

# Formale, digitale Methoden und Modelle in den Geschichtswissenschaften. Am Beispiel digitaler Rechnungsbücher

Christopher Pollin

6. Oktober 2019, Graz

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Aufbau der Arbeit und Literaturübersicht . . . . .	2
1.2	Related Work . . . . .	4
<b>2</b>	<b>Formale Methoden und Modelle in den Geschichtswissenschaften</b>	<b>6</b>
2.1	Von der erzählenden zur theorieorientierten Geschichtswissenschaft . . . . .	7
2.1.1	Geschichte und Geschichtswissenschaft: eine Definition . . . . .	8
2.1.2	Hermeneutik - Empirismus . . . . .	9
2.1.3	”Klassische” Methoden . . . . .	10
2.2	(Digitale) Quantifizierer und Traditionalisten: Entwicklung einer formalen, digitalen Methodik . . . . .	12
2.2.1	Interpretation und Theoriebildung . . . . .	12
2.3	Modellbildung als Kern der Digital History . . . . .	14
2.3.1	Forschungsdaten in den Geschichtswissenschaften . . . . .	17
2.3.2	Modelle in den Geschichtswissenschaften . . . . .	19
<b>3</b>	<b>Historische Rechnungsbücher als Quelle</b>	<b>20</b>
3.1	Finanztransaktionen . . . . .	22
3.2	DEPCHA . . . . .	23
3.2.1	The George Washington Financial Papers . . . . .	23
3.2.2	Stageville . . . . .	23
3.2.3	The Wheaton Accounts . . . . .	23

<b>4</b>	<b>Web of Data</b>	<b>29</b>
4.1	Geschichte und Vision des Web of Data . . . . .	29
4.2	Web of Data Stack . . . . .	30
4.2.1	Resource Description Framework . . . . .	31
4.2.2	Resource Description Framework Schema (RDFs) . . . . .	33
4.2.3	Taxonomien: Simple Knowledge Organisation (SKOS) . . . . .	35
4.2.4	Abfragesprache: SPARQL . . . . .	35
4.3	Ontologien . . . . .	36
4.3.1	Vom Wissen, über das Semantische Netz zur Ontologie . . . . .	37
<b>5</b>	<b>Anwendung formaler Methoden auf Basis einer Ontologie</b>	<b>40</b>
5.1	Die Bookkeeping-Ontology als konzeptuelles Model für digitale Editionen von historischen Rechnungsbücher . . . . .	40
5.2	Digitale Edition von Historischen Rechnungsbüchern . . . . .	43
5.2.1	Digitale Edition . . . . .	43
5.2.2	Bookkeeping-Ontology . . . . .	43
5.2.3	TEI Beispiele DEPCHA . . . . .	44
5.3	Informationsvisualisierung . . . . .	45
5.3.1	Statistische Auswertung . . . . .	49
<b>6</b>	<b>Textfragmente</b>	<b>49</b>
<b>7</b>	<b>Zusammenfassung</b>	<b>49</b>

# 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.<sup>1</sup> Für Historiker\*innen 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 die 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 Geschäftsmann 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 anderen Akteuren. 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”<sup>2</sup> - mit einer gewissen Datengrundlage zu untermauern. Auf dieser Datengrundlage, die stets offen und nachvollziehbar, sowie in die fachwissenschaftliche Diskussion eingebettet sein muss und deswegen einem konzeptuellen Modells bedarfs, können die Interpretationen der Historiker\*innen 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 Modellen 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)*<sup>3</sup> – umfasst eine

---

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

<sup>2</sup>Vgl. THALLER, Manfred: Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte ,unscharfer ‘Systeme [1984]. *Historical Social Research/Historische Sozialforschung*. Supplement, 2017, S.159.

<sup>3</sup>NHPRC-Mellon Implementation Grants: Digital Edition Publishing Cooperatives, <https://www.archives.gov/nhprc/announcement/depc>; Finanziert durch The Andrew W. Mellon Foundation <https://mellon.org>.

Kooperation von Partnern aus der USA, koordiniert durch das *Wheaton College Massachusetts*, 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 Herangehensweise von Historiker\*innen an Forschungsfragen in den Geschichtswissenschaften. Einen zentralen Stellenwert darin hat die sogenannte *Bookkeeping Ontology*, die als konzeptuelles Modell die "natürliche", semantische Struktur von Rechnungsunterlagen beschreibt und im Sinn des *Web of Data* (aka. Semantic Web) implementiert ist. **Die Forschungsfrage dieser Arbeit befasst sich damit, inwieweit Technologien des *Web of Data* geschichtswissenschaftliche Domänen, in diesem Fall historische Rechnungsbücher, beschreiben können, damit die Anwendung formaler Methoden nachvollziehbar ist.**

## 1.1 Aufbau der Arbeit und Literaturübersicht

Am Anfang der Arbeit steht neben einer allgemeinen theoretischen Diskussion zu Theorien und Methoden in den Geschichtswissenschaften<sup>4</sup> eine Auseinandersetzung mit der Anwendung formaler Methoden in den Geschichtswissenschaften<sup>5</sup> unter Einbeziehung von Inhalten aus dem Bereich der historischen Fachinformatik,<sup>6</sup> der digitalen Geisteswissenschaften,<sup>7</sup> Digital History.<sup>8</sup> Ziel ist es einen Überblick darüber zu verschaffen, welche Ziele die Geschichtswissenschaft verfolgt. Dabei wird besonders Rücksicht auf die "digitale" Komponente<sup>9,10</sup> in diesem Zusammenhang genommen, um zu zeigen, dass es keine Gegensätze zwischen "analoger" und "digitaler" Geschichtswissenschaft gibt, sondern Synergien.

Um Modelle<sup>11</sup> möglichst transparent zu halten, kann man sich dem *Web of Data* bedienen. Im zweiten 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,<sup>12</sup> technische<sup>13</sup> und kritische<sup>14</sup>

---

org/ 05.06.2019.

<sup>4</sup>JORDAN, Stefan: Theorien und Methoden der Geschichtswissenschaft. UTB GmbH, 2018.

<sup>5</sup>THALLER: Historical Social Research/Historische Sozialforschung. Supplement 2017.

<sup>6</sup>THALLER, Manfred: Historical Information Science: Is there such a Thing? New Comments on an old Idea [1993]. Historical Social Research/Historische Sozialforschung. Supplement, 2017.

<sup>7</sup>JANNIDIS, Fotis/KOHLE, Hubertus/REHBEIN, Malte: Digital humanities: eine Einführung. Springer-Verlag, 2017.

<sup>8</sup>GRAHAM, Shawn/MILLIGAN, Ian/WEINGART, Scott: Exploring big historical data: The historian's macroscope. World Scientific Publishing Company, 2015.

<sup>9</sup>FÖHR, Pascal: Historische Quellenkritik im Digitalen Zeitalter. Dissertation, Universität Basel, 2017.

<sup>10</sup>WINTERGRÜN, Dirk: Netzwerkanalysen und semantische Datenmodellierung als heuristische Instrumente für die historische Forschung. Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2019.

<sup>11</sup>Dem Modellbegriff folgend nach: STACHOWIAK, Herbert: Allgemeine Modelltheorie. 1973.

<sup>12</sup>BERNERS-LEE, Tim/HENDLER, James/LASSILA, Ora: The semantic web. Scientific American, 284 2001, Nr. 5.

<sup>13</sup>BERNSTEIN, Abraham/HENDLER, James/NOY, Natalya: A New Look at the Semantic Web. Commun. ACM, 59 2016, Nr. 9 (URL: <http://doi.acm.org/10.1145/2890489>).

<sup>14</sup>SWARTZ, Aaron: Aaron Swartz's A Programmable Web: An Unfinished Work. Synthesis lectures on the semantic web: Theory and Technology, 3 2013, Nr. 2.

Diskussion dieses Themenbereiches notwendig. Ein zentraler Aspekt dabei sind die Themen Wissensmodellierung,<sup>15</sup> Ontologien<sup>16</sup> und *Linked Open Data*,<sup>17</sup> die dazu dienen können Modelle und Forschungsdaten<sup>18</sup> über das Web auszutauschen.

Das 3. Kapitel versteht sich als Quellenstudie und behandelt den Quellentypus historischer Rechnungsbücher<sup>19</sup> und ihre Edition.<sup>20</sup> In einem zweiten Schritt werden konkreten Editionsprojekten aus dem Projektkontext DEPCHA angeführt. Diese Auswahl umfasst drei Projekte, die *George Washington's Financial Papers*,<sup>21</sup> die *Wheaton Family Papers*<sup>22</sup> und die *Stagville Accounts*.<sup>23</sup>

Der gemeinsame Nenner diese Projekte besteht neben der zeitlichen und regionalen Einordnung der Quellen, auch der methodische Zugang der Erschließung. Diese erfolgt im Sinne einer digitalen Editionen.<sup>24</sup> Der dabei verwendete Standard ist die *Text Encoding Initiative (TEI)*.<sup>25</sup> dieses Kapitel widmet sich, wie digitale Edition von historischen Rechnungsbüchern umgesetzt werden können<sup>26,27</sup> Die digitale Edition in dieser Arbeit versteht sich als quellen-orientierte und inhaltsbezogene Edition, in der nicht jedes textuelle Phänomen ediert werden muss, sondern die semantische Struktur der Quellen im Vordergrund stehen.<sup>28</sup> Zur formalen Beschreibung semantischer Strukturen kann das *Web of Data* herangezogen werden. Dies soll am Beispiel der *Bookkeeping Ontologie*,<sup>29</sup> 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 Informationsvisualisierung<sup>30</sup> beschrieben und reflektiert.

---

<sup>15</sup>KELLY, Richard: Practical knowledge engineering. Elsevier, 2016.

<sup>16</sup>STUCKENSCHMIDT, Heiner: Ontologien: Konzepte, Technologien und Anwendungen. Springer-Verlag, 2009.

<sup>17</sup>BAUER, Florian/KALTENBÖCK, Martin: Linked open data: The essentials. Edition mono/monochrom 2011.

<sup>18</sup>NEHER, Günther/RITSCHER, Bernd: Semantische Vernetzung von Forschungsdaten. In STEPHAN BÜTTNER, HANS-CHRISTOPH HOBOHM, Lars Müller (Hrsg.): Handbuch Forschungsdatenmanagement. Band 6, 2011.

<sup>19</sup>GLEBA, Gudrun/PETERSEN, Niels: Wirtschafts- und Rechnungsbücher des Mittelalters und der Frühen Neuzeit- Formen und Methoden der Rechnungslegung: Städte, Klöster und Kaufleute. Universitätsverlag Göttingen, 2015.

<sup>20</sup>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: [http://www.zfdg.de/sb001\\_007](http://www.zfdg.de/sb001_007)).

<sup>21</sup>The George Washington Financial Papers Project, <http://financial.gwpapers.org>

<sup>22</sup>Wheaton Family Papers, <https://digitalrepository.wheatoncollege.edu/handle/11040/7928>

<sup>23</sup>StagvilleAccounts, <https://fromthepage.com/agbedavies/stagville-accounts>

<sup>24</sup>SAHLE, Patrick: Digitale Editionsformen: Textbegriffe und Recodierung. Band 3, BoD–Books on Demand, 2013.

<sup>25</sup>CUMMINGS, James: The text encoding initiative and the study of literature. A Companion to Digital Literary Studies 2013.

<sup>26</sup>TOMASEK, Kathryn/BAUMAN, Syd: Encoding financial records for historical research. Journal of the Text Encoding Initiative 2013, Nr. 6.

<sup>27</sup>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.

<sup>28</sup>VOGELER, Georg: The 'assertive edition'. International Journal of Digital Humanities, 2019.

<sup>29</sup>Bookkeeping Datamodel for Historical Accounts, [gams.uni-graz.at/o:depcha.bookkeeping](https://gams.uni-graz.at/o:depcha.bookkeeping)

<sup>30</sup>FRANK, Ingo: Visualisierungswerkzeuge zur Erklärung historischer Ereignisse: geschichtstheoretische Anforderungsanalyse und zeichentheoretisches Rahmenwerk. INF-DH-2018 2018.

## 1.2 Related Work

Es lässt sich eine Vielzahl von Projekten festmachen, die quantitative, formale, und/oder digitale Verfahren zur Berbeitung und Erschließung von historischen Quellen verwenden und darüber reflektieren. Dabei gibt es auch zahlreiche Beispiele für den Quellentypus der historischen Rechnungsbücher, digitaler Edition, sowie der Anwendung von *Web of Data* Technologien.

VASOLD hat in seiner Arbeit in den 1990er Jahren mit der computergestützten Verarbeitung der Itinerar<sup>31</sup> 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 diese Daten im quellenorientierten Datenbanksystem Kleio<sup>32</sup> verarbeitet. 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.<sup>33</sup>

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.<sup>34</sup> So wird das Fundament für künftige Erforschungen von beispielsweise Preisen, Münzen oder prosopographische, wie auch onomastische<sup>35</sup> Studien gelegt.

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 konzeptuellen Modell semantisch beschrieben wurden. Täter, Tat, Opfer, Strafen und personenbezogene Information werden so maschinenlesbar strukturiert und können weiter ausgewertet werden.<sup>36</sup>

Bei Rechnungsbüchern handelt es sich um pragmatisches, aus dem Alltag der Menschen stammendes Schriftgut, das zur Dokumentation und als Gedächtnisstütze dient. Die einheitliche und ”natürliche” Struktur, sowie der Inhalt solcher Quellen, eignen sich für die Anwendung quan-

---

<sup>31</sup>Als Itinerar versteht man nicht nur Reiseberichte, sondern auch die Methode der Rekonstruktion der getätigten Wege einer Person oder Gruppe.

<sup>32</sup>ToDo: Definition + Literatur zu Kleio

<sup>33</sup>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.

<sup>34</sup>PERSTLING, Matthias: Multimediale Dokumentation und Edition mehrschichtiger Texte : das steirisch-landesfürstliche Marchfutterurbar von 1414/1426. Dissertation, 2013.

<sup>35</sup>Onomastik bezeichnet die Namenforschung, die die Verbreitung, Bedeutung und Herkunft von Eigennamen erforscht

<sup>36</sup>POLLIN, 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.

titativer (und computergestützter) Methoden. So lassen sich diesbezüglich eine Vielzahl von Projekten festmachen.

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

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.<sup>38</sup>

Diese umfassen beispielsweise Schuld- und Rechnungsbücher des Deutschen Ordens um 1400,<sup>39</sup> oder Rechnungsunterlagen der Grafen von Savoyen aus dem 13. Jahrhundert.<sup>40</sup>

Leuven-Beispiel

Mannheim-Beispiel

---

<sup>37</sup><https://www.siebenbuerger.de/zeitung/artikel/kultur/14043-rechnungsbuecher-von-bistritz-aus-den.html>, 15.07.2019

<sup>38</sup>VOGELER et al.: The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies, S.11-13, <https://gams.uni-graz.at/srbas>.

