Buchführung. Hier lassen sich auch Wörter wie "*Bill*", "*Paid*" oder "*Settled*" vorfinden, die zeigen, dass der Status einer Transaktion "auf Rechnung" erfolgte, bereits bezahlt oder anderweitig beglichen wurde.

Die mittlere Spalte nennt die zweite Person, neben Laban Wheaton, die an der Transaktion beteiligt ist: den Käufer. Jede hier angeführte Person steht in einem Geschäftsverhältnis mit Wheaton. Jeweils darunter werden die einzelnen Transaktionen gelistet.

In den vier Spalten rechts werden die Geldbeträge gelistet: in der ersten Spalte die Dollar und in der zweiten die Cent, sowie bei der Zusammenfassung mehrerer Transaktionen zu einer Person auch eine Summe.

Im Textfluss lassen sich zwischen den Einträgen entweder einfache Zahlen "12", "24", oder neue Tagesdaten finden, die eben nun alle folgende Transaktionen einem neuen Datum zuordnen. So steht "12" und "24" auf dieser Seite für den 12. und 24. September und "Oct 8th 1831" ordnet alle folgenden Einträge dem 8. Oktober 1831 zu.

Rechts neben den Akteuren in den Einträgen lässt sich meistens ein "D" bzw. "Dr", seltener ein "C" bzw. "Cr" vorfinden, was für "Debitor" und "Creditor" steht. Dies markiert eine Transaktion mit diesem Partner, welche auf die HABEN bzw. die SOLL Seite gebucht wird.

¹⁵¹TOMASEK/BAUMAN: Journal of the Text Encoding Initiative, Nr. 6, 2013, S.7-9.

<div> <div>9¹⁰⁰</div> <div>Norton Tuesday Sept 6 1831</div> </div>			
	Wheaton Wheeler		
393	To 1/2 Burrill Corn	50	
	Samuel Mory	0	
395	By Pension received	120	50
	By Pension for certificate &c. Dr	100	
	" Do. of power	5	04
	" Cash for him \$ 52.00	53	50
395	Thompson Drupp	D	
	Do cash of at Smithy	1	00
395	Thompson Drupp	D	
	Do cash 80/	5	00
	Oct. 8 th 1831		
398	Timothy Smith	D	
	To 6 lbs. Lard 104		
	" 7 1/4 lbs. do 104	1	33
	Oct 10 th 12 " "		
392	William Hodges	Dr	
	to 5 th 11 th Lard 10 pr lb 71		
	to 4 th Butter 18 pr lb 72	1	73
	Oct 11 th " "		
398	Timothy Smith	Dr	
	to 6 lbs 14 th Butter 1/ pr lb 1.12		
	to 14 qts new milk -- 56		
	to 6 qts do. 26	1	97
396	Samuel Mory	D	
	Do cash twenty five Dollars 25.00	25	00
settled	15		
	Wm. Babcock	0	
	By work from this day	\$	2 00
	24		

Abbildung 9: Seite 100 des Daybook, 1828-1859a

Es folgt hier die vollständige Transkription der ersten 8 Einträge der Seite 100. Diese zeigen die soeben beschriebenen Strukturen.

Norton Tuesday Sept 6 1831

	Wheaton Wheeler		
393	Dr To 1/2 Burhell Corn		50
	12		
	Samuel Morey Cr		
396	By pension received	120	
	To Expense pd. for certificates L. D100		
	"Do. of power 50x		
	Cash pd. him \$62.00	53	50
	Thompson Tripp DR		
395	To cash 6/ at Smiths	1	00
	24		
	Thompson Tripp DR		
395	To cash 30/	5	00
Oct. 8th 1831			
	Timothy Smith DR		
368	To 6 lb. Lard 10x 6 lb. Lard"7 1/4 lb do 10x	1	33
Oct 10th			
	Hiram Hodges Dr		
392	To 5 lb11 oz Lard - 19d pr lb		
	71 5 lb11 oz Lard To 4 lb Butter 13 pr lb 72	1	73
Oct. 11th			
	Timothy Smith Dr		
368	To 6 lbs 14 oz Butter 1/ pr lb 1.15		
	To 14 Qts New Milk 56		
	To 6 Qts & 1 Pt. do 26	1	97
	Samuel Morey DR		
396	To cash twenty five dollares 25.00	25	00
	15		
	Mrs. Babcock DR		
Settled	By work from this day pd	\$2	00

Transkription der Seite 100 des Daybook

TOMASEK verfolgt dabei einen TEI/XML-Ansatz, um die textuelle Struktur der Quelle und die inhaltliche Struktur der Transaktionen zu beschreiben. Letztere ist schwerer durch reines TEI Markup beschreibbar bzw. für eine weitere Analyse schwerer nachnutzbar, da man für eine algorithmische Auswertung diskrete Daten braucht.

Das "*Slave-Ledger*", das *Wheaton Daybook* und die Finanzaufzeichnung von George Washington lassen sich auf Grund der regionalen und zeitlichen Einordnung gut miteinander vergleichen und weisen ähnliche textuelle und semantische Strukturen auf. Ihre Kontextualisierung, sowie an die Quellen gestellten Forschungsfragen variieren, lassen sich aber auch wieder auf wiederkehrende Entitäten, wie Personen, Orte und Waren- bzw. Geldflüsse zurückführen. Um eine bessere maschinelle Analyse der Information in diesen Quellen zu ermöglichen, sowie historische Rechnungsunterlagen zu beschreiben, die eine andere Struktur aufweisen, bedarf es standardisierter Modelle. Im nächsten Kapitel wird eine Vorgehensweise und Technologie-Stack beschrieben, der dies ermöglicht.

4 Web of Data

Das *Semantic Web*, oder wie es seit 2013 vom *World Wide Web Consortium* (W3C) genannt wird, *Web of Data* umfasst einen Stack an Standards und Technologien.¹⁵² Dieser Stack dient dazu, Daten und ihre Struktur ausdrucksstark zu beschreiben, sodass Daten maschinenlesbar über das Web verteilt und genutzt werden können. Für die Geisteswissenschaften und die Geschichtswissenschaften im Speziellen eröffnet dieser Technologie Stack, dass Forschungsdaten ausführlich strukturiert, kontextsensitiv, sich selbst beschreibend, genutzt werden können.

Dieses Kapitel dient dazu anfangs die Geschichte, die Idee des und Kritik am *Web of Data* zu beschreiben, geht dann auf den Technologie Stack ein und setzt dann einen Schwerpunkt im Bereich Ontologien. Diese Schwerpunktsetzung erfolgt, da Ontologien eine maschinenlesbare Umsetzung von Modellen ermöglichen und Kapitel 5 sich mit der sogenannten *Bookkeeping Ontology* beschäftigt, die für das Mapping von historischer Information in historischen Rechnungsbüchern verwendet wird.

4.1 Geschichte und Vision des *Web of Data*

*”The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users.”*¹⁵³

Beim *TED Talk* im Jahre 2009 fordert Tim Burners-Lee das Auditorium auf, mit ihm gemeinsam die Worte zu rufen: *”Raw Data Now!”*.¹⁵⁴ Die Vision von Burners-Lee, dem Erfinder des World Wide Web, ist das sogenannte *Semantic Web* (= *Web of Data*). Im Gegensatz zum klassischen Web, das als ein Web von Dokumenten betrachtet werden kann, versucht das *Web of Data* Daten aus unterschiedlichen Quellen zu integrieren und miteinander zu verknüpfen. Daten sollen so vorliegen, dass nicht nur Menschen diese in neuen Kontexten nutzen können, sondern auch Softwareagenten. Maschinen sollen in der Lage sein selbständig die Struktur von Daten ”verstehen” zu können, um bestimmte Aufgaben umzusetzen. Oder anders formuliert: Maschinen sollen in die Lage versetzt werden Inhalte im Web soweit selbstständig verarbeiten zu können, dass Automatisierung auf Ebene der Bedeutung möglich ist. Ein konkretes Anwendungsszenario, das mit Hilfe des *Web of Data* umgesetzt wird sind verbesserte Suchfunktionalitäten für Informationssysteme, die auch die semantische Ebene miteinbeziehen. TOCHTERMANN und MAURER führen ein Beispiel eines Informationsbedürfnisses einer Person an, die gerne einen Termin mit einem Arzt in Graz vereinbaren möchte, der gleichzeitig auch ein Hautarzt ist. Diese Person stellt eine Suchanfrage in einer dafür geeigneten Suchmaschine, bestehend aus 3 Wör-

¹⁵²Linked Data, w3.org/standards/semanticweb/data.html, 23.11.2019

¹⁵³BERNERS-LEE/HENDLER/LASSILA: *Scientific american*, Nr. 5, Bd. 284, 2001, S.3.

¹⁵⁴BERNERS-LEE Tim: *The Next Web*, ted.com/talks/tim_bern timers_lee_on_the_next_web?language=de, 05.04.2019.

tern: "Ärzte, Hautarzt, Graz". Da in einer klassischen Suchmaschine nur das Vorkommen der Wörter berücksichtigt wird, muss die suchende Person sich noch durch die angezeigten Suchergebnisse arbeiten, bis sie den passenden Treffer gefunden hat. Eine Suchmaschine im *Web of Data* ist in der Lage sogenannte Wissensbasen zu befragen, um welche Begriffe es sich hinter den Zeichenketten handelt.¹⁵⁵ So wie sich dieses Alltagswissen formal beschreiben lässt, lässt sich auch Wissen über eine Fachdomäne, wie eben historische Rechnungsbücher, formalisieren. Dabei handelt es sich aber weder um Maschinen die selbstständig lernen, oder gar eine künstliche Intelligenz, sondern um formalisiertes, maschinenlesbares Wissen.¹⁵⁶ Der Wissensbegriff in diesem Zusammenhang entspringt einer informationswissenschaftlichen Perspektive wie bei WERSIG,¹⁵⁷ KUHLEN¹⁵⁸ oder FAUVR-BULLE¹⁵⁹ und sowie im Kapitel 2.3.3 und 4.3 kurz angeführt.

4.1.1 Kritik am *Web of Data*

"I will never touch that crap of RDF or the semantic web; this is a pipe dream of reality ignoring academics and I will not have it. I will only use JSON-LD."

Ist ein Zitat eines Vortrages von Phil ARCHER aus dem Jahr 2014. Viele IT-Entwickler sehen *Web of Data* Standards kritisch und so ist das *Web of Data* über zehn Jahre nach Burners-Lees Aufruf noch immer nicht erreicht.¹⁶⁰; zumindest nicht für die Allgemeinheit. Die großen Internetriesen hingegen verfügen über ihr eigenes "Semantic Web", in denen sie ihre eigenen Agenten mit ihrer großen Datenmenge arbeiten lassen. So verwendet Google beispielsweise seinen *Knowledge Graph*, um Ergebnisse für Suchanfragen zu verbessern.¹⁶¹

Es gib weitere Stimmen, die das *Web of Data* in seiner jetzigen Überlegung kritisieren und es wurde schon vor Jahren als Zukunftstechnologie totgesagt. Die Kritik richtet sich daran, dass die Überlegungen zu akademisch und technisch zu kompliziert sein. SHIRKY kritisiert, dass der top-down Ansatz, die unflexible, strenge Art der Formalisierung und die zu starke Spezifikation auf einzelne Domänen von Ontologien nicht sinnvoll in einem dezentralen Web funktionieren können.¹⁶² Die Kritik von SWARTZ, der sich ein *Semantic Web* herbeisehnt, richtet sich an

¹⁵⁵ PELLEGRINI, Tassilo/BLUMAUE, Andreas: Semantic Web. Wege zur vernetzten Wissensgesellschaft 2006, S.1-2.

¹⁵⁶ ebd., S.1-6.

¹⁵⁷ WERSIG: Information-Kommunikation-Dokumentation.

¹⁵⁸ WELLER, Katrin: Ontologien. In KUHLEN, Rainer/SEMAR, Wolfgang/STRAUCH, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und praxis. 6. Auflage. Berlin and Bosten, 2013.

¹⁵⁹ FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation.

¹⁶⁰ WETTLAUFER, Jörg: Der nächste Schritt? Semantic Web und digitale Editionen. Digital Scholarship in the Humanities, 30 2018 Nr. 1 (URL: http://www.zfdg.de/sb002_007).

¹⁶¹ Whatever Happened to the Semantic Web?, twobithistory.org/2018/05/27/semantic-web.html

¹⁶² SHIRKY, Clay: Ontology is overrated. Categories, Links and Tags 2005.

den, in seinen Augen noch nicht ausgereiften, Technologie Stack und die zu komplizierten und aufgedröselten Standards. Beispielsweise ist für SWARTZ die Aufzeichnungssprache XML im Gegensatz zu JSON (und XML/RDF im Vergleich zu JSON-LD) zu überladen und kann nur mit mehr Aufwand verarbeitet werden.¹⁶³

Festzuhalten ist aber, dass im Aufkommen der AI-Forschung, der Künstlichen Intelligenz, Themen wie *Knowledge Graphs* und Wissensbasen, und so auch das *Web of Data*, wieder interessanter werden. Sie helfen dabei AI-Algorithmen und Entscheidungen zu unterstützen und "Bedeutung" in großen Datenmengen zu extrahieren und zu formalisieren.¹⁶⁴

¹⁶³SWARTZ: Synthesis lectures on the semantic web: Theory and Technology, Nr. 2, Bd. 3, 2013.

¹⁶⁴BERNSTEIN/HENDLER/NOY: Commun. ACM, Nr. 9, Bd. 59, 2016.

4.2 Web of Data Stack

Um diese Vision in die Tat umzusetzen bedarf es mehrerer aufeinander aufbauender, technischer Grundlagen, die sich im *Web of Data Stack* manifestieren. In Abbildung 10 werden die Technologien und Standards des *Web of Data* im sogenannten *Semantic Web Stack* dargestellt. Es gibt unterschiedliche Darstellungen dieses Stacks, dennoch zeigt diese Version alle grund-

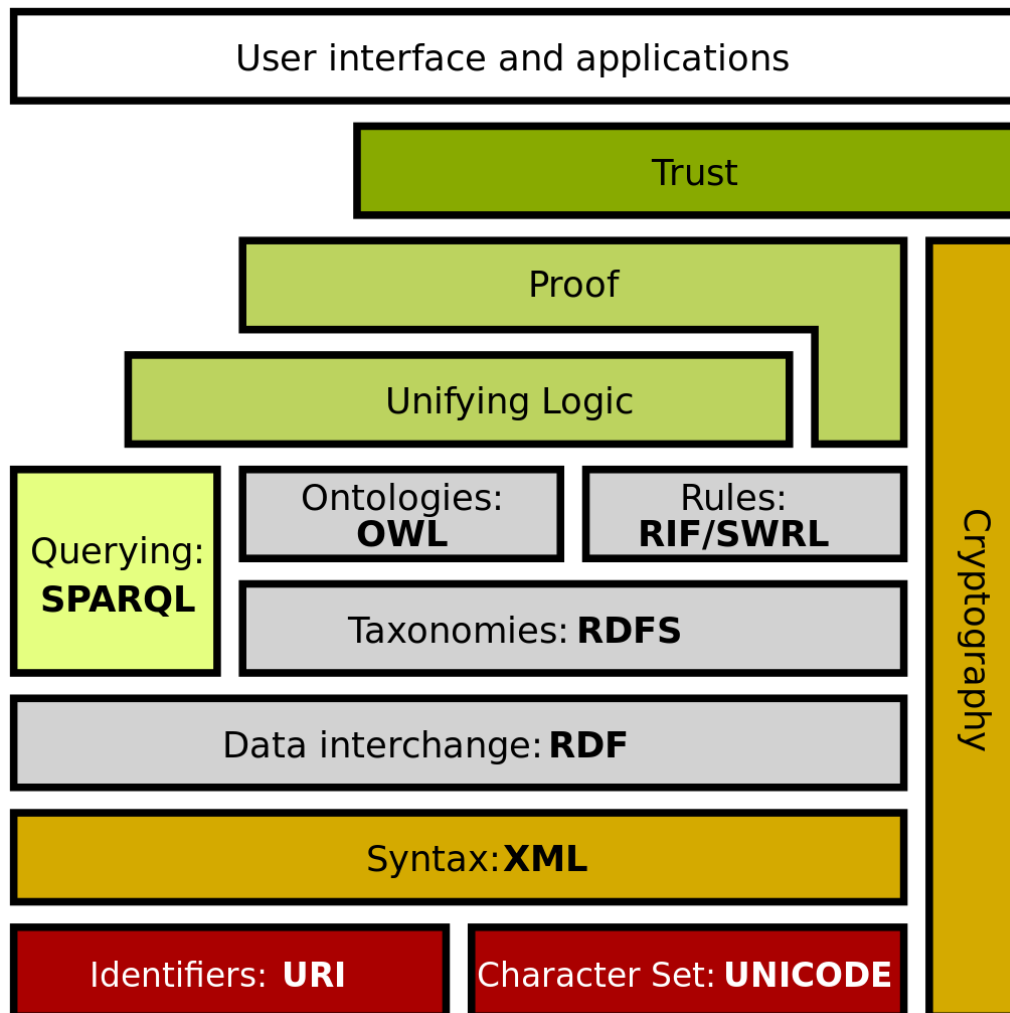


Abbildung 10: Web of Data Stack

legende Konzepte. Der *Web of Data Stack* ist von unten nach oben zu lesen. In den Kästchen in denen eine Abkürzung, wie URI oder RDF, angeführt werden, sind Technologien bereits im Einsatz. Die Themen *Trust*, *Proof* etc. sind noch Bestandteil gegenwärtiger Forschung. Grob zusammengefasst beschreiben diese Konzepte, wie eine vertrauenswürdige und 'wahre' Kommunikation zwischen Akteuren im *Web of Data*, sowie die Verschlüsselung und Anwendung sichergestellt werden kann.

