HAL
open science

Methodological Guide on Sources in AtoM

Kathleen Walker-Meikle, Jörg Lehmann, Francesca Morselli, Lorenza Tromboni, Andrei Zamoiski

To cite this version:

Kathleen Walker-Meikle, Jörg Lehmann, Francesca Morselli, Lorenza Tromboni, Andrei Zamoiski. Methodological Guide on Sources in AtoM. 2016, 10.13140/RG.2.2.22529.25441 . hal-01419138

HAL Id: hal-01419138

https://hal.science/hal-01419138v1

Submitted on 22 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CENDARI
COLLABORATIVE EUROPEAN DIGITAL ARCHIVE INFRASTRUCTURE

METHODOLOGICAL GUIDE ON SOURCES IN ATOM

AUTHOR(S)	Kathleen Walker-Meikle
COLLABORATOR(S)	Jörg Lehmann, Francesca Morselli, Lorenza Tromboni, Andrei Zamoiski
THEMES	World War One and Medieval Culture
PERIOD	21st Century

CENDARI is funded by the European Commission's 7th Framework Programme for Research
This Archival Research Guide has the DOI: 10.13140/RG.2.2.22529.25441

SEVENTH FRAMEWORK PROGRAMME

TABLE OF CONTENTS

6 ABSTRACT

6 INTRODUCTION

Note of the author

7 THE USE OF ATOM WITHIN THE CENDARI PROJECT

Methodology used in the CENDARI project for selecting and entering archival institutions and descriptions

Examples of 'hidden archives'

Discovering the Central Asian archives

"Trophies of the War": The Russian state military archive

Material in the Archival Directory AtoM: Numbers

16 WHAT IS NOT INCLUDED – AND WHY IT IS A "WORK IN PROGRESS"

The limitations of the CENDARI project

18 BIBLIOGRAPHY

ABSTRACT

This Archival Research Guide (ARG) aims to provide the user with a broad overview of the Archival Directory, a key component of the CENDARI Virtual Research Environment. The CENDARI Archival Directory consists of all available data in the AtoM application ("Access to Memory") and data in the main repository, which is derived from international or national aggregators as well as national archives or other data providers. All data, whether manually inputted via the AtoM application or collected from institutions, form together the Archival Directory. In this guide the term "Archival Directory" refers only to the application AtoM and it hence will be used in this narrow sense. The term "Archival Directory AtoM" reinforces this definition. Unlike most of the other ARGs in CENDARI, this is not a thematic overview of a historical subject but instead a practical guide to the Archival Directory AtoM's aims and how it was created and organised.

This guide will present the methodology used by researchers associated with the CENDARI project when creating archival description and archival institution entries in CENDARI. It explains how material was selected to be entered into the Archival Directory AtoM, which institutions were considered as 'hidden archives' by the CENDARI team, what material was not included, the limitations of the project and the future of the Archival Directory AtoM. The CENDARI Archival Directory AtoM can be accessed both directly from the website and indirectly via the link provided in the Note-Taking Environment (NTE).

INTRODUCTION

The Archival Directory AtoM is an open access tool which serves as the means by which users can add descriptions of institutions as well as descriptions of holdings, collections, records, and single items. A detailed manual on how to do this can be found here. Moreover, the Archival Directory AtoM presents all the data manually encoded by the CENDARI team and is publicly accessible for every user of the internet. For all historical research, any source needs to have indications of the context in which it was created and distributed. Moreover, it should have information regarding its history (that is, use, storage and authenticity). Most cultural heritage institutions (archives, libraries and museums) possess the legal authority on the credibility and reliability of their source material. Their catalogues and finding aids contain the necessary information on single holdings and record groups, integrating essential information on how to use any given inventory and how to interpret its contents. The software used for the manual establishment of descriptions of sources is AtoM. It aims to be user-friendly and conforms to the standardized format Encoded Archival Description (EAD). This archival standard is used in almost all European and many non-European countries. It is approved by the International Council on Archives (ICA), the institution that collaborated in the development of the AtoM software. AtoM is free web-based application, and as a multilingual multi-repository interface and database for archival descriptions, all elements in AtoM can be translated into multiple languages using the translation interface.

Note of the author

As with all Archival Research Guides in CENDARI, users can add material to this document, along with adding or editing entries in the Archival Directory AtoM. Suitable contributions could include explanations on what they added to the Archival Directory AtoM in the course of their research, and the methodology used. Similarly, users could detail their experience with working with the AtoM software or when uploading or updating archival entries. Contributors might also expand on why they decided to add entries to the Archival Directory, and what were the gaps in the extant entries that made them decide to upload new entries into the system. Additionally, contributors might wish to note any relevant updates, whether in regard to the software, archival standards or entries in the Archival Directory AtoM itself.

THE USE OF ATOM WITHIN THE CENDARI PROJECT

The CENDARI project set up the Archival Directory AtoM according to the needs of its two case studies on the First World War and Medieval culture. The main objective was to enhance the accessibility and visibility of unique historical archives across Europe through the creation of a single directory that could be used by users in the CENDARI infrastructure.

The Archival Directory covers, in a representative manner, different types of institutions with archival holdings (archives, libraries, research institutes and museums) in all European and many non-European areas, based on the two case studies.

CENDARI placed an emphasis on "hidden archives" in Europe, especially in Central, Eastern and South East Europe. The project defined "hidden" both archives with poor digital presence and those institutions with digital information that is either not visible or not searchable. Those institutions and their holdings often remain neglected due to both the historical divisions of the Cold War and language barriers.

An institution's visibility can depend on the strength of its digital representation and the accessibility it provides to all available information on its holdings or items. Even some national archives, virtual key players in the field of First World War Studies and Medieval Culture, can be considered "hidden" if they do not provide accessible and structured information on their collections and holdings.

Regarding World War One research, CENDARI selected relevant institutions. Therefore, institutions in former front areas, capital cities, and heavy industry regions were given priority. Military archives and museums were considered crucial for inclusion in the Archival Directory. National (and State) archives, libraries and museums in all belligerent countries were prioritized over those in neutral and non-belligerent countries.

Additionally, a representative coverage of all administrative levels was achieved, to include municipal, regional, provincial institutions (e.g. "Länder", départements, etc.) while also incorporating non-state institutions like church archives, archives of the Red Cross, and archives universities and non-governmental societies or War Graves Commissions.

This was done so that complementary material from different administrative level creators and stored in different archives, would be available.

Apart from the holdings of national archives and libraries, the case study on Medieval Culture included a vast number of ecclesiastical organisations that hold relevant sources. In comparison to the immense amount of archival material available for research on the First World War, the number of available medieval manuscripts and archival material is more manageable, and the coverage of this material is more evenly distributed. This required and enabled a more granular approach to medieval sources and at the same time allowed for a more straightforward sharing of data. This was facilitated by the relatively clear and uncomplicated legal status of most of the medieval sources.

Methodology used in the CENDARI project for selecting and entering archival institutions and descriptions

Archival institutions relevant for the two case studies (First World War and Medieval Culture) were identified according predefined criteria.

World War I archives selection criteria:

- All countries that participated in the First World War were included.
- Special attention has been given to records especially in Eastern and South East Europe, in order to describe so-called 'hidden archives', which did not have any digital representation.
- Archival institutions with significant holdings were described, with priority given to central national archives, national military archives, national war museums etc.
- Archives that possess a large selection of digitized finding aids which are available online, but could not be included in the repository.

Medieval culture archives selection criteria:

- Archives in all European countries, including the USA and Australia, thus covering the most part of the Western countries holding medieval sources.
- Special attention paid to archives in Eastern and South East Europe.
- Collections of great interest to the academic community
- Archival institutions that have digitized a large part of their collections.

More than 1,200 descriptions of cultural heritage institutions were added to the Archival Directory AtoM. The work of the CENDARI team was not limited to only describing institutions, but also covered holdings, collections, records, and single items. These sources were selected by focusing on two case studies of the First World War and Medieval Culture, as well as the thematic Archival Research Guides, while giving special attention to archival descriptions from "hidden archives". More than 2.500 archival descriptions were inputted manually, selected from a transnational perspective and in different languages.

Information on these holdings was discovered in different ways. This could include information from archivists, research gathered by on-site visits, or the adaptation of descriptions from repositories, guidebooks and finding aids. Information on these holdings was translated into English in order to increase their visibility to the historical and archival community.

Regarding the First World War, "hidden" archives included those without digitally accessible information, in a language rarely mastered by researchers, or if these holdings contained different media types like films, photographs, audio files, which had been neglected by research. Even if these archival records have not been digitized and their availability is therefore limited, the consolidation of these sources has a value in itself. For example, if films are available only in a single archive and can be only viewed there, their description in the Archival Directory AtoM widens their visibility. Researchers are made aware of their existence, the possibility of consulting them, and encouraged to not limit their approach to a national point of view. The Archival Directory contains descriptions of film collections from various countries, thus supporting research from a comparative, transnational perspective.

Medieval collections in the Archival Directory have been selected mainly according to two criteria. For the aim of completeness, the CENDARI team selected descriptions using traditional scholarly reference tools, such as (but not limited to) Latin Manuscript Books Before 1600: A List of the Printed Catalogues and Unpublished Inventories of Extant Collections by Paul Oskar Kristeller and Sigrid Krämer (München, 19934) and Iter Italicum. A finding list of uncatalogued or incompletely catalogued humanistic manuscripts of the Renaissance in Italian and other libraries, by Paul Oskar Kristeller (London – Leiden – New York – København – Köln 1963-1997). Scholarly digital repertoires such as the databases and authority lists in the Mirabile platform and in the Integrated Archive for the Middle Ages provided by the Società Internazionale per lo Studio del Medioevo Latino and the Fondazione Ezio Franceschini were also used. The National Czech Library together with the Università di Cassino and other international partners (including the Italian Ministry for Cultural Activities, the French IRHT – Institut de Recherche et d'Histoire des Textes, the Belgian Bulletin Codicologique Scriptorium, etc.) provided additional support in this activity.

In particular, the authority lists and the shelfmarks available in the Mirabile platform had a basic relevance for selecting collections and institutions. These authority lists of libraries, archives, collections and single manuscripts are based on the bibliography collected in Medioevo Latino. Bollettino bibliografico della cultura europea da Boezio a Erasmo (secoli VI–XV) (Firenze, 1985), a bibliographical bulletin which covers the last 30 years of scholarly publications (journals, books, printed and digital resources) written in English, Spanish, Portuguese, Italian, French, German and pointing to some 300,000 records related to many disciplines in the field of medieval studies. Analysing critical editions of Medieval and Renaissance texts (written until 1536) and the secondary literature, Medioevo Latino helped to identify the most relevant collections and institutions.

Examples of 'hidden archives'

Discovering the Central Asian archives

Vast archival materials preserved in the countries of Eastern Europe and Central Asia are still not described properly by archivists and historians. Many of these holdings are not represented properly in archival guides and only a small part of these holdings have been digitized.

In the course of the project, CENDARI researchers found many archival collections relevant for First World War research, "hidden" in archival institutions of post-Soviet states. There are huge differences between the presentation of archival holdings between former Soviet republics, nowadays independent states. On one side there is the excellent example of the National Archives of Estonia, which provides open access to their archival holdings (see also The CENDARI White Book of Archives). On the other hand, there is little information on archives of Central Asia and their holdings. Only Kazakhstan has made progress in presenting archival sources online. This catalogue contains descriptions in Russian; however, they are rather brief, with few details on scope and content of holdings.

Central State Archive Republic of Kazakhstan

A complicated situation exists in the case of archival institutions in Central Asian states such as Tajikistan, Uzbekistan, and Turkmenistan. In this case, researchers will face difficulties in finding what archival data exists and whether it can be relevant for their research.

The below-mentioned national archives have not yet provided information on their collections at all in the internet:

Central State Archives of Tajikistan

Central State Archives of Turkmenistan

Central State Archive of Republic Uzbekistan

At the moment a researcher would have to visit these archival institutions in person to gain complete information on the holdings in situ. In the archives there are no English-language catalogues. In the reading rooms, the guidebooks could be available both in national languages and in Russian. Some guidebooks of archival holdings in Russian were published in the Soviet period before 1991. Due to political reasons (as the Civil war in Tajikistan in the 1990s or political isolation of some republics) and economic hardships many archival institutions are unable to create new archival guidebooks. Even those that have been published in recent decades are not available in libraries abroad nor online. That is why the Soviet archival catalogues are currently the best guides to some archival collections of the region of Central Asia.

Some archival collections may have been renamed and could be available under a new reference code. Furthermore, the archival fonds could have been merged or divided between other archival fonds or collections. In addition, contacting and negotiating with these archives is not an easy task as some of these countries are still politically isolated.

In the case of the Central Asian archives, the information on the archival collections (fonds) was reconstructed by research based on printed materials and secondary records on the history of Central Asia during the Soviet era. The primary sources, which had been published in the Soviet period, were helpful to define relevant archival fonds (collections) with proper data (reference code, title, timeframes). Articles produced by some Russian archivists and historians familiar with Central Asian archival holdings were also consulted.

"Trophies of the War": The Russian state military archive

Another example of a "hidden archive" are the collections of the Russian State Military Archive (RGVA) in Moscow. This archive preserves a great amount of archival sources from Europe, which were seized by the Soviet Army at the end of the WWII. During the Second World War many archival collections were seized by the fighting sides. Many state and private archives were looted by the Nazis, and at the end of the war were discovered by the Soviet and U.S. armies. Such archival collections were considered as "trophies of the War". The Soviet special command transported the archival documents to Moscow. They were placed separately in a newly established secret Special Archive of the USSR (in Russian Osobyi Arkhiv). The Special Archive received and housed confiscated German state records as well as captured Belgian, Polish, Austrian, French, Dutch and other state and institutional records. There were also private archives, which had been confiscated by the Nazis and then seized by the Red Army, NKVD and other Soviet organs in Eastern and Central Europe. The Special Archive also holds many Jewish collections looted by the Nazis. These include the records of the Jewish community of Salonika, papers of the Rothschild families of Paris and Vienna, and documents of the Alliance Israélite Universelle in France.

The building of the former Special Archives, now a part of the Russian State Military Archive (photograph taken by Andrei Zamoiski, October 2014).

The Special Archive was unknown and inaccessible to historians in the Soviet period until Perestroika. In 1999 the Archive was merged into the Russian State Military Archive (the main building of the RGVA is located nearby). In the last decade some archival collections were restituted to their rightful owners (i.e. Austrian archives).

In the online catalogue, a researcher can find only brief information about the holdings, such as the reference code and the title of the collection. There is no online description of scope and content. Regarding some German collections seized by the Red Army, see the website Sonderarchiv Moskau. The German historian Sebastian Panwitz has transcribed some collections of German origin and placed them online.

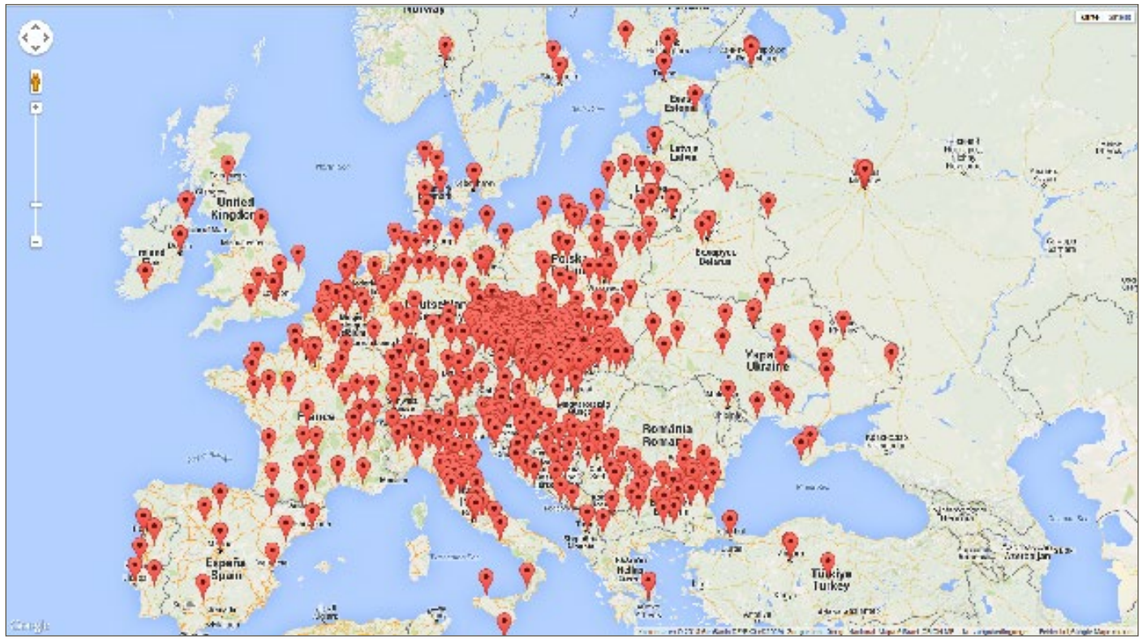
In the reading room of the RGVA one can look through a registry (opis) of the archival collection in Russian and order several folders per day.

Material in the Archival Directory AtoM: Numbers

As of September 2015, the Archival Directory AtoM contained information on more than 1,200 institutions and more than 4,750 archival descriptions relevant for research on Medieval Culture and on the First World War.

Regarding archival institutions, all European countries are covered, with an emphasis on countries referred to in the selection criteria. Furthermore, institutions in many non-European countries were also included in the research, if they were, for example, participants in the First World War. The following maps show the distribution of archival institutions contained in the CENDARI archival directory as of May 2015:

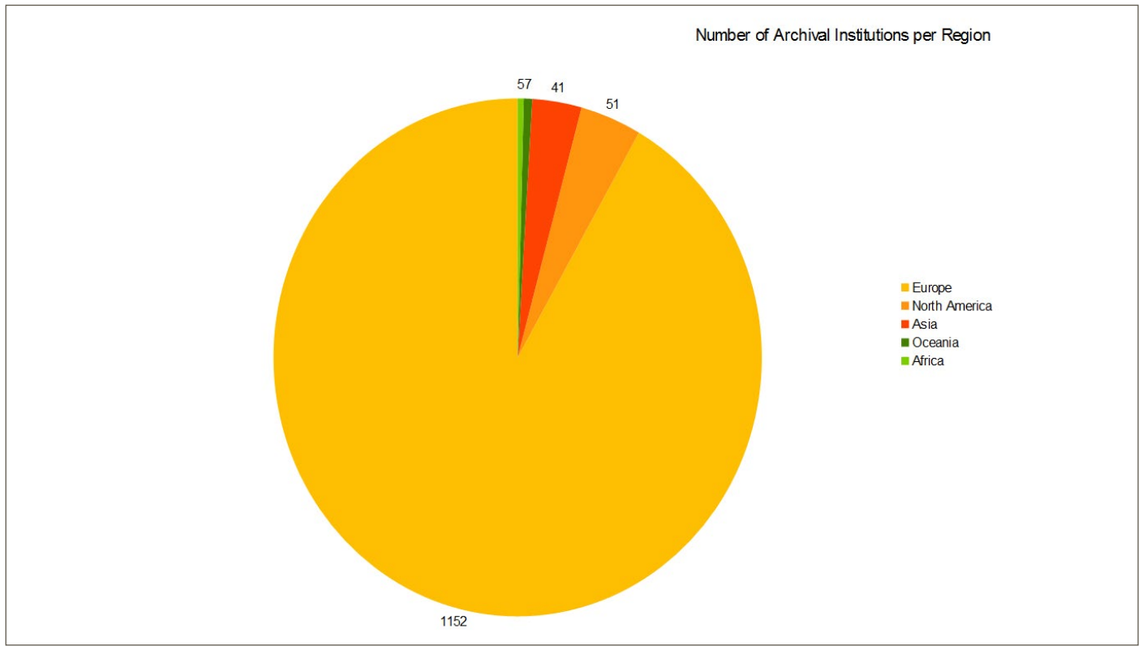
Worldwide distribution of institutions

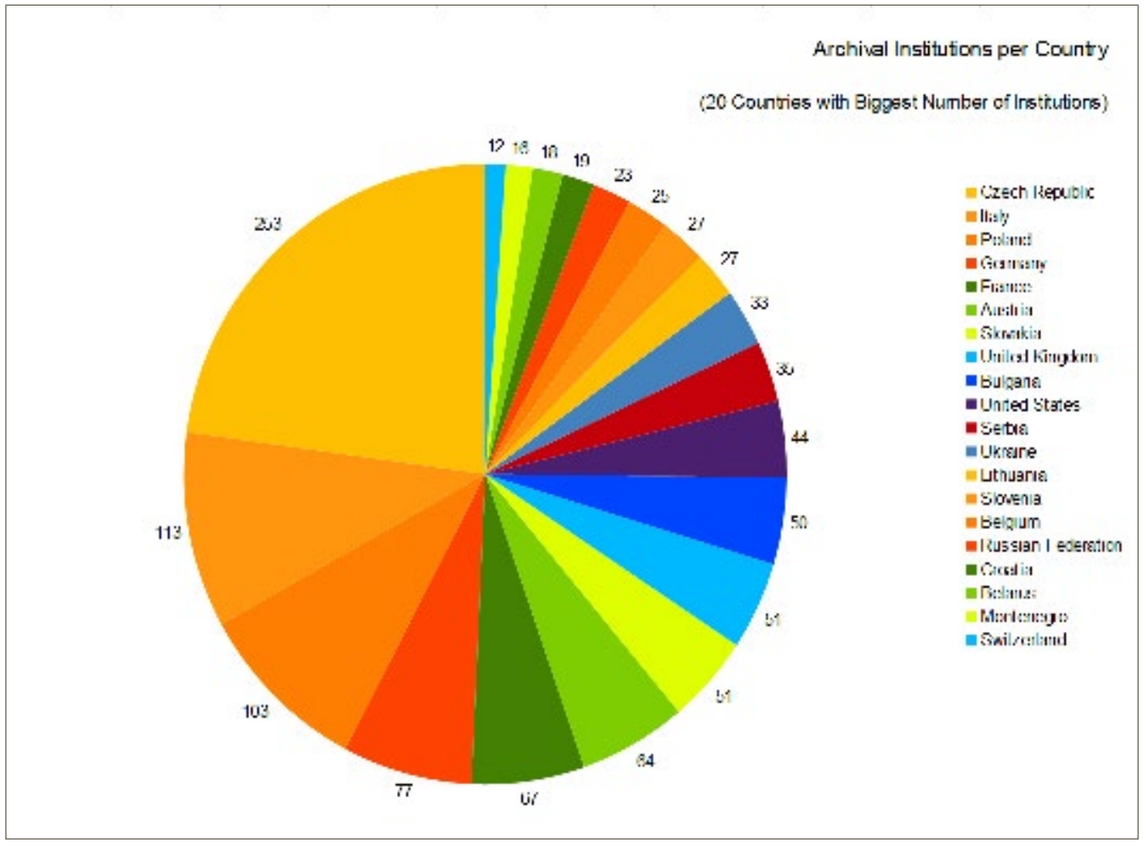


Distribution of institutions within Europe

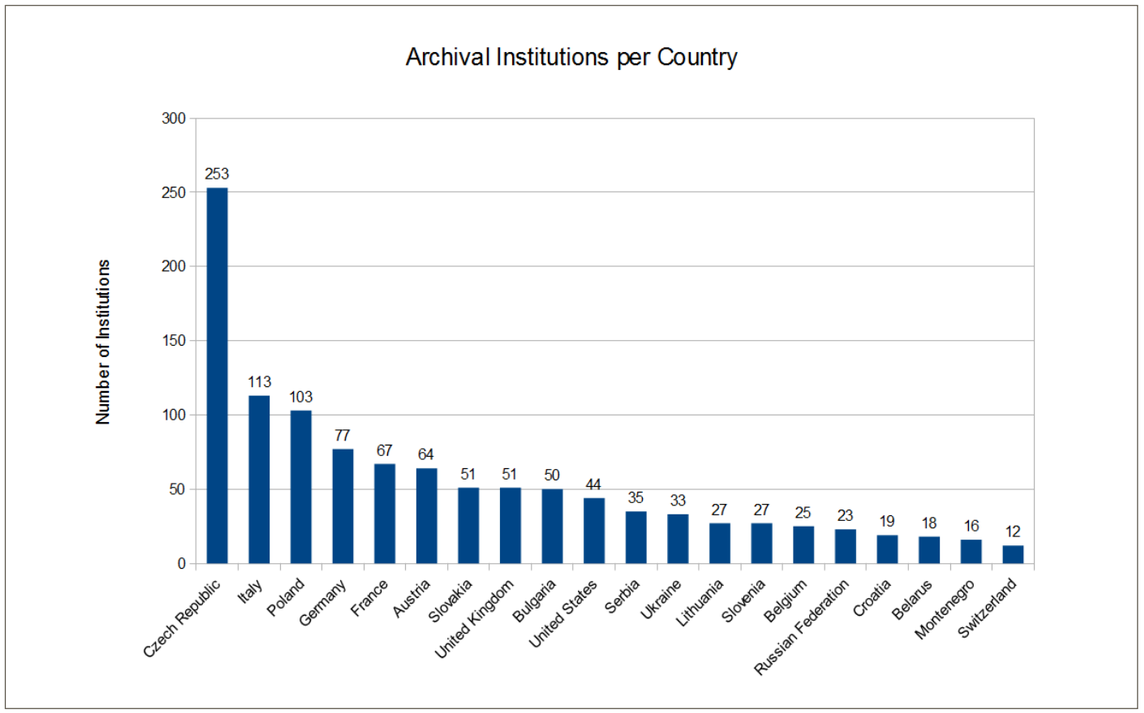
The following charts allow a closer look at how many institutions per country were recorded. The numbers show the status as of September 2015:

Less than 10% of all archival institutions described in the Archival Directory are not located in Europe. Most non-European institutions were added because some of their holdings are relevant for First World War research, such as archives in India, Australia or South Africa, all of them being nations participating in World War One.



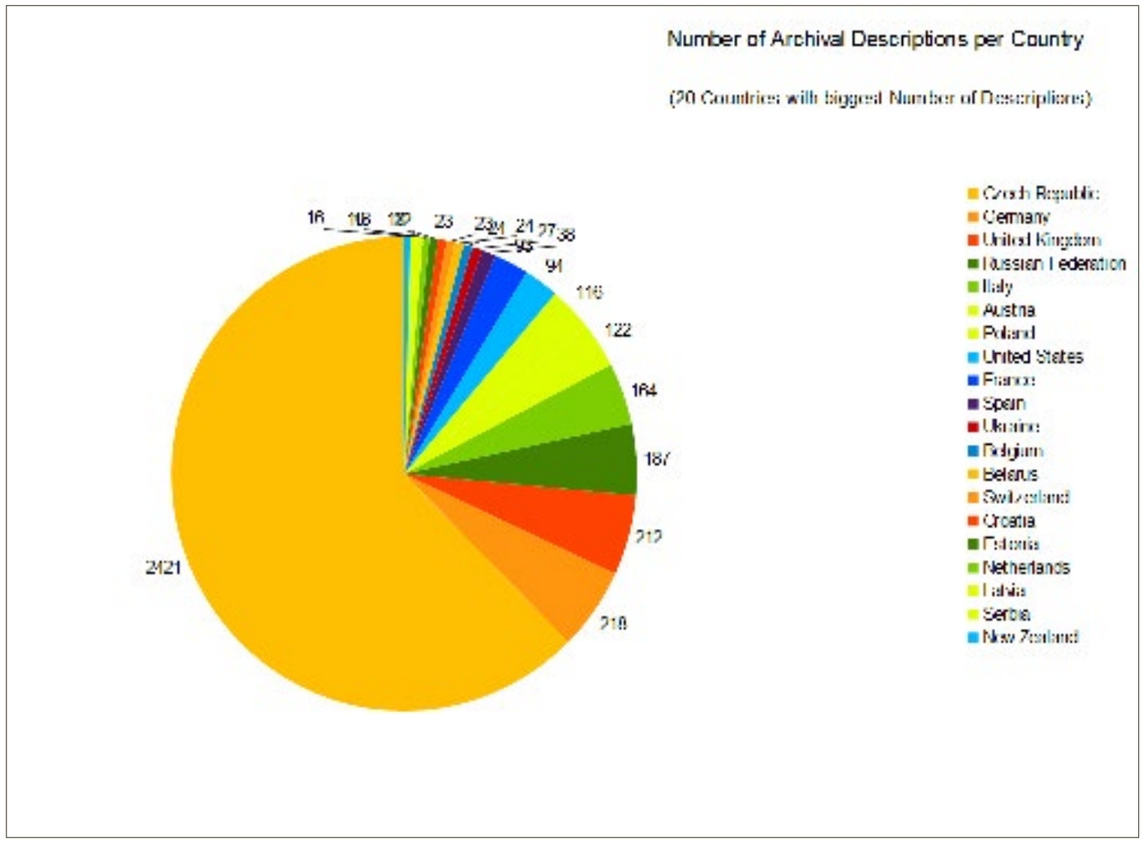


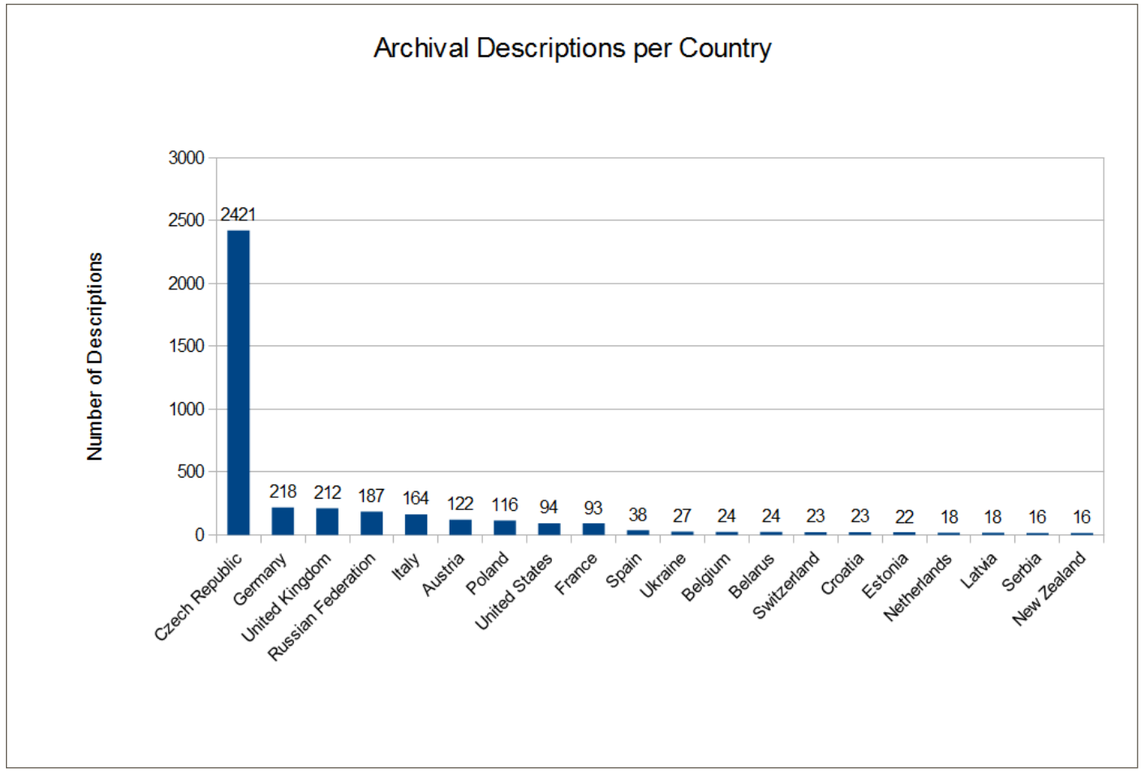
The pie chart above shows the number of institutions in 20 European countries.



The same numbers as in the pie chart can be seen in the histogram above. Here the large amount of archival institutions from the Czech Republic is even clearer. The large number of institutions in the Czech Republic (and in Italy) can be explained by two factors. Firstly, the medieval Czech and Italian teams within CENDARI focused on institutions which they know best and which they considered to be relevant to their own Archival Research Guides. Secondly, an automated import from the Czech Department of Archives Administration led to a disproportionately high number of Czech institutions in the Archival Directory AtoM. It was common practice within the CENDARI project to ingest data from cultural heritage institutions into the CENDARI repository. However, in this case an exception was made. As the Czech data providers were interested in publicizing the collaboration with CENDARI and needed something to exhibit, it was agreed that the Czech records would be made publicly visible and searchable. This normally would not have been the case, since the CENDARI repository can only be searched by logged-in users.

Archival descriptions were added to the Archival Directory at the level of collections and holdings for the case study on the First World War, and on the level of one single item (manuscript) for the case study on Medieval Culture. The following charts show how many archival descriptions per country were established in the Archival Directory. The numbers provided present the status as of September 2015. Of the 4,750 archival descriptions contained in the archival directory, about 2,700 were established through manual encoding, whereas about 2,050 datasets were automatically imported from the Czech Department of Archives Administrations, mostly regarding the First World War.





Finally, regarding what proportion of these archival descriptions can be ascribed to World War One or Medieval Culture, around 500 archival descriptions do not have period indications since these were missing from the consulted archives descriptions. Around 3,000 records can be ascribed to the period of the First World War, and 1,000 records can be attributed to the medieval period. This disproportion stems from the automated import of files from the Czech Department of Archives Administration mentioned above, where were all relevant to the First World War. In other words: The CENDARI team produced about 1,000 manually established archival descriptions for each case study on the First World War and on Medieval Culture.

WHAT IS NOT INCLUDED – AND WHY IT IS A "WORK IN PROGRESS"

Many cultural heritage institutions are currently in the process of establishing digitally available descriptions of their holdings. Therefore not all archival material is available in digital format, not even on the level of collections and holdings, and the CENDARI project cannot aim to provide an exhaustive and all-encompassing directory of archival material. The selection of sources always depends on the thematic focus and methodological approach chosen by the researcher. With historiographical approaches, focal points and methods being newly established or shifting in time, different types of sources are being examined. The best example for this is the inclusion of private diaries, letters and memoirs in historical research in the last decades, parallel to the establishment of archives specialized on collecting private writings and other ephemeral records or the emergence

of crowd-sourced collections in digital form, such as has been provided by the project "Europeana Collections 1914-1918". Nevertheless, the CENDARI project aimed to find a balance between the focus on 'hidden archives' and their holdings and the inclusion of main collections known to scholars. This goal has been achieved by the parallel work of establishing archival descriptions manually in the Archival Directory AtoM and at the same time the automated ingestion of metadata into CENDARIs repository.

The limitations of the CENDARI project

While it was possible to encode information on more than 1,200 institutions manually, this workflow could not be maintained at the level of describing holdings, record groups, collections or single items and artifacts. Moreover, in many cases the relevant digital information already exists and the provision of archival descriptions in digital format does not need to be repeated. It was therefore the task of the CENDARI team to make sure that available data were incorporated into the CENDARI repository via an automated ingestion of source information in order to broaden the archival directory further and ensure the usability of existing data.

Notable examples regarding the ingestion of available metadata from cultural heritage institutions are the cases of examples of Europeana and The European Library together with the projects "Europeana Collections 1914–1918", "Europeana Newspaper" and "European Film Gateway". Other major contributors to CENDARI's repository were, amongst others, the German Bundesarchiv, Archives Hub UK, ICRC Archives Geneva, the Istituto Centrale per gli Archivi. In general, no digitized objects were incorporated into the repository, only metadata. This was due to technical limitations (storage) as well as the legal framework in which digitized media often has a copyright claim, even if there are no copyright restrictions for the object itself. If an institution provides digitized objects (for example, Europeana Collections 1914–1918), links to these digitized media are provided in the metadata in order to enable users of the CENDARI infrastructure to quickly assess this material.

Researchers can follow their research interests and find relevant material in collections and archives not yet contained in the archival directory. Archival institutions and descriptions can be added and need not be limited to the CENDARI case studies on the First World War and Medieval Culture. The scope of the content can be broadened into potentially all historical domains. The 1,200 institutions and 4,750 collections in AtoM thus provide for a starting point as well as examples for best practice. Users are invited to add descriptions of archival institutions and archival sources, enlarging in this way the knowledge base. Moreover, all newly added descriptions in the Archival Directory AtoM are migrated to the CENDARI repository and indexed over there on an automated basis. This service makes new material available within the whole CENDARI infrastructure. Researchers can find these descriptions via the Note Taking Environment, via a faceted search or the CENDARI ontologies. In this way enriched search results will be made available for all the users of the CENDARI virtual research environment. For these reasons the Archival directory can be seen as a continuous "work in progress", to the benefit of all future users.

BIBLIOGRAPHY

- David E. Fishman, *Securing Our Inheritance: the fate and state of Jewish documentary heritage in Europe*. London : Rothschild Foundation (Hanadiv) Europe, 2015.
- David E. Fishman, Mark Kupovetsky, Vladimir Kuzelenkov, eds. *Nazi-Looted Jewish Archives in Moscow: A Guide to Jewish Historical and Cultural Collections in the Russian State Military Archive*. Scranton: University of Scranton Press, 2010.
- Michael Kurtz, *America and the Return of Nazi Contraband: The Recovery of Europe's Cultural Treasures*. Cambridge: Cambridge University Press, 2006.
- Mikhail Mitsel. *The Final Chapter: Agro-Joint in the years of the Great Terror* [in Russian]. Kiev: Dukh i Litera, 2012.
- ISAD(G): General International Standard Archival Description (2 ed.), (Ottawa 2000)
- Jessica Bushey, International Council on Archives (ICA) "Access to Memory" (AtoM): Open-source software for archival description.
- International Council on Archives on ISAD(G) (also includes accepted translations of all of the terms)
- International Council on Archives Committee on Descriptive Standards
- UNESCO Archives Portal (provides information on archival institutions all over the world)
- International Standard Archival Authority Record. For Corporate Bodies, Persons and Families
- ISDIAH: International Standard For Describing Institutions with Archival Holdings

- International Standard Archival Authority Record. For Corporate Bodies, Persons and Families