ISSN 0204–2061. KNYGOTYRA. 2008. 51

DIGITIZATION IS NOT ONLY MAKING IMAGES: MANUSCRIPT STUDIES AND DIGITAL PROCESSING OF MANUSCRIPTS

ZDENĚK UHLÍŘ

National Library of the Czech Republic
Klementinum 190, 110 00 Praha 1
E-mail: Zdenek.Uhlir@nkp.cz

Author deals with the link between the digital processing of historical documents, especially manuscripts and the manuscript studies, codicology and bibliology and cultural history as well. The greatest part of the paper applies to the case study about the Manuscriptorium digital library which is provided by the National Library of the Czech Republic.

Key words: digitization, manuscript studies, National Library of the Czech Republic.

INTRODUCTION

Digitization of historical documents and/or holdings has been in progress for approximately fifteen or twenty years, since about the end of eighties or the beginning of nineties of the twentieth century. At the earliest, during the first half of nineties digitization meant creating a surrogate, an alternative carrier, something like "better microfilm". Goal of such a digitization was very simple, namely to be a preservation aid. Up to this day for some people digitization still counts to preservation. These people do not understand the spirit of the information, the knowledge of/and society and do not see the challenge of the information and communication technologies that comes with it.

It is a question whether digitized historical documents are the preservation aids because they are also promotion aids. Such a "digital promotion" of historical documents addresses not to specialists but more likely to a general public that is neither interested in historical nor similar studies but in a mere information about the past. Briefly the general public does not want to "consult" historical document, to study its internal and/or external features but it prefers to see a historical document as a thing, as a physical object, that illustrates the more or less known past instead. General public is not interested in a sophisticated difference between "the past" and "the history" (*res gestae* and *historia rerum gestarum* using Hegelian words), so it is not interested in digital cop-

ies; a digital copy is for such a public simple caption, a sign for something else, i.e. for exhibition of original document. Thus, such an understanding of digitization goes in a vicious circle: the supposed objective was to exclude originals from lending, on the other hand the factual result is a more massive lending of originals. Thus, this primeval understanding of digitization results in an internal discrepancy.

A wide expansion of the Internet in the mid nineties was a great challenge for digitization of historical documents and/or holdings. At the earliest, it was not accepted well by the academic community because of suspicion of commercialization, at least in most European countries; sometimes it was understood negatively as an "American invention". In the first years of its spread in Europe, Internet was accepted positively mostly by professional communities and some memory institutions (especially by libraries and librarians in the Czechlands). A keyword of that time was not "preservation" yet but "democratization". It had to mean that keepers, i.e. memory institution professionals ought not to make borders for any interested person, no matter if experts or the general public. The Internet presentation of a historical document ought to be accessible for everybody and particularly for "vulgarians" (*hommes de la rue*, *gemeine Menschen*). Of course, it was another illusion because "vulgarians" are interested in other things than in digital copies of historical documents, medieval manuscripts etc. On the other hand, intellectuals although not experts in historical documents could be interested in that. Again public worldwide could be and is interested in it. Thus, "democratization" means global accessibility independently on a social, cultural, and national environment but dependently on education and/or erudition. And a global accessibility under another aspect means a free access; not necessarily free of charge but absolutely free of any discriminatory balk.

Thus, digitization means a global dissemination of historical documents and/or holdings, a challenge for inter-cultural studies in general and an impulse for manuscript studies in particular. Consequently, digitization is not simply making images; it is and must be much more. Perhaps it is not seen when digitizing small number of manuscripts, several units or at the utmost several tens. It is however seen after digitizing critical mass of manuscripts, i.e. perhaps a few hundred. Then, simple images, i.e. image sequences are unmanageable for the end user as well as for the information system administrator. Compound digital document/s that make relations between data (digital images) and metadata (descriptive catalogue record, structural and technical metadata as well) must be created. And it is only the first step, only the simplest form of the compound digital document, of course. After overstepping the critical mass of compound digitized documents there must be created a sophisticated information system that enables orientation, e.g. showing both the whole database and partial collections, as well as navigation,

e.g. enabling not only standard full text search but also combined search using operators, expert search using filters, search implementing graphical variants etc. Further, more complex form of the compound digital document may contain also full texts that should be correlated according to the appropriate parts or passages to the digital images. It may contain several full texts (e.g. edition of the original text and its translation) that should be correlated to each other and both full texts to the digital images. It may contain also audio documents that should be correlated to the full texts and digital images as well etc. Not only editions of original historical, i.e. primary documents and their translations but also secondary documents, i.e. documents about the primary documents may be integrated into the compound digital documents. And there are possibilities that can be multiplied practically *ad infinitum*. It is clear that such a digitization of historical documents, or let us say manuscripts, means a complex digital processing of manuscripts and that it is very important for manuscript studies, codicology, bibliography etc.

THEORETICAL SUPPOSITIONS

Accordingly, digital processing of manuscripts is about a paradigm shift. It can be understood in triple way: firstly in relation to the historical auxiliary sciences (*historische Hilfswissenschaften*), respectively in relation to the so-called quantitative codicology in opposite to the traditional archeology of book; secondly in relation to the historical methodology of collective, mass, aggregate, wholesale phenomena against the individual ones; and thirdly in relation to the so-called pragmatical edition in opposition to the traditional critical and/or semi-critical edition. Of course, such a paradigm shift is quite a long run rather than a fast and brief occurrance, it is "scientific revolution" in Kuhnean words. Therefore digitization of historical documents and/or holdings is currently an activity in advance operations of which are both routine and under research and development at present. Traditional ways of representation historical documents and/or holdings consisted of complex and sophisticated descriptions utilizing terminology with very hard semantic reduction that passes by natural language/s and that not always distinguishes between things or objects on one side and ideas, notions, concepts etc. on the other. Paradoxically this traditional terminology is based on vernacular, i.e. substantially natural languages and simultaneously deprecates its own basis, such as translation between linguistic utterances and consequently national discourses is sometimes very difficult. Digitization by making more or less accurate copies of historical documents enables to overstep some of these crucial problems. Again, digitization facilitates a goal-directed navigation and control and management of a huge amount of data by preparing electronic full texts. It is impossible through traditional ways.

Some technical conditions are necessary for the paradigm shift, namely dividing

data from software, standardization of data, and interoperability of tools and systems. Firstly, dividing data from software, i.e. mutual independence of data and software is a substantial condition that enables the other two ones. Digital, i.e. Internet environment is principally heterogeneous, not homogeneous, the tools and systems are various, indeed different, so that sharing data among tools and systems could be a problem if the data were software, i.e. tool and/or system dependent. Such software, tools and/or systems could not import and/or export data from one another and they would not interoperate. Problems could arise also by using communication protocols because of a need to implement profiles according to the regular, i.e. open, non proprietary, not software dependent data standards. Thus, dividing data from software is *conditio sine qua non* for the digital representation of historical documents and/or holdings.

Secondly, standardization of data is a necessary condition for applying manuscript studies, codicology and bibliography in the digital, i.e. Internet environment. Standardization means to use open, not proprietary standards for data preparation/creation, respectively to use regular, ordinary, conventional formats for data exchange. In other words, standardization means to use various software tools for creating the same data output[1] as for the form and for achieving the same result as for the content. Thus, for data creator a principle of standardization means on one hand a flexibility in the use of software tools, editors, processors etc., on the other hand a variety of choices in the information depth; and when using markup languages represents an inner structure of document, the principle of standardization means to use various manners of concrete application of markup and consequently to represent various information and information levels from the same primary evidence. Thus, standardization of data is the middle of how to make data readable for machines and generally understandable for humans.

Thirdly, interoperability of tools and systems consists of inter-tool and/or inter-system communication, i.e. in importing and/or exporting data, in data mining, harvesting etc. It is a principal condition for working complex modular systems, for cooperation of their inner components and for cooperation with external tools as well. Is particularly important to guarantee

[1] See various XML editors for creating descriptive catalogue records and fulltexts as well, e.g. jEdit, access through Internet: <http://www.jedit.org/>; Peter´s XML editor, access through Internet: <http://www.iol.ie/~pxe/>; OpenOffice in combination with SourceForge, access through Internet: <http://www.openoffice.cz/stahnout> and <http://www.openoffice.cz/stahnout>; GNU Emacs, access through Internet: <http://www.gnu.org/software/emacs/emacs.html>; NoteTabLight, access through Internet: <http://www.webmasterfree.com/NoteTab_Light__d7660.html>. See also special software tools, e.g. MEdit, access through Internet: <http://www.manuscriptorium.com/Site/ENG/medit_eng.asp>; MTool, access through Internet: <http://www.manuscriptorium.com/Site/ENG/mtool_eng.asp>.

interoperability between internal tools of the system and the external tools because it oversteps traditional approach to the information system and/or digital library and comes close to a virtual research environment. In this case a virtual research environment means a possibility to create and process a personal collection of digital documents (in all probability full text editions of original historical, i.e. primary documents) using special tools that enable to get to information from point of view of e.g. computational linguistics etc. Or virtual research environment means to create a grid of resources among which one of them is the main and/or (relatively) independent one and other are collateral and/or dependent. Thus, interoperability is the most apparent condition for practical work with historical documents and/or holdings.

DIGITAL HUMANITIES AND PARADIGM SHIFT: IDEAS AND PRACTICE

The practical work with historical documents and/or holdings is not a work of technicians (or if you like computer scientists) and librarians (or if you like information scientists) but it is more likely a work of scholars and researchers in humanities, especially historians and philologists. Of course, there are various historical and philological specializations and sub-disciplines, on the other hand theoretical and paradigmatic foundations of all branches of historical and/or philological scholarship are the same so that I can speak about history and/or philology generally. Since the theoretical transformations that are influenced by an application of information and communication technologies concern the deepest level of humanities in general and history and/or philology in particular, we can speak in the matter of these theoretical transformations about a paradigm shift that asserts its rights just now. Relevant questions that concern paradigm shift in humanities, history, philology, library and archival science are very complex but some of them can be set aside as fundamental for digital libraries presenting written cultural heritage [3; 9].

Firstly, a big problem is that codicology, i.e. discipline that deals with manuscript books, literary manuscripts and alike, is understood almost solely as archeology of book. Archeology of book means that questioning and research are oriented to a physical condition of manuscript book (but the same is valid also for the printed book and consequently for bibliography that deals with this type of book material) and that questioning and research oriented to the intellectual content and consequently to the cultural history is more or less extinguished. Thus, the traditional codicology and/or bibliography which is mainly archeology of book is contradictory to the idea of digital representation of written cultural heritage which is based on an effort to research a content not a container. Both conceptions, i.e. archeology of book and researching a content not a container as well are equally connected with idea of evaluation of historical sources, only from

different points of view: on one hand physicality of the printed environment enables distinguishing original and its copy – and so physicality is for archeology of book an essential notion; on the other hand virtuality of the digital environment does not enable distinguishing original and its copy but it very well enables comparing different contents – and so virtuality is for digital humanities a basic notion [6; 13].

Secondly, there is a new conception of content types and/or content levels as articulated in the IFLA document Functional Requirements for Bibliographic Records (shortly called FRBR) [5]. It describes a bibliographic conceptual framework that negates the traditional one that is based on subordination to the idea of printed edition. FRBR´s conceptual framework is much more complex and articulates four gradual levels characterizing the whole existence of the literary work, i.e. an item (a physical object-individual book, i.e. a purely actual object), a manifestation (a printed edition-bibliographic unit, i.e. a multiplicity of items), an expression (a version, translation, adaptation, mutation, i.e. concrete wording), and work (an artifact as an aspect of a personality, i.e. a purely virtual object). In the consequence of a pure virtuality of work some authors acknowledge that the work level has only a theoretical importance and that the level of work and that of expression are perceived and understood simultaneously as only one level called "worxperssion". However, it is too sophisticated in this context. On the other hand it is very important that only the item level has a real and simultaneously actual i.e. concrete existence while the other levels (manifestation, expression, work) are as for their existence in some respect real and virtual and perhaps abstract. Thus, a consequence of cataloguing and bibliographical work concerning written cultural heritage is crucial.

Thirdly, there is a concept of a fluid text that is contradictory to the concepts of archetype, "Urtext", and so-called best wording as well. Such a contradiction is very important for historians and/or philologists because the idea of critical edition is based on these notions. While archetype, "Urtext", or so-called best wording means that there is a strongly given text that can be, or indeed has to be the base for the critical edition, fluid text means that there isn't any possibility like this. In other words, while archetype, "Urtext", or so-called best wording means that their wording is transparent to the ideal sense, fluid text means that there is no possibility like this because the proper reality is the massiveness of the individual records that combine the fluid text. A combination of qualitative and quantitative analysis is necessary so that we can understand reality of texts and reality of life as well. Computational linguistic tools are more appropriate for solving such problems than traditional philological methods and so the way for digital history and/or digital philology is opened [10].

Fourthly, the idea of versioning has been popular during last roughly twenty years. It

means that preparing critical editions has no sense because each individual wording, each individual manuscript record, each individual glossed adaptation has its own sense that cannot be compensated for one wording that is presumably the right in a contradiction to the other wordings. The idea of versioning is one of the basic ideas of the so-called new philology. Accords and differences between various/all versions are important. A research in the area of versioning is concurrently a research in the area of cultural history, in the area of a content in a contradiction to the area of a container. It enables us to see all the linguistic versions at the same level, i.e. coordinated mutations rather than subordinated translations. This method finds its inspiration in the era of both European integration and globalization. Thus, the computational linguistics is again against the traditional one so that a place for digital history and/or philology is opened (compare [1; 19]).

MANUSCRIPTORIUM DIGITAL LIBRARY: A CASE STUDY

What I said hitherto is a mere theoretical conception (more concisely see [24]) that must be only realized practically. There are only a very few attempts in the practical realization of the virtual research environment concerning historical documents and/or holdings. One of them – perhaps the most sophisticated and the largest worldwide – is Manuscriptorium digital library (access through Internet: <http://www.manuscriptorium.com>) provided by the National library of the Czech Republic (access through Internet: <http://www.nkp.cz>) from the content part nad AIP Beroun Ltd. (access through Internet URL: <http://www.aipberoun.cz>) from the technical part. National library of the Czech Republic started its mass digitization in 1995, respectively 1996 and in 1997 and 1998 created standard DOBM [8], which was adopted as a UNESCO recommendation in 1999. National library of the Czech Republic was in 1999–2001 one of the full partners of the European project MASTER (Manuscript Access through Standard for Electronic Records) [23; 20; 21; 22], whose goal was to make a TEI (Text Encoding Initiative; access through Internet: <http://www.tei-c.org/index.xml>) compatible standard and in 2002 National library of the Czech Republic crated the MASTER+ standard (msnkpaip.dtd), a standard (access through Internet: <http://digit.nkp.cz>) the goal of which was to enable the digital document, i.e. to connect descriptive catalogue record and images, respective sequences of images that represent copy of the original historical document. In 2003 opened catalogue of historical holdings arose and in early 2004 operation of the Manuscriptorium digital library was introduced. In December 2007 work on the European project ENRICH (European Networking Resources and Information concerning Cultural Heritage; Access through Internet: <http://enrich.manuscriptorium.com>; <http://enrich-data.manuscriptorium.com>) [18] started the goal of which is to integrate resources

that provide historical documents and especially manuscripts as much as possible. Thus, Manuscriptorium digital library is based on a wide and rich experience, so it is able to offer and/or provide a relatively advanced service.

Heart of the Manuscriptorium digital library is a catalogue of base records in a format according to the MASTER standard. When catalogue records made by other creators or offered by other providers are imported into the Manuscriptorium base records database they are converted into a format according to the MASTER standard if they are originally created in another format (usually MARC21, UNIMARC, Dublin Core, MODS and other). MASTER does not prescribe what information depth whatever descriptive catalogue record should have; it depends on each cataloguer's resolution. It can be a problem from the traditional manuscript studies, codicology or bibliography point of view of because the database content can be unbalanced. On the other hand delivery of descriptive catalogue records by individual providers or partners can be regularized according to the concrete feasibility and capability. The biggest advantage of the MASTER standard is a very detailed fragmentation of data and consequently a big search consistency. When possibility of subsequent replacement of an existing descriptive catalogue record for the better one is ensured (which in the case of the Manuscriptorium database) then the difficulty is only in theory but not in practice because cataloguing is not creating products that are once and for all finished but it is a continual process that is never done. Last but not least the Manuscriptorium database of the descriptive records offers several options: firstly, database of the base records can contain various language mutations; secondly, base records can be alternated by special records of lesser weight expressive in specific points of view, e.g. codicology, art history, musicology etc.; and thirdly each of existing catalogue records can be improved. Thus, the Manuscriptorium database of base records is very flexible. Moreover, its search system allows searching with character tolerance, with the use of graphical variants with use various operators or phrases etc. It is robust enough to fulfill well the needs of the end user.

At present the most important part of the Manuscriptorium digital library are images, i.e. digital copies of original historical documents although we see that digitization is not only and perhaps also not chiefly making images. Of course, the images are scanned with the highest sensible resolution and in the archival quality. On the other hand, Manuscriptorium digital library is not an archival system; it is a presentation system, so that images of the excellent, i.e. archival quality are not provided via the Manuscriptorium. All the same images of more quality levels are provided, typically gallery thumbnails, previews, low/Internet quality, normal quality and black and white optimization for more pointed contrast for better reading if needed. Five quality levels are typically provided. However, it is not

obligatory for all potential partners, it depends on partner's capacity and ability. We can say that three quality levels are desirable, i.e. gallery thumbnails, previews and one sequence of images of better quality. Of course, there are still more possibilities concerning quality levels of images but they can be widely used in future rather than nowadays; it comes down to such a kind of scanning that would enable to see watermarks, to read palimpsests, text under blotch etc. The question is if this kind of image scanning should be a standard offer of the Manuscriptorium digital library. According to my opinion it may not be and it will be better to provide such qualities in some specific collateral resource; exploitation of such images is namely not common.

Another part of the Manuscriptorium content are full texts [17]. Full text data are prepared according to the TEI standard, existing document type definitions are for prose, poetry and factual prose (access through Internet: <http://digit.nkp.cz>). At present the full text content of Manuscriptorium consists of scholar editions, or should I say transcriptions of primary, i.e. original historical documents; implementation of full texts of secondary documents, i.e. scholarly papers etc. is in process of testing. Editions and/or transcriptions of primary, original historical documents have various forms at present and also in the future. Some of them are digitally born pragmatical editions of individual manuscripts or versions of texts (see e.g. http://www.manuscriptorium.com – query – title: Homiliarium quod dicitur Opatovicense; repository: Národní knihovna České republiky; shelf mark: III F 6), some of them are retro-conversions of traditional printed critical editions (see e.g. http://www.manuscriptorium.com – query – title: Codex gigas; reporitory: Kungl. Biblioteket – Sveriges nationalbiblioteket; shelf mark: A 148), some of them are translations of original (at present only Latin) texts (at present only into Czech) (see e.g. http://www.manuscriptorium.com – query – title: Codex gigas; reporitory: Kungl. Biblioteket – Sveriges nationalbiblioteket; shelf mark: A 148), some of them are mere hand made transcriptions without any scholarly apparatus (see e.g. http://www.manuscriptorium.com – query – title: Paměti Jednoty Bratrské z let 1530-1546; repository: Národní knihovna České republiky; shelf mark: XVII C 3), some of them are made through scanning and OCR with subsequent human control (see e.g. http://www.manuscriptorium.com – query – title: Článkové všeobecného sněmovního snešení; repository: Parlamentní knihovna; shelf mark: F 5042). Such differences as for quality of full texts that are implemented within Manuscriptorium digital library can be seen as very inconsistent. On the other hand such variability as for quality levels of full texts is very practical and operative. To prepare editions of historical texts is in any case very difficult, paleographical and linguistical skills are necessary. There are few people that are able to prepare such editions more so when the skills concerning computing in humanities are not generally disseminated.

Nowadays the problem of digital full text editions is truly up-to-date and very important as well [14; 16]. To present only digital image copies of original historical documents is at any rate not enough – to present digital texts of historical primary documents is necessary for the conversion of media which is nowadays one of main tasks of humanities. Theories concerning so-called fluid text occur [2], conceptions concerning versioning as a presentation of multiple texts appearance [15], versioning software tools are created (see "v-machine", access through Internet: <http://v-machine.org/>). The philological and historical work rise on one hand and computing in humanities dissolve on the other. The new ways of study of text transmission and manuscript tradition with utilization of specific striking character expand [7]. Accordingly, building digital libraries is a big challenge for scholarly edition work. The team of Manuscriptorium digital library collaborates with the team of the Old Czech Department of the Institute of the Czech Language in Prague (see http://www.ujc.cas.cz/oddeleni/index.php?page=staroces>). This collaboration brings great benefits for both parts.

A music notation is also a text while it is seen in semiotic sense, of course, such that music editions are also full text editions in a wider sense. Therefore the Manuscriptorium team is testing possibilities how to do it now. Unfortunately there is no best practice and no workflow available, so that various ways must be examined. Two standards that should well enable to represent music notation using XML, MusicXML [12] and MEI (Music Encoding Initiative) that is compatible with TEI (Text Encoding Initiative) [11] were identified. Other potential standard for representing music notation is developed by the CMME (Corpus Mensurabilis Musice Electronicum) [4]. It already has been established that the MusicXML standard is a proprietary standard and moreover it is more likely a standard for the exchange format for representation of the music notation than for the archival one. Thus, although its software support is relatively good, it would not be presumably the best solution to use it. The MEI standard is probably a better choice for the archival format for representation of the music notation but its software support is at present rather poor. On the other hand there is a hopeful factor, namely that community using the MEI is growing, so that the software support could be better in the future. The same applies to the community using the CMME standard. It will take some time, of course, until the best practice for music editors will be ready.

Also audio/music document/s can be a part of the compound digital document within the Manuscriptorium digital library. From the technical side it is no problem for many years. On the other side it is a big problem from the copyright point of view. Although the problem with copyright does not concern composer/s (because the compositions that are covered are already public domain), it concerns still performer/s. Thus, implementation of audio/music documents

into the Manuscriptorium digital library is possible only by way of a trial if agreed by performer/s. The audio-music performance is an interpretation of the original source, certainly not the original source itself. On the other hand, music is a performance, not a written source. Consequently the audio representation of music is fundamental for the end user together with the representation of the written expression of music. Thus, testing mutual correlation between images and music editions, i.e. full texts and audio/music documents as well is an important step in creating virtual research environment from a more complex point of view. However, all questions and tasks of this activity will be solved only when simultaneously the problem of the copyright in the Internet is solved.

There are some tasks that are between the Manuscriptorium digital library and the individual research. These tasks concern virtual research environment in proper sense because they connect Manuscriptorium as a general resource with the research area of individual persons. Main representatives of these task results are virtual collections and virtual documents that are in the phase of preparing now and next year they will be tested. The goal of the virtual collection/s is to choose everybody's own collection for his/her individual research and such a virtual collection can be made accessible simultaneously for somebody else even for everybody if the creator of the virtual collection wants. Personalizing tools and/or content creation as well is an important task of developing Internet nowadays. Through such personalized collections and/or documents a valuable content that would be close and inaccessible in other cases can be made accessible. Collective and concurrently individual and/or personal nature of research comes through virtual collection and virtual document. It is fundamental that especially virtual collection can be created very flexibly, i.e. both by choosing and book-marking already existing documents and by choosing documents from the growing database according to the general query. Consequently virtual collection may not be static, it can be dynamic too and so virtual collection is a heuristic tool for a wide application. While virtual collection is or can be a heuristic tool, virtual document is more likely a research result, product, publication in the traditional sense. Thus, virtual collection and virtual document as well should be very hopeful and promising outputs of the virtual research environment in a near future.

There are still other ideas of development of the Manuscriptorium digital library in long-term view. It concerns external tools that can interoperate with the Manuscriptorium. Some of them are prepared for future testing now. Use of some software tools based on the computational linguistics is most interesting in this case. There are already some full texts contained in the Manuscriptorium and the full texts, especially editions of primary, original historical documents will grow increasingly; full text editions were already identified as a substantial part of the Manuscriptorium

digital library. Making full text editions available is a high priority among end user requirements. Therefore a further work with the full text editions is also desirable. So we may very well imagine an export of full text edition/s from the Manuscriptorium database and its/their import into a special corpus of texts where a software tool based on the computational linguistics, e.g. a corpus manager can be applied to the newly created corpus of texts. As for corpus manager is quite sophisticated tool both for searching within texts and for displaying results, in some cases e.g. also for clustering the texts it is very convenient for heuristic purposes of cultural and literary historian as well as of other specialists. Using such tools and also other external tools is an outlook of several years.

And ultimately a futurological vision: some computer scientists believe that OCR can also be applied to manuscripts...

CONCLUSION

Theoretical conception and practical experience of the National library of the Czech Republic during the methodical digitization process concerning historical documents and/or holdings have proved that digitization is really not simply making images. Digitization of historical documents and/or holdings means not only and not firstly creating mere digital copies for preservation purpose but complex activity concerning presentation of cultural heritage and representation of historical sources. Therefore images must be accompanied by other types of documents, i.e. by full texts, audio/music documents, multimodal/multimedia documents etc. Digitization of historical documents and/or holdings leads to the paradigm shift this way and it is one way to the information and knowledge society.

REFERENCES

1. BROM, Vlastimil. *Der deutsche Dalimil*: Untersuchungen zur gereimten deutschen Übersezung der alttschechischen Dalimil-Chronik. Brno: Masarykova univerzita, 2006. 281 p. ISBN 80-210-4211-7;

2. BRYANT, John. *The Fluid Text*: a Theory of Revision and Editing for Book and Screen. Ann Argot: University of Michigan Press, 2002. 198 p. ISBN 0472068156.

3. CHODOROW, Stanley. The Medieval Future of Intellectual Culture: Scholars and Librarians in the age of Elektron. In: *ARL*: A Bimonthly Newsletter of Research Library Issues and Actions. Issue 189, December 1996 [accessed 6 June 2008]. Access through Internet: <http://www.arl.org/bm~doc/medieval.pdf>.

4. *CMME*: dynamic early music editions. Access through Internet: <http://www.cmme.org/>.

5. *Functional Requirements for Bibliographic Records*: Final Report. München: K.G.Saur, 1998. 136 p. ISBN 3-598-11382-X [accessed 6 June 2008]. Access through Internet: <http://www.ifla.org/VII/s13/frbr/frbr.htm>.

6. GIESECKE, Michael. *Der Buchdruck in der frühen Neuzeit*: Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien. Frankfurt am Main: Suhrkamp, 1998. 957 p. ISBN 3-518-28957-8.

7. KALISZUK, Jerzy. *Mędrcy ze Wschodu:* legenda i kult Trzech Króli w średniowiecznej Polsce. Warszawa: Efekt, 2005. 332 p. ISBN 8387338249.

8. KNOLL, Adolf; MAYER, Tomáš; PSOHLAVEC, Stanislav; VOMLEL, Jan. *Digitization of Rare Library Materials. Storage of and Access to Data*: The Solution for the Compound Document, Manuscripts and Old Printed Books [CD-ROM]. Praha: Národní knihovna České republiky, 1997.

9. MICHELUCCI, Pascal; MARTEINSON, Peter: Paradigma Lost? Electronic Publishing and the Renewal of Research. In: *CHWP*: A 12, Publisher April 1998. [Jointly published with *TEXT Technology*, 8.1 (1998), Wright State University.]. [Accessed 6 June 2008]. Access through Internet: <http://www.chass.utoronto.ca/epc/chwp/micheluc/>.

10. MOELLER, Bernd-Stackmann, Karl. *Städtische Predigt in der Frühzeit der Reformation*: Eine Untersuchunng deutscher Flugschriften der Jahre 1525 bis 1529. Göttingen: Vandenhoeck-Rupprecht, 1996. 383 p. ISBN 3-525-82436-X.

11. *The Music Encoding Initiative* (MEI), access through Internet: <http://www.lib.virginia.edu/digital/resndev/mei/>.

12. *MusicXML Definition.* Access through Internet: <http://www.recordare.com/xml.html>.

13. O´DONNEL, James J. *Avatars of the Word*: From Papyrus to Cyberspace. Cambridge, Mass.; London: Harvard University Press, 2000. 210 p. ISBN 0-674-00194-X.

14. REHBEIN, Malte. *Editionen als Softwareproblem: die "dynamische Textedition"* [accessed 4 Juny 2008]. Access through Internet: <http://www.denkstaette.de/files/Rehbein_Editionen_als_Softwareproblem.pdf>.

15. REIMAN, Donald H. "Versioning": The Presentation of Multiple Texts. In REIMAN, Donald H. *Romantic Texts and Contexts*. Columbia: University of Missouri Press, 1987, p. 167–180. ISBN 0826206492.

16. ROBINSON, Peter. Current issues in making digital editions of medieval texts – or, do electronic scholarly editions have a future? *Digital Medievalist*, 1.1 (Spring 2005). ISSN 1715-0736 [accessed 4 Juny 2008]. Access through Internet: http://www.digitalmedievalist.org/article.cfm?RecID=6>.

17. UHLÍŘ, Zdeněk. Manuscriptorium: evropská digitální knihovna rukopisů s plnými texty. In *Problematika historických a vzácných knižních fondů Čech, Moravy a Slezska, 2007*. Sborník z 16. odborné konference, Olomouc, 14.-14. listopadu 2007. Olomouc: Vědecká knihovna v Olomouci; Brno: Sdružení knihoven ČR, 2008, p. 103–108. ISBN 978-80-7053-276-8 (VKOL); 978-80-86249-47-6 (SDRUK).

18. UHLÍŘ, Zdeněk: Manuscriptorium na cestě k evropské digitální knihovně. In *Knihovny současnosti 2007*. Brno: Sdružení knihoven ČR, 2007, p. 136–144. ISBN 978-80-86249-44-5.

19. UHLÍŘ, Zdeněk. Nově objevený zlomek latinského překladu Kroniky tak řečeného Dalimila. *Knihovna*, 16, 2005, Nr. 2, p. 137–169. ISSN 1801-3252.

20. UHLÍŘ, Zdeněk: Projekt MASTER a jeho aplikace v NK ČR. In *CASLIN 2002*. Ochrana a sprístupňovanie dokumentov: nové trendy. Grand hotel Permon, Podbanské 18,

032 42 Pribylina, Vysoké Tatry – Slovenská republika, 23.–27. júna 2002. Martin: SNK, 2002, p. 47–51.

21. UHLÍŘ, Zdeněk. Projekt "MASTER" a problematika elektronického zpracování středověkých rukopisů. In *Ikaros* [online], 1999, č. 8. ISSN 1212-5075 [accessed 4 Juny 2008]. Access through Internet: <http://www.ikaros.cz/node/391>.

22. UHLÍŘ, Zdeněk. Projekt MASTER a standardizace v oblasti zpracování rukopisů. *Národní knihovna*: knihovnická revue. 10,1999, Nr. 3, p. 109–113. ISSN 1214-0678 [accessed 4 Juny 2008]. Access through Internet: <http://full.nkp.cz/nkkr/NKKR9903/9903109.html>.

23. UHLÍŘ, Zdeněk. Standard MASTER: katalogizace rukopisů v XML. *Národní knihovna*: knihovnická revue. 13, 2002, Nr. 2, p. 84–101. ISSN 1214-0678 [accessed 4 Juny 2008]. Access through Internet: <http://full.nkp.cz/nkkr/Nkkr0202/0202084.html>.

24. UHLÍŘ, Zdeněk. *Teorie a metodologie elektronicko-digitálního zpracování rukopisů a hybridní knihovna* [The theory and methodology of electronic-digital processing of manuscripts and the hybrid library.]. Praha: Národní knihovna České republiky, 2002. 324 p. ISBN 80-7050-410-2.

SKAITMENINIMAS – NE TIK ATVAIZDŲ GAMYBA

ZDENĚK UHLÍŘ

Santrauka

Autorius straipsnyje aptaria senųjų istorinių dokumentų, tokių kaip viduramžių ir naujųjų amžių pradžios rankraščių, inkunabulų (knygų, išleistų iki 1500 m.), senųjų spausdintų knygų (iki 1800 m.), istorinių žemėlapių, jei reikia, archyvinių knygų, dokumentų ir pan. skaitmeninimą. Svarbiausia mintis, kad skaitmeninimas – ne tik atvaizdų gamyba, t. y. archyvavimui ir reprezentavimui nepakanka pagaminti istorinio dokumento atvaizdą. Reikia jį papildyti kitais duomenimis, t. y. metaduomenimis, taip, kad būtų galima perteikti visą informaciją apie istorinį dokumentą, suteikti prieigą prie jos, nes skaitmeninimas – tai ne paprasta techninė veikla, o pasaulinė istorinių dokumentų sklaida. Pagrindinį skaitmeninimo tikslą padeda pasiekti trys teoriniai pagrindimai ir techninės sąlygos: pirma, reikia atskirti duomenis nuo programinės įrangos; antra, standartizuoti duomenis; trečia, įrankiai ir sistemos turi tarpusavyje derintis. Taigi skaitmeninimas – tai perkėlimas iš spausdintų dokumentų tradicinės informacinės, komunikacinės ir žinių aplinkos į virtualią erdvę. Toks perkėlimas humanitariniuose moksluose taip pat reiškia paradigmos pokytį. Humanitarinių mokslų, ypač istorijos ir filologijos, paradigmos pokyčiai matomi įvairiose vietose ir vyksta įvairiomis formomis. Skaitmeninant istorinius dokumetus, keturi iš jų yra svarbiausi. Pirmiausia, tai kodikologijos transformacija iš vadinamosios knygos archeologijos į kiekybinę kodikologiją kultūros istorijos prasme; antra, nauja turinio tipų ir/ar turinio lygių koncepcija, išreikšta IFLA dokumente *Funkciniai bibliografinio įrašo reikalavimai* (sutrumpintai FRBR); trečia, takaus teksto koncepcija, prieštaraujanti ankstesnėms archetipo koncepcijoms, „Urtext" ir vadinamajai tiksliausiai formuluotei; ketvirta,

daugybinių versijų idėja, kuri buvo populiari pastaruosius dvidešimt metų.

Toliau autorius supažindina su Čekijos nacionalinės bibliotekos iniciatyva – skaitmenine biblioteka *Manuscriptorium*. Šiandien tai viena didžiausių pasaulyje skaitmeninių bibliotekų, pristatanti senuosius istorinius dokumentus. Ją sudaro skaitmeninių dokumentų kūrimas ir jų integravimas į įvairius kitus išteklius, naudojančius skirtingus duomenis ir metaduomenų standartus. *Manuscriptoriumo* širdis – katalogo įrašų, suderintų su MASTER standartu, duomenų bazė. Kai partneriai naudoja kitus metaduomenų formatus, jie konvertuojami į MASTER taip, kad, viena vertus, egzistuotų heterogeniška terpė, o antra vertus, būtų kuriamas homogeniškas centrinis katalogas. Paprastas aprašomasis MASTER standartas šiai skaitmeninei bibliotekai buvo papildytas struktūriniais metaduomenimis, reprezentuojančiais originalaus dokumento struktūrą, ir techniniais duomenimis, aprašančiais skaitmeninimo procesą. Išplėstinis MASTER leidžia jungti individualius analitinius objektus/vaizdus/puslapius į vientisą dokumentą/virtualią knygą. Naudodamas METS standartą, vientisas dokumentas leidžia pirmiausia sujungti kelis katalogo įrašus, reprezentuojančius tą patį dokumentą, antra, sujungti visus tekstus ir koreliuoti juos skaitmeniniais atvaizdais, trečia, sukurti išsklaidytą dokumentą su sąsajomis į nutolusius atvaizdus. Tokiu būdu *Manuscriptoriumo* skaitmeninė biblioteka gali būti iš tikrųjų išsklaidoma bet kur virtualioje erdvėje. Šiuo metu ketinama šiai skaitmeninei bibliotekai pradėti taikyti kompiuterinės lingvistikos įrankius ir ontologijas, tačiau tai jau ateities uždaviniai.

Įteikta 2008 m. gegužės mėn.