

Virtual German Charter Network:

A Virtual Research Environment for the handling of medieval charters

Daniel Ebner, Jochen Graf, Manfred Thaller

University at Cologne / HK Informationsverarbeitung
Albertus-Magnus-Platz, D-50931 Köln
Germany
{daniel.ebner,jochen.graf,manfred.thaller}@uni-koeln.de

Abstract

Recently the project *Monasterium* (<http://www.monasterium.net/>) has developed into the central collaborative repository for digitized medieval charters from Central European Archives. Based on the experiences with the technical platform for this project, a new project has been started in Germany, which is intended to develop a consistent virtual research environment for charter research based on the experience with the original design of *Monasterium*. Besides presenting this new project, which for the time being is centered in Germany, we also discuss our understanding of what constitutes a “virtual research environment” (VRE) in the Humanities. To support a VRE in the Humanities – or any discipline – we consider it important to have an abstract understanding, a model, of the research process within that discipline. Such a model, understanding editorial work in the Humanities as a layered process of adding preliminary and potentially shifting interpretations upon a basic supply of interpreted sources, is presented and made more concrete by discussing individual layers.

Background I

During the last few years the site <http://www.monasterium.net/> has developed into one of the large collaboratories for medieval source material. It has been started by Dr. Thomas Aigner of the Episcopal archive at St. Pölten, Austria, originally as a resource to make the charters of monastic archives of Austria available in digital form. In the meantime it has grown into an international effort, that brings ca. 80 archives from a dozen European countries together, organized into a non profit organization: <http://www.icar-us.eu/>. Between them, they have made ca. 250.000 medieval charters available, all in the form of digital facsimiles, many of them connected to editorial texts. Part of the environment offered contains an online editor hiding XML, which allows registered users to work collaboratively on these charters, producing XML texts following the Charter Encoding Initiative standard (cf. <http://www.cei.uni-muenchen.de/>), basically a TEI extension, providing the categories of studies used within medieval diplomatics.

Background II

The German National Research Council (*Deutsche Forschungsgemeinschaft – DFG*) has in 2009 asked for proposals for projects to explore the potential of Virtual Research Environments as part of its program for the development of the German research infrastructure. This proposal identified a specific problem of current virtual research environments, worrying funding institutions like the German National Research Council: Particularly in the Humanities it is not always clear, whether Virtual Research Environments actually reach the intended audience, whether they are really accepted and used by the academic community they claim to serve. It has been a precondition of that funding program, that a consortium submitting a proposal would have to consist of a group of memorial institutions, a group of research projects which are using content held in these institutions for concrete academic interests and a group of technology providers,

establishing a research environment, which connects the interests of content holders and researchers. The chair for Humanities Computer Science (*Historisch-Kulturwissenschaftliche Informationsverarbeitung*) at the *Universität zu Köln*, which has for some years developed and maintained the platform behind *Monasterium*, proposed to extend that platform into a Virtual Research Environment. This has been accepted by a group of major German archives, with significant holdings of medieval charters: The *Generaldirektion der Staatlichen Archive Bayerns*, the *Landesarchiv Baden-Württemberg* and the *Landesarchivverwaltung Rheinland-Pfalz*, augmented by a number of important city and regional archives, the *Stadtarchiv Mainz*, the *Stadtarchiv Speyer*, the *Stadtarchiv Worms*, the *Stadtarchiv Würzburg* and the *Archiv des Bistums Speyer*. The consortium has been joined by three research projects connected intimately to the material provided by these archives: The *Forschungsstelle für Vergleichende Ordensgeschichte (FOVOG, Dresden)* uses the charters provided to study the structure of medieval religious orders and monasteries with the main emphasis upon the networks between them; the *Institut für Geschichtliche Landeskunde (IGL, Universität Mainz)* builds research upon the politics of the archbishops of Mainz from Adolf I. until Johann II. von Nassau (1373-1419) on the accessible material; the chair for *Historische Grundwissenschaften und Historische Medienkunde* (Historical Auxiliary Sciences and Historical Media) at the *Universität München* studies the development of the notaries in Germany in the later middle ages.

The consortium submitted a successful proposal for a *Virtuelles deutsches Urkundennetzwerk (VdU)* – a Virtual Network of German Charters. Work on the project started in fall of 2010, a very preliminary website has been created, which at the time of this writing is mainly intended for usage within the consortium (<http://www.vdu.uni-koeln.de>). Discussions about the future usage of this VRE as a potential platform for the systematic presentation of German medieval charters

beyond the archives currently in the consortium have been started. It is also clear, that all tools developed within this project will be made available also for the next version of the platform of Monasterium. At the “Supporting Digital Humanities 2011” conference we present only technical / methodical aspects of the project and touch upon all content related questions of medieval studies only in as far as they are necessary for understanding design decisions.

