

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

Christopher Pollin

1. Juni 2019, Graz

Inhaltsverzeichnis

1	Einleitung	1
2	Formale Methoden und Modelle in den Geschichtswissenschaften	3
2.1	Theorien und Methoden der Geschichtswissenschaft	3
2.2	Die historische Entwicklung formaler Methoden: "Traditionalisten" vs. "Quantifizierer"	5
2.3	Modellbildung in den Geschichtswissenschaften	5
2.3.1	Modellbegriff	6
2.3.2	Was ist Informationsmodellierung	8
2.3.3	System	8
2.4	Forschungsdaten in den Geschichtswissenschaften	8
2.5	Modelle in den Geschichtswissenschaften	9
3	Web of Data	10
3.1	Geschichte und Vision des Web of Data	10
3.2	Web of Data Stack	11
3.2.1	Resource Description Framework	11
3.2.2	Resource Description Framework Schema (RDFs)	14
3.2.3	Taxonomien: Simple Knowledge Organisation (SKOS)	16
3.2.4	Abfragesprache: SPARQL	16
3.3	Ontologien	17
3.3.1	Vom Wissen, über das Semantische Netz zur Ontologie	18
3.3.2	Ontology Engineering	20
3.3.3	Reasoning	20
3.3.4	Linked Open Data	20

4	Historische Rechnungsbücher als Quelle	20
4.1	DEPCHA	21
4.2	The George Washington Financial Papers	22
4.3	The Wheaton Accounts	22
5	Digitale Edition von historischen Rechnungsbücher	27
5.1	Digitale Edition	27
6	Ein Informationsmodell, formales Modell, konezptuelles Modell für Rechnugsbücher: die Bookkeeping Ontology	27
7	Linked Open Data und Geschichte	28
8	Zusammenfassung	28

1 Einleitung

Seit dem Zeitpunkt an dem es für Menschen den Bedarf gibt sich an Ereignisse oder Erkenntnisse, die für sie wichtig waren, zu erinnern, existieren auch Formen und Möglichkeiten Wissen darüber zu repräsentieren. Ein natürliches Beispiel dafür findet sich bereits in den ersten Hochkulturen als Menschen dokumentierte, dass jemand bei jemand anderen in der Schuld steht. Bereits um 3000 vor Christus lassen sich auf Tontafel Schuld und Kredit finden: *"Alok schuldet dem Sumar 10 Eimer Korn"*. Wir können also feststellen, dass die Dokumentation ökonomischer Abhängigkeiten einer der ersten strukturierten Formen waren, um Wissen, wer wem etwas schuldet, festzuhalten und nieder zu schreiben.[ToDo Dan MCCreary]

Für Historiker*innen liefert uns diese Zeile einen Indiz auf ein Ereignis in der Vergangenheit, das, wenn es in einen bestimmten historischen Kontext gestellt wird, für Forschungsfragen in den Geschichtswissenschaften relevant ist. Innerhalb der schriftlichen Überlieferung - der historischen Quelle - stecken semantische Strukturen, die dazu dienen können, ein Ereignis aus der Vergangenheit rekonstruieren zu können.

Die Beschreibung der 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" - mit einer gewissen Datengrundlage zu untermauern.

Ziel dieser Arbeit soll eine theoretisch und praktische Auseinandersetzung mit den Herausforderungen und Möglichkeiten, die formalen, und somit auch digitale, Methoden und Modellen in der Geschichtswissenschaft mit sich bringen. Der praktische Anteil liegt dabei bei der Beschreibung der semantischen Anreicherung von digital edierten historischen Rechnungsbüchern in einem konkreten Projektkontext. Dieser Projektkontext – Digital Edition Publishing Cooperative for Historical Accounts (DEPCHA) – umfasst eine Kooperation zur Entwicklung eines Publikationshub von semantisch angereicherten, digitalen Editionen historischer Rechnungsbücher. Im Zuge dieser Masterarbeit soll zur Veranschaulichung zwei Editionsprojekte aus diesem Projektkontext, die *George Washington's Financial Papers*, die *Wheaton Family Papers* und die *Stagville Accounts* als Beispiele dienen.

Am Anfang der Arbeit steht neben einer allgemeinen theoretischen Diskussion zur Anwendung empirischer, formaler Methoden in den Geschichtswissenschaften, in denen unter anderen Manfred THALLER^{1,2} oder XXX viel beigetragen haben.

Im XXX Kapitel wird auf eine, Technologie Stack, das *Web of Data* (aka Semantic Web) eingegangen, das die Grundlage zur formalen Beschreibung von Information im *World Wide Web*

¹THALLER, Manfred: Digital Humanities als Wissenschaft. In Digital Humanities. Springer, 2017.

²THALLER, Manfred: Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte ,unscharfer 'Systeme [1984]. Historical Social Research/Historische Sozialforschung. Supplement, 2017.

liefert. Dafür ist eine konzeptionelle^{3,4} und technische⁵ Erörterung dieses Themenbereiches notwendig. Ein notwendiger Schwerpunkt liegt dabei auf Wissensmodellierung,⁶ Ontologien⁷ und *Linked Open Data*^{8,9} liegt, sowie einer kritischen Auseinandersetzung damit. *Web of Data*,¹⁰ das *Ontology Engineering*¹¹ und das Reasoning, das automatisierte logische Schlussfolgern,¹² spielen dabei eine hervorgehobene Rolle.

Das XXX Kapitel versteht sich als Quellenstudie und versucht den Quellentypus historischer Rechnungsbücher herauszuarbeiten. Dabei wird auf die Projektkontexte und Forschungsfragen im Kontext des Projektes DEPCHA eingegangen, eines durch die *Andrew W. Mellon Foundation* geförderte und durch das *Wheaton College Massachusetts* koordinierte Kooperation des Zentrums für Informationsmodellierung und Partnern aus den USA.

Darauf aufbauend wird die digitale Edition allgemeiner behandelt, und auf digitale Editionen historischer Rechnungsbüchern am Beispiel der angeführten Projektkontexte und deren Umsetzung eingegangen. Der dabei verwendete Standard ist die Text Encoding Initiative (TEI) auf dessen Grundlage die angeführten Editionen basieren.

Die digitale Edition in dieser Arbeit versteht sich als geschichtswissenschaftliche, quellenorientierte und inhaltsbezogene Edition, in der nicht jedes textuelle Phänomen ediert werden muss, sondern die semantische Struktur der Quellen im Vordergrund stehen. Zur formalen Beschreibung semantischer Strukturen kann das *Web of Data* herangezogen werden. Dies soll am Beispiel des DEPCHA Projektkontext und der *Bookkeeping Ontologie*, einem konzeptuellen Modell zur Beschreibung von Transaktionsprozessen in historischen Rechnungsunterlagen, durchgeführt werden.

