

Archives Portal Europe – A Challenge of Harmonisation and Outreach

*Karin Bredenberg, Silke Jagodzinski
APEX project*

The Archives Portal Europe

The portal

The Archives Portal Europeⁱ gives joint access to European archives in the Internet. As such it allows a comprehensive and cross-border research in archival descriptions and according context information from 27 European countries.ⁱⁱ Next to other European portals, like the cross-domain site Europeanaⁱⁱⁱ for digital objects in archives, museums, libraries, and audiovisual collections, and the libraries' portal The European Library^{iv}, the Archives Portal Europe gives easy access to the rich European cultural heritage. Users can start their research in the archives directly and they often might find new connections and materials and maybe get new ideas for their research.

„The Archives Portal Europe allows us to make new discoveries without knowing what the end result will be. Users often are surprised about what they find out.“^v

Search and content in the portal

The Archives Portal Europe provides a multilingual surface for research and browsing. The simple full text search is mostly used to access the portal's information. An auto-completion function provides similar terms with the same root, which helps to search with multiple terms in different languages. A conducted simple search leads to results in three types of provided information:

archival material, what are in fact results in archival descriptions of the material;

names of and information about record creators, which can be persons, corporate bodies and families,

institutions, which are information about the participating institutions in the portal.

Most important is, of course, the search for **archival material**. The archival material is described by archivists and these descriptions are searchable in the Archives Portal Europe. A search engine indexes the texts and analyses the provided data. Therewith, next to the full text search, refinements for defined facets, like date of creation, hosting institution or contained digital objects are possible. Furthermore, the results can also be displayed in a context view, which shows the hierarchical correlation of a file or an item in its archival description. This context view is based on the three layers principle of the Archives Portal Europe, which allows the hierarchical browsing from the European to the national archival landscapes to the single institution with its holdings guide down to the fonds and their respective structures. The same approach of hierarchical access to archival materials is given within the advanced search that allows users to not just conduct a search, but to explore the fonds and file descriptions along these structures beforehand.

The detailed archival description is displayed with the according title, reference number, dates and all other descriptive information provided by the hosting institution. Of course, the institution itself, most of the time identical with the content provider, is named, too. Users can print the full information or share via social networks and they can send feedback to the archive with comments or questions about the record or item they have found.

Next to the search for archival material there is a search for **record creators** available. Access to archival material via their creators is often more intuitive and easier for users. Furthermore,

the tradition of archival descriptions states the *principles of provenance*, which means the definition of fonds by the criterion of their creators in order to understand the context of archival material.^{vi} Names and other provided information about creators are part of the full text search index. Again the list of search results can be filtered with refinements, like dates of existence, language of the description or the entity type. The most useful information for users are probably the provided relations from one record creator to archival material or to archives keeping this material or even to other record creators. Further developments in the portal are going to visualise networks between persons, families and corporate bodies and between these entities and fonds and archival material.

The third search option is about **institutions**. A directory provides detailed information about the archival institutions participating in the portal. Next to address and contact information, a lot of information about visitor and user services and descriptions about the holdings and history of an archive are given. Users can search for this information and furthermore find an institution via a map. Taken up the archival landscape again, users can browse from European level via country through the archival landscape down to an institution.

The portal service is topped off with the **featured document**, which presents the most interesting and exciting digital objects of archival items from the partner institutions. This digital exhibition gives users an insight in the archival collections all over Europe and provides a stage for archives to show their treasures.

Users and archives

The Archives Portal Europe was visited by more than 20,000 people each month in the first six months of 2014.^{vii} The number of users is increasing strongly each month since the end of 2013. The portal's content is also indexed and can be found by search engines like Google, Bing and Yahoo, which also drives a lot of traffic to the portal. Most visitors come from the various countries in Europe. But also some oversea countries like the US and Brazil are part of the top twenty visiting countries of the portal.

A non representative survey among the portal users in the beginning of 2014 showed that most of the users are archivists or information professionals. Apart from that, the virtual user of archives seems to be the same as the real user in the reading rooms where history enthusiasts, like genealogists or hobby researchers for local history, scientific researchers and students are the three biggest user groups.^{viii}

Content providers in the Archives Portal Europe come from 27 countries, whereas institutions from 32 European countries are currently partner in the portal's network. The numbers and types of archives differ from one country to another, depending on the ability and willingness of the institutions. Whereas the portal started with the national archives and national archives administrations as the initial project partners, all kind of archival institutions are now part of the portal. Even archives in research institutions, museums and archival collections in libraries provide their content.