What is a Virtual Research Environment?

As many concepts, which are directly related to national and international funding programs, a *Virtual Research Environment* is in our opinion not completely clear, for reasons of funding often used for activities which a few years ago would have been described by other concepts, at that time in the focus of research funding. We therefore start with the definition *we* have been using within the project.

We assume, that a VRE is an instance of the eHumanities paradigm, providing a workable eHumanities solution for a clearly defined disciplinary community. eHumanities in turn are derived from the eScience paradigm. For both eScience as well as eHumanities we do not see any universally recognized definitions. Indeed, most representations of the concepts we have seen at conferences tend to be loose or define them by giving examples (cf. http://www.jisc.ac.uk/programme_vre.html). We assume however, that VREs can be discussed with any precision only, if there is an abstract understanding of what components have to be there to make them complete¹.

In the case of eScience not the accessibility of expensive equipment to a community which is not restricted at a specific institution, but defined by abstract criteria is the central point; expensive physics equipment has been shared by researchers from different institutions already in pre-network days, and there is a reason, why the WWW has been developed at CERN. The important criterion is, that the whole research process is embedded into and supported by an integrated environment, which supports all three, the acquisition of information, its analysis and the publication of new information derived from the results of this analysis within one integrated IT environment, which allows the researcher to work without changing the medium for different steps of the research process. For eScience the three steps can easily be identified: Provision of research data by expensive equipment, analysis on high performance computers and publication in data repositories and preprint / open access publication facilities. We note in passing, that in the eHumanities for *some* of the stages an equivalent can be found quite easily: A digital library – particularly when we expand the notion beyond the printed book – is a good Humanities replacement of expensive sensors; when it comes to the publication of research data and research results in the form of papers and monographs on digital platforms, the differences between eScience and

eHumanities (almost) disappear. We notice however, that we are aware of very few instances, where within the eHumanities genuine IT based analytical abilities are provided. An XML editor may be a tool which supports the fixation of an analytical process, but that analytical process is still exclusively based on the Human analytical capability. We can not claim to remove this deficit in our own designs, but we are aware of it and hope to provide at least some possibilities for the inclusion of analytical tools within them.

A concept for VRE supporting research based on manuscript resources.

We understand research based on manuscript resources as a process, by which interpretative assumptions about the meaning of the source can be connected to its digital representation in such a way, that the *assumptions* can be changed, communicated and published, without changing the *representation*. Such assumptions can have completely different conceptual scope: The statement, that a specific letter is an “f” and not an “s” is quite frequently not final within the handling of source and be the valid subject of a critical discussion. The statement, whether one *Ulricus comes* within a charter is the same as in another, is as valid a subject of discussion. The statement, that a portion of text is a quotation from another document, is clearly as valid. On the other hand the techniques to discover the problems behind these discussions or to decide them, both mentally, as when supported by computational tools, are clearly different – and have to be linked into a research environment supporting all of them on different levels.

To support all of them, we use the following layered representation of the research process:

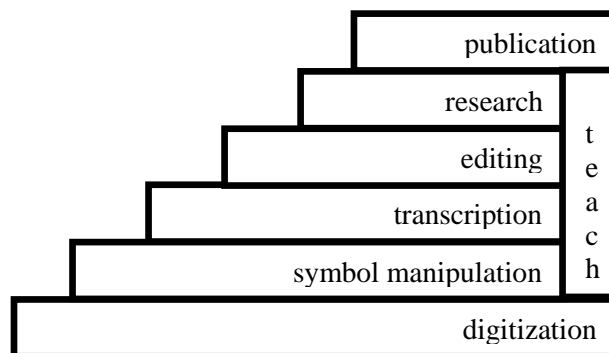


Figure 1: Layer model of research process

We start with a description of the processes and activities we locate within these individual layers; the hierarchy of the five bottom layers (*digitization* through *research*) depicts the degree to which the individual activities within a layer are open for substitution by automated processes. I.e.: While we think, that the process of symbol manipulation – as represented, e.g., by recognizing the forms of letters identifying a specific scribe – can within a relatively short period of time be automated considerably, the research process in a more narrow definition will continue more heavily on unaided human intellectual work. This does *not* imply any statement about the intrinsic intellectual value of an individual activity. Defining the characteristics, by which a scribe can be

¹ <http://www.d4science.eu/vre> claims to give an “explicit abstraction for the Virtual Research Environment“. We fail to see this abstraction and find us in many examples for details.

identified, is as challenging as the process of drawing an abstract conclusion from the content of a document. Once the characteristics of a graphic form *are* identified, it becomes much more easy to apply them automatically to a larger body of documents, however, than it would be to generalize the reasoning behind the structural interpretation of the reality described by one specific textual fragment to a large number of documents.