<sup>39</sup>todo

<sup>40</sup>todo

## 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 Historiker\*innen fusst.<sup>41</sup>

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 werden. Die Menge an unterschiedlichen Fragestellungen ist groß und wird in Kapitel ?? 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 Historiker\*innen.

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 "historischen 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.

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 dieser fachlichen Ausrichtungen ist die Erarbeitung und Anwendung von Modellen, um **historische Forschungsdaten** sinnvoll in eine verarbeitbare Strukturen überführen zu können. So setzt sich dieses Kapitel auch mit einer grundlegenden Erörterung der **Modelltheorie** und

---

<sup>41</sup>VOGELER: Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert?, 06.06.2019.



Modellbildung 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”*<sup>42</sup>

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 werden stets positiv hervorgehoben. Sie fordern eine stärkere Theoriebildung in der Geschichtswissenschaft und orientieren sich verstärkt 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 Theorieorientierung in den Geschichtswissenschaften wichtiger Bestandteil. KOCKA versteht darunter einen expliziten Gebrauch von Modellen, Theorien und Begriffen zur Strukturierung des Gegenstandes.<sup>43</sup> Je stärker man sich beispielsweise mit historischen Fragestellungen, die soziale oder wirtschaftliche Phänomene untersuchen, wie die Beziehungen zwischen sozialen Klassen oder wie ökonomisches Wachstum, steigt auch die Notwendigkeit mehr über Fragestellungen und Theorien der Sozial- und Wirtschaftswissenschaft zu wissen..<sup>44</sup> So ist es unabdingbar über ein grundlegendes Verständnis über Modelle der Konjunkturtheorie oder klassentheoretische Konzepte der Sozialgeschichte zu verfügen.<sup>45</sup> Der vermehrt Einsatz formale Verfahren zur Erschließung, Auswertung und Darstellung von Information in historischen Quellen, befördert diese Entwicklung weiter.

<sup>42</sup>Vgl. MANN, Golo: Plädoyer für die historische Erzählung. In KOCKA, Jürgen/NIPPERDEY, Thomas (Hrsg.): Theorie und Erzählung in der Geschichte. München, 1979, S.53.

<sup>43</sup>MAGERSKI, Christine et al.: Wie schreibt man Geschichte? Zagreber Germanistische Beiträge, 2009, Nr. 18, S.2.

<sup>44</sup>KOCKA, Jürgen: Theorien in der Geschichtswissenschaft. In Theoriendebatte 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.6-8.

<sup>45</sup>SCHMIEDER, Felicitas/SOKOLL, Thomas: Theorien und Methoden der Geschichtswissenschaft., S.1.

### 2.1.1 Geschichte und Geschichtswissenschaft: eine Definition

Gegenstandsbereich der **Geschichte** ist das Handeln der Menschen in der Vergangenheit. 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, z.B. durch eine Ausgrabung gefundenen Kleidungsstück als Überreste, und bewusst überlieferten Quellen, wie etwas ein Denkmal, als Tradition.<sup>46</sup> 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.<sup>47</sup>

Jedenfalls ist der Begriff mehrdeutig. Geschichte kann als Vergangenheit, Geschichtsforschung oder Geschichtsschreibung verstanden werden. 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.<sup>48</sup>

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 früher 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.

Im 16. Jahrhundert entwickelt sich der Begriff Geschichte, der anfangs nur für den Gegenstandsbereich der Historie verwendet wurde.<sup>49</sup> Das moderne Verständnis des Wortes Geschichte umfasst sowohl die Gesamtheit von realen Geschehnissen, als auch den Bericht 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.<sup>50</sup> Jede und jeder soll in der Lage sein, feststellen zu können, ob eine Interpretation eines historischen Sachverhalts durch eine angeführte Quelle belegbar ist. Forschungsgegenstand und Forschungsfrage sind in den Geschichtswissenschaften nicht klar abgegrenzt und die Interessen von Historiker\*innen sind

<sup>46</sup>OPGENOORTH, Ernst/SCHULZ, Günther: Einführung in das Studium der Neueren Geschichte. Paderborn, München, Wien, Zürich, 2010, 49–55.

<sup>47</sup>KIRN, Paul: Einführung in die Geschichtswissenschaft. Berlin, Boston, 2015, S.7-12.

<sup>48</sup>FRANK: INF-DH-2018 2018, S.5-7.

<sup>49</sup>OPGENOORTH/SCHULZ: Einführung in das Studium der Neueren Geschichte, S.57-58.

<sup>50</sup>DEMANDT, Alexander: Philosophie der Geschichte von der Antike zur Gegenwart. 2011, S.13-32.

interdisziplinär, weswegen auch die angewandte Methodik vielfältig ist.<sup>51</sup> Auf die Methoden in den Geschichtswissenschaften wird in Kapitel 2.1.3 detaillierter eingegangen. Zwei grundlegende Strömungen in der Erkenntnis- und Wissenschaftstheorie sind wichtig und werden im Folgenden weiter ausgeführt: die Hermeneutik und der Empirismus.

### 2.1.2 Hermeneutik - Empirismus

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. Im großen Unterschied zu naturwissenschaftlichen Fragestellungen, ist die Vergangenheit nicht in einem Experiment beweis- bzw. widerlegbar. Geprägt wird die Hermeneutik, unter anderen, von den Arbeiten von SCHLEIERMACHER, der sie zu einer *universellen Theorie des Verstehens* weiterentwickelte, in der jedes Verstehen ein individueller Prozess ist. Aus dem neuen Geschichtsverständnis des 19. Jahrhunderts entwickelte DILTHEY die Hermeneutik zu einer historischen Methodik. Bei ihm wird Verstehen zur Methode in der Geisteswissenschaften. Wo Naturwissenschaften die Welt von außen betrachten und versuchen sie zu erklären, ist die Aufgabe der Geisteswissenschaften, die Welt von innen zu verstehen. In der Hermeneutik ist Verstehen essentiell. Unter GADAMER wird die Hermeneutik zur Philosophie weiterentwickelt. Es bedarf immer ein Vorverständnis, um in einen Prozess des Verstehens eintauchen zu können, an dessen Ende nur eine Annäherung an eine Wahrheit möglich ist.<sup>52</sup>

Nicht nur unser Wissen über historische Quellen, sondern alles Wissen beruht auf einem Verstehen, das in einer Auslegung unseres Wissens erläutert wird. Verstehen in diesem Sinne ist zirkelförmig. Jemand tritt mit einem gewissen Vorwissen an eine Fragestellung heran. Es werden in der Auseinandersetzung mit der Materie Erfahrungen 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 besseres Bild zu einem Thema. Eine Problemstellung muss zuerst gefunden, erkannt und verstanden werden, damit sie weiter verarbeitet werden kann.<sup>53</sup>

In folgendem Beispiel nach STANGL sind die Wörter „*vihe*“, „*verwürckt*“ oder „*fewer*“ ohne jeglichen Kontext unverständlich. Werden sie aber in ihrem Kontext gelesen, wird die Bedeutung des Textes offensichtlich:

*"Item so eyne mensch mit eyne vihe, mann mit mann, weib mit weib, vnkeusch treiben, die haben auch das leinen verwürckt, vnd man soll sie der gemeinen gewonheit nach mit dem fewer vom leben zum todt richten"*

<sup>51</sup>REICHE, Ruth/BECKER, Rainer/BENDER, Michael: Verfahren der Digital Humanities in den Geistes- und Kulturwissenschaften. Niedersächsische Staats- und Universitätsbibliothek Göttingen, 2014, S.13.

<sup>52</sup>OPGENOORTH/SCHULZ: Einführung in das Studium der Neueren Geschichte, S.19-30.

<sup>53</sup>STANGL, Werner: Geschichte der Hermeneutik, <https://arbeitsblaetter.stangl-taller.at/ERZIEHUNGSWISSENSCHAFTGEIST/HermeneutikHistorie.shtml>, 01.06.2019.

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 Teilstück vom Ganzen her verstanden, korrigiert oder erweitert wird und sich umgekehrt das Ganze von den Teilen her bestimmt.

## Der Empirismus...

### Gegenüberstellung Hermeneutik - Empirismus

Im Zusammenspiel mit empirischer Methoden ist die Hermeneutik für die Hypothesenbildung und für die Interpretation der Ergebnisse von großer Bedeutung. Bei der empirischen Überprüfung von Sachverhalten spielt die Hermeneutik eine wesentliche Rolle bei der Quantifizierung von qualitativen Aussagen. Auch die Interpretation von empirischen Resultaten ist, so STANGL, ein hermeneutischer Vorgang.<sup>54</sup>

”Das Verhältnis zwischen Hermeneutik und empirischer Wissenschaft ist in der philosophischen Debatte bis heute kontrovers. Insbesondere in den Sozialwissenschaften wurde diese Debatte zwischen Vertretern einheitswissenschaftlicher Positionen, wie sie die Vertreter des Kritischen Rationalismus Karl Popper und Hans Albert einnehmen, und alternativen Positionen (etwa der Kritischen Theorie um Max Horkheimer und Theodor W. Adorno), die sich gegen eine ihrer Meinung nach „blinde“ Übertragung naturwissenschaftlicher Erkenntnismodelle auf die Sozial- und Geisteswissenschaften gewehrt haben, intensiv in den 1960er und 70er Jahren ausgetragen (vgl. den sogenannten Positivismusstreit).”

### 2.1.3 ”Klassische” Methoden

Methoden hingegen sind und waren ein stets anerkanntes und nicht wegzudenkendes Werkzeug in den Geschichtswissenschaften. Ein Blick auf die Vielfalt historischer Hilfswissenschaften, die sich mit der Erschließung von bestimmten Quellentypen beschäftigt, zeigt dies. Die Hilfswissenschaften bzw. Grundwissenschaften – ”als Abzweigungen der allgemeinen historischen Arbeitsform”[Vgl.][S.10]<sup>55</sup> – mit ihrem Ziel die historischen Quellen aufzubereiten, erstrecken sich von der Paläografie über die Diplomatik und Aktenkunde bis hin zur Genealogie.

Die Historische Fachinformatik versteht sich auch als Hilfswissenschaft innerhalb der Geschichtswissenschaften und beschäftigt sich mit der Anwendung und Reflexion von formalen Verfahren im Fach.<sup>56</sup>

<sup>54</sup>STANGL, Werner: Der hermeneutische Zirkel, <https://arbeitsblaetter.stangl-taller.at/ERZIEHUNGSWISSENSCHAFTGEIST/HermeneutikZirkel.shtml>, 01.06.2019.

<sup>55</sup>VON BRANDT, Ahasver: Werkzeug des Historikers: eine Einführung in die historischen Hilfswissenschaften. Band 33, W. Kohlhammer Verlag, 2007.

<sup>56</sup>Was ist HFI?, Kurzbeschreibung des Faches und seiner Positionierung <http://hfi.uni-graz.at/basisinformation/was-ist-hfi/>, 23.05.2019.

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. Historiker\*innen formulieren eine Fragestellung, als Ergebnis eines spezifischen Forschungsstandes, um eigenen 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 Fachkollegen\*innen verglichen und in übergeordnete Kontexte eingeordnet. Für DROYSEN mündet die Reflexion auf die historische Methode somit gleichsam wie von selbst in eine Theorie der historischen Erkenntnis, der Hermeneutik. Aus diesem Verständnis heraus erwuchs auch die Anwendung der Hermeneutik als Methode der Geschichtswissenschaft.

In der Geschichtswissenschaft von heute gibt es eine Vielzahl von Methoden, deren Abgrenzung zu Theorien oft verschwimmen. Diskursanalyse oder Konflikttransformation, um zwei moderne Beispiele anzuführen, lassen sich sowohl als Theorie, also auch als Methode verstehen. Desweiteren verschwimmt die Unterscheidung zwischen qualitativer und quantitativer Methode. Erstere getragen von einem hermeneutischen, Zweitere von einem empirischen Verständnis.<sup>57</sup>

**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 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.<sup>58</sup> 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.

<sup>59</sup> Gerade in den Geschichtswissenschaften der 1980er war diese eine Auseinandersetzung zwischen "Traditionalisten" und "Quantifizierern".

---

<sup>57</sup> SCHMIEDER/SOKOLL: Theorien und Methoden der Geschichtswissenschaft, S.2-5.

<sup>58</sup> WOLF, Willi: Qualitative versus quantitative Forschung. Bilanz qualitativer Forschung, 1 1995, S.309-329.

<sup>59</sup> JARAUSCH, Konrad Hugo/ARMINGER, Gerhard/THALLER, Manfred: Quantitative Methoden in der Geschichtswissenschaft. Wissenschaftliche Buchgesellschaft, 1985, S.203-206.

## 2.2 (Digitale) Quantifizierer und Traditionalisten: Entwicklung einer formalen, digitalen Methodik

”Aber alle Autoren halten daran fest, dass sich die Quantifizierung als ein gewichtiges Werkzeug historischer Analyse erwiesen habe”.<sup>60</sup>

In den 1980er 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.<sup>61</sup> 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. [ToDo: vl an diese Stelle ein Beispiel, in dem jemand historische Daten ’falsch’ quantitativ ausgewertet hat] 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 wissenschaftliche Arbeit.

ToDo LATOUR<sup>62</sup>

Es wird empfohlen die grundlegenden Annahmen und Definitionen, die in der Interpretation von Quellen verwendet werden in einer, und so haben es bereits Vertreter der Historischen Fachinformatik gefordert, gemeinsamen *knowledge domain* zu formalisieren.<sup>63</sup> Die standardisierten Konzepten und Technologien des *Web of Data* (aka. Semantic Web), sowie des *Linked Open Data* haben zum Ziel Wissensdomänen als konzeptuelle, maschinenlesbare Modelle zu formalisieren und nach diesen Modellen 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 wir aber erst später in Kapitel 4 und 5 weiter eingegangen.

### 2.2.1 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 Historikerin in die Lage größere Mengen von Belegen zu verarbeiten. Genau darin besteht der Mehrwert qualitativer Methodik und 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

<sup>60</sup>JARAUSCH/ARMINGER/THALLER: Quantitative Methoden in der Geschichtswissenschaft, S.191-206.

<sup>61</sup>ebd., S.XX-XX.

<sup>62</sup>LATOUR, Bruno: Zirkulierende Referenz. Bodentischproben aus dem Urwald am Amazonas. 2016.

<sup>63</sup>THALLER: Historical Social Research/Historische Sozialforschung. Supplement 2017, S.XX-XX.

