UDC: 005.922.52:930.25

Georg Vogeler
University of Graz
Centre for Information Modelling – Austrian Centre for Digital Humanities
Merangasse 70, A-8010 Graz, Austria
georg.vogeler@uni-graz.at

DIGITAL DIPLOMATICS:
WHAT COULD THE COMPUTER CHANGE
IN THE STUDY OF MEDIEVAL DOCUMENTS?

Abstract: The essay discusses the research possibilities created by the computer for diplomatic research. Following the steps of traditional scholarly work with medieval documentary heritage (heuristics, documentation, comparison, grouping, establishing rules, diplomatic critique, scholarly edition) it evaluates recent studies applying digital methods to charters. It argues that the easier accessibility of images online, the possibility to process large numbers of documents and a new conceptualisation of editing bring significant changes to diplomatics. It tries to identify the major changes in five fields: (1) verbal description is characterised by classification, rather than representation; (2) visual evaluation is enhanced by easy access to images and by the help of image analysis software; (3) full text search supported by natural language processing software makes it possible to connect a single charter to others hidden in large charter corpora characteristic of the Late Middle Ages; (4) statistical methods and visualization of large data sets helps to see the single charter in its full contemporary context; (5) digital editing can aggregate work done separately in one common representation and makes the editors' knowledge on the documents explicit. Raising the question if this "digital diplomatics" is fundamentally different from pre-digital diplomatics, it comes to the conclusion that digital diplomatics helps to relieve the research of medieval documents from epistemological constraints determined by the conceptualisation of its results as to be represented in the form of a printed book.

Keywords: diplomatics, computer, digital humanities, methodology, critical edition, palaeography.

Everything is going digital these days: the libraries, the archives, even the whole field of humanities. This papers deals with the "digitisation" of a research branch of the humanities which has a long tradition – diplomatics.

The paper therefore can begin with one of the founding fathers of modern critical study of legal documents dealing with single cases.[1] It starts with a model of diplomatics research developed mostly by Theodor Sickel in the mid-nineteenth century.[2] In this model, the work of the diplomatist starts with finding the documents to be studied (heuristics). That meant travelling through archives, following hints of colleagues and vague entries in former publications and even in the local finding aids, trying to convince the owner to give access to the archives and similar issues. When sitting in front of an original or even a copy of a charter the scholar knew he would have to leave the archives soon and therefore worked on a documentation of the item from as much perspectives as he could think to be interesting: transcription, damages, existence of validation means like seals, time and method of copy. There are many publications in the 19th and 20th century whose scientific value consists of making a newly found document known to the scientific comunity (*Neufund*). For the critical diplomatic study Sickel developed the idea of *Kanzleimäßigkeit*, i.e. that a document can be considered as a forgery if it doesn't follow the rules of charter production in the issuing institution. Sickel knew that not every prince, not every small lord had the possibility to employ personal clerks who developed traditions or can even be identified by their hand, but had to rely on the scribes of their documents' recipients. Thus the major concept of Sickel's diplomatic critique was not so much the establishment of a chancery rule, but the reconstruction of the usual production process in the specific historical situation in which a given document was created. To achieve this, the diplomatist had to compare the documents appearing to be issued from one person/institution, and to try to build groups among them (e.g. by diplomatic type, hand, textual style, author, subject, region of production, writing sup-

---

[1] Luciana DURANTI, From Digital Diplomatics to Digital Records Forensics, *Archivaria* 68 (2009) 39–66, is trying to develop a time independent theory of diplomatics, mixing the pre-modern concepts of juridical diplomatic judgement with problems of archival management. She uses therefore the term "digital diplomatics" in a different way than this paper does. She considers the critical analysis of electronic records in the archives and their transfer into the archives as "Digital Diplomatics", which could be translated as "archival sciences of the electronic document" and "forensics of the electronic document". Her ahistorical approach is not discussed in this paper.

[2] On Sickel cf. O. REDLICH, Theodor Sickel, *Mitteilungen des Instituts für Österreichische Geschichtsforschung* 42 (1927/28) 153–164. For his methods of diplomatics research see TH. VON SICKEL, Programm und Instruktionen der Diplomata-Abteilung, *Neues Archiv der Gesellschaft für ältere deutsche Geschichtskunde* 1 (1876), 427–482 and *Die Urkunden Konrad I., Heinrich I. und Otto I.*, ed. TH. SICKEL, MGH. Die Urkunden der Deutschen Könige und Kaiser 1, Hannover 1879, I–XIX.

port, means of authentication, etc.). From these groups, the scholar could deduce characteristics, which could be termed as rules although they might be valid only for a short period of time. Documents which showed deficiencies in these rules were considered spurious or, if enough hints came together, blatant forgeries. This judgement is an important step in critical diplomatic work (*discrimen veri ac falsi*). The major form in which all this work is published is the critical scholarly edition. These seven steps of the model of diplomatic research as it was developed in the 19th century will be the guideline to compare the current situation of scholarly work on medieval and early modern legal documents and what computer technologies contribute to it. The question behind that is whether the use of the computer with its currently available methods could change the working model of diplomatics as far as to justify the notion of "Digital Diplomatics" as a special branch of "Diplomatics".

Heuristics

Envisaging the diplomatist in his everyday environment means an image of him sitting in a library reading books hoping to find hints on the existence of documents in which he is interested. These books don't have to be primarily printed but can be manuscript too. Works like the *Regesta imperii* by Johann Friedrich Böhmer[3] or the *Regesta pontificum Romanorum* by August Potthast[4] can be considered as good examples of the results of this kind of work – they are a kind of specialised analytic index to printed works but the researcher got a well ordered short list of documents and references to their publication place, and therefore a starting point for further diplomatic and historical research.

But we certainly have to envisage the diplomatist sitting in archival study rooms or often working as an archivist in the archival vaults themselves. The researcher would have to rely on the inventories created by the archivists or would have to browse the documents themselves. As an archivist he could follow the organisation of the documents by the archival holders and would create inventories of charters, registers or chartularies to help the future researcher. The researcher would have to read them selectively to find the charters of one issuer, on one subject, of a specific type.

---

[3] The "first" volume of the still ongoing project: J.-FR. BÖHMER, *Regesta chronologico-diplomatica regum atque Romanorum inde a Conrado I. usque ad Heinricum VII. Die Urkunden der Römischen Könige und Kaiser von Conrad I. bis Heinrich VII. 911-1313*, Frankfurt a. M. 1831 with pp. III–XIV on the concept of the work.

[4] *Regesta Pontificum Romanorum inde ab a. post Christum natum 1198 ad a. 1304*, ed. A. POTTHAST, vols. 1–2, Berlin 1874–1875 (repr. Graz 1957[2]).

The access to reference works has changed with their digitisation. Online archival inventories, charter databases and digitised prints help the researcher to find the documents he's interested in, because there is more information available and the check takes less time. The diplomatist can check the existence of potentially interesting documents before he travels to the holding archives and he can browse a library of books less readily available in print. The heuristics as further improved with efficient search facilities. While the simplest way of digitising printed references is to create an image of each page which can only be read by a person, with automatic character recognition software you can add the possibility of full text search by keywords. This adds a new way of working with the references as the search for a freely chosen keyword is as fast – or even faster – as the search in the selection of keywords in the analytical index, and certainly faster than to read a complete book to find a single document. Keyword search therefore opens the digitally accessible charter corpus to new research questions.