Digitization

In many ways this layer represents the sensors employed within research environments operating in eScience. The transfer of manuscripts into digital form provides data which can then be processed. To separate this process from the following layers is not as self explanatory as might be assumed, however. In the current practice of the digitization of cultural heritage material, there is frequently a strong emphasis of very high quality metadata as part of the digitization. It is, e.g., known from reports of the German project which works towards the digitization of the complete printed library holdings of the 18th century, that the cataloguing of the holdings is about to become or has already become more expensive than the actual digitization of the books. A collaborative effort, as is usually implied by a VRE, would have to deviate from this. One of the founding reasons for the existence of Monasterium has been, that the charters are digitized with highly optimized workflows and presented “as is”: If high quality finding aids are available in traditional form, they are also made available digitally. If not, only such information gets added as is absolutely necessary – which can be cut down to the assignment of a URN (<http://tools.ietf.org/html/rfc1737>) or a similar identifier establishing a persistent identity of the document in the digital space. This does not preclude considerations of the requirements of later levels. It is, e.g., well understood that the mechanical extraction of forms is supported by digitization following specific characteristics, which are similar to those required for high quality OCR. Digitization is aware of the state of the art of the applications, which can later be applied to the documents; by itself it is agnostic of any intellectual efforts to be supported later by material processed.

Symbol manipulation

Medieval studies have always differentiated between internal characteristics of a document, which are the result of an analysis of the language used within it, and the external ones, which are represented by material, form of writing and other properties observable (in principle) even by somebody who is unable to understand the language of the document. This level represents what is traditionally the domain of *paleography*: Identifying scribes or drawing conclusions about the temporal and spatial origin of a document from the significant forms of the letters used. Digital technology allows an extension of this concept: While history of art has a tradition of looking at the characteristics of illuminated characters and other embedded visual features of manuscripts, such studies are usually not seen as paleography. In as far, as art history is able, however, to define the rules for distinguishing between different styles of visual embellishments of a manuscript in a sufficiently strict way to allow a program to find such a form within a body of raw digital images,

the borders between the two fields get blurred, however. The same is, of course, also true for all tools which do not intend to automate the recognition of shapes or writing styles defined as a family of shapes, but try to support the process by which human recognition of such features shall be supported. And we may point out that one of the research projects to be supported by the VRE we are building, is interested in the history of notaries – which identify themselves by graphic symbols, which cannot directly be related to a textual representation.

Transcription

Not everybody reading a document intends to transcribe it; nevertheless we consider this a layer of work with manuscripts which is present in any research process. Reading is a precondition of a transcription and the transcription itself is the product of reading a document sufficiently carefully to be able to represent it in another form, which can more easily understood by others. The difference between this layer and the following one is, that we assume it to be focused mainly on information contained within the document we are working upon. While other documents in the same or similar hand writing can shed light upon the correct reading of the document currently being transcribed, the ultimate decision of the correct transcription of a specific graphic form has to be based upon the context in which it appears. Talking about *transcribing* – not *transliterating* – on the other hand means, that we have to be aware of forms of writing within a specific body of material. The Icelandic manuscripts, reputed to consist of up to 20 percent of abbreviations, come to mind. Resolving such abbreviations correctly requires access to the lists of understood ones, which goes beyond the scope of an individual manuscript – the lists of previously identified abbreviations and the shapes of them which have been observed operate still on a mechanic level, however, while the choice of the correct resolution of a shape into the long form of the abbreviated text relies on the string of other characters within which the shape of an abbreviation has been identified. So looking at the start of the letter ‘m’ in Cappelli’s *Lexicon Abbreviatarum* (1928) we see that a frequent abbreviation for *magister* is virtually indistinguishable from that for *mandamus*. Which is the correct one, only the linguistic context can clarify-

Editing

Editing in our definition describes such research activities, which can in no way be confined to the handling of an individual document. “Abbreviations” are a good example. Whether a squashed letter “m” stands for *magister* or *mandamus* can be decided within, and only within, the context of a specific handwritten page. To which specific well known philosophic conclusion the phrase “*Ergo etc.*”² refers within a medieval treatise on philosophy can never be decided by reading this individual document. It requires access to a considerable body of further texts, which repeat the same premises as