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 zusammensetzen. Fehlerhafte Vorbedingungen führen zu fehlerhaften Ergebnissen, die wiederum verzerrte Interpretationen mit sich führen können. Gehört ein Kaufmann in die Gruppe des Besitzbürgertum oder zum alten Mittelstand? Es ist notwendig, dass das Erkenntnisziel eines Verfahren 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.<sup>64</sup> Bereits 1988 veranschaulichte THALLER das in seinem sogenannten "Preußenbeispiel". Der Begriff "Preußen" bezieht sich im Jahre 1670 auf deutlich andere Koordinaten, als im Jahre 1770. Wer als also "Preuße" zu kategorisieren ist, ist abhängig zu welcher Zeit und an welchem Ort eine Person lebt bzw. geboren wurde.<sup>65</sup> 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 Kontexten ab. Dies umfasst zum Beispiel politischen und administrativen Grenzen, wie im "Preußenbeispiel" angeführt, oder in Hinblick auf historische Rechnungsunterlagen beispielsweise Währungs- und Maßsysteme.

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 Verfahren auch 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 keine 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.

ToDo: Theorien in der Quantifizierung.

Theoriebedürftigkeit der Historie als Wissenschaft unterstrichen. Der Ruf nach mehr Theorie war weitgehend durch den Versuch motiviert, das Paradigma des Historismus, der die Einzigartigkeit und Individualität einer Entwicklung betont und behauptet, dass Historiker Phänomene in ihrer eigene Zeit und ihrem eigenen Kontext verstehen sollten, abzulösen. KOCKA definiert Theorien als eine expliziten und konsistenten Satz von verwandten Begriffen, die benützt werden können, historische Zusammenhänge, die nicht aus dem Studium der Quelle alleine abgeleitet werden können, zu strukturieren und zu erklären.

---

<sup>64</sup>JARAUSCH/ARMINGER/THALLER: Quantitative Methoden in der Geschichtswissenschaft, S.182-193.

<sup>65</sup>THALLER: Historical Social Research/Historische Sozialforschung. Supplement 2017, S.264-266.

aber n Die Funktionen der historischen Forschung:

- Klärung von Problemformulierungen
- Definition von überprüfbaren Hypothesen für Zusammenhänge von Faktoren
- K
- 
- Aufwerfen neuer Fragestellungen.

66

### 2.3 Modellbildung als Kern der Digital History

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 und die **angewandten digitale Geisteswissenschaften**, die die Anwendung dieser Mittel und Methoden zur Erstellung konkrete formaler Modelle in den geisteswissenschaftlichen Disziplinen befördert.<sup>67</sup>

*”Im wissenschaftlichen wie außerwissenschaftlichen Sprachgebrauch hat gegenwärtig der Modellbegriff zunehmend Relevanz erlangt. Bei zahlreichen passenden – leider auch unpassenden – Gelegenheiten ist von ”Modellen” die Rede.”*<sup>68</sup>

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”*.<sup>69</sup>

Auch KOBLEK 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 zur Definition des Begriffes, Konzepte wie Abstraktion, Entität oder System verwendet werden, die wiederum in anderen Bereichen terminologisch unterschiedliche verwendet werden. Im deutschen Sprachgebrauch lässt sich eine Doppelbedeutung des Modellbegriffs festmachen: als Abbild von etwas, sowie Vorbild für etwas (jemand steht Modell beim Malen).<sup>70</sup>

<sup>66</sup>JARAUSCH/ARMINGER/THALLER: Quantitative Methoden in der Geschichtswissenschaft, S.182-191.

<sup>67</sup>PIOTROWSKI, Michael: Digital Humanities. Computational Linguistics, and Natural Language Processing. Lectures on Language Technology and History 2016.

<sup>68</sup>STACHOWIAK: Allgemeine modelltheorie, S.1.

<sup>69</sup>ebd.

<sup>70</sup>ebd., S.129.



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 ein Bereich eingeschränkt werden muss. 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.

Ein Modell hat eine normative, deskriptive und explorative Funktion.<sup>71</sup> Modellierungen erfolgt auf mehreren Ebenen: als modellierte Instanz, das heißt es ist ein Datenabbild eines Gegenstandes oder Textes (Instanz). Datenmodell sind Ausdruck eines Musters, das auf mehrere Instanzen anwendbar ist. Sowie Metamodell, die ein Ausdruck eines Musters ist, das auf mehrere Datenmodelle anwendbar ist, darstellen.<sup>72</sup>

In den unterschiedliche Disziplinen haben Modelle eine abweichende Funktion. In den Ingenieurwissenschaften steht ein Modell für ein formales Modell, das einen Sachverhalt beschreibt und erklärt. 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 Aufgaben eingesetzt. Zusammengefasst kann man eine große Menge an Modelltypen festmachen.<sup>73</sup>

Den gleichen Prinzipien folgende lassen sich, wie bei den Methoden, 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 Reihung 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. 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.<sup>74</sup>

**Statische Modelle** beschreiben den Zustand eines Systems zu einem bestimmten Zeitpunkt, wogegen **dynamisches Modell** prozesshafte Strukturen eines Systems beschreiben.<sup>75</sup>

Eine weitere Unterscheidung ist in der Einteilung in informellen bzw. nicht-formale, semiformale und formale Modelle gegeben. Existiert keine eindeutige Beschreibungssyntax für ein Modell und es wird mittels natürlicher Sprache beschrieben, so spricht man von einem **informellen Modell**. Ein Sachverhalt in den Geschichtswissenschaften, wie er beispielsweise in einer Publikation skizziert wird, kann als ein solches Modell betrachtet werden. Oft ist dies aber, begründet

---

<sup>71</sup>ToDo

<sup>72</sup>ToDo

<sup>73</sup>KOBLER, Maximilian: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung. Logos Verlag Berlin GmbH, 2010, S.41-44.

<sup>74</sup>WOLF: Bilanz qualitativer Forschung 1 [1995], S.309–329.

<sup>75</sup>ToDo

durch die Verwendung natürliche 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 1 veranschaulicht dies.



**Abbildung 1:** 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 EEntity-Relationship-Modell (ER) oder UML (ToDo) Diagrammen.

Das Entity-Relationship-Modell basiert auf drei Begriffen: Relationen, Attributen und Entitäten. Die Entitäten – alle potenziellen Untersuchungsobjekte – verfügen über Eigenschaften, sogenannte Attribute. Weiters stehen sie zu anderen Entitäten in unterschiedlichen Beziehungen, sogenannten "Relationen".

UML ToDo

**Formale Modelle** verfügen über eine konkrete Syntax und Semantik. Das erlaubt die formale definition von Modellen auf Basis deskriptiver Logiken.<sup>76</sup> Für formale Modelle lassen sich Algorithmen formulieren, die es erlauben Daten, die durch ein solches Modell beschrieben sind, automatisch zu validieren.

Im Gegensatz zum Computer kann der Mensch auch mit informellen Modellen arbeiten. Man

<sup>76</sup>Es gibt unterschiedliche deskriptive Logiken, die sich in ihrer Ausdrucksmächtigkeit und Entscheidbarkeit unterscheiden.

kann stark strukturierte und formalisierte Modelle, die algorithmisch genutzt werden können, offenen Modellen gegenüberstellen, in die sich ein Mensch hineindenken kann.

Konzeptuelle (konzeptionelle, ist ein Synonym) Modelle (conceptual model) bzw. Informationsmodelle (information model) sind beispielsweise in der Wirtschaftsinformatik von zentraler Bedeutung, da Sie als Schnittstelle zwischen Informatik und Wirtschaftsinformatik besonders gut geeignet sind.<sup>77</sup> Ich denke man kann diese genauso auch auf die Digitalen Geisteswissenschaften ummünzen.

huhu<sup>78</sup>

### 2.3.1 Forschungsdaten in den Geschichtswissenschaften und historische Information

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”*.<sup>79</sup> 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.<sup>80</sup> Folgende Darstellung zeigt die Abhängigkeit dieser 4 Begriffe zueinander, mit einem anwachsen der Komplexität der Struktur, sowie der Anbindung an einen kognitiven Agenten.<sup>81</sup>

---

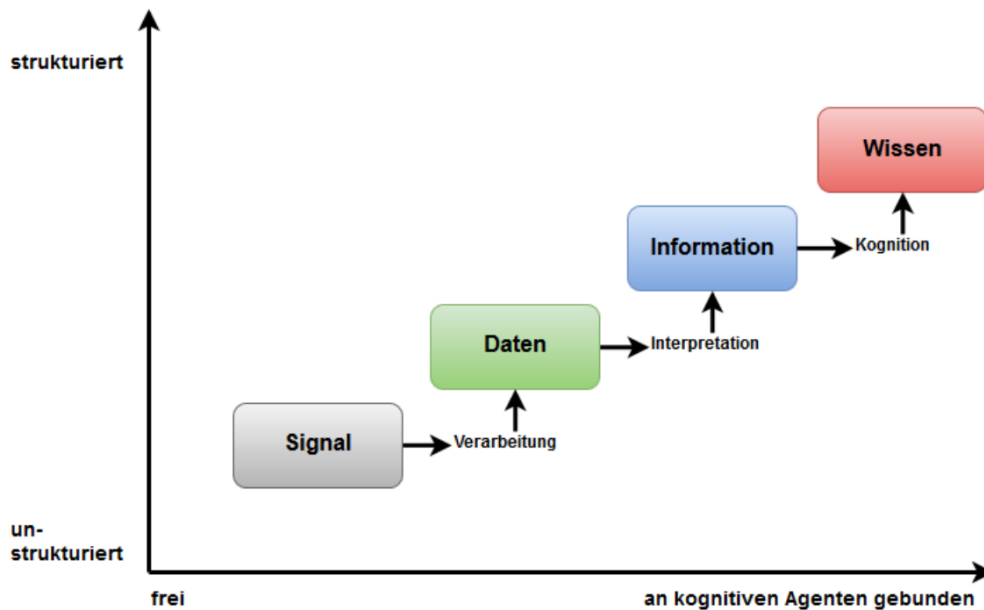
<sup>77</sup> KOBLER: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung, S.44-47.

<sup>78</sup> Vgl JANNIDIS/KOHLER/REHBEIN: Digital humanities: eine Einführung, S.99-108.

<sup>79</sup> WERSIG, Gernot: Information-Kommunikation-Dokumentation. Band 5, Verlag Dokumentation, 1971, S.74.

<sup>80</sup> FAVRE-BULLE, Bernard: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001, S.93-97.

<sup>81</sup> Eine ausführlichere Auseinandersetzung mit den Begriffen Daten, Information und Wissen findet sich in meiner ersten Abschlussarbeit. POLLIN: Vom Suchen, Stöbern und Finden : Information Retrieval am Beispiel der Digitalen Sammlung des Hans Gross Kriminalmuseums, Masterarbeit Graz, S.20-28.



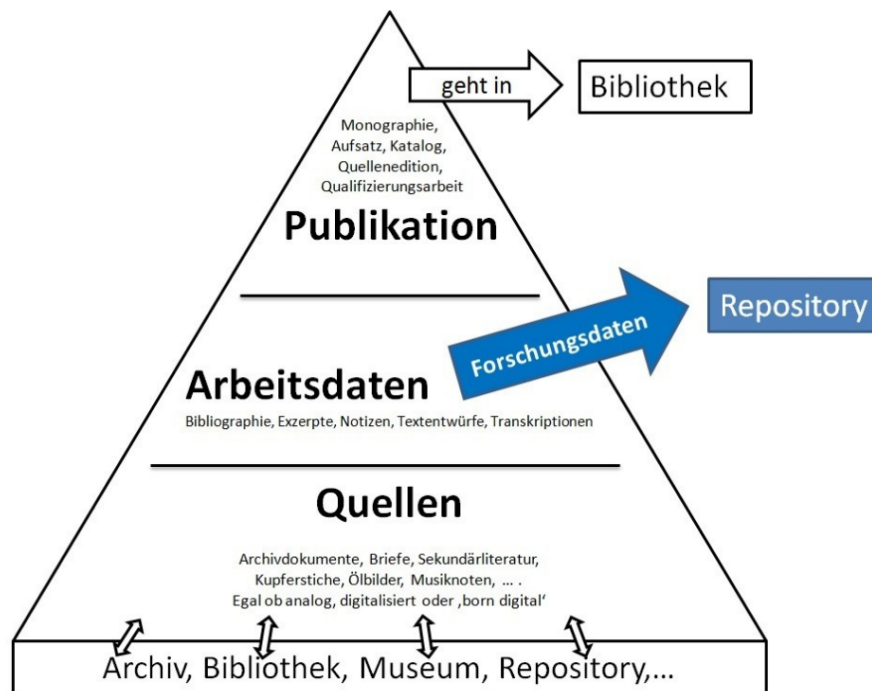
**Abbildung 2:** Signal, Daten, Information und Wissen

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.<sup>82</sup>

ANDORFER spricht sich für einen nicht *”inflationären Gebrauchs dieses Begriffes”* und einer *”präzisere 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 Fachwissenschaftler\*innen vertraute Begrifflichkeit bleiben und .<sup>83</sup>

<sup>82</sup>KINDLING, Maxi et al.: Forschungsdatenmanagement an Hochschulen: Das Beispiel der Humboldt-Universität zu Berlin. LIBREAS. Library Ideas, 2013, Nr. 23 (URL: <https://libreas.eu/ausgabe23/07kindling>), 09.06.2019.

<sup>83</sup>ANDORFER, Peter: Forschungsdaten in den (digitalen) Geisteswissenschaften. Versuch einer Konkretisierung.(DARIAH-DE Working Papers 14). Göttingen: DARIAH-DE. Verfügbar unter urn: nbn:de:gbv, 2015 (URL: <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-14.pdf>).



**Abbildung 3:** 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 nicht eindeutig. HILTMANN sieht einen engeren und einen weiteren Begriff von Forschungsdaten in den Geschichtswissenschaften. Im engeren Sinne versteht man Metadaten zu und Annotationen von historischen Quellen in all ihren Ausprägungen. Als Beispiel ist die archivalische 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.<sup>84</sup>

Im Zuge dieser Arbeit soll aber dem engeren Begriff folge geleistet werden. Forschungsdaten wie Geschichtswissenschaften sind sowohl innere wie äußere merkmale historische Quellen wieso adverbien speichert werden verfügung stehen die interpretierbar und Maschinen verarbeitbar.

### 2.3.2 Modelle in den Geschichtswissenschaften

Damit Forschungsdaten nachnutzbar und nachvollziehbar sind bedarf es eines mitgelieferten Modells. Diese Modelle wiederum müssen auf Standards beruhen. Es ist zeitensiv die Datenmodelle andere zu verstehen und oft ist es leichter sein eigenes Modell zu entwickeln. Ein gene-

<sup>84</sup>Vgl. HILTMANN, Torsten: Forschungsdaten in der (digitalen) Geschichtswissenschaft. Warum sie wichtig sind und wir gemeinsame Standards brauchen. 2018 (URL: <https://digigw.hypotheses.org/2622>), 09.06.2019..

relles Problem mit Informationssystem und Forschungsdaten ist, dass sie nur für eine bestimmte Fragestellung bzw. ein bestimmtes Projekt entwickelt wurden. 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). Wenn auch Umsetzungen der TEI nicht einem 'klassischen' Standard entsprechen, so sind sie Überlegungen, wie man Text strukturieren und annotieren kann, sehr wohl standardisiert.

