

Medievalia: a Place Where Archival Science and History Meet and Then Separate Again

BOGDAN-FLORIN POPOVICI, Ph.D.

archivist, National Archives of Romania, Braşov County Division, Braşov, str. G. Bariţiu nr. 34, 500025,
Braşov, România
e-mail: bogdanpopovici@arhivelenationale.ro

Medievalia: a Place Where Archival Science and History Meet and Then Separate Again

ABSTRACT

This paper seeks to present some aspects from the experience of Medievalia Project, undertaken by University of Bucharest - Faculty of History and National Archives of Romania, aiming to offer a national portal for medieval records. The paper will offer a presentation of some strategic approach taken for digitization and for retrieval of “documents”.

Key words: digitisation, National Archives of Romania, University of Bucharest, Norwegian Archives; Medievalia Project, scopeArchiv

Medievalia: un luogo dove l'archivistica e la storia si incontrano e poi si separano di nuovo

SINTESI

Il presente articolo cerca di presentare alcuni aspetti derivanti dall'esperienza del progetto Medievalia, intrapreso dall'Università di Bucarest-Facoltà di Storia e dall'Archivio nazionale di Romania, con lo scopo di offrire un portale nazionale per i documenti medievali. L'articolo presenta di alcuni approcci strategici adottati per la digitalizzazione e la ricerca di documenti.

Parole chiave: digitalizzazione, Archivio nazionale di Romania, Università di Bucarest, Archivi norvegesi, progetto Medievalia, scopeArchiv

Medievalia: Kraj, kjer se srečata in znova razideta arhivska znanost in zgodovina

IZVLEČEK

Namen prispevka je predstaviti izkušnje projekta Medievalia, ki ga izvajata Univerza v Bukarešti - Fakulteta za zgodovino in Nacionalni arhiv Romunije, s ciljem ponuditi nacionalni portal srednjeveških zapisov. V prispevku bo predstavljen strateški pristop, sprejet za digitalizacijo in za ponovni priklic dokumentov.

Ključne besede: digitalizacije, Državni arhiva Romunije, Univerza v Bukarešti, norveški arhiv; projekt Medievalia, scopeArchiv

1 Introductory notes

Between 2015-2016, University of Bucharest-Faculty of History, University of Cluj Napoca, National Archives of Romania (hereafter *NAR*) and National Archives of Norway ran a project named “Digitizing medieval documents from the National Archives of Romania”, aiming to migrate into digital the image of records held by NAR before 1600. The project was developed within the framework of the Programme PA16/RO12 (“Conservarea și revitalizarea patrimoniului cultural și natural”) and financed by the Financial Mechanism SEE 2009-2014, contract PA16/RO12-LP-8/18.12.2014.

The project implied scanning old records with various documentary forms and carriers, creating

a metadata profile for them and displaying them online, from a database. In this paper we shall not question some strategic decisions (such as creating a new database for this project, whilst NAR already have implemented a fully functional IT system, including a robust archival information system, scopeArchiv), nor that description adopted declined any use of archival standard in use. In the following, I shall try to examine a presumed special “historical perspective” in search and retrieve of medieval documents. It must be emphasized from the very beginning the fact that, since the project faced delays, we could not have relevant statistical information about the usage of the “database” until the moment of authoring this paper, therefore the analysis below will be based on personal consideration of the author, who may be biased by his profession as archivist.

2 The workflow

The scanning the medieval documents was made using some flat A2 scanners belonging to NAR and some others acquired in the Project. The documents were always accompanied by a scan target (for colour calibration and dimensions bearing) and by a label containing the extended reference code of the record. By scanning, two outputs were generated one TIFF and one JPG files, at 600 dpi (for charters) and 400 dpi (for book registers). The images were used for palaeography specialist to create descriptions of the documents. After finishing the task, the metadata were imported into the NAR system. From the JPG file it were derived 2 versions of the image, one JPG file for web visualisation and one PNG file for thumbnail. All files were stored on disks as filesystem storage, in separate locations for master and derivative. Using the NAR system, images were automatically attached to the descriptions and then exported in the new database.

Files	Purpose	Parameters
TIFF	Master	Color, 24 bits, 400 or 600 dpi
Jpeg	Reproduction Copy	Color, 24 bits, 400 or 600 dpi
Jpeg_a	Internet view	Color, 24 bits, 300 dpi, 2000 x 200 pixels
png_t	Thumbnails	Color, 24 bits, 72 dpi, 200 x 200 pixels

Metadata and images are accessible via a website (Arhiva medieval a Romaniei, 2015-2016). The search is using a Google-approach: text box where the string should be input. When the list of hits appears, the results are listed at the document or component¹ level. They may be grouped based on holder, issuer or date and sorted by relevance. Also, there is a possibility to narrow the search using facets (with statistics): holder, hierarchy, issuer, data, subjects, places, names and language.

In detailed view, the header contains: date, issuer, title and abstract, followed by narrative reference code, previous reference code, coded reference code, names, places, subjects, status of transmission, structure, carrier, type of record (document or book-register), seal, language, bibliography, observation and author of the description. There is also a thumbnail directing to the image viewer, that displays the digital, zoom-able, image of the document.

3 A review of the workflow. A critique of the assumptions of the project team

While the NAR have had an archival information system since 1970s, its decommissioning at the end of 1990s let the institution for more than a decade out of use of modern technology. The implementation of a modern system in 2013 open the way for presence on the internet, and dissemination of archival description. In this regard, Medievalia Project was welcomed as an opportunity to increase the number of records available online. Between various initial expectations and the final outcome a gap opened, and some perspectives prevailed over others.

A first remark about the project and the database created is that it started and it was implemented having in mind a limited goal: “database for researchers”. In our opinion, this perspective is, from the very beginning, discriminatory and limited. It must be stated, for contextualisation, that in the

1. Within the framework of the project, a component may mean an intellectual (e.g. transsumptum) or physical (page) element of a record.

past in Romanian Archives there was a separation made between “serious researchers” and “amateurs”; in some moments in the past, for accessing special fonds or collections it was required a recommendation from the scientific body where the user came from. NAR made serious steps away from this perspective, enlarging the accessibility and removing formal barriers in front of the users. It is, in the end, a measure of democracy: access to information is a constitutional right and any citizen is equal entitled to study records. I do believe the role of the National Archives is to enhance the access of the users to their holdings. Also, I do believe medieval charters should be protected from frequent handling in the reading room, so their digitisation will help in a high degree to their preservation.. But I do not believe stating that a database is for academic use is helping in any way the National Archives fulfilling its role in society, promoting its collection to the public or promoting a modern approach to the users.

Of course, there is no formal barrier or obstacle for anyone to use the database. In the end, anyone accessing the web address may search in the database. But exactly here resides the main issue. While in the reading room there is an archivist helping and directing the user (*Duff and Johnson 2002, 480-485*), on the internet there is not such a help. Contextual information is, in this regard, vital. This contextual information is very much distilled, if not completely absent, from the website. For instance, the only way to access the information is to create a query. That means, due to “Google approach”, a user must know what (s)he is looking for: a historical character, a topic, a place, a date. Without such preliminary data, a user can hardly go “beyond the entrance door”. Presuming though a user can go over, (s)he will find a very limited abstract; it really gives only a flavour about the information in the document. The large burden of work during a project might be an excuse for such succinct presentation for the content, but only for those who does not know the team refuses to use the archival abstract existent in NAR finding aids! Basically, instead of using summaries of half a page, with detailed presentation of the content, it was decided to re-create an abstract of a few words, with the explicit intention in mind that “the true researcher will read the document, not the abstract”. Leaving aside the fact that archival finding aids were compiled, in many cases, by local archivists, more aware of the local history, documentary context, literature on the topic etc.; the fact itself that a bigger amount of information was intentionally avoided is, at least from my perspective, a barrier for the access to information and a serious betrayal of archival ethics².

But, on the other hand, I am wondering what would be the profile of the ideal “researcher”. Should (s)he be historian, a Middle Age examiner? Very likely so. But what about a statistician or anthropologist—are they “researchers” enough to be able to fully exploit the database? What about students, with limited experience in research? Hard to say; it is more easy to notice you may not *discover* in the database, but *retrieve*; meaning you must know before what you’re looking, which is hardly a way to enhance the use of records. In the same regards, one can notice the interface and indexes are designed only for Romanian, Hungarian, German and English languages, leaving aside languages of populations that heavily interacted with Romanian history and whose interests in the database might be high: Turkish, Bulgarian, Greek. It seems the Project team set another filter for the users...

While already 30 years ago some authors underscored “the power of the principle of provenance” in retrieval the archival information (*Bearman and Lytle 1985*), for *Medievalia* team this aspect is almost completely irrelevant. “Almost” because in the search and displaying of metadata the presentation of the archival context of the document is very limited (basically, it only indicates the titles of the archival units of description). There is no description of fonds or subfonds or series; even when filtering by facets, the only result is a flat list of abstracts. While I can agree with the advantages of “subject access” (vs. “provenance access”) (*Yeo 2010, 102-103*), I truly believe that complete lacking of description of upper level of a certain document not only break the archival descriptive principle and standards, but create a barrier to the contextualisation and deep understanding the record³ (*Yeo 2010, 102-103; Duff 2010, 122-124*). And, for me, raises serious questions about the profile of a historian who does not believe in context...

But, as Wendy Duff, among others, noticed, subject access “requires ... [to] create subject indexes and describe records using topical terms” (*Duff 2010, 124*). Examining the way metadata are

2. “Archivists should promote the widest possible access to archival material and provide an impartial service to all users” (ICA Code of Ethics, 1996, 6.).

3. For instance, in relation document > transsumptum, the only possible navigation is bottom up, from transsumptum to the document, while reverse navigation is not possible!

presented, some critical remarks can be made. Firstly, the indexed filed, like issuer, for instance, but also Places, Subjects or Names are basically, *tags*. By this, we understand strings, as they appear in the document (or as they were typed by author of the description), with no authority control, with no contextualisation, no explanations, no consolidation in a thesaurus, for instance. The same issuer may be types in different ways (“Steven, the ruler”, “Steven, the voivode”, Steven the 3rd” etc.) and a user must be well initiated in the topic to see it is the same character. Of course, no ISAAR(CPF) description was taken into consideration... In my opinion, such an approach reduces not only the possibilities of management (basically, noting is related with nothing, so each string should be checked manually), but also the retrieval is heavily affected.

A final remark concerns the possibility to cite the items. Descriptions of the documents have no reference at all; there is no possibility for a regular user to quote the description displayed, as from a publication. The explanation resides on the technology used for this website (Oracle Webcenter). For an initiated user, copying a part of the link displayed is the way to quote, but this is a technical subtlety not explained to public. About the images, as I mentioned above, the approach taken was to scan also a label with the narrative indication of the reference code. While it might look like a good idea to “froze” the reference code, in some specific cases it will be disastrous. In this project documents from non-processed fonds were scanned; therefore, the reference code may be temporary. When the processing will end, the document may have a complete different reference code, but the image cannot be changed except by image processing, which is not feasible due to amount of files.

4 A different approach?

The difficulties faced by the project team were somehow understandable, since presenting archival materials raises challenges due to their complex nature. Despite that, ignoring professional literature and experience was not a good option and undeserved lowered the impact and final results. This “reinvent the wheel” approach is as much as incredible since the Team had a good example for publishing archival documents in the system NAR is currently using. ScopeArchiv, like an archival information system, allows for multiple possibilities of search: free text (“Google approach”), hierarchy, descriptors or field search. It allows to model hierarchies, and describe various levels. The descriptors are modelled in a very powerful way, within thesauri, with the possibility of consolidated management and deep description (implementing ISAAR(CPF), making links to various supportive authority sources of persons, places, maps etc.). Even the reference code could be configured to be watermarked on the image for presentation, while the master would stay intact, allowing for future reprocessing the documents. The most powerful feature, though, would be the capacity of multiple relationships. Here it is a presentation of relations in scopeArchiv and in Medievalia.

5 Conclusions

The implementation of Medievalia Project means for NAR a new possibility to disseminate its valuable medieval archival treasure. More than 40.000 documents are intended to be published online and more than 100.000 images attached. The project implied a great effort, and the results are impressive. Despite that, as shown in this paper, the approach taken for presentation the documents showed, more than once, some unhappy choices. The complete denial of archival expertise in the field, obsessive orientation towards a special type of researcher (which, I suspect, is not even a real embodied as imagined) decrease the value of the output. Maybe a future project will allow for an improvement, helping various categories of users to understand a more faithful representation of vast network of relationships and context of archival materials.

Images

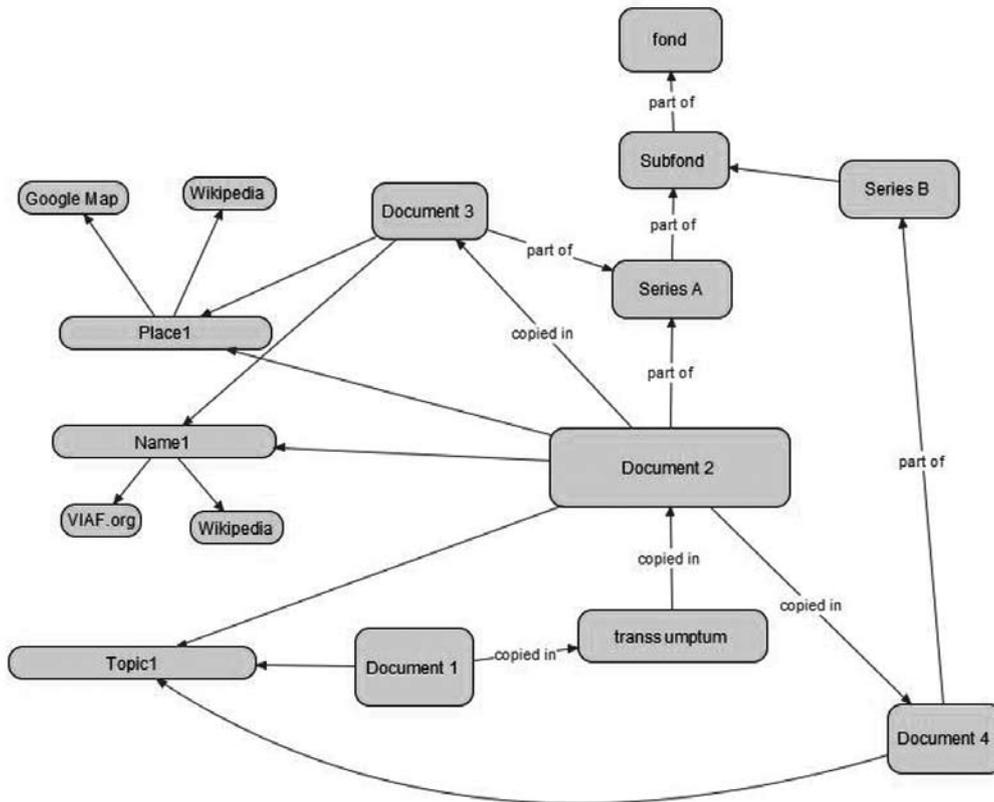


Image 1: Modelling relationships in scopeArchiv

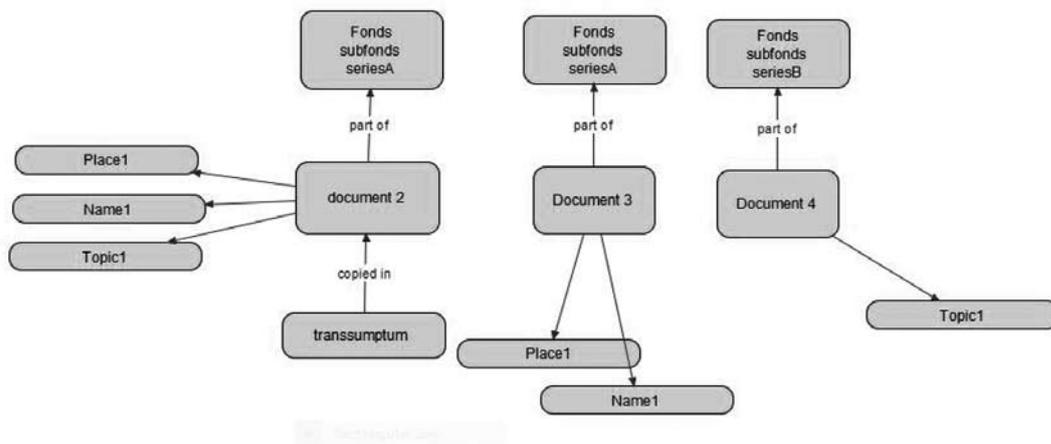


Image 2: Modelling relationships in Medievalia

References

- Arhiva medievala a Romaniei (2015-2016). Available at: <http://www.arhivamedievala.ro> (accessed May 2016).
- Bearman, D., Lytle, R (1985). The Power of the Principle of Provenance. *Archivaria* 21, nr. Winter (1985): pp. 14-27.
- Code of Ethics* (1996). International Council on Archives.
- Duff, W., Johnson, C. A. (2000). Accidentally Found on Purpose: Information-Seeking Behaviour of Historians in Archives. *The Library Quarterly*, October 2002: pp. 472-496.
- Duff, W. (2010). Archival Mediation. In *Currents of Archival Thinking*, de Terry Eastwood, & Heather MacNeil, pp. 115-136.
- Yeo, G. (2010). Debates about Description. In *Currents of Archival Thinking*, de Heather MacNeil Terry Eastwood, pp. 89-114.

SUMMARY

Between 2015-2016, University of Bucharest-Faculty of History, University of Cluj Napoca, National Archives of Romania (hereafter NAR) and National Archives of Norway ran a project named “Digitizing medieval documents from the National Archives of Romania”, aiming to migrate into digital the image of records held by NAR before 1600. The project was developed within the framework of the Programme PA16/RO12 (“Conserarea și revitalizarea patrimoniului cultural și natural”) and financed by the Financial Mechanism SEE 2009-2014, contract PA16/RO12-LP-8/18.12.2014. This paper presents the workflow for digitizing, describing and publishing documents online, and critically examined some of the choices taken by the project team. In this regard, it was criticized the orientation towards a special category of user (“THE researcher”), the reduced amount of descriptive information, the lack of authority controls.

Typology: 1.04 Professional Article

Submitting date: 24.01.2016

Acceptance date: 20.02.2016