

APE_x



Deliverable

Project Acronym: APE_x
Grant Agreement number: 297355
Project Title: Archives Portal Europe network of excellence

D6.1 First Analysis report: Applying Web 2.0 solutions in archival applications

Revision: [Final]

Authors:

Kuldar Aas (NAE)
Go Sugimoto (NANETH)
Silke Jagodzinski (BA)
Urmaz Tamm (NAE)
Daniel Jeller (ICARUS)
Zoltan Lux (NAH)

Project co-funded by the European Commission within the ICT Policy Support Programme



Dissemination Level

P	Public
---	--------

STATEMENT OF ORIGINALITY

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. It reflects only the authors' views and the European Community is not liable for any use that might be made of the information contained therein.

TABLE OF CONTENT

1.	INTRODUCTION	4
PART I: SURVEY ON WEB 2.0 AND ARCHIVAL USERS		5
2.	OVERVIEW OF THE SURVEY	5
3.	UPTAKE OF USER-RELATED TOOLS	5
4.	ARCHIVAL USERS	8
5.	MOST BENEFICIAL TOOLS OR TECHNOLOGIES	10
6.	SUMMARY AND OUTLOOK	11
PART II: APPLYING USER ORIENTED TOOLS IN ARCHIVAL APPLICATIONS.....		12
7.	PERSONAL RESEARCH SPACE.....	13
7.1.	<i>Short outline</i>	13
7.2.	<i>Good examples</i>	13
7.3.	<i>Proposed components for Archives Portal Europe</i>	16
7.3.1.	Saving and managing information	18
7.3.2.	Extend and enrich search results	19
7.3.3.	Sharing and collaborating	20
7.4.	<i>Architecture</i>	21
7.5.	<i>Use cases</i>	23
8.	FACETED SEARCH	25
8.1.	<i>Short outline</i>	25
8.2.	<i>Good examples</i>	25
8.3.	<i>Main problems and research questions</i>	28
8.4.	<i>Proposed components for Archives Portal Europe</i>	28
8.4.1.	Types	30
8.4.2.	Dates	31
8.4.3.	Content providers	31
8.4.4.	Places, person names and subjects.....	32
9.	FEEDBACK TOOLS.....	34
9.1.	<i>Short outline</i>	34
9.2.	<i>Good examples</i>	34
9.3.	<i>Main problems and research questions</i>	37
9.4.	<i>Possible components</i>	38
9.5.	<i>Possible implementation</i>	40
9.5.1.	User management and misuse prevention	40
9.5.2.	Scope of feedback and structure of feedback form	40
9.5.3.	Managing feedback	41
10.	TAGGING	43
10.1.	<i>Short outline</i>	43
10.2.	<i>Good examples</i>	44
10.3.	<i>Main problems and research questions</i>	45
10.4.	<i>Proposed components and architecture</i>	50
11.	LINKED OPEN DATA	52
11.1.	<i>Short outline</i>	52
11.2.	<i>Good examples</i>	53
11.3.	<i>Main problems and research questions</i>	54

12. SUMMARY AND OUTLOOK 58
ANNEX I: APEX WEB 2.0 QUESTIONNAIRE..... 59

LIST OF FIGURES

Figure 1: Number of national archives using different tools or technologies6

Figure 2: Relevance of archival user groups9

Figure 3: Draft of data architecture of the personal research space in Archives Portal Europe.....22

Figure 4: Feedback form of the National Archives of UK.....35

Figure 5: Feedback form of the Photo Documentary Database36

Figure 6: ICPSR feedback form.....37

LIST OF TABLES

Table 1: Most beneficial tools for researchers and archival institutions (ordered) 10

Table 2: List of personal research space functions 16

Table 3: Comparison of functions and data entities 22

Table 4: Possible facets for archival portals 29

Table 5: Possible component functions of feedback tools 38

1. Introduction

One of the main aims of the APE_x project is to develop Archives Portal Europe¹ into an environment which allows users to find, manage, reuse and share information relevant to him or her in the best possible way. In order to deal with this task the project has started research to answer to the following questions:

- Which are the most crucial user groups of Archives Portal Europe?
- What is the state of the art on user-related tools and technologies in archival portals and institutions?
- Which user-related tools and technologies should be implemented in an international portal like Archives Portal Europe?
- Which considerations should be taken into account when implementing these tools and technologies?

This deliverable gives a first overview on the outcomes of such research carried out between June and December 2012. It is necessary to highlight that the current report should be seen as a basis for further discussions around the topics and the contents are rather preliminary. The work on specifying and possibly solving the issues raised in this document will continue throughout 2013, resulting in an update to this deliverable in early 2014.

In addition, the deliverable does not intend to provide novel concepts and advance the current state of the art but just synthesise and summarise the current situation in this area.

Another important note is the (mis)use of the term “Web 2.0” in the title of this deliverable. The project takes a rather wide approach and defines Web 2.0 as “all kinds of user collaboration and contribution tools and technologies”. We believe that this helps us to concentrate more on the needs of our users and partner institutions instead of just sticking with more restricted and limited definitions.

The deliverable consists of two major parts:

- Part I (chapters 2 - 6) describes the results of a survey carried out among APE_x partners on the importance of users and uptake of novel tools and technologies;
- Part II (chapters 7 – 11) describes some of the tools and technologies which seem to be most reasonable for implementation in an international research portal and outlines some of the design and implementation considerations of these.

¹ <http://www.archivesportaleurope.net/>

Part I: Survey on Web 2.0 and archival users

2. Overview of the survey

The initial task inside the Web 2.0 research was to carry out a small survey on the use of user-oriented tools and technologies in partner archives. The survey was prepared in June 2012 and the answers were gathered between July and September.

The request to participate in the survey was sent out to all 28 European national archives participating in APE_x project and a total of 22 responses were received. Thus it can be argued that despite the fact that the survey was held over the vacation season the results are still sufficient to give a general overview on the main tendencies in European countries.

However the limited coverage of participants in the survey does at the same time provide a major drawback. It is obvious that national archives represent only a small fraction of archival institutions in Europe and therefore it cannot be claimed that the survey results would provide a holistic view on the use of and plans around user related tools in European archives.

Another aspect to keep in mind when reading this part of the deliverable is that European national archives are quite different in their roles. While some function as major content holders and access providers to archival records, others (especially in larger countries) are more in the role of a general directorate and thus act more as legislative and governing bodies. This separation also implies different attitudes and understandings towards archival users and their needs, which may have resulted in slightly different answers to the survey.

To summarise, this survey can only be used as a baseline and limited decision mandate for specifying further actions inside APE_x. The (largely intentional) limitations mentioned above were tolerated because of time considerations and the fact that a more comprehensive survey will be carried out in the end of 2013 after the project will have gathered experiences with implementing and using Web 2.0 related tools in Archives Portal Europe.

The survey itself consisted of three main parts which are described and analysed in more detail in the following sub-chapters:

- Uptake of novel tools in archival portals and information systems (sub-chapter 2.2);
- Relevance of different user groups (sub-chapter 2.3);
- Relevance of different tool/technology categories (sub-chapter 2.4).

Technically the survey was published as a Google Docs Form and the original questions and introductory texts are available in Annex I:.