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 ableiten lassen müssen. THALLER fordert deswegen, dass die Standards zur Beschreibung von Modellen eher beschreibend statt vorschreibend sind.<sup>85</sup>

### 3 Historische Rechnungsbücher als Quelle

In diesem Kapitel wird das Rechnungsbuch als historische Quelle dargestellt und sowohl inhaltliche wie auch äußerliche Eigenschaften festgemacht. In einem weiteren Schritt wird dann auf konkrete digitalen Editionsprojekte von Rechnungsbüchern eingegangen.

Bei Rechnungsbüchern handelt es sich um pragmatisches, aus dem Alltag der Menschen stammendes Schriftgut, das zur Dokumentation und als Gedächtnisstütze diente, bis eine Schuld getilgt wurde. Zu meist auf Papier und ohne übergeordneten Standards verfasst, folgte sie keinen repräsentativen Zweck. Dennoch haben sie zumeist eine einheitliche Struktur, die zeigt, dass es sich um ein "Werkzeug des Alltages" gehandelt hat, damit es auch für andere verständlich und nachnutzbar ist. GLEBA und PETERSEN heben eine Vielfalt der Inhalte in Rechnungsbüchern, sowie die Möglichkeiten der "interdisziplinären Erschließung der Quelle 'Rechnungsbuch'" hervor. Die Herausforderungen der Erschließung der Quelle sind vielfältig: Maß-, Größen- und Gewichtsordnungen und Information zu Geldwerten sind sehr stark an Region und Zeit gebunden. Dennoch lassen sich viele Forschungsfragen aus solchen Quellen ableiten: wie haben sich Güter, Dienstleistungen und Geldbeträge über die Zeit hinweg verändert, welche wirtschaftlichen Abhängigkeiten zwischen Akteuren gibt es? Rechnungsbücher gehen "weit über wirtschafts- und verwaltungsgeschichtliche Aspekte hinausgehen"<sup>86</sup> und so haben Fragestellungen dieser Art weitere soziale, gesellschaftliche und wirtschaftliche Implikationen, die uns ein Verständnis geben können, wie das Leben der Menschen in der Vergangenheit ausgesehen haben kann.<sup>87</sup>

<sup>85</sup>THALLER, Manfred: The Need for Standards: Data Modelling and Exchange [1991]. Historical Social Research/Historische Sozialforschung. Supplement, 2017, S.204.

<sup>86</sup>Nach KLAPP, BRUCH, Julia: Die Kunst, Daten in Informationen umzuwandeln. 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.14.

<sup>87</sup>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, S.7-10.

BRUCH sieht Rechnungsbücher sehr geeignet dafür Daten zu liefern, damit sie mit quantitativen Methoden weiterverarbeitet werden können.<sup>88</sup>

MEDEA-Paper: Historical accounting documents have taken many forms over the course of human history, and their formats have changed in ways that have not always been reflected in the historical literature. While a shorthand reference to the development of so-called double entry accounts based on the practices of early modern Italian merchants might prevail among historians of the modern era, considerable evidence demonstrates a more nuanced process of change over time. In fact, historians have used the term “accounts” for a wide range of primary sources. Long before the fifteenth century, people all over the world tracked exchanges of goods, services, and rights by writing on a variety of materials that included clay and wax tablets, slate, tallies, papyrus, parchment, and paper, and they used a range of tools that were appropriate to these different materials. At various times and places, the people making these records had literacy skills ranging from basic numeracy through full literacy; their purposes in creating the records extended from simple memory aids through complex professional evaluations. The common element among these highly differentiated sources was their focus on economic activities. Thus, we can deduce that the form and inner structure of the accounts evolved in relation to available materials and tools, in association with diverse purposes, and in connection to the development of economic activities. Historians have demonstrated that those responsible for keeping track of the funds of cities, estates, businesses, charities, and other human organizations developed their accounting practices as they encountered new situations and problems. Medieval and early modern accounts from French- and German-speaking regions reveal some of the processes of this non-linear, locally influenced development (Arlinghaus 2000, Mersiovsky 2000). And research on nineteenth-century business accounts has shown that only in the final decades of the century did accounting become a professional pursuit in which such contemporary notions as cost accounting and depreciation of business machinery were standard management principles that reputable firms and organizations were expected to follow (McGaw 1985, Yates 2000). In the following four examples, MEDEA participants illustrate areas where digital scholarly editions of accounts can add to knowledge about the evolution of accounting practices as well as about daily life and the sorts of factors that have been the subject of social and economic history for the past eighty years.<sup>89</sup>

Historische Rechnungsbüchern sind einen Teil einer Menge von Primärquellen, die Aufzeichnungen von Handel mit Waren und Dienstleistungen darstellen. Sie werden seit längerer Zeit als Primärdatensätze für historische Forschung genutzt. Sie liefern reichhaltige und strukturierte

---

<sup>88</sup>BRUCH: Die Kunst, Daten in Informationen umzuwandeln.

<sup>89</sup>CLIFFORD, Anderson et al.: MEDEA (Modeling semantically Enhanced Digital Edition of Accounts) as Historical Method. 2016, S.2.

Daten, die oft längere Zeiträume abdecken und die als Aggregation vieler Einzelinformationen enthalten zu Beantwortung unterschiedlicher Forschungsfragen herangezogen werden können. Sie beinhalten beispielsweise Information über den Kauf und Verkauf von Waren, Werkzeugen und Rohstoffen eines bestimmten Gewerbes, sowie über die ausgezahlten Löhne von Arbeitnehmern\*innen. Eine andere Art von Rechnungsbüchern stellen Haushaltskonten dar. Diese beziehen sich auf den Kauf von Lebensmitteln und Bekleidung sowie für Zahlungen für Dienstleistungen in Form von Kochen, Waschen oder andere Haushaltsarbeiten. HFRs im Zusammenhang mit der Haushaltsführung auch im Zusammenhang mit der Haushaltsführung enthalten Listen der für Hausarbeit gezahlten Löhne, und solche Listen finden sich manchmal auf der Seite Rückseiten von Taschenkalendern und nicht in formalen Kontenbüchern oder Ledgern. In Regionen wo Arbeiter versklavt wurden, finden sich HFRs in den Buchhaltungsbüchern und in der persönlichen Aufzeichnungen über Sklavenhalter und Sklavenhändler.<sup>90</sup>

Eine Transkription allein reicht nicht aus um die unterschiedlichen Dimensionen einer solchen Quelle abzudecken: die linguistische/textuelle, die quantifizierbare und die semantische Dimension. Für Forschungszwecke unterliegen historische Quellen einem Transformationsprozess hin zu (vernetzten) Informationsquellen, die in verschiedenen Forschungsszenarien genutzt werden können. Um dies zu veranschaulichen, werden drei Fallstudien von Projektpartnern und ihren jeweiligen Forschungsinteressen diskutiert, die weit über wirtschaftliche und administrative Aspekte hinausgehen.

Händler haben finanzielle Aufzeichnungen geführt, um Käufe und Verkäufe von Waren seit den Handelsökonomien des alten Mesopotamiens zu verfolgen, und der Impuls, regelmäßige Konten zu führen, führte zu Standards für die Erfassung von Finanzinformationen. Im Laufe der Jahrhunderte boten verschiedene einflussreiche Texte gewöhnlichen Geschäftsleuten die Möglichkeit, zu lernen, wie man das macht. Für das frühneuzeitliche Europa skizzierte Fra Luca Pacioli in seiner Abhandlung *The Rules of Double-Entry Bookkeeping* von 1494 Rechnungslegungsgrundsätze, die die Handelsrechnung in ganz Europa beeinflussten. Mitte des achtzehnten Jahrhunderts veröffentlichte der schottische Mathematiker John Mair ein einflussreiches Lehrbuch mit dem Titel *Buchhaltungsmethodik*; oder, eine methodische Abhandlung von Handelskompten, nach der italienischen Form. Bei der Präsentation von Pacioli's System für Englischsprachige durchlief Mair's Text zahlreiche Ausgaben zwischen den folgenden Sprachen 1736 und 1808.<sup>91</sup>

### 3.1 Finanztransaktionen

*Extensible Business Reporting Language* (XBRL)<sup>92</sup> ist ein XML-basierter Standard und frei verfügbare Sprache für die Beschreibung und den Austausch inhaltlicher und technischer In-

---

<sup>90</sup>TOMASEK/BAUMAN: *Journal of the Text Encoding Initiative* 2013, S.2.

<sup>91</sup>ebd., S.3.

<sup>92</sup>ENGEL, Phillip et al.: *Extensible Business Reporting Language (XBRL) 2.1*. XBRL International 2003.



formation der Geschäftsberichterstattung. XBRL verwendet verschiedenen Spezifikationen und das Herzstück bildet ein *instance document*, das die finanziellen Fakten beschreibt und eine *taxonomy* das die Konzepte erfasst. Kodierung eines Finanzbericht in XBRL erfordert die Verbindung von Instanzdokumenten mit Taxonomien und zugehörige *Linkbases*, die Sammlungen von eingehenden und externen Links enthalten. CLIFFORD et. al führen an, dass es für die Beschreibung von Transaktionen in historischen Rechnungsbüchern geeignet ist, kritisieren aber das es für die Foramlisierung historischer Daten nicht entwickelt wurde. Beispielsweise ist es nicht möglich historische Währungen in XBRL zu beschreiben, da nur die durch ISO 2015 definierten Währungen erlaubt sind..<sup>93</sup>

## 3.2 DEPCHA

Daten - aus unterschiedlichen Formaten - sollen auf einer gemeinsamen Plattform zusammengeführt werden und adäquate Formen des Retrievals, Discoverys und der Visualisierung eröffnen, um die Arbeit mit den Quellen zu erleichtern. Die Überführung nach RDF auf Basis der im Projektkontext entwickelten *Bookkeeping-Ontologie*, die Transferprozesse historischer Rechnungsbücher formalisiert, erlaubt die Interoperabilität, Verlinkung und Zusammenführung der Informationen im Sinne des *Web of Data* und *Linked Open Data*.

### 3.2.1 The George Washington Financial Papers

The George Washington Financial Papers (1748-1799) 3 gives insight into the life of George Washington and other topics such as the material culture, social history, manufacturing and agriculture. The financial papers exist as digital edition, created and published via an open-source, Drupal 4 based editorial platform, and aim to make Washington's records freely accessible. The platform allows editing and publishing financial documents and gives the users the possibility to perform simple analytical functionalities. Samples of research questions that could be of interest to historians are: How much money did Washington spent annually and for which specific commodities? Which role slave trade plays in his business? How did the price of certain commodities fluctuate? What did the network of partners look like and who did business with him? How was the value of tobacco calculated through different currencies [St14]?

### 3.2.2 Stageville

### 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* und seiner Familie, einem Geschäftsbesitzer eines *dry goods store* aus Norton, Massachusetts in den Vereinigten Staaten

---

<sup>93</sup>CLIFFORD et al.: MEDEA (Modeling semantically Enhanced Digital Edition of Accounts) as Historical Method, S.7-9.

von Amerika, dokumentieren. Unter *dry goods* versteht man einen aus dem 17. Jahrhundert stammenden historischen Begriff, der je nach Region leicht unterschiedliche Güter zusammenfasst. Darunter werden im Allgemeinen aber Produkte verstanden, die "trocken" sind, wie etwa Textilien, Konfektionskleidung, Tabak oder bestimmten Lebensmittel, wie etwa Kartoffel.<sup>94</sup>

Die Stadt Norton ist eine charakteristische landwirtschaftlich und von Manufakturen geprägte Kleinstadt im Nordosten der USA im Hinterland der regional bedeutenden Häfen Boston, New Bedford, Newport, and Providence. In einer Geschichte der Stadt Norton von CLARK wird *Laban Morey Wheaton* namentlich genannt. Er lebte von 14. September 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*. Er war verheiratet und Angehöriger des Kongregationalismus, einer Form der christlichen Gemeindeverfassung. Die Wheaton Familie besaß eine Milchviehherde sowie Fabriken zur Produktion von Baumwollwatte, die *Wheaton Manufacturing Company*, und eine zur Produktion von Strohhüten<sup>95</sup> und so gehörte Laban Wheaton zu einem der wohlhabenden Männer der Stadt. Sogar ein Portrait von ihm ist bei CLARK überliefert.<sup>96</sup>



**Abbildung 4:** Portrait von Laban Morey Wheaton

Aus diesem Kontext heraus, der über *Laban Morey Wheaton* überliefert ist 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, Rechnungsbücher von Wheaton zeigen nicht nur Angaben zu Gütern,

<sup>94</sup>Definition von *Dry Goods*, <https://chestofbooks.com/reference/Dictionary-of-Dry-Goods/Dry-Goods.html>, 23.05.2019, Vgl. COLE, George S: A complete Dictionary of Dry Goods and History of Silk, Linen, Wool and other Fibrous Substances. 1892. 2015.

<sup>95</sup>TOMASEK/BAUMAN: Journal of the Text Encoding Initiative 2013, S.6.

<sup>96</sup>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.496.

Dienstleistungen und Geldbeträgen in den einzelnen Transaktionen, sondern auch Namen einer Vielzahl von Individuen und ihrer Familien, die in Beschäftigungsverhältnissen mit Wheaton stehen, Objekte und Gebäude mieteten oder Handel betrieben. TOMASEK führt an, dass diese Rechnungsbücher detaillierte Informationen über das tägliche Leben der Einwohner\*innen einer Stadt in New England, geprägt durch eine gemischte Land-, sowie Industriegewirtschaft im zweiten Quartal des 19. Jahrhunderts, darstellen.<sup>97</sup>

Das *Wheaton College Massachusetts* stellt die *Wheaton Family Papers* Dokumente in ihrem digitalen Repositorium online zur Verfügung.<sup>98</sup>

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 Paper, 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.<sup>99</sup> Teile davon wurden im Zuge des DEPCA Projektes für die Umsetzung eines Prototyps zur Publikation von semantisch angereicherten, digitalen Edition von Rechnungsbüchern verwendet.

Verkauf und Kauf von Gütern des täglichen Lebens listet und ein *Ledger*, ein Hauptbuch, das die Finanzflüsse an einer zentralen Stelle zusammenfasst transkribiert und mit der TEI ausgezeichnet.<sup>100</sup>

Im Alltag einer Kleinstadt wie Norton zu dieser Zeit 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 Doppelteebuchführung und trennte HABEN und SOLL auf zwei Bücher. Transaktionen zwischen lokalen Geschäftspartner\*innen und seinen Kund\*innen wurden chronologisch im *Daybooks* erfasst, jeweils mit einem Verweis in der linken Spalte auf die Seite im Hauptbuch, auf der der Geschäftsmann die laufende Schuld des Kunden und die Zeiten, zu denen die Schuld beglichen wurde, erfasst

---

<sup>97</sup>CLIFFORD et al.: MEDEA (Modeling semantically Enhanced Digital Edition of Accounts) as Historical Method, S.5.

