

Francia – Forschungen zur westeuropäischen Geschichte

Bd. 40

2013

DOI: 10.11588/fr.2013.0.40966

Copyright

Das Digitalisat wird Ihnen von perspectivia.net, der Online-Publikationsplattform der Max Weber Stiftung - Deutsche Geisteswissenschaftliche Institute im Ausland, zur Verfügung gestellt. Bitte beachten Sie, dass das Digitalisat urheberrechtlich geschützt ist. Erlaubt ist aber das Lesen, das Ausdrucken des Textes, das Herunterladen, das Speichern der Daten auf einem eigenen Datenträger soweit die vorgenannten Handlungen ausschließlich zu privaten und nicht-kommerziellen Zwecken erfolgen. Eine darüber hinausgehende unerlaubte Verwendung, Reproduktion oder Weitergabe einzelner Inhalte oder Bilder können sowohl zivil- als auch strafrechtlich verfolgt werden.

GEORG VOGELER

VON DER TERMINOLOGIE ZUR ONTOLOGIE

Das »Vocabulaire international de la diplomatie« als Ressource des Semantic Web

Dieser Beitrag beschreibt den Weg von einer analogen zu einer digitalen Ressource, die Begriffe der Diplomatie darstellt. Es geht nicht um konkrete Urkunden und ihre digitalen Repräsentationen, sondern um die Art und Weise, wie die Diplomaten darüber sprechen¹. Es geht um einen möglichen Weg, wie man das »Vocabulaire international de la diplomatie« (im Folgenden: VID) digitalisieren und dadurch anders nutzbar machen kann. »Digitalisieren« meint hier, es in ein elektronisches Format zu verwandeln, das dem Computer mehr Informationen zur Verfügung stellt als nur die den Text repräsentierende Abfolge von Zeichen. Der Beitrag umfasst deshalb drei Teile: Als erstes ist das »Vocabulaire« selbst vorzustellen und von seiner »Retrodigitalisierung« zu berichten. Der zweite Teil diskutiert die digitalen Formate, in die das VID danach umgesetzt wurde, und skizziert, wie man diese in der Arbeit mit Urkunden nutzen könnte. Abschließend werden Perspektiven aufgezeigt, die eine noch weiter gehende Digitalisierung der Inhalte des VID bieten könnte.

Das »Vocabulaire international de la diplomatie« ist ein Projekt der Commission Internationale de Diplomatie (CID). Die CID hatte sich nach ihrer Gründung vorgenommen, Vorschläge für einheitliche Editionsregeln zu erarbeiten, über die Robert-Henri Bautier 1974 zum ersten Mal auf dem Kolloquium der CID berichtete. 1984 legte die CID eine erweiterte und abschließende Fassung dieser Editionsregeln vor². Noch während der Arbeiten an ihnen nahm die Commission ein »Vocabulaire« in Angriff. Die Arbeiten hatten drei Ziele: Erstens sollten Unklarheiten über den Gebrauch der Begriffe ausgeräumt und diese normalisiert werden. Zweitens sollten durch die Entsprechungen in allen europäischen Wissenschaftssprachen fremdsprachige Texte besser verständlich werden. Drittens sollte durch die Definitionen sichergestellt werden, dass die Arbeiten der Diplomaten durch Dritte richtig genutzt werden können, indem sie Missverständnisse über den Gebrauch der Begriffe ausräumen³. Dazu organisierte die CID zwischen 1978 und 1982 sowie 1988 und 1989 mehrere *Colloques techniques*⁴.

1 Zum aktuellen Stand digitaler Urkundenrepräsentationen vgl. zuletzt meinen Beitrag: Georg VOGELER, Digitale Urkundenbücher. Eine Bestandsaufnahme, in: Archiv für Diplomatik 56 (2010), S. 363–392 und Digitale Urkundenpräsentationen, hg. von Joachim KEMPER, Georg VOGELER, Norderstedt 2011 (Schriften des Instituts für Dokumentologie und Editorik, 5).

2 Normalisation internationale des méthodes de publication des documents latins du Moyen Âge. Colloque de Barcelone, 2–5 octobre 1974, Paris 1977. Erweitert: Commission Internationale de Diplomatie. Normes internationales pour l'édition des documents médiévaux. Einleitung von Robert-Henri BAUTIER, in: Diplomatica et sigillographica. Travaux préliminaires de la CID et de la commission internationale de sigillographie pour une normalisation internationale des éditions de documents et un Vocabulaire international de la diplomatie et de la sigillographie, Zaragoza 1984 (Folia Caesaraugustana, 1), S. 13–64.

3 Carlsruhards BRÜHL, Préface, in: Vocabulaire international de la diplomatie, hg. von Maria MLAGROS CÁRCEL ORTÍ, Valencia 1994, 2. verb. Aufl. 1997 (Col·lecció Oberta), S. 13 und Robert-Henri BAUTIER, Avant-propos, *ibid.*, S. 15–17.

4 Vgl. die Übersicht auf der Webseite der CID: <<http://www.cidipl.eu/congresses.html>>.

Eine erste Vorarbeit zu einer mehrsprachigen Terminologie legten Stefania Ricci und Robert-Henri Bautier schon 1984 vor⁵. 1990 hatten sie ein »Vocabulaire de la sigillographie« erarbeitet⁶. Das Ergebnis der Arbeit am eigentlichen »Vocabulaire de la diplomatie« erschien 1994 zum ersten Mal. Eine zweite verbesserte Auflage kam 1997 auf den Markt⁷. Das »Vocabulaire« enthält 641 Stichwörter mit Definitionen und Verweisen. Zu den französischen Begriffen liefern die Artikel lateinische, deutsche, englische, spanische und italienische Entsprechungen. Ein Register verweist noch zusätzlich vom Katalanischen, Dänischen, Ungarischen, Niederländischen, Portugiesischen, Rumänischen und Tschechischen auf die Definitionen.

Das »Vocabulaire« erhielt nicht nur Lob. Olivier Guyotjeannin wies in seiner Rezension ebenso auf fehlende Begriffe (»Kurzarenga«, »data topica«) hin wie Rolf Große (»Reskript«). Große wunderte sich, dass bestimmte historische Phänomene, wie insbesondere die deutsche Königsurkunde, weitgehend ausgeblendet wurden. Beide waren von den gefundenen Definitionen nicht restlos überzeugt. So fragte sich Guyotjeannin, weshalb das »Vocabulaire« nicht der belgischen Tradition gefolgt sei, von *action juridique* statt von *acte juridique* zu sprechen, was die Verwechslung mit *acte écrit* erschwert hätte. Große kritisierte definitorische Inkonsistenzen zum deutschsprachigen »Lexikon des Mittelalters« beim Stichwort »Landesherrliche Urkunden«, die das VID als öffentliche Urkunden und nicht als Privaturkunden definiert, und die eingeschränkte Definition von »Privileg« als kirchlicher Urkunde, ferner den Verlust einmal gewonnener Trennschärfe in Ahasver von Brandts Unterscheidung zwischen »Vidimus« und »Transumpt«, die das VID nicht kennt⁸.

Was ist das VID also? Auch wenn das VID insbesondere über die Kanzleien viele konkrete historische Aussagen macht, war es kein Ziel, ein europäisches Handbuch der Diplomatie als historische Darstellung vorzulegen. Die Kritik von Guyotjeannin und Große zeigt, dass das VID aber auch keine vollständige Liste der mit Begriffen belegten Konzepte der diplomatischen Forschung ist, wenn es diesem Ziel auch nahe kommt. Obwohl die Commission Internationale de Diplomatie den Text sanktioniert hat, ist es schließlich auch keine autoritative Referenz geworden, unter anderem weil es einzelne Definitionen vorschlägt, die gewichtige Gruppen von Forschern nicht akzeptieren. Wie aber auch die Rezensenten klarstellen, ist es auf jeden Fall ein Hilfsmittel, um sich sicher zwischen den Sprachen zu bewegen.

Diese Funktion war auch der erste Grund dafür, eine digitale Fassung anzufertigen. In Vorbereitung der Tagung »Ein Standard für die Digitalisierung mittelalterlicher Urkunden« am 5./6. April 2004 in München⁹ ist der gedruckte Text mit den klassischen Mitteln der Retrodigitalisierung in eine elektronische Fassung verwandelt worden, d.h. es wurden Scans der Seiten angefertigt und aus den dabei entstandenen digitalen Bildern mit einer Texterkennungssoftware (Abbyy Finereader 7) elektronische Volltexte gewonnen. Der elektronische Text diente den Dolmetschern der Tagung zur Vorbereitung, weshalb in Handarbeit die Stichwörter extrahiert und in Listen übertragen wurden¹⁰. Der elektronische Text lag schließlich in zwei Tabellen vor, eine, die die Nummerierung des VID und die Definition des Begriffs enthält, und

5 Robert-Henri BAUTIER, Travaux préliminaires pour l'établissement du vocabulaire international menés sous l'égide du Comité international de sigillographie, Commission internationale de diplomatie, in: *Diplomatica et sigillographica* (wie Anm. 2), S. 169–214.

6 *Vocabulaire international de la sigillographie*, hg. von Robert-Henri BAUTIER, Stefania RICCI, Rom 1990 (Pubblicazioni degli Archivi di Stato, 3).