3. Uptake of user-related tools

The first section of the survey touched upon the use of user related tools in APE_x member archives. The respondents were asked about the uptake of nine distinct access and user related tools or

technologies in their institutions². For each of the areas the respondents were asked to provide a short description and (if possible) links to the public interface of the tool.

The nine areas asked for were:

- Social networking tools (forums, blogs, wikis, etc);
- Feedback tools for reporting errors in archival descriptions;
- Personalised research tools (bookmarking, documenting research results, personal databases etc);
- Crowdsourcing (user-based transcribing and linking records, storytelling, etc);
- Tagging of archival descriptions and records;
- Tools for organising virtual exhibitions;
- Faceted search tools;
- Tools for providing multilingual access;
- Tools for publishing archival collections as linked data;

It is again visible from the list of tools and technologies that the definition used for Web 2.0 was rather widespread. Especially the last four areas are not that much related to user interaction in general (virtual exhibition, faceted search and multilingual access) or just represent potential for future more powerful user services (linked data).

A general overview of the uptake of these tools or technologies is shown on Figure 1.

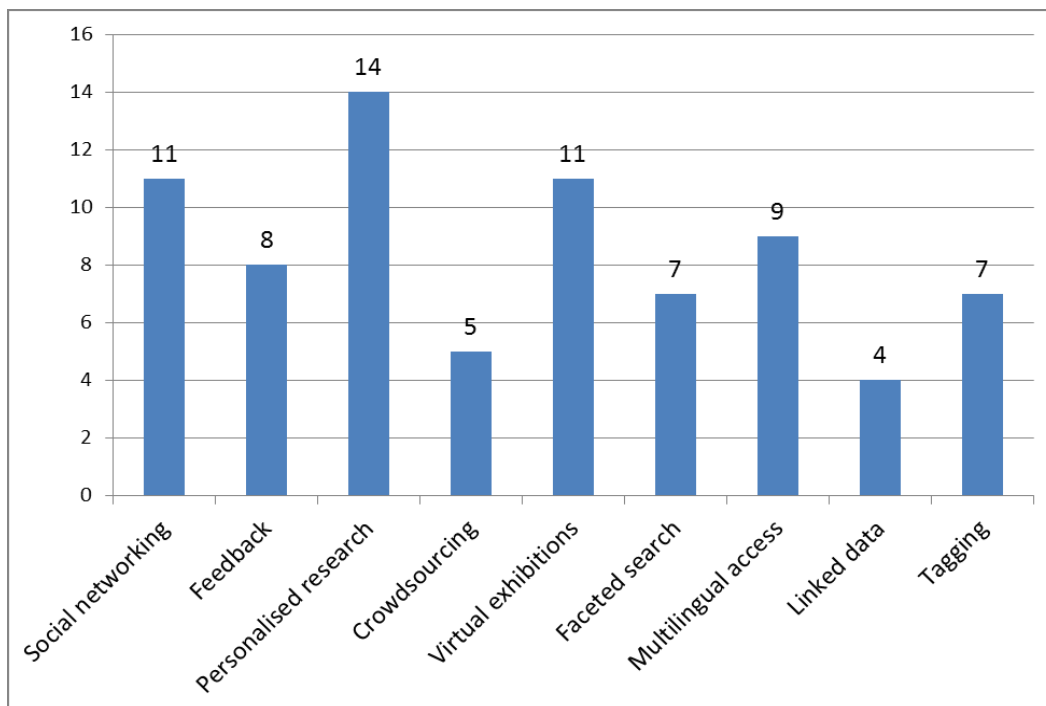


Figure 1: Number of national archives using different tools or technologies

² Note that in some cases respondents also outlined solutions applied in their respective country in other institutions.

When looking at the single answers we are also able to draw some additional conclusions on the nine areas.

Social networking tools (available in 11 archives out of 22 respondents): The main activity in the area of social networking is the presence on social media tools such as Facebook or Twitter and a majority of the positive answers mentioned these sites as the main tool for interaction with their users. In addition a few archives mentioned also dedicated forums being set up at their institutional portals but rather surprisingly this formed a minority of the respondents. A possible explanation for this could be that archival users are not willing to take up an additional forum as there are already quite a few such environments available. Thus it could be argued that it is indeed more reasonable for archives to build their virtual presence into already established sites and environments.

Feedback tools (8/22): While many archives are concerned about the quality of their archival descriptions the tools which might help to involve users for reporting errors and adding comments are rather less spread. Some archives also noted a reason for this – potential misuse and also high popularity of such tools could result in a possible overflow of feedback which the employees are not able to review in an appropriate timeframe. However, there is a visible growing trend in this area as quite many archives said to be working on the development of such tools.

Personalised research space (14/22): Being the most popular area of interest, it is visible that many national archives are quite concerned about providing their researchers with personalised virtual services. While in most cases the solutions consist of simple bookmarking and bookmark classification tools there are also archives which have put more effort into the area and offer a range of services including the possibility of virtual copy-requests, creating simple databases by users and ordering (paper) records into reading rooms. There is also visible progress going on as multiple respondents' highlighted future plans and on-going developments in this area.

Crowdsourcing tools (5/22): Crowdsourcing is still a rather novel approach applied only in a handful of countries. The outlined examples deal mostly with the indexing of genealogical records but some good solutions are available in the area of transcribing records.

Virtual exhibitions (11/22): While not necessarily a user collaboration and contribution related area, virtual exhibitions are quite popular. In most cases archives have built possibilities for virtual exhibitions directly into their institutional web pages or access portals but there are also some examples of exploiting publicly available tools for this purpose (Facebook, Flickr).

Faceted search (7/22): According to the survey surprisingly few archives offer possibilities for faceted search. It was also highlighted by some respondents that as more advanced search functionalities depend on the technical solution of respective catalogue systems it is mainly the high cost of developing new systems which hinders the delivery of faceted search functionalities. Looking at the available solutions the main facets used are dates/periods, archival creators and reference codes. Depending on the availability of appropriate tags also persons, events, subjects and places were mentioned.

Multilingual access (9/22): The respondents mentioning multilingual access tools usually highlighted search interfaces or access portals being available in multiple languages. However, there are unfortunately no examples of multilingual access to archival descriptions.

Linked data (4/22): Linked data, as a rather novel concept, is the area least implemented in archival institutions thus the number of available examples could be described as surprisingly high even when adding that most of the examples only deal with a very limited part of archival holdings.

Taken the number of problem areas ranging from archival data quality to data reuse licensing the application of linked data principles in archival institutions is far from straightforward. However, the area itself seems interesting to quite a few archival institutions and thus there is hope for new emerging solutions in the near future.

Tagging (7/22): Tools for tagging archival descriptions and records are available in some archives but most examples only allow this for internal users. In addition there was only one example for “free tagging”, i.e. tagging solutions which would not be based on prescribed dictionaries.

In general it can be said that user oriented technologies have not yet been widely implemented in European national archives. Only three of the asked areas have been applied in half or more of the institutions (social networking, personalised research, virtual exhibitions). In addition there are only a handful of archives which have implemented multiple of the more advanced technologies. In most institutions the interaction with users relies only on simple external solutions if applied at all.