Die Basis des Stacks bilden *Uniform Resource Identifier* (URI), die Ressourcen im Web eindeutig identifizieren und *UNICODE* als internationaler Standard zur Zeichenkodierung. Der

Syntax Bereich wird mit der *Extensible Markup Language* (XML), einer Auszeichnungssprache zur Strukturieren von Text, ermöglicht, wobei diese Technologie immer mehr durch die *JavaScript Object Notation* (JSON) ersetzt wird. Das *Resource Description Framework* dient dazu Daten im Web auszutauschen und ermöglicht es Daten als Graphen zu beschreiben. Die formalen Regeln für RDF werden mittels *Resource Description Framework Schema* RDFS und *Web Ontology Language* (OWL) beschrieben. RDFS erlaubt es Klassen und Beziehungen zwischen Klassen zu beschreiben und OWL ermöglicht es weitere logische Regeln in Modellen abzubilden. *SPARQL Protocol And RDF Query Language* (SPARQL) ist, vergleichbar mit SQL in der relationalen Datenbankwelt, die graphenbasierte Abfragesprache für RDF Daten.¹⁶⁵ In den folgenden Unterkapiteln werden nun die angeführten Technologien weiter beschrieben. Die Konzepte ohne Technologien, sowie *RIF/SWRL*, werden, da sie für diese Arbeit nicht essentiell sind, nicht behandelt.

4.2.1 Daten als Graph: Resource Description Framework

RDF ist ein Datenmodell zur Darstellung und für den Austausch von Daten im Web. Daten werden in diesem Modell als Ressourcen definiert, wobei eine Ressource alles sein kann: ein Dokument, eine Person, ein physisches Objekt oder ein abstraktes Konzept. Über Ressourcen werden Statements der Form Subjekt-Prädikat-Objekt formuliert. Jedes Statement drückt eine Beziehung zwischen zwei Ressourcen aus. Das Subjekt und das Objekt stehen dabei für die beiden miteinander verbundenen Ressourcen; das Prädikat beschreibt die Art ihrer Beziehung. Diese Zusammensetzung von Subjekt, Prädikat und Objekt werden als Triples bezeichnet. Betrachtet man den Satz "*Bob ist befreundet mit Alice*", dann lässt sich folgendes Triple extrahieren: *<Bob>* als Subjekt, *<ist befreundet mit>* als Prädikat und *<Alice>* als Objekt. Ob *Alice* mit *Bob* befreundet ist, geht aus diesem Statement noch nicht hervor, da jeder Relation in RDF nur eine Richtung definiert.¹⁶⁶ In der graphischen Darstellung wird schnell klar, dass es sich beim RDF Datenmodell um einen gerichtete Graphen handelt, der aus Knoten (Subjekt und Objekt), sowie aus Kanten (Prädikat) besteht, wie Abbildung 11 zeigt.

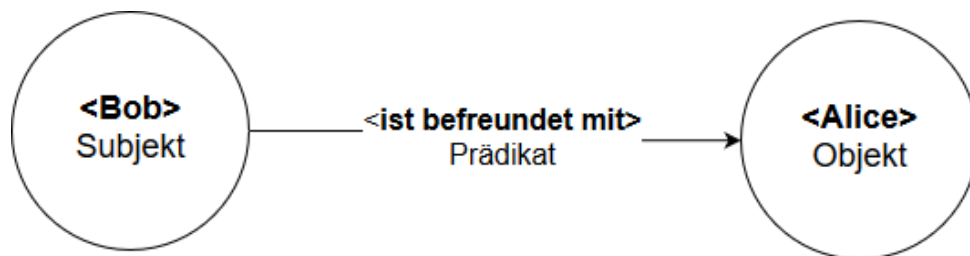


Abbildung 11: Beispiel eines Tripel, eigene Darstellung, 17.07.2019

¹⁶⁵HORROCKS, Ian et al.: Semantic web architecture: Stack or two towers? In International Workshop on Principles and Practice of Semantic Web Reasoning. Springer 2005.

¹⁶⁶POWERS, Shelley: Practical RDF: solving problems with the resource description framework. O'Reilly Media, Inc., 2003, S.16-21.

SCHREIBER und RAIMOND¹⁶⁷ erklären RDF in einem ausführlichen Beispiel an Hand folgenden Aussagen:

Bob ist eine Person.

Bob ist befreundet mit Alice.

Bob ist geboren am 4. Juli 1990.

Bob interessiert sich für die Mona Lisa.

Die Mona Lisa wurde von Leonardo da Vinci entworfen.

Jede dieser Zeilen steht für ein Triple. *Bob* ist Subjekt in vier der oben genannten Tripeln, *Mona Lisa* tritt zweimal als Objekt und einmal als Subjekt auf. Dies ermöglicht es eine beliebige Menge an Triple zu einem komplexeren Graphen zusammenzusetzen und somit komplexere Sachverhalte beschreiben zu können. Abbildung 12 veranschaulicht das.

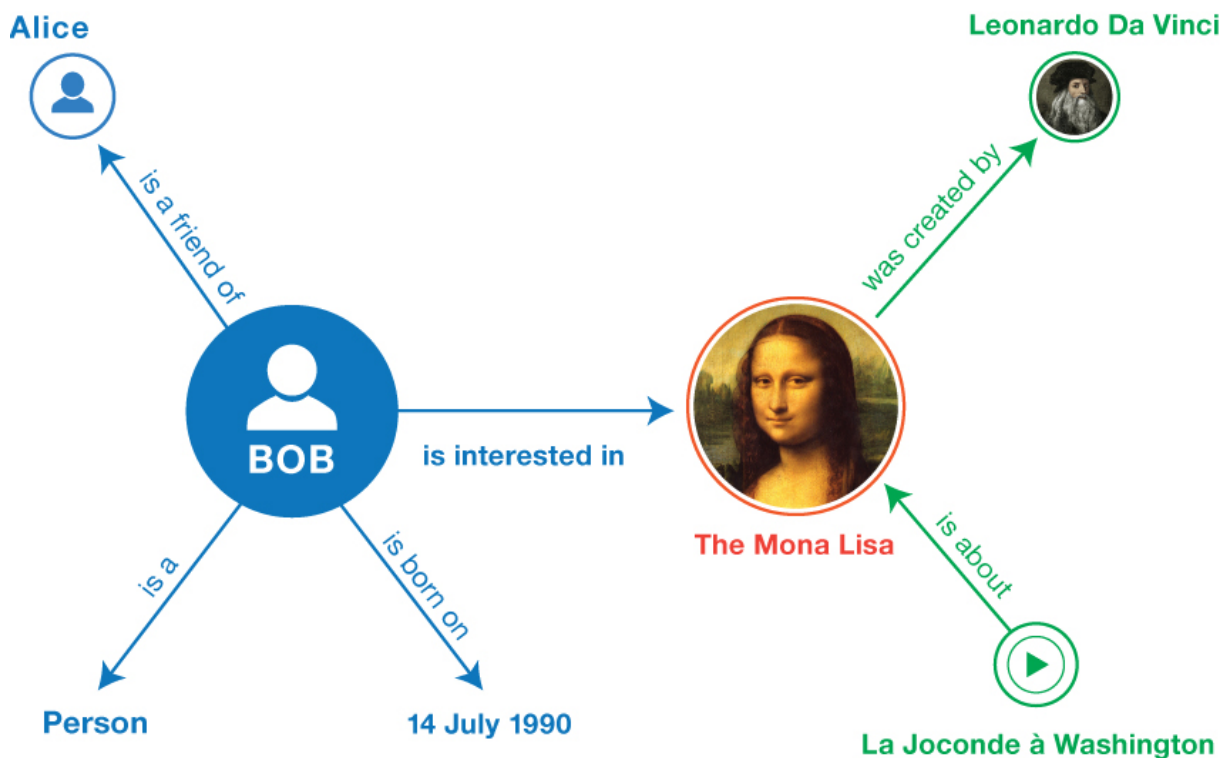


Abbildung 12: Visualisierung eines Graphen auf Basis eines RDF-Datensatz.

URI können in allen drei Positionen eines Triple erscheinen. Somit ist jeder Ressource, sowie jeder Beziehung zwischen Ressourcen durch eine URI identifizierbar. URIs sind global gültige Namen für digitale Ressourcen der Konzepte und dienen der eindeutigen Identifikation und Referenzierung. Sie sind durch ein erweiterbares Schema definiert, damit Ressourcen im Internet eindeutig adressiert werden können.

¹⁶⁷SCHREIBER, Guus/RAIMOND, Yves: RDF 1.1 Primer. W3C working group note 2014.

Um dabei die Einheitlichkeit zu gewährleisten, folgen sie einem vordefinierten Satz von Syntaxregeln, der fünf Komponenten beinhaltet:¹⁶⁸

$$URI = scheme:[//authority]path[?query][#fragment]$$

- **scheme**: Definiert den Kontext und Typ. Bekannte Schemata sind die Webprotokolle *Hyper Text Transfer Protocol* (HTTP) oder das *File Transfer Protocol* (FTP), sowie Notationskonzepte wie *Uniform Resource Name* (URN) oder *Digital Object Identifier* (DOI).
- **authority**: Verwaltet Instanzen in einem vom Schema definierten Interpretationsraum. Ein Beispiel ist das *Domain Name System*, wie etwa *gams.uni-graz.at*
- **path**: Der Pfad enthält – oft hierarchisch organisierte – Angaben, die zusammen mit dem Abfrageteil eine Ressource identifizieren und beschreibt den Weg durch die Filestruktur auf einem Server zu einem Dokument.
- **query**: Der Abfrageteil beinhaltet Daten zur Identifizierung von solchen Ressourcen, deren Ort durch die Pfadangabe allein nicht genau angegeben werden kann, wie beispielsweise ein Datensatz aus einer Datenbank.
- **fragment**: Ist der optionale Fragmentbezeichner und referenziert eine Stelle innerhalb einer Ressource. Der Fragmentbezeichner bezieht sich immer nur auf den unmittelbar vorangehenden Teil des URI und wird von einem Hash (#) eingeleitet und entspricht oft einer ID in einem HTML Dokument.

Weiters werden URI in *Uniform Resource Locator* (URL) und *Uniform Resource Name* (URN) unterteilt. Wo URN Namen von Ressourcen eindeutig identifizieren, wie etwa bei ISBN Nummern von Büchern, sind URL die gängigsten URI's, die den Ort einer Ressource adressieren und über einen Webbrowser auch aufrufen können.¹⁶⁹ Für das Triple *<Bob> <interessiert sich für> <die Mona Lisa>* wird jeder Teilbestand eine URI und in der *Turtle* Serialisation von RDF ergibt es folgenden Code:

```
1 BASE <http://example.org/>
2 PREFIX foaf:<http://xmlns.com/foaf/0.1/>
3 PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
4 PREFIX schema:<http://schema.org/>
5 PREFIX dcterms:<http://purl.org/dc/terms/>
6 PREFIX wd:<http://www.wikidata.org/entity/>
7
8 bob#me a foaf:Person ;
```

¹⁶⁸BERNERS-LEE, Tim/FIELDING, Roy/MASINTER, Larry: Uniform resource identifier (URI): Generic syntax. 2004 – Technischer Bericht (URL: www.rfc-editor.org/info/rfc3986).

¹⁶⁹POWERS: Practical RDF: solving problems with the resource description framework, S.21-22.

```

9      foaf:knows <alice#me> ;
10     schema:birthDate "1990-07-04"^^xsd:date ;
11     foaf:topic_interest wd:Q12418 .
12
13 wd:Q12418 dcterms:title "Mona Lisa" ;
14     dcterms:creator <http://dbpedia.org/resource/Leonardo_da_Vinci> .

```

BASE und *PREFIX* definiert die Namespaces, die als Kurzschreibweise für die *authority* in der URI verwendet werden und jeweils auf ein RDF oder OWL verweisen. Jede Zeile entspricht nur der Formalisierung eines Tripels. Überall stehen URIs, mit Ausnahme von "1990-07-04" und "Mona Lisa". Hierbei handelt es sich um Literale, die bestimmte Datentypen, einmal ein Datum und einmal reinen Text, umfassen.

Verbalisiert bedeutet der Code folgendes: eine Person (bob#me), die als Person nach der *Friend of a Friend* (FOAF) Ontologie definiert ist, kennt eine andere Person (<alice#me>), die auch durch FOAF definiert ist. Weiters hat sie ein Geburtsdatum in Form eines normalisierten Datums und ist interessiert an dem Thema (wd:Q12418). Dieses Thema wird nun weiters beschrieben und es handelt sich dabei um den Datensatz zur Mona Lisa in Wikidata, einer freien und offenen Wissensbasis, die von Menschen und Maschinen befüllt und verwaltet wird.¹⁷⁰

4.2.2 Klassen und Beziehungen: Resource Description Framework Schema (RDFS)

Im vorhergehenden Beispiel wurden <Bob> und <Alice> der Klasse <Person> zu geordnet. Sie sind Instanzen der Klasse <foaf:Person>. Desweiteren wurde eine Relation zwischen diesen beiden Personen definiert: *Bob ist befreundet mit Alice*.

Um die Beziehungen zwischen Ressourcen zu beschreiben, liefert das **Resource Description Framework Schema (RDFS)**¹⁷¹ eine semantische Erweiterung für RDF. Dies umfasst die Möglichkeit Klassen und Relationen, sogenannte Properties, zu definieren und folgt dem Paradigma der Objektorientierung. Es lassen sich auf diese Weise Instanzen von Klassen erzeugen, die alle Eigenschaften der Klasse und ihrer übergeordneten Klassen erben. Blickt man auf das FOAF-Vokabular,¹⁷² das als RDFS umgesetzt ist, so kann man feststellen, dass *foaf:Person* eine Unterklasse von *foaf:Agent* ist. Weitere Unterklassen davon sind *foaf:Group* und *foaf:Organization*. Alle drei Unterklassen haben bestimmte Eigenschaften gemeinsam: sie setzen Handlungen in der Welt. Sie unterscheiden sich aber in den Relationen mit denen sie selbst, oder in Abgrenzung zu anderen Klassen, beschrieben werden können. Eine *foaf:Group* besteht aus mehreren *foaf:Person*. Dies wird mittels der Property *foaf:member* ausgedrückt. Im Gegenzug verfügt *foaf:Person* über eine Relation *foaf:knows*, die ausdrückt, dass sich zwei Personen kennen.

¹⁷⁰Wikidata, wikidata.org, 04.01.2019.

¹⁷¹BRICKLEY, Dan/GUHA, Ramanathan V/MCBRIDE, Brian: RDF Schema 1.1. W3C recommendation, 25 2014.

¹⁷²FOAF-Specification, xmlns.com/foaf/spec/, 29.05.2019.

Die Richtung dieser Relation - wer wen kennt - wird mit den in RDFS mittels den Termen *rdfs:Domain* und *rdfs:Range* definiert. Für *foaf:knows* wird *Domain* und *Range* auf die Klasse *foaf:Person* gesetzt: eine Person kennt eine andere Person. Für die Property *foaf:member* wird *Domain* auf *foaf:Group* und *Range* auf *foaf:Person* gesetzt: eine Gruppe besteht aus Personen. Weiter führt RDFS Datentypen ein. Damit lässt sich beschreiben, um welche Art eines Literals es sich handelt. Es kann für die maschinelle Verarbeitung sehr wichtig sein zu wissen, ob es sich um eine Zeichenkette, eine Zahl, eine Datumsangabe oder eine XML Struktur handelt, da mit unterschiedlichen Datentypen andere Operationen einhergehen.

Mit den RDF Properties *rdfs:label* und *rdfs:comment* lassen sich Properties und Classes benennen und beschreiben. Das ist deswegen nötig, da der Fokus einer Klasse nur in bestimmten Kontexten Sinn macht. *rdfs:label* definiert einen menschenlesbaren Namen einer Ressource. Es besteht stets die Möglichkeit in RDF die einzelnen Labels einer Ressource mit Sprachkürzel zu versehen. Die Property *rdfs:comment* erlaubt es eine verbale Beschreibung zu einer Klasse hinzuzufügen. Am Beispiel von *foaf:Person* ist das Label "Person" und die Beschreibung:

'The Person class represents people. Something is a Person if it is a person. We don't nitpic about whether they're alive, dead, real, or imaginary. The Person class is a sub-class of the Agent class, since all people are considered 'agents' in FOAF. '.

Daneben existiert auch ein Konstrukt *rdfs:seeAlso*, um ausdrücken, dass es unter folgender URL noch weitere Information zu dieser Ressource gibt.¹⁷³ Folgende Darstellung veranschaulicht die soeben beschriebenen Konstrukte in RDFS und zeigt Unterklassen der Superklasse *foaf:Agent*, zwei Instanzen der Klasse *foaf:Person* und stellt graphisch - durch den Pfeil - *Domain* und *Range* einer Property dar.¹⁷⁴

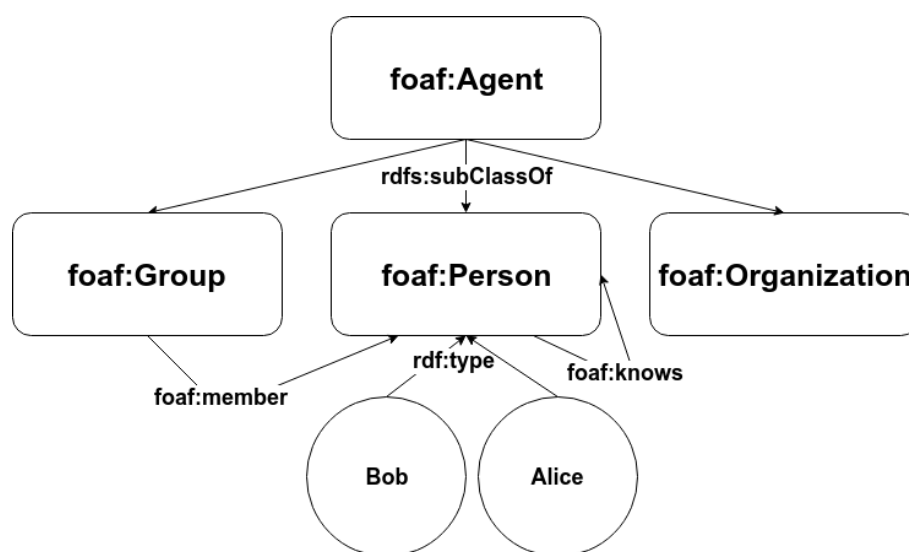


Abbildung 13: RDFS-Beispiel auf Basis des FOAF-Vokabulars

¹⁷³BRICKLEY/GUHA/MCBRIDE: W3C recommendation, Bd. 25, 2014.