From the point of diplomatic research this software still has drawbacks, as it usually relies on a linguistic check after identifying a possible range of characters on a single spot in the image. Medieval Latin and old forms of the vernacular aren't very well documented in digital dictionaries and morphologic analyzers.[5] Additionally the specific diplomatic methods of coding observations with abbreviations and special signs (e.g. the $\overset{X}{\underset{X}{}}$ of the *elongata* in the MGH charter editions) can be a challenge to commercial OCR software.[6] To get to a better full text quality (the aim for commercial OCR software is to reduce the error rate to 1 per 1000 characters or less), methods of typing in the text have been applied to several projects, e.g. the dMGH, the CAO, the Regesta Imperii.

The diplomatist is interested in single charters – not in books containing charters. To fulfill this research need a reproduced printed page as

---

[5] For the Latin in general the situation is much better: *LemLat* by the Istituto di Linguistica computazionale, CNR, Pisa <http://www.ilc.cnr.it/lemlat/>; the corpus of the *Perseus Digital Library* at Tufts-University in Medford/Somerville <http://nlp.perseus.tufts.edu/syntax/treebank/latin.html> or W. WHITACKER, WORDS. A Latin-English Dictionary Program, Version 1.97FC, 2006 <http://archives.nd.edu/whitaker/words.htm>.

[6] For checking the quality of OCR of printed charter editions in the digital texts you can compare almost every charter edition converted by commercial OCR software in the internet archive, e.g. *The Coucher Book or Chartulary of Whalley Abbey*, ed. W. A. HULTON. Vol. 4: Historical and Literary Remains Connected with the Palatine Counties of Lancaster and Chester, Manchester 1849, <http://archive.org/stream/coucherbook orch04hultgoog/coucherbookorch04hultgoog_djvu.txt>.

searchable full text is helpful but the identification of the single charter in the print would have to be the main goal of retro-digitisation for diplomatics research.[7] While the dMGH reproduces the text page by page on the internet it doesn't give a visible hint how to navigate to a single document. In fact there is a way to address each single charter – invisible for the user the dMGH has built in a reference table between pages and traditional citing styles of charters; it can be accessed via a resolving mechanism based on the canonical references for the charters: http://www.mgh.de/dmgh/diplomata/resolving/D_FI_219 is the URL for charter 219 in the edition of the documents of Emperor Frederic I by Heinrich Appelt, starting at page 365 of the first volume published in 1975.[8] The *Württembergisches Urkundenbuch* is an example for a digitised charter edition which has done the next step – the text of the printed version is stored in a database in which each charter is one entry, and the major divisions of a charter edition (number, abstract, comments, text, notes) are extracted as single sub-entries to the charter which can be searched separately.[9] Only recently experiments have been made to include this step into an automatic process with digitised charter editions in the Google books collection. They show that the identification of a single charter can be made automatically while the extraction of the parts of a charter edition is currently an unsolved problem.[10]

Converting printed editions into charter databases is a kind of reverse engineering. As many of the charters haven't been edited or even described in print, the diplomatics scholar has to do a search in the archives. The major entry points are the archival inventories. Almost all major archives are working on putting their inventories online and they are increasingly using

---

[7] G. VOGELER, Digitale Urkundenbücher. Eine Bestandsaufnahme, *Archiv für Diplomatik* (= *AfD*) 56 (2010) 363–392.

[8] See the description of the interface by C. RADL (MGH) at <http://www.mgh.de/dmgh/linking/#dd>.

[9] <http://www.wubonline.de>; on the project cfr. M. M. RÜCKERT – N. WURTHMANN, Das „Württembergische Urkundenbuch Online" im Netzwerk digitaler Urkundenpräsentationen, *Digitale Diplomatik, Neue Technologien in der historischen Arbeit mit Urkunden*, ed. G. VOGELER, *AfD* – Beiheft 12, Köln 2009 (= *Digitale Diplomatik* 2009), 130–145.

[10] Cfr. the results of the *German Virtual Charter Network* (*Virtuelles deutsches Urkundennetzwerk*) at <http://www.vdu.uni-koeln.de> with the Google data at <http://www.vdu.uni-koeln.de/vdu/collections>. The software was developed by the team of the chair for Historisch-Kulturwissenschaftliche-Informationsverarbeitung Prof. M. Thaller, in particular by J. Lenz and A. Streicher. An example with good results is *Tabulae Ordinis Theutonici, ex tabularii regii Berolinensis codice potissimum*, ed. E. STREHLKE, Berlin 1869 at <http://www.vdu.uni-koeln.de/vdu/OrdinisTheutonici/collection>.

database technologies to do that. Consequently, in these databases charters are represented as single items with structured descriptions. Some of the archives are even working on the digitisation of their holdings as images and often they start with their charter collections. The *Diplomatico* of the Florence State Archive can serve as an example that it is possible to publish a significant number, as they have already put almost two thirds of their charter collection online (85.000 of 140.000 documents). It has been created in several digitisation campaigns in five years between 1997 and 2002 and published online in 2008.[11] A much older project is the Hungarian State Archives *Diplomatic Collection*, which is based on a documentation of the scattered sources of medieval history of Hungary started already in the 1930s with microfilms and with database since the 1970s. Digital images were created of all the 140.000 documents and put online in 2010. It is currently the most extensive charter database online representing the collection of a single archive. Apart from its pure size already the archivists in the 1930s have made a step further by aggregating documents from different sources.[12]

The reduced costs for creation of digital images have given birth to other archival projects with a similar approach: the Soprintendenza per gli Archivi di Puglia started in 1999 to digitise the charters of 18 ecclesiastical and 17 smaller public and private archives in Apulia and published them since 2006 online.[13] The project aimed at facilitating the access to documentary heritage of these smaller institutions. A similar approach was the beginning of the monasterium.net project which started in the early 2000s as a project of Lower Austrian monastic archives. Meanwhile it has grown to a European charter database to which 106 archives from 13 European countries contribute information on their charters including images (cur-

---

[11] F. KLEIN, Il progetto 'Diplomatico' dell' Archivio di Stato di Firenze. Un archivio digitale di dati e immagini in costruzione, *Reti Medievali* 1 (2000) <http://www.storia.unifi.it/_rm/rivista/mater/Klein.htm>; IDEM, Il progetto 'Diplomatico' dell'Archivio di Stato di Firenze, *Le pergamene nell'era digitale*, ed. E. ANGIOLINI, Atti dei convegni di Spezzano (3 settembre 2004) e di Ravenna (24 settembre 2004), Modena 2005, 7–15.

[12] C. REISZ, Collectio Diplomatica Hungarica. Mittelalterliche Urkunden aus Ungarn online, *Digitale Urkundenpräsentationen*, eds. J. KEMPER – G. VOGELER, Schriften des Instituts für Dokumentologie und Editorik 6, Norderstedt 2011, 35-40; G. RÁCZ, Collectio Diplomatica Hungarica. The Online Portal of the National Archives of Hungary on Medieval Charters, *AfD* 56 (2010) 423–444.

[13] The project, managed by C. PALMA, was originally called *pergamopuglia* and has now moved in the context of the general archival information system under the name of "*Pergamene di Puglia online*" <http://www.sapuglia.it/index.php?option=com_content&view=article&id=46&Itemid=66>.

rently over 430.000 images).[14] For the heuristics of diplomatic research these projects show a significant difference to the pre-digital ways, as they give access to several charter collections preserved physically in different places. It can be considered as another step of separating the charter as an object of diplomatics study from its physical place of preservation.

A final source of charter heuristics are the diplomatic portals dedicated to special collections. In particular the older originals attracted the interest of researches proposing digitisation campaigns and so projects like the Belgian *Thesaurus diplomaticus*,[15] the German digitised *Lichtbildarchiv älterer Originalurkunden*[16] or the textual database of French *Chartes originales antérieurs à 1121* came into existence.[17]

In conclusion the process of digitisation has contributed three major possibilities to the heuristics in diplomatics research: increased availability of descriptions, keyword searches, and a high number of charter images available.