<sup>98</sup>Wheaton Family Papers, <https://digitalrepository.wheatoncollege.edu/handle/11040/7928>, 19.05.2019.

<sup>99</sup>Vgl. TOMASEK Kathryn: The Wheaton College Digital History Project: Undergraduate Research in a Local Collection, <https://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.

<sup>100</sup>TOMASEK/BAUMAN: *Journal of the Text Encoding Initiative* 2013.

hat.<sup>101</sup> Folgende Abbildung 5 zeigt Seite 100 und 101 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 "*Burrell 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, wo Information zum Status der Transaktion vorzufinden sind und zeigt somit die Doppeltebuchführung. 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. Jeder hier angeführte Person steht als in einem Geschäftsverhältnis mit Wheaton. jeweils darunter werden die einzelnen Transaktionen gelistet. In den 4 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 folgenden Transaktionen einem neuen Datum zuordnen. So steht "12", "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, für "Debitor" und "Creditor" steht. Dies markiert eine Transaktion mit diesem Partner, dass sie auf die HABEN bzw. die SOLL Seite gebucht wird.

Die vollständige Transkription der ersten 8 Einträge der Seite 100. Diese zeigen die oben beschriebenen Strukturen.

---

<sup>101</sup> TOMASEK/BAUMAN: Journal of the Text Encoding Initiative 2013, S.7-9.

Norton Tuesday Sept 6 1831		Norton Nov. 28 <sup>th</sup> 1831.	
393	Wheaton Wheeler To 1/4 Burrhells Corn	77	David C Bates To 15 ft. Pine planks
395	Samuel Mory By Pension receive	Bill	Summer Knapp To my own this day to draw wood
395	By Pension receive To cash of at Smithy	397	Richard J. Sweet To 2 ft 1/2 L. Pl. 1st. Brown
395	Thompson Sniff To cash of at Smithy	Bill	Summer Knapp To my own this day 3/
395	Thompson Sniff To cash of	Bill	David C Bates By 1 gall L. Wine
398	Oct 8 <sup>th</sup> 1831 Timothy Smith To 6 lbs. Lard	395	Thompson Sniff To 1/2 pr. Buckskin clothing
392	" 7 1/4 lbs. do - 10 <sup>th</sup> Oct 10 <sup>th</sup> 1831 Hiram Hodges To 6 lbs. Lard	398	Timothy Smith To 3 lbs. honey
392	to 4 lbs. Butter 13 pr 12 Oct 11 <sup>th</sup> 1831 Timothy Smith to 6 lbs. 4 pr Butter 1/2 pr & Lard	Bill	Oliver Clapp To 1 gallon 8 pr 1/2 3/4 lbs
398	to 14 lbs. new milk to 6 pr 1/2 do.	395	Thompson Sniff To cash for 1/2 lb. 5 <sup>th</sup>
396	Samuel Mory To cash buying from Doherty	Bill	" 1 pr Blk. Panty
392	Oct 15 <sup>th</sup> By work from this day	Bill	Henry Clement To 1/2 pr 8 pr 1/2 at 8 <sup>th</sup>
392	Hiram Hodges To 1/4 lb. Rain	395	Henry Clapp To 1 pr 1/2 pr 1/2 taken without charge
395	Thompson Sniff To cash 20 <sup>th</sup>	395	December 9 <sup>th</sup> 1831 Summer Knapp To my own a large 1/2 day 2/
392	Hiram Hodges To 3 lbs. Butter	395	Samuel Mory To 1 lb. 1/2 of wood
395	Thompson Sniff To cash of	392	Hiram Hodges To 1 pr 1/2 pr 1/2 to 1/2 pr 1/2
77	David C Bates To 50 best Buck	395	Samuel Mory To 1 pr 1/2 of wood
392	Hiram Hodges To 1 lb. fresh Pork	393	Wheaton Wheeler To 1/2 pr 1/2 pr 1/2
392	John Smith By 2 hours work raising Stone	395	Thompson Sniff To cash 54 <sup>th</sup>
392	To 1/2 pr 1/2 pr 1/2 to 1/2 pr 1/2 pr 1/2	395	Wheaton Wheeler By cash of
392	By work making Stone	395	Samuel Mory To 1 pr 1/2 of wood
392	David C Bates To cash 1/2 of	395	Wheaton Wheeler To 1 pr 1/2 of wood

Abbildung 5: Seite 100 des Daybook, 1828-1859a

## 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

TOMASEK verfolgt dabei einen TEI/XML-Ansatz, um drei Ebenen dieser Quelle zu beschreiben. Das Layout, die textuelle Struktur der Quelle und die abstrakte Ebene, inhaltliche Ebene der Transaktionen. Letztere ist nur schwer durch die TEI Markup beschreibbar.

Es erweitert den Bereich der Fragen zu historischen Narrativen und geografischen Informationen. Es ist interessant zu folgen, eine Person oder eine Familie, wie sie im Laufe der Zeit im Tagesbuch erscheinen und ihre Lebensweise rekonstruieren. sozialer Hintergrund für eine historische Erzählung. Gleiches gilt für geografische Informationen. die Möglichkeit, geografische

Beziehungen von Personen oder die Herkunft von Waren zu verfolgen.

Ziel der Masterarbeit soll es nicht sein die Projektinhalte zu dokumentieren, sondern sich mit theoretischen und praktischen Fragestellungen zu formaler Modellen und formaler Methoden in den Geschichtswissenschaften auseinanderzusetzen, wiewohl Quellen, Workflows und Daten aus dem DEPCHA Projekt einfließen sollen.

## 4 Web of Data

Das Web of Data kann als Stack an Standards und Technologien aufgefasst werden. Flexibel und ausdrucksstark auf der einen Seite und mit dem Ziel maschineneververständliche Information anzubieten, scheint es geeignet als Standard um Modelle zu beschreiben und zu verteilen.

### 4.1 Geschichte und Vision des Web of Data

Beim *TED Talk* im Jahre 2009 fordert Tim Burners-Lee das Auditorium auf mit ihm gemeinsam die Worte zu rufen: *”Raw Data Now!”*.<sup>102</sup> Die Vision von Burners-Lee, dem Erfinder des World Wide Web, ist das sogenannte *Semantic Web*:

*”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.”*<sup>103</sup>

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 verknüpft. 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 umsetzen zu können. Oder anders formuliert sollen Maschinen Inhalte im Web soweit verarbeiten können, dass Automatisierung auf Ebene der Bedeutung möglich ist. Ein konkretes Anwendungsszenario, das mit Hilfe des *Web of Data* umgesetzt werden könnte, wären 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 Homöopath ist. Diese Person stellt eine Suchanfrage in einer dafür geeigneten Suchmaschine, bestehend aus 4 Wörtern: ”Ärzte, Homöopathie, Stadt Graz”. Da in einer klassischen Suchmaschine nur das Vorkommen der Wörter berücksichtigt wird, muss die suchende Person sich noch durch die

<sup>102</sup>BERNERS-LEE Tim: The Next Web, [https://www.ted.com/talks/tim\\_bern timers\\_lee\\_on\\_the\\_next\\_web?language=de](https://www.ted.com/talks/tim_bern timers_lee_on_the_next_web?language=de), 05.04.2019.

<sup>103</sup>BERNERS-LEE/HENDLER/LASSILA: Scientific American 284 [2001], S.3.



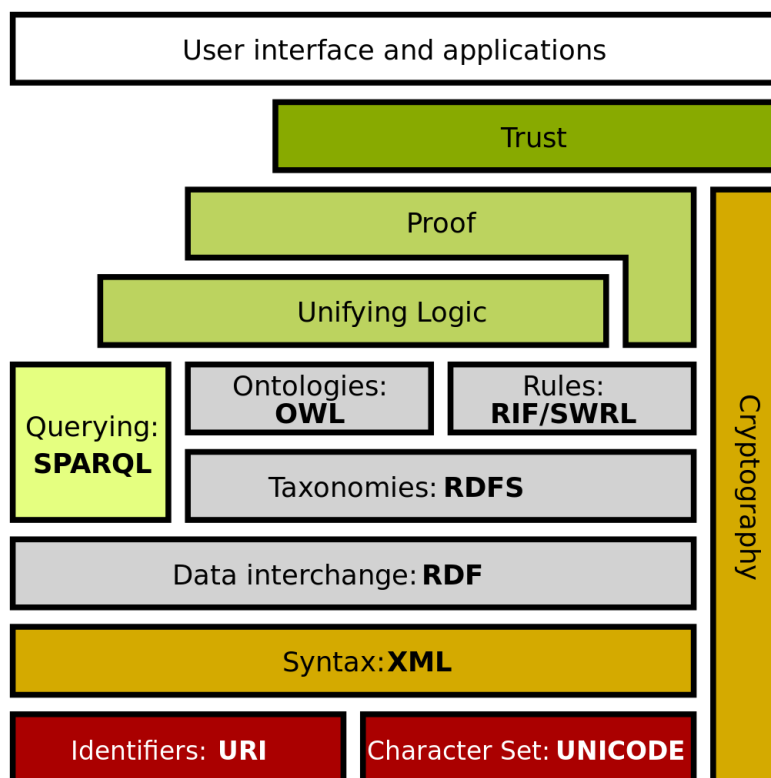
angezeigten Suchergebnisse arbeiten, bis sie den passenden Treffer gefunden hat. Eine Suchmaschine im *Web of Data* ist in der Lage sogenannte Wissensbasen zu befragen, um welche Begriffe es sich hinter den Zeichenketten handelt.<sup>104</sup>

Dabei handelt es sich aber weder um Maschinen die selbstständig lernen, oder gar eine künstliche Intelligenz, sondern um formalisiertes, maschinenlesbares Wissen. Der Wissensbegriff in diesem Zusammenhang entspringt einer informationswissenschaftlichen Perspektive wie bei WERSIG, KUHLEN oder FAUVR-BULLE.

a<sup>105</sup>

## 4.2 Web of Data Stack

Um diese - noch nicht erreichte Vision - in die Tat umzusetzen bedarf es mehrere aufeinander aufbauender technischer Grundlagen, die sich im *Semantic Web Stack* manifestieren. Auf dessen Basis, dargestellt in Abbildung ??, soll in diesem Kapitel die grundlegenden Technologien und Standards des *Web of Data* erörtert werden.



**Abbildung 6:** Visualisierung eines Graphen auf Basis eines RDF-Datensatz.

Das Semantic Web ist nicht erreicht. Zumindest nicht für die Allgemeinheit. Die großen Internetriesen hingegen verfügen über "ihre Semantic Web", in denen sie ihre eigenen Agenten mit

<sup>104</sup> PELLEGRINI, Tassilo/BLUMAUEER, Andreas: Semantic Web. Wege zur vernetzten Wissensgesellschaft. Berlin [ua] Springer 2006, S.1-2.

<sup>105</sup> ebd., S.1-6.



ihrer großen Datenmenge arbeiten lassen.<sup>106</sup>

#### 4.2.1 Resource Description Framework

Das *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.<sup>107</sup> 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 7 zeigt. SCHREIBER und RAIMOND<sup>108</sup> erklären RDF in einem ausführlichen Beispiel an Hand

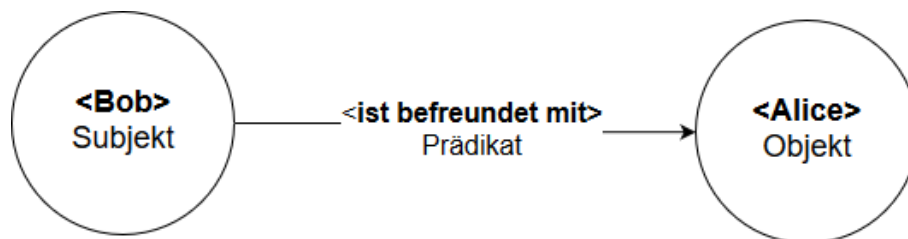


Abbildung 7: Semantic Web Stac

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.*

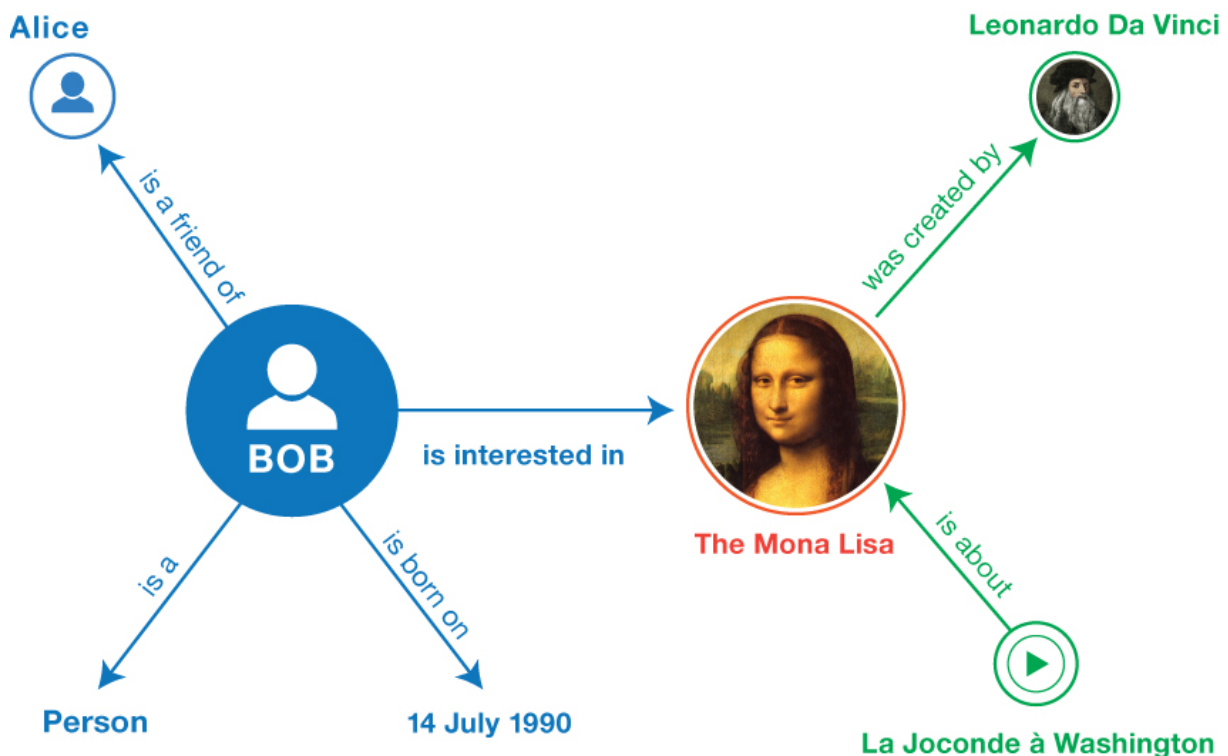
Jede dieser Zeilen steht für ein Triple. *Bob* ist Subjekt in vier der oben genannten Tripeln, *Mona Lisa* tritt zweimal als Objekt und einmal als Subjekt auf. Dies ermöglicht es eine beliebige Menge an Triple zu einem komplexeren Graphen zusammenzusetzen und somit komplexe-

<sup>106</sup><https://twobithistory.org/2018/05/27/semantic-web.html>

<sup>107</sup>Vgl. POWERS, Shelley: Practical RDF: solving problems with the resource description framework. Ö'Reilly Media, Inc.", 2003, S.16-21.