¹⁷⁴HITZLER, Pascal et al.: Semantic Web: Grundlagen. Springer, 2007, S.33-88.

Hinter jeder Klasse, Property und jeder Instanz steht eine URI. Folgendes RDF-Snippet zeigt wie die Klasse *foaf:Person* und *foaf:knows* im FOAF-Vokabular mittels RDFS definiert werden.

```
1 PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#> .
2 PREFIX foaf:<http://xmlns.com/foaf/0.1/> .
3 PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
4
5 foaf:Person a rdfs:Class ;
6   rdfs:label "Person" ;
7   rdfs:comment "A person." ;
8   rdfs:subClassOf foaf:Agent .
9
10 foaf:knows a rdf:Property;
11   rdfs:label "knows" ;
12   rdfs:comment "A person known by this person (indicating some
13   level of reciprocated interaction between the parties).";
14   rdfs:domain foaf:Person ;
15   rdfs:range foaf:Person .
```

4.2.3 Abfragesprache: SPARQL

Wie man in der Welt der relationalen Datenbanken mit der Abfragesprache SQL Datenbankabfragen formulieren kann, so kann man mit *SPARQL* *r*(SPARQL Protocol And RDF Query Language) RDF Daten bzw. Triple in Graphdatenbanken abfragen.¹⁷⁵

Folgendes Snippet einer SPARQL-Abfrage zeigt die Syntax dieser Abfragesprache. Ziel dieser Abfrage ist es, alle Ressourcen in einer Graphdatenbank abzufragen, die über eine *foaf:name* und eine *foaf:knows* Property verfügen. Das Ergebnis wird nach den Personen und Namen gruppiert, also es werden keine Dubletten von URI's zurückgegeben. Als Ausgabe erfolgt eine Tabelle (im XML, JSON oder CSV Datenformat) in der die Namen und die Anzahl der Freunde, definiert als alle Knoten auf die *foaf:know* referenziert.

```
1 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
2 SELECT ?name (COUNT(?friend) AS ?count)
3 WHERE {
4   ?person foaf:name ?name .
5   ?person foaf:knows ?friend .
6 }
7 GROUP BY ?person ?name
```

¹⁷⁵GROUP, W3C SPARQL working et al.: SPARQL 1.1 Overview. 2013.

Mit PREFIX werden die Namespaces definiert. Das reservierte Wort *SELECT* definiert alle Variablen. Diese werden durch ein vorangestelltes Fragezeichen gekennzeichnet, die als Rückgabewert definiert werden. Daneben gibt es in der Version 1.1 von SPARQL Operatoren und Funktionen, wie etwas *COUNT()*, das alle Treffer der *?friend* Variabel zählt und in der *?count Variabel* speichert. Im *WHERE* Bereich werden alle Bedingungen für die Abfrage definiert. Diese Bedingungen entsprechen der Definition eines Teilgraphen, der wiederum eine Teilmenge des gesamten Datenbestandes abbildet. Hier stehen weiter Konstrukte wie *OPTIONAL*, einem logischen Oder, so wie *UNION* einem logischen Und zur Verfügung. Über eine SPARQL-Endpoint können so Userinnen über das Web Datenbestände abfragen und mit diesen arbeiten.¹⁷⁶

¹⁷⁶DUCHARME, Bob: Learning SPARQL: querying and updating with SPARQL 1.1. O'Reilly, 2013, S.1-45.

4.3 Formalisierung von Modellen: Ontologien

Der Gegenstandsbereich der **Ontologie als Disziplin in der Philosophie** umfasst alles, das existiert. Das Erkenntnisziel, so MEIXNER, ist auf allgemeiner begrifflicher Ebene zu finden und beschäftigt sich mit der Einteilung des Seins und den Grundstrukturen der Wirklichkeit, sowie der Frage nach dem Wesen der Existenz. Die Ontologie verfolgt nicht das Ziel Erkenntnis über ein Objekt zu erhalten, es beispielsweise zu vermessen oder zu beschreiben, sondern stellt sich die Frage nach welchen allgemeinen Kriterien Objekte im Verhältnis zu ontologischen Begriffen wie Sein, Aktualität, Universalie, Exemplifikation, Sachverhalt oder Individuum stehen.¹⁷⁷

Der Begriff **Ontologie in der Informationswissenschaft bzw. Informatik** umfasst ein pragmatisches Konzept zum Austausch und zur Wiederverwendung von formalisierten und gemeinschaftlich verwendeten Wissensstrukturen durch ein gemeinsames Vokabular. Ziel dabei ist es, Informationssysteme zu implementieren. Die Spezifikation eines solchen Vokabulars für eine bestimmte Domäne, ob übergeordnet und generalisierend, oder fachspezifisch, nennt man Ontologie.

Der Begriff wird in zwei Disziplinen mit jeweils unterschiedlichem Fokus verwendet. Dennoch können Gemeinsamkeiten festgemacht werden. Beide setzen sich mit der Frage auseinander, wie die Welt sinnvoll strukturiert werden kann, damit wir uns besser darin zurecht finden können. In diesem Kapitel wird die informationswissenschaftliche Dimension des Ontologie-Begriffs und seiner Nutzung in den digitalen Geisteswissenschaften diskutiert und der Frage nachgegangen, ob Ontologien ein geeignetes Werkzeug zur Formalisierung von geschichtswissenschaftlichen Domänen darstellen. Dabei soll anfangs "Wissen" kurz aus informationswissenschaftlicher Sicht definiert werden und über das semantische Netz eine Brücke zur Ontologie geschlagen werden.

4.3.1 Vom Wissen, über das Semantische Netz zur Ontologie

*"An ontology is an explicit specification of a conceptualization"*¹⁷⁸

Wissen ist eine systeminterne Repräsentation vorliegender Erfahrungen eines kognitiven Agenten zu einer bestimmten Zeit, die einem zu überprüfenden Anspruch auf Gültigkeit ausgesetzt sein muss. Wissen prägt das Handeln und Denken eines Agenten auf den unterschiedlichsten Ebenen und dient zur Lösung von Problemen im weitesten Sinne. Das jeweils aktuelle Wissen bildet einen kontextuellen Rahmen, in dem bestehende und ankommende Information interpretiert und zu neuen Erfahrungen verarbeitet werden. Diese Definition von Wissen – eine stärker informationswissenschaftliche – hat seinen, neben vielen Definitionen in anderen Fachbereichen, legitimen Ursprung. Unterschiedliche Disziplinen haben andere Fragestellungen und

¹⁷⁷MEIXNER, Uwe: Von der Wissenschaft der Ontologie. Logos (neue Folge), 1 1994.

¹⁷⁸HOEKSTRA, Rinke: Ontology Representation: design patterns and ontologies that make sense. Band 197, Amsterdam: IOS Press, 2009, S.69.

benötigen dafür ein anderes theoretisches Gerüst. Ein Wissensbegriff in der Philosophie, beispielsweise, ist weiter gefasst, als ein Wissensbegriff in der Informationswissenschaft, dessen Aufgabe darin besteht als Hilfsmittel in der Entwicklung und Umsetzung von Informationssystemen zu fungieren.¹⁷⁹

Mittels Ontologie lässt sich "Wissen" als Netzwerk beschreiben. Ein Netzwerk ist ein gerichteter Graph, bestehend aus einer Menge von Knoten und einer Menge von Kanten, die die einzelnen Knoten miteinander verknüpfen. Damit lassen sich (fast) beliebige Entitäten und deren Verknüpfungen miteinander abbilden. Die Überlegungen zu einem **semantischen Netz**, als gedanklichen Vorgänger der Ontologie, stammen von QUILLIAN, der damit ein formales Erklärungsmodell für *'die menschliche Repräsentation von Wissen über Worte und ihre Bedeutung als Netzwerk von Begriffen und ihren Relationen'*¹⁸⁰ beschreibt. Semantische Netze können einen Kompromiss zwischen menschenverständlicher Repräsentation einer Domäne und der formalen Verarbeitbarkeit durch eine Maschine darstellen.¹⁸¹ Das ist dadurch gegeben, dass die Struktur des Graphen (= Netz), sich einfach in Rechnern als Matrizen abbilden lassen. Dies entspricht den Erörterungen von informellen Modellen in Kapitel 2.3.1.

Die Ontologie ist eine Erweiterung des semantischen Netzes und nach GRUBER kann sie durch ein **4-Tupel** definiert werden. C ist eine Menge von **Klassen** (concepts, classes - Mengen von Entitäten aus der Realität), R eine Menge von **Relationen** (properties - Beziehungen zwischen Klassen), I eine Menge von **Instanzen** (individuals - einzelne Entität aus einer Menge) und A eine Menge von **Axiomen** (axioms - logische Regel).¹⁸² C und R lassen sich dabei stets als Graph abbilden. Ein Beispiel zur Veranschaulichung:

Es existiert eine Klasse (C) "Katzen", die mit der Relation "ist ein" (R) mit einer Klasse "Säugetiere" verbunden ist. Die Individuen (I) "Garfield" und "Tom" sind Instanzen der Klasse "Katzen" und erben alle Eigenschaften, die in der Klasse "Katzen" definiert wurden. Eine Regel kann definiert werden (A), sodass immer wenn eine Klasse eine "ist ein" Verbindung zu einer Klasse wie "Säugetiere" hat, es ausgeschlossen ist, dass es eine zweite "ist ein"-Verbindung gibt, die auf eine andere Klasse wie etwa "Vögel" referenziert.

¹⁷⁹FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation.

¹⁸⁰STUCKENSCHMIDT: Ontologien: Konzepte, Technologien und Anwendungen.

¹⁸¹REICHENBERGER, Klaus: Kompendium semantische Netze. Konzepte, Technologie, Modellierung. Springer, 2010.

¹⁸²JOOST BREUKER, Pompeu CASANOVAS/KLEIN, MC/FRANCESCONI, Enrico: The flood, the channels and the dykes: Managing legal information in a globalized and digital world. Law, Ontologies and the Semantic Web: Channelling the Legal Information Flood, 188 2009.

Der Begriff der Ontologie wird terminologisch unscharf verwendet.¹⁸³ Zwar sind die Unterschiede klein, aber sie sind dennoch entscheidend und sollen im Folgenden diskutiert werden. Eine der ersten Definitionen des Begriffs der Ontologie stammt von GRUBER:

"An ontology is an explicit specification of a conceptualization"

Eine *"conceptualization"* beschreibt den Prozess einer Vereinfachung, aber Fokussierung, eines bestimmten Aspekts der Realität. So kann eine Ontologie als Dokumentation eines wissenschaftlichen Prozesses agieren, in dem die Wirklichkeit abstrahiert und reduziert wird und gleichzeitig die Domäne bzw. Forschungsfrage hervorgehoben und amplifiziert wird.¹⁸⁴ Unter *"explicit"* versteht man, dass die Bedeutungen aller von der Ontologie erfassten Begriffe klar und eindeutig definiert sein müssen. Dies beinhaltet alle ihre Eigenschaften, Beschränkungen und Beziehungen, innerhalb, als auch außerhalb der Domäne.¹⁸⁵ BORST erweitert GRUBERS Definition um *"formal specification of a shared conceptualization"*.¹⁸⁶ *"Formal"* ergänzt dabei die Definition um die Notwendigkeit, dass Ontologien maschinenlesbar sein müssen. Erst diese Eigenschaft hebt sie von anderen Methoden zur Formalisierung von konzeptionellen Datenmodellen hervor. Der Zusatz *"shared"* reflektiert die Tatsache, dass eine Ontologie Wissen erfasst, das durch den Konsens einer Gruppe - z.B. durch einen wissenschaftlichen Diskurs - akzeptiert wird. Eine Ontologie sollte nicht im Stillen von einer Person alleine entwickelt werden, sondern in einem iterativen Prozess (**Ontology Engineering**) des Austausches und der Diskussion mit anderen entstehen. Ein solcher Prozess kann wie folgt ablaufen:

- Definition der Notwendigkeit und des Zieles einer Ontologie
- Strukturierung des Wissens und konzeptionelle Entwicklung
- Implementierung und Modellierung
- Evaluierung und Dokumentation
- Iteration dieses Prozesses im Austausch mit anderen

Allgemeiner betrachtet definieren LINCKELS & MEINEL eine Ontologie als ein Datenmodell zur Darstellung eines Sets miteinander vernetzter Konzepte innerhalb einer (Fach-)Domäne.¹⁸⁷ WELLER spricht von einer formalen und schematischen Darstellung einer Wissensdomäne auf Basis definierter Regeln und Vokabulars.¹⁸⁸

¹⁸³GRUBER, Thomas: A translation approach to portable ontology specifications. Knowledge acquisition, 5 1993 Nr. 2, S.1.

¹⁸⁴THALLER: Historische Sozialforschung. Supplement 29 2017.

¹⁸⁵SURE, York/STUDER, Rudi: Methodology, tools & case studies for ontology based knowledge management. 2003.

¹⁸⁶BORST, Willem Nico: Construction of Engineering Ontologies for Knowledge Sharing and Reuse. Dissertation, Netherlands, 1997.

¹⁸⁷LINCKELS, Serge/MEINEL, Christoph: E-librarian service: user-friendly semantic search in digital libraries. Springer, 2011.

¹⁸⁸WELLER: Ontologien.

Zusammengefasst kann gesagt werden, dass sich mittels Ontologien komplexere Sachverhalte so darstellen lassen, dass Mensch und Maschinen in der Lage sind Strukturen, die durch eine Ontologie definiert und standardisiert sind, weiterverarbeiten zu können. Der Mehrwert kann vor allem in der Möglichkeit automatisierter Schlussfolgerungen, im Information Retrieval oder anderen formalen Methoden zur Verarbeitung von Daten liegen.¹⁸⁹

4.3.2 Reasoning

Der Ontology Editor Protégé erlaubt es, eine Ontologie und die darin enthaltenen Daten (Individuals) einem Reasoning - dem Abarbeiten aller vorhandenen Regeln in einer Ontologie auf Basis einer deskriptiven Logik - zu unterziehen. Für solche Zwecke gibt es natürlich auch API's und Bibliotheken in Programmiersprachen.¹⁹⁰ Das Reasoning gilt als ein essentieller Baustein im Design, der Entwicklung, der Wartung und in der praktischen Anwendung einer Ontologie. Das Ergebnis davon sind Inferenzen. Inferenzen sind neu hergeleitete Schlussfolgerungen auf Basis der formalen Regeln einer Ontologie.¹⁹¹ Die Überprüfung strukturierter Daten mittels logischen Schlussfolgerungen kann dazu dienen, größere Datenmengen auf ihre Konsistenz und somit auch auf ihre Qualität hin zu prüfen, da logische Inkonsistenzen als Fehlermeldung angezeigt werden.

¹⁸⁹JANNIDIS/KOHLE/REHBEIN: Digital Humanities: Eine Einführung, S.162-178.

¹⁹⁰MUSEN, Mark A: The Protégé Project: A Look Back and a Look Forward. AI matters, 1 2015 Nr. 4.

¹⁹¹DENTLER, Kathrin et al.: Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web, 2 2011 Nr. 2.

5 Anwendung formaler Methoden auf Basis einer Ontologie

Das abschließende Kapitel verbindet die theoretischen Überlegungen mit den technischen Ausführungen und skizziert die Arbeit, die im Zuge des Projektes **Digital Edition Publishing Cooperative for Historical Accounts (DEPCHA)**¹⁹² umgesetzt wurde. In diesem Projekt wird ein gemeinsamer Publikation-Hub für historische Rechnungsbücher implementiert. Im Zentrum steht die Entwicklung und Nutzung einer Ontologie, einer formalen Beschreibung eines konzeptuellen Modells zur Standardisierung von Buchungstransaktionen in historischen Rechnungsunterlagen. Dabei wird ein im *Web of Data*, sowie ein *Linked Open Data* Zugang verfolgt. Das heißt, dass neben der Transkription und digitalen Edition der Rechnungsbücher, sowie der Zurverfügungstellung der Digitalisate der Quellen, hochstrukturierte RDF-Daten existieren, die auf Grundlage der sogenannten *Bookkeeping-Ontology* formalisiert sind und Referenzen auf Normdaten und LOD-Vokabularien setzen. Mit diesem Zugang soll die Nachprüfbarkeit und die Nachnutzbarkeit der Forschungsdaten und aller darauf aufbauenden Funktionalitäten und deren Ergebnisse gewährleistet sein. Dabei spielt die *Bookkeeping-Ontology* – im THALLER’schen Sinne als *knowledge domain* – eine zentrale Rolle.

In diesem Kapitel wird zuerst auf die digitale Edition von historischen Rechnungsbüchern und die dazu nötigen Überlegungen eingegangen. Darauf aufbauend wird die semantische Anreicherung der digital edierten historischen Quellen im Projektkontext von DEPCHA diskutiert, wofür eine detaillierte Beschreibung der domänenspezifischen *Bookkeeping-Ontology* notwendig ist. Abschließend wird auf die Repräsentation dieser Daten am Beispiel der Informationsvisualisierung von historischer Information in Rechnungsbüchern eingegangen.