Still, it is possible to see that there is a general wish to continue working on these areas and quite many archives indicate for developing new information systems which would include some of the Web2.0 tools.

4. Archival users

While developing or researching user related tools and technologies it is reasonable to learn also more about the actual users of archives. To achieve that the second part of the questionnaire included also three questions about the relevance of 14 prescribed user groups for the archival institutions:

Please rate on a scale of 1–5 how important do you regard the following user groups for your institution (1 – least important; 5 – most important)

Please rate on a scale of 1–5 how important do you regard the following user groups for the Archives Portal Europe (1 – least important; 5 – most important)

Please rate on a scale of 1–5 which user groups would in your opinion be most interested in using social collaboration and contribution tools (1 – least interested; 5 – most interested)

Therefore the intention was to look into the general relevance of certain user groups but in addition to compare the differences between the users of institutional and international portals as well as to the relevance of Web 2.0 tools for those groups.

The groups used in the survey as well as the results as a comparative graph are visible on Figure 2.

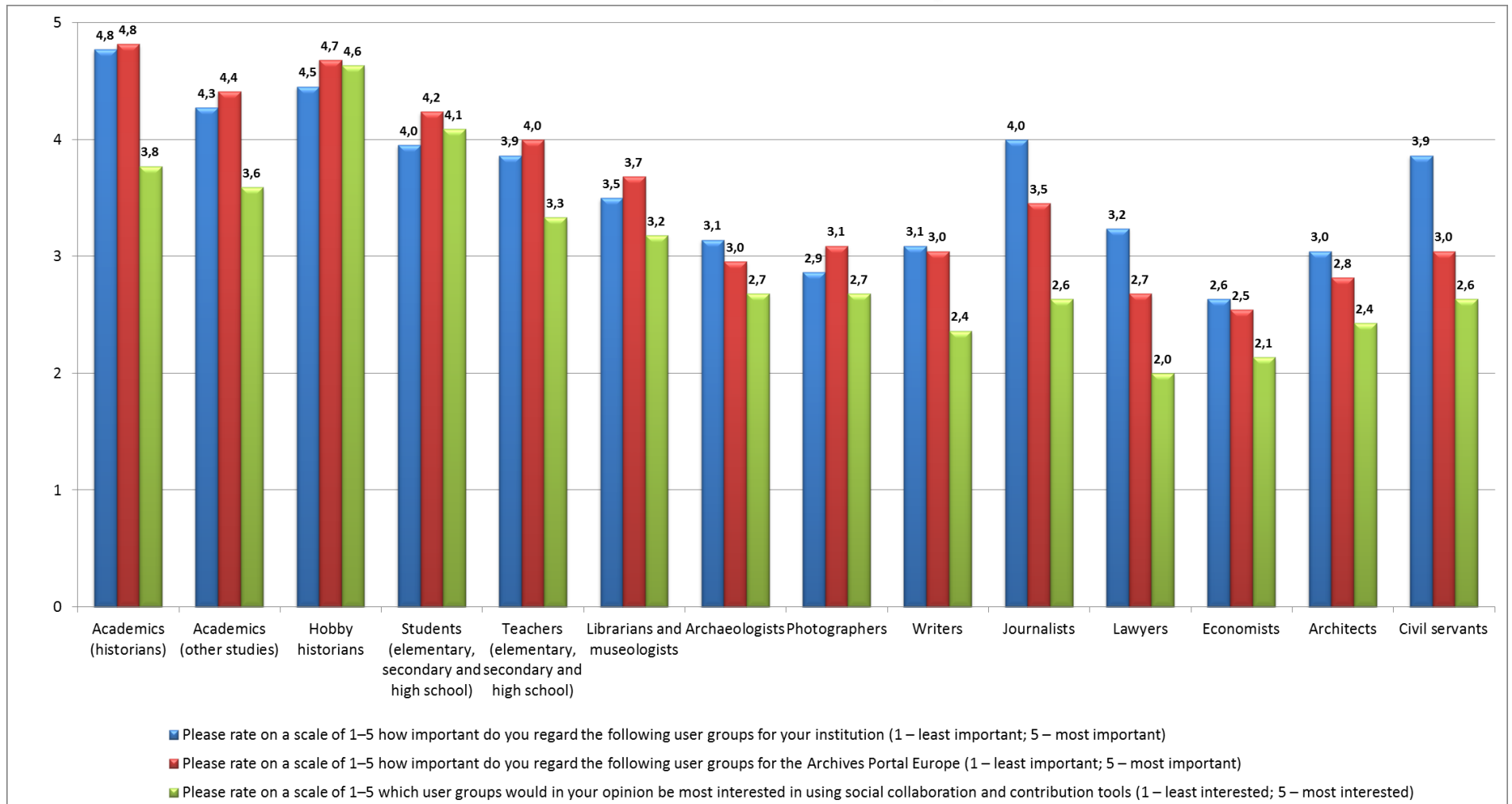


Figure 2: Relevance of archival user groups

When looking at the results we would like to draw your attention to the following items:

- Please note that the results reflect the opinion of national archives’ employees. In addition, the APE_x project plans to carry out an end-user survey which will be used to extend and verify the preliminary results and conclusions highlighted below;
- Main user groups for both institutions and Archives Portal Europe are academics (both archival and other studies), hobby historians, students and teachers;
- In all the mentioned top 5 groups the Archives Portal Europe scores slightly better. In other words – respondents from national archives have the opinion that these user groups would be slightly more important for the portal than for the institutions themselves;
- Two user groups having significant importance for the institutions – journalists and civil servants – have a significantly lower score on the importance for the Archives Portal Europe. For civil servants this is also quite obvious as they would mainly use archival resources in the framework of respective national legislation and therefore would prefer using local tools and portals instead of an international one;
- The user groups possibly most interested in using user-oriented solutions are students and hobby historians, followed by academics and teachers.

Despite the limitations of this survey the results are sufficient to provide reasonable input to take the next steps in providing user oriented tools and technologies inside Archives Portal Europe and verify the assumption that the tools developed should primarily serve hobby historians (genealogists, local history enthusiasts etc.), tightly followed by educational and research purposes.

5. Most beneficial tools or technologies

Please rate on a scale of 1–5 which of the listed tools would your researchers find most beneficial for their work? (1 – least beneficial; 5 – highly beneficial)

Please rate on a scale of 1–5 which of the listed types of tools / technologies would your institution find most beneficial for its operations? (1 – least beneficial; 5 – highly beneficial)

The last part of the survey concentrated on the benefit of certain tools or technologies for either the users or the archival institutions themselves. The main purpose of this part was to prioritise the areas mentioned used also in the first part of the questionnaire (see Chapter 3) and propose a more limited number of tools for closer inspection and research.

The average results upon the different categories are highlighted in Table 1.