<sup>108</sup>Vgl. SCHREIBER, Guus/RAIMOND, Yves: RDF 1.1 Primer. W3C working group note 2014.

re Sachverhalte beschreiben zu können. Abbildung 8 veranschaulicht das. *Uniform Resource*



**Abbildung 8:** Visualisierung eines Graphen auf Basis eines RDF-Datensatz.

*Identifier*(URI) können in allen drei Positionen eines Triple erscheinen. Somit ist jeder Resource, sowie jeder Beziehung zwischen Ressourcen durch eine URI identifizierbar. URI's sind durch ein erweiterbares Schema definiert, damit Ressourcen im Internet eindeutig adressiert werden können. Um dabei die Einheitlichkeit zu gewährleisten, folgen sie einem vordefinierten Satz von Syntaxregeln, der 5 Komponenten beinhaltet:<sup>109</sup>

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

- **scheme:** Definiert den Kontext und Typ. Bekannte Schemata sind beispielsweise die Webprotokolle *Hyper Text Transfer Protocol* (http) oder das *File Transfer Protocol* (ftp), sowie Notationskonzepte wie *Uniform Resource Name* (URN)urn oder *Digital Object Identifier* (doi).
- **authority:** Verwaltet Instanz in einem bestimmten vom Schema angegebenen Interpretationsraum, wie etwa das *Domain Name System*.
- **path:** Der Pfad enthält – oft hierarchisch organisierte – Angaben, die zusammen mit dem Abfrageteil eine Ressource identifizieren.

<sup>109</sup>Vgl. BERNERS-LEE, Tim/FIELDING, Roy/MASINTER, Larry: Uniform resource identifier (URI): Generic syntax. 2004 – Technischer Bericht (URL: <http://www.rfc-editor.org/info/rfc3986>).

- **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, abgerufen werden
- **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.

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.<sup>110</sup> 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 ;
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> .

```

#### 4.2.2 Resource Description Framework Schema (RDFs)

Im vorhergehenden Beispiel wurden *<Bob>* und *<Alice>* der Klasse *<Person>* zugeordnet. 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)** eine semantische Erweiterung für RDF. Dies umfasst die Möglichkeit Klassen und Relationen, so genannte 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-

<sup>110</sup>Vgl. POWERS: Practical RDF: solving problems with the resource description framework, S.21-22.

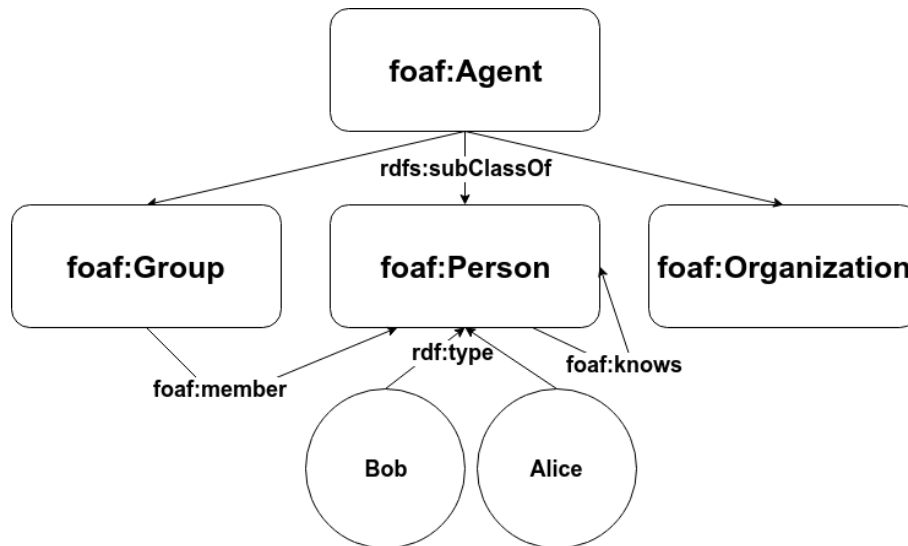
Vokabular,<sup>111</sup> das als RDFs umgesetzt ist, so kann man feststellen, dass *foaf:Person* eine Unterklasse von *foaf:Agent* ist. Weitere Unterklassen davon sind *foaf:Group* und *foaf:Organization*. Alle drei Unterklassen haben bestimmte Eigenschaft gemeinsam: sie setzen Handlungen in der Welt. 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. 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 Lables 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 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.<sup>112</sup> Folgender 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 Propertie dar.

<sup>111</sup>FOAF-Specification, <http://xmlns.com/foaf/spec/>, 29.05.2019.

<sup>112</sup>BRICKLEY, Dan/GUHA, Ramanathan V/MCBRIDE, Brian: RDF Schema 1.1. W3C recommendation, 25 2014.



**Abbildung 9:** RDFs-Beispiel auf Basis des FOAF-Vokabulars

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 .

```

ToDo Hitzler<sup>113</sup>

### 4.2.3 Taxonomien: Simple Knowledge Organisation (SKOS)

### 4.2.4 Abfragesprache: SPARQL

Wie man in der Welt der relationalen Datenbanken mit der Abfragesprache SQL Datenbankabfragen formulieren kann, so kann man mit *SPARQL Protocol And RDF Query Language* RDF

<sup>113</sup>HITZLER, Pascal et al.: Semantic Web: Grundlagen. Springer-Verlag, 2007.

Daten bzw. Triple in Graphdatenbanken abfragen.<sup>114</sup>

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* Propertie verfügen. Das Ergebnis wird nach den Personen und Namen gruppiert, also es werden keine Dupletten von URI's zurückgegeben. Als Ausgabe erfolgt eine Tabelle (bzw. eine 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 } GROUP BY ?person ?name
```

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 SPARQL1.1 Version 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 User über das Web Datenbestände abfragen und mit diesen Arbeiten.<sup>115</sup> Ein bekanntes Beispiel dafür ist der Endpoint von Wikidata.

## 4.3 Ontologien

TODO

formationsmodelle umfassen die Verwendung einer expliziten Sprache zur Informationsdarstellung.<sup>116</sup>

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

<sup>114</sup>GROUP, W3C SPARQL working et al.: SPARQL 1.1 Overview. W3C Recommendation. W3C 2013.

<sup>115</sup>DUCHARME, Bob: Learning SPARQL: querying and updating with SPARQL 1.1. O'Reilly Media, Inc.", 2013, S.1-45.

<sup>116</sup>KOBLER: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung.

die Frage nach welchen allgemeinen Kriterien Objekte im Verhältnis zu ontologischen Begriffen wie Sein, Aktualität, Universalie, Exemplifikation, Sachverhalt oder Individuum stehen.<sup>117</sup> 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 nennt man Ontologie.

Der Begriff wird in zwei Disziplinen mit jeweils unterschiedlichen Fokus verwendet. Dennoch sehe ich Gemeinsamkeiten. 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 informationswissenschaftlichen Dimension des Ontologie-Begriffs und seiner Nutzung in den digitalen Geisteswissenschaften diskutiert und der Frage nachgehen, 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

**Wissen** ist eine systeminterne Repräsentation vorliegender Erfahrungen eines Menschen zu einem bestimmten Zeitpunkt, die einem zu überprüfenden Anspruch auf Gültigkeit ausgesetzt sein muss. Als solches prägt Wissen das Handeln und Denken eines Menschen auf den unterschiedlichsten Ebenen und dient zur Lösung von Problemen. Das jeweils aktuelle Wissen bildet einen kontextuellen Rahmen, in dem ankommende und bestehende Information interpretiert und zu neuen Erfahrungen verarbeitet werden.<sup>118</sup>

Diese Definition von Wissen – eine stärker informationswissenschaftliche – hat seinen, neben vielen Definitionen in anderen Fachbereichen, legitimen Ursprung. Unterschiedlichen Disziplinen haben andere Fragestellungen und benötigen dafür ein anderes theoretisches Gerüst. Ein Wissensbegriff in der Philosophie, beispielsweise, sollte viel weiter gefasst sein, als ein Wissensbegriff in der Informationswissenschaft, dessen Aufgabe darin besteht als Hilfsmittel in der Entwicklung und Umsetzung von Informationssystemen zu fungieren.

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*

---

<sup>117</sup>MEIXNER, Uwe: Von der Wissenschaft der Ontologie. Logos (neue Folge), 1 1994.

<sup>118</sup>FAVRE-BULLE: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation.

als Netzwerk von Begriffen und ihren Relationen' <sup>119</sup> beschreibt. Semantische Netze können einen Kompromiss zwischen menschenverständlicher Repräsentation einer Domäne und der formalen Verarbeitbarkeit durch eine Maschine darstellen. <sup>120</sup> Das ist dadurch gegeben, dass die Struktur des Graphen (=Netz), sich einfach in Rechnern als Matrizen abbilden lässt.

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). <sup>121</sup> 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 *SSäugetier* verbunden ist. Die Individuals (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 *SSäugetier* hat, es ausgeschlossen ist, dass es eine zweite *ist ein* Verbindung gibt, die auf eine andere Klasse wie etwa "Vögel" referenziert.

Der Begriff der Ontologie terminologisch unscharf verwendet. <sup>122</sup> Die Unterschiede sind klein, aber 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"* <sup>123</sup>

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. <sup>124</sup> 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. <sup>125</sup> BORST erweitert GRUBERS Definition um '*formal specification of a shared conceptualization*'. <sup>126</sup> "Formal" ergänzt dabei

<sup>119</sup> STUCKENSCHMIDT: Ontologien: Konzepte, Technologien und Anwendungen.

<sup>120</sup> REICHENBERGER, Klaus: Grundlagen semantischer Netze. In Kompendium semantische Netze. Springer, 2010.

<sup>121</sup> 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.

<sup>122</sup> Vgl. GRUBER, Thomas: A translation approach to portable ontology specifications. Knowledge acquisition, 5 1993, Nr. 2, S.1.

<sup>123</sup> HOEKSTRA, Rinke: Ontology Representation Design Patterns and Ontologies that Make Sense. In Proceedings of the 2009 conference on Ontology Representation: Design Patterns and Ontologies that Make Sense. Ios Press 2009, S.69.

<sup>124</sup> Vgl. THALLER: Historical Social Research/Historische Sozialforschung. Supplement 2017.

<sup>125</sup> SURE, York/STUDER, Rudi: Methodology, tools & case studies for ontology based knowledge management. 2003.

<sup>126</sup> BORST, Willem Nico: Construction of engineering ontologies for knowledge sharing and reuse. 1997.



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 darf nicht im Stillen von einer Person alleine entwickelt werden, sondern sollte 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 dieser Punkte 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.<sup>127</sup> WELLER spricht von einer formalen und schematischen Darstellung einer Wissensdomäne auf Basis definierter Regeln und Vokabulars.<sup>128</sup>

Zusammengefasst kann man sagen, dass sich mittels Ontologien komplexere Sachverhalte so darstellen lassen, dass Mensch und Maschinen in der Lage sind Strukturen, die durch eine Ontologie definierte und standardisierte 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.

Der Ontology Editor Protégé erlaubt es, eine Ontologie und die darin enthaltenen Daten (Individuals) einem Reasoning - dem Abarbeiten aller Vorhanden Regeln in einer Ontologie auf Basis einer deskriptiven Logik - zu unterziehen. Für solche Zwecke gibt es natürlich auch API's und Bibliotheken in Programmiersprachen.<sup>129</sup> 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.<sup>130</sup> Die Überprüfung strukturierter Daten mittels logischen Schlussfolgerungen kann dazu dienen, größere Datenmengen auf ihre Konsistenz und

<sup>127</sup> LINCKELS, Serge/MEINEL, Christoph: E-librarian service: user-friendly semantic search in digital libraries. Springer Science & Business Media, 2011.

<sup>128</sup> 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.

<sup>129</sup> MUSEN, Mark A: The protégé project: a look back and a look forward. AI matters, 1 2015, Nr. 4.

<sup>130</sup> DENTLER, Kathrin et al.: Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web, 2 2011, Nr. 2.

somit auch auf ihre Qualität hin zu prüfen, da logische Inkonsistenzen als Fehlermeldung angezeigt werden.

huhu Ontologie zitieren<sup>131</sup>