³BERNERS-LEE, Tim/HENDLER, James/LASSILA, Ora: The semantic web. *Scientific American*, 284 2001, Nr. 5.

⁴CARDOSO, Jorge: The semantic web vision: Where are we? *IEEE Intelligent systems*, 22 2007, Nr. 5.

⁵BERNSTEIN, 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>).

⁶KELLY, Richard: *Practical knowledge engineering*. Elsevier, 2016.

⁷STUCKENSCHMIDT, Heiner: *Ontologien: Konzepte, Technologien und Anwendungen*. Springer-Verlag, 2009.

⁸RIETVELD, Laurens et al.: *Linked Data-as-a-Service: The Semantic Web Redeployed*. In GANDON, Fabien et al. (Hrsg.): *The Semantic Web. Latest Advances and New Domains*. Cham: Springer International Publishing, 2015.

⁹BAUER, Florian/KALTENBÖCK, Martin: *Linked open data: The essentials*. Edition mono/monochrom 2011.

¹⁰SWARTZ, Aaron: *Aaron Swartz's A Programmable Web: An Unfinished Work*. Synthesis lectures on the semantic web: Theory and Technology, 3 2013, Nr. 2.

¹¹HITZLER, Pascal et al.: *Ontology Engineering with Ontology Design Patterns: Foundations and Applications*. IOS Press, 2016.

¹²BURSZTYN, D. et al.: Reasoning on web data: Algorithms and performance. In *2015 IEEE 31st International Conference on Data Engineering*. April 2015.

2 Formale Methoden und Modelle in den Geschichtswissenschaften

Am ersten August des Jahres 1808 hat ein gewisser James Haley 1/4 viertel Pfund Pulver, 1 Pfund Kugeln und einen 1 Pfund Zucker zum Preis von 2 Schilling und 6 Pence im Laden der *Stagville Plantage* in *North Carolina* käuflich erworben. Solche Information über die Vergangenheit lassen 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 die weiter fachwissenschaftliche Arbeit der Historiker*innen ruhen kann.

Damit diese Datengrundlage ausreichend groß ist und Aussagen nicht nur auf Ausschnitte des Quellenkorpus reduziert sind, ist es notwendig formale Verfahren zu Erschließung und Analyse zu verwenden. Auf historische Rechnungsbücher bezogen, könnte man so Berechnungen anstellen, wie sich etwa der Preis einer Ware über eine bestimmte Zeitspanne entwickelt hat. Setzt man diese Erkenntnisse in einen größerer historischen Zusammenhang, kann die 'klassische' Arbeit von Historiker*innen beginnen.

Im Mittelpunkt dieses Kapitels steht eine Diskussion über Theorien und Methoden der Geschichtswissenschaften und einer darauf aufbauenden Auseinandersetzung und der historischen Entwicklung der Anwendung formaler Verfahren innerhalb des Faches beschrieben werden. Aus dieser Diskussion heraus sollen die theoretischen Grundlagen der historischen Fachinformatik erörtert werden, wobei ein Fokus auf Modelltheorie und Modellbildung in den Geschichtswissenschaften und auf historische Forschungsdaten gelegt wird. Weiters werden informationswissenschaftliche Grundlagen in diese Diskussion einbezogen.

2.1 Theorien und Methoden der 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"*¹³

Führte MANN im Jahr 1979 an, um auszudrücken, dass durch die Einführung neuartiger theoretischer Ansätze die Hauptaufgabe der Geschichtswissenschaften, das Erzählen von Geschichten, aufgeweicht wird. Diese neuartigen Ansätze etablierten sich ausgehend von Historikern wie Mommsen, Wehler oder Kocka in den 1960er Jahren, die *"eine Reform der Geschichtswissenschaft jenseits des Historismus"* propagierten. Sie versuchten die Theoriebildung, wie in den systematischen Nachbardisziplinen, wie etwa der Soziologie, in den Geschichtswissenschaften zu etablieren.

¹³TODO Seitenanzahl und .In MANN, Golo: Plädoyer für die historische Erzählung. Theorie und Erzählung in der Geschichte/Hrsg. von J. Kocka, T. Nipperdey. München 1979, S.40–56.

Wheeler Golo mann Auseinandersetzung: theorieorientierte, moderne Geschichtswissenschaft vs. erzählende Art¹⁴

”*Erzählende Geschichte bedarf heute einer Theorie der Erzählung*” beschreibt SOKOLL die heutige Situation, in der historische Forschung ohne theoretische Fragestellung nicht mehr möglich ist. 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]¹⁵ – mit ihrem Ziel die historischen Quellen aufzubereiten erstrecken sich von der Paläografie über die Diplomatik und Aktenkunde bis hin zu Genealogie. Die Historische Fachinformatik versteht sich auch als Hilfswissenschaft innerhalb der Geschichtswissenschaften und beschäftigt sich mit der Anwendung und Reflexion von formalen Verfahren in diesem Fach.¹⁶

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 ein Thema 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 – als Lehre des deutenden Verstehens – der sogenannten Hermeneutik. Aus diesem Verständnis heraus erwuchs ’die’ Theorie 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, interpretierender und quantitativer Methode. Erstere getragen von einem hermeneutischen, zweitere von einem empirischen Verständnis.¹⁷

¹⁴MAGERSKI, Christine et al.: Wie schreibt man Geschichte? Zagreber Germanistische Beiträge, 2009, Nr. 18, S.1-6.

¹⁵VON BRANDT, Ahasver: Werkzeug des Historikers: eine Einführung in die historischen Hilfswissenschaften. Band 33, W. Kohlhammer Verlag, 2007.

¹⁶Was ist HFI?, Kurzbeschreibung des Faches und seiner Positionierung <http://hfi.uni-graz.at/basisinformation/was-ist-hfi/>, 23.05.2019.

¹⁷Vgl, TODO SOKOLL, Thomas: Einführung: Theorien und Methoden – an praktischen Beispielen. Grundlagen der Geschichtswissenschaft, S. 1–5.