Table 1: Most beneficial tools for researchers and archival institutions (ordered)

For researchers		For institutions	
<i>Tool / technology</i>	<i>Mean value</i>	<i>Tool / technology</i>	<i>Mean value</i>
Personal research space	4,14	Feedback tools	4,19

Feedback tools	4,09	Linked data	4,05
Tagging	4,00	Tagging	3,86
Linked data	3,90	Virtual exhibitions	3,76
Faceted search	3,86	Faceted search	3,67
Virtual exhibitions	3,59	Personal research space	3,50
Social networking	3,50	Social networking	3,33
Multilingual access	3,45	Crowdsourcing	3,33
Crowdsourcing	3,14	Multilingual access	3,14

Based on the ranking the project formed a top five of user oriented tools to be examined closer (see also Part II of this report):

- Personal research space³
- Feedback tools
- Tagging
- Linked data
- Faceted search

However, together with the end-user part of the survey, the relevance of tools is also going to be further discussed. Furthermore, additional end-user evaluations are going to be executed once first examples of user oriented tools are available on the Archives Portal Europe.

6. Summary and outlook

The main outcome of the survey was that it helped the project to decide on the next steps of the Web 2.0 research inside of the APE_x project and (as visible in the next chapter) the work continues mainly on the five areas of tools highlighted in chapter 5, and is mostly oriented on the user groups of academics, hobby historians, students and teachers. In addition it provides a reasonable basis for future surveys including possibly a larger amount of archival institutions as well as end-users.

³ It is reasonable that this area scores low in the category of institutional benefit. However, being on the top spot for researcher benefit it certainly has to be included in further research of the APE_x project.

Part II: Applying user oriented tools in archival applications

This part of the report describes in more detail the research done by the project in the five focus areas highlighted in chapter 5.

Taken the best practice found through the survey of APE_x member archives there are not too many areas where good solutions would be available. Therefore the research went also beyond the analyses of tools mentioned in the survey and included a more in-depth search for tools and solutions available in other regions and professions.

The five areas in question are also at different levels regarding both the maturity of such tools and their technical complexity. The areas of personal research space, faceted search and feedback tools can be said to be rather well developed and covered with available solutions in archival institutions. Therefore the according chapters below concentrate more on delivering more specific discussions around the implementation of these. It is also recommended to the APE_x project to implement tools in these three areas inside Archives Portal Europe during 2013.

At the same time the areas of tagging and linked open data can be said to be more complex and thus the majority of the according chapters include initial discussions and a mapping of research topics and questions. Therefore, instead of offering specific solutions and architectures the research in these two areas will continue throughout 2013 and will be documented in more detail in the next version of this deliverable (early 2014).

While analysing the five focus areas special attention has also been drawn towards gathering and describing available best practices. Each of the chapters below includes a special section for good examples in the respective area.

7. Personal research space

7.1. Short outline

Personal research space is obviously an area which is not possible to be developed as a singular tool but should rather be seen as a work environment which combines a bundle of more specific tools for the user to perform certain tasks. Possible features in a personal research space can be distinguished between working with archival descriptions and content and use the workspace as a virtual desk for dealing with information around the research, e.g. bookmark pages of interest, create small databases and write articles.

Also the social collaboration aspect should be taken into account, as in some cases users might have the need to collaborate in collecting information with these tools or simply want to share it among a closed group or publicly. Especially when looking at the key user groups of hobby historians and students it might be assumed that these are rather interested in collaborating in their tasks.

Having a personal space also means that it can only be available for authorised users. As there is a trend towards archival institutions being present in other social collaboration environments (Facebook, Twitter, Foursquare etc.) it seems reasonable that also the users would be able to use their already available accounts to authorise themselves in the Archives Portal Europe. In addition it might also be reasonable that the social collaboration actions mentioned above take place not only inside Archives Portal Europe but also in other environments. To put it simple – users of the Archives Portal Europe should be able to log in with their social network accounts as well as share their research results (link books, personal databases, etc.), for example, within Facebook. In the longer perspective it might also be valuable to make the singular personal research space tools available as “Apps” which might easily be integrated into other environments than Archives Portal Europe.

7.2. Good examples

- **National Archives of Estonia, Virtual Research Lab (VAU):** <http://www.ra.ee/vau>

The National Archives of Estonia has created a “single sign-on” research environment which brings together access to all resources as well as offers a few central tools like a link book and personal databases.

Link book – When users browse through different databases and/or digitised resources they have the possibility to easily bookmark pages of interest and add small descriptions to the bookmarks. There is also an option to group bookmarks into collections. For bookmarking pages in various Nat. Archives’ portals one must pay attention to a small group of links usually situated at the top right corner where clicking on the links bookmarks the current page or opens the link book for managing.

Personal databases – As an extension of the link book users have the possibility to define a simple table structure and use it to document their research. Small and quite basic databases can be generated with types of text, integer, URL and date for field values. The functionality is

mainly used by genealogists who use the personal databases to write down the pieces of information they find in various archival records (as an example by using fields like: Name of person; Year of birth; ID of the record where the information was found; link to the digitised image where the information has been transcribed from).

- **The National Archives of America:** <http://www.archives.gov/social-media/>

The National Archives and Records Administration is probably the leading authority in applying social media and personal research tools. One of the most interesting tools in this area is the “Storify” (<http://storify.com/>). It is a social media curation tool that allows anyone to create stories by drawing together content from other social media sites and archival records. Users can re-order the elements of their stories, add text to help give context to the readers, and build a larger story from the individual elements. It might be reasonable also for Archives Portal Europe to associate their accounts with the “Storify” account to link or subscribe to stories in the “Storify” portal and publish them in their personal view.

- **Stack Overflow** <http://stackoverflow.com/>

Stack overflow can be seen as a specialised forum. Users can post questions about various topics and others can provide solutions to them. Answers can be rated as “useful” or “misleading”. Such a solution could also be used next to archival research environments for users to help each other out as an example to understand the content of materials.

- **gahetNA - Friends of the national archives (NL):** <http://www.gahetna.nl>

In gahetNA recent searches are saved automatically in a list for logged in users. Single search results can be marked as favourites and can be saved this way. The archival material itself resp. an image cannot be saved in the personal area. Archival material can also be ordered into the archives’ reading room.

Registered users can contact archivists as well as other users to communicate about topics and archival material. They can participate in indexing handwritten text, asking and answering questions and give feedback about the descriptions.

- **Archives départementales du Puy-de-Dôme (FR):**
<http://www.archivesdepartementales.puydedome.fr/>

The archives in Puy-de-Dôme offer saving and printing of search results and saving of (digitized) archival material in an album. Single search results can be marked and downloaded in an EAD file.

After signing an extra agreement registered users are allowed to transcribe handwritten texts and to publish these transcriptions. Another interesting service is the saving of the permalink of the image and the possibility to give feedback via email for each image.

- **MICHAEL Culture Project (EU):** <http://www.michael-culture.org/>

On their (old) portal the MICHAEL Culture Project offers a specific area for registered users. Users can save their research but no single search results, archival material or links. In addition users can upload external documents and share them with other registered users, repository staff or the public (non-registered users). Users can create descriptions for saved items and can export information as XML files. There also tools to import information from XML files, e.g. bibliographic information.

- **Prometheus Bildarchiv (DE):** <http://prometheus-bildarchiv.de/>

Prometheus is an image archive with contents from different archives and museums. Registered users can save single images (archival records) into collections and presentations. It is possible to share these collections and presentations with the public and other users. The user can also determine whether other users and the general public can only view or also edit the collections. Within a collection comments can be added to single images and these comments can be published on the portal. Information can be printed and downloaded.