## **5 Anwendung formaler Methoden auf Basis einer Ontologie**

### **5.1 Die Bookkeeping-Ontology als konzeptuelles Model für digitale Editionen von historischen Rechnungsbücher**

MEDEA-Paper: A CIDOC-CRM Based Ontology of Economic Activities All these XML formats represent parts of what humans did when they exchanged goods and services and recorded these economic activities in written form. Each method has its advantages. But still it seems that behind all these forms is a common model which could be represented conceptually in the following way: There are agents agreeing on a transfer of something measurable and executing this transfer. Usually the transfer is mutual, i.e. at least two transfers aggregate to a transaction, in which usually money is exchanged for commodities, services or rights. Additionally one or the other party creates a documentation of this transfer, which can be an artifact like a tally, but usually is a written accounting document. It is important to notice that the financial document is not necessarily representing the transfer or the transaction itself; a record can even be fictitious. It also adds a level of interpretation: The transactions are organized in the document following a specific scheme meaningful to the agent commissioning the record, the “categories”, “rubrics”, or “accounts”, which can be abstracted to dimensions; often the accountant aggregates the individual transactions into totals, balancing the financial state or demonstrating profits and losses; and some transactions are not recorded at all. While the historical process of economic transactions can be modelled as seen above, the source material thus has a complex relationship to these historical events we assume to have happened. As in all manuscript analysis, a page is written by one or more hands (writers) in one or more sessions. One specific record can be found on (parts of) one or more pages. So we have an activity performed by one or more actors at a period of time, which may have been discontinuous, which resulted in a record. The traces of this activity can be found in the document and can be established as such by competent readers, as we have seen in the examples above. Each record represents a claim about something that happened in the real world, an economic transaction of some sort. This is another activity, connected to but distinct from the record creation event. In an event-based modelling of historical accounts these two activities will form the core of a network of information, and information—about actors, time, location, the historical account document, and other events connected to the transaction such as goods, money, currency information—become nodes attached to this core. Formalizing the conceptual model

---

<sup>131</sup>Vgl JANNIDIS/KOEHLE/REHBEIN: Digital humanities: eine Einführung, S.162-178.

resulting from these considerations would allow us to map the formalizations described above for their procession as complete set. This goal seems reasonable as of course the records of an economic activity in the ledger of one company should find their counterpart in the ledger of its exchange partner. A meta search on digital representations of various Hanse records could therefore allow extracting networks and flows of trade activities. The research on long term price series or their regional diversion would profit from a common access to the records. The recent solutions historians at the University of Tokyo (Kokaze et al, 2016) have found to represent chronology in historical accounts confirm this: They study the relationship between agents mentioned in the accounts of the British Lay-Osborn Flotilla and Chinese administration from the 1860s in time. The conceptual model covers completely, what a TEI based transactionography (hsfr) and XBRL encoding describe: The `hfrs:transaction/medea:transaction` consists of `hfrs:transfers`, each of them involving `medea:agents/@hfrs:fra` and `@hfrs:til`. The `hfrs:transfer` is defined as ‘A transfer of something of value’, which is a `tei:measure/medea:measurable`. The `medea:documentation` of the transfer is encoded with `hfrs:attested` as a reference to a TEI text structure. The `hfrs:listAccount` and `hfrs:account` express the interpretative organization of transactions in `medea:schemes` and `medea:dimensions`. `hfrs:listTransaction` and `hfrs:transferGrp` are generic grouping functionalities which are reflected in the `medea:dimensions` of documentation and the `medea:total` and `medea:balance` concepts. The relationship between the suggested conceptual model and the abstract level in XBRL is even more obvious: The XBRL ‘financial facts’ are a superclass to `medea:transfer`, `medea:transaction`, or `medea:total`. They are recorded in ‘instances’ (`medea:document`) and the ‘taxonomies’ describe their interpretative dimensions (`medea:dimension`, `medea:schema`). The currently most adaptable formal representation for this kind of data is the Resource Description Framework (RDF) of the W3C. Its graph-based model is flexible but structured enough to express complex data structures. The basic conceptualization similar to simple natural language sentences as subject-predicate-object triples makes it fairly easy to understand. RDFs and OWL<sup>7</sup> offer rich possibilities for formal modelling. A preliminary RDFs schema for the conceptual model has therefore been created and published on github (<https://github.com/GVogeler/bookkeeping>). The CIDOC-CRM, as an event-based ontology, is a good structure for encoding such information, as shown through the use of the standard for several types of cultural-historical evidence. As a matter of fact, it serves well to cover the main parts of the conceptual model described above, with its two event types: 1. The creation of the accounting document 2. The transfer as the recorded economic activity While the first is clearly a subclass of the `E65_Creation` class, the latter can be considered as a subtype of `E7_Activity` and of `E28_Conceptual_Object`. Both are connected by the document written by the accountant and attesting a transfer. The example from Kokaze et al. (2016) demonstrates that the general properties of CIDOC-CRM events have practical research relevance. Given the specific need in modelling these documents and the activities documented in them, some extensions to CIDOC-CRM should be made. This is also in line with what we have seen in other areas including archaeological excavations<sup>8</sup> and buildings,<sup>9</sup> and scientific observations<sup>10</sup> esta-

blishing a CIDOC-CRM based family of standards. In the development of this CIDOC-CRM based extended model it will be an important question what the target for the modelling is, that is, what time periods, geographical areas, and specific document types will function as evidence for the modelling. This will determine how specific the model can be: the larger the scope, the looser and more flexible the model must be. The two event types described above have different relationships to truth. As for the records in the historical account we know they exist and were written, at least as long as the original document exists. What is questionable is who wrote it, when, and also the level of unity, that is, what makes one record separate from other records. For the economic transaction documented in the record the question of existence is somewhat different than from the record. We only know there is a claim for an economic transaction, not that any transaction actually took place, for which one often needs additional evidence, knowledge, and analysis to determine. This creates an interesting question of modelling principles: should one model the worldview expressed by the documents or should one express the worldview seen as most likely to be correct by the historians (Ciula, Spence, Vieira 2008)? The choice of modelling standards should not determine that choice. However, as CIDOC-CRM is a tool for information integration it is more common to use it for models aiming towards historical truthfulness. That does not mean that the worldview expressed in the documents should be hidden, just that the activities expressed in the model will be the ones the historians believe to have taken place. Modelling and encoding the transactions recorded in historical accounting documents as stand-off reference between textual evidence and description of the economic and social ‘facts’ even makes possible multiple interpretations of the textual evidence, which would lead to considering the addition of a factoid-based approach (Bradley and Pasin 2015) to the model.

**Conclusions** The current results of the work done by the MEDEA collaborative are therefore fivefold: 1. The textual genre of accounts shows a great variety in form in its history with a substantial common ground in concepts and human activities represented in these texts. 2. These concepts and human activities are of high interest in the humanities and historical social science research. 3. Based on the work of the DH community in the TEI and of contemporary business reporting standardization activities in XBRL, it is possible to develop a conceptual model for the documentary representation of human activities in the field of economics. 4. This model can be expressed as a graph structure, and it can be integrated into the larger conceptual framework of the CIDOC-CRM family of standards. 5. It is worthwhile to continue the development of the formal model and make suggestions about how to integrate the formal model into encoding practice and data publication of historical accounting records. Certainly, the solutions discussed here are only a few of all the possible ways that web technologies can enhance research with historical accounting records. The MEDEA scholars look forward to suggestions for modifications of the ontology, additions of other possible methods for encoding economic facts represented in accounting documents, more efficient setups for execution of projects, and a thorough scrutiny of the conceptual consistency of our proposal. The MEDEA scholars will continue their work on the ontology and will try to set up a collection of examples showing how

we think the integration of the representation of content expressed in RDF and scholarly edition of accounting documents can be realized.<sup>132</sup>

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)**<sup>133</sup> umgesetzt wurde. In diesem Projekt wird ein gemeinsamer Publikations-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 eben der Transkription und Edition der Rechnungsbücher, sowie der Verfügugstellung der Digitalisate der Quellen, hochstrukturierte RDF-Daten existieren, die mittels der Bookkeeping-Ontologie formalisiert sind und Referenzen auf Normdaten und LOD-Vokabularien setzen. Mit diesem Zugang soll die Nachprüfbarkeit, die Sicherstellung und die Nachnutzbarkeit der Forschungsdaten und aller darauf aufbauenden quantitativen Methoden und deren Ergebnisse gewährleistet sein. Dabei spielt die Bookkeeping-Ontologie, im THALER'schen Sinne als *knowledge domain* eine zentrale Rolle. In diesem Kapitel soll zuerst auf die digitale Edition historischen Rechnungsbüchern und die dazu nötigen Standards, vor allem die TEI, eingegangen werden. Danach folgt die semantische Anreicherung und die Repräsentation der Inhalte in den Transaktionen, sowie eine detaillierte Beschreibung der domänenspezifischen Bookkeeping-Ontology.

## 5.2 Digitale Edition von Historischen Rechnungsbüchern

### 5.2.1 Digitale Edition

Eine Einführung

### 5.2.2 Bookkeeping-Ontology

Die *Bookkeeping Ontology* ist ein konzeptionelles Modell zur formalen Beschreibung von Transaktionen in historischen Rechnungsunterlagen. Es wurde in einem Ontologie-Engineering-Prozess entwickelt, an dem Historiker\*innen, Softwareentwickler\*innen und digitale Geisteswissenschaftler\*innen beteiligt waren.

Eine Transaktion besteht aus mindestens einem Transfer. Jeder Transfer umfasst den Austausch von Dingen, die quantifizierbar sind, von einem Akteur zum anderen. Solche sogenannte *Measurable* lassen sich in Geldbeträge, sowie wirtschaftliche Güter unterteilen. Geldbeträge sind dadurch gekennzeichnet, dass sie aus eine Zahl und einer Währung bestehen, wie etwa 10 US

---

<sup>132</sup>CLIFFORD et al.: MEDEA (Modeling semantically Enhanced Digital Edition of Accounts) as Historical Method, S.9-12.

<sup>133</sup>[gams.uni-graz.at/depcha](http://gams.uni-graz.at/depcha)

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 oder Kontent sein. Die Ontologie erlaubt es explizit zu formulieren von wem an wen *Measurable* 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 zeitliche, räumliche und inhaltlichen Dimension, also der Zordnung zu einem bestimmten, durch die Forschungsfrage geprägten Kontext, zu beschreiben.

Ein bk:Transfer kann von einem bk:Agent durchgeführt werden, also jemand übernimmt den Transferprozess. Bei der Verbuchung der Transaktion wird eine Transaktion im Sinne der doppelten Buchhaltung auf die Haben oder Soll Seite gebucht.

### 5.2.3 TEI Beispiele DEPCHA