² Siger of Brabant, *Quaestiones in tertium De anima*, more than hundred times simply writes *Quare etc.* when the next step of an intellectual argument is obvious (for him and, hopefully, his contemporary readers). Assuming that more people understand *ergo* than *quare* we have modified the quote.

the original document and give the conclusion in more detail. In the context of charters “editing” in our definition refers to all activities which require a comparison with the context of at the very least other charters, usually also with other documents of the same period. This can refer to stylistic considerations, e.g. the question, whether the *arenga* of a specific charter shows influences from that of another issuer or, more mundanely: Whether an individual named within one charter is identical to an individual within another; to which modern location a topographical term used in a medieval document refers to; to which calendar date the day of a saint quoted for dating purposes refers to in the dioceses in which the charter has been issued; whether the seal affixed to the charter shows a motif also contained on other charters of the same spatiotemporal frame of reference; to what value a monetary amount specified within a document should be converted in the same frame of reference; whether a document quoted within another one has survived; which other charters have been testated by the same chancellery etc.

Research

Within editorial activities the document, its reconstruction within a potentially corrupted tradition, the resolution of unique phrases stays in the centre of the intellectual exercise. Colloquially this is certainly part of the research process. We consider *research* in the narrower sense to be comprised of all activities, however, which deal with abstractions which go beyond the reconstruction and interpretation of the text contained within one specific document. Identifying a charter as relating to an individual monastery, though that monastery is not mentioned explicitly by its name, but by some phrase, which can be connected to it based upon the content of another charter, is an *editorial* activity within our model. Interpreting the fact, that this charter documents a gift of a specific piece of land as an expression of a ruler’s policy to strengthen his power base by relying more strongly on clerics, is a *research* activity within this narrower sense of our model.

Publication

This layer represents on the one hand the obvious: Making results obtained by a researcher from the documents contained within the VRE available, as digital publication or as utterance in some communication system – mailing list, blog, etc. We assume however, that this layer of a VRE should embrace a broader sense of publishing: We subsume here also the necessity to provide access to the content on all levels of the research process from other information systems. These can be other VREs or, e.g., digital libraries, which present to the non specialist a snapshot of the documents as they have currently been prepared by the research community using the VRE. We could speak about export functions, rather than a publication layer; we do not, however, as we assume that catering for the needs of a closely supported research community will lead to tools – and representations of the content – which are not necessarily attractive to wider communities. So publication, both for Human readers of the information, as well as for other information systems processing the information further for the needs of the communities they server, involves a process of packaging

which is different from the simple provision of the content as is.

Teach

Teaching in our model could be described as a specific form of publication. That is, we assume that a proper VRE in the Humanities creates links to learning tools at all levels from symbol manipulation through to research. This can be realized by creating interfaces between the resources provided within the content administered, but also by integrating access to a learning module into the environment of the VRE. We notice at the same time, that for university teaching the possibility to access a system geared towards collaboration leaves a very blurred line between teaching and research to begin with. Within Monasterium, three universities are currently using the environment to train students in the handling of medieval charters. As the result of their coursework can be released as a contribution towards the transcriptions, descriptions and encoded texts of Monasterium, if the quality is deemed sufficient, they are actually being taught by inclusion into the research process, the “teaching interface” being the academic teacher.

A concept for collaboration

The Virtual Network of German Charters inherits the model of cooperation which has been used within Monasterium. Monasterium has started as project to make charters available, which are being held within small archives of monasteries, some of them in rather remote areas. The original idea of the founder and originator of the project, Dr. Thomas Aigner director of the *Diözesanarchiv St. Pölten*, has been to use the internet to make these holdings more easily accessible. This he has accomplished admirably by two efforts: To provide affordable quality digitization for archives which are too small to organize digitization campaigns themselves, and to integrate such documentation of these charters as has been available in local catalogues and cartularies into a database. When a few years ago the chair for Humanities Computer Science at the *Universität zu Köln* has been invited to renew and enhance the technical infrastructure of the project, the first target has been to attract specialists to transcribe, describe and edit the many charters, for which no or only little searchable information has been available before, only digital images of the charters being included into Monasterium. As there are millions of such charters in Europe and a very large number of the charters within Monasterium also have only very little searchable information, it is absolutely clear, that it is totally impossible to plan for a projects which would provide a few hundred thousand person years of academic staff to provide digital editions according to the state of the art. Even if the model of Monasterium has been so attractive, that the majority of the holdings do not come from the original small monastic archives any more, but from the big collections of the National Archives of the central European states.

The only viable way has been to provide an interface, which allows volunteers to work with the charters, transcribing the ones existing in the data base only as images, inserting appropriate markup into digitized versions of older cartularies etc. Such volunteers are

available – the best specialist for the charters of the institutions, both clerical as well as secular, of a region outside the big centers of medieval power may still very well be a teacher at a local gymnasium or, indeed, the archivist of a local institution. And, as we have mentioned in the context of teaching, creating transcriptions of charters and marking them up according to the rules of diplomatics, the branch of historical studies devoted to *diplomata* or charters, can be a very valuable part of academic training. Neither gymnasium teachers, nor distinguished local archivists and not even students specializing in diplomatics, have a very close affinity to markup theory or XML. It was clear from the very beginning, therefore, that that kind of volunteers could only be recruited when they had an interface available, which would allow them to produce markup without being aware of it. So the first contribution of the Cologne outfit, when taking responsibility for the technical infrastructure of Monasterium has been the creation of an editor, which allows to transcribe a charter which is displayed on the screen and insert markup into that charter by simply highlighting a portion of the text and select the categories applicable to that. Such an editor has been created with three goals in mind: (a) It should of course handle context sensitivity, allow the selection of only such tags, which are legal at a specific point, considering the markup inserted already; (b) it should allow for “partial markup”, highlighting and marking such terms as are important for the most general purposes, basically topographic terms and proper nouns; (c) it should also support markup for the full range of diplomatics. For the later purpose we could fortunately use the *Charter Encoding Initiative (CEI)* (<http://www.cei.uni-muenchen.de/>) schema proposed and maintained by Dr. Georg Vogeler, than at the University of Munich, currently at the University of Graz. The CEI encoding scheme can be understood as a layer of semantic tags for the specific conceptual categories used by diplomatics, embedding a specific subset of the TEI.

A general statement on that, as we see this as an important and occasionally overlooked principle: Virtual Research Environments for the Humanities should not focus on teaching Humanities’ scholars the joy of XML. They should provide tools, which make the technological advantages of XML available for the community of the Humanities, without the later having to bother about its complexities.

Collaboration in the case of Monasterium means to open the field to people at many institutions and actively recruit volunteers outside of the archival or historical profession. There are many of these, which are willing to contribute content of high quality. Nevertheless, a Virtual Research Environment for charters, should give the implicit guarantees for the quality of its content, which is associated with traditional archival finding aids and printed publications on diplomatics. For Monasterium this has been realized by implementing a model of moderated contributions. For this a user of the system can claim a small number of charters to work on them exclusively for a short period and edit a private copy of their descriptive data. When they have finished their work, they can select a moderator who will check it for academic quality and

release it into the public area of the system (or refuse to do so).

Implementation and technical issues

Underlying technology

Monasterium is based on an eXist (<http://www.exist-db.org/>) data base; this will continue to be the case. Indeed one should note, that the platform for Monasterium and the platform for the German Virtual Charter Network will remain closely related. The first version of the German Virtual Charter Network has been built upon the platform used for Monasterium; when the expanded platform developed for the VRE, Monasterium will be transferred to a copy of the improved platform. The differences are mainly, that Monasterium has started with an architecture, where eXist as a data base engine has been augmented by a Java applet based editor, which now is slowly aging and creates a bottleneck by its long loading time. The new VRE is based on eXist and an XForms (<http://www.w3.org/TR/xforms/>) oriented interface, particularly for the new version of the editor, based on the betterForm (<http://www.betterform.de/>) implementation. As for the charter network mixed content would be necessary for the XForms which they do not support, an extension of the XForm standard for that purpose has been developed within the project by Daniel Ebner³. This extension runs rather smoothly: As browsers so far do not support XForms directly, as the W3C intended, betterForms essentially translates them into a Ajax/JavaScript set of classes, so on the implementation level the extension adds to these classes. The system is therefore based upon the XQuery, REST, XForms (XRX) architecture (De Jonge, 2010; <http://en.wikibooks.org/wiki/XRX>)⁴. An Atom (<http://en.wikibooks.org/wiki/XRX>) based interface is already realized as technical baseline for the publication concept described in the conceptual section of this paper.

Participating archives are offered server space for their charters, quite frequently hold the images on their own servers, though. The metadata repository is currently being linked to a preservation repository of the University at Cologne.