5.1 Digitale Edition von historischen Rechnungsbüchern

„Edition ist die erschließende Wiedergabe von historischen Dokumenten“¹⁹³

Grundlage geisteswissenschaftlicher Forschung ist die Erschließung und Verfügbarmachung von textuellen Quellen. Das *Institut für Dokumentologie und Editorik* (IDE) versteht unter *Erschließung* die historisch-kritische Auseinandersetzung mit einem Text, unter *Wiedergabe* die Repräsentation bereits existierender Texte, unter *Dokument* ein allgemeines, pluralistisches Textverständnis und unter *historisch*, die Auseinandersetzung mit Texten, die eine historische Distanz zum Edierenden aufweisen. Die Überwindung dieser Distanz, so das IDE, ist das Ziel der Edition.¹⁹⁴

¹⁹²gams.uni-graz.at/depcha

¹⁹³SAHLE, Patrick: Vom editorischen Fachwissen zur digitalen Edition: Der Editionsprozeß zwischen Quellenbeschreibung und Benutzeroberfläche. In THALLER, Manfred (Hrsg.): FUNDUS - Forum für Geschichte und ihre Quellen. Universität Köln, 2003 (URL: webdoc.sub.gwdg.de/edoc/p/fundus/2/sahle.pdf), S.76.

¹⁹⁴Editionsbegriff des Institut für Dokumentologie und Editorik, i-d-e.de/themen/editorik, 23.11.2019.

Dieser Zugang wird von Disziplinen wie der Editionsphilologie, oder eben den historischen Hilfswissenschaften verfolgt. Geprägt von der Anwendung computergestützter Methoden zur Erstellung, Erforschung und Verbreitung von Quellen, spricht man in den digitalen Geisteswissenschaften von **digitalen Editionen**. Für die Umsetzung kommen weitere Disziplinen, wie die Datenvisualisierung, Historische Fachinformatik oder Computerlinguistik hinzu. Ein Standardwerk zum Thema digitale Edition hat SAHLE verfasst. In diesem dreibändigen Werk diskutiert er den Editions begriff auf theoretischer und technologischer Ebene und arbeitet einen pluralistischen Textbegriff heraus, der in SAHLES Textrad veranschaulicht wird.¹⁹⁵

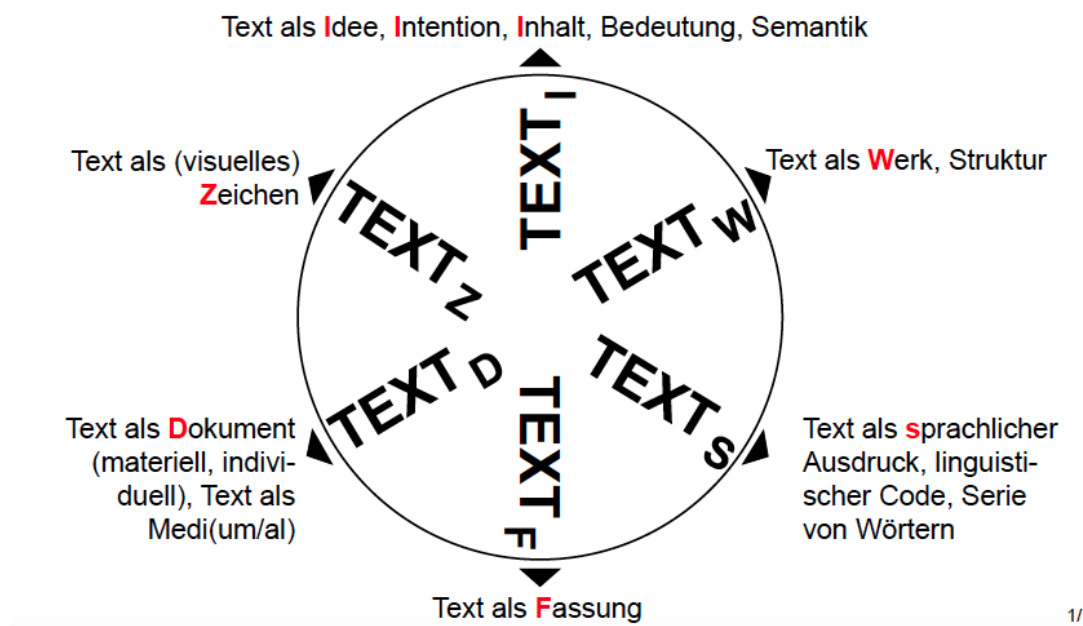


Abbildung 14: Pluralistische Textbegriff nach SAHLE

Abbildung 14 zeigt fünf unterschiedliche Ebenen, die für das Verständnis von Text relevant sind:

- Text als Inhalt: die Idee für den Roman "Schachnovelle" von Stefan Zweig.
- Text als Werk: die Struktur eines Werkes als Sequenz in einem Buch.
- Text als Sprache: der "linguistische Code" der Schachnovelle als sprachliche Äußerung.
- Text als Fassung: verschieden Fassungen des Textes, wie man sie in Zweigs Nachlass findet, wie etwas handschriftlich in einem Manuskript, als Typoskript oder die Korrekturen des Typoskripts.
- Text als Dokument: die materielle Erscheinungsform eines Exemplars in einer Bibliothek.
- Text als Zeichen: die Deutung des Visuellen.¹⁹⁶

¹⁹⁵ SAHLE: Digitale Editionsformen: Textbegriffe und Recodierung.

¹⁹⁶ ebd., S.46.

Die **kritische Texteditionen** hat eine lange Tradition und wird zur Beantwortung unterschiedlicher Fragestellungen in den Geschichtswissenschaften herangezogen. Es handelt sich um eine Methode, die darauf abzielt, die Abweichungen unterschiedlicher Überlieferungen eines Textes zu bearbeiten und versucht das Material für eine weitere inhaltliche Analyse bereitzustellen. Dabei handelt es sich um eine hermeneutische Methode. In ihr wird die Quelle eingeordnet, dokumentiert und hinsichtlich ihrer Authentizität, sowie Vertrauenswürdigkeit überprüft. Gerade das Schaffen von Vertrauen, sodass nicht jede Quelle erneut auf Punkt und Beistrich geprüft werden muss, ist eine Kernaufgabe der Edition und wichtiger Bestandteil des wissenschaftlichen Forschungsprozesses und wichtiges Kriterium für die wissenschaftliche Auseinandersetzung mit ihr.¹⁹⁷

Die Idee einer computergestützten Edition wird seit den 1970er Jahren diskutiert. Der Text wird formal in einer Datenbank abgelegt und aufbereitet, damit Abfragen an den Text, oder besser gesagt die Daten, die der Text umfasst, möglich sind. Besonders die Verwendung der Auszeichnungssprache XML und des de facto Standards zur Textauszeichnung TEI hat sich für den Datenaustausch und die Langzeitarchivierung von digitalen Editionen als brauchbar erwiesen.¹⁹⁸ Diese Kombination ermöglicht es editorische Phänomene in Texten mittels einem maschinenlesbaren und kollektiven, also von einer Fachcommunity weiterentwickelten, Vokabulars auszuzeichnen. Nicht eine reine technische Lösung erlaubt es Text formal zu beschreiben und auszuwerten, sondern liegen viele Überlegungen in den Daten selbst und sind so offen für unterschiedliche technische Lösungen.¹⁹⁹ Die digitale Edition²⁰⁰ ist durch folgende Paradigmen^{201, 202} gekennzeichnet:

- **Offenheit:** die digitale Edition erlaubt es den ursprünglichen Text und seiner Annotationen in beliebige weitere Kontexte abzubilden.
- **Vom Produkt zum Prozess:** eine Printfassung ist ein Produkt eines wissenschaftlichen Prozesses. Die digitale Edition aber ist eine Form eines wissenschaftlichen Prozesses, der leicht veränderlich und dokumentierbar ist. Jedoch stellt diese Möglichkeiten noch ein offenes Forschungsthema der digitalen Geisteswissenschaften dar.
- **Vernetzung und Zugänglichkeit:** Wissenschaft lebt von der Auseinandersetzung mit anderen und somit ist die Vernetzung von Information und Wissen ein wichtiger Bestandteil.

¹⁹⁷ AUMANN, Stefan et al.: From digital archive to digital edition. Historische Sozialforschung, 1999, S.101-144.

¹⁹⁸ BURNARD, Lou: What is the Text Encoding Initiative?: How to add intelligent markup to digital resources. OpenEdition Press, 2014, 3.

¹⁹⁹ VOGELER, Georg: Religion aus Daten? : Zur digitalen Edition von religiösen Grundlagentexten. 2018, S.180-181.

²⁰⁰ Zu diesem Thema siehe auch: PIERAZZO, Elena: Digital scholarly editing: Theories, models and methods. Routledge, 2016.

²⁰¹ SAHLE: Digital Humanities. Eine Einführung 2017, S.240-241.

²⁰² SCHOLGER, Martina: Assoziationsprozessen auf der Spur: digitale Edition der Notizbücher von Hartmut Skerbis. Dissertation, Universität Graz, 2018, S.102-108.

Die digitale Edition, über das Web verteilt und als Prozess verstanden, erlaubt es Inhalte mit andern leichter zu vernetzen.

- **Multimedialität:** da es keine Begrenzung gibt, wie in einer gedruckten Edition, können sämtliche Ebenen der Edition in unterschiedlichen Formen zugänglich gemacht werden. Etwa alle Faksimile eine Quelle und nicht nur ausgewählte Seiten, sowie zusätzliche Präsentations und Repräsentationsformen, wie etwas Informationsvisualisierung als Zusammenfassung bestimmter Strukturen in einem Text.

Die Umsetzung einer digitalen Edition ist auch mit neuen Herausforderungen verbunden. Es ist notwendig, dass die langfristige Sicherung der Forschungsdaten, die Zitierbarkeit und Authentizität der Edition gewährleistet sind. Umso mehr ist es notwendig, dass sorgsam methodisches Vorgehen dokumentiert wird und in einer Form als *knowledge domain* abgebildet wird, um eine formale Quellenkritik und -analyse transparent halten zu können.²⁰³

5.1.1 Digitale Edition und Web of Data

*"Für die geschichtswissenschaftliche Forschung stehen die Inhalte der Dokumente im Vordergrund."*²⁰⁴

Dieser Aussage von SAHLE folgend bedarf es eines Werkzeuges im Kontext der digitalen Edition, um die Inhalte, die semantischen Strukturen, einer Quelle formal zu beschreiben. Dieses Werkzeug kann, so VOGELER, das *Web of Data* sein und überführt die digitale Edition in eine sogenannte **"Assertive Edition"**.

Historikerinnen haben ein besonderes Interesse an "historischen Fakten". So sind Methoden notwendig, deren Ziel es ist, formale Darstellungen der durch den Text vermittelten Informationen in strukturierten Datenbanken abzubilden. Um auch den Ansprüchen der textuellen Ebenen einer Quelle gerecht werden zu können, schlägt VOGELER die Verwendung von eingebetteten Referenzen im TEI-Markup vor, die es erlauben aus dem Text extrahierte Aussagen in eine RDF-Repräsentation zu überführen. Die Logik dieser Referenzen, sprich der jeweiligen Domäne bzw. Forschungsfrage, wird als Ontologie abgebildet.²⁰⁵

MEYER²⁰⁶ vergleicht 52 digitale Editionsprojekte hinsichtlich ihres Expressivitätsgrades und ihres Abstraktionsgrades. Unter Expressivität versteht er wie komplex und ausdrucksstark die Inhalte bzw. Daten in einer Edition ausgedrückt werden und kategorisiert diese in drei Gruppen:

²⁰³ KROPAČ, Ingo: Theorien, Methoden und Strategien für multimediale Archive und Editionen. *Mediaevistik und Neue Medien* 304 2004, S.306-315.

²⁰⁴ SAHLE: *Digital Humanities. Eine Einführung* 2017, S.236.

²⁰⁵ VOGELER: *International Journal of Digital Humanities*, Bd. 1, 2019.

²⁰⁶ MEYER, Till: *Datenmodellierung für digitale Editionen*. Diplomarbeit, Humboldt-Universität zu Berlin, 2019.

Referenzierung über Normdaten (wie etwa GND oder GeoNames), Aufbereitung als und Verknüpfung mit LOD-Ressourcen und Modellierung interner Bezüge. Die drei Abstraktionsgrade, die den Expressivitätsgraden gegenüberstehen, beschreiben

- konkrete Entitäten in stark/schwach strukturierten Texten (Person in einem Roman)
- konkrete und abstrakte Entitäten in stark strukturierten Texten (Person oder Transaktion in einem Rechnungsbuch)
- abstrakte Entitäten in schwach strukturierten Texten (philosophisches Konzept in einem Manuskript).

Er kommt zu dem Schluss, dass *Web of Data* Zugänge es ermöglichen, tiefergehende Inhalte zu beschreiben und nutzbar zu machen, sieht aber auch noch keine Standardlösung zur Umsetzung des Technologie Stack des *Web of Data* und die damit verbunden Paradigmen in digitalen Editionen.

Die Anbindung an ***Linked Open Data*** Hubs, wie beispielsweise *Wikidata* oder *DBpedia.org* spielen eine hervorgehobene Rolle in der Vernetzung von Entitäten und Information über Editi-
onsprojekte hinweg.²⁰⁷ Dort kann eine stabile und über URIs adressierbare kollaborative Wissensbasis genutzt werden, um Daten auf übergreifende Konzepte zu referenzieren, was für eine weitere Anreicherung, Normalisierung und Aufbereitung für das Finden genutzt werden kann. Im Kapiteln 5.2 wird dieser Zugang am Beispiel des DEPCHA Projektes detaillierter ausgeführt.

²⁰⁷WETTLAUER: Digital Scholarship in the Humanities, Nr. 1, Bd. 30, 2018.

5.2 Die Bookkeeping-Ontology

Die *Bookkeeping Ontology*²⁰⁸ ist ein konzeptionelles Modell zur formalen Beschreibung von Transaktionen in historischen Rechnungsunterlagen. Es wurde in einem Ontologie-Engineering-Prozess entwickelt, an dem Historikerinnen, Softwareentwicklerinnen und digitale Geisteswissenschaftlerinnen beteiligt waren. Die Ontologie befindet sich nach wie vor in der Entwicklung (Stand 2020). Sie liegt der Arbeit von VOGELER zugrunde.²⁰⁹

5.2.1 Grundlagen

Ausgangspunkt ist das *Resources, events, agents* (REA) Modell,²¹⁰ eine Ontologie, die Buchhaltungsprozessen virtualisiert. Wie der Name sagt, besteht es aus Ressourcen (Gütern, Dienstleistungen, Geld), die in Ereignissen, also Vereinbarungen oder Geschäftstransaktionen, von einem Akteur (Person, Organisation) zum anderen transferiert werden. Im Mittelpunkt jedes REA-Modells steht in der Regel ein Ereignispaar, das durch eine Austauschbeziehung verbunden ist. Eines dieser Ereignisse repräsentiert gewöhnlich den Abgang, die andere den Empfang einer Ressource. Im Verkaufsprozess wäre zum Beispiel ein Ereignis ein "Verkauf" - bei dem Waren abgegeben werden - und das andere Ereignis ein "Geldeingang", bei dem Bargeld empfangen wird. Diese beiden Ereignisse sind miteinander verbunden: ein Geldeingang erfolgt im Austausch gegen einen Verkauf und umgekehrt.

Eine weitere Grundlage bilden die Arbeiten im Kontext von *Modeling Semantically Enhanced Digital Edition of Accounts* (MEDEA).²¹¹ Diese Arbeitsgruppe hebt gerade die Möglichkeiten der Normalisierung und fachwissenschaftliche Annotation von Information in historischen Rechnungsbüchern hervor und erörtert auch die Rolle der *extensible Business Reporting Language* (XBRL),²¹² ein XML-basierter Standard "for reporting financial information". Weiters wird die Ableitung eines konzeptuellen Modells zur Beschreibung von Transaktionen von der Top-Level-Ontology CIDOC-CRM²¹³ diskutiert.²¹⁴ Diese Überlegungen fließen in die Entwicklung der Bookkeeping-Ontology ein. Eine andere Ontologie, die sich in diesem Umfeld bewegt und wirtschaftliche Aktivitäten beschreibt sind die Überlegungen von VAJDA et.al..²¹⁵ Eine tiefergehende Auseinandersetzungen mit diesen Überlegungen steht noch aus.

²⁰⁸Bookkeeping Ontology, <https://gams.uni-graz.at/o:depcha.bookkeeping>, 24.08.2020

²⁰⁹Ontology for the RDF description of historical accounting, <https://github.com/GVogeler/bookkeeping>, 08.09.2020

²¹⁰MCCARTHY, William E: The REA accounting model: A generalized framework for accounting systems in a shared data environment. Accounting review, 1982.

²¹¹<https://medea.hypotheses.org/>, 07.09.2020

²¹²XBRL, <https://www.xbrl.org/>, 25.08.2020

²¹³CIDOC Conceptual Reference Model (CRM, <http://www.cidoc-crm.org/>, 25.08.2020

²¹⁴CLIFFORD et al.: Modeling semantically Enhanced Digital Edition of Accounts as Historical Method, S.9-12.

²¹⁵VAJDA, Jonathan/MERRELL, Eric/SMITH, Barry: Toward an ontology of commercial exchange. 2019.