```

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
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 <#pers.2>
35  a bk:Between;
36  bk:name "James Haley".

```

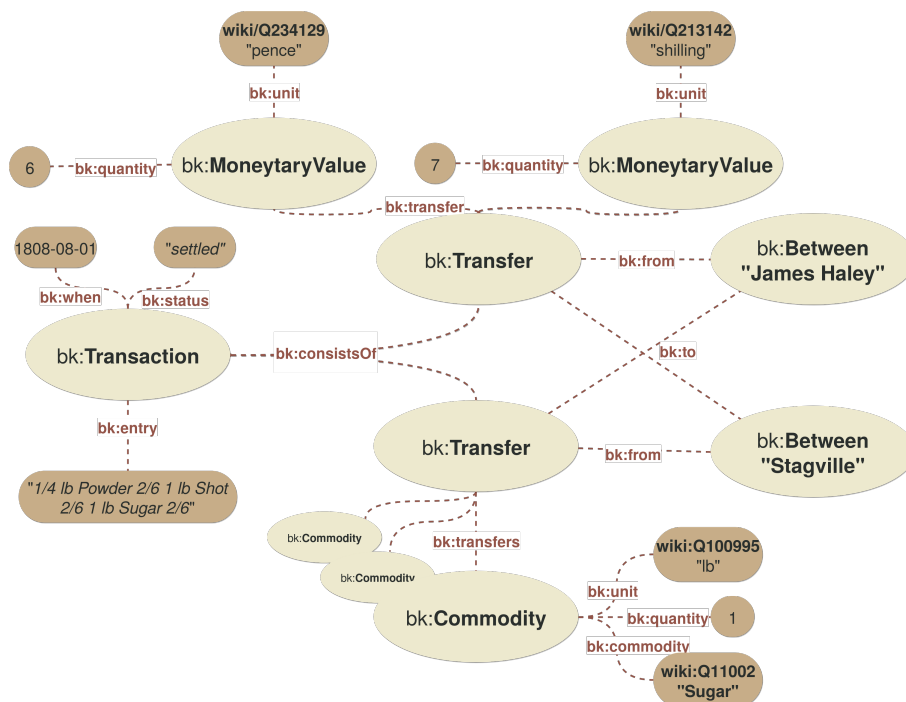


Abbildung 10: Graphische Darstellung des RDF

### 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"*<sup>134</sup>

<sup>134</sup>BURDICK, Anne et al.: Digital\_Humanities. Mit Press, 2012, S.119.

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 auf die Aggregation und Synthese der einzelnen Belge. 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.<sup>135</sup>

Als Visualisierung versteht man – neben dem eigenständigen Fach aus dem Bereich *Human-Computer Interaction* – die Anwendung computerbasierter, interaktiver und visueller Repräsentation von abstrahierten Daten, die eine Hilfestellung bei der inhaltlichen Erfassung eines Themas liefern. Die Daten können sowohl quantitativer, 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) Räumen, wie etwa in einem Koordinatensystem dargestellt. MUNZNER fasst dies zusammen als "*Its scientific visualization when the spatial representation is given. Its Information Visualization when the spatial representation is chosen*".<sup>136</sup> Im Falle von Rechnungsbüchern, und den Geschichtswissenschaften generell, kann man stets von einer Informationsvisualisierung ausgehen. Eine naturgetreue Darstellung historischer Ereignisse ist nie möglich.

FRANK 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 dient. Im politikwissenschaftlichen CEWS-Projekt (Alker et al., 2001) behandelt multiperspektivische Konfliktforschung und leistet seinen Beitrag zur Methoden- und Werkzeugentwicklung. So wird im Projekt Modallogik als formaler Rahmen für die kontrafaktische Analyse von möglichen Konfliktverläufen eingesetzt und die auf Husserl zurückgehende generative Phänomenologie wird zur Analyse der Perspektiven von historischen Akteuren (Konfliktparteien) auf Konfliktereignisse vorgeschlagen. Pellon (2010) kommt aus dem Bereich In

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.

---

<sup>135</sup>FRANK: INF-DH-2018 2018, S.3-5.

<sup>136</sup>MUNZNER, Tamara: Process and pitfalls in writing information visualization research papers. In Information visualization. Springer, 2008, S.134-153).



Sprache ist leichter als Bilder verwenden in einer Kommunikation.

Wissensgewinnung durch Informations-Verbalisierung.

In der Zwischenzeit braucht's neue Aggregatzustände für das Wissen. Datafication, die Dinge werden auch digital erfasst, neue Ansichten auf Daten und auf das Wissen.

Wissensgewinnung durch Informationsvisualisierung

Informationsvisualisierung dann, wenn man ein Thema nicht verbalisieren kann?

Menschliche Wahrnehmung ist multimodal. Es entstehen unterschiedliche Repräsentationen in unserem Bewusstsein von aufgenommenen Wahrnehmungen. Ein produktives Wechselverhältnis herstellen zwischen Sprache und Text.

Close Viewing: Blick auf einen Eintrag Distant Viewing: Blick auf alle Einträge

Können wir da was draus lernen? Aus der Distant Viewing Darstellung?

Übersetzung eines Themas in Daten und die Übersetzung der Daten in Bilder. Wie soll das Interface aussehen am Ende: in Beschreibungen, Skizzen und Worten.

Ein Blick ist selten genug: Multiple Views Das Prinzip der Multiperspektivität.

Information-Seeking-Mantra: Ben Shneiderman (in frühesten Tagen sich bereits damit beschäftigt)

Overview first ein Orientierungswissen anbieten then zoom and filter das interessiert mich v.l. Details on demand damit arbeite ich

Informationsvisualisierung will nicht (nur) hübsch sein, sondern Problemstellungen lösen.

Visuelle Repräsentationen können einerseits auf ein realistisches Bild hinziehen (Dinge nachmodellieren), oder diagrammatisch (in eine andere Sprache überführen, mit einer Syntax und einer Semantik; menschengemacht)

Scientific Visualization und Information Visualization

A Tour through the Visualization Zoo (Jeffrey Heer, Michael Bostock et. al. 2010)

Visualization are also interesting for non-expert users. Because while they are interacting with them, their pattern-mind is learning. Even though they don't know the numbers behind that. [Valdis Krebs, 2011]

Papers:

Choreographien der Existenz: Zur multimodalen Erweiterung biographischer Forschung und Lehre durch Verfahren der visuellen Analyse und Synthese

Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges

Paper: Distributed Cognition as a Theoretical Framework for Information Visualization

“The purpose of information visualization is to amplify cognitive performance, not just to create interesting pictures.”

“Information visualization should do for the mind what automobiles do for the feed. [Card et. al 2008, p.539]

Es ist die Rede von einer Synergie zwischen Kognition und Computer.

Wo sind die Grenzen und Herausforderungen der menschlichen Aufnahmefähigkeit. Things far away, in the past, in the future, very large - very small.

Komplexität: viele Elemente, unterschiedlicher Art haben viele Relationen multidimensional , multivariable und kann auch noch dynamisch sein. Überall brauchen wir Visualisierung nicht, sondern nur da, wo die Komplexität größer ist. Definition des Begriffes der Komplexität, aus der Komplexitätsforschung.

Historische Rechnungsbücher = das Ganze Elemente: Transaktionen Transfer Akteure Güter, Dienstleistungen, als Wirtschaftsgüter Geldbeträge Preise und Steuern Datum

Beziehungen: Von nach Wieviel, Währung, Maßeinheit

Für die Informationsvisualisierung ist es wichtig zuerst ein konzeptuelles Modell zu entwickeln und durchdenken, um daraus die Anwendungsszenarien ableiten zu können.

SciVis - InfoVis

Wenn die räumliche Einordnung gegeben ist, wenn die räumliche Darstellung erst geschaffen wird.

Auch ein Fotoapparat macht SciVis

Worlds, not stories, paper [Moritz Stefaner]

Informationsvisualisierung ist die Welt nur mit anderen Regeln repräsentiert, um abstrakte Daten sichtbar zu machen.

Microsoft timelinestoryteller

Alle Einträge-Datum + Legende = alle Personen.

Treemap

Inspiziert durch tree diagram: also folder mit underfolder;

Es gibt Mengen. Die sind Rechtecke, und je größer die Menge, desto größer die Rechtecke. Und alle darin befindlichen Untermengen (Säugetiere -> Hund, -> Katze) sind im Rechteck und zeigen, wie viel Platz sie innerhalb der übermenge einnehmen.

Hierarchien, Vergleich, Teil eines Ganzen

Rechnungsbuch: Alle Einträge und die Wirtschaftsgüter in einer Hierarchie:

Alle Lebensmittel, darin sind die Erdäpfel, mais Alle Textilien Produkte: Kleidung Alle Dienstleistungen - Arbeit,

Graphs, Maps, Trees: Franco Moretti

### 5.3.1 Statistische Auswertung

## 6 Textfragmente

Das *Data for History Consortium*<sup>137</sup> geht einen vergleichbaren Weg und versucht ein gemeinsames Set an Methoden im *Web of Data* zu entwickeln, um Daten in den Geschichtswissenschaften zu modellieren, verknüpfen und auszutauschen.

hehe<sup>138</sup> huhu<sup>139</sup>

Kapitel in dem digitale Edition, TEI und How to Bookkeep behandelt wird. Und die Besonderheiten editorische Arbeit mit Rechnungsbüchern.

TOMASEK und BAUMAN beschreiben ein Modell eines interpretativen Markups, um Beziehungen zwischen Individuen, Geld- Güter- und Dienstleistungstransfer, die Doppeleintrag-Buchhaltung umfassenm auszuzeichnen. Die Auszeichnung basiert auf den ausdrucksstarken Richtlinien der TEI.<sup>140 141</sup>

## 7 Zusammenfassung

---

<sup>137</sup>BERETTA, Francesco/BRUSEKER, George: The dataforhistory. org project: a proposal. In Workshop on the creation of an international Data for History consortium. 2017.

<sup>138</sup>Vgl. SAHLE: Digitale Editionsformen: Textbegriffe und Recodierung.

<sup>139</sup>Vgl. JANNIDIS/KOHLE/REHBEIN: Digital humanities: eine Einführung, S.234-252.

<sup>140</sup>Vgl. TOMASEK/BAUMAN: Journal of the Text Encoding Initiative 2013, S.1-2, <http://journals.openedition.org/jtei/895>, 08.03.2018.

<sup>141</sup>KOBLER: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung, S.41-44.

# 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, 368–389
- ANDORFER, Peter:** Forschungsdaten in den (digitalen) Geisteswissenschaften. Versuch einer Konkretisierung.(DARIAH-DE Working Papers 14). Göttingen: DARIAH-DE. Verfügbar unter urn: nbn: de: gbv, 2015 [〈URL: http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-14.pdf〉](http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-14.pdf)
- BAUER, Florian/KALTENBÖCK, Martin:** Linked open data: The essentials. Edition mono/monochrom 2011
- BERETTA, Francesco/BRUSEKER, George:** The dataforhistory. org project: a proposal. In Workshop on the creation of an international Data for History consortium. 2017
- BERNERS-LEE, Tim/FIELDING, Roy/MASINTER, Larry:** Uniform resource identifier (URI): Generic syntax. 2004 – Technischer Bericht [〈URL: http://www.rfc-editor.org/info/rfc3986〉](http://www.rfc-editor.org/info/rfc3986)
- BERNERS-LEE, Tim/HENDLER, James/LASSILA, Ora:** The semantic web. *Scientific American*, 284 2001, Nr. 5, 34–43
- BERNSTEIN, Abraham/HENDLER, James/NOY, Natalya:** A New Look at the Semantic Web. *Commun. ACM*, 59 2016, Nr. 9, 35–37 [〈URL: http://doi.acm.org/10.1145/2890489〉](http://doi.acm.org/10.1145/2890489)
- BORST, Willem Nico:** Construction of engineering ontologies for knowledge sharing and reuse. 1997
- BRICKLEY, Dan/GUHA, Ramanathan V/MCBRIDE, Brian:** RDF Schema 1.1. W3C recommendation, 25 2014, 2004–2014
- BRUCH, Julia:** Die Kunst, Daten in Informationen umzuwandeln. 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, 14–44
- BURDICK, Anne et al.:** *Digital Humanities*. Mit Press, 2012
- 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.:** MEDEA (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, 71–87
- DUCHARME, Bob:** Learning SPARQL: querying and updating with SPARQL 1.1. Ö'Reilly Media, Inc.", 2013
- ENGEL, Phillip et al.:** Extensible Business Reporting Language (XBRL) 2.1. XBRL International 2003
- FAVRE-BULLE, Bernard:** Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001
- 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
- GLEBA, Gudrun/PETERSEN, Niels:** Wirtschafts-und Rechnungsbücher des Mittelalters und der Frühen Neuzeit-Formen und Methoden der Rechnungslegung: Städte, Klöster und Kaufleute. Universitätsverlag 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
- GRAHAM, Shawn/MILLIGAN, Ian/WEINGART, Scott:** Exploring big historical data: The historian's macroscope. World Scientific Publishing Company, 2015
- GROUP, W3C SPARQL working et al.:** SPARQL 1.1 Overview. W3C Recommendation. W3C 2013

- GRUBER, Thomas:** A translation approach to portable ontology specifications. Knowledge acquisition, 5 1993, Nr. 2, 199–220
- HILTMANN, Torsten:** Forschungsdaten in der (digitalen) Geschichtswissenschaft. Warum sie wichtig sind und wir gemeinsame Standards brauchen. 2018 (URL: <https://digigw.hypotheses.org/2622>)
- HITZLER, Pascal et al.:** Semantic Web: Grundlagen. Springer-Verlag, 2007
- HOEKSTRA, Rinke:** Ontology Representation Design Patterns and Ontologies that Make Sense. In Proceedings of the 2009 conference on Ontology Representation: Design Patterns and Ontologies that Make Sense. Ios Press 2009, 1–236
- JANNIDIS, Fotis/KOHLE, Hubertus/REHBEIN, Malte:** Digital humanities: eine Einführung. Springer-Verlag, 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, 3
- JORDAN, Stefan:** Theorien und Methoden der Geschichtswissenschaft. UTB GmbH, 2018
- KELLY, Richard:** Practical knowledge engineering. Elsevier, 2016
- KINDLING, Maxi et al.:** Forschungsdatenmanagement an Hochschulen: Das Beispiel der Humboldt-Universität zu Berlin. LIBREAS. Library Ideas, 2013, Nr. 23 (URL: <https://libreas.eu/ausgabe23/07kindling>)
- KIRN, Paul:** Einführung in die Geschichtswissenschaft. Berlin, Boston, 2015
- 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, 7–27
- LATOUR, Bruno:** Zirkulierende Referenz. Bodienstichproben aus dem Urwald am Amazonas. 2016

- LINCKELS, Serge/MEINEL, Christoph:** E-librarian service: user-friendly semantic search in digital libraries. Springer Science & Business Media, 2011
- MAGERSKI, Christine et al.:** Wie schreibt man Geschichte? Zagreber Germanistische Beiträge, 2009, Nr. 18, 3–19
- MANN, Golo:** Plädoyer für die historische Erzählung. In **Kocka, Jürgen/Nipperdey, Thomas (Hrsg.):** Theorie und Erzählung in der Geschichte. München, 1979, 40–56
- MEIXNER, Uwe:** Von der Wissenschaft der Ontologie. Logos (neue Folge), 1 1994, 375–399
- MUNZNER, Tamara:** Process and pitfalls in writing information visualization research papers. In Information visualization. Springer, 2008, 134–153
- MUSEN, Mark A:** The protégé project: a look back and a look forward. AI matters, 1 2015, Nr. 4, 4–12
- NEHER, Günther/RITSCHEL, Bernd:** Semantische Vernetzung von Forschungsdaten. In **Stephan Büttner, Hans-Christoph Hobohm, Lars Müller (Hrsg.):** Handbuch Forschungsdatenmanagement. Band 6, 2011, 169–190
- 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. Berlin [ua] Springer 2006
- PERSTLING, Matthias:** Multimediale Dokumentation und Edition mehrschichtiger Texte : das steirisch-landesfürstliche Marchfutterurbar von 1414/1426. Dissertation, 2013
- 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. 2017
- POLLIN, Christopher/VOGELER, Georg:** Semantically Enriched Historical Data. Drawing on the Example of the Digital Edition of the *Ürfehdebucher* der Stadt Basel". In WHiSe@ISWC. 2017, 27–32
- POWERS, Shelley:** Practical RDF: solving problems with the resource description framework. Ö'Reilly Media, Inc.", 2003
- REICHE, Ruth/BECKER, Rainer/BENDER, Michael:** Verfahren der Digital Humanities in den Geistes-und Kulturwissenschaften. Niedersächsische Staats-und Universitätsbibliothek Göttingen, 2014

- REICHENBERGER, Klaus:** Grundlagen semantischer Netze. In Kompendium semantische Netze. Springer, 2010, 3–19
- 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
- SAHLE, Patrick:** Digitale Editionsformen: Textbegriffe und Recodierung. Band 3, BoD–Books on Demand, 2013
- SCHMIEDER, Felicitas/SOKOLL, Thomas:** Theorien und Methoden der Geschichtswissenschaft. Springer-Verlag, 2009
- SCHREIBER, Guus/RAIMOND, Yves:** RDF 1.1 Primer. W3C working group note 2014
- STACHOWIAK, Herbert:** Allgemeine modelltheorie. 1973
- STUCKENSCHMIDT, Heiner:** Ontologien: Konzepte, Technologien und Anwendungen. Springer-Verlag, 2009
- SURE, York/STUDER, Rudi:** Methodology, tools & case studies for ontology based knowledge management. 2003
- SWARTZ, Aaron:** Aaron Swartz’s A Programmable Web: An Unfinished Work. *Synthesis lectures on the semantic web: Theory and Technology*, 3 2013, Nr. 2, 1–64
- THALLER, Manfred:** Historical Information Science: Is there such a Thing? New Comments on an old Idea [1993]. *Historical Social Research/Historische Sozialforschung. Supplement*, 2017, 260–286
- THALLER, Manfred:** The Need for Standards: Data Modelling and Exchange [1991]. *Historical Social Research/Historische Sozialforschung. Supplement*, 2017, 203–220
- THALLER, Manfred:** Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte ,unscharfer ‘Systeme [1984]. *Historical Social Research/Historische Sozialforschung. Supplement*, 2017, 138–159
- TOMASEK, Kathryn/BAUMAN, Syd:** Encoding financial records for historical research. *Journal of the Text Encoding Initiative* 2013, Nr. 6
- VASOLD, Gunter:** Das Itinerar Erzbischof Konrads IV. von Salzburg: 1291-1312; computergetstützte 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: http://www.zfdg.de/sb001\\_007〉](http://www.zfdg.de/sb001_007)
- VOGELER, Georg:** The 'assertive edition'. *International Journal of Digital Humanities*, 2019, 1–14
- 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, 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.207–218
- WERSIG, Gernot:** Information-Komunikation-Dokumentation. Band 5, Verlag Dokumentation, 1971
- WINTERGRÜN, Dirk:** Netzwerkanalysen und semantische Datenmodellierung als heuristische Instrumente für die historische Forschung. Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2019
- WOLF, Willi:** Qualitative versus quantitative Forschung. *Bilanz qualitativer Forschung*, 1 1995, 309–329

# Abbildungsverzeichnis

1	Nichteindeutigkeit nicht-formaler Modelle, KÖNIG Bettina : Vorlesung "Modellierungsmethoden der Informatik", <a href="http://www.ti.inf.uni-due.de/fileadmin/public/teaching/mod/slides/ws201112/einfuehrung.pdf">http://www.ti.inf.uni-due.de/fileadmin/public/teaching/mod/slides/ws201112/einfuehrung.pdf</a> , 10.06.2019 .	16
2	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 . . . . .	18
3	Datenpyramide geisteswissenschaftlicher Forschungsdaten im institutionellen Kontext, ANDORFER, Peter: Forschungsdaten in den (digitalen) Geisteswissenschaften: Versuch einer Konkretisierung, 2015, S.14 . . . . .	19
4	Portrait von Laban Morey Wheaton, Vgl. 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. . . . .	24
5	Seite 100 des <i>Daybook</i> , <a href="http://hdl.handle.net/11040/17982">http://hdl.handle.net/11040/17982</a> , 1828-1859	27
6	Visualisierung eines Graphen auf Basis eines RDF-Datensatz, <a href="https://www.w3.org/TR/rdf11-primer/">https://www.w3.org/TR/rdf11-primer/</a> , 10.04.2019. . . . .	30
7	Semantic Web Stac . . . . .	31
8	Visualisierung eines Graphen auf Basis eines RDF-Datensatz, <a href="https://www.w3.org/TR/rdf11-primer/">https://www.w3.org/TR/rdf11-primer/</a> , 10.04.2019. . . . .	32
9	RDFs-Beispiel auf Basis des FOAF-Vokabulars, eigene Darstellung. . . . .	35
10	Graphische Darstellung des RDF, eigene Darstellung, 01.06.2019. . . . .	45