Generalizing the Data Model and its support

Monasterium, as has been described, uses the CEI Encoding scheme for the individual charters. The charters are usually collected in subsets within the data base, which reflect individual archives. These collections are structured by machine readable versions of such finding aids as exist within the archives, which are replaced by a skeleton version of a finding aid if none exists. In some cases existing printed cartularies have taken the place of a finding aid. These structures of finding aids or cartularies into which the encoded descriptions of the charters have been embedded, have been encoded within a not very well defined set of rules for direct HTML encoding. The new VRE replaces these HTML based introductions to the

³ This is part of the thesis project of Daniel Ebner.

⁴ Within the project Jochen Graf is responsible for the concrete overall architecture.

individual collections by *Encoded Archival Description (EAD)* (<http://www.loc.gov/ead/>) encoded descriptions of the finding aids themselves. Leaving aside the question on how exactly these levels of encoding are represented in the data base, we can describe the encoding used henceforth by an outer envelop of EAD for the archival level, into which CEI encoded charters are embedded, which in turn draw upon the TEI for features not related to the specifics of diplomatics descriptions. So overall the VdU uses an “onion model” of markup schemes as shown in figure 1.

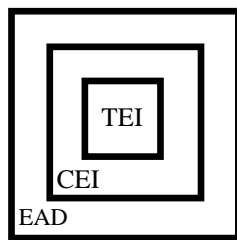


Figure 2: Onion model of markup schemes

To support that markup scheme, the architecture of the editor used within the VRE has changed considerably. While the original one has been hardcoded for the CEI scheme in Java, the new one has been designed from the start to allow the support of any structure described by an XML Schema, so on the editor level support of a different schema would be quite simple – which is important, if the range of sources covered by the VRE is extended beyond medieval charters. The CEI, effectively the layer implementing the semantics of the class of sources administered, would have to be exchanged for a markup scheme representing the semantics of a different class of sources. In this sense the new editor, or rather: Its underlying technique, the extension classes to betterForms, is reusable. A word of caution, though: This does not remove the need to adapt the data base structure feeding the editor.

Generalizing the Model of Collaboration

To adapt the system better to the needs of a larger number of users, as well as to a more clearly defined relationship between holdings of an individual archive administered in the system, the simple “user – moderator” model is currently extended towards a four-layered rights model with an orthogonal extension.

Users continue to have the right to work exclusively on a small number of charters for a restricted period of time, working during that time on a copy of the reserved charters administered under their userid.

Moderators continue to have the right to check the quality of the work of a user, before it is transferred into the public space. They can define, whether they are willing to act in that capacity for all users of the VRE or only for a defined group of users (their students, e.g.). Users will in the future have the right to appeal against a negative decision to another moderator. While it is not clear, whether that can be realized under current funding, we furthermore plan to provide possibilities for the coexistence of differently marked up charters, which have been authorized by different moderators, to allow for differences in matters of intellectual interpretation.

Archivists have additional rights to handle restrictions of specific groups of documents, granting the right to work on them or decide about the quality of such work only to specific users and moderators. For the holdings originating from their archive they are always also acting as moderators.

Administrators have the right to create users, moderators and archivists and to delegate the right to create users of users to moderators and archivists.

Independently of this hierarchy, *every* user has the right to use the system for “private” versions of charters. These remain accessible only to him or her, this private work does *not* imply that other users would not be allowed to work on the same charters in the regular user – moderator relationship, however. Additionally a user can administer private notes on charters in an unrestricted full text format. Private versions of transcriptions, descriptions and mark up of a charter, together with notes relating to them, form a private view, which is exempt from the moderation process. The components of a private view can be exported into appropriate XML formats as well as PDF files. Appropriate components of a private view can be made public, if a moderator agrees to it. A specific contribution which has been made public previously, can be withdrawn by a user from public visibility. If a user is removed from the system, copies of all his private views existing at the time are created and transferred to his last known contact address.

Concrete tools offered under the defined conceptual layers

As we assume, that our attempts at analyzing the research process to be supported by a VRE and identifying layers, which can be targeted by clearly defined tools, are more easily transferable to other VREs than the actual tools offered, we have desisted from specifying concrete practical applications when defining the conceptual layers. A definition of the tools intended for the individual layers follows below. As these are quite numerous, we should clearly differentiate, what will be attempted within the funding provided by for the *Virtuelle deutsche Urkundennetzwerk* and where we talk about interfaces into tools which may exist already, or which are being developed by others. The core development consists of what has been described above: A flexible system for holding the documents and working on them, operating under a clear rights model, with well defined possibilities to export parts of the content. This is the main focus of the development work and it defines what is available for the members of the consortium now and will become publicly available as a first release soon after the publication of this paper. In all the other cases mentioned below, we are strictly working under a “principle of subsidiarity”. When the team developing the VRE becomes aware of an existing tool, which fulfills one of the functions below, it will try to embed it or link to it; if not, it will try to develop its own solution. In some cases we illustrate the functions described below by referring to projects, which have implemented functions like the one we describe. As we have only recently brought the redesign of the platform to a stage, where practical interfaces can be discussed, these references should not be interpreted as indications of well advanced contacts, however.