5.2.2 Beschreibung

In der derzeitigen Fassung definiert die *Bookkeeping Ontology* folgendes:

Eine Transaktion besteht aus mindestens einem Transfer. Jeder Transfer umfasst den Austausch von Dingen, die quantifizierbar sind, von einem Akteur zum anderen. Solche, sogenannte *Measurables*, lassen sich in Geldbeträge, sowie wirtschaftliche Güter unterteilen. Geldbeträge sind dadurch gekennzeichnet, dass sie aus einer Zahl und einer Währung bestehen, wie etwa 10 US Dollar. Weiters gibt es spezielle Formen von Geldbeträgen, wie etwa Steuern, die eine einseitige Geldabgabe definieren, sowie Preise, die den monetären Wert eines Wirtschaftsgutes beschreiben. Die wirtschaftlichen Güter wiederum werden aufgeteilt in Waren – bestehend aus Menge, Einheit und Art, wie 1 Sack Erdäpfel – und Dienstleistungen, die eine zeitliche Komponente mit sich bringen können. Die Akteure diesen Transfers können Individuen, Gruppen, Organisationen, Konten oder Kategorien innerhalb von Konten, wie beispielsweise alle Einnahmen einer Stadt auf den Ausschank von Wein, sein. Die Ontologie erlaubt es explizit zu formulieren von wem an wen *Measurables* transferiert werden.

Weiters ist die Verbindung zur historischen Quelle bzw. zum jeweiligen Eintrag wichtiger Bestandteil der Ontologie. Ein Eintrag ist ein Informationsfragment und Beleg eines Ereignisses einer Transaktion in der Vergangenheit. Neben dieser Referenz besteht die Möglichkeit jede Transaktion in seiner zeitlichen, räumlichen und inhaltlichen Dimension, also der Zordnung zu einem bestimmten, durch die Forschungsfrage geprägten, Kontext, zu beschreiben.

Ein Transfer, als Teilbestand einer Transaktion, kann auch von einem *Agent*, also einer weiteren Person, durchgeführt werden: jemand übernimmt den Transferprozess für jemand anders. Bei der Verbuchung der Transaktion wird eine Transaktion im Sinne der doppelten Buchhaltung auf die Haben oder Soll Seite gebucht.

5.2.3 Beispiel

Zu Beginn dieser Arbeit (Kapitel 2) wurde ein Beispiel angeführt, indem ein Akteur, James Haley, am 01.08.1808 ein 1/4 Pfund Pulver, 1 Pfund Munition und 1 Pfund Zucker zum Preis von 2 Schilling und 6 Pence im Laden der *Stagville Plantage* in *North Carolina* käuflich erworben hat. Am folgendem Beispiel wird gezeigt, wie die Annotation des XML/TEI aussieht, um eine Transaktion im Sinne der Bookkeeping-Ontologie zu beschreiben und wie die daraus extrahierten RDF-Triple aussehen.

XML/TEI Beispiel mit Bookkeeping-Annotation

Ausgangspunkt ist die Definition eines Containers, der die textuelle Struktur *bk:Entry* und die semantische Struktur *bk:Transaction* zusammenführt. Für die Annotation im XML/TEI wird das Attribut *@ana* verwendet, das eine oder mehrere Interpretationen eines Elements und dessen Inhalts erlaubt. Der *bk:Entry* stellt oft in Rechnungsunterlagen entweder eine Zeile in einer

Tabelle oder einen Eintrag in einer Liste dar, ist aber nicht darauf reduziert.

Die hier dargestellte *bk:Transaction* besteht aus zwei *bk:Transfers*: der *bk:Transfers* von Pulver, Zucker und Schrot, und der Fluss des Geldes, also der *bk:Transfers* von 7 Schilling und 6 Pence.

```
1 <head>
2   <name ana="bk:to" ref="depcha:stagville">Stagville</name>
3   <date ana="bk:when" when="1808-08-01"> August 1st 1808</date>
4 </head>
5 <table>
6   <head>
7     69 <name ana="bk:from" ref="depcha:pers.2">James Haley</name>
8     <span ana="bk:status">Self Dr.</span>
9   </head>
10  <row ana="bk:entry" xml:id="Transaction-0">
11    <cell>
12      <measure ana="bk:commodity" commodity="wd:Q12861" quantity="0.25"
13        unit="wd:Q100995">1/4 lb Powder <measure ana="bk:price"
14        quantity="0.33" unit="wd:Q213142">2/6</measure>
15    </measure>
16  </cell>
17  <cell>
18    <measure ana="bk:money" quantity="7" unit="wd:Q213142">7</measure>
19  </cell>
20  <cell>
21    <measure ana="bk:money" quantity="6" unit="wd:Q234129">6</measure>
22  </cell>
23 </row>
24 </table>
```

Neben trivialen Markup, wie beispielsweise der Auszeichnung des Datums mittels `<date ana = "bk:when">`, wird der Geldfluss durch die Werte `@ana = "bk:from"` und `@ana = "bk:to"` ausgedrückt. Da in den meisten Quellen dieser Art ein Kauf einer Ware dokumentiert ist, handelt es sich hierbei um die Default-Annahme. Daraus lässt sich schließlich der Fluss von Wirtschaftsgütern – einer Ware oder Dienstleistung – ableiten. Einzelne Einträge nennen in der Regel nur einen Akteur einer Transaktion, während die zweite Seite, der Ladenbesitzer des Stagville-Stores, indirekt, aus dem Kontext der Fachwissenschaftlerin, oder auf einer vorangestellten Seite, abgebildet ist. Darüber hinaus kann der Text *"Self Dr."* in der Kopfzeile der Buchung als Transaktions- und Buchungsstatus interpretiert werden und drückt aus, dass die Buchung von dieser Person auf der Sollseite eines doppelten Buchungsstils vorgenommen wurde.

Ein weiterer wichtiger Aspekt ist die Normalisierung von Entitäten wie Personen, Orten, Währungen, Waren oder Dienstleistungen, die im Zusammenhang mit einer solchen Quelle sehr komplex werden können. So sind beispielsweise bestimmte Gewichtseinheiten oder Währungen von regionalen und zeitlichen Unterschieden betroffen. Aufgrund der großen Vielfalt kon-

textsensitiver Entitäten wird die Normalisierung der Daten durch die Verknüpfung mit *Wikidata.org* oder durch die Erweiterung von Workflows für andere LOD-Vokabulare erreicht. Dies löst natürlich keineswegs die komplexen geschichtswissenschaftlichen Herausforderungen, die mit diesen Normalisierungen einhergehen, die Verwendung von Wikidata könnte aber im Sinne einer großen Wissensbasis ein helfendes Werkzeug sein.

Wikidata erlaubt es eigenständige neue Konzepte zu erstellen, die noch in keinem Vokabular vertreten sind. Das folgende Beispiel verdeutlicht dies: ein weiterer in DEPCHA aufgenommener Datensatz enthält Listen von Waren aus der Stadt Regensburg im 13. Jahrhundert.²¹⁶ In diesem Datensatz wird der Begriff "*Scheffen Hafer*" verwendet und bezieht sich auf eine ganz bestimmte zeitliche und regionale historische Einheit. Der oben beschriebene Arbeitsablauf erlaubt es, solche Begriffe direkt mit *Wikidata.org*-Konzepten zu verbinden, indem die Q-Nummer direkt an ein Attribut angehängt wird. Dieser pragmatische Ansatz wird durch Werkzeuge wie OpenRefine²¹⁷ für die halbautomatische Identifizierung unterstützt.

Als Ergebnis dieses Workflows extrahiert eine XSL-Transformation die annotierten Strukturen im XML/TEI und überführt sie in ein RDF, dessen Grundlage die Bookkeeping-Ontologie darstellt. Das folgende RDF/Turtle-Snippet zeigt das verarbeitete XML/TEI:²¹⁸

RDF-Extraktion

Danach folgen die RDF-Triple, die aus dieser Annotation extrahiert werden können.

```
1 @prefix bk: <https://gams.uni-graz.at/o:depcha.ontology> .
2
3 <#Transaction-0>
4   a bk:Transaction ;
5   bk:consistsOf <#Transfer-1>, <#Transfer-2> ;
6   bk:when "1808-08-01" ;
7   bk:text "1/4 lb Powder 2/6 1 lb Shot 2/6 7 6" .
8
9 <#Transfer-1>
10  a bk:Transfer ;
11  bk:transfers <#Measurable-1> ;
12  bk:from <#stagville> ;
13  bk:to <#pers.2> .
14
```

²¹⁶StadtA Regensburg, Kameralien 3

²¹⁷Open source Desktop Anwendung zur Datenbereinigung und Datentransformation, <https://openrefine.org>, 26.08.2020

²¹⁸POLLIN: Digital Edition Publishing Cooperative for Historical Accounts and the Bookkeeping Ontology, S.10-15.

```

15 <#Transfer-2>
16   a bk:Transfer ;
17   bk:transfers <#Measurable-2-1> ;
18   bk:from <#pers.2> ;
19   bk:to <#stagville> .
20
21 <#Measurable-1>
22   a bk:Commodity ;
23   bk:unit <wiki:Q100995> ;
24   bk:quantity "0.25" ;
25   bk:commodity <wikivQ2908004> ;
26   bk:price <#Price-1> ;
27   bk:text "1/4 lb Powder" .
28
29 <#Measurable-2-1>
30   a bk:Money ;
31   bk:unit <wiki:Q213142> ;
32   bk:quantity "7" .
33
34 <#stagville>
35   a bk:Between;
36   bk:name "Stagville Store".
37
38 <#pers.2>
39   a bk:Between;
40   bk:name "James Haley".

```

Abbildung 15 stellt die angeführte Transaktion schematisch dar. Die hellen Knoten definieren Instanzen von Klassen der Bookkeeping-Ontologie, die im Sinen von *Object Properties* mit anderen Instanzen verknüpft sind, bzw. durch weitere Attribute, eben *Data Properties*, beschrieben werden. Angedeutet wird die Normalisierung der Währungen und Maßeinheiten über Wikidata mit den angeführten Q-Identifikatoren. Die RDF-Daten bilden genau diese Graphenstruktur ab, und können so in einer Graphdatenbank verwendet werden.

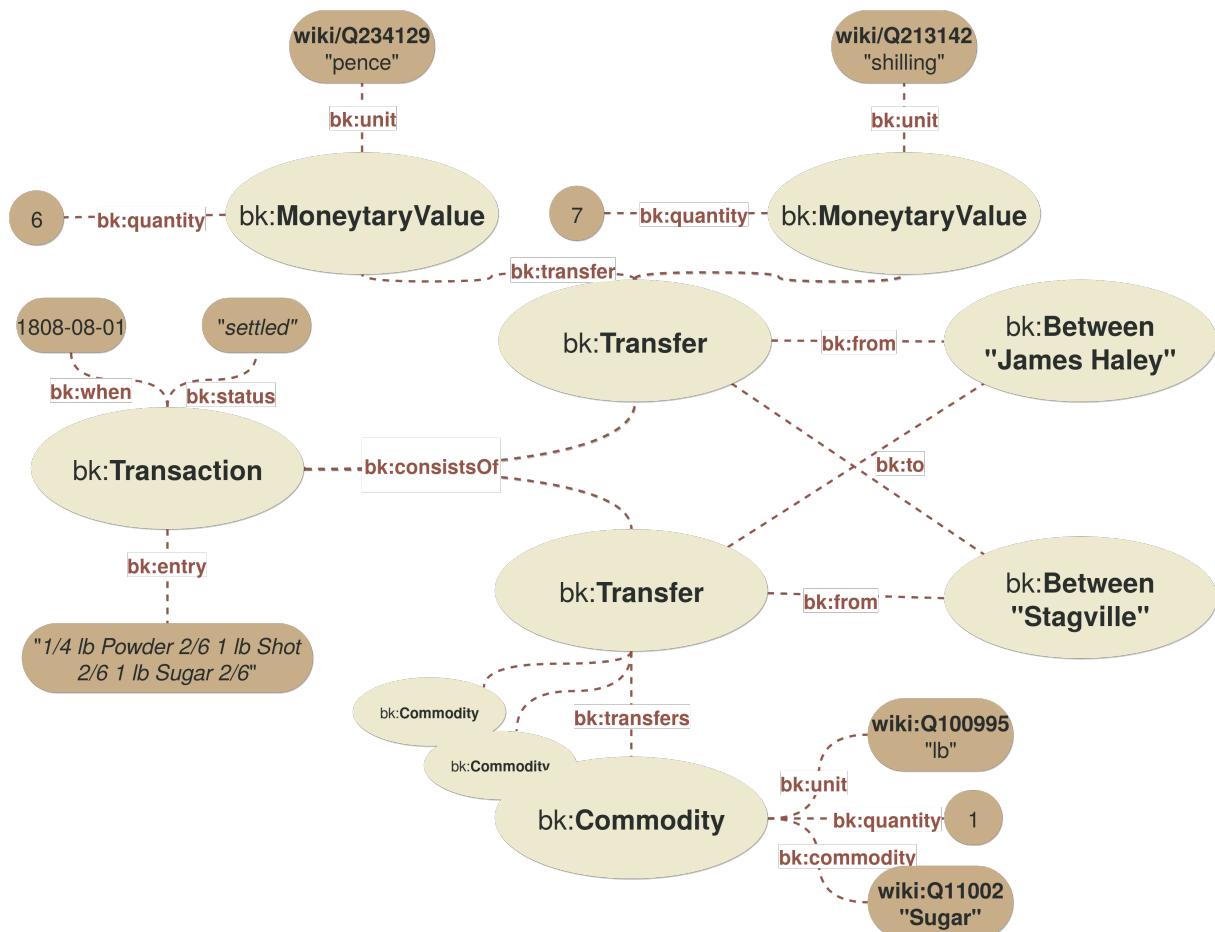


Abbildung 15: Graphische Darstellung des RDF

5.3 Informationsvisualisierung

*”The time of diagrammatic thinking is upon us. We need graphical interfaces for multidimensional and multimedia authoring that take advantage of computers’ abilities to aggregate, synthesize, and organize arguments along multiple axes”*²¹⁹

Arbeitet man mit historischen Rechnungsbüchern, so arbeitet man mit einer Vielzahl von Informationsfragmenten. Für Forschungsfragen ist oft der Blick auf das Ganze interessant und so auch die Aggregation und Synthese der einzelnen Belege. BURDICK sieht gerade im Aufkommen der digitalen Geisteswissenschaften die Zeit reif, die Ergebnisse der computergestützten Verarbeitung multidimensional zu visualisieren. Die Informationsvisualisierung ist eine Methode, die dabei helfen kann Erklärungen historischer Ereignisse, die aus der Zusammenschau vieler Quellen und Entitäten zusammenkommen, zu argumentieren und sichtbar zu machen.²²⁰ Sie dient nicht dazu eine ”historische Wahrheit” abzubilden, die auf einer Zahlenbasis errechnet wurde, sondern kann vielmehr ein Werkzeug im hermeneutischen Prozess sein, um eine Domäne besser verstehen zu können.

Als Visualisierung versteht man – neben dem eigenständigen Fach aus dem Bereich *Human-Computer Interaction* – die Anwendung computerbasierter, interaktiver und visueller Repräsentation von abstrakten, nicht-physikalischen Daten, die eine Hilfestellung für den Erkenntnisgewinn sein können. Die Daten können sowohl quantitativer, wie etwa geographische Koordinaten oder Messwerte, sowie qualitativer Natur sein. Die Fachliteratur unterscheidet zwei Herangehensweisen. Wo die **Scientific Visualization** die Welt so darstellt, wie sie ist, beispielsweise die naturgetreue Darstellung des menschlichen Körpers, wird in der **Information Visualization** ein Datenbestand in neuen (abstrakten) Informationsräumen, wie etwa in einem Koordinatensystem dargestellt. MUNZER fasst dies zusammen als *”Its scientific visualization when the spatial representation is given. Its Information Visualization when the spatial representation is chosen”*.²²¹ Im Falle von Rechnungsbüchern und den Geschichtswissenschaften generell, kann man stets von einer Informationsvisualisierung ausgehen. Eine naturgetreue Darstellung historischer Ereignisse ist nie möglich.

5.3.1 Voraussetzungen

Die Menschliche Wahrnehmung ist multimodal. Es entstehen unterschiedliche Repräsentation in unserem Bewusstsein auf Basis der aufgenommen Wahrnehmungen unserer Umgebung. Bei der Umsetzung von Anwendungen der Informationsvisualisierung sind für PREIM und DACH-

²¹⁹BURDICK, Anne et al.: *Digital Humanities*. Mit Press, 2012, S.119.

²²⁰FRANK: INF-DH-2018 2018, S.3-5.

²²¹MUNZNER, Tamara: *Process and pitfalls in writing information visualization research papers*. In *Information visualization* Springer, 2008, S.134-153).

SELT folgende Dimensionen zu berücksichtigen:²²²

- Zielgruppe: für wen ist die Visualisierung konzipiert?
- Aufgabe: welches Ziel wird mit der Darstellung verfolgt; was soll dargestellt werden?
- Daten: auf welchen Grundlage fußt die Darstellung?
- Repräsentation: welche Formen der Repräsentation werden für die Darstellung genutzt?
- Medium: wo wird es ausgegeben; wo kann man damit arbeiten?

Auf das DEPCHA Projekt gemünzt, wird mit einer Visualisierung das Ziel verfolgt, für ein fachwissenschaftliches Publikum eine größere Menge historischer Information von Transaktionen über das Web so aufzubereiten, dass aus der Aggregation der historischen Information neue Erkenntnisse gewonnen werden können. So kann die visuelle Repräsentation von historischer Information in den Rechnungsbüchern zur Bearbeitung von relevanten Forschungsfragen herangezogen werden. Die Zielsetzung der Visualisierungen liegt darin, die Dimensionen der Rechnungsbücher, das heißt die beteiligten Akteure, die Wirtschaftsgüter, die Geldbeträge jeweils in ihrem historischen Kontext, für Anwenderinnen, im Sinne einer quellenbasierten Arbeitsplattform erfahrbar zu machen. Diese Arbeitsplattform soll Fachwissenschaftlerinnen erlauben, sowohl den Einzelbeleg aus der Quelle, als auch die Aggregation der Information einer ausgewählten Menge von Belegen zu erkunden und selbstständig organisieren zu können. Diese zwei Zugangsarten kann man als *Close Viewing*, der "klassische" Zugang punktuelle Quellen manuell zu analysieren und *Distant Viewing*, die eine Form der formalen Methodik mit sich bringen muss, bezeichnen.²²³

Allgemein kann man sagen, dass das Ziel der Informationsvisualisierung der Gewinn neuer Erkenntnis ist. Wichtig ist das Entdecken neuer Zusammenhänge, die erst in einem größeren Kontext sichtbar werden (*Discovery*), das Treffen von zuverlässigen Entscheidungen (*Decision making*), die explorative Analyse von Informationsräumen (*Exploration*), sowie das Sichtbarmachen von Erklärungen (*Explanation*). Gerade für eine fehlerhafte, unscharfe oder inhomogene Datenbasis kann die menschliche Exploration von Daten sehr effektiv sein.²²⁴ Das Organisieren in DEPCHA umfasst die Selektion und das Filtern von Ergebnissen, die durch das *Discovery* der Visualisierungen gegeben sind. Die explorative Analyse, sowie eine Sichtbarmachung von Erklärungen durch Darstellungen kann vorher entworfene Thesen zu einem aus einem Rechnungsbuch abgeleiteten Sachverhalt stützen.

²²²PREIM/DACHSELT: Interaktive Systeme: Band 1: Grundlagen, Graphical User Interfaces, Informationsvisualisierung, S.437-440.

²²³SCHUBERT, Charlotte: Tagungsbericht: HT 2014: Close Reading and Distant Reading. Methoden der Altertumswissenschaften in der Gegenwart. In H-Soz-Kult Göttingen, 2014 (URL: www.hsozkult.de/conferencereport/id/tagungsberichte-5625).

²²⁴PREIM/DACHSELT: Interaktive Systeme: Band 1: Grundlagen, Graphical User Interfaces, Informationsvisualisierung, S.439-444.

Eine weitere Zielsetzung liegt darin, dass eine Visualisierung dabei helfen kann, die Datenqualität der digitalen Edition zu verbessern. Größere Mengen an ähnlich strukturierter Information bergen ein erhöhtes Risiko, dass sich während der manuellen Annotation Fehler einschleichen. Durch die Visualisierung der Daten ist es leichter Workflows aufzusetzen, in dem die Visualisierung des Quellenkorpus genutzt werden kann, um fehlerhafte Transkriptionen oder falsche Zuordnungen bestimmter Entitäten – wie etwa die Normalisierung von Personen, Gütern oder Maßeinheiten – besser entscheiden zu können.

Um die notwendige Dimension, nach denen eine mögliche Visualisierung umgesetzt werden kann, festmachen zu können, muss ein Modell definiert werden. Da im Projekt DEPCHA Rechnungsbüchern aus unterschiedlichen Kontexten zusammengeführt werden, bedarf es eines konzeptuellen Modells, um dieses Mapping herzustellen. Die Kategorien, die im Mapping abgebildet werden sollen, entsprechen auch den Dimensionen, die für eine Visualisierung der Informationsobjekte, ihrer Attribute und Relationen, relevant sind. Geldbeträge und Mengenangaben werden als quantitative Daten verstanden, da sie arithmetische Operationen erlauben, wohingegen Kategorien von Gütern und Währungen qualitative Daten darstellen, die für die Beschreibung, Ordnung und Gruppierung dienlich sind.²²⁵

5.3.2 Multidimensionale Rechnungsbücher – Multiple Views und Interaktion

GLEBA und PETERSEN heben die Vielfalt der Inhalte in Rechnungsbüchern, sowie die Möglichkeiten der *”interdisziplinären Erschließung der Quelle Rechnungsbuch”* hervor²²⁶ und auch VOGELER sieht in Rechnungsbüchern multidimensionale Quellen.²²⁷

Ein Credo der Informationsvisualisierung ist es, eine Domäne in mehreren Dimensionen darzustellen, vom Großen ins Detail zu gehen (*overview first*), dort zu filtern und in die Daten eintauchen zu können, um anschließend auf die Detailebene vorzudringen. Um die Vielfalt der Ebenen und Dimensionen von historischen Rechnungsbüchern auch nutzen zu können, bedarf es auch einer Vielzahl an Repräsentationen. Deswegen bedarf es multipler Perspektiven auf ein Thema. Dies wird in der Informationsvisualisierung als *Multiple Views* zusammengefasst.²²⁸

Ein gelungenes Projekt zur Datenvisualisierung wirtschaftlicher Daten, das *Multiple Views* anbietet, ist der *Atlas of Economic Complexity*.²²⁹ Es soll stellvertretend für eine moderne Webanwendung stehen, die ähnliche Informationsobjekte repräsentiert. Eine Anlehnung von Dar-

²²⁵PREIM/DACHSELT: Interaktive Systeme: Band 1: Grundlagen, Graphical User Interfaces, Informationsvisualisierung, S.448-450.

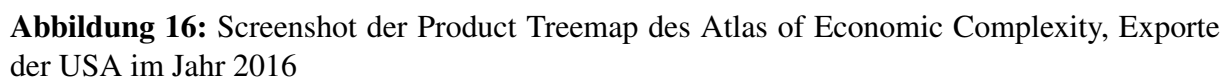
²²⁶GLEBA/PETERSEN: Wirtschafts-und Rechnungsbücher des Mittelalters und der Frühen Neuzeit-Formen und Methoden der Rechnungslegung: Städte, Klöster und Kaufleute, S.7-10.

²²⁷VOGELER et al.: The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies

²²⁸PREIM/DACHSELT: Interaktive Systeme: Band 1: Grundlagen, Graphical User Interfaces, Informationsvisualisierung, S.452-462.

²²⁹Atlas of Economic Complexity, atlas.cid.harvard.edu, 09.12.2019

What did United States of America export in 2016?



Eine weitere Form der Repräsentation sind **Bar Charts** (Säulendiagramme). Diese Repräsentationsform ist ein klassischer Weg, um Häufigkeitsverteilungen darzustellen, jedoch ist sie auf eine bestimmte Anzahl an Ausprägungen (Säulen) begrenzt. Jede dieser Säulen auf der X-Achse

kann beispielsweise ein Geschäftsjahr mit den Einnahmen bzw. Ausgaben in einem Rechnungsbuch darstellen. Da der Zeitraum eines einzelnen Buches sich mit bis zu 15 Säulen abdecken lässt, jede Säule entspricht einem Jahr, kann dies eine geeignete Darstellung sein. Ein **Stacked Chart** erlaubt es, eine weitere Dimension in ein Bar Chart einzufügen, um Einkünfte und Ausgaben weiter kategorisieren zu können. Beispielsweise könnten diese Dimensionen die einzelnen Akteure sein, die an den Geldflüssen beteiligt sind. Rechnungsbücher können aus einer Vielzahl von Perspektiven betrachtet werden: welche Güter, Akteure und Geldbeträge kommen vor und welche sozialen und gesellschaftlichen Implikationen könnten daraus abgeleitet werden.

Prototypisch wurde auf der in RDF modellierten Datengrundlage mittels SPARQL eine Datenbankabfrage formuliert und das so gewonnene JSON Ergebnis als Input an die Datenvisualisierungssoftware *Tableau Desktop*²³⁰ übergeben. Bei den folgenden zwei Beispielen handelt es sich um Visualisierungen, wie sie sinnvollerweise für Rechnungsbücher umgesetzt werden könnten.

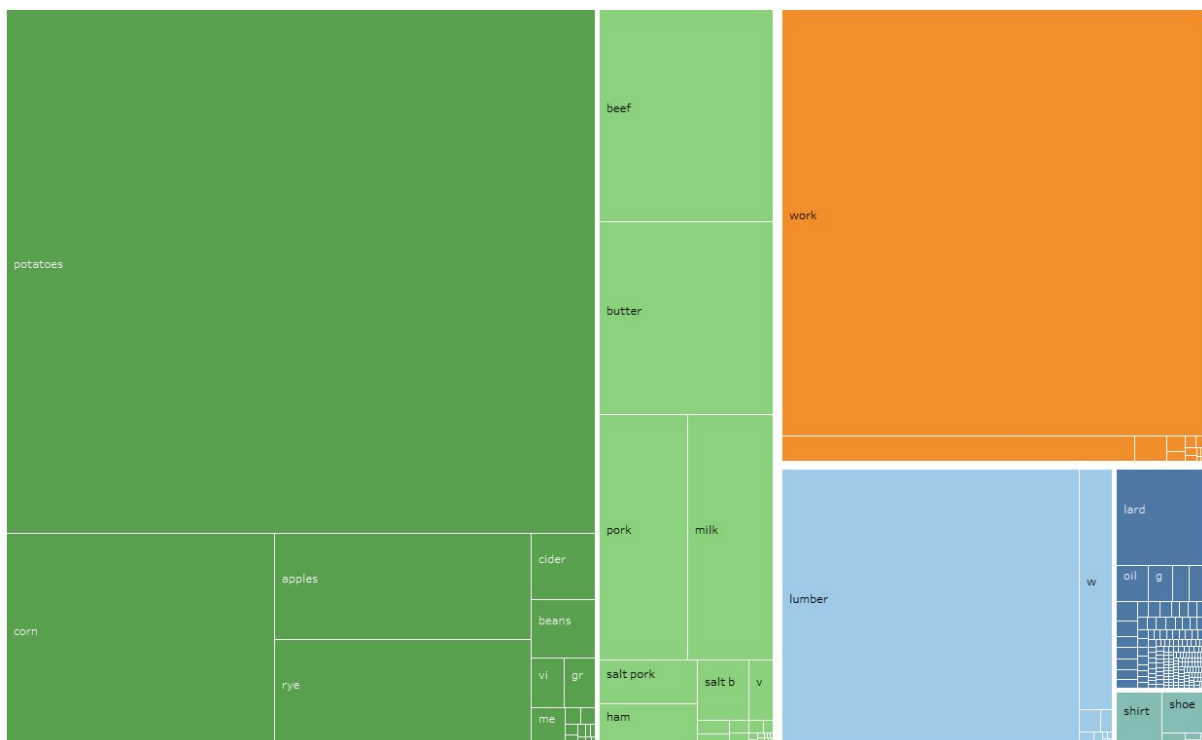


Abbildung 17: Treemap nach Häufigkeit von Waren und Dienstleistungen. Gruppiert nach tierischen (hellgrün) und pflanzliche (dunkelgrün) Lebensmittel, Dienstleistungen (orange), Rohstoffen (hellblau), Textilien (türkis) und sonstigen Waren (dunkelblau)

In dieser Treemap Ansicht der wirtschaftlichen Güter, entspricht die Größe der Rechtecke der Häufigkeit wie oft ein Gut in allen Transaktionen des *Wheaton Daybooks* vorkommt. Eine besondere Problemstellung dabei ist die Repräsentation bzw. Umrechnung von unterschiedlichen

²³⁰Tableau Desktop, tableau.com/de-de/products/desktop, 09.12.2019.

Maßeinheiten auf normalisierte Werte, die dann gegenübergestellt werden können. So stünden 100 *bushel potatoes*, 20 *feet lumber* gegenüber. Deswegen wurde in dieser Darstellung nur die absolute Anzahl, wie oft etwas genannt wird, dargestellt. In dieser Ansicht werden 3 Hauptgruppen unterschieden: Lebensmittel (grün), Dienstleistungen (orange) und Waren (blau). Diese wurden, um eine weitere Hierarchieebene einzufügen, weiter unterteilt in pflanzliche (dunkelgrün) und tierische (hellgrün) Lebensmittel, Rohstoffe (hellblau), Textilien (dunkelblau) und andere Waren (türkis). Im DEPCHA Projekt sind die Hierarchien, nach denen die Rechtecke angeordnet werden, durch Thesauri, die von den Fachwissenschaftlerinnen definiert werden, vorgegeben. Diese Hierarchien müssen demzufolge sinnvoll auf die Quellen und die Forschungsfragen abgestimmt werden.

In Abbildung 17 zeigt sich sofort der Schwerpunkt der Geschäfte von Wheaton: pflanzliche Produkte, vor allem Kartoffel, Mais, Äpfel und Roggen, sowie Fleisch und Milchprodukte, Arbeit als Dienstleistung und verarbeitetes Holz (*lumber*) als Rohstoff. Daneben gibt es eine Vielzahl von Gütern, die selten gehandelt werden, wie etwa Bücher. Diese befinden sich im stark geschachtelten dunkelblauen Bereich.

Neben den Warenflüssen, sind die Geldflüsse ein zentraler Bestandteil der Daten. Folgende Darstellung zeigt die Geldflüsse an L.M. Wheaton für jedes Jahr. Jede Farbe auf den einzelnen Säulen repräsentiert einen Akteur. Personen wie etwa Wheaton Wheeler (hellroter Balken), ein Verwandter von L.M. Wheaton, kommt sehr regelmäßig in den Laden, kauft aber nur Lebensmittel und Alltagsgegenstände zu niedrigeren Geldbeträgen, gleichzeitig verrichtet er auch öfters Arbeit für L.M. Wheaton.

Es lässt sich daraus schließen, und solche Interpretationen sind die Zielsetzung dieser Darstellungen, dass Wheaton Wheeler aus einem niederen sozialen Status stammt und in einer gewissen Abhängigkeit zu L.M. Wheaton steht. Weiters erhält ein User so einen Überblick, in welchen Jahren niedrigere oder größere Einnahmen zustande kamen, und über die einzelnen Akteure und ihrer Frequenz.

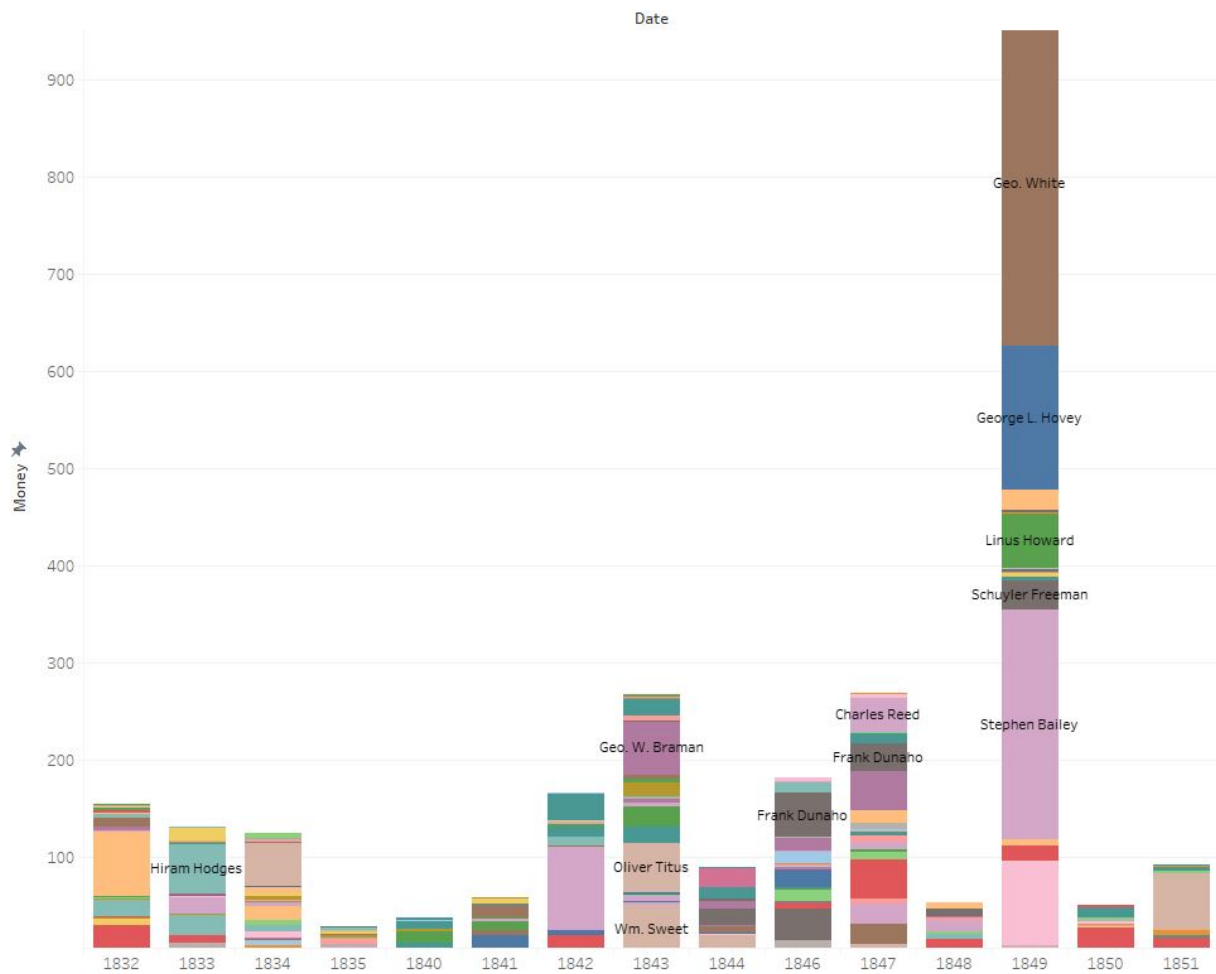


Abbildung 18: Geldfluss nach Jahren (x-Achse) und Summe der Geldflüsse (y-Achse). Stacks nach einzelnen Akteuren.