7 Siehe oben Anm. 3.

8 Olivier GUYOTJEANNIN, in: *Revue historique* 296 (1996), S. 431 f.; Rolf GROßE, in: *Francia* 23/1 (1996), S. 236 f.

9 Vgl. dazu Georg VOGELER, Ein Standard für die Digitalisierung mittelalterlicher Urkunden. Bericht zum Workshop (München 5./6. April 2004), in: *Archiv für Diplomatik* 50 (2004), S. 23–34.

10 Diese Arbeit hat vorwiegend und unentgeltlich Dagmar Groß übernommen, der dafür noch einmal ausdrücklich zu danken ist.

eine andere, welche die Wörter und die Sprache, aus der sie stammen, auflistet. Diese beiden Tabellen lassen sich leicht in eine relationale Datenbank verwandeln, die das Konzept und seine sprachlichen Repräsentanten als Entitäten kennt und über interne automatische Identifikatoren verbindet. Relationale Datenbanken sind Kern vieler Softwareprodukte, und es lässt sich mit heutigen Programmier-techniken leicht eine Anwendung schreiben, die Funktionalitäten wie eine Suche oder die Extraktion von Korrespondenzlisten für die Dolmetscher ermöglicht. Die von mir gewählte Lösung verwendet MySQL als Datenbankback-end¹¹ und PHP als Programmiersprache¹² für die unter <http://www.cei.lmu.de/VID/> im Web zugängliche Anwendung. Die Datenbank enthält als dritte Komponente noch die Gliederung des VID. Seit Sommer 2011 haben auch der Herausgeber und die Commission Internationale de Diplomatique als Inhaber der intellektuellen Rechte zugestimmt, dass die so entwickelte Online-Ressource öffentlich verfügbar ist. Sie kann als sprachliches Hilfsmittel genutzt werden, indem unter anderem der gesuchte Begriff direkt in der URL angegeben wird (z.B. <http://www.cei.lmu.de/VID/?Aussteller>).

Der Zeitpunkt dieser Retrodigitalisierung ist in mehrfacher Hinsicht von Bedeutung: Zunächst einmal ging es, wie erwähnt, darum, den Dolmetschern eine mehrsprachige Terminologie zur Verfügung zu stellen. Zum anderen war das Ziel des Münchener Treffens dem der CID zu Beginn ihrer Arbeit am VID sehr nahe: Auf der Tagung waren die Bearbeiter derjenigen Projekte vertreten, die sich der XML-Technologie bedienten, um Urkunden online zu publizieren. Auf dem Treffen wollten sie sich über gemeinsame elektronische Editionsprinzipien verständigen, also die Ziele, die sich das CID für die gedruckten Editionen gesetzt hatte, im Bereich der digitalen Techniken realisieren. Eines der Ergebnisse war, dass sich die Gruppe zur Charters Encoding Initiative zusammenschloss, um an einem Vorschlag für eine mögliche XML-Codierung von Urkundenbeschreibungen zu arbeiten¹³.

Die Retrodigitalisierung des VID blieb davon zunächst unabhängig. Die technische Aufbereitung in einer relationalen Datenbank hat jedoch seine intellektuellen Grundstrukturen noch einmal offengelegt: Auch wenn das VID wohl am meisten Verwendung als Wörterbuch findet, ist der Text dennoch als Terminologie konzipiert. Eine Terminologie legt die Bedeutung eines Wortes fest, um wissenschaftlichen Sprachgebrauch abzusichern. Um das zu erreichen, kann sie eine sprachliche Definition liefern; sie kann Begriffe zueinander in Beziehung setzen, in dem sie z.B. festlegt, dass ein Terminus einen anderen einschließt (Hierarchisierung), oder sie kann hinweisend den Bedeutungsgehalt beschreiben, in dem sie Beispiele gibt oder den Bedeutungsgehalt bildlich wiederzugeben versucht. Das VID verwendet die beiden ersten Methoden. So ordnen die Gliederung und der Nummerierungsstil, der Unterordnungen mit an die Nummern angehängten Buchstaben ausdrückt, die Definitionen hierarchisch. Die Definitionen enthalten zusätzlich unspezifische Verweise. Es ist in diesem Zusammenhang interessant, dass das VID anders als die Versuche der Paläographen und Kodikologen auf Bilder als Weg der Bedeutungsbeschreibung gänzlich verzichtet¹⁴. Das ist bei jüngst veröffentlichten diplomati-

11 <http://www.mysql.com>.

12 <http://www.php.net>.

13 Zur Tagung und ihren Ergebnissen vgl. Georg VOGELER, Ein Standard für die Digitalisierung mittelalterlicher Urkunden. Bericht zum Workshop (München 5./6. April 2004), in: AHF-Informationen 2004 Nr. 023 <http://www.ahf-muenchen.de/Tagungsberichte/Berichte/pdf/2004/023-04.pdf> und DERS., Standard (wie Anm. 9). Die Webseite der Charters Encoding Initiative <http://www.cei.lmu.de> gibt einen Einblick in die weitere Arbeit der CEI.

14 Die Arbeit am »Lexique international de Codicologie« hat das CIPL in der gleichen Zeit begonnen. Erste Ergebnisse liegen auf Französisch, Italienisch und Spanisch im Druck vor: Denis MIZERELLE, *Vocabulaire codicologique du français. Répertoire méthodique des termes français relatifs au manuscrit*, Paris 1985 (Rubricae, 1), Marilena MANIACI, *Terminologia del libro manoscritto*, Roma, Milano 1996 (Addenda, studi sulla conoscenza, la conservazione e il

schen Handbüchern wie dem von Benoît-Michel Tock, Olivier Guyotjeannin und Jacques Pycke oder dem von Thomas Frenz nicht der Fall¹⁵. Im Gegensatz zum Wörterbuch stehen bei einer mehrsprachigen Terminologie nicht die Beziehung zwischen Wörtern und ihr Gebrauch im Vordergrund, sondern die Beziehung zwischen den Sachverhalten und den Ausdrücken dafür in den verschiedenen Sprachen.

Die Beobachtung, dass das VID mehr als nur ein Wörterbuch ist, führt zur Suche nach passenden Technologien, die es ermöglichen sollten, die Beziehungen zwischen den Bezeichnern und den Definitionen sowie zwischen den Definitionen untereinander mit dem Computer auswertbar zu machen. Die relationale Datenbank, in der das VID derzeit abgelegt ist, geht dabei nicht weit genug, denn Texte sind in Tabellen und Spalten aufgeteilt, die für die Maschine vollständig gleichwertig sind. Welche Tabelle den Sachverhalt, welche Spalte den Terminus oder die Sprache, aus welcher der Terminus stammt, enthält, ist eine freie Entscheidung des Datenbankentwicklers und nur mit seinen Erläuterungen verständlich. Es ist also sinnvoll, die Daten in ein Format zu verwandeln, in dem die Bedeutungen der Strukturen öffentlich bekannt sind, die möglichst viele Forscher benutzen und die von allgemeinen Softwarelösungen berücksichtigt werden. Es geht dabei nicht um ein konkretes Dateiformat, sondern um die Datenmodelle, die damit ausgedrückt werden, also um die abstrakte Organisation der Informationen, welche die Funktionslogik der damit umgehenden Software und die möglichen Verwendungszwecke bestimmen. Zwei Modelle sollen hier näher diskutiert werden, nämlich die Beschreibungssprache der Text Encoding Initiative (TEI)¹⁶ und das vom W3C entwickelte Simple Knowledge Organization System (SKOS)¹⁷. Zu weiter gehenden Formalisierungen wie z.B. mit der vom W3C vorgeschlagenen Web Ontology Language OWL möchte ich im Ausblick noch kurze Bemerkungen machen¹⁸.

Für die Verwendung der TEI spricht, dass Geisteswissenschaftler diese Sprache zur Darstellung von geisteswissenschaftlichen Texten entwickelt haben, und ihre große Verbreitung unter den Anwendern von Computertechniken für geisteswissenschaftliche Fragestellungen. Auch

restauro del materiale librario, 3), Pilar OSTOS, M.^a Luisa PARDO, Elena E. RODRÍGUEZ, Vocabulario de codicología. Versión española revisada y aumentada del Vocabulaire codicologique de Denis Muzerelle, Madrid 1997 (Instrumenta bibliologica). Denis Muzerelle hat auch eine Webseite mit französischen, italienischen, spanischen und englischen Begriffen erarbeitet: DENIS MUZERELLE, Vocabulaire Codicologique. Répertoire méthodique des termes français relatifs aux manuscrits, avec leurs équivalents en anglais, italien, espagnol, 2002/3 <<http://vocabulaire.irht.cnrs.fr/>>. J. Peter Gumbert arbeitet an einer englischen Version und hat 2010 die ersten Entwürfe den Kollegen zur Kommentierung bereitgestellt: Words for Codices. A Codicological Terminology in English, an attempt by J. P. GUMBERT, <<http://www.cei.lmu.de/extern/VocCod/WOR10-1.pdf>>, <http://www.cei.lmu.de/extern/VocCod/WOR10-2.pdf> und <http://www.cei.lmu.de/extern/VocCod/WOR10-3.pdf>.

- 15 Benoît-Michel TOCK, Olivier GUYOTJEANNIN, Jacques PYCKE, *Diplomatique médiévale*, Turnhout 2006 (*L'atelier du médiéviste*, 2); Thomas FRENZ, *Papsturkunden des Mittelalters und der Neuzeit*, Stuttgart 2000 (*Historische Grundwissenschaften in Einzeldarstellungen*, 2). Anders hingegen Giovanna NICOLAJ, *Lezioni di diplomatica generale*, Roma 2007, die auf Bilder und Graphiken verzichtet
- 16 <<http://www.tei-c.org/>>.
- 17 SKOS Simple Knowledge Organization System. Reference, hg. von Alistair MILES, Sean BECHHOFFER, W3C Recommendation 18 August 2009 <<http://www.w3.org/TR/skos-reference/>> und SKOS Simple Knowledge Organization System-Home Page, hg. von W3C, 2004ff. <<http://www.w3.org/2004/02/skos/>>.
- 18 OWL Web Ontology Language. Reference, hg. von Mike DEAN, Guus SCHREIBER, W3C Recommendation 10 February 2004 <<http://www.w3.org/TR/owl-ref/>>. 2009 ist eine Version 2 von OWL entstanden: OWL 2 Web Ontology Language. Document Overview, hg. von d.W3C OWL Working Group, W3C Recommendation 27 October 2009 <<http://www.w3.org/TR/owl2-overview/>>.

im Bereich der Diplomatik ist die TEI verbreitet, wie sie z. B. im Projekt der »Chartes originales antérieures à 1121 conservées en France«¹⁹ oder bei der digitalen Edition der »Anglo-Saxon Charters« Anwendung findet²⁰. Die TEI bietet eine sehr große Zahl von möglichen Beschreibungselementen für Textstrukturen an. Interessanterweise sind darunter keine, die explizit den vorliegenden Fall beschreiben. In den Leitlinien der Verwendung der Beschreibungselemente sind folgende Abschnitte einschlägig: »2.3.6 The Classification Description«²¹, »3.3.4 Terms, Glosses, Equivalents, and Descriptions«²², »3.7 Lists«²³ und »9 Dictionaries«²⁴. Die im Abschnitt zur »Classification Description« vorgestellten Beschreibungselemente sind als Teil des `teiHeader` primär ein Instrument, auf das Dokument anwendbare bibliographische Taxonomien zu referenzieren. Die Elemente können aber auch dazu verwendet werden, selbst eine Taxonomie zu definieren²⁵. Der Abschnitt über »Dictionaries« zielt auf die Strukturen von Lexika, in denen hoch komprimiert grammatische und semantische Informationen kodiert sind. Von allgemeinem Interesse ist, dass die TEI ihr Konzept von einsprachigen Wörterbüchern ableitet und Übersetzungen als Zitat einer Bedeutungsvariante versteht, mithin also entgegen ihrer Tendenz explizite Beschreibungselemente typisierten vorzieht und kein Element für »Übersetzung« kennt, sondern die Übersetzung als Zitat des Typs »translation« versteht. Die Definitionen von »Termini« (`<term>`) und »Glossen« (`<gloss>`) in den Richtlinien der TEI²⁶ entsprechen eher der Funktion des VID, wobei die Definition von `<gloss>` als »a phrase or word used to provide a gloss or definition for some other word or phrase« die Vorstellung von knapper Interlinearglossierung evoziert, die den ausführlichen Texten des VID nicht entspricht. Eine Konstruktion des VID als Liste hat in der TEI schließlich die Schwäche, dass die Richtlinien nur zweiwertige Listenpunkte kennen, bestehend aus einem Bezeichner und einem Wert. Die Konstruktion des VID, das einen numerischen Bezeichner, eine Beschreibung und eine Liste an Termini kennt, ist damit nur in einer verschachtelten Liste auszudrücken. Die Struktur des VID mit Hilfe der TEI auszudrücken, scheint also Behelfskonstruktionen vorauszusetzen.

Dass die Guidelines in ihrer derzeitigen Form nicht unbedingt geeignet sind, Sachkonzepte zu beschreiben, hat schon 2004 zur Gründung einer Ontologies Special Interest Group geführt. Sie diskutiert, wie die TEI so zu erweitern sein könnte, dass sie nicht nur zur Strukturierung von Text, sondern auch zur Beschreibung von Sachverhalten eingesetzt werden kann²⁷. Sie hat sich bislang jedoch vorrangig um die typischerweise in Registern ausgedrückten Personen und Orte und um die Verwendung der TEI als Beschreibung von Museumsmaterial kümmern können. Für eine Terminologie wie dem VID sind die Ergebnisse dieser Arbeitsgruppe derzeit noch nicht nutzbar.

Wenn man das VID also mit den Elementen der TEI beschreiben will, rücken seine Eigen-

19 Bearb. v. Cédric GIRAUD, Jean-Baptiste RENAULT, Benoît-Michel TOCK, Nancy 2010 <<http://www.cn-telma.fr/originaux/index/>>.

20 ASChart-Project, bearb. von Dame Janet NELSON, Harold SHORT, Paul SPENCE <<http://www.aschart.kcl.ac.uk/>>.

21 In: TEI P5: Guidelines for Electronic Text Encoding and Interchange, hg. von TEI Consortium, Version 1.9.1, 5.3.2011 <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD55>>.

22 Ibid. <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COHQU>>.

23 Ibid. <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COLI>>.

24 Ibid. <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html#DITPTR>>.

25 Vgl. <<http://www.cei.lmu.de/VID/tei/taxonomy.php>>.

26 TEI P5: Guidelines (wie Anm. 21) <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gloss.html>>.

27 TEI Ontologies SIG, betreut von Øyvind EIDE, Christian Emil ORE, 2004–2010 <<http://www.tei-c.org/SIG/Ontologies/>>.

schaften als Text in den Vordergrund, und der Codierer muss auf allgemeine linguistische Beschreibungselemente ausweichen. In Anlehnung an die gedruckte Vorlage wird dann z.B. jeder Sachverhalt als ein Abschnitt (<div>) behandelt, dem die Definition als Beschreibung (<desc>) und ein Absatz mit den Termini (<term>) zugeordnet sind²⁸.

Leichter fällt die Übersetzung der Daten aus dem VID in die Regeln von SKOS, die das W3C seit 2009 ausdrücklich für die Gestaltung von Terminologien vorschlägt. Auch wenn vorwiegend human-, natur- und sozialwissenschaftliche Terminologien in diesem Format gespeichert sind, liegen auch geisteswissenschaftliche Terminologien in SKOS vor. Das Format wird von einer Reihe von Programmen unterstützt²⁹. Ob diese für diplomatische Fragestellungen nutzbar sind, werde ich unten kurz diskutieren. Hier geht es zunächst um den Vergleich der Struktur der SKOS-Vorschläge und des VID. SKOS kennt genau wie das VID Bezeichner (*label*, als *prefLabel* [*preferred label*] und *altLabel* [*alternative label*]) und Beschreibung (*definition*) für ein abstraktes Konzept und ordnet die Konzepte hierarchisch. Es erfordert also keinen nennenswerten Programmieraufwand, um aus der oben beschriebenen relationalen Datenbank SKOS-Daten in den beiden dominierenden Notationsformen für SKOS – Turtle und RDF/XML – zu machen³⁰. Die Kapitelstruktur, die Nummern und Unternummer des VID stehen dabei für Konzepte, die hierarchisch miteinander in Beziehung stehen. Die Termini sind *label* für diese Konzepte. Die im »Vocabulaire« mit Asterisk oder mit der Abkürzung »cf.« markierten Texte sind unbestimmte Verweise. Listing 1 gibt dafür ein Beispiel.

```
vid:VID_1 rdf:type skos:Concept;
  skos:altLabel "Urkundenlehre"@de;
  skos:altLabel "Urkundenforschung"@de;
  skos:altLabel "Diplomatik"@de;
  skos:altLabel "Diplomatic"@en;
  skos:altLabel "Diplomática"@es;
  skos:prefLabel "Diplomatique"@fr;
  skos:altLabel "Diplomática"@it;
  skos:definition
    "La Diplomatie
    est la science qui étudie la *tradition, la *forme et l'*élaboration des
    *actes écrits. Son objet est d'en faire la critique, déjuger de leur
    *sincérité, d'apprécier la qualité de leur texte, de dégager des formu-
    les tous les éléments du contenu susceptibles d'être utilisés par l'his-
    torien, de les dater, enfin de les éditer."@fr;
  skos:inScheme vid:VID;
  skos:related vid:VID_24;
  skos:related vid:VID_108;
  skos:related vid:VID_3;
  skos:related vid:VID_265;
  skos:related vid:VID_131 .
vid:VID_23 rdf:type skos:Concept;
  skos:altLabel "Regesten"@de;
```

28 <<http://www.cei.lmu.de/VID/tei>>. Dort können die VID-Daten in drei alternativen TEI-Formaten eingesehen werden: als div-term Konstruktion unter <<http://www.cei.lmu.de/VID/tei/div.php>>, als list/label/item-Konstruktion unter <<http://www.cei.lmu.de/VID/tei/list.php>> und als taxonomy-Ausschnitt unter <<http://www.cei.lmu.de/VID/tei/taxonomy.php>>.

29 Vgl. die Übersicht des W3C <<http://www.w3.org/2004/02/skos/tools>> mit weiterführenden Links.

30 <<http://www.cei.lmu.de/VID/skos>>.

```

skos:altLabel "chancery rolls"@en;
skos:altLabel "regesta"@es;
skos:altLabel "catálogo"@es;
skos:prefLabel "regestes"@fr;
skos:altLabel "regesti"@it;
skos:altLabel "regestenlijst"@nl;
  skos:definition
    "Des regestes
    sont, réunies en un recueil chronologique, des *analyses d'actes provenant
    d'un même *chancellerie ou concernant un même objet, accompagnées éventuel-
    lement des indications historiques illustrant l'itinéraire d'un prince,
    les événement de son principat, les faits concernant la personne physique
    ou morale concernée."@fr;
  skos:inScheme vid:VID;
  skos:narrower vid:VID_23-1;
  skos:related vid:VID_39;
  skos:related vid:VID_268.

```

Listing 1: Ein Ausschnitt aus dem VID als SKOS-Ressource, in Turtle-Syntax

Können diese beiden Kodierungsformen des VID in der praktischen Arbeit mit Urkunden eingesetzt werden? Die in den Definitionen der Beschreibungselemente der TEI angelegte Logik beschränkt sich technisch auf Unter- und Überordnungen sowie darauf, dass bestimmte Elemente immer in Verbindung mit anderen verwendet werden. Jede weitere Bedeutung, die mit den Beschreibungselementen der TEI geschaffen wird, kann erst mit explizit dafür entwickelter Software nutzbar gemacht werden. Mir sind keine spezifischen Instrumente bekannt, die sich gerade auf die für die Kodierung des VID verwendeten Beschreibungselemente beziehen. In der Tools-Übersicht des TEI-Konsortiums finden sich für die oben benannten Beschreibungselemente fertige Softwaremodule nur für die Publikation³¹. Eine Codierung des VID nach den Regeln der TEI orientiert sich also am dominierenden Standard des Computereinsatzes in den Geisteswissenschaften, hat aber nur einen eingeschränkten technischen Nutzen.

Das ist bei der Codierung nach den Regeln von SKOS anders. Auch im geisteswissenschaftlichen Bereich gibt es Referenzbeispiele für die Benutzung von SKOS als Thesaurusformat. Sowohl der Object Type Thesaurus des National Monuments Record Centre in der Association (MDA) der English Heritage³² wie der UNESCO-Thesaurus³³ verwenden das Format. Eine SKOS-Terminologie des VID würde sich also gut in solche Angebote und die damit verbundenen Softwarewerkzeuge integrieren lassen. Die Integration ist sinnvoll, denn die für Urkunden einschlägigen Stellen in den beiden Thesauri sind unzureichend. Beide Thesauri werden verwendet, um Objekte des kulturellen Erbes zu klassifizieren und ihre Beschreibungen in großen Datenbanken zu normieren. Im MDA-Thesaurus gibt es z.B. die Konzepte SEAL und DOCUMENT, die aber nicht miteinander verbunden sind. Ein DOCUMENT ist Teil der Klasse WRITTEN ITEM, die Teil von WRITTEN COMMUNICATIONS ist. Als Unterklasse gibt der Thesaurus gerade einmal BULLA an. Privilegien, *instrumenta*, *chartae/notitiae*, Breven,

31 <<http://www.tei-c.org/Tools/>>, <<http://www.tei-c.org/Activities/SIG/Tools/>> und <<http://wiki.tei-c.org/index.php/Category:Tools>>. Vgl. zum Verhältnis der TEI zu Ontologien auch Christian-Emil ORE, Øyvind EIDE, TEI and cultural heritage ontologies: Exchange of information?, in: *Literary and Linguistic Computing* 24 (2009), S. 161–172.

32 <http://thesaurus.english-heritage.org.uk/thesaurus.asp?thes_no=144>. MDA steht für die »Museum Documentation Association«.

33 <<http://www2.ulcc.ac.uk/unesco>>.

Mandate sind ebensowenig vorgesehen wie die Unterscheidung zwischen Siegelurkunden und Notariatsinstrumenten. Der UNESCO-Thesaurus definiert noch allgemeiner DOCUMENTS als Teil des Abschnittes »MT 5.30 Information sources« und damit als Oberbegriff zu »Primary Documents« und »Manuscripts«, die wiederum keine speziellere Unterscheidung kennen. Das VID könnte hier also an das Konzept »Document/Primary Document« anknüpfen und eine Vielfalt an Differenzierungen ermöglichen.

Darüber hinaus hat die SKOS-Ressource, in die das VID mit relativ geringem Aufwand verwandelt werden kann, ein eigentlich diplomatisches Anwendungsgebiet, und zwar im Umfeld der digitalen Urkundenpräsentation als Regest oder Edition. Die dafür derzeit eingesetzten technischen Lösungen strukturieren ihre Daten nämlich unterschiedlich. Die Entscheidung für die eine oder die andere Struktur lässt sich aus diplomatischer Sicht gewöhnlich gut rechtfertigen. Als Beispiel möge ein Ausschnitt aus dem »Codice diplomatico della Lombardia medievale« dienen. Michele Ansani hatte schon relativ früh auf die Codierung der Texte mit XML gesetzt, zu einer Zeit, da die Vorschläge der TEI für eine solche Codierung noch deutliche Schwächen bei den Phänomenen aufwiesen, die eine diplomatische Edition ausmachen³⁴. Unter anderem war das diplomatische Formular nicht sinnvoll codierbar³⁵. Schon 1999 hielt er das VID für ein sinnvolles Hilfsmittel für diese Art von Codierung³⁶. Listing 2 zeigt einen Ausschnitt und macht deutlich, dass die Namen der Beschreibungselemente vom italienischen diplomatischen Fachvokabular abgeleitet sind.

```
<TENOR>
  <PROTOCOLLO>(ST) <DTCRON>Anno dominice incarnationis milleximo centesimo trigesimo secundo, mense decembris, indicione undecima.
</DTCRON></PROTOCOLLO>
  <TESTO><DISPOSITIO>Tibi <RECIP><PERSONA nm="Pigiriolus qui dicitur da Nerviano">Pigiriolo <NOTA>(a)</NOTA> qui dicitur da <TOP nm="Nervianum" id="Nerviano, Mi" tipo="loc" cert="certo">Nerviano
</TOP></PERSONA><REST>et </REST><LB/><PERSONA nm="Boregnianus qui dicitur da Nerviano" pat="Alfeus qd"> Boregniani filio <PERSONA nm="Alfeus qd" fil="Boregnianus qui dicitur da Nerviano">quondam Alfei</PERSONA> qui dicitur da Nerviano</PERSONA></RECIP> promitto atque spondeo me ego <AUCT><PERSONA nm="Ardericus" pat="Iohannes qui dicitur de Cantone de loco Vogenzade qd" lex="langobarda">Ardericus filius <PERSONA nm="Iohannes qui dicitur de Cantone de loco Vogenzade qd" fil="Ardericus">quondam Iohannis qui dicitur de Cantone de loco <TOP nm="Vogenzade" id=" Guanzate, Co" tipo="loc" cert="certo"><REST>Vogen</REST><LB/>zade</TOP></PERSONA>, qui professus sum lege vivere Longobarda</PERSONA></AUCT>
  (...)
```

Quidem et <NOTA>(b)</NOTA> ad hanc adfirmandam <LB/> promissionis cartam accepi ego qui supra Arderico a te qui supra Pingiriolo exinde launehil-

- 34 Vgl. Michele ANSANI, Procedure di lavoro: la codifica XML dei testi, in: Codice diplomatico della Lombardia medievale, hg. von DEMS., Pavia 2000–2011, <<http://cdlm.unipv.it/progetto/codifica-xml>>.
- 35 Vgl. auch Georg VOGELER, Towards a standard of encoding medieval charters with XML, in: Literary and Linguistic Computing 20 (2005), S. 269–280.
- 36 Michele ANSANI, Diplomatica (e diplomatisti) nell'arena digitale, in: Scrineum 1 (1999) <<http://scrineum.unipv.it/biblioteca/ansani.htm>>, gekürzt auch in: Archivio Storico Italiano 158 (2000), S. 349–379.

```

di lempum de manstruca <REST>una</REST>. <FORMULAE>Quia sic in<LB/>ter
nos convenit.
</FORMULAE></TESTO>
<ESCATOCOLLO>
  <DTTOP>Actum in suprascripta civitate Mediolani.</DTTOP>
  <P/>
  <SMR>Signum + manum <NOTA>(a)</NOTA> suprascripti Arderici qui hanc
  cartam promissionis ut supra fieri rogavit.</SMR>
  <P/>
  <SMT>Signum +++ <NOTA>(d)</NOTA> manuum <PERSONA nm="Tedaldus Vuitar-
  dus">Tedaldi Vuitardi</PERSONA>, <PERSONA nm="Oddelus qui dicitur Girvi-
  nus">Oddeli qui dicitur Girvini</PERSONA>, testium.</SMT>
  <P/>
  <COMPLETIO>(ST) Ego <SCRIPT><PERSONA nm="Otto notarius sacri palatii">
  Otto notarius sacri palatii</PERSONA></SCRIPT> scripsi, postraditam con-
  plevi et dedi.</COMPLETIO>
</ESCATOCOLLO>

```

Listing 2: Codierungsbeispiel des »Codice diplomatico della Lombardia medievale« (Urkunde des Mailänder Domkapitels, Dezember 1132, ed. Maria Franca Barone, <<http://cdlm.unipv.it/edizioni/mi/milano-chiesamaggiore/carte/chiesamaggiore1132-12-00>>

Man kann den Text aber auch mittels anderer XML-Standards strukturieren. Seit 2005 kodiert z. B. das Monasterium-Projekt seine Daten gemäß den Vorschlägen der Charters Encoding Initiative, und es wäre sicherlich sinnvoll, wenn auch die lombardischen Urkunden in einem überregionalen Urkundenportal wie »Monasterium« recherchierbar wären³⁷. Die obige Urkunde würde nach den Regeln der CEI wie in Listing 3 aussehen:

```

<tenor>
  <protocol><notariusSign /> <datatio>Anno dominice incarnationis mil-
  leximo centesimo trigesimo secundo, mense decembris, indicione undeci-
  ma. </datatio></protocol>
  <context><dispositio>Tibi <recipient><persName reg="Pigiriolus qui
  dicitur da Nerviano">Pigiriolo <ref type="note">(a)</ref> qui dicitur da
  <placeName reg="Nervianum" id="Nerviano, Mi" type="loc" certainty="cer-
  to">Nerviano </placeName></persName><supplied>et </supplied><lb /><pers-
  Name reg="Boregnianus qui dicitur da Nerviano Alfei qd"> Boregniani filio
  <persName reg="Alfeus qd Boregniani qui dicitur da Nerviano">quondam Al-
  fei</persName> qui dicitur da Nerviano</persName></recipient> promitto
  atque spondeo me ego <issuer><persName reg="Ardericus Iohanni qui dici-
  tur de Cantone de loco Vogenzade qd">Ardericus filius <persName reg="Io-
  hannes qui dicitur de Cantone de loco Vogenzade qd">quondam Iohannis qui
  dicitur de Cantone de loco <placeName reg="Vogenzade" id="Guanzate, Co
  "type="loc" certainty="certo"><supplied>Vogen</supplied><lb/>zade</place-

```

37 Vgl. zu Monasterium.net jüngst Karl HEINZ, Monasterium.net. Auf dem Weg zu einem europäischen Urkundenportal, in: Regionale Urkundenbücher. Die Vorträge der 12. Tagung der Commission Internationale de Diplomatique, hg. von Theo KÖLZER, St. Pölten 2010 (Mitteilungen aus dem Niederösterreichischen Landesarchiv, 14), S. 139–145. Zur Technik vgl. mom-wiki, hg. v. Jochen GRAF, Georg VOGELER, 2008–2011 <<http://www.mom-wiki.uni-koeln.de/>>.

```
Name></persName>, qui professus sum lege vivere Longobarda</persName></
issuer>,
(...)
```

```
Quidem et <ref>(b)</ref> ad hanc adfirmandam <lb /> promissionis cartam
accepi ego qui supra Arderico a te qui supra Pingiriolo exinde launechil-
di lempum de manstruca <supplied>una</supplied>. <setPhrase type="formu-
la">Quia sic in<lb />ter nos convenit. </setPhrase></dispositio></cont-
ext>
```

```
<eschatocol>
  <issuedPlace>Actum in suprascripta civitate Mediolani.</issuedPlace>
  <p />
  <notariusSub>Signum <pict type="cross"/> manum <ref>(a)</ref> supra-
scripti Arderici qui hanc cartam promissionis ut supra fieri rogavit.</no-
tariusSub>
  <p />
  <notariusSub>Signum <pict type="triple cross" /> <ref>(d)</ref> ma-
num <persName reg="Tedaldus Vuitardus">Tedaldi Vuitardi</persName>,
<persName reg="Oddelus qui dicitur Girvinus">Oddeli qui dicitur Girvi-
ni</persName>, testium.</notariusSub>
  <p />
  <notariusSub><notariusSign /> Ego <notarius><persName reg="Otto no-
tarius sacri palatii">Otto notarius sacri palatii</persName></notarius>
scripsi, postcoraditam complevi et dedi.</notariusSub>
</eschatocol>
</tenor>
```

Listing 3: Codierungsbeispiel der Urkunde des Mailänder Domkapitels, Dezember 1132, ed. Maria Franca Barone, <<http://cdlm.unipv.it/edizioni/mi/milano-chiesamaggiore/carte/chiesamaggiore1132-12-00>>, in CEI

Die Text Encoding Initiative erarbeitet einen allgemeinen Datenaustauschstandard in den Geisteswissenschaften. 2007 stellten Paul Spence und Ariana Ciula eine ganz auf der TEI basierende Lösung für die »Anglo-Saxon Charters« vor³⁸. Eine Codierung des obigen Textes nach ihren Vorschlägen würde vermutlich wie in Listing 4 aussehen:

```
<text>
  <ab type="protocoll">(ST) <date>Anno dominice incarnationis milleximo
centesimo trigesimo secundo, mense decembris, indicione undecima.
</date></seg>
  <ab type="context"><seg type="disposition">Tibi <seg type="beneficiary"><persName>Pigiriolo <ref>(a)</ref> qui dicitur da <placeName key="-
Nerviano, Mi" certainty="certo">Nerviano </placeName></persName><sup-
plied>et </supplied><lb/><persName> Boregniani filio <persName key="Alfeus
qd">quondam Alfei</persName> qui dicitur da Nerviano</persName></seg>
```

38 Vgl. auch Ariana CIULA, Paul SPENCE, Threads of Integration. The Anglo-Saxon Charters pilot project at the Centre for Computing in the Humanities, in: Digitale Diplomantik. Neue Technologien in der historischen Arbeit mit Urkunden, hg. von Georg VOGELER, Köln u.a. 2009 (Archiv für Diplomantik. Beiheft, 12), S. 40–55.

```

promitto atque spondeo me ego <seg type="issuer"><persName key="Ardericus">Ardericus filius <persName key="Iohannes qui dicitur de Cantone de loco Vogenzade qd">quondam Iohannis qui dicitur de Cantone de loco <placeName key="Guanzate, Co" certainty="certo"><supplied>Vogen</supplied>
<lb/>zade</placeName></persName>, qui professus sum lege vivere Longobarda</persName></seg>
( . . . )

```

```

Quidem et <ref>(b)</ref> ad hanc adfirmandam <lb/> promissionis cartam
accepi ego qui supra Arderico a te qui supra Pingiriolo exinde launehildi
lempum de manstruca <supplied>una</supplied>. <seg type="Clausula">-
Quia sic in<lb/>ter nos convenit.</setPhrase> </seg></ab>
  <ab type="eschatocoll">
    <placeName>Actum in suprascripta civitate Mediolani.</placeName>
    <seg type="subscription">Signum + manum <ref>(a)</ref> suprascripti
    Arderici qui hanc cartam promissionis ut supra fieri rogavit.</seg>
    <seg type="subscription">Signum +++ <ref>(d)</ref> manum <persName
    key="Tedaldus Vuitardus">Tedaldi Vuitardi</persName>, <persName key="Od-
    delus qui dicitur Girvinus">Oddeli qui dicitur Girvini</persName>, tes-
    tium.
  </seg>
  <seg type="subscription">(ST) Ego <seg type="scriptor"><persName
  key="Otto notarius sacri palatii">Otto notarius sacri palatii</persName>
  </seg> > scripsi, postraditam conplevi et dedi.</seg>
</ab>
</text>

```

Listing 4: Codierungsbeispiel der Urkunde des Mailänder Domkapitels, Dezember 1132, ed. Maria Franca Barone, <<http://cdlm.unipv.it/edizioni/mi/milano-chiesamaggiore/carte/chiesamaggiore1132-12-00>>

Die drei Texte unterscheiden sich nur in der Art und Weise, mit der ihre Struktur markiert ist. Das können andere Namen für die Strukturelemente sein oder, wie im Fall der TEI-Version, allgemeinere Bezeichner, die mithilfe des `type`-Attributs näher bestimmt werden: `<tei:div type="protocoll"> = <cei:protocoll> = <cdlm:protocollo>`. Eine Terminologie mit Definitionen ermöglicht es, diese unterschiedlichen Strukturierungsvorschläge daraufhin zu prüfen, ob und wo sie die gleichen Sachverhalte meinen. Sie werden damit leichter ineinander konvertierbar. Zum Austausch muss dann nur das Beschreibungsmuster mit den Konzepten des VID verbunden werden. Im besten Fall referenziert schon die Schema-Definition auf das VID. Die TEI sieht z.B. in ihrer Spezifikation von Schemabeschreibungen (Abschnitt 22, auch geläufig unter dem Akronym ODD) vor, Beschreibungselemente der TEI mit Hilfe des `equiv`-Attributs mit einer externen Ressource zu verknüpfen³⁹. In einem TEI-Dokument

39 Vgl. zum Problem den Vortrag von Sebastian Rahtz auf dem »Treffen der Ontologies SIG« in Zadar im November 2010 <<http://tei.oucs.ox.ac.uk/Oxford/2010-11-12-CRM/>> und die Experimente von K. Faith LAWRENCE, Crowdsourcing Linked Data From Shakespeare To Dr Who. pp. 1–3, in: Proceedings of the ACM WebSci'11, June 14–17 2011, Koblenz, Germany, 2011 <<http://journal.webscience.org/499/>> und Michael O. JEWELL, Semantic Screenplays: Preparing TEI for Linked Data, in: Digital Humanities 2010, London 7.–10.7.2010 <<http://www.contextus.net/node/94>>.

könnte das VID z.B. als Taxonomie in den *classification declarations* eingebunden werden, und könnten key- und type-Attribute darauf referenzieren⁴⁰.

Die Verwendung bestimmter Annotations-elemente durch die verschiedenen Beschreibungsformate macht auch Aussagen über die hierarchische Struktur von Texten. So versteht das Beschreibungsschema des CDLM die Klauseln z.B. als von der eigentlichen Disposition unabhängigen Teil des Urkundentextes. Das VID definiert dagegen die Klauseln als Teil der Dispositio⁴¹. Der Unterschied mag marginal wirken, ist aber für eine Suche nach Stichwörtern in der Dispositio von Relevanz. Wenn nun die Definitionen des VID technisch vollständig abgebildet sind, dann kann eine Urkundensuchmaschine⁴² aus dem Verweis in der Beschreibungsdefinition einer digitalen Urkundenpräsentation auf das VID schließen, dass im CDLM in eine Suche in der Dispositio auch das Element <CLAUSULAE> einzubinden ist.

Auch jenseits der Struktur der Codierung können sich Anwendungsmöglichkeiten des VID in digitalen Urkundenpräsentationen ergeben. So können mithilfe des VID sprachliche Unschärfen vermieden werden. Alleine im VID gibt es sechs Konzepte, auf die der Bezeichner *cédule* zutrifft. Eine Beschlagwortung von Urkundentypen mit diesem Wort alleine lässt also offen, ob es sich um protokollarische Notizen der Karolingerzeit oder Anlagen handelt. Ein formalisiertes Schlagwort wie »Cédule (VID_412Ad)«, das für die protokollarischen Notizen in karolingischer Zeit steht, ist dagegen klar von »Cédule (VID_411a)« unterscheidbar, das ein Schreiben einer Autorität an die von ihr Abhängigen meint⁴³.

Auch die Hierarchisierung des VID ermöglicht Differenzierung. Gerade in italienischen Urkundeneditionen ist es z.B. üblich, jede Urkunde mit einer typologischen Überschrift zu versehen⁴⁴. Dabei ist über Grenzen der Editionen hinweg nicht klar, ob sich diese Überschrift auf den Inhalt oder auf die diplomatische Form bezieht. Eine Klassifikation als »donatio (VID_425)« macht hingegen klar, dass es sich um eine inhaltliche Typologie handelt, denn VID_425 ist Teil des Kapitels 6 »Nature juridique des actes«. Sie wäre damit aus einer Liste der formal diplomatischen Gattungen auszuschließen, in der »instrumentum (VID_295)« oder »privilegium (VID_393)« zu finden wäre. Eine Suche nach Rechtsinhalten kann damit die Überschrift »donatio (VID_425)« miteinbeziehen, die Überschrift »instrumentum (VID_295)« nicht.

40 Zu diesem Zweck ist das VID als *tei:taxonomy*-Ausschnitt unter <<http://www.cei.lmu.de/VID/tei/taxonomy.php>> verfügbar.

41 <<http://www.cei.lmu.de/VID/VID.php?205>>: »clauses (lat.: *clausula*) est une *disposition particulière d'un acte, dont l'ensemble constitue le *dispositif de cet acte (...).«

42 Zu den Konzepten für eine solche Urkundensuchmaschine vgl. Markus HELLER, Georg VOGELER, Modern information retrieval technology for historical documents, in: Humanities, Computers and Cultural Heritage. Proceedings of the XVIth international conference of the Association for History and Computing, Amsterdam 2005, S. 143–148 (online: http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20051064.pdf) und Georg VOGELER, Fachspezifische Indexierung von historischen Dokumenten. Quellen zwischen Zeichenketten und Information – Beispiel Urkunden, in: Forschung in der digitalen Welt. Sicherung, Erschließung und Aufbereitung von Wissensbeständen, hg. von Rainer HERING u.a., Hamburg 2006 (Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt Hamburg, 20), S. 43–58, online: <http://hup.sub.uni-hamburg.de/opus/volltexte/2008/77/chapter/HamburgUP_Forschung_Vogeler.pdf>.

43 Ich verwende im Beitrag der Verständlichkeit halber häufiger eine Notation, die den französischen Terminus wiederholt, um das abstrakte Konzept zu bezeichnen. Innerhalb einer wirklichen Anwendung ist ein abstrakter Bezeichner »VID_412Ad« vorzuziehen, den die Benutzeroberfläche in die Sprache des Benutzers übersetzt.

44 Alessandro PRATESI, Una questione di metodo. L'edizione delle fonti documentarie, in: DERS., Tra carte e notai: Saggi di diplomazia dal 1951 al 1991, Roma 1992 (Miscellanea della società romana di storia patria, 35), S. 7–31 verweist schon auf die »terminologia consacrata dalla diplomazia« (S. 21).

Im Kontext der Regestierung und Edition von Papsturkunden wird der Bedarf an dieser Art von personen- und projektübergreifenden Referenz besonders deutlich. Die Masse des zu erfassenden Materials und die breite geographische Streuung seiner Überlieferung leistet einer regionalen oder wenigstens nationalen Arbeitsweise Vorschub. Die Regestierung der Papsturkunden vor 1198 ist ebenso wie der Nachweis der Originale nach 1198 im »Censimento« national oder gar nach Lagerort organisiert. Allen Mitarbeitern an diesen Projekten ist aber bewusst, dass sie Ergebnisse einer zentralen Produktionsstätte bearbeiten. Diese Annahme kann den Blick auf Unterschiede verstellen, die innerhalb eines Staates, eines Archivs oder auch nur der Aufgaben eines Bearbeiters nicht auftreten. So hat Hans-Henning Kortüm erst aus dem Vergleich der Papsturkunden mehrerer Regionen mit den Privaturkunden in diesen Räumen schließen können, dass die Sprache der älteren Papsturkunden auch vom Empfänger abhängt⁴⁵. Umgekehrt sind statistische Auswertungen wie die von Stefan Hirschmann⁴⁶ erst möglich, wenn die dezentral erstellten Regesten und Editionen projektübergreifende Kriterien und Kategorien verwenden. Die gemeinsamen Richtlinien für die Erfassung von Papsturkunden müssen also beim Übergang in eine digitale Repräsentationsform möglichst auch technisch zusammengefasst werden, z.B. in einem Datenschema, das sich auf das VID bezieht, oder in einem gemeinsamen Thesaurus für die Beschlagwortung, der als SKOS-Ressource für alle Forscher und damit für Anfragen, die den Bearbeitungsraum übergreifen, zugänglich ist. Dass die Mehrsprachigkeit des VID für ein gemeinsames Suchportal der Papsturkunden bei einem europäischen Forscherpublikum auch von ganz praktischem Nutzen ist, sei nur am Rande angemerkt.

Das digitale VID kann aber auch als Instrument dienen, die Beziehungen zwischen Urkunden auszudrücken, die an verschiedenen Stellen im Internet verfügbar sind. Es macht damit die unter dem Stichwort »Semantic Web« zusammengefassten Techniken für die Diplomatik nutzbar⁴⁷. Tim Berners Lee hatte schon 1999 vorgeschlagen, das auf Hypertext basierende World Wide Web zu einem Semantic Web weiterzuentwickeln, in dem nicht nur die Dokumente unspezifisch aufeinander verweisen (»Internet der Dokumente«), sondern in dem die Verweise vom Computer auswertbar sind (»Internet der Dinge«)⁴⁸. Es handelt sich dabei nur um ein Teilgebiet dessen, was unter »Künstlicher Intelligenz« firmiert, denn die Verfahren, mit denen die semantisch aufgeladenen Verknüpfungen ausgewertet werden, sind ein eigenständiger Forschungsbereich. Es geht stattdessen ausschließlich darum, die Dokumente, die im Internet verfügbar sind, durch spezifische Verweise auf andere Dokumente genauer zu erschließen. Deshalb fußt die Kerntechnologie für das »Semantic Web« auch auf der Theorie der »gerichteten

45 Hans-Henning KORTÜM, Zur päpstlichen Urkundensprache im frühen Mittelalter. Die päpstlichen Privilegien 896–1046, Sigmaringen 1995 (Beiträge zur Geschichte und Quellenkunde des Mittelalters, 17).

46 Stefan HIRSCHMANN, Statistische Anmerkungen zu den Papsturkunden Lucius' III., Urbans III. und Gregors VIII. (1181–1187), 1997 <http://www.historik-hirschmann.de/hauptteil_Aufsatz.htm>; DERS., Die päpstliche Kanzlei und ihre Urkundenproduktion (1141–1159), Frankfurt am Main u.a. 2001 (Europäische Hochschulschriften, III/913).

47 Zum Einsatz der Konzepte des Semantic Web im Bereich der Historischen Grundwissenschaften vgl. u.a. Toby BURROWS, Applying Semantic Web Technologies to Medieval Manuscript Research, in: Kodikologie und Paläographie im Digitalen Zeitalter, Bd. 2, hg. von Franz FISCHER, Christiane FRITZE, Georg VOGELER, Norderstedt 2010 (Schriften des Instituts für Dokumentologie und Editorik, 3), S. 117–131 (online: <http://kups.ub.uni-koeln.de/4346/>; urn:nbn:de:hbz:38-43464), Robert KUMMER, Semantic Technologies for Manuscript Descriptions? Concepts and Visions, *ibid.*, S. 133–154 (online: <http://kups.ub.uni-koeln.de/4347/>; urn:nbn:de:hbz:38-43472) und die Beiträge zur Tagung »Digital Diplomatics. Tools for the Digital Diplomatist«, Neapel 29.9.–1.10.2011, zur Publikation vorgesehen im »Archiv für Diplomatik«.

48 Tim BERNERS-LEE, James HENDLER, Ora LASSILA, The Semantic Web, in: Scientific American Magazine 17 (Mai 2001) online: <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>.

Graphen«, die sich mit der Logik solcher Verweisarten beschäftigt⁴⁹. Diese Technologie firmiert unter dem Namen »Resource Description Framework« (RDF), die das W3C 1998 zu einem der Internetstandards erklärt hat⁵⁰. Es legte dabei Regeln fest, wie digitale und abstrakte Objekte so zueinander in Beziehung zu setzen sind, dass sie nach dem Modell »Subjekt – Prädikat – Objekt« satzähnliche Aussagen formulieren. Die folgende Zeichenkette `<artem:charte4902>` `<vid:51_confirmation>` `<google-books:chartes-de-la-Picardie-132>`, würde damit aussagen, dass die im ARTEM-Projekt transkribierte Urkunde 4902 eine Bestätigung einer Urkunde ist, die in einer Urkundensammlung der Picardie gedruckt und so z.B. bei Google-Books online verfügbar wäre. Sie würde dem Computer mitteilen, dass Informationen über zwei Urkunden an verschiedenen Stellen im Internet miteinander in Beziehung stehen⁵¹. Sie würde ihm schließlich sagen, dass diese Beziehung den Regeln von »Vocabulaire international de la diplomatie« folgt. Es gibt eine große Anzahl von Datenbankprogrammen für solche Ausdrücke, die unter der Gattungsbezeichnung »Triple Store« laufen⁵². Als Anfragesprache setzt sich SPARQL durch, das ein rekursives Akronym für SPARQL Protocol and RDF Query Language ist. Das W3C hat die Sprache 2008 in seinen Standardempfehlungen aufgenommen⁵³.

Von technologischer Seite ist also RDF gut unterstützt. Um diese Technologien aber sinnvoll für diplomatische Forschungen einsetzen zu können, bedarf es dreierlei: 1. digitale Repräsentationen von Urkunden, 2. eine Beschreibung der Urkunden in Formaten, aus denen RDF-Statements generiert werden können, und 3. für den Diplomatiker aussagekräftige Verbindungen in Form von Taxonomien und Ontologien. Über den Entwicklungsstand des ersten Punkts in dieser Agenda habe ich an anderer Stelle berichtet⁵⁴. Beim zweiten und dritten Punkt kommt das digitale VID ins Spiel: bei der Strukturierung und als Referenzziel von RDF-Sätzen über die digitale Urkundenpräsentation, wie oben ausgeführt.

Wie macht das praktischen Sinn? Wenn das VID das gesamte Spektrum an diplomatischen Kategorien, Fragestellungen und Methoden abdecken würde, dann könnte man annehmen, dass alle diplomatischen Aussagen über eine Urkunde als RDF-Sätze formuliert werden könnten. Das ist schon aus dem systematischen Grund nicht richtig, dass diplomatische Aussagen wie andere geisteswissenschaftliche auch Interpretationen und argumentative Vorschläge sind. Sie benötigen also auch die Freiheit rhetorischer Überzeugungsversuche und lassen sich nicht vollständig formalisieren. Sehr viel praktischer ist die Überlegung, dass es bei der Editions- und Regestierungsarbeit stark formalisierte Aussagen gibt, die sich gut in RDF-Sätze übertragen ließen. So können Aufzählungen von Beglaubigungsmitteln, die Bestimmung der Urkundenart sowie Listen von Ausstellern, Empfängern, Notaren oder Zeugen formalisiert und unter Verweis auf das VID dargestellt werden. Es ist dabei nicht notwendig, dass diese Aussagen das RDF-Format haben, solange sie die drei Positionen eines RDF-Triples eindeutig bezeichnen und die Texte an den Positionen den Regeln von RDF entsprechen. Eine Codierung eines Zeugen als `<cei:charter id="n_142">` ... `<vid:263_témoins>`Bonushomo de Vero-

49 Zur Graphentheorie vgl. Reinhard DIESTEL, Graphentheorie, Heidelberg 2010 und Peter TRITTMANN, Graphentheorie: Eine anwendungsorientierte Einführung, Leipzig 2003.

50 Resource Description Framework (RDF), hg. von W3C, 1999–2004 <http://www.w3.org/RDF/>.

51 Genau genommen muss zusätzlich noch festgelegt werden, dass die ARTEM-Edition und der Druck, den Google-Books digitalisiert hat, eindeutige Referenzen für jeweils eine bestimmte Urkunde sind.

52 Eine Übersicht liefert das W3C in seinem Semanti-Web-Angebot unter »Category: Triple Store«, letzte Änderung 2009 http://www.w3.org/2001/sw/wiki/Category:Triple_Store.

53 SPARQL Query Language for RDF, W3C Recommendation, hg. von Eric PRUD'HOMMEAUX, Andy SEABORNE, 15 January 2008 <http://www.w3.org/TR/rdf-sparql-query/>.

54 VOGELER, Digitale Urkundenbücher (wie Anm. 1).

na</vid:263_témoins> ... <cei:charter> kann z.B. leicht in das RDF-Triple <gall-pont_1:n_142> <vid:263_témoins> "Bonushomo de Verona". umgewandelt werden. Diese Art der Auszeichnung kann im Übrigen ebenso dafür verwendet werden, Register zu erstellen. Die Akkumulation solcher Register und ihre Verbindung mit Namensrepositorien wie der von den Bibliotheken entwickelten »Gemeinsamen Normdatei« (GND) würde die digitale Urkundenpräsentation mit nicht diplomatischen Informationen über andere Dokumente im Web verknüpfen, also z.B. mit Literaturangaben zu Personen oder mit Archivfindbüchern⁵⁵. Für die Vernetzung der anwachsenden Zahl an digitalen Urkundenpräsentationen wäre es darüber hinaus erstrebenswert, diplomatisch klassifizierbare Verweise zu anderen Urkunden so zu gestalten, dass sie in RDF-Triple umgewandelt werden können, also so, dass die Art und Weise der Beziehung mit einer Referenz auf das digitale VID und das Ziel als stabile kanonische Referenz formuliert sind⁵⁶.

Alles das ist auf dem derzeitigen Stand der Digitalisierung des VID schon realisiert oder realisierbar. Der letzte Teil dieses Beitrags soll jedoch eine Perspektive aufzeigen, wohin die Digitalisierung des VID führen könnte. Es ginge darum, noch mehr vom logischen und semantischen Gehalt des »Voculaire« technisch auszudrücken. Dafür stünde die Web Ontology Language OWL zur Verfügung. Das W3C hat die Web Ontology Language im Jahr 2004 zu einer offiziellen Empfehlung gemacht⁵⁷. OWL ist eine besonders detaillierte Methode, Sachverhalte und Bezüge zwischen ihnen für den Computer nachvollziehbar darzustellen. Sie dient dazu, Klassen und Eigenschaften zu benennen, diese miteinander zu Axiomen zu verbinden und konkrete Instanzen damit in Beziehung zu setzen, um so Inkonsistenzen finden und automatische Schlüsse ziehen zu können. OWL orientiert sich an der Prädikatenlogik und hat diese in zwei Formaten (OWL Lite und OWL-Description Logic) umgesetzt, welche der Computer in endlicher Zeit entscheiden kann, die also in Software umsetzbar sind⁵⁸. Eine wichtige Eigenschaft dieser Sprache ist, dass sie die in SKOS sehr eingeschränkte Menge an Arten der Beziehungen zwischen den Objekten näher zu beschreiben vermag, sodass insbesondere Ausschlusskriterien formuliert werden können.

Diese Darstellung ist für das VID deshalb von Interesse, weil die im VID definierten Begriffe weniger über ihre Hierarchie, also über Teil-Ganzes-Beziehungen, Bedeutung gewinnen als über die Definitionen selbst, die sich mit SKOS nicht ausdrücken lassen⁵⁹. Das nächste Ziel der Digitalisierung des VID müsste also sein, seine Definitionen in eine formale Sprache zu übersetzen. Damit ist z.B. die Unterscheidung des VID zwischen VID_54 »copie authentique« und VID_55 »copie informée« als exklusiv zu beschreiben, während beide Typen z.B. auch als VID_56 »copie figurée« gestaltet sein können. Die Liste der verschiedenen Arten von Kopien ist ein gutes Beispiel, welches Potential in der Beschreibungslogik steckt. Eine Kopie, auf welche die Definition von VID_61 »Une copie sous le sceau« zutrifft (»est une copie dont l'authenticité est attestée par l'apposition du sceau d'une autorité«) ist z.B. immer auch eine »copie authentique« (VID_54), weil sie ein Siegel als »élément de validation« trägt, wie in der Definition von VID_54 gefordert: »Une copie authentique (lat.: *copia authentica*) est une copie qui, établie dans une certaine forme, comporte des éléments de validation destinés à lui donner pleine foi.« Derartige Schlussmechanismen sind Teil von Software, welche die Bezeichnung »Rea-

55 Zur GND vgl. die Projektvorstellung der Deutschen Nationalbibliothek <<http://www.d-nb.de/standardisierung/normdateien/gnd.htm>>, dort mit weiterführender Literatur.

56 Zur kanonischen Referenzierung vgl. Bernhard ASSMANN, Sind die kanonischen Zitierweisen der Geisteswissenschaften als nachhaltige Komponenten digitaler Repositorien geeignet?, Magisterarbeit Köln 2005 <<http://www.cei.lmu.de/pub/MagArbAssmann.pdf>>.

57 OWL Web Ontology Language. Reference (wie Anm. 18).

58 Das dritte Format »OWL-Full« ist nach dem derzeitigen Stand der Technik nicht zuverlässig in realistischer Zeit berechenbar.

59 Vgl. Pascal HITZLER u.a., *Semantic Web. Grundlagen*, Heidelberg u.a. 2008.

soner« trägt⁶⁰. Die Umsetzung des VID in OWL würde es also ermöglichen, solche Instrumente mit in die diplomatische Analyse von digitalen Urkundenpräsentationen einzubeziehen.

Besonders interessant ist ein solcher Ansatz natürlich bei kontinuierlich wachsenden und verteilt erarbeiteten Projekten wie z.B. der Regestierung und Edition der Papsturkunden. Ein digital konzipiertes Regestierungsprojekt würde dazu z.B. einerseits eine gemeinsame Personendatenbank aufbauen, die ontologisch versucht, aus einer Vielfalt von Varianten der Namen für eine Person, aber auch aus der paläographischen Identifikation von Händen abstrakte Personenidentitäten festzulegen. Andererseits würde es die Verbindung zwischen den Regesten und den Personen mithilfe einer Ontologie wie dem VID benennen, d.h. die Rollen der Personen im Rechtsgeschäft, bei der Produktion der Urkunde und für ihre Überlieferung mit eindeutigen Namen belegen, die für bestimmte Konzepte stehen. Dann könnten schon während der Erfassung leicht »Ausreißer« gesucht und gefunden werden. So trifft z.B. die Definition von VID_286d-1 (ingrossierender Notar der Papstkanzlei vor den Reformen Innocenz' III.) nur auf Urkundenschreiber (VID_274) zu, wenn die in diese Kategorie eingeordneten Personen stabil mit VID_286b (*notarii papae*) korrespondieren. Eine Hand, die einmalig eine Urkunde für einen anderen *notarius papae* ausfertigt, fiel also auf.

Mit einem größeren Fundus an Daten könnten auch Versuche unternommen werden, Annahmen über Regeln in der Papstkanzlei zu treffen und diese mit der regestierten Kanzleiproduktion zu vergleichen. Solche diplomatischen Fragestellungen werden zwar auch mit klassischen Methoden angegangen, aber nicht umsonst hat die Masse der Papsturkunden schon des 12. Jahrhunderts Stefan Hirschmann dazu bewegt, sie mit statistischen Methoden auszuwerten⁶¹. Das VID und die Erschließung der Urkunden mit seiner Hilfe würde komplexere Zusammenhänge auszudrücken versuchen als die zeitliche Verteilung einzelner Merkmalsausprägungen und so dabei helfen, die Arbeitsweise der Papstkanzlei im 12. Jahrhundert aus den alleine für den Zeitraum 1181–1186 mehreren Tausend Urkunden zu rekonstruieren.

SPARQL-Anfragen machen indirekte Verbindungen zwischen Urkunden, Personendatenbank und Konzepten des VID sichtbar, die formalisiert notiert und in einer RDF-TripleStore gespeichert sind. Der Anfrage »Zeige alle Urkunden, in denen die Kombination aus dem Verweis unter der Kategorie VID_274 (>Schreiber<) mit Unterkategorien auf eine bestimmte Person und dem Verweis unter der Kategorie VID_286b (*notarius papae*) auf eine bestimmte Person nur einmal vorkommt!« würde dem SPARQL-Ausdruck in Listing 5 entsprechen:

```
PREFIX regpont: <http://regesta-pontificia.net/>
vid: <http://www.cei.lmu.de/VID/skos/>
skos: <http://www.w3.org/2004/02/skos/core#>

SELECT ?urkunde ?notariusPP ?scribe
WHERE {
  ?urkunde vid:VID_3 regpont:charter ;
  ?vid_schreiber ?scribe ;
  vid:VID_286b ?notariusPP .
  OPTIONAL { ?vid_schreiber skos:narrower vid:VID_274 . }
  OPTIONAL { ?vid_schreiber a vid:VID_274 . }
  GRAPH ?s { ?scribe a regpont:person . }
```

60 Vgl. Marko LUTHER u.a., Who the Heck Is the Father of Bob? A Survey of the OWL Reasoning Infrastructure for Expressive Real-World Applications, in: The Semantic Web: Research and Applications. Lecture Notes in Computer Science 5554 (2009), S. 66–80.

61 HIRSCHMANN, Statistische Anmerkungen (wie Anm. 46) und DERS., Päpstliche Kanzlei (wie Anm. 46).

```

GRAPH ?n { ?notariusPP a regpont:person . }
FILTER (?s != ?n)
}
GROUP BY ?notariusPP ?scribe
HAVING count(?urkunde) = 1

```

Listing 5

Für komplexere Zusammenhänge wären abstrakte Hypothesen als spezielles Regelsystem der Kategorien des VID zu formulieren. Dafür stehen Regelsprachen wie die Semantic Web Rule Language (SWRL) zur Verfügung⁶². Ein Reasoner kann diese Regeln mit der wachsenden Zahl an digital, also formalisiert erschlossenen Urkunden vergleichen und Inkonsistenzen feststellen. Die formalisierte Hypothese über die Arbeitsweise der Papstkanzlei würde also mit jeder neu hinzugefügten Urkunde überprüft werden. Neben Hypothesen über das Verhältnis von Konzipient und Schreiber könnte man auch Karrieren in der Papstkanzlei zu ermitteln versuchen und das Verhältnis von inneren und äußeren Merkmalen zum Kanzleigang oder zur Typologie der Urkunden näher erforschen.

Das seit 2011 online verfügbare digitale VID kann also wie sein gedruckter Vorläufer als Übersetzungshilfsmittel dienen. Es kann als Hilfsmittel bei der Beschlagwortung und bei der Strukturierung von Text in digitale Urkundenpräsentationen eingebaut werden. Beim Austausch von Daten zwischen digitalen Urkundenpräsentationen und bei der projektübergreifenden Suche kann es als Referenz für die diplomatisch bedeutsamen Formatbestandteile dienen und damit Teil von diplomatischen Forschungsfragen an große Datenmengen werden. Dieser Einsatz hängt entscheidend davon ab, ob die Definitionen des VID für die Diplomater als Referenz akzeptabel sind. Die Rezensionen von Guyotjeannin und Große haben Schwächen deutlich gemacht haben, die nicht nur darin bestehen, dass bestimmte Phänomene nicht erfasst sind, sondern auch darin, dass die einzelnen Definitionen hinter den Stand der diplomatischen Forschung zurückfallen. Beide Phänomene scheinen jedoch nur einen kleinen Ausschnitt des VID zu betreffen, sodass es sinnvoll erscheint, eine dritte, verbesserte und rein digitale Auflage des »Vocabulaire« ins Auge zu fassen, welche die Kritik aufnimmt.

Würde man bei einer gründlicheren Untersuchung des Definitionsbestandes jedoch zu dem Ergebnis kommen, dass substantielle Mängel vorliegen, müsste man alternativ an ein größeres angelegtes Forschungsprojekt denken, das die einschlägige Handbuchliteratur zur europäischen Diplomatie auswertet und eine neue Ressource aufbaut⁶³. Bis auf Weiteres werde ich die existierende digitale Repräsentation des VID weiter betreuen, mich um die Korrektur der OCR-Fehler kümmern und an Beispielanwendungen arbeiten. Ob sich das digitale VID durchsetzen kann, inhaltlich verbessert wird und sich damit die mit der Druckfassung beabsichtigten Zwecke, diplomatische Kategorien zu klären, sie für Außenstehende zugänglich zu machen und zur Vereinheitlichung der Editionsarbeit beizutragen, im Digitalen erfüllen lassen, das wird die Zukunft erweisen.

62 Vgl. dazu Ian HORROCKS u.a., SWRL: A Semantic Web Rule Language. Combining OWL and RuleML, W3C Member Submission 21 May 2004 <<http://www.w3.org/Submission/SWRL/>> und Peter F. PATEL-SCHNEIDER, A Proposal for a SWRL Extension towards First-Order Logic, W3C Member Submission 11 April 2005 <<http://www.w3.org/Submission/SWRL-FOL/>>.

63 Ein solches Projekt wurde von Luciana Duranti, Antonella Ghignoli, Olivier Guyotjeannin und mir im Jahr 2009 schon einmal konzipiert, konnte die Geldgeber aber leider nicht überzeugen.