Generally speaking the “vision” of the technical work on the VRE realizing the *Virtuelle deutsche Urkundennetzwerk* can be expressed as follows:

- (1) Create a platform, where as many charters can be brought together, as institutions are willing to contribute.
- (2) Connect them to as many tools, as are fit to be connected.

Digitization

Digitization is understood to be a preparatory process to the introduction of sources into the VRE, so it is not part of the tools described here. Entering data into the VRE is usually done by converting finding aids from contributing archives; within Monasterium this is implemented by import procedures negotiated individually with contributing institutions (their individuality being actually one of the most significant bottlenecks for further expansion). With the current archival partners of VdU a set of import procedures has been defined, which allows the automatic import of EAD encoded finding aids. There is no broad standard used within existing systems to administer electronic descriptions and content of individual charters (starting with OCRed cartularies and going through Excel based catalogues created over the years all the way to proprietary and mutually exclusive systems of vendors of software for the administration of collections). So the import of digitized resources currently represents a significant drain on the resources of any VRE supporting charters. We hope, of course, that the facilities of the VdU to edit EAD conforming finding aids interactively on the Web will remove or reduce that drain on individually designed import procedures.

What the VdU *does* support for the integration of holdings, is the assignment of URNs to the important charters which guarantee the possibility to quote them permanently. The VRE strictly maintains a concordance between these assigned URNs and the shelf marks or call signed used within the originating archives. Within research there is a long tradition of quoting archive-held documents by their call marks. So providing an easy way to follow a quotation within a printed book to the digital representation of the document quoted is a design principle for the VdU, by making these *canonical references* directly addressable.

Symbol manipulation

On this level tools for basic image improvement are under preparation. The major challenge at the moment consists in finding tools, which allow embedding enhancement functions in such a way, that the users can select them on the basis of descriptions, which are related to enhancement problems typical for digitized charters, without forcing them to understand the criteria for the selection of a specific image enhancement method for a specific problem. A similar interface, pointing to methods particularly appropriate for problems related to the legibility of seals, is provided in parallel. A major goal of the next stage of implementation is a tool for the creation of “symbol catalogues”, originally for the private view of individual users. Roughly speaking these represent significant alphabets of individual scribes, connected to tools which allow the reduction of selected characters /

symbols to their basic graphic shape and providing support for the machine supported comparison of such characters and shapes. This is one of the areas, where we hope to gain most from connecting to existing tools, concretely the ones for the support of paleographic work presented at recent conferences (Rehbein 2009; Fischer 2010).

Transcription

In the strict sense, this is the domain of the editor being part of the core system of the VRE. The main additions here deal with four areas:

- (a) On a very mundane level connections to existing digital versions of traditional tools of the disciplines, like Cappelli (e.g. http://inkunabeln.ub.uni-koeln.de/vdibProduction/handapparat/nachs_w/cappelli/cappelli.html), will be created.
- (b) On the other hand support for linking individual sections of the transcribed charters to parts of the marked up text are examined. It remains to be seen, whether any of the tools recently being presented by other projects can be reused outside of their context.
- (c) Furthermore, we are examining better support to cartularies which have been converted by OCR into machine readable form outside of the project described here. The chair for Humanities Computer Science currently uses the funding received from a Google Humanities Award for that purpose.
- (d) Finally we watch with great interest the work produced by MONK (<http://monk.target.rug.nl/>), i.e., the possibilities to create indexed searching for forms graphically similar to one extracted from a document within a collection of digitized ones. Again the possibility to use the development outside of its original context may be the biggest problem.