- **Monasterium (EU):** <http://www.monasterium.net/>

Monasterium is a European portal for charters which has the goal to motivate users participating actively in the in-depth editing and indexing of archival records. The personal space at Monasterium does fulfil these demands by offering an editing tool. Registered users can save search results (bookmark) and archival records (usually digitised images). Transcriptions and descriptions of the charters will be saved and reviewed by authorised staff before publishing. Information can be printed (PDF) and downloaded (XML, PDF).

- **CENDARI (EU):** <http://www.cendari.eu/>

The enquiry environment at the heart of CENDARI will incorporate tools which allow researchers to engage with geographically dispersed archives via multilingual searches, custom visualisation, shared research and collaboration spaces, and personalized virtual environments. As the work is still in progress the use of elements or technologies developed by CENDARI must be decided later on.

- **Zotero:** <http://www.zotero.org>

Zotero is a browser plug-in tool which allows the researcher to save and organise research documentation as a library and to create groups with common libraries. While being an external plug-in it still seems reasonable to give Archives Portal Europe users the possibility to use Zotero libraries and groups seamlessly inside the portal. In fact, The European Library (<http://www.theeuropeanlibrary.org>) uses this service on their website.

7.3. Proposed components for Archives Portal Europe

The examples of the different personal research spaces include a range of different functions. Most common tools are based around archival contents like saving and managing links, ordering records to reading rooms and printing. At the same time there is a growing number of tools for the users to share information and to collaborate with each other in archival research.

A few good examples are also available for integrating different research environments to allow for seamless access to and mash-up of resources gathered within different locations. Also functions bordering on crowdsourcing like the creation of personal databases and transcriptions are becoming more available.

An overview of possible tools and functionalities in a personal research space is available in Table 2. Each of the functions is discussed in more detail below.

Table 2: List of personal research space functions

Function	Description	Comments
<i>Login, saving and managing information</i>		
Registration and login	Users should be able to register themselves and log in to be able to access personalised services	This is a necessary requirement for offering personalised services This function is already available at the Archives Portal Europe.
	Associate social media accounts with Archives Portal Europe for login	
Create and manage collections	The authorised user should be able to create and describe named collections	The collections can be the highest level which serves as thematic bracket for the other features.
Save searches	Save search terms and applied limitations for future reuse	This function is already available at the Archives Portal Europe.
Save search results	Save all or single selected search results into a collection or link book	
Create and manage link books	To collect and manage saved links, the user should be able to create link books	
Save links	Bookmark links to additional resources found inside the portal	In Archives Portal Europe digital objects are not stored but only linked. Therefore the functionality would allow to bookmark also the links to the original digital objects

Function	Description	Comments
	Bookmark links to additional resources found outside the portal	External resources might be archival descriptions and relevant documents (e.g. dictionaries)
Create personal databases	Define simple database structures to document research	Note that the personal databases need to be integrated to link books / bookmarks, i.e. it should be possible to easily use saved links / bookmarks for creating a database entry
Export information	Export the information gathered in the portal (personal files, bookmark collections etc.) into common formats like PDF and CSV.	This function also supports the printing of material when downloaded in PDF or other visual formats. Also integration of other tools like Zotero and Storify should be considered
Extend and enrich search results		
Upload and import additional information	Upload files found in other sources into the user's personal space	In general a user should also be able to upload files relevant to his/her research and organise these
	Import additional information from other sources	The personal research space should support simple upload of other information like link books created in other applications. Also integration of other tools like Zotero and Storify should be considered
Create descriptions for saved items	Add personal descriptions and comments to the saved items (e.g. archival descriptions, bookmarks, files, saved search)	
Sharing and collaborating		
Create and manage groups	The users should be able to create groups and manage those (invite other users, assign rights etc.)	This function could later be extended to support groups in other social media applications (e.g. invite friends from Facebook)
Share and publish information	Share information (e.g. search results, bookmarks, files, descriptions) with other registered users or defined groups	This function should be available in both Archives Portal Europe and on other social media environments. Next to this functionality we also need a "browse shared / public information from other users" possibility which would allow seeing of and searching for contributions from other users.
Add questions and comments	Posting general questions and comments on internal or external forums	Another possibility should be to post questions about certain resources into a general forum. As an example – users might have a button "Post to forum" next to archival descriptions on Archives Portal Europe which allows posting a link and a relevant question or commenting to the appropriate forum topic (see Stack Overflow for an example). This function could later be also integrated

Function	Description	Comments
		with other social media environments (Facebook, Foursquare etc)
Share stories created with the "Storify" tool.	User can embed "Storify" content into his/her articles, publications etc.	First goal should be implementing the basic functionality - just to share the "Storify" content. In the future if proves itself not too complicated also creating a "Storify" environment inside the personal workspace as a plugin, similar to Wordpress' "Storify" plugin.

7.3.1. Saving and managing information

Create and manage collections and link books: Users should have the possibility to group their links and saved items in an own structure. For that reason they should create and name collections as an umbrella to manage the information for own purposes. Another option would be the usage of tagging. Users could add keywords to their links and items and group them by search results.

When saving a link or item a form can ask for additional data like the title, description, assignment to a specific collection or link book and/or tags and keywords. The user might also be prompted to choose whether the link is going to refer to the resource within the portal or to the material in the original catalogue.

Basic search functionality should also exist with an option to search in all collections/tags or within a specific collection/tag. Sorting search results and changing their order in the collection should also be enabled. A more detailed overview of this component in action is given in UC3.

The implementation should be a type of a "plugin" component that can be placed in all views of the portal. That is to bookmark links within the Archives Portal Europe. For saving links outside the portal a simple form can be used.

Save searches: Once an user started a search in the Archives Portal Europe, (s)he must be able to save the search terms as well as any applied limitations after logging in. The user can take the same settings to repeat the search later on. It may be possible that the results differ because new content was uploaded by the archives or old content as be updated meanwhile.

Save search results: Saving search results is one of the main functionalities that an archival portal needs. Users must be able to save single archival descriptions as output of a research with all necessary information. This functionality goes further than simply save a link inside the portal but allows users to work with structured metadata.

The search results must be selectable to choose single results from the list (selecting the whole list could be possible as well). Users can add own content to each archival description. All saved description can be exported to an own created finding aid which can be saved as XML and PDF file.

Save links: Saving links includes two functions which can be helpful for users:

- Saving links **within the portal** (no search results)

Links within the portal which are not displayed as search results can be saved. These can be links to archival descriptions (found by browsing) at any level, to service information on the portal or to editorial content.

- Saving links **outside the portal**

Links pointing to pages outside the portal can be saved. These can be links to the digital object on the content providers' websites and any other webpage as well. Here we should differ between materials related to APE_x portal and those that are not. Categories "internal", "external" and "unrelated" could be used.

Create personal databases: Users might need a way to structure their work into a database to better manage and search amongst data, also share it with others. For that a simple interface should exist where users can define the structure of their databases and afterwards import data from uploaded files. A restriction for file types should be implemented, for example limiting it to CSV files only. A more detailed overview of this component is given in UC6.