5.3.3 Informationsvisualisierung und Web of Data

FRANK²³¹ führt 4 Beispiele von Projekten an, die die Methoden der Informationsvisualisierung so nutzen, dass sie zentraler Bestandteil zur Bearbeitung einer geisteswissenschaftlichen Forschungsfrage und zur Überprüfung der Ergebnisse herangezogen werden können. Ein Beispiel daraus ist das Projekt CEWS. Dieses Projekt behandelt multiperspektivische Konfliktforschung mit Fokus auf politikwissenschaftlichen Forschungsfragen. So wird im Projekt Modallogik als formaler Rahmen für die kontrafaktische Analyse von möglichen Konfliktverläufen eingesetzt und die generative Phänomenologie wird zur Analyse der Perspektiven von historischen Akteuren, den Konfliktparteien, auf Konfliktereignisse vorgeschlagen.

Gemeinsam haben alle diese Projekte, dass der Ausgangspunkt für die Visualisierung zum einen Daten sind, zum anderen ein Modell, dass den Daten Struktur verleiht und so die formale Verarbeitung in Hinblick auf eine Forschungsfrage erlaubt. Ein solches Modell kann unter Verwendung von *Web of Data* Technologien und unter Verwendung von *Linked Open Data* beschrieben und formalisiert werden. Beispielsweise werden *Knowledge Graphs* verwendet um explorative Suchen größerer, kontextsensitiver Daten umzusetzen.²³²

Es veranschaulicht, welche Ziele und Herausforderungen mit Methoden der Informationsvisualisierung einhergehen. Dabei handelt es sich um ein starkes Werkzeug, um bestimmte Inhalte in Sammlungen von historischen Quellen, beispielsweise Rechnungsbüchern, hervorzuheben. Es muss aber auf vielen Ebenen gut durchdacht sein, damit die vergangene "Wirklichkeit" nicht verzerrt wird. Von der Forschungsfrage (was soll gezeigt werden), der Darstellung (wie wird es dargestellt), den Daten (woraus wird es erzeugt), sowie der Interaktion damit (was kann ich damit tun) bedarf eines durchdachten Plans. Darstellungen dürfen nicht zu speziell sein und etwas repräsentieren, das nur für eine einzige Forschungsfrage bzw. einzelnes Rechnungsbuch von Interesse ist, dürfen aber auch nicht zu allgemein sein, sodass daraus keine generischen Lösungen ableitbar sind.

²³¹FRANK: INF-DH-2018 2018.

²³²SARRAFZADEH, Bahareh/VECHTOMOVA, Olga/JOKIC, Vlado: Exploring knowledge graphs for exploratory search. In Proceedings of the 5th Information Interaction in Context Symposium. 2014.

6 Zusammenfassung und Ausblick

Die Gegebenheiten der Wissensproduktion in der Geschichtsforschung haben sich auf Grund der Digitalisierung gewandelt und beeinflussen wie Geschichte gedacht, erarbeitet und interpretiert wird. Dies geschieht nicht erst bei der ausgefeilten maschinellen Analyse von Daten, sondern beginnt bereits beim Sammeln und Entdecken von Primär- und Sekundärquellen über das Web. Computergestützte Methoden zur Analyse, sowie die Zurverfügungstellung der digitalisierten Quellen, erlauben neue Perspektiven im Umgang mit historischen Quellen. Aber Digitalisierung ist vielmehr ein mächtiges Werkzeug, das die Arbeit mit größeren Datenmengen effizienter gestaltet, nicht aber ein Paradigmenwechsel in den Geschichtswissenschaften.²³³ Dies ist aber noch eine offene Frage, die in der Fachgemeinschaft diskutiert wird.

In seinem *”Grundriß der Historik”*, gibt DROYSEN einen Überblick über die Methoden der Geschichtswissenschaft. In der Arbeit mit formalen, digitalen Methoden und Modellen, sowie Informationssystemen für die Geschichtswissenschaft, erscheinen immer noch die Überlegungen von DROYSEN sehr relevant. Das historische Denken und Forschen, eingeteilt in Heuristik, Kritik, Interpretation und Topik (Formen der Darlegung der Ergebnisse), sind nach wie vor essentieller Bestandteil der Geschichtswissenschaften.²³⁴

Auch FICKERS führt diese vier grundlegenden Arbeitsschritte an, sieht aber, dass der Einfluss der Digitalisierung auf die Entwicklung neuartiger Problemstellungen noch offen ist. Weiter führt FICKERS an: *”Die suggerierte Objektivität, die mit diesen Darstellungen von Informationen [z.B. Visualisierungen] einhergeht, gilt es zu dekonstruieren. Zur digitalen Hermeneutik gehört deshalb auch eine digitale Werkzeugkritik.”*²³⁵ Oder eben eine geschichtswissenschaftliche Datenkritik, wie er es nennt.²³⁶

Betrachtet man beispielsweise die Beschreibung der Heuristik bei DROYSEN, so lässt sich vieles davon auf die heutige Arbeit ummünzen. *”Die Kunst der Heuristik ist, das historische Material zu erweitern und zu ergänzen, und zwar:”*²³⁷

- (a) *”durch divinatorisches Suchen und Entdecken;”*
- (b) *”durch Kombination, [...], durch richtige Einreihung [...]”*
- (c) *”durch Analogie [...]”*

²³³KÖNIG, Mareike: Geschichtsforschung und Archive im digitalen Zeitalter. Chancen, Risiken und Nebenwirkungen. Archivar, 73. Jahrgang 2020 Nr. 3.

²³⁴HARDTWIG: Über das Studium der Geschichte, S.85-116.

²³⁵Digitale Hermeneutik - Ein Update für das kritische Denken in den Geisteswissenschaften, <https://www.science.lu/de/forschungstrends/digitale-hermeneutik-ein-update-fuer-das-kritische-denken-geisteswissenschaften>, 27.08.2020

²³⁶FICKERS: Zeithistorische Forschungen, Nr. 1, Bd. 17, 2020, S.2.

²³⁷HARDTWIG: Über das Studium der Geschichte, S.96.

(d) ”durch Hypothese, deren Beweis die Evidenz des Ergebnisses ist [...]”

Die Arbeit im Projektkontext DEPCHA und an der *Bookkeeping-Ontology* kann genau ein Mittel sein, um beispielsweise die Werkzeugkritik oder die Anforderungen der Heuristik, zumindest in Teilen, zu erleichtern bzw. zu ermöglichen. Die *Bookkeeping-Ontology* erlaubt es Transaktionen, also den Fluss von wirtschaftlichen Gütern und Geldbeträgen von einem Akteur zum anderen, zu beschreiben.

Dabei bildet der iterative Prozess der Entwicklung eines Modelles im *Ontology Engineering* einen hermeneutischen Prozess, in dem sich beteiligte Fachexpertinnen mit einer geschichtswissenschaftlichen Forschungsfrage beschäftigen, Quellen sammeln, bearbeiten, Ergebnisse verwerfen und verfeinern. In diesem Prozess spiegelt sich die *Analogie* wider. Transaktionsprozesse sind ein alltägliches Phänomen in unserer heutigen Lebenswelt. Es fällt uns leicht diese ”natürliche” Struktur aus heutiger Zeit in die Rekonstruktion der Vergangenheit einzupflegen. Am Ende dieses Prozesses steht ein maschinenlesbares Modell, das als Abbild eines Forschungsprozesses verstanden werden kann. Die eigentlichen Entitäten, sprich Daten bzw. die kontextualisierte Information aus den Quellen sind aus diesem Modell abgeleitete Instanzen. Wodurch die Anforderung an die *Kombination* nach DROYSEN erfüllt wird.

Eine Möglichkeit der Implementierung ist durch den Einsatz von XML/TEI und RDF gegeben. Beide Standards, basierend auf elementaren Datenstrukturen (Baum und Graph), haben selbst beschreibenden Charakter. Dadurch sind sie für die Langzeitarchivierung, im Sinne von Forschungsdaten, in einem digitalen Repositorium geeignet. Wo TEI die Annotation der textuellen Quelle ermöglicht und so die Grundlage für die digitale Edition stellt, ermöglicht es RDF semantische Strukturen formal zu beschreiben. Letzteres ist wichtig, da die Publikation und Veröffentlichung, sowie die Vernetzung der Information im Sinne von Linked Data, die Nutzbarkeit und den wissenschaftlichen Konsens (oder Dissens) abbilden kann. Gerade die Publikation geht dann mit einem Mehrwert einher: Formen der Informationsvisualisierung lassen sich aus diesen Strukturen umsetzen und damit Formen des *Suchens und Entdeckens*.

Die *Hypothese* trifft den empirischen Teil der Geschichtswissenschaften. Diese Daten entsprechen nicht einer absoluten Wahrheit, sondern sind das Produkt eines hermeneutischen Prozesses einer Fachwissenschaftlerin, auf deren Basis weitere Arbeit und Interpretation ruhen kann. Die so entstandenen Daten können aber unter Verwendung von quantitativen Methoden analysiert werden, um Hypothesen zu stützen oder zu falsifizieren. Dadurch, dass die Daten im Web und ihre *knowledge domain* im Web zur Verfügung steht, können Hypothesen an geschichtswissenschaftliche Fragestellungen effektiver bearbeitet werden.

Zusammengefasst kann gesagt werden, dass folgende Vorteile mit der Verwendung von Ontologien einhergehen:

- Ein gemeinsames Vokabular bzw. Modell erleichtert die Kommunikation in der Wissensproduktion und kann von Fachkolleginnen und Kollegen nachgenutzt werden.
- Die Wiederverwendung eines Modells befördert den fachwissenschaftlichen Diskurs, in dem bestehende Konzepte kritisiert und fehlende Konzepte ergänzt werden können. Dies führt zu einem allgemein akzeptierten und verständlicheren konzeptionellen Modell.
- *Linked Open Data*, wie etwa auf einem Hub wie Wikidata, ermöglicht Interoperabilität der Daten in verschiedenen wissenschaftlichen Disziplinen.

Literatur

- ALEXANDER, Bryan/DAVIS, Rebecca Frost:** Should liberal arts campuses do digital humanities? Process and products in the small college world. *Debates in the digital humanities*, 2012, S.368–389
- ANDORFER, Peter:** Forschungsdaten in den (digitalen) Geisteswissenschaften: Versuch einer Konkretisierung. DARIAH-DE, Working Papers 14 2015 (URL: webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-14.pdf)
- AUMANN, Stefan et al.:** From digital archive to digital edition. *Historische Sozialforschung*, 1999, 101–144
- BAUER, Florian/KALTENBÖCK, Martin:** Linked open data: The essentials. A Quick Start Guide for Decision Makers. Wien: Edition mono/monochrom, 2011
- BERNERS-LEE, Tim/FIELDING, Roy/MASINTER, Larry:** Uniform resource identifier (URI): Generic syntax. 2004 – Technischer Bericht (URL: www.rfc-editor.org/info/rfc3986)
- BERNERS-LEE, Tim/HENDLER, James/LASSILA, Ora:** The semantic web. *Scientific american*, 284 2001 Nr. 5, S.34–43
- BERNSTEIN, Abraham/HENDLER, James/NOY, Natalya:** A New Look at the Semantic Web. *Commun. ACM*, 59 2016 Nr. 9, S.35–37
- BOER, Víctor de/MERONO PENUELA, A/OCKELOEN, CJ:** Linked data for digital history: lessons learned from three case studies. *Anejos de la Revista de Historiografía*, 2016 Nr. 4, S.139–162
- BORST, Willem Nico:** Construction of Engineering Ontologies for Knowledge Sharing and Reuse. Dissertation, Netherlands, 1997
- BRICKLEY, Dan/GUHA, Ramanathan V/MCBRIDE, Brian:** RDF Schema 1.1. W3C recommendation, 25 2014, S.2004–2014
- BRUCH, Julia:** Die Kunst, Daten in Informationen umzuwandeln. Zur Auswertung eines zisterziensischen Rechnungsbuchs aus dem 13. und 14. Jahrhundert und den Herausforderungen in der Analyse serieller Wirtschaftsquellen. *Wirtschafts- und Rechnungsbücher des Mittelalters und der Frühen Neuzeit*, 2015, 13–44
- BRUMFIELD, Ben:** The Stagville Accounts in DEPCHA: Plantation Financial Records as Linked Data. From the Page, 2019 (URL: content.fromthepage.com/stagville-accounts-in-depcha/)

- BRUMFIELD, Ben/AGBE-DAVIES, Anna:** Encoding Account Books Relating to Slavery in the U.S. Modeling semantically Enriched Digital Edition of Accounts, 2015 (URL: medea.hypotheses.org/182)
- BURDICK, Anne et al.:** Digital_Humanities. Mit Press, 2012
- BURNARD, Lou:** What is the Text Encoding Initiative?: How to add intelligent markup to digital resources. OpenEdition Press, 2014, 3
- CHALMERS, Alan F:** Wege der Wissenschaft: Einführung in die Wissenschaftstheorie. Springer, 2007
- CHEN, Peter Pin-Shan:** The entity-relationship model—toward a unified view of data. ACM Transactions on Database Systems (TODS), 1 1976 Nr. 1, 9–36
- CLARK, George Faber:** A History of the Town of Norton, Bristol County, Massachusetts, from 1669-1859. Crosby, Nichols, and Company, and author at Norton, 1859
- CLIFFORD, Anderson et al.:** Modeling semantically Enhanced Digital Edition of Accounts as Historical Method. 2016
- COLE, George S:** A complete Dictionary of Dry Goods and History of Silk, Linen, Wool and other Fibrous Substances. 1892. 2015
- CUMMINGS, James:** The text encoding initiative and the study of literature. A Companion to Digital Literary Studies 2013
- DEMANDT, Alexander:** Philosophie der Geschichte von der Antike zur Gegenwart. 2011
- DENTLER, Kathrin et al.:** Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web, 2 2011 Nr. 2, S.71–87
- DENZEL, Markus A/HOCQUET, Jean Claude/WITTHÖFT, Harald:** Kaufmannsbücher und Handelspraktiken vom Spätmittelalter bis zum beginnenden 20. Jahrhundert. Franz Steiner Verlag, 2002, 163
- DETEL, Wolfgang:** Geist und Verstehen. Historische Grundlagen einer modernen Hermeneutik. Philosophische Abhandlungen Band 104, 2011
- DUCHARME, Bob:** Learning SPARQL: querying and updating with SPARQL 1.1. O'Reilly, 2013
- EVANS, Richard J:** Fakten und Fiktionen: über die Grundlagen historischer Erkenntnis. Campus-Verlag, 1998

- FAVRE-BULLE, Bernard:** Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001
- FICARA, Elena:** Texte zur Hermeneutik. Von Plato bis heute. Reclam, 2015
- FICKERS, Andreas:** Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik? Zeithistorische Forschungen, 17 2020 Nr. 1, S.157–168 (URL: <http://hdl.handle.net/10993/44086>)
- FÖHR, Pascal:** Historische Quellenkritik im Digitalen Zeitalter. Dissertation, Universität Basel, 2017
- FRANK, Ingo:** Visualisierungswerkzeuge zur Erklärung historischer Ereignisse: geschichtstheoretische Anforderungsanalyse und zeichentheoretisches Rahmenwerk. INF-DH-2018 2018
- FROMMER, Sören:** Grundbegriffe aus Theorie, Methode und Wissenschaftsgeschichte. (URL: www.historische-archaeologie.de/downloads/grundbegriffe.pdf)
- GLEBA, Gudrun:** Rechnen. Wirtschaften. Aufschreiben. Vernetzte Schriftlichkeit – Wirtschafts- und Rechnungsbücher als Quellen klösterlicher Alltagsgeschichte. In **Pätzold, Stefan/Stumpf, Marcus (Hrsg.):** Mittelalterliche und frühneuzeitliche Rechnungen als Quellen der landesgeschichtlichen Forschung. Münster, 2016, 51–65
- GLEBA, Gudrun/PETERSEN, Niels:** Wirtschafts-und Rechnungsbücher des Mittelalters und der Frühen Neuzeit-Formen und Methoden der Rechnungslegung: Städte, Klöster und Kaufleute. Göttingen, 2015
- GLEBA, Gudrun/PETERSEN, Niels:** Zur Einleitung. In **Gleba, Gudrun/Petersen, Niels (Hrsg.):** Wirtschafts-und Rechnungsbücher des Mittelalters und der Frühen Neuzeit-Formen und Methoden der Rechnungslegung: Städte, Klöster und Kaufleute. Universitätsverlag Göttingen, 2015, 7–11
- GOTTFRIED, Gabriel:** Fakten oder Fiktion? Zum Erkenntniswert der Geschichte. Oldenbourg
- GRAHAM, Shawn/MILLIGAN, Ian/WEINGART, Scott:** Exploring big historical data: The historian's macroscope. London, 2015
- GROUP, W3C SPARQL working et al.:** SPARQL 1.1 Overview. 2013
- GRUBER, Thomas:** A translation approach to portable ontology specifications. Knowledge acquisition, 5 1993 Nr. 2, S.199–220
- HARDTWIG, Wolfgang:** Geschichtskultur und Wissenschaft. München, 1990