Editing

For editorial work in the sense defined above we mainly plan for direct and convenient interfaces into existing tools for the identification of standardized names – proper nouns and topographical references, as provided by the authorities files promoted mainly by the library communities (e.g. <http://www.loc.gov/catdir/pcc/naco/>). Fortunately recently very promising systems for such have been started also within the context of infra structural institutions of the Humanities, e.g. the iMGH (<http://www.mgh.de/dmgh/imgh/>) of the *Monumenta Germaniae Historica* drawing upon the extensive knowledge about historical topographical terminology represented within its extensive series of source editions. More challenging are tools for the search for complex phrases, which help in the identification of similar *formulae*, given the vagaries of medieval orthography and grammar. Here tests with tools developed for other projects of the chair for Historical Computer Science at Cologne are currently designed. In the context of the VD18 (<http://vd18-prot.bibliothek.uni-halle.de/>) project we have been responsible for trying to find titles of books, which in the data bases to be consulted have occasionally been entered in full length, occasionally abbreviated to a fraction of their length and in some cases been distributed across a number of data base categories. As they have been transcribed in various degrees of faithfulness to the

original – i/j, u/v etc. – the variance between representations of the same title is high. In work not yet published a combination of indexed access based on an indexing mechanism allowing for spelling variation with a comparison of candidate titles by Levenshtein-based (1966) algorithms has been very successful. The main problem to apply these results to the charters is, that titles have a very clear start; formulaic components of a charter have so only after diplomatics markup has been added, which reduces the number of documents significantly.

Research

As mentioned initially, we see the question of truly analytic support of Humanities Research within VREs as one of the generally difficult question. So what we have to offer initially consists mainly of the possibility to attach content related keywords in a user specified number of criteria to individual documents, which allow the definition of private de-facto collections of charters. We are contemplating the possibility to expand these towards the creation of private ontologies connecting the charters. But, as we have mentioned before, we do not think that teaching Humanities scholars XML should be the point of a VRE, we think neither the teaching of formalisms of ontologies should be so. So far, though, we have not discovered a convincing model, where these formalities could be hidden behind a intuitive metaphor.

We are very proud to announce however, that in the case of the statistics modules of DEEDS (<http://www.utoronto.ca/deeds/>) preliminary contacts towards an integration of such modules into the VdU environment have been made.

Publication

This is the field, where the work plans are particularly concrete, and the creation of “publication” facilities is one of the main focal points of the phase of development just starting. We are currently working on a PDF export facility for charter descriptions, following the traditional formatting for cartulary entries. An export facility for the charter descriptions within either RTF or OOXML (<http://www.ecma-international.org/publications/standards/Ecma-376.htm>) or ODF (<http://www.oasis-open.org/committees/download.php/12572/OpenDocument-v1.0-os.pdf>) is being analyzed, may be dropped as too complex at the moment, however. There is obviously a conversion tool for presenting the XML encoded descriptions in HTML existing already, which HTML encodes them for the WWW. We try to make these independent of the context of the current user interface, so these HTML encodings can be exported. XSLT filters to extract from the “onion model of markup” described above into pure EAD or CEI or TEI are planned. This is connected to the current work on the implementation of an OAI PMH (<http://www.openarchives.org/OAI/openarchivesprotocol.html>) interface, which, however, will first be restricted to DC (<http://dublincore.org/>) and EDM (<http://www.europeanlabs.eu/attachment/wiki/WP1CommunityMeetingArchives/EDM%20v5.1-100406.pdf>) metadata. We are considering publication facilities for private notes into Excel sheets; as these are depending on the private notes mechanism, which has not

been realized yet, this will probably not be realized within the current stage of work.

All these levels of publishing components of individual charters will contain references to the URN of the charter as well as to its canonical form of quotation.

Teach

No concrete interfaces to teaching systems are planned for the currently guaranteed funding, with the exception of a German system for the teaching of paleography (<http://www.palaeographie-online.de/>).

Acknowledgments

The funding for the work described in this paper has been granted by the German Research Council, DFG, under its funding line *Virtuelle Forschungsumgebungen*.

References

- Cappelli, A. (1928). *Lexicon Abbreviaturum*. Leipzig: J.J. Weber.
- De Jonge, A. (2010). *XXR Using Xpath 2.0*. Apress.
- Fischer, F. et al. (edd.) (2010). *Kodikologie und Paläographie im digitalen Zeitalter 2 / Codicology and Palaeography in the Digital Age 2*. Norderstedt: Books on Demand, 2010. (available from: http://kups.ub.uni-koeln.de/4337/1/kpdz2online_gesamt.pdf)
- Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet Physics Doklady*, 10(8) 707–710.
- Rehbein, M. et al. (edd.) (2009). *Kodikologie und Paläographie im digitalen Zeitalter / Codicology and Palaeography in the Digital Age*. Norderstedt: Books on Demand, 2009. (available from: <http://kups.ub.uni-koeln.de/2939/1/kpdz%2DOnlineFinal.pdf>)

All URLs quoted in the text have been accessed on October 28th 2011.