The use of the portal is free of charge for the users and for all content providers.

The projects behind the portal and future plans

The portal was developed by two international projects funded by the European Commission. The Archives Portal Europe network (APEnet) project, from 2009 to 2012, designed the portal in its logical model and technical structure and launched the first public version in January 2012. The successive project Archives Portal Europe network of excellence (APEx), from 2012 to 2015, is expanding, enriching, enhancing and sustaining the portal. Whereas one goal is to gain more archival institutions as content providers, to thereby extend the searchable archival descriptions in the portal, and, with this, to reach a higher number of portal users, the project

follows up on further technical developments in order to meet users' requirements and content providers' demands.

In 2009 the **APENet** project started with twelve partners and seven associated members that joined throughout the project lifetime to build an „Internet Gateway for Documents and Archives in Europe“, which was recommended by the European Commission as early as 2005.^{ix} Meanwhile, the work of the APENet project continues in the successor **APEX** with 33 partner institutions from 32 countries as per today. The work is divided into eight work packages, led by the partners in the Netherlands, Germany, Spain, France, Estonia and Sweden. A technical coordinator, a scientific coordinator and a development coordinator supervise the progress and development. The project board, consisting of the work packages leads, is managing the daily work with all partners. The Executive Steering Committee with representatives from each partner is the highest leading and decisive body for the whole project.

In order to ensure the mid- and long-term existence and progress of the portal, the APEX partners are establishing a foundation. The Archives Portal Europe **Foundation** will be located in The Hague, Netherlands, and will not only host the portal but also work on further technical developments, on business models and on raising more partners, content and users.

Providing content to the Archives Portal Europe

From the very beginning, the Archives Portal Europe network's approach was to enable content providers to manage their data in the portal independent from a technical crew or from an ingestion team. Within both projects, two tools have been developed, which allow content providers to prepare their data for the portal, to upload their files and to define when the data should be published. With this decentralised **workflow**, each partner can of course also decide to remove data from the portal or the whole system. The same principle works for the delivery of archival descriptions to Europeana. Each content provider can define, if, how and when to send their files to the cross-domain portal.

The two **tools** mentioned above are provided for data preparation for the portal. The usage and definition of international archival standards for the portal, which are outlined in the next chapter, require providing the possibility to turn the content providers' data into these formats. Both tools have the same set of functionalities and workflows, and even a similar navigation. The **Data Preparation Tool** is a standalone tool, which can be downloaded and installed locally. A connection to the Internet is not necessary for usage. Content providers are able to convert their files into the defined formats for the Archives Portal Europe and for Europeana and furthermore, they can create files with basic information about the archival institution and about record creators via forms. The **dashboard** is the main central instrument for data management and at the same time also a tool for data preparation. In both tools, content providers can validate their files, which is a check for technical validation; furthermore in the dashboard, the upload and publication of these files as well as their delivery to Europeana is possible.

The login to the dashboard is given to a person responsible for the data of one contributing archive, called the Institution Manager, and guarantees access to the data management. The main contact person for an Institution Manager is the according Country Manager, who acts a representative of the Archives Portal Europe for all contributors in one country and who is in contact with the Archives Portal Europe team. A Country Manager community was established by the APENet and the APEX projects during the last five years to enable them to work together in, for example, marketing questions.

The Challenge of Harmonisation

One of the main objectives for the Archives Portal Europe is to bring archivists from Europe together on a technical basis. Finding aids and archival descriptions from different countries and

different institutions must have identical definitions for the structured information and the portal must ensure to present and use this information in the same way, no matter which language, script or tradition of archival description a content provider has used.

International archival standards

The International Council on Archives defined common standards for the description of archival material and for the description of archival authority records. These descriptive standards provide general guidance for the preparation of archival descriptions and define specific elements of information about archival material and for preparing archival authority records.^x

Both standards have a corresponding definition for encoding this information in Extensible Mark-up Language (XML) for technical purposes. These technical standards are called Encoded Archival Description (EAD) for archival material, based on the General International Standard Archival Description (ISAD(G)), and Encoding Archival Context – Corporate bodies, Persons, Families (EAC-CPF) for authority records, based on the International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR (CPF))^{xi}. Both standards define the structure of an XML file with the aid of a schema in order to encode the demanded information. Files, which are created according to the EAD schema or the EAC-CPF schema, can be exchanged between archives, portals and/or databases.

EAD and EAC-CPF were developed as international archival standards for data exchange. At least EAD is established and a de-facto standard in the archival world since more than ten years. Having an international standard available, gives the most suitable solution for an international project and for an archives portal to use these standards, as the APEnet and APEx projects did. EAD is used for the ingest of archival descriptions, like finding aids, holdings guides and source guides, into the portal. The data are used for the „Search in archives“, ie the search in archival material. EAC-CPF is used for the ingest of record creators' descriptions, which enables the „Search in names“ function in the portal.

But even if the standards already define the encoding for finding aids and authority records, the variety of using the elements and attributes is very broad. Therefore the Archives Portal Europe has defined its own stricter profiles of the international archival standards. The so-called ape-profiles define EAD and EAC-CPF furthermore also in a technical sense. For example some elements and attributes are not used or for specific attributes certain values have been defined. With this definition of our own, schema files were created to ensure content providers to follow this definition. The schema files are the basis for conversion and validation during data preparation with the Archives Portal Europe tools.^{xii}

Defining standard schema profiles for the Archives Portal Europe

In both projects, APEnet and APEx, a work package for defining standards and guidelines has existed. Within this work package archivists from all over Europe work together on the standards' definition for Archives Portal Europe purposes and on their implementation. During the APEnet project, a profile for EAD, called apeEAD, was defined. For this task a workflow was created and successfully established to align the partners' opinions, demands and experiences and the input from the technical team, the portal's programmers.

First a survey was spread among all APEnet and APEx partners. In general the survey asked for usage of the standard, experiences and interest to develop the according ape-profile. The comprehensive and detailed feedback shows, how many project members can contribute to this task and in which way. Based on this, tasks are distributed among smaller working groups, sometimes during a face-to-face meeting in form of a think pool.

After that, all interested partners have been asked to provide their schema file for the standard or, if a schema is not available, a list of used elements and attributes. A spreadsheet with all

possible elements and attributes proved to be helpful for archivists to answer this question. The replies about the elements and attributes have then been compared in order to finalise the central profile. Next to the elements that are already mandatory in the general standard, at least half of the partners have had to use an element to define that for the ape-profile. If an element is not mandatory and not used by the partners, there is no need to take care about its mapping and conversion. Of course, convincing use cases and arguments could lead to exceptions from this policy. An active and useful discussion among the archivists in the project leads to practical solutions.

In a third step, a draft schema was created and all partners were asked to encode their information in the ape-format with the help of the tools provided. At the same time, a discussion for functionalities in the portal, based on the standards' definition, started among the work package task groups. The questions: What is possible? What is feasible? What is useful? were discussed.

At the end the draft profile with example files and use cases, a verbal description of the profile, which explains the usage for the portal, and proposals for functions in the portal were published in a state of the art report and delivered to the European Commission.

This workflow, developed for apeEAD was repeatedly used for apeEAC-CPF and apeMETS, the Archives Portal Europe's profile of the Metadata Encoding and Transmission Standard^{xiii}. A similar procedure was used for the development of an own definition for Encoded Archival Guide (EAG), which is based on the descriptive archival standard International Standard for Describing Institutions with Archival Holdings (ISDIAH)^{xiv}.

The apeEAD profile

Setting up a schema profile of our own based on an existing one means to decide between possible options and to possibly further restrict the usage of elements and/or attributes. For apeEAD, plans for technical implementation in the portal and intended functions were an important parameter for the definition.^{xv}

With this in mind, the list of mandatory elements, ie the amount of **mandatory information** for delivering data to the Archives Portal Europe, was extended. These additional mandatory elements are mainly needed with regard to the central data management and processing. The tools for data preparation add this information automatically in case they do not already exist in the content provider's file.

For example, the element <eadid>, already mandatory to exist according to the general standard, is required to provide the identifier for the EAD file and must be delivered with an identifier of the institution in the attribute @mainagencycode. Ideally, this is an registered ISIL code^{xvi}, otherwise a standardised syntax based on ISO standard 15511 is required. Also a two-letter code for the country, where the institution maintaining the described archival material is located, is necessary to be available in the attribute @countrycode. The identifier of the institution and the one for the EAD document compose the required global identifier in the attribute @identifier for the EAD document within the portal.

The **values for attributes** can be fixed or limited in the schema. apeEAD defines to use only codes and formats according to certain international standards for dates, the language, a country, a script or a repository.^{xvii} Another example for defined attribute values is for the attribute @level in the element <archdesc>. The element <archdesc> indicates the beginning of the archival description, in contrast to the element <eadheader>, which contains administrative data about the finding aid and the EAD file as such. The attribute @level indicates the character of the described unit. Whereas the general EAD schema allows a list of terms here, apeEAD defines all archival descriptions as descriptions for fonds. So, „fonds“ is the only term allowed as attribute value at this level of description.

Language information is an important aspect within the multilingual Archives Portal Europe. In order to read and use language information technically, it is necessary to use a language code according to the defined standard additionally or instead of the word itself. Therefore it is defined to always use to the element <language>, which allows the attributes @langcode for language information and @scriptcode for script information. In the general EAD schema, information about the language of the described material can be described directly in the element <langmaterial> as well, but that does not allow to enter a language code.

Also the encoding for the **levels of description** is more restricted in the apeEAD schema than in the general one. In the Archives Portal Europe, the schema does not use possible enumeration from <c01> up to <c12> in order to allow a deeper hierarchical structure, if needed. Furthermore the attribute values for @level, which describes the character of the component, is limited to "fonds", "series", "subseries", "file", and "item".^{xviii} During the data preparation the conversion adds the allowed values automatically according the defined mapping between local data formats and apeEAD.

The apeEAC-CPF profile

The same aspects, which influenced the definition of apeEAD, have concerned the work on apeEAC-CPF. In order to limit the variety of encodings for the entities description, some information has been defined as mandatory additionally to the general EAC-CPF schema and attribute values.

A very consentaneous decision between the discussing European archivists in the APEx project was the **rejection** of the possibility to describe multiple identities of an entity within one EAC-CPF file. For reasons of acknowledging the provenance of the EAC-CPF file, ie the content provider having created this information, as well as handling of EAC-CPF files in the Archives Portal Europe, the element <multipleIdentities> was restricted from use in the apeEAC-CPF schema. It has been anticipated, that having constituent files instead would allow for more flexibility in the central system.

Furthermore, for data processing it was necessary to define additional elements as **mandatory**. Most of them can be filled automatically during the export of files from a database or during the conversion to apeEAC-CPF. These are mostly administrative data for the description, like the maintaining agency or the name of the agent, who created the description.^{xix} Additionally mandatory in the entities' descriptions are the dates of existence. Within the Archives Portal Europe there are plans to process EAC-CPF data automatically in the future. Matching newly uploaded files with already stored files from other content providers or with external authority files is also planned later on. For this, additional information to the name of an entity is necessary. If an institution delivers data without dates of existence, they must be added or, in case they are unknown, defined as unknown.

apeEAC-CPF has more **limited vocabulary** for attribute values, than apeEAD has. A good example might be the definition for names. Multiple names of an entity can be encoded as single name entries. These might be translations, pseudonyms or even abbreviations of names. Each name entry can be split in parts. To define the single parts of a name entry they can contain an attribute @localType. In apeEAC-CPF the values for the attribute @localType are limited and therefore name parts can only be: name of a corporate body, family name, a person's name, first name, surname, birth name, title, prefix, suffix, alias, patronymic or the legal form of a corporate body. These definitions are used for search options and refinements, as described in the chapters above, and for a proper name display in the portal. For this the order of single <part> elements has been defined. Of course, the attribute @localType is not mandatory and partners can deliver their data without this information. But for data quality and consistence, it is nice to have and its use is therefore recommended.

Other international encoding standards in the Archives Portal Europe

Without doubt, apeEAD and apeEAC-CPF are the most prominent and important standards in the portal. Next to these two, a new and extended version of EAG was defined and published in 2012. First EAG was created by the Spanish Ministry of Culture in 2002 for the Censo Guía de Archivos de España y Iberoamérica^{xx}. In order to enhance information about archives and to provide a directory for archival institutions in Europe, the first EAG version was revised by the APEX project. The updated EAG 2012 is already implemented in the Archives Portal Europe.^{xxi}

On request of some partners, the possible usage of METS^{xxii} for the portal was evaluated. METS is a standard for describing digital objects and their structural connections. The structural connection can be used, for example, for creating a view of images described in the METS file. Even if only few archives need to create their archival descriptions in connection with METS files, an own profile was defined as apeMETS. More interesting is the related profile for apeMETSRights, which enables archivists to assign specific legal information about the rights of use and re-use to a single digital object described in the METS file.^{xxiii}

One of the tasks of the Archives Portal Europe in the landscape of European cultural heritage gateways is data aggregation for Europeana. In close collaboration with Europeana, the APENet project defined a mapping for archival descriptions to the Europeana data format, Europeana Semantic Elements (ESE). This mapping was updated after Europeana changed its data model to the Europeana Data Model (EDM) in 2012. In 2014 the APEX project again revised the conversion from apeEAD to EDM in order to provide proper hierarchical information for archival descriptions. With that latest update, the Archives Portal Europe delivers not only descriptions for digital objects, but contextual information as provided in the finding aid, as well. Furthermore record creators data from apeEAC-CPF will in future be added for their integration in Europeana.

The Challenge of Outreach

Within the two projects APENet and APEX, the Archives Portal Europe has established a network of archivists and information professionals all over Europe. Cooperations, collaborations and discussions exist with non-European bodies and non-archivists' projects. One of the main benefits of the Archives Portal Europe is the professional exchange of opinions and knowledge among its members.

Especially the work with archival standards is an excellent example for trans-European collaboration and formation of opinions. During the discussions and detailed work with international archival encoding standards and the implementation in the portal, the network was able to give valuable feedback towards the standards to the according responsible maintaining bodies. Also for the revision of EAD to a new version, EAD3, the partners of the Archives Portal Europe were able to contribute a common European note.

Next to these positive aspects, there still are challenges for European archivists and the portal. A question, which not only raises in the Archives Portal Europe but among other archivists and information professionals and among users working in and with online portals, is the quality of the provided data. The Archives Portal Europe tries to constrain a certain quality by the ape-profiles definition for archival standards. But of course, this can only effect the technical usage of elements. The way, in which they are filled, and the information, that are provided, still are the content providers' business. Titles, which are in fact only quotation marks, because in a printed finding aid, they indicate a repetition, are not useful for anybody. There is no information left. If there is no additional description, users cannot even search for this item. The same goes for titles like „Foto“ or „Photo“ and no additional description or the plain description „album“. Better awareness from content providers, how the information can be used outside the local system and how they are presented, would be necessary. The usage of quality management could be

one of the challenges coming along with the publication and dissemination of archival descriptions.

Of course, the main reason of the portal's existence is to provide archival information for users. Therefore, attracting new and more users stays a very important aspect of the portal. Not only more and new information from archives, but new and state-of-the-art functions and technologies are needed. Constant progress, evaluation and development of the portal will be the biggest challenge for the Archives Portal Europe Foundation in the next years.

Notes

ⁱ The Archives Portal Europe is accessible via URL <www.archivesportaleurope.net>. [Consulta: 01/09/2014].

ⁱⁱ At the moment of creating this text, the number of countries with archival descriptions in the Archives Portal Europe is 27. The portal is going to grow each week, so the number might be higher by the time of your reading.

ⁱⁱⁱ Europeana is accessible via URL <www.europeana.eu>. [Consulta: 01/09/2014].

^{iv} The European Library is accessible via URL <<http://www.theeuropeanlibrary.org>>. [Consulta: 01/09/2014].

^v MENNE-HARITZ, Angelika (2014): "Cross-border Discoveries in the Archives Portal Europe". <<http://www.apex-project.eu/index.php/en/articles/189-cross-border-discoveries-in-the-archives-portal-europe>>. [Consulta: 01/09/2014].

^{vi} See LAFUENTE URIÉN, Aranzazu (2014): "Archival authority control: an introduction to Encoded Archival Context for Corporate Bodies, Persons and Families (EAC-CPF)". <<http://www.apex-project.eu/index.php/en/articles/184-archival-authority-control-an-introduction-to-encoded-archival-context-corporate-bodies-persons-and-families>>. [Consulta: 01/09/2014].

^{vii} The number of visits along with other aspects are analysed regularly with Google Analytics.

^{viii} The results of the survey were forwarded to the European Commission with the internal Usability evaluation report for release 1.2 of the Archives Portal Europe in May 2014.

^{ix} Official Journal of the European Union, 29.11.2005; 2005/535/EC

^x See International Council on Archives, "General International Standard Archival Description (ISAD(G)", 2nd ed., 1999, p 7, <<http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html>> [Consulta: 01/09/2014]. Information about ISAAR(CPF) are available in "International Standard Archival Authority Record For Corporate Bodies, Persons and Families (ISAAR(CPF))", 2nd ed., 2003, <<http://www.icacds.org.uk/eng/ISAAR%28CPF%292ed.pdf>>. [Consulta: 01/09/2014].

^{xi} Information on the current version of EAD, EAD 2002, is accessible via URL <<http://www.loc.gov/ead/index.html>>. [Consulta: 01/09/2014]. Please note, that in Winter 2014/15 a new version of EAD, EAD3, will be published. Information on EAC-CPF is accessible via URL <<http://eac.staatsbibliothek-berlin.de/>>. [Consulta: 01/09/2014].

^{xii} See chapter above with regard to the Archives Portal Europe tools. The schema files for apeEAD and apeEAC-CPF are accessible via URL <<http://www.archivesportaleurope.net/Portal/profiles/apeEAD.xsd>> and <<http://www.archivesportaleurope.net/Portal/profiles/apeEAC-CPF.xsd>> respectively. [Consulta: 01/09/2014].

^{xiii} See online at <<http://www.loc.gov/standards/mets/>> for general information on METS and at <<http://www.archivesportaleurope.net/Portal/profiles/apeMETS.xsd>> for the current draft of the apeMETS profile. [Consulta: 01/09/2014].

^{xiv} See International Council on Archives "International Standard for Describing Institutions with Archival Holdings (ISDIAH)", 1st ed., 2008, at <<http://www.ica.org/download.php?id=1657>> and the schema of EAG at <http://www.archivesportaleurope.net/Portal/profiles/eag_2012.xsd>. [Consulta: 01/09/2014].

^{xv} Further details and more examples for specifications in apeEAD are published in the best practise guide APEnet, "Mapping towards and normalisation in APEnet EAD", 2011 <http://www.apenet.eu/images/docs/apenet_mapping_normalisation_guide.pdf>. [Consulta: 01/09/2014]

^{xvi} The International Standard Identifier for Libraries and related organizations (ISIL), comprises a set of standard identifiers used for the unique identification of libraries, archives, museums and related organizations. Cf. ISO 15511:2011 <http://www.iso.org/iso/catalogue_detail?csnumber=57332>. [Consulta: 01/09/2014].

^{xvii} The named attributes use the ISO standards, which are recommended by EAD in general, as default values.

^{xviii} See the whole list of possible values for this attribute in the EAD Tag Library, 2002 <<http://www.loc.gov/ead/tglib/elements/c.html>>. [Consulta: 01/09/2014].

^{xix} A list of mandatory elements in apeEAC-CPF is mentioned in the article BREDENBERG, Karin, "Records creators - Use of EAC-CPF in Archives Portal Europe", 2014 <<http://www.apex-project.eu/index.php/en/articles/157-records-creators-use-of-eac-cpf-in-archives-portal-europe>>. [Consulta: 01/09/2014].

^{xx} The Censo-Guía is available via URL <<http://censoarchivos.mcu.es/CensoGuia/portada.htm>>. [Consulta: 01/09/2014].

^{xxi} Read more about EAG on the APEX project website: <<http://www.apex-project.eu/index.php/en/outcomes/standards>>. [Consulta: 01/09/2014].

^{xxii} Read more about METS on the METS homepage: <<http://www.loc.gov/standards/mets/>>. [Consulta: 01/09/2014].

^{xxiii} Information on apeMETS as used in the Archives Portal Europe can be found on the APEX project website: <<http://www.apex-project.eu/index.php/en/outcomes/standards>>. [Consulta: 01/09/2014].