- HARDTWIG, Wolfgang:** Über das Studium der Geschichte. Deutscher Taschenbuch Verlag, 1990
- HILTMANN, Torsten:** Forschungsdaten in der (digitalen) Geschichtswissenschaft. Warum sie wichtig sind und wir gemeinsame Standards brauchen. 2018 [〈URL: digigw.hypotheses.org/2622〉](http://digigw.hypotheses.org/2622)
- HITZLER, Pascal et al.:** Semantic Web: Grundlagen. Springer, 2007
- HOEKSTRA, Rinke:** Ontology Representation: design patterns and ontologies that make sense. Band 197, Amsterdam: IOS Press, 2009
- HOFFMANN, Barbara/HORN, Klaus:** Ehepaare, Eheverläufe und Lebenslauf in Leipzig 1580-1730 KLEIO-Anwendung in einem historischen Forschungsprojekt. Historische Sozialforschung, 15 1990 Nr. 3 (55, 171–198
- HORROCKS, Ian et al.:** Semantic web architecture: Stack or two towers? In International Workshop on Principles and Practice of Semantic Web Reasoning. Springer 2005, 37–41
- JANNIDIS, Fotis:** Grundlagen der Datenmodellierung. In Digital Humanities Springer, 2017, 99–108
- JANNIDIS, Fotis/KOHLE, Hubertus/REHBEIN, Malte:** Digital Humanities: Eine Einführung. Springer, 2017
- JARAUSCH, Konrad Hugo/ARMINGER, Gerhard/THALLER, Manfred:** Quantitative Methoden in der Geschichtswissenschaft. Wissenschaftliche Buchgesellschaft, 1985
- JOOST BREUKER, Pompeu CASANOVAS/KLEIN, MC/FRANCESCONI, Enrico:** The flood, the channels and the dykes: Managing legal information in a globalized and digital world. Law, Ontologies and the Semantic Web: Channelling the Legal Information Flood, 188 2009, S.3
- JORDAN, Stefan:** Vetorecht der Quellen. 2010 [〈URL: http://dx.doi.org/10.14765/zzf.dok.2.570.v1〉](http://dx.doi.org/10.14765/zzf.dok.2.570.v1)
- JORDAN, Stefan:** Theorien und Methoden der Geschichtswissenschaft. 4. Auflage. UTB GmbH, 2018
- KINDLING, Maxi et al.:** Forschungsdatenmanagement an Hochschulen: Das Beispiel der Humboldt-Universität zu Berlin. LIBREAS. Library Ideas, 2013 Nr. 23 [〈URL: libreas.eu/ausgabe23/07kindling〉](http://libreas.eu/ausgabe23/07kindling)
- KIRN, Paul:** Einführung in die Geschichtswissenschaft. Berlin, Boston, 2015

- KLAPP, Sabine:** Die “Äbtissinnenrechnungen” des Klosters St. Klara auf dem Werth. Alltag und Festtag einer geistlichen Frauengemeinschaft Straßburgs am Ausgang des Mittelalters. Zeitschrift für die Geschichte des Oberrheins Bd. 159, 2011, 211–248
- KÖNIG, Mareike:** Geschichtsforschung und Archive im digitalen Zeitalter. Chancen, Risiken und Nebenwirkungen. Archivar, 73. Jahrgang 2020 Nr. 3, 245–251
- KOBLER, Maximilian:** Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung. Logos Verlag Berlin GmbH, 2010
- KOCKA, Jürgen:** Theorien in der Geschichtswissenschaft. In Theoriedebatte und Geschichtsunterricht: Sozialgeschichte, Paradigmawechsel und Geschichtsdidaktik in der aktuellen Diskussion. Vorträge auf der Jahrestagung des Landesverbandes nordrhein-westfälischer Geschichtslehrer am 23. November 1981 in Essen Paderborn: Schöningh, 1982, S.7–27
- KROPAČ, Ingo:** Theorien, Methoden und Strategien für multimediale Archive und Editionen. Mediaevistik und Neue Medien 304 2004
- LINCKELS, Serge/MEINEL, Christoph:** E-librarian service: user-friendly semantic search in digital libraries. Springer, 2011
- MAGERSKI, Christine et al.:** Wie schreibt man Geschichte? Zagreber Germanistische Beiträge, 2009 Nr. 18, S.3–19
- MANN, Golo:** Plädoyer für die historische Erzählung. In **Kocka, Jürgen/Nipperdey, Thomas (Hrsg.):** Theorie und Erzählung in der Geschichte. München, 1979, S.40–56
- MCCARTHY, William E:** The REA accounting model: A generalized framework for accounting systems in a shared data environment. Accounting review, 1982, S.554–578
- MCCARTY, Willard:** Modeling: A Study in Words and Meanings. In **Susan Schreibman, Ray Siemens, John Unsworth (Hrsg.):** A Companion to Digital Humanities. John Wiley and Sons, Ltd, 2007. – Kapitel 19, 254–270
- MEIXNER, Uwe:** Von der Wissenschaft der Ontologie. Logos (neue Folge), 1 1994, S.375–399
- MEYER, Till:** Datenmodellierung für digitale Editionen. Diplomarbeit, Humboldt-Universität zu Berlin, 2019
- MUNZNER, Tamara:** Process and pitfalls in writing information visualization research papers. In Information visualization Springer, 2008, S.134–153
- MUSEN, Mark A:** The Protégé Project: A Look Back and a Look Forward. AI matters, 1 2015 Nr. 4, S.4–12

- NEHER, Günther/RITSCHER, Bernd:** Semantische Vernetzung von Forschungsdaten. In **Büttner, Stephan/Hobohm, Hans-Christoph/Müller, Lars (Hrsg.):** Handbuch Forschungsdatenmanagement. Band 6, 2011, S.169–190
- NEUWÖHNER, Andreas:** Haushalt und Finanzen im Spiegel kommunaler Rechnungen: die Stadt Paderborn im 17. Jahrhundert. In **Pätzold, Stefan/Stumpf, Marcus (Hrsg.):** Mittelalterliche und frühneuzeitliche Rechnungen als Quellen der landesgeschichtlichen Forschung. Münster, 2016, 11–50
- NÜSSEL, Friederike et al.:** Offenbarung als Geschichte: Implikationen und Konsequenzen eines theologischen Programms. Vandenhoeck & Ruprecht, 2018
- OPGENOORTH, Ernst/SCHULZ, Günther:** Einführung in das Studium der Neueren Geschichte. Paderborn / München / Wien / Zürich, 2010
- PELLEGRINI, Tassilo/BLUMAUER, Andreas:** Semantic Web. Wege zur vernetzten Wissensgesellschaft 2006
- PERSTLING, Matthias:** Multimediale Dokumentation und Edition mehrschichtiger Texte : das steirisch-landesfürstliche Marchfütterurbar von 1414/1426. Dissertation, Universität Graz, 2013
- PIERAZZO, Elena:** Digital scholarly editing: Theories, models and methods. Routledge, 2016
- PIOTROWSKI, Michael:** Digital Humanities. Computational Linguistics, and Natural Language Processing. Lectures on Language Technology and History 2016
- POLLIN, Christopher:** Vom Suchen, Stöbern und Finden : Information Retrieval am Beispiel der Digitalen Sammlung des Hans Gross Kriminalmuseums. Diplomarbeit, Universität Graz, 2017
- POLLIN, Christopher:** Digital Edition Publishing Cooperative for Historical Accounts and the Bookkeeping Ontology. In Proceedings of the Doctoral Symposium on Research on Online Databases in History. 2019, S.7–14
- POLLIN, Christopher/VOGELER, Georg:** Semantically Enriched Historical Data. Drawing on the Example of the Digital Edition of the "Urfehdebucher der Stadt Basel". In WHiSe, ISWC. 2017, S.27–32
- POWERS, Shelley:** Practical RDF: solving problems with the resource description framework. O'Reilly Media, Inc., 2003
- PREIM, Bernhard/DACHSELT, Raimund:** Interaktive Systeme: Band 1: Grundlagen, Graphical User Interfaces, Informationsvisualisierung. 2. Auflage. Berlin / Heidelberg, 2010

- REHBEIN, M.:** Reconstructing the textual evolution of a medieval manuscript. *Literary and Linguistic Computing*, 24 08 2009, S.319–327
- REICHE, Ruth/BECKER, Rainer/BENDER, Michael:** Verfahren der Digital Humanities in den Geistes-und Kulturwissenschaften. Niedersächsische Staats-und Universitätsbibliothek Göttingen, 2014
- REICHENBERGER, Klaus:** Kompendium semantische Netze. Konzepte, Technologie, Modellierung. Springer, 2010
- ROBSON, Keith:** Accounting numbers as “inscription”: Action at a distance and the development of accounting. *Accounting, Organizations and Society*, 17 1992 Nr. 7, 685–708
- RUMBAUGH, James/JACOBSON, Ivar/BOOCH, Grady:** Unified modeling language reference manual, the. Pearson Higher Education, 2004
- SAHLE, Patrick:** Vom editorischen Fachwissen zur digitalen Edition: Der Editionsprozeß zwischen Quellenbeschreibung und Benutzeroberfläche. In **Thaller, Manfred (Hrsg.):** FUNDUS - Forum für Geschichte und ihre Quellen. Universität Köln, 2003 (URL: webdoc.sub.gwdg.de/edoc/p/fundus/2/sahle.pdf), 76–102
- SAHLE, Patrick:** Digitale Editionsformen: Textbegriffe und Recodierung. Band 3, Norderstedt, 2013
- SAHLE, Patrick:** Digitale Edition. Digital Humanities. Eine Einführung, 2017, S.234–249
- SARRAFZADEH, Bahareh/VECHTOMOVA, Olga/JOKIC, Vlado:** Exploring knowledge graphs for exploratory search. In *Proceedings of the 5th Information Interaction in Context Symposium*. 2014, S.135–144
- SCHMIEDER, Felicitas/SOKOLL, Thomas:** Theorien und Methoden der Geschichtswissenschaft.
- SCHOLGER, Martina:** Assoziationsprozessen auf der Spur: digitale Edition der Notizbücher von Hartmut Skerbisch. Dissertation, Universität Graz, 2018
- SCHREIBER, Guus/RAIMOND, Yves:** RDF 1.1 Primer. W3C working group note 2014
- SCHRÖDER, Wilhelm Heinz:** Historische Sozialforschung: Forschungsstrategie Infrastruktur Auswahlbibliographie. *Historical Social Research* 1988 Nr. 1
- SCHUBERT, Charlotte:** Tagungsbericht: HT 2014: Close Reading and Distant Reading. Methoden der Altertumswissenschaften in der Gegenwart. In *H-Soz-Kult Göttingen*, 2014 (URL: www.hsozkult.de/conferencereport/id/tagungsberichte-5625)
- SHIRKY, Clay:** Ontology is overrated. Categories, Links and Tags 2005

- STACHOWIAK, Herbert:** Allgemeine Modelltheorie. 1973
- STERTZER, Jennifer:** Working with the Financial Records of George Washington: Document vs. Data. Digital Studies/le Champ Numérique, 2014 [⟨URL: doi.org/10.16995/dscn.57⟩](https://doi.org/10.16995/dscn.57)
- STUCKENSCHMIDT, Heiner:** Ontologien: Konzepte, Technologien und Anwendungen. Springer, 2009
- SURE, York/STUDER, Rudi:** Methodology, tools & case studies for ontology based knowledge management. 2003
- SWARTZ, Aaron:** A Programmable Web: An Unfinished Work. Synthesis lectures on the semantic web: Theory and Technology, 3 2013 Nr. 2, S.1–64
- THALLER, Manfred:** Historical Information Science: Is there such a Thing? New Comments on an old Idea [1993]. Historical Social Research. Supplement 29, 2017, 260–286
- THALLER, Manfred:** The Need for Standards: Data Modelling and Exchange [1991]. Historische Sozialforschung. Supplement, 2017, S.203–220
- THALLER, Manfred:** Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte 'unscharfer' Systeme [1984]. Historische Sozialforschung. Supplement 29, 2017, S.138–159
- TOMASEK, Kathryn/BAUMAN, Syd:** Encoding financial records for historical research. Journal of the Text Encoding Initiative 2013 Nr. 6
- UTE, Daniel:** Kompendium Kulturgeschichte. Theorien, Praxis, Schlüsselwörter. 5. Auflage. Frankfurt am Main
- VAJDA, Jonathan/MERRELL, Eric/SMITH, Barry:** Toward an ontology of commercial exchange. 2019
- VASOLD, Gunter:** Das Itinerar Erzbischof Konrads IV. von Salzburg: 1291-1312; computergestützte Itinerarerstellung und Itineraranalyse. Selbstverl. d. Inst. für Geschichte d. Karl-Franzens-Univ. Graz, 1996
- VOGELER, Georg:** Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert? In **Baum, Constanze/Stäcker, Thomas (Hrsg.):** Grenzen und Möglichkeiten der Digital Humanities. 2015 [⟨URL: www.zfdg.de/sb001_007⟩](http://www.zfdg.de/sb001_007)
- VOGELER, Georg:** Religion aus Daten? : Zur digitalen Edition von religiösen Grundlagentexten. 2018, S.178–198

- VOGELER, Georg:** The 'assertive edition'. On the consequences of digital methods in scholarly editing for historians. *International Journal of Digital Humanities*, 1 2019, S.309–322
- VOGELER, Georg et al.:** The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies. In **Sarnowsky, Jürgen (Hrsg.):** Konzeptionelle Überlegungen zur Edition von Rechnungen und Amtsbüchern des späten Mittelalters. Göttingen, 2016, S.13–41
- VON BRANDT, Ahasver:** Werkzeug des Historikers: eine Einführung in die historischen Hilfswissenschaften. Band 33, W. Kohlhammer Verlag, 2007
- WELLER, Katrin:** Ontologien. In **Kuhlen, Rainer/Semar, Wolfgang/Strauch, Dietmar (Hrsg.):** Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und praxis. 6. Auflage. Berlin and Bosten, 2013, S.S.207–218
- WELLMER, Albrecht:** Georg Henrik von Wright über "Erklären" und "Verstehen". 1979
 〈URL: <https://www.jstor.org/stable/42571419>〉
- WERSIG, Gernot:** Information-Kommunikation-Dokumentation. Band 5, Verlag Dokumentation, 1971
- WETTLAUFER, Jörg:** Der nächste Schritt? Semantic Web und digitale Editionen. *Digital Scholarship in the Humanities*, 30 2018 Nr. 1, S.187–198 〈URL: http://www.zfdg.de/sb002_007〉
- WINTERGRÜN, Dirk:** Netzwerkanalysen und semantische Datenmodellierung als heuristische Instrumente für die historische Forschung. Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2019
- WOLF, Willi; König, E./Zedler, P. (Hrsg.):** Qualitative versus quantitative Forschung. Band Bd. I: Grundlagen qualitativer Forschung, Beltz, 1995, S.309–329
- WÜRZ, Simone:** Methoden der Digital Humanities in der Bearbeitung und Erforschung mittelalterlicher Rechnungsbücher. Möglichkeiten und Grenzen am Beispiel der digitalen Edition der Augsburger Stadtrechnungsbücher. In **Pätzold, Stefan/Stumpf, Marcus (Hrsg.):** Mittelalterliche und frühneuzeitliche Rechnungen als Quellen der landesgeschichtlichen Forschung. Münster, 2016, 101–113
- ZIEGLER, Walter:** Studien zum Staatshaushalt Bayerns in der zweiten Hälfte des 15. Jahrhunderts. Die regulären Kammereinkünfte Niederbayerns 1450-1500. München, 1981

Abbildungsverzeichnis

1	Zusammenhang zwischen Induktion und Deduktion in empirischen Wissenschaften, de.wikipedia.org/wiki/Empirie , 09.11.2019.	14
2	Nichteindeutigkeit nicht-formaler Modelle, KÖNIG Bettina: Vorlesung "Modellierungsmethoden der Informatik", ti.inf.uni-due.de 10.06.2019	24
3	Klassendiagramm nach UML, de.wikipedia.org/wiki/UML , 10.06.2019	26
4	Signal, Daten, Information und Wissen. POLLIN Christopher: Vom Suchen, Stöbern und Finden : Information Retrieval am Beispiel der Digitalen Sammlung des Hans Gross Kriminalmuseums, Masterarbeit Graz, S.21	28
5	Datenpyramide geisteswissenschaftlicher Forschungsdaten im institutionellen Kontext, ANDORFER, Peter: Forschungsdaten in den (digitalen) Geisteswissenschaften: Versuch einer Konkretisierung, 2015, S.14	29
6	1. Seite von George Washington, 1750-72, <i>Ledger Book A</i> , Library of Congress, Manuscript Division, STERTZER Jennifer: doi.org/10.16995/dscn.57 . .	36
7	Seite 3 des Objektes <i>00133_sv0061</i> der Stagville Accounts, fromthepage.com/agbedavies/stagville-accounts/00133-sv0061/display/7884 . .	37
8	Portrait von Laban Morey Wheaton, CLARK, George Faber: A History of the Town of Norton, Bristol County, Massachusetts, from 1669-1859. Crosby, Nichols, and Company, and author at Norton, 1859, S.497.	39
9	Seite 100 des <i>Daybook</i> , hdl.handle.net/11040/17982 , 1828-1859	41
10	Web of Data Stack, en.wikipedia.org/wiki/Semantic_Web_Stack , 29.12.2019. .	47
11	Beispiel eines Tripel	48
12	Visualisierung eines Graphen auf Basis eines RDF-Datensatz, w3.org/TR/rdf11-primer , 10.04.2019.	49
13	RDFS-Beispiel auf Basis des FOAF-Vokabulars, eigene Darstellung.	52
14	Pluralistische Textbegriff als Rad, SAHLE: Digitale Editionsformen: Textbegriffe und Recodierung, S.45-49	60
15	Graphische Darstellung des RDF, eigene Darstellung, 01.06.2019.	69
16	Screenshot der Product Treemap des Atlas of Economic Complexity, Exporte der USA im Jahr 2016, atlas.cid.harvard.edu 09.12.2019.	73
17	Treemap nach Häufigkeit von Waren und Dienstleistungen. Gruppiert nach tierischen (hellgrün) und pflanzliche (dunkelgrün) Lebensmittel, Dienstleistungen (orange), Rohstoffen (hellblau), Textilien (türkis) und sonstigen Waren (dunkelblau), eigene Darstellung, erstellt mit Tableau Desktop	74
18	Geldfluss nach Jahren (x-Achse) und Summe der Geldflüsse (y-Achse). Stacks nach einzelnen Akteuren, eigene Darstellung, erstellt mit Tableau Desktop . . .	76