Export information: Users should be able to also export the information from the portal. The possibilities for export should include at least:

- PDF: most users would probably like to export information for printing or sending to other people. The best way for this is to define certain export designs for the lists of search results, links, and other information we have in the portal. Using these designs users could generate PDF files and print / download / send if necessary;
- Structured data: a more ambitious idea is the download of structured data. In case of databases the portal can offer a simple table export, e.g. CSV, XLS, ODS, XML. In case of search results resp. archival descriptions the metadata download in xml format should be aimed. That means the user would be able to download the archival description of a single item as well as any other level, including sublevels, of the EAD hierarchy. On the other hand the user should be able to download his/her collection of archival descriptions, e.g. saved search results, all together in an own EAD file, what means the user creates his/her own finding aid. Sometimes EAD files could be imported in reference management software. The same should be possible for defined content from a EAG file and EAC-CPF file. In both cases there are structured data as well. An export can be possible in xml, of course, but in CVS or XLS/ODS as well.

7.3.2. Extend and enrich search results

Upload and import additional information: External information can be classified in 2 groups:

- Upload **files:** For using the personal research space as online desktop, files in several formats can be uploaded. This could be a text, image, audio/video or any other file, e.g. digital objects, which the user wants to have near his/her information in the portal.
- Import **structured information:** In a second approach users are allowed to upload structured information. The portal can offer to use these data:
 - a. in a database

Structured information in any structured format can be uploaded, e.g. CSV, XLS, ODS, in a user created database (see above).

b. similar to research results

Structured information in any xml standard can be uploaded to add metadata to a topic, an archival description, a file or to create a bibliography.

Create descriptions for saved items: User generated text can be added to the saved items (e.g. archival descriptions, bookmarks, files, saved search). The user generated information can be added in with a simple standardized form asking for a tag, a description and a link. An important feature is to give the possibility to export and share/publish all of the added information together with the item.

7.3.3. Sharing and collaborating

Create and manage groups: User should be able to create a group within the portal by sending messages to other users to join. Users can accept or reject requests sent to them. If an user accepts a request (s)he has access to data shared among the group. A user can administrate his/her groups by adding new members, ejecting them etc. If a user deletes one of the groups data shared within, it shouldn't be deleted as well. If needed a way of sharing information with groups created outside the portal, like Facebook, should be possible. That requires the functionality needed for the user to share his/her work in user's Facebook groups.

Share and publish information: The difference between sharing and publishing information lies in the defined group. Published information goes to all people using the portal - to the public. Shared information goes to a defined group of portal users. The users can define the group (or only a single person) themselves. Users can share their information to all registered users of the portal by inviting them (email address = user name). Additionally the portals can offer an automatically created group, like "all registered users".

- Sharing via social media

The easiest way for some users might be the idea of sharing information via their own social media accounts. Like and share buttons of different applications might be possible to implement.

- Sharing inside the portal

Another possibility is to create a simple browsing possibility for all information which (s)he has access to. An inspiration of such a function could be the Google Drive solution which provides a common overview on both the information created by the user as well as shared by others. As for the sharing users would just need to set appropriate access rights to relevant groups and/or single users.

Another important aspect is that the portal has to decide what kind of data the user should be able to share or publish: search results (single or list), export formats (PDF, CSV, XML files), link books / collections.

Add questions and comments: Sometimes users might run into problems when working with foreign materials and not being able to understand the content. Therefore an option for posting questions and answers is something that could prove worthy. This can be solved using a simple forum but it would be more convenient if the question could be added via a simple form on the spot instead of starting a new thread inside the forum. Added questions can be organized in a separate view for a better overview. Questions can be marked as “solved” or “opened for discussion”. Users can up rate or down rate answers added whether they prove themselves useful or misleading. A more detailed overview of this component in action is given in UC4.

Share stories created with the “Storify” tool: With the “Storify” tool one can easily create articles using materials from various sources. The simplest functionality to implement in the first phase would be the sharing of pre-created stories by embedding the “Storify” content into APE_x documents: simply pasting the URL into the document to make it appear in the text. The next phase would be the creation of new stories inside the APE_x environment, the edition of the existing ones etc. For that the APE_x user must associate his/her account with his/her “Storify” account. It would be good, if a plugin exists for tasks as such like the Wordpress “Storify” plugin. Otherwise if it turns out to be too complex to create from scratch only the most simple functionality should be implemented.

7.4. Architecture

On a more technical perspective the information in question can be classified into four broad areas (arranged according priority):

1. Archival area: search results / archival descriptions, (own created) finding aids
2. Link book: internal and external links
3. File area: uploaded files
4. Database: structured data

Whereas link books and databases are suggested to be created as distinct applications inside the personal research space, the archival area and the file area are topics for several functions but can become applications as well. In this case a user generated “collection” could be the umbrella for the areas.

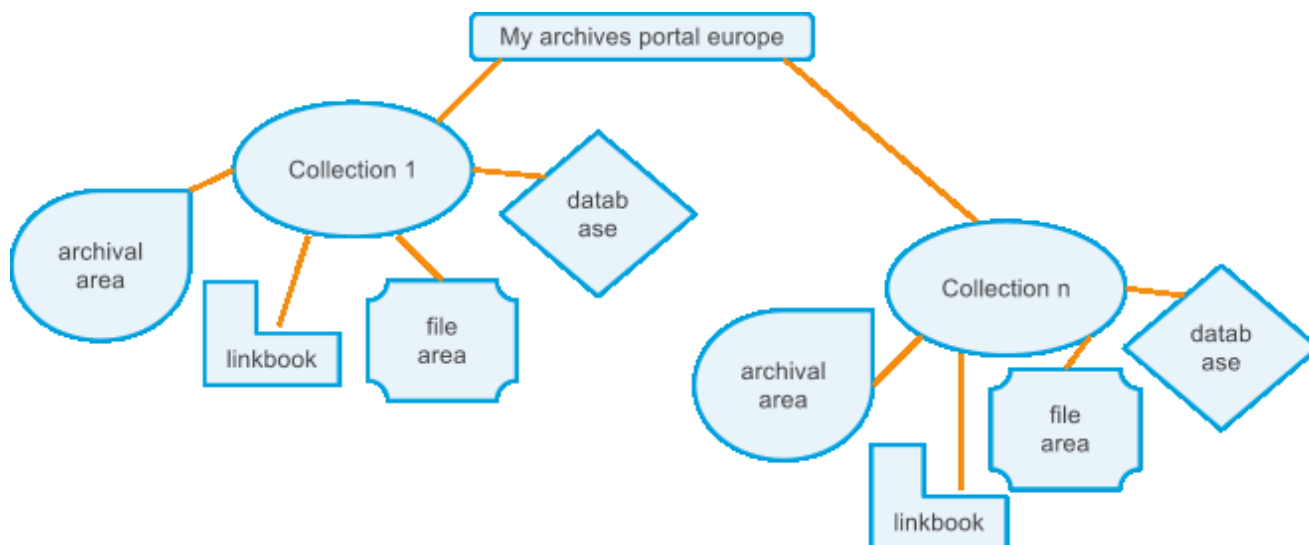


Figure 3: Draft of data architecture of the personal research space in Archives Portal Europe

It is also visible that not all of the functions need to be applied on all the data entities. In Table 3 the project team proposes a matrix which compares the relevance of the core functions to the above data model. The table entries are additionally colour coded according to their implementation priority:

- Red: priority 1
- Blue: priority 2
- Green: priority 3
- Grey: priority 4

Table 3: Comparison of functions and data entities

	Export	Share	Publish	Add own content
Archival area				
Search	no	no	No	no
single search result	yes	no	No	yes
list of search results	yes	yes	Yes	yes
(own created) finding aids	yes	yes	Yes	yes
Link book				
Link collections	yes	yes	Yes	yes
Single links	yes	no	No	yes
File area				
uploaded files	yes	yes	No	yes
Database	yes	yes	Yes	yes

7.5. Use cases

To help visualise the implementation of a personal research space the project has also provided six sample use cases.