Documentation

The Belgian *Thesaurus diplomaticus* and the French database of originals are not only reference tools but have to be considered as first results

---

[14] <http://www.monasterium.net>; A. KRAH, Monasterium.net - das virtuelle Urkundenarchiv Europas. Möglichkeiten der Bereitstellung und Erschließung von Urkundenbeständen, *Archivalische Zeitschrift* 91 (2009) 221–246; A. SCHLICHTE, Urkunden im Netz: das internationale Urkundenportal monasterium.net. *Arbido. Fachportale für Kulturgut* 1 (23. 2. 2012) 30–31. A. AMBROSIO, Il progetto „monasterium" in Italia: le prime esperienze a Napoli, *Archivi* 5 (2010) 129–145; K. HEINZ, Monasterium.net - Auf dem Weg zu einem europäischen Urkundenportal, *Regionale Urkundenbücher*. Die Vorträge der 12. Tagung der Commission Internationale de Diplomatique, ed. by Th. KÖLZER – W. ROSNER – R. ZEHETMAYER, Mitteilungen aus dem Niederösterreichischen Landesarchiv 14, St. Pölten 2010, 139–145.

[15] *Thesaurus Diplomaticus*, created by PH. DEMONTY, CD-ROM Turnhout 1997; PH. DEMONTY, Le „Thesaurus Diplomaticus", un instrument de travail pour une nouvelle approche en diplomatique médiévale, *La diplomatique urbaine en Europe au Moyen Âge*, eds. TH. DE HEMPTINNE – W. PREVENIER, Studies in Urban Social, Economic and Political History of the Medieval an Early Modern Low Countries 9, Louvain – Apeldoorn 2000, 123–132.

[16] <http://lba.hist.uni-marburg.de/lba/>. S. MÜLLER, LBA-online. Datenbank, Strukturen und Online-Archiv des Digitalisierungsprojekts Lichtbildarchiv älterer Originalurkunden Marburg, *Digitale Urkundenpräsentationen*, 41–49.

[17] *Chartes originales antérieures à 1121 conservées en France*, eds. C. GIRAUD – J.-B. RENAULT – B.-M. TOCK, Telma, Orléans 2010. <http://www.cn-telma.fr/origi naux/index/>; on the project cfr. B.-M. TOCK, La base de donées des chartes originales antérieures à 1121 conservées en France, *Resourcing sources*, ed. K. S. B. KEATS-ROHAN, Prosopographica et genealogica 7, Oxford 2002, 153–163.

of diplomatic research. They give transcriptions, abstracts and descriptions of external features to each document. Transcribing and creating an abstract was already an important part of the work of the diplomatist in the pre-digital world. This information usually was kept private to prepare a critical edition. Walter Koch and his team e.g. are working on the edition of the charters of Emperor Frederic II since 1986. They have collected documentation on paper of all the ca. 4000 charters known. The files in the Munich offices include photographs, preliminary transcriptions, cards with the textual tradition and literature and extensive formal descriptions following a strict scheme of description.[18] This documentation is enhanced by something not written on paper. All the collaborators working on the documentation knew the rules of transcription applied, translated well known Latin names to modern names or recorded the dates of the documents in a normalised form.

The researchers in the pre-digital world had pencils and papers to produce this kind of documentation. They had to copy it by hand or with photocopying machines into new form. To implement the documentation schema established by the editors of the charters of Emperor Frederic II, the team used photocopied forms representing the structure of information to be collected. This kind of work is very similar to the digital databases which form the starting point of the *Chartes originales* or the *Lichtbildarchiv älterer Originalurkunden*. The organisation of the information in databases makes the access to the data easier as the data can be selected and sorted immediately.

A database already organises the documentation in a way that facilitates collaboration: the whole corpus can be queried simultaneously by different users and even modified if the users don't work on the same record. This cooperation can be done through internet interfaces. In the last few years software has been developed which facilitates the input of descriptive information on historical documents. The T-Pen project for example includes tools to transcribe an image "inline", i.e. it offers input fields displayed directly on the image.[19] The software developed for the monasterium.net-portal, called "Monasterium-Collaborative Archive" (MOM-CA) offers an online editor with fields developed for classical charter description (image 1). It includes mark-up of issuer and recipient in the abstract or the diplomatic discourse in the transcription.

The resources I've described above contain information on charters from different points of view. Connected, they can give a comprehensive

---

[18] Cfr. W. KOCH: Vorrede, in: *Die Urkunden Kaiser Friedrichs II.*, Vol. 1: 1198–1212, ed. W. KOCH, MGH Diplomata 14/1, Hannover 2002, XII.

[19] <http://t-pen.org>.

image of the charter. Thus an archival description could be reused for the preparation of a diplomatic edition, the transcription in a database of originals could be integrated into the edition of documents of a selected chancery which would include documents which have survived only as copies too. To achieve this interoperability a common structure is needed. In 2004 began the work on a proposal for a common vocabulary to describe the structure of diplomatic databases in XML (eXtensible Markup Language) – the Charters Encoding Initiative (CEI).[20] It is similar to a wider standard of structuring humanities texts with XML, the Text Encoding Initiative (TEI), which offers in particular a model how to describe the transcription of primary sources in a way that it can be read by the computer.[21] The CEI adds to the standards of the TEI description standards for validation methods and the diplomatic discourse. Both models still have to be developed further to include more description models for charter production.[22]

Already in the pre-digital world the scholarly community in diplomatics had developed a tool to facilitate information interchange on diplomatics – the Vocabulaire Internationale de Diplomatique (VID).[23] Since 2012 it is online at <http://www.cei.lmu.de/VID>. The online version offers not only the text and some methods to navigate through the terms for the human reader, but a version of the data in the Resource Description Framework (RDF) language defined by the W3C. It thus can be accessed from computers with the techniques of the so called "Semantic Web". These techniques offer some potential for digital diplomatic descriptions. First it can be used as an automatic dictionary of diplomatic terms. Thus, it can solve a problem of purely verbal description, such as homographs, i.e. one word that describes different phenomena. In the online version of the Vocabulaire Internationale de Diplomatique, identification codes of the

---

[20] <http://www.cei.lmu.de>, cfr. G. VOGELER, Ein Standard für die Digitalisierung mittelalterlicher Urkunden mit XML. Bericht von einem internationalen Workshop in München 5./6. April 2004, *AfD* 50 (2004) 23–34; IDEM, Towards a Standard of Encoding Medieval Charters with XML, *Literary and Linguistic Computing* 20 (2005) 269–280; IDEM, Charters Encoding Initiative (CEI). Zu Möglichkeiten der Integration mit Hilfe eines Standards für Urkundendigitalisierung, *Alte Archive – Neue Technologien*, eds. TH. AIGNER – K. WINTER, St. Pölten 2006, 181–198.

[21] <http://www.tei-c.org>.

[22] E.g. the model for Byzantine imperial charters presented by S. POULIMENOU – SP. ASONITIS – M. POULOS, Metadata Encoding for the Documents Based on the Rules of Diplomatics Science, *Metadata and Semantics*. Proceedings of the 2nd International Conference on Metadata and Semantics Research 11–12 October 2007, eds. A. SICILIA – M. D. LYTRAS, Berlin 2009, 91–100.

[23] *Vocabulaire international de la diplomatique*, ed. M. MILAGROS CÁRCEL ORTÍ, València 1997[2].

printed version are used to make these concepts discernable: while a "bull" can mean a papal document and a metal seal, <http://www.cei.lmu.de/VID /#VID_506> (metal seal) is not the same as <http://www.cei.lmu.de/VID/# VID_421-1> (papal document). Secondly a database which stores these unique identifiers for each concept can extract from the RDF representation of the VID multilingual terms and thus give the possibility to create multilingual search interfaces. A third possibility of the VID online is that it organises the definitions in hierarchies and thus gives the possibility to search with hierarchical inclusions. A database software making use of these possibilities could therefore expand a search for "notarii" (http:// www.cei.lmu.de/VID/#VID_73) into a search which would find "clerici curiae episcopi" (http://www.cei.lmu.de/VID/#VID_293) too.[24]

The content of a charter is usually described by the abstract (*regestum*). This can be considered as a rule-based linguistic representation, which in the end still has to be read by humans. Aleksandrs Ivanovs and Aleksey Varfolomeyev are studying the possibility to create 'translations' of charter text into a computer readable language, the "Attempto Controlled English" (ACE).[25] Their work is still in an experimental phase. However, computer linguists have already developed well established methods to extract so called Named Entities, i.e. parts of text which can be understood as names of individuals. The method of Named Entity Recognition can form the base to create analytic indices, or create complex research environments like the ChartEx project currently does.[26]

---

[24] On the conversion of the VID to a RDF-resource see G. VOGELER, Von der Terminologie zur Ontologie. Das „Vocabulare international de la diplomatique" als Ressource des Semantic Web, *Francia* 40 (2013) 281–297.

[25] Cfr. A. IVANOVS – A. VARFOLOMEYEV, Wiki Technologies for Semantic Publications of Old Russian Charters, paper given at the *Digital Humanities 2012*, Hamburg http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/wiki-technologies -for-semantic-publication-of-old-russian-charters/ and IDEM, Some Approaches to the Semantic Publication of Charter Corpora. The Case of the Diplomatic Edition of Old Russian Charters, *Digital Diplomatics. The Computer as a Tool for the Diplomatist?*, eds. A. AMBROSIO – S. BARRET – G. VOGELER, *AfD* – Beiheft 14, Köln 2014 (forthcoming). On ACE cfr. J. L. DE COI et AL., Controlled English for Reasoning on the Semantic Web, *Semantic Techniques for the Web*, eds. F. BRY – J. MAŁUSZYŃSKI, Heidelberg 2009, 276–308.

[26] On NER in general see D. JURAFSKY – J. H. MARTIN, *Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition*, Prentice Hall 2009, 759–768. An example of applying the methods to charters give H. STOYAN – S. SCHMIDT, Textmining in Beständen digitalisierter Urkunden, *Digitale Diplomatik* 2009, 318–324. For the ChartEx project (2012/13) see the homepage at <http://www.chartex.org>.
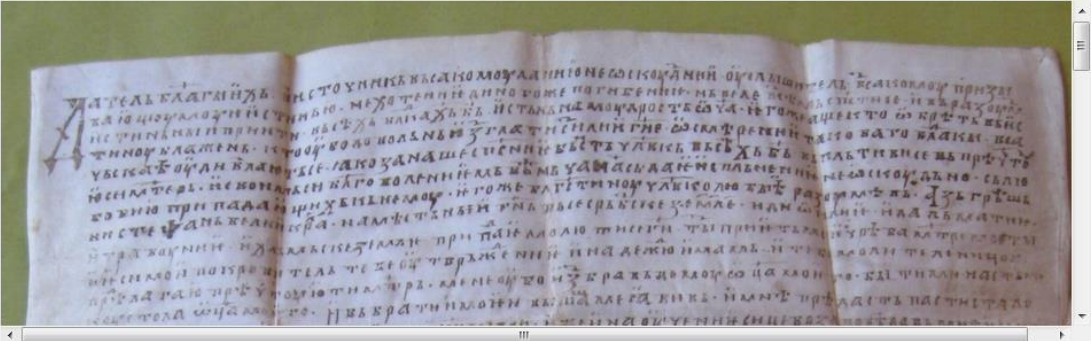
Image 1: monasterium.net editor mode of the charter of king Stefan Nemanjić to the monastery of the Holy Mother of God on the island of Mljet, 1217/27.
<http://www.mom-ca.uni-koeln.de/mom/SerbianRoyalDocumentsDubrovnik/1217xxxx_tpq
_%E2%80%93_Stefan_Nemanji%C4%87/charter>

Image 2a: Seal of Pernhart Graf zu Schawnberg, Landmarschall in Österreich, 13. September 1455, Wien , BayHStA Raitenhaslach Klosterurkunden 744 in monasterium.net
<http://www.mom-ca.uni-koeln.de/mom/DE-BayHStA/KURaitenhaslach/1455_09_13/charter>

Image 2b: Seal from the charter by Georg, Propst und Archidiakon von Baumburg, September 9, 1491, BayHStA Raitenhaslach Klosterurkunden 877 in monasterium.net
<http://www.mom-ca.uni-koeln.de/mom/DE-BayHStA/KURaitenhaslach/1491_09_09/charter>

One could argue that some of the descriptions of charters with a precise terminology could be replaced with digital images which have become much cheaper to produce than a formal description. Measuring the height and length for example can be done by the automatic separation of foreground and background.[27] Verbal descriptions are usually much less precise than the very image. Considering the two seals in Image 2a and 2b one can agree that a description "damaged seal in red wax" covers both of them. However both seals differ significantly in their value for art history research as the damage has destroyed the image on 2a almost completely. This difference is easy to see, but hard to describe. Therefore it is easier to conduct a specific research interested in the seal by looking at the image than by creating a detailed description which would cover all possible research interests. Images could therefore replace detailed descriptions.

But the description has still its justification: finding seals in red wax in good shape is made much easier if the description is reduced to a small controlled vocabulary in which the word "damage" for example is exclusively used for damaged seals and not for "no damage". Thus the more generic short description is rather a classification than an exhaustive verbalisation of the visual impressions. A description of the seal as "seal of the issuer" contributes an intellectual interpretation which makes it much more useful to research as it classifies the seal as a the symbolic representation of the legal will.

The software used to organise and display the charters in the monasterium.net collection has a built-in image annotation tool. With this tool you can add comments to specific parts of the charter's image and link it to the description of the charter.[28] A description made with the help of this tool can add significantly to the simple verbal description of the charter as it creates a link between a visual impression and its verbal expression. The indicative statement "This is the monogram of the document" is more preferable than the descriptive „The document carries a monogram".

Compare

Diplomatic judgement is based on the classical two categories of characteristics of a document: the "external" features, i.e. palaeographical and codicological, and the "internal" features, i.e. the linguistics and style of the text. Up to the second half of the 19th century, when photography

---

[27] The methods necessary for this are described in general introductions into image processing, like W. BURGER – M. J. BURGE, *Principles of Digital Image Processing: Core Algorithms*, London 2009, 5–48.

[28] A description of the tool in MOM-CA at <http://www.mom-wiki.uni-koeln.de/VDU/Imagetools/Usage>. The tool was developed by André Streicher, Cologne.

became a possibility to reproduce a visual impression, the pre-digital diplomatist had to rely greatly on his memory to compare the external features. A judgement such as "written in a 14th century book script", based on the experience of the scholar, was reliable as far as the skill of the scholar reached. Notes could support his memory and were the major tool for study of features like the content, the material of the seal etc. Transcriptions were the major tools to compare the internal features of the documents. The number of document transcriptions which could be compared at the same time was restricted by the possibilities to keep an overview. Comparison started usually with two documents. The corpus of documents collected and documented in the first steps took part in the comparison through the intuition of the scholar, "remembering" something to be checked directly.

Since transcriptions are made with the help of the computer, the handling of a large corpus has been facilitated extremely. Full text search has been enhanced by methods of proximity search, by display of the results in context and recently by fuzzy methods which expand the search keyword to "similar" words e.g. by allowing small differences in the lettering.[29] The research on how to compute natural language goes back a long time and therefore computer linguists have developed many tools which are useful in diplomatic research. Morphological and grammatical databases for example, can help searching when the search term and the words in the corpus are expanded to their respective lemma. These databases exist not only for modern languages but for Latin too.[30] They store the lemma, its morphological word forms and information on the grammatical type. Therefore they can be used to create a syntactic analysis of a sentence. With these methods, searching can be enhanced, e. g. to allow the search for phrases.

Diplomatics follows four major aims when comparing charter texts: (1) find citations from other charters, (2) identify the use of model texts by finding common set phrases, (3) establish an original text by collating several textual traditions and (4) find stylistic similarities in charters probably caused by the same person composing the text. As the analysis of differences between two versions of computer code is an old problem of com-

---

[29] An overview of approximate string matching algorithms can be found in Th. CORMEN – R. LEISERSON, *Introduction to Algorithms*, Cambridge (Mass.) 2001[2], 364–367; G. NAVARRO – R. BAEZA-YATES – E. SUTINEN – J. TARHIO, Indexing Methods for Approximate String Matching, *IEEE Data Engineering Bulletin* 24/4 (2001) 19–27; L. BOYTSOV, Indexing Methods for Approximate Dictionary Searching: Comparative Analysis, *Journal of Experimental Algorithmics* 16/1 (2011) 1–91.

[30] See note 5.

puter science[31] there are already applications to create collations automatically. One of the most advanced is Juxta <http://www.juxtasoftware.org/> which allows the user to put texts online, get visualisations of their similarities and differences and download an encoded version of the text comparison.[32] This method supports the second and the third task of the list above, because it detects textual similarities of given texts.

For the first task, the diplomatist can make use of the computer's searching capabilities. The comparison of features encoded for the computer can be done very quickly, as far as the existence of a single item or simple patterns is concerned. Finding all documents, which contain a given string of characters, in the 350.000 documents of the monasterium.net database for example, takes about three seconds. Comparing with the help of search engines makes it possible to compare a given document to documents yet unknown in the individual's diplomatic research. Els de Parmentier used this for her work on the chancery of the counts of Flanders and Hainaut (1191–1244).[33]

The comparison of visual features is supported not only by the easy access to digital images but by software tools too. Already standard image manipulation tools allow the user to extract zones of interest by zooming into images of high resolution, to separate letters from their background by changing colour values or by detecting edges. A set of these tools has been implemented into the MOM-CA software, giving the diplomatist easy access to them via the web. The image manipulation tool of MOM-CA adds the functionality to do the comparison itself – the user can create transparent extracted image zones and overlay them.[34]

---

[31] The oldest descriptions of "diff"-algorithms are E. JAMES – W. HUNT – M. D: MCILROY, An Algorithm for Differential File Comparison, *Computing Science Technical Report. Bell Laboratories* 41 (1976); E. W. MYERS, An O(ND) Difference Algorithm and Its Variations, *Algorithmica* 1/2 (1986) 251–266; W. MILLER – E. W. MYERS, A File Comparison Program, *Software — Practice and Experience* 15/11 (1985) 1025–1040; E. UKKONEN, Algorithms for Approximate String Matching, *Information and Control* 64 (1985) 100–118.

[32] Another webservice for textual comparison is CollateX <http://collatex.source forge.net/>, which currently offers a REST-API but hasn't a graphical user interface.

[33] Cfr. E. DE PAERMENTIER, Le poids des mots. Le dictamen de la chancellerie des comtes et comtesses de Flandre et de Hainaut (1191–1244). Méthode d'analyse assistée par ordinateur, résultats et opportunités, Bibliothèque d'École des Chartes 169 (2011), forthcoming and EADEM, *Diplomata Belgica*. Analysing Medieval Charter Texts (dictamen) Through a Quantitative Approach. The Case of Flanders and Hainaut (1191–1244), *Digital Diplomatics*. *The Computer as a Tool for the Diplomatist?*, forthcoming.

[34] A description of the functionalities at <http://www.mom-wiki.uni-koeln.de /VDU/Imagetools/Usage>.

With these tools the diplomatist is able to perform traditional tasks of charter comparison faster than he could without the computer and thus is able to process a greater number of samples. This velocity and the potential numbers of documents suggest a statistical approach. Therefore, the 'grouping' step in diplomatics work looks much different in digital diplomatics research than in the pre-digital age.

Grouping

Comparison between original documents in diplomatic research has the purpose to find documents written by the same scribe (similar letter forms and ductus) or composed by the same notary (same style). From repetitive use of phrases the researcher can deduce the use of a common model e.g. a common written resource like a formulary. Reoccurring clauses can be the result of common legal concepts, the diplomatic discourse part of chancery tradition. Copies of the same text have to be compared to establish the textual tradition in a stemma. Textual comparison helps to identify reused text from older privileges confirmed. A major step in diplomatic research is therefore grouping documents which share common features, stem presumably from the same chancery or answer the same historical need in society.

Statistical methods are fundamental to the work of Michael Gervers and his Toronto team. In the DEEDS project they are looking for a method to date the English private charters of the 13th century which usually don't carry a date. They built a database of these documents containing the full text of about 9000 charters. In this database they can calculate the distribution of arbitrary word pairs for the whole corpus and deduce from the dated documents the chance that a specific word pair is used in a specific time. New documents can be dated by matching their text to the database.[35]

The statistical approach gives Nicolas Perreaux the possibility to group documents typologically by their vocabulary and to distinguish groups by author. He built a database of 150.000 documentary texts from the 7th to the 14th century. He applied methods of correspondence analysis

---

[35] Cfr. M. GERVERS, The DEEDS Project and the Development of a Computerised Methodology for Dating Undated English Private Charters of the Twelfth and Thirteenth Centuries, *Dating Undated Medieval Charters*, ed. M. GERVERS, Woodbridge 2000, 13–35; M. GERVERS – M. MARGOLIN, Application of Computerised Analyses in Dating Procedures for Medieval Charters, *Le médiéviste et l'ordinateur* 42 (2003) <http://lemo.irht.cnrs.fr/42/mo42_01.htm>; IDEM, Managing Meta-data in a Research Collection of Medieval Latin Charters, *Digitale Diplomatik* 2009, 271–282; G. TILAHUN – A. FEUERVERGER – M. GERVERS, Dating Medieval English Charters, *Annals of Applied Statistics* 6 (2012) 1615–1640.

on the text and created a table of matching and not-matching textual features. This table can be clustered into four groups which correspond very well with the social status of the author, i.e. papal, imperial/royal, episcopal and private charters.[36]

A last example of computer based grouping is given by the Groningen Intelligent Writer Identification System (GIWIS) developed by Lambert Schomaker and his team. The system calculates statistics of images which can be used to identify hands. It contains functionalities to group a collection of images by similarities in script features like broadness of stroke or angularity of script.[37] It has been used for diplomatics research by Jinna Smit for her study of the Dutch chancery documents between 1299 and 1345. The software suggested groups of documents probably written by the same hand which highly correspond with her judgement as an expert in the chancery.[38]

All these methods calculate similarities in the distribution of letters or pixels of the digital representation of a document. They suggest clusters of documents in a corpus. The cited research has shown that these clusters come close to a diplomatics judgement and therefore are useful for diplomatics research. But they show too, that the results created by the computer still need evaluation of the human expert. Their main advantage lies in processing huge amounts of documents.

Another exploration approach is the visualisation of non-statistical data. Plotting documents on a map as the Berlin group editing documents of Charles IV is one example.[39] Image 3 is a visual representation of data

---

[36] Cfr. N. PERREAUX, De l'accumulation à l'exploitation? Expériences et propositions pour l'indexation et l'utilisation des bases de données diplomatiques, *Digital Diplomatics. The Computer as a Tool for the Diplomatist?*, forthcoming.

[37] Cfr. A. BRINK – M. BULACU – J. SMITS – L. SCHOMAKER, Writer Identification Using Directional Ink-trace Width Measurements, *Pattern Recognition* 45 (2012), 162–171 and M. AUSSEMS – A. BRINK, Digital Palaeography, *Kodikologie und Paläographie im Digitalen Zeitalter*, Schriften des Instituts für Dokumentologie und Editorik 2, Norderstedt 2009, 293–308.

[38] J. SMIT, The Death of the Palaeographer? Experiences with the Groningen Intelligent Writer Identification System (GIWIS), *AfD* 57 (2011) 413−426; EADEM, Meten is weten? De toepassing van het Groningen Intelligent Writer Identification System (GIWIS) op Hollandse kanselarij-oorkonden, 1299-1345, *Bulletin de la Commission royale d'Histoire* 176/2 (2010) 343–359 and her presentation at the conference *Digital Diplomatics 2011* in Naples < http://www.cei.lmu.de/digdipl11/slides-rep/Smit/index.pdf>.

[39] <http://telota.bbaw.de/constitutiones/> and description in the poster by Alexander CZMIEL and Mathias LAWO presented at the Digital Diplomatics 2011 conference in Napels: <http://www.cei.lmu.de/digdipl11/posters-rep/czmiel.pdf>. On the geographic representation of charter data in maps see also C. RADL, Die Urkundeneditionen innerhalb der dMGH, *Digitale Diplomatik* 2009, 101–115, pp. 112–114.

extracted from Gunter Vasolds itinerary of archbishop Konrad IV of Salzburg (1291-1312). It connects the issuing places of the charters and the people present at the legal act. It illustrates groups of persons accompanying the archbishop. It demonstrates that visualising doesn't mean only graphics of statistics.[40]

Rules

Creating groups implies building rules, but it doesn't mean that they are explicit. In the field of data mining, computer scientists are working on creating computer tools to establish rules.[41] They haven't been applied to diplomatics research yet. In diplomatics, it is still the researcher who draws the conclusions. The computer could check on consistency of a formally described rule set. It could establish an empirical rule set ("there is no evidence of the existence of a fact in the data"), but it can't do the interpretation as the suggestion of a convincing theory. That converges with the tendency of pre-digital diplomatics to remain narrative and idiographic, i.e. to tell the story of a single document, a single chancery hand or a single phrase. Even chancery studies or research on regional diplomatic traditions are conceived as a collective story.

Identify forgeries

A major motivation for grouping and establishing rules in diplomatics is to find the documents which are singled out by the grouping process. In the concept of "Kanzleimäßigkeit", documents which weren't produced by chancery scribes, thus lacking common paleographical and linguistic features and means of authentication, are suspect of having been forged. Many of these suspicions could be explained by identifying the document as a product from outside the chancery and therefore would have to be supported by arguments from the content and from contextual considerations (interests).

If you consider the groups created with the help of the computer as described above, it is clear that documents not fitting into these groups are suspicious. In terms of computer science that means highlighting docu-

---

[40] The screenshot has been taken from the preliminary website at <http://ginko.uni-graz.at/konrad4/rgraph>. The data is based on G. VASOLD, *Das Itinerar Erzbischof Konrads IV. von Salzburg 1291–1312. Computergestützte Itinerarerstellung und Itineraranalyse*, Schriftenreihe des Instituts für Geschichte 8, Graz 1996.

[41] E.g. with the help of the lattice theory developed by G. BIRKHOFF, *Lattice Theory*, Providence (Rhode Island) 1940, 1948[2], 1967[3], cfr. *General Lattice Theory*, ed. by G. GRÄTZER. Basel 2002[2]; B. GANTER – R. WILLE, *Formal Concept Analysis: Mathematical Foundations* (transl. C. FRANZKE), Heidelberg 2005.
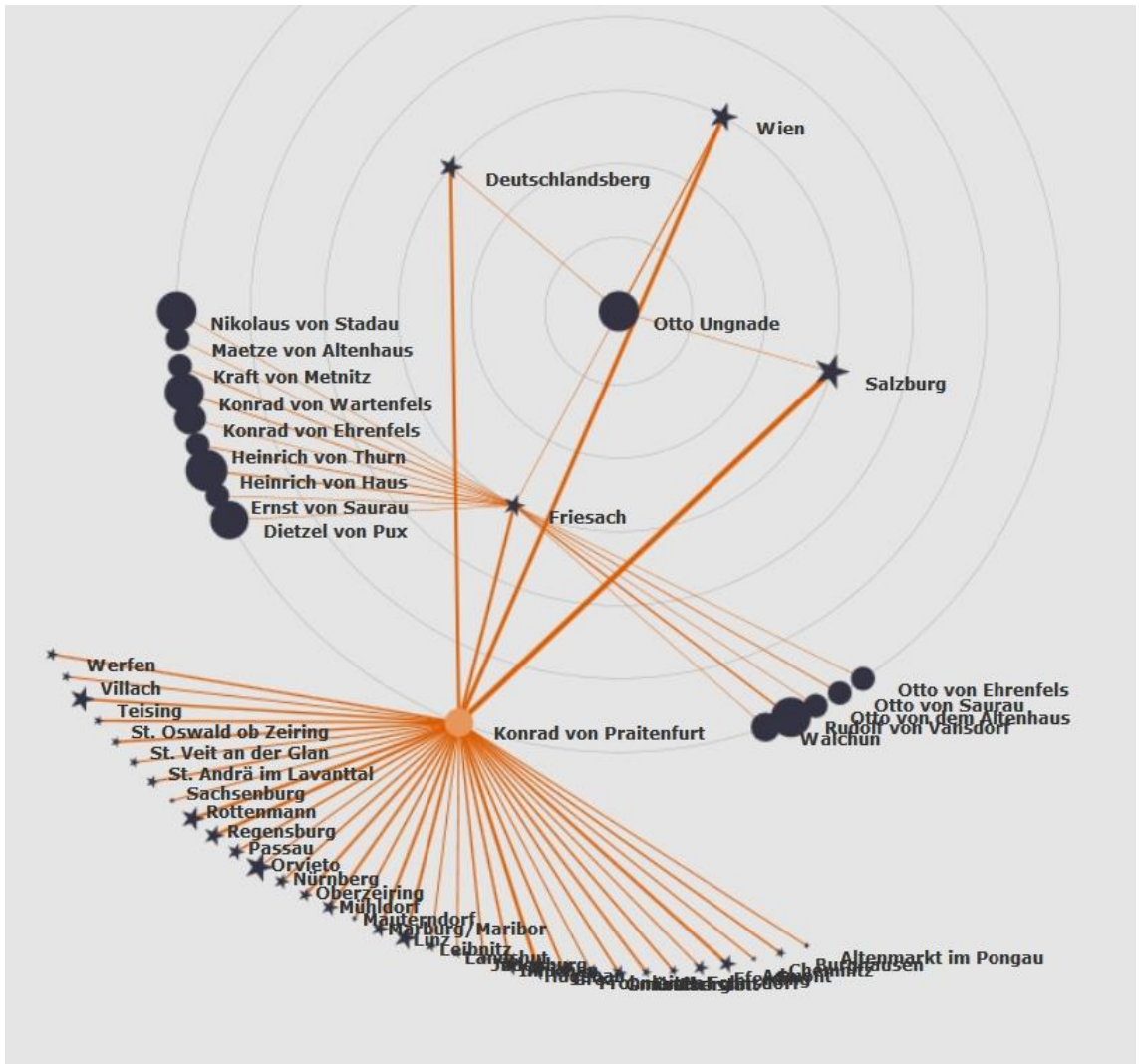
Image 3: Places of charter issuing, naming Konrad of Praitenfurt in his itinerary and the persons named together with him and Otto Ungnade; data from VASOLD, *Itinerar* (see note 40).

ments in the corpus with singular features. Documents produced outside the chancery for the issuer should then carry features found in groups of the corpus not connected to the chancery in question. This automatic approach hasn't been taken so far but Annegret Fiebig already in 2000 suggested the identification of forgeries in the Kuenringer corpus by analyzing the use of punctuation, a suggestion that accords to the paleographical human judgement.[42]

Critical Edition

Much more work has been done to transfer the methods of critical edition into the digital world. The aim of a printed critical edition is to make the probably original text available to future research. That includes that the critical edition encompasses all the relevant things for this text by documenting the textual tradition, by adding diplomatic considerations on the text and referring to the variants in the textual tradition. The critical edition helps using the text in scholarly practice by giving canonical citation, analytical indices and abstracts. In the print world the charters are organised in volumes which form a comprehensive corpus. The user can assume that the critical edition includes all documents by well-defined selection criteria in a meaningful order. These criteria have been transferred to digital editions, certainly in the process of conversion of printed into digital resources.