2.2 Die historische Entwicklung formaler Methoden: "Traditionalisten" vs. "Quantifizierer"

In den 1980er Jahren standen sich im 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, übernahmen die "Quantifizierer" formale Methoden, wie etwa statistische Verfahren, aus den Sozialwissenschaften, um sie auf Quellenkorpora anzuwenden und die daraus gewonnenen empirischen Fakten für die Interpretation zu nutzen.¹⁸ Ergebnis dieser Auseinandersetzung war es, dass bei der Anwendung formaler Methoden besonders auf die Nachvollziehbarkeit geachtet werden muss, damit nicht Dinge, die nicht empirisch beweisbar sind, auch nicht so missverstanden werden können. Aus diesem Grund muss die Quelle ohne jegliche Vorannahmen zur Verfügung gestellt werden und einsehbar sein und das angewandte formale Modell bzw. die formale Methode in ihrer Gänze offen gelegt werden. Gerade der letztere Aspekt definiert den eigentlichen Kern der Geschichtswissenschaften: die Interpretationen der Vergangenheit. Diese Interpretation sollte für andere nachvollziehbar sein und wissenschaftlichen Kriterien standhalten. Es wird empfohlen die grundlegenden Annahmen und Definitionen, die in der Interpretation von Quellen verwendet werden in einer gemeinsamen *knowledge domain*, wie THALLER es nannte, zu formalisieren.¹⁹

Zentrales Konzept in einem "digitalen" Verständnis formaler Arbeitstechniken in den Geschichtswissenschaften in diesem Zusammenhang sind das *Web of Data* (aka. Semantic Web) und das *Linked Open Data* Paradigma. Diese Technologien ermöglichen es konzeptuelle Modelle in den Geisteswissenschaften mensch- und maschinenlesbar zusammen mit der Quelle nachvollziehbar und nach nutzbar zu machen. Diese Technologien erleichtern den Austausch konzeptueller Modelle über das Web können sie so leichter verteilt und nach genutzt werden.

2.3 Modellbildung in den Geschichtswissenschaften

*"Im wissenschaftlichen wie außerwissenschaftlichen Sprachgebrauch hat gegenwärtig der Modellbegriff zunehmend Relevanz erlangt. Bei zahlreichen passenden – leider auch unpassenden – Gelegenheiten ist von "Modellen" die Rede."*²⁰

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 Wirklich-*

¹⁸JARAUSCH, Konrad Hugo/ARMINGER, Gerhard/THALLER, Manfred: Quantitative Methoden in der Geschichtswissenschaft. Wissenschaftliche Buchgesellschaft, 1985, S.XX-XX.

¹⁹THALLER, Manfred: Historical Information Science: Is there such a Thing? New Comments on an old Idea [1993]. Historical Social Research/Historische Sozialforschung. Supplement, 2017, S.XX-XX.

²⁰STACHOWIAK, Herbert: Allgemeine modelltheorie. 1973, S.1.

keit”.²¹

2.3.1 Modellbegriff

Auch KOBLEER führt an, dass der Modellbegriff in fast jeder wissenschaftlichen Disziplin vorzufinden ist, 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).²² 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 der menschlichen Kultur. Der Fokus in dieser Arbeit betrachtet nun nur externalisierte, niedergeschriebene Modelle, die sich theoretisch in Datenmodelle überführen lassen. In den unterschiedlichen 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 sind Modelle oft beschreibender Funktion. In der Wissenschaftstheorie werden Theorien in Form von mathematischen Modellen dargestellt. In der Wirtschaftsinformatik werden Modelle für vielfältige Aufgaben eingesetzt. Zusammengefasst kann man eine relativ große Menge an Modelltypen festmachen.²³

Eine Klassifikation lässt sich im Grad der syntaktischen und semantischen Präzisierung umsetzen. Unterscheiden kann man informelle Modelle ohne eindeutige Beschreibungssyntax, wie etwa natürliche Sprachen, semiformale Modelle mit definierter Syntax, aber weitgehend ohne semantische Konstruktionsregeln, wie etwa ER- oder UML-Modelle, oder formale Modelle, die neben der Syntax auch die Semantik ausreichend beschreiben, wie Modelle auf Grundlage deskriptiver Logiken. Für formale Modelle lassen sich dann Algorithmen definieren, die Daten, die durch ein solches Modell beschrieben sind, automatisch zu prüfen. Im Gegensatz zum Computer kann der Mensch aber auch mit nicht formalen Modellen arbeiten. Es stehen also stark strukturierte, formalisierte Modelle, die algorithmisch genutzt werden können, offenen Modellen gegenüber, 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

²¹ STACHOWIAK: Allgemeine modelltheorie.

²² ebd., S.129.

²³ KOBLEER, Maximilian: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung. Logos Verlag Berlin GmbH, 2010, S.41-44.

gut geeignet sind.²⁴ Ich denke man kann diese genauso auch auf die Digitalen Geisteswissenschaften ummünzen.

huhu²⁵

Unterschieden werden eine normative, deskriptive explorative Funktion eines Modells. Modellierungen geschieht 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.

Es gibt vier verschiedene Möglichkeiten, Metamodelle zu gestalten:

Das Entity-Relationship-Modell (ER) baut auf den drei Begriffen Entität, Attribut und Relation auf. Die Entitäten (alle möglichen Untersuchungsobjekte) haben Eigenschaften („Attribute“) und stehen zu anderen Entitäten in Beziehung („Relation“) Für die Beziehungen spielen die Kardinalitäten/Multiplizitäten eine wichtige Rolle, das heißt, dass ein Objekt unterschiedlich viele Beziehungen haben kann. Die Umsetzung eines ER erfolgt in der Informatik normalerweise in Tabellen, in denen die Beziehungen durch Verweise ausgedrückt werden, und mehrfach vorkommende Daten verringert werden.

Das Graphenmodell baut nur auf zwei Begriffen auf, nämlich der Entity und der Property (Eigenschaft) Rollen; die Eigenschaft ist eine Referenz auf eine andere Entität. Also es geht um „Knoten“ und Verbindungen zwischen Knoten („Kanten“). Entitäten können literal (also eine Zahl oder ein Text) oder eine Ressource (abstrakter Bezeichner für eine Entität) sein. Solche „Graphen“ können sowohl gerichtet als auch ungerichtet sein.

Klassen abstrahieren Entitäten. Das bedeutet, dass jedes Mitglied einer Klasse mindestens eine gemeinsame Eigenschaft mit den anderen Mitgliedern einer Klasse besitzt. Ohne dieser Eigenschaft ist es kein Mitglied. Um das zu bestimmen, muss die Eigenschaft, welche eine Klasse ausmacht, immer genau definiert werden.

Man kann seine Objekte aber auch als Hierarchien modellieren, bei denen Objekte und Eigenschaften Teil eines anderen Objektes sind. Wenn man hier etwas löscht, dann werden auch alle Unterelemente gelöscht.

²⁴KOBLER: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung, S.44-47.

²⁵Vgl JANNIDIS, Fotis/KOHLER, Hubertus/REHBEIN, Malte: Digital humanities: eine Einführung. Springer-Verlag, 2017, S.99-108.