UC 1: Simple MyPages

The user logs in and starts a search with 3 terms. Beforehand (s)he choose 4 archives in 3 European countries. (S)he gets 439 search results. He/she saves his search terms and selects 45 of the results for saving. The user adds comments to the 5 most useful results and downloads the results plus description for his/her personal source guide.

UC 2: Advanced MyPages

The user logs in via his/hers social network account. (S)he starts a search with 3 terms. Beforehand (s)he choose 4 archives in 3 European countries. Our user gets 439 search results. He/she saves his search terms and selects 45 of the results for saving. The user adds comments to the 5 most useful results and bookmarks 7 external web pages which include related content. Then (s)he finds (analogue) books in a library database and imports the MARC-xml/RFD data to MyPages. Additionally (s)he uploads a draft of his/her new article and an article from an e-journal. He/she creates a thematic guide with sources, books, article and links. (S)he shares the dossier with some fellows in his/her social network and a group in Archives Portal Europe. For his/her own use (s)he downloads his collection.

In one of the 45 results (s)he finds a description that (s)he can complete because (s)he has seen this item once in the reading room. (S)he adds a comment to this description and publishes this for all users.

UC 3: Adding links to link book

The user logs in and starts a search. The user opens a search result and presses the button “Bookmark link”. A small window opens where the user provides some additional information: the link’s title, description, category and keywords. It is possible for the user to manage his/her links in the same window, create a new collection and tag on the fly. The user has an option to choose whether this link is going to refer to a resource within the portal or the archives that owns this material. “Internal” is the default value but the user changes this option and the link changes automatically. The user presses the button “Save link” and the window closes. The user creates also some links referring to materials that aren’t related to the Archives Portal Europe at all (some wiki pages). For that (s)he uses a simple form in the link book view with the same fields but with the “link” field being editable, i.e. not read-only.

UC4: Adding comments to materials

The user logs in and opens the “Questions and Answers” view. In this view the user can look at his/her Q&A history, start new threads etc. The user is also able to sort all threads by topic, answer count, time etc. The user chooses a thread and clicks on it which directs him/her to the material with the discussion displayed at the bottom. The user posts his/her own answer and rates the answers posted by others. As (s)he is the owner of the on-going discussion he or she chooses the

option “Close discussion”. He then uses the button “Back to QA” which takes the user back to previous page.

UC5: Save search results

The user starts an advanced search and refines his/her list of search results. In the final list of search results the user marks several entries with a checkbox. Marked entries are saved in the user’s archival area. When the user goes to the archival area (s)he sees all saved entries in a list in a simple or an extended view. The user can add own content to each archival description. All saved descriptions can be exported to an own created finding aid which can be saved as xml and PDF file.

UC6: Create personal database

The user logs in and opens the “My Databases” view. There the user has various options for viewing and managing his or her already existing databases. The user clicks the “Create new Db” button and gets directed to a new view where the fields need to be filled with database name, brief description, detailed description, and whether the database is public. The user submits general data about the database and gets directed to the next view where database structure needs to be defined, i.e. names and field types. Integer, text, date and URL are the values between one can choose. The user clicks “Add new field” to add new database field, a row with name and type parameters. After finishing describing the fields, the user submits the form and gets directed to the next view where one can upload files containing structured data. The user selects appropriate files from the local computer and uploads them to the database. First the file type gets checked and an error message gets displayed if the file type doesn’t match with the ones allowed. Fields inside the uploaded file then get checked line-by-line to see whether they match the field types defined before. If a value cannot be casted to a type the value should be excluded and a warning message gets added to the report displayed in the end.

8. Faceted search

8.1. Short outline

Faceted search is a technique for clustering information into categories according a classification system. It allows users to filter items, e.g. search results, by choosing given values. Precondition for faceted search is the classification of information, i.e. the provider of this service needs information to classify. It can be defined if faceted search uses only classification regarding content information or technical / formal information as well.

The web search world has offered two different ways to explore its whole content.

1. Direct search, in which the user is required to give an exact search expression (string) in a text box. The portal's search engine creates a query and runs it on the databases of the portal's content. Usually in case of specific databases (e.g. archival finding aids) this way of search can be made more complicated. The expert user may know more precisely in which type of data (s)he would search. For this purpose usually the systems are providing a so called "advanced search" option, where the user is allowed to put search words into different text boxes and it is usually also allowed to select logical operators to connect them. (Google)
2. The second way for exploring the content of a portal or a database is the navigation in a predetermined hierarchical structure. The most common examples are the archival information systems in which the users can browse (up and down) the tree structure of the archival material.

Over the last few years the direct search has gained dominance, almost all websites provide such an option. Recently a new approach has emerged especially in content specific databases. This approach combines both above mentioned type of exploring a "digital content". This type of search is called faceted search. Or as defined in Wikipedia: *Faceted search enables users to navigate a multi-dimensional information space by combining text search with a progressive narrowing of choices in each dimension.*⁴

8.2. Good examples

- **National Archives of Sweden (SE):** <http://www.nad.riksarkivet.se>

The Swedish database NAD offers the possibility to filter options for search results, but not before entering the search.

The search offers four main facets: Archive (with sub-options Type & Provenance), Register, Institution, Location. Within an option the facets are ordered hierarchical.

⁴ http://en.wikipedia.org/wiki/Faceted_search

Each facet displayed shows the number of hits. Some options and facets contain a tooltip with explanations.

- **The National Archives, The new Catalogue (UK):**
<http://discovery.nationalarchives.gov.uk/SearchUI/>

The National Archives has two additional options to their simple search: enter a date (range) and/or click a subject. Afterwards the users can filter their search results.

Among the search results six facets are available: collections, subjects, dates (ranges), catalogue level, closure status, and record opening date. The facets within the options are ordered by the number of hits which are shown at each facet.

- **NARA National Archives and Record Administration Online Public Access (USA):**
<http://www.archives.gov/research/search/>

The archives offer a Google-like simple search which can be switched to a very detailed but a little advanced search. The search results are grouped by main types of archival material, e.g. online holdings, unit of descriptions, special holdings like Presidential Libraries or NARA's authority records, etc. At that point it can be switched to the advanced search in order to refine the search. After any kind of search which results a list, a structured list of several types of categories appears. The users can refine the search results by those categories. The facets are e.g. the clusters, data sources, level of descriptions, type of archival materials, date, format, location, etc.