Born digital scholarly charter edition follow the basic rules of printed critical edition. An example which is very close to the forms deve-loped in the print world is the *Codice diplomatico della Lombardia medievale*, organised by Michele Ansani.[43] The resemblance to a printed edition isn't surprising if one considers that the collection started with the conversion of printed editions into digital texts and continued the work by adding new digital editions. But the CDLM advanced from this by adding information which wasn't part of the printed editions. The text is structured in terms of the contemporary diplomatic discourse as it has been established by diplomatics research. The electronic text of the charters contains therefore 'tags' marking start and end of individual parts of the discourse (e.g. listing 1). This means that his edition has encoded more information than is actually visible. This information can be used for searching and grouping,

---

[42] A. FIEBIG, *Urkundentext. Computergestützte Auswertung deutschsprachiger Urkunden der Kuenringer auf Basis der eXtensible Markup Language (XML)*, Schriften zur südwestdeutschen Landeskunde 33, Leinfelden – Echterdingen 2000, on the forgeries cfr. p. 206–208.

[43] <http://cdlm.unipv.it/>.

i.e. for diplomatic research, either by the search interface or – with much richer search possibilities – by downloading the encoded source of the documents.[44] The idea of encoding the charter text to make it computer readable and add human interpretation into the pure text is a major concept in digital edition.[45]

```
<intitulatio>Johannes episcopus, servus servorum dei</in-
titulatio>, <inscriptio>venerabilibus fratribus Pragensi et
Jauriensi episcopis ac dilecto filio preposito ecclesie Ne-
vnburgensis, Pataviensis diocesis</inscriptio>, <salutatio>
salutem et apostolicam benedictionem</salutatio>. <arenga>
<damage>M</damage>ilitanti ecclesie licet inmeriti dispone-
nte domino presidentes circa curam ecclesiarum et monaste-
riorum omnium solercia reddimur indefessa solliciti, ut iu-
xta debitum pastoralis officii eorum occurramus dispendiis
et profectibus divina cooperante clemencia salubriter in-
tendamus</arenga>.
```

Listing 1: Pope John XXII to the bishops of Prague and Raab, July 4th, 1318, HHStA Wien, AUR
<http://www.mom-ca.uni-koeln.de/mom/AggOCart/1318_VII_04/charter>

The idea of a book as an intellectual work forced the pre-modern editors to select and arrange the charters by consistent criteria. This led to the situation that a document by Emperor Henry II dated December 31, 1014, in the archives of Bergamo in the first half of the 11th century, was edited twice – in the regional selection of *Codice Diplomatico della Lombardia Medievale* as n. 255,[46] and almost one hundred years earlier by Harry Bresslau in the MGH edition of the charters of the Emperor, i.e. a collection by issuer.[47] In an online publication this could be replaced by linking editions from different places. In MOM-CA, the software used for

---

[44] M. ANSANI, Procedure di lavoro: la codifica XML dei testi, *Codice diplomatico della Lombardia medievale*, Pavia 2000 <http://cdlm.unipv.it/progetto/codifica-xml>. The encoded source of the texts isn't linked from the document, but it can be accessed by adding '.xml' to the URL, e.g. <http://cdlm.unipv.it/edizioni/bg/bergamo-pergamene2-1/carte/bgpergg1014-12-31.xml>.

[45] Cfr. P. SAHLE, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, vol. 2, Schriften des Instituts für Dokumentologie und Editorik 8, Norderstedt 2013, 70–76.

[46] *Le pergamene degli archivi di Bergamo aa. 1002–1058*, eds. M. CORTESI – A. PRATESI, Carte medievali bergamasche 2/1, Bergamo 1995; *Curia vescovile – Archivio generale mensa*, ed. C. CARBONETTI VENDITELLI, Pavia 2005, n. 255, online at <http://cdlm.unipv.it/edizioni/bg/bergamo-pergamene2-1/carte/bgpergg1014-12-31>

[47] MGH H.II, n. 293, ed. H. BRESSLAU 1900/03, online at <http://www.mgh.de/dmgh/diplomata/resolving/D_HII_293>

the monasterium.net charters portal, this concept led to the development of a functionality by which documents from several archival fonds can be aggregated into a research collection.[48]

Critical editing is a process. Many of the results in diplomatic analysis as described above are joined together. As existing online documentation of charters often represents only a part of some of these results, a concept of digital edition can be developed that would join the separate results in a kind of "meta edition" created from different sources. Let's take the text of the charter issued by Emperor Henry V in 1108 for Göttweig abbey as an example: The Monumenta Germaniae Historica already published an abstract and the text of the document online at <http://www.mgh.de/ddhv /dhv_38.htm>. The catalogue of facsimiles of charters of European sovereigns by 1200 edited by Irmgard Fees lists in its online version at <http:// www.hgw-online.net/abbildungsverzeichnis/dh-v-38> a print of the document in Sickel, *Monumenta graphica* (1882), and refers to the archival documentation including an image in the monasterium.net portal at <http: //www.mom-ca.uni-koeln.de/mom/AT-StiAG/GoettweigOSB/1108_IX_06 /charter> with indications of several other editions and some diplomatic commentary and identifications of the persons and places in the document. In monasterium.net you can find at <http://www.mom-ca.uni-koeln.de/mom /AT-StAKrems/Krems/StaAKr-0001/charter> a copy of the document. A pseudo code for this kind of aggregation into a virtual document could look like in Listing 2.

```
 1: <edition>
 2:     <regest from="http://dmgh.de/dhv38/abstract" />
 3:     <textualtradition>
 4:           <orig from="http://monasterium/AT-StiAG/Goettwe
                  igOSB/1108_IX_06/witnessOrig" />
 5:           <copy from="http://monasterium/AT-StiAG/Goettwe
                  igOSB/1108_IX_06/witness#Krems" />
 6:     </textualtradition>
 7:     <diplomaticComments from="http://monasterium/AT-StiAG
                  /GoettweigOSB/1108_IX_06/diplomaticAnalysis"/>
 8:     <facsimile from="http://hgw-online.net/abbildungsver
                  zeichnis/dhv_38/facsimile" />
 9:     <text from=" http://dmgh.de/dhv38/tenor" />
10: </edition>
```

Listing 2: pseudo-code of a mash up of the charter of Emperor Henry V for Göttweig abbey, September 6, 1108

---

[48] An example for this aggregation is presented by the collection of charters of the Teutonic Order <http://www.mom-ca.uni-koeln.de/mom/DO/collection>.

Conclusion

At the current state of development, digital diplomatics can change diplomatics in three fields: (1) the methods of documentation of charters, (2) the amount of information that can be taken and processed in a single study and (3) analytical tools.

The methods of documentation are in particular enhanced by the falling price of images. Producing and copying images has technically become so cheap, that whole archival fonds can be published as images. This seems to be a gradual change as photography is a tool for diplomatics study since the 19th century.[49] But it imposed constraints in the diplomatics methodology, which is one reason why it focused very much on philological method. With the cheap reproduction of digital images this can change. When many external features can be better documented by the digital image,[50] verbal description becomes a method of classification. This verbal description will therefore have to rely much more on fixed terminologies than it had in narrative descriptions. Digital representation of diplomatic terminology has already started with the online version of the VID. Its benefits are to be studied.

Digital editing has introduced two new concepts into documentation of diplomatics phenomena: annotation and collaboration. Annotation gives the possibility to introduce code into the very charter text. This makes diplomatic interpretation of the text explicit. Collaboration can make the documentation of charters more efficient. That includes aggregating information from different sources. In a digital environment a new transcription of a charter printed in a 19th century edition is only necessary if it adds new knowledge, e.g. by being more precise or using transcription rules better fitting to new research questions.

---

[49] Which led to the creation of several photographic collections like the Hungarian mentioned above (see note 12). On photographic collections of charters in general see *Fotographische Sammlungen mittelalterlicher Urkunden in Europa. Geschichte, Umfang, Aufbau und Verzeichnungsmethoden der wichtigsten Urkundenfotosammlungen, mit Beiträgen zur EDV-Erfassung von Urkunden und Fotodokumenten*, ed. P. RÜCK, Historische Hilfswissenschaften 1, Sigmaringen 1989. On the discussion about the epistemological value of photography against lithography see L. SANTIFALLER, Über Pause und Photographie und über Sickels Kontroverse mit Pflugk-Harttung, in: TH. VON SICKEL, *Römische Erinnerungen. Nebst ergänzenden Briefen und Aktenstücken*, ed. L. SANTIFALLER, Wien 1947, 92–96.

[50] For the problems of visual representation of manuscript materials see M. TERRAS, Artefacts and Errors. Acknowledging Issues of Representation in the Digital Imaging of Ancient Texts, *Kodikologie und Paläographie im Digitalen Zeitalter* 2, eds. F. FISCHER – Ch. FRITZE – G. VOGELER, Schriften des Instituts für Dokumentologie und Editorik 3, Norderstedt 2010, 43–61.

Digital diplomatics can process huge data sets. A corpus of 9000 documents as in the DEEDS project can be studied as a whole. Statistical methods can become part of diplomatics research. The distribution of linguistic features across time and space already helps to solve problems of dating. Hand identification is supported by statics of image patterns recognisable to the computer. But it is not only statistics. Comparative and exploration studies profit from the easy access to a high amount of documents. As common features in a digital corpus can be extracted – and checked – the comparison between whole *Urkundenlandschaften* is facilitated. Studies like Els de Parmentier's show that the digital representation favours exploration of an unknown corpus rather than comparison of a small collection of known charters.

The computer offers many new analytical tools to the diplomatist: e.g. making quick comparisons of texts with Juxta, selecting images of 'similar' writing hands with GIWIS, finding specific clauses in a corpus with search engines introducing Latin stemming software. But the current state of the discussion tends to be careful about diplomatic interpretation of the outcomes of these tools as far as they make unsupervised diplomatic decisions – an image judged 'similar' by the machine can be similar by characteristics which aren't results of the individual writing hand.[51]

So why call it "digital diplomatics"? Isn't this traditional diplomatics with new tools? The answer to the question depends on some general observations of the history of diplomatics. Firstly, the prevalence of philology can be considered as a consequence of printed culture. As I showed above, digital technology can solve several technological constraints which were part of the research practice in pre-digital diplomatics. These constraints favoured philological questions and a deep description of few documents. In pre-digital diplomatics philological questions acquired prevalence because they could be answered with printed reproductions. The easy reproduction of images, the digital search engines, the possibility to count fast and easily, the display of research results in interactive graphics and all the methods to interact with the charters in their digital reproduction aren't possible in the printed form. Hence, research in diplomatics has to be conceived not only to be presented in the form of a book. Printing is still a possible solution for publishing research results and has many advantages, but it isn't the only one anymore.

Secondly, digital diplomatics can work with thousands of documents. Considering the progress made in the *Papsturkundenwerk* it seems

---

[51] On some problems of automatic palaeographic classification see D. MUZERELLE, À la recherche d'algortihmes experts en écritures médiévales, *Gazette du Livre Médiévale* 56–57 (2011) 5–20.

currently impossible to create a printed critical edition of all papal documents only of the 12th century. Photographing all of the about 18.000 documents documented in Jaffe's *Regesta pontificum*[52] seems to be a small task compared to the 85.000 images online from the Florence State Archives. With digital tools diplomatics can take into account an entire corpus of surviving charters, studying it as a whole but also using it to compare a single document under thorough diplomatic scrutiny to other documents similar by region, time, chancery, content, people involved, mode of conservation, etc. Digital diplomatics provides therefore much more diplomatic practice than the analysis of a single document.

This doesn't mean that digital diplomatics overturns everything that is well-established in diplomatics scholarship. The changing forms of scholarly documentation of charters and the changing methods to work with them break some technological constraints which had become part of the mindset of pre-digital diplomatics research, but it doesn't change the main interest of diplomatists – what can historical documents in their whole, as texts and physical objects, tell us about the people who produced them, kept them and used them? Digital diplomatics tries to understand historical charters better by using the computer.

Георг Фогелер

ДИГИТАЛНА ДИПЛОМАТИКА: ШТА БИ КОМПЈУТЕР МОГАО ДА ПРОМЕНИ У ПРОУЧАВАЊУ СРЕДЊОВЕКОВНИХ ДОКУМЕНАТА?

Резиме

Студија расправља о истраживачким могућностима које компјутер отвара пољу дипломатичких истраживања. Пратећи кораке традиционалног научног рада са средњовековном документарном баштином (хеуристика, документовање, поређење, груписање, успостављање правила, дипломатичка критика, научно издање) оцењују се скорашња дела у којима се дигитална метода примењује на обраду повеља. Доказује се да лак приступ сликама на интернету, могућност процесирања великог

---

[52] *Regesta Pontificium Romanorum. Ab condita ecclesia ad annum post christum natum MCXCVIII*, ed. by Ph. JAFFÉ, 2. ed. rev. by S. LOEWENFELD – F. KALTENBRUNNER – P. EWALD, Leipzig 1885–1888.

броја докумената и ново осмишљавање издања доприносе значајним променама у дипломатици. Покушавају се идентификовати крупне промене на пет поља: (1) класификација пре него репрезентативност одлика је усменог описивања; (2) лак приступ сликама и помоћ софтера за њихову анализу побољшавају визуалну евалуацију; (3) претрага пуног текста, подржана софтвером за природно језичко процесирање, омогућава повезивање појединачних повеља са другима скривеним у великим дипломатичким корпусима, особеним за раздобље позног средњег века; (4) статистичке методе и визуализација великих скупова података помажу да се појединачна повеља сагледа у пуноћи свог савременог контекста; (5) дигитално приређивање омогућава заједничко приказивање резултата рада, који се иначе обављао одвојено, и недвосмислено потврђује знања издавача о документима. Постављајући питање да ли је оваква „дигитална дипломатика" темељно другачија у односу на ону пре-дигиталну, долази се до закључка да она помаже да се истраживање средњовековних докумената ослободи епистемолошких ограничења која су налагала да се резултати хеуристичког процеса представљају у облику штампане књиге.

Кључне речи: дипломатика, компјутер, дигиталне хуманистичке науке, методологија, критичко издање, палеографија.

185