2.3.2 Was ist Informationsmodellierung

2.3.3 System

Der Begriff System bezeichnet einen explizit von seiner Umgebung abgegrenzt Gedankenbereich. Ein System besteht dabei aus einer Menge von Elementen, die miteinander in Beziehung stehen und über spezifische Eigenschaften verfügen. Ein Element ist dabei nicht weiter spezifiziert, es handelt sich nur um atomare Bestandteile innerhalb eines Systems. Ein System, hat zu einer bestimmten Zeitpunkt einen bestimmten Systemzustand und kann auf Grund des Verhaltens eines Systems eine Änderung (interner externer Ereignisse) beschreiben. Systeme können aus Systemen bestehen.²⁶

2.4 Forschungsdaten in den Geschichtswissen

Unter Daten versteht man ganz allgemein Zeichen, die Informationen darstellen und dem Zweck der weiteren Verarbeitung dienen. Die Informatik fasst den Begriff noch etwas spezifischer und versteht darunter Informationseinheiten in einer formalisierten und damit maschinenlesbaren und -bearbeitbaren Form. Das heißt, Informationen werden durch Daten so repräsentiert, dass sie vom Computer automatisch verarbeitet werden können.

Was diese Daten am Ende repräsentieren, ist für die Definition erst einmal egal, und eigentlich auch für den Computer. Oder, wie es die Informatiker immer wieder betonen: Egal, ob die Daten Bilder, Texte, Filme, Objektbeschreibungen oder was auch immer repräsentieren, am Ende sind alles Daten und können als solche verarbeitet werden.

Nun soll es hier aber um Forschungsdaten gehen. Als Forschungsdaten wiederum werden allgemein alle Daten definiert, die während eines Forschungsprozesses entstehen oder sein Ergebnis sind. Doch, und das ist an dieser Stelle die Frage, lässt sich diese Definition denn so einfach auf den geschichtswissenschaftlichen Forschungsprozess übertragen?

Tatsächlich scheint diese Definition stärker an Fächern orientiert, die auf Grundlage von Experimenten und der dabei gewonnenen Messdaten arbeiten. Historische Arbeiten jedoch beruhen vor allem auf einem hermeneutischen Prozess, der auf der Lektüre und Verarbeitung einer größeren Zahl von Texten, Bildern und anderen Quellen und Informationen aufbaut. Hier stehen vielmehr die Quellen selbst im Vordergrund als die an ihnen vollzogenen "Messungen" und die sich daraus ergebenden "Messdaten" (die es natürlich auch gibt).

Während der Tagung wurde deutlich, dass die Teilnehmer letztlich mit zwei unterschiedlichen Begriffen von "geschichtswissenschaftlichen Forschungsdaten" operierten: einem engeren und einem weiteren Begriff. Während der engere Begriff tatsächlich nur Daten einbezog, die im konkreten Forschungsprozess der einzelnen Projekte entstehen, wie z.B. die Textannotationen oder Metadaten, schloss der weitere Begriff auch die dem historischen Forschungsprozess zugrundeliegenden Quellen als Ganzes mit ein – so wie sie von Bibliotheken, Archiven und Museen

²⁶KOBLE: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung, S.52-53.

wissenschaftlich aufbereitet und damit der potentiellen Nutzung durch die Forschung zur Verfügung gestellt werden.

Dieser breitere Begriff macht dabei zugleich die Komplexität der Diskussion um "historische Forschungsdaten" deutlich. Denn die Feststellung, dass im Sinne eines funktionalen Quellenbegriffs letztlich alles, was über vergangene Kulturen und Gesellschaften Auskunft geben kann, zur Quelle werden kann, gilt schließlich auch für deren digitale Repräsentation. Damit kann letztlich auch jede digitale Repräsentation von Texten, Bildern und Objekten sowie ihrer Präsenz in Zeit und Raum zu einem Forschungsdatum werden. Mit dem Blick auf eine digital organisierte und kommunizierende Gesellschaft heißt das auch, dass am Ende jedes Byte zu einer historischen Quelle und damit zu einem Forschungsdatum werden kann.²⁷

2.5 Modelle in den Geschichtswissenschaften

Das Ziel: Aus den vielen Fäden ein allgemeingültiges Modell für die Repräsentation historischer Informationen in der Informationstechnologie zu machen.

What we need is a discussion of the communalities between the underlying information models and the identification of properties, for which clear conceptual models can be devised. Such clear conceptual models are a prerequisite for technical solutions, which ultimately can enable the exchange of data across the different approaches. Such conceptual models, therefore, are what we need as standards.

Analyzing someone else's dataset is difficult. In fact, it is so difficult that it is often simpler and less costly to create one's own dataset from scratch rather than use one prepared by another researcher from the same underlying documentary sources.

Thaller sagt, dass viel gewi Projekte einfach nur für eine Fragestellung ein System gebaut haben. Es ist verdammt schwer, die Daten aus einem anderen Projekt herauszuholen, zu verstehen, zu nutzen. Die Quelle kann für unterschiedliche Forscher, unterschiedlich interessant sein. Er hebt hervor, dass es nnotwendig ist, dass wir Standardisierungen auf Eben konzeptueller Modelle durchführen.

The approach to standardization proposed by the Text Encoding Initiative (TEI) and developed elsewhere in this volume is I feel, prescriptive.

A standard must accommodate each of these very different aims, but at the same time facilitate automatic conversion of data structured for very different uses. From this perspective, standardization needs to be descriptive rather than prescriptive.²⁸

²⁷Vgl. HILTMANN, Torsten: Forschungsdaten in der (digitalen) Geschichtswissenschaft. Warum sie wichtig sind und wir gemeinsame Standards brauchen. 2018 (URL: <https://digigw.hypotheses.org/2622>).

²⁸THALLER, Manfred: The Need for Standards: Data Modelling and Exchange [1991]. Historical Social Research/Historische Sozialforschung. Supplement, 2017, S.204.

3 Web of Data

3.1 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!"*.²⁹ 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."*³⁰

Im Gegensatz zum klassischen Web, das als ein Web von Dokumenten betrachtet werden kann, versucht das *Web of Data* Daten aus unterschiedlichen Quellen zu integrieren und miteinander zu verknüpfen. Daten sollen so vorliegen, dass nicht nur Menschen diese in neuen Kontexten nutzen können, sondern auch Softwareagenten. Maschinen sollen in der Lage sein selbständig die Struktur von Daten "verstehen" zu können, um bestimmte Aufgaben 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 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.³¹

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.

DATEN-INFORMATION-WISSEN

²⁹BERNERS-LEE Tim: The Next Web, https://www.ted.com/talks/tim_bern timers_lee_on_the_next_web?language=de, 05.04.2019.

³⁰BERNERS-LEE/HENDLER/LASSILA: Scientific American 284 [2001], S.3.

³¹PELLEGRINI, Tassilo/BLUMAUER, Andreas: Semantic Web. Wege zur vernetzten Wissensgesellschaft. Berlin [ua] Springer 2006, S.1-2.

3.2 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.

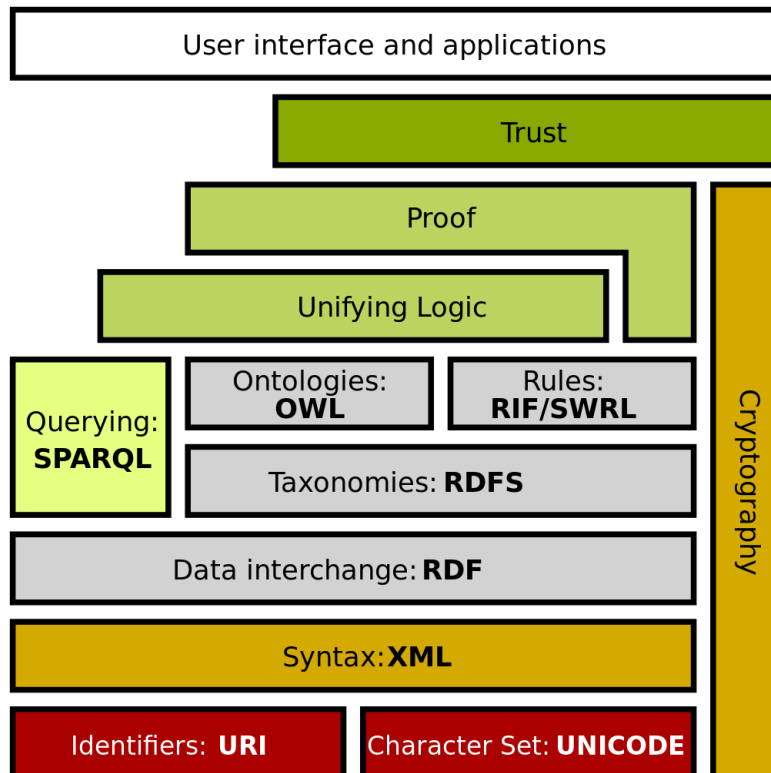


Abbildung 1: Visualisierung eines Graphen auf Basis eines RDF-Datensatz.

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 ihrer großen Datenmenge arbeiten lassen.³³

3.2.1 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

³²PELLEGRINI/BLUMAUER: Wege zur vernetzten Wissensgesellschaft. Berlin [ua] Springer 2006, S.1-6.

³³<https://twobithistory.org/2018/05/27/semantic-web.html>

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.³⁴ In der graphischen Darstellung wird schnell klar, dass es sich beim RDF Datenmodell um einen gerichteten Graphen handelt, der aus Knoten (Subjekt und Objekt), sowie aus Kanten (Prädikat) besteht, wie Abbildung 2 zeigt. SCHREIBER und RAIMOND³⁵ erklären RDF in einem ausführlichen Beispiel an Hand

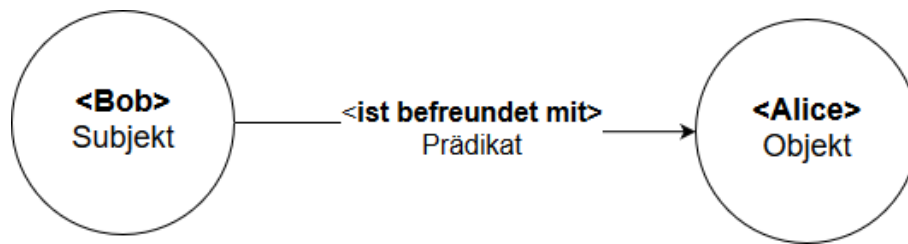


Abbildung 2: Semantic Web Stac

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 Sachverhalte beschreiben zu können. Abbildung 3 veranschaulicht das. **Uniform Resource 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:³⁶

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

³⁴Vgl. POWERS, Shelley: Practical RDF: solving problems with the resource description framework. Ö'Reilly Media, Inc.", 2003, S.16-21.

³⁵Vgl. SCHREIBER, Guus/RAIMOND, Yves: RDF 1.1 Primer. W3C working group note 2014.

³⁶Vgl. BERNERS-LEE, Tim/FIELDING, Roy/MASINTER, Larry: Uniform resource identifier (URI): Generic syntax. 2004 – Technischer Bericht (URL: <http://www.rfc-editor.org/info/rfc3986>).

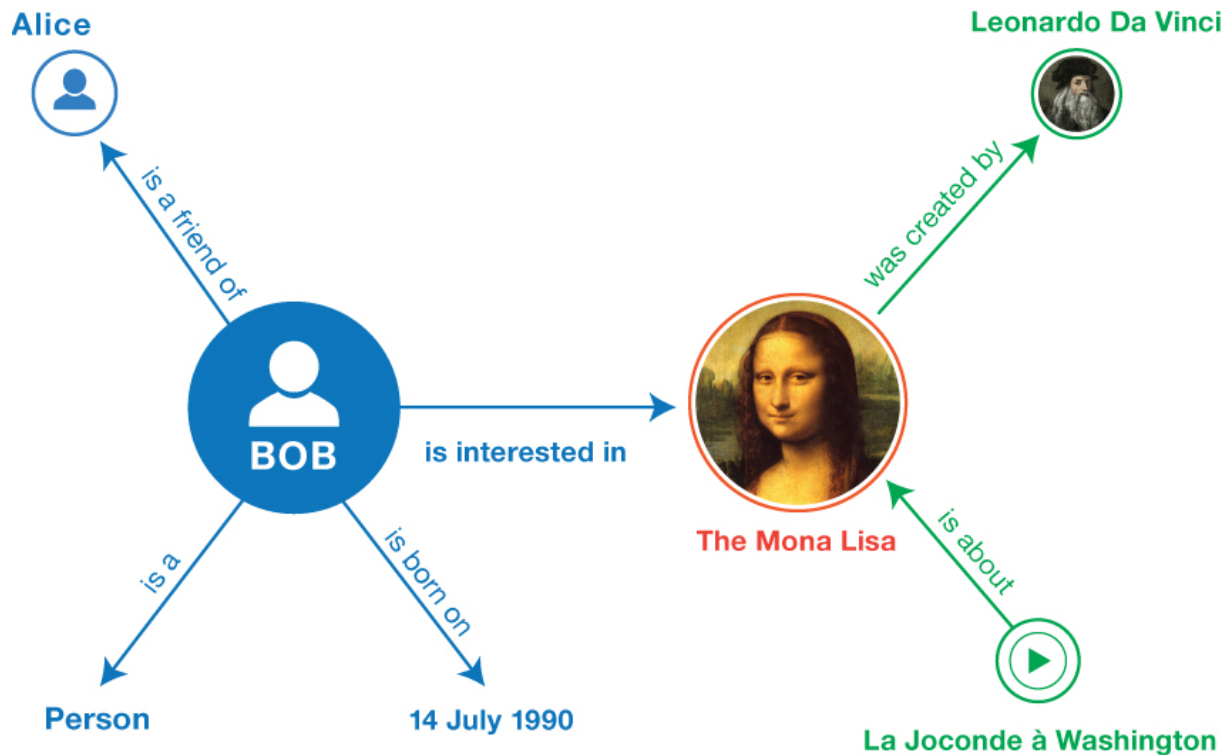


Abbildung 3: Visualisierung eines Graphen auf Basis eines RDF-Datensatz.

- **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.
- **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.³⁷ 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> .
```

3.2.2 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-Vokabular,³⁸ 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

³⁷Vgl. POWERS: Practical RDF: solving problems with the resource description framework, S.21-22.

³⁸FOAF-Specification, <http://xmlns.com/foaf/spec/>, 29.05.2019.

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 Konstrukut *rdfs:seeAlso*, um ausdrücken, dass es unter folgender URL noch weitere Information zu dieser Ressource gibt.³⁹ 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.

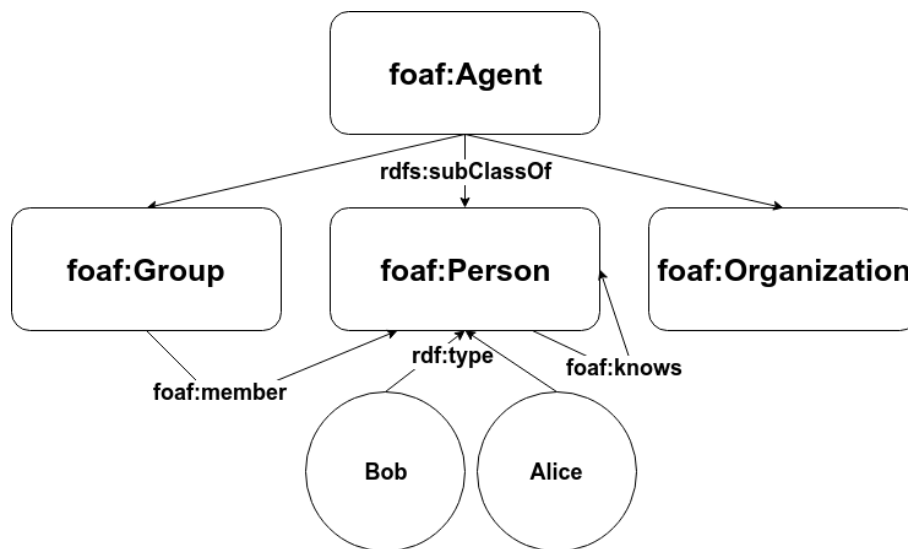


Abbildung 4: RDFs-Beispiel auf Basis des FOAF-Vokabulars

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#> .

³⁹BRICKLEY, Dan/GUHA, Ramanathan V/MCBRIDE, Brian: RDF Schema 1.1. W3C recommendation, 25 2014.

```

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 Hitzler⁴⁰

3.2.3 Taxonomien: Simple Knowledge Organisation (SKOS)

3.2.4 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 Daten bzw. Triple in Graphdatenbanken abfragen.⁴¹

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 Rück-

⁴⁰HITZLER, Pascal et al.: Semantic Web: Grundlagen. Springer-Verlag, 2007.

⁴¹GROUP, W3C SPARQL working et al.: SPARQL 1.1 Overview. W3C Recommendation. W3C 2013.

gabewert 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.⁴² Ein bekanntes Beispiel dafür ist der Endpoint von Wikidata.

3.3 Ontologien

ToDo

formationsmodelle umfassen die Verwendung einer expliziten Sprache zur Informationsdarstellung.⁴³

Der Gegenstandsbereich der **Ontologie als Disziplin in der Philosophie** umfasst alles, das existiert. Das Erkenntnisziel, so MEIXNER, ist auf allgemeiner begrifflicher Ebene zu finden und beschäftigt sich mit der Einteilung des Seins und den Grundstrukturen der Wirklichkeit, sowie der Frage nach dem Wesen der Existenz. Die Ontologie verfolgt nicht das Ziel Erkenntnis über ein Objekt zu erhalten, es beispielsweise zu vermessen oder zu beschreiben, sondern stellt sich die Frage nach welchen allgemeinen Kriterien Objekte im Verhältnis zu ontologischen Begriffen wie Sein, Aktualität, Universalie, Exemplifikation, Sachverhalt oder Individuum stehen.⁴⁴

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.

⁴²?, .

⁴³KOBLER: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung.

⁴⁴MEIXNER, Uwe: Von der Wissenschaft der Ontologie. Logos (neue Folge), 1 1994.

3.3.1 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.⁴⁵

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 als Netzwerk von Begriffen und ihren Relationen'*⁴⁶ beschreibt. Semantische Netze können einen Kompromiss zwischen menschenverständlicher Repräsentation einer Domäne und der formalen Verarbeitbarkeit durch eine Maschine darstellen.⁴⁷ 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).⁴⁸ 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 "Säugetier" 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

⁴⁵FAVRE-BULLE, Bernard: Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001.

⁴⁶STUCKENSCHMIDT: Ontologien: Konzepte, Technologien und Anwendungen.

⁴⁷REICHENBERGER, Klaus: Grundlagen semantischer Netze. In Kompendium semantische Netze. Springer, 2010.

⁴⁸JOOST 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.

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 "Vögelreferenziert.

Der Begriff der Ontologie terminologisch unscharf verwendet.⁴⁹ 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"*⁵⁰

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.⁵¹ 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.⁵² BORST erweitert GRUBERS Definition um '*formal specification of a shared conceptualization*'.⁵³ "Formal" ergänzt dabei die Definition um die Notwendigkeit, dass Ontologien maschinenlesbar sein müssen. Erst diese Eigenschaft hebt sie von anderen Methoden zur Formalisierung von konzeptionellen Datenmodellen hervor. Der Zusatz "*shared*" reflektiert die Tatsache, dass eine Ontologie Wissen erfasst, das durch den Konsens einer Gruppe - z.B. durch einen wissenschaftlichen Diskurs - akzeptiert wird. Eine Ontologie 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

⁴⁹Vgl. GRUBER, Thomas: A translation approach to portable ontology specifications. Knowledge acquisition, 5 1993, Nr. 2, S.1.

⁵⁰HOEKSTRA, 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.

⁵¹Vgl. THALLER: Historical Social Research/Historische Sozialforschung. Supplement 2017.

⁵²SURE, York/STUDER, Rudi: Methodology, tools & case studies for ontology based knowledge management. 2003.

⁵³BORST, Willem Nico: Construction of engineering ontologies for knowledge sharing and reuse. 1997.

Allgemeiner betrachtet definieren LINCKELS & MEINEL eine Ontologie als ein Datenmodell zur Darstellung eines Sets miteinander vernetzter Konzepte innerhalb einer (Fach-)Domäne.⁵⁴ WELLER spricht von einer formalen und schematischen Darstellung einer Wissensdomäne auf Basis definierter Regeln und Vokabulars.⁵⁵

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.⁵⁶ 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.⁵⁷ Die Überprüfung strukturierter Daten mittels logischen Schlussfolgerungen kann dazu dienen, größere Datenmengen auf ihre Konsistenz und somit auch auf ihre Qualität hin zu prüfen, da logische Inkonsistenzen als Fehlermeldung angezeigt werden.

huhu Ontologie zitieren⁵⁸

3.3.2 Ontology Engineering

3.3.3 Reasoning

3.3.4 Linked Open Data

4 Historische Rechnungsbücher als Quelle

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

⁵⁴LINCKELS, Serge/MEINEL, Christoph: E-librarian service: user-friendly semantic search in digital libraries. Springer Science & Business Media, 2011.

⁵⁵WELLER, 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.

⁵⁶MUSEN, Mark A: The protégé project: a look back and a look forward. AI matters, 1 2015, Nr. 4.

⁵⁷DENTLER, Kathrin et al.: Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web, 2 2011, Nr. 2.

⁵⁸Vgl JANNIDIS/KOHLE/REHBEIN: Digital humanities: eine Einführung, S.162-178.

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.⁵⁹

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 1808.⁶⁰

4.1 DEPCHA

Daten - aus unterschiedlichen Formaten - sollen auf einer gemeinsamen Plattform zusammengeführt werden und adäquate Formen des Retrievals, Discoverys und der Visualisierung eröffnen,

⁵⁹TOMASEK, Kathryn/BAUMAN, Syd: Encoding financial records for historical research. *Journal of the Text Encoding Initiative* 2013, Nr. 6, S.2.

⁶⁰ebd., S.3.

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*.

4.2 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]?

4.3 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 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.⁶¹

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 eine Fabrik zur Produktion von Baumwollwatte und Strohhüten⁶²

⁶¹Definition 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.

⁶²TOMASEK/BAUMAN: Journal of the Text Encoding Initiative 2013, S.6.

und so gehörte Laban Wheaton zu einem der wohlhabenden Männer der Stadt. Sogar ein Portrait von ihm ist bei CLARK überliefert.⁶³



Abbildung 5: 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.

Das *Wheaton College Massachusetts* stellt die *Wheaton Family Papers* Dokumente in ihrem digitalen Repositorium online zur Verfügung.⁶⁴

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.⁶⁵ Teile davon wurden im Zuge des DEPCHA Projektes für die Umsetzung eines Prototyps zur Publikation von semantisch angereicherten, digitalen Edition von Rechnungsbüchern verwendet.

⁶³CLARK, 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.

⁶⁴Wheaton Family Papers, <https://digitalrepository.wheatoncollege.edu/handle/11040/7928>, 19.05.2019.

⁶⁵Vgl. 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.

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.⁶⁶

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 hat.⁶⁷ Folgende Abbildung 6 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 Satus der Transaktion vorzufinden sind und zeigt somit die Doppelteebuchführung. 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 Oerson 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 mehrere Transaktionen zu einer Person auch eine Summe.

Im Textfluss lassen sich zwischen den Einträgen entweder einfache Zahlen "12", "24" oder neue Tagesdaten finden, die eben nun alle folgende Transaktionen einem neuem Datum zuordnen. So steht "12", "24" auf diese 100 f+r 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, das sie auf die HABEN bzw. die SOLL Seite gebucht wird.

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

⁶⁶TOMASEK/BAUMAN: Journal of the Text Encoding Initiative 2013.

⁶⁷ebd., S.7-9.

Norton Tuesday Sept 6 1831		Norton Nov. 28 th 1831.	
393	Wheaton Wheeler To 1/2 Burrhells Corn 50	77	David C Bates D To 15 ft. Pine planks 34 45
395	Samuel Mory By Pinner receive 120 50 By Pinner for interest 20 " of same 24 " Cash for 82 00 58 50	Bill	Summer Knapp D To my own this day to draw 50
395	Thompson Sniff D To cash of at Smith 1 00	397	Richard J Sweet D To 2 ft 1/2 L. Pl. 1st. Brown 44
395	Thompson Sniff D To cash 80/ 5 00	Bill	Summer Knapp D To my own this day 3/ 50
398	Oct 8 th 1831 Timothy Smith D To 6 Hrs. Lard 104 " 2 1/4 Hrs. do 104 1 33 Oct 10 th 1831	Bill	David C Bates D By 1 gall L. Wine 8/ 1 33
392	William Hodges Dr to 5 Hrs. Lard 104 to 4 Hrs. Butter 13 for 72 1 18 1 33 Oct 11 th 1831	395	Thompson Sniff D To 1/2 for Buckskin clothing 42
398	Timothy Smith Dr to 6 Hrs. Lard 104 to 14 Hrs. new milk 56 to 6 Hrs. do 20 1 97	398	Timothy Smith D To 3 Hrs. honey 1/ 50
396	Samuel Mory D To cash twenty five Dollars 25 00	Bill	Oliver Clapp D To 1 Gall. 8 Hrs. 8 1/2 Hrs 68
392	Oct 15 By work from this day 2 00	395	Thompson Sniff D To cash for 1/2 Hrs. 100 1 00 " 1/2 for Bk. Party 1 00
392	William Hodges Dr To 1/2 Hrs. Rain 4	Bill	Henry Clement D To 1/2 Hrs. 104 1 14
395	Thompson Sniff D To cash 204 20	395	Henry Clapp D To 1/2 for 1/2 Hrs. taken without 104 " 1/2 Hrs. 104 " 1/2 Hrs. 104
392	William Hodges D To 3 Hrs. Butter 104 54	December 9 th 1831	
395	Thompson Sniff Dr To cash 10 1 00	Bill	Summer Knapp D To my own a large 1/2 day 2/ 33
77	David C Bates D To 50 best Buck 604 60	395	Samuel Whigley D To 1/2 Hrs. 1/ 17
392	William Hodges D To 1/2 Hrs. fresh Pork 104 53	392	William Hodges D To 1/2 Hrs. to 1/2 Hrs. 104 54
395	John Smith D By 2 Hrs. work raising Stone 40	395	Samuel Mory D To 1/2 Hrs. of wood 3 50
395	David C Bates D To 1/2 Hrs. 104 1 00	393	Wheaton Wheeler D To 1/2 Hrs. 104 50
395	Thompson Sniff D To cash 54/ 9 00	395	Thompson Sniff D To cash 54/ 9 00
395	Wheaton Wheeler D By cash 5/ 50	395	Wheaton Wheeler D By cash 5/ 50
395	Samuel Mory D To 1/2 Hrs. 104 3 94	395	Samuel Mory D To 1/2 Hrs. 104 50

Abbildung 6: 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.

5 Digitale Edition von historischen Rechnungsbücher

hehe⁶⁸ huhu⁶⁹

5.1 Digitale Edition

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. ^{70 71}

6 Ein Informationsmodell, formales Modell, konezptuelles Modell für Rechnugsbücher: die Bookkeeping Ontology

Im Zuge des Projektes **Digital Edition Publishing Cooperative for Historical Accounts (DEPCHA)**⁷² wird ein gemeinsamer Publikations-Hub für historische Rechnungsbücher umgesetzt. Im Zentrum steht die Entwicklung und Nutzung einer Ontologie zur Formalisierung und Standardisierung von Buchungstransaktionen in historischen Rechnungsunterlagen, sowie ein *Linked Open Data* Zugang. Der Mehrwert entsteht durch Funktionalitäten des Retrievals, der Visualisierung und Analyse der eingespielten Datenbestände.

Für alle diese Projekte existieren hochstrukturierte RDF-Daten, die jeweils mit einer domänenspezifischen Ontologie beschrieben sind. Diese Ontologie wurde in einem iterativen *Ontology Enginneering*-Prozess mit den FachkollegInnen, basierend auf den fachspezifischen Forschungsfragen, generiert.

⁶⁸Vgl. SAHLE, Patrick: Digitale Editionsformen: Textbegriffe und Recodierung. Band 3, BoD–Books on Demand, 2013.

⁶⁹Vgl JANNIDIS/KOHLE/REHBEIN: Digital humanities: eine Einführung, S.234-252.

⁷⁰Vgl. TOMASEK/BAUMAN: Journal of the Text Encoding Initiative 2013, S.1-2, <http://journals.openedition.org/jtei/895>, 08.03.2018.

⁷¹KOBLER: Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung, S.41-44.

⁷²gams.uni-graz.at/depcha

Das *Data for History Consortium*⁷³ 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.

The common knowledge domain of these documents is formalized in a "bookkeeping" ontology, based on the REA model and compliant with the CIDOC CRM. As a conceptual data model, the ontology is developed in an iterative process. It formalizes the interpretation of transactions of money, commodities and services from one actor to another, and further properties that can be found in historical accounts. The RDF data extracted from the accounts becomes therefore a highly structured and self describing data set, being interoperable and reusable for researchers in diverse fields. The RDF representation can link to URI's of commodities, places, persons or other LOD vocabularies. Additionally the RDF representation contributes to the LOD. Thus, all formal methods applied in the DEPCHA project can be transferred to other data conforming to the proposed ontology and any kind of combined data set.

7 Linked Open Data und Geschichte

8 Zusammenfassung

⁷³BERETTA, Francesco/BRUSEKER, George: The dataforhistory. org project: a proposal. In Workshop on the creation of an international Data for History consortium. 2017.

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
- 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>)
- 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>)
- 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
- BURSZTYN, D. et al.:** Reasoning on web data: Algorithms and performance. In 2015 IEEE 31st International Conference on Data Engineering. April 2015, 1541–1544
- CARDOSO, Jorge:** The semantic web vision: Where are we? *IEEE Intelligent systems*, 22 2007, Nr. 5, 84–88
- 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
- COLE, George S:** A complete Dictionary of Dry Goods and History of Silk, Linen, Wool and other Fibrous Substances. 1892. 2015
- DENTLER, Kathrin et al.:** Comparison of reasoners for large ontologies in the OWL 2 EL profile. *Semantic Web*, 2 2011, Nr. 2, 71–87
- FAVRE-BULLE, Bernard:** Information und Zusammenhang: Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Springer, 2001

- 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.:** Ontology Engineering with Ontology Design Patterns: Foundations and Applications. IOS Press, 2016
- 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
- KELLY, Richard:** Practical knowledge engineering. Elsevier, 2016
- KOBLER, Maximilian:** Qualität von Prozessmodellen: Kennzahlen zur analytischen Qualitätssicherung bei der Prozessmodellierung. Logos Verlag Berlin GmbH, 2010
- 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. Theorie und Erzählung in der Geschichte/Hrsg. von J. Kocka, T. Nipperdey. München 1979
- MEIXNER, Uwe:** Von der Wissenschaft der Ontologie. Logos (neue Folge), 1 1994, 375–399

- MUSEN, Mark A:** The protégé project: a look back and a look forward. *AI matters*, 1 2015, Nr. 4, 4–12
- PELLEGRINI, Tassilo/BLUMAUER, Andreas:** Semantic Web. Wege zur vernetzten Wissensgesellschaft. Berlin [ua] Springer 2006
- POWERS, Shelley:** Practical RDF: solving problems with the resource description framework. "Ö'Reilly Media, Inc.", 2003
- REICHENBERGER, Klaus:** Grundlagen semantischer Netze. In *Kompodium semantische Netze*. Springer, 2010, 3–19
- RIETVELD, Laurens et al.:** Linked Data-as-a-Service: The Semantic Web Redeployed. In **Gandon, Fabien et al. (Hrsg.):** The Semantic Web. Latest Advances and New Domains. Cham: Springer International Publishing, 2015, 471–487
- SAHLE, Patrick:** Digitale Editionsformen: Textbegriffe und Recodierung. Band 3, *BoD–Books on Demand*, 2013
- SCHREIBER, Guus/RAIMOND, Yves:** RDF 1.1 Primer. W3C working group note 2014
- SOKOLL, Thomas:** Einführung: Theorien und Methoden – an praktischen Beispielen. *Grundlagen der Geschichtswissenschaft*
- 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:** Digital Humanities als Wissenschaft. In *Digital Humanities*. Springer, 2017, 13–18
- 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
- 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

Abbildungsverzeichnis

1	Visualisierung eines Graphen auf Basis eines RDF-Datensatz, https://www.w3.org/TR/rdf11-primer/ , 10.04.2019.	11
2	Semantic Web Stac	12
3	Visualisierung eines Graphen auf Basis eines RDF-Datensatz, https://www.w3.org/TR/rdf11-primer/ , 10.04.2019.	13
4	RDFs-Beispiel auf Basis des FOAF-Vokabulars, eigene Darstellung.	15
5	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.	23
6	Seite 100 des <i>Daybook</i> , http://hdl.handle.net/11040/17982 , 1828-1859	25