The faceted search option provides a very detailed way for refine a given list of search result.

An interesting point is "topic clusters" which means: "Document clustering is the automatic organization of search results into groups of related content that are known as clusters."⁵ The system automatically creates clusters from different information. E.g. it clusters keywords (nouns) from title and from description; provenance and document type. User can "remix" these clusters if they are not useful enough.

- **National Archives of Estonia MAPS (EE):** <http://www.ra.ee/kaardid/index.php/en>

The Maps database of the National Archives of Estonia offers three possibilities - a simple, an advanced and an extended search. When using the simple search the facets title, place and date created can be used.

⁵ Topic Clusters: <http://www.archives.gov/research/search/using-opa.html#clusteropa>

In the advanced search additional options are available; one remarkable possibility is to use geosearch, i.e. select a region on Google Maps and search for maps which overlap or are within the selected region. Also possibilities for searching based on archival reference codes and authors of maps are available.

In the extended search a range of expert options is added like the conservation status, physical form etc.

- **The Archives of Aube (FR):** http://www.archives-aube.fr/arkotheque/wiki_archives/ressource.php

The French archive offers a browsing possibility by facets; rather interestingly users cannot enter a search term.

The browsing facility allows filtering all possible results with 5 options: Location, Person, Subject, Time period and Techniques. Inside the facets location, time period and techniques sub-options are possible to be selected. As an example for location the possibilities are world (countries), France (departments) and Aube (cities, municipality).

In facets Person and Subject users can input a search term but besides this also have the possibility to browse the lists in alphabetical order.

Each option and facet displayed shows the number of hits. Chosen facets can be deleted by one-click (trash sign).

- **La recherche aux Archives de l'État (BE):** <http://search.arch.be>

The national archives of Belgium offers a simple search without any limitations. The search results are shown within four classes: persons, archives, creators and sources.

Within the classes archives, creators and persons the user can use the advanced search function, which differs in each class.

Class archive: The user can enter search terms, dates, and choose from a classification list. The search can be refined for fonds with digital objects and digital-born documents. The search results can be filtered according the archive, the language, the date range and a classification.

Class creators: The user can enter search terms, date ranges and names, and (s)he can choose from classification list and from a location list. Additionally users can browse within offered filter classes: type of creator (person, family, corporate body), language, classification, source, type of archive, date range. The same filters can be used to refine search results.

Class person: the search for a person can be very deeply refined. The simple search offers any search term as well as date and location (taken from a list). At the same time the user can search for first and last name, gender, profession and person's role (e.g. mother, son, or grandmother). The search results cannot be refined afterwards.

- **Archiveshub (UK):** <http://www.archiveshub.ac.uk>

Archiveshub from the UK offers some possibilities to refine a search beforehand. User can choose a specific archive and (s)he can choose one of the option for defining the search term: keywords, titles, dates, creators, reference numbers, subjects, names of people, organisations and places, and media types. The search results can be filtered by subjects and repositories.

- **EUROPEANA (EU):** <http://www.europeana.eu>

Europeana offers a simple search without any refinements. Search results can be filtered by media type, language, date, country, copyright and provider.

- **Economic and Social Data Service (UK):** <http://www.esds.ac.uk/Lucene/Search.aspx>

The search of the Economic and Social Data Service offers the possibility to enter a general term or to define the term by: title, subject, keyword, geography, data creator, methodology, publications, and UKDA study number. Search results can be filtered by document type of the data, subject and date.

8.3. Main problems and research questions

The number of possible facets to be implemented is quite huge and thus a crucial question is which facets and how to implement, even some of them are already available in the Archives Portal Europe. As the answer to this question can only be based on the actual needs of the users, it is not possible to give a specific set of recommendations in this document but only list reasonable facets in archival applications (see Chapter 8.4). Before the actual implementation of the facets starts, probably an end-user survey needs to be done to select the most crucial and beneficial facets.

Another problem is that the facets and their usage in resource discovery are often controlled by pre-defined vocabularies and ontologies. Due to the multilingual nature of Archives Portal Europe it is not easy to first agree on a certain vocabulary and even if it would become possible the vocabulary needs to be mapped into all available languages of archival descriptions. This is also necessary for an appropriate user experience – as the user interface of Archives Portal Europe supports a multitude of European languages the users would also expect that the search facets delivered as lists or selections are provided in the same languages.

However, as the problems of multilingualism and ontologies are also present in the areas of tagging and linked open data the APE_x project will continue to seek at least partial solutions to achieve a reasonable level of support for some vocabularies and ontologies.

8.4. Proposed components for Archives Portal Europe

With regard to faceted search the main components are the actual facets and their implementation details. The facets extracted from the best-practice research are summarised in Table 4 and discussed in more detail below.

Table 4: Possible facets for archival portals

Facet	Description	Comments
Types		
Type of the archival material	Archival material can be classified by technical, formal or content related aspects	This facet should be used for refining search results. Especially the formal criteria (see below) should be used
File format	File format of the archival material, eg jpg, gif, mp4, pdf, mp3	For archival portals probably the type (text, image, audio) is more relevant
Hierarchical level of the archival description	Search is limited only to certain levels of the archival description	Should be used for refining search results and as an advanced search option. This facet is already partly available in APE
Digital object	Search is limited to archival descriptions with digital objects	Is already available at APE
Digital object type	Search is limited to certain types of digital objects: Image, Sound, Text, Video, Unspecified	Is already available at APE
Date type	User can specify the type of the date: full date, no date, descriptive date	Is already available at APE
Dates		
Dates input as numbers	User is able to input dates as numbers or select from a built-in calendar	Is already available at APE
Dates input from a timeline with slider	User is able to select a date or epoch using a slider	This would add to the previous and make the search more user-friendly and visually appealing
Verbalised periods	The user is able to input or select a verbalised period	This possibility demands a date/epoch vocabulary to be available
Content providers		
Name of the archival institution	Institution which holds the fonds is defined in search settings	Is already available at APE
Fonds	User specifies the name of the fonds/collections to be searched	It is already possible at APE to refine search results (button "View only results from this finding aid")
Country	List of countries where the content provider comes from	Is already available at APE
Places, names & subjects		
Place and location	The search uses an index of place names	The implementation of such a facet depends on the availability of indexed place names

Facet	Description	Comments
Geosearch	The user is able to view search results on a map or alternatively select areas from the map to limit search results on other facets	This is probably not possible to be achieved in a pan-European portal as it demands for geographical coordinates to be available in object descriptions or related to the place names mentioned in the descriptions.
Names	The user is able to search for indexed names of persons, families, corporate bodies or browse a list of these	The implementation of such a facet depends on the availability of indexed names
Provenance / creators	As a sub-facet of names this facet allows to search for records from specific creators	
Subjects and keywords	The user is able to search on specific subjects and keywords	This functionality depends on the availability of a subject and/or keyword list

8.4.1. Types

Type of the archival material: In general the possible types used as facets can be grouped as follows:

- **Technical criteria:** extent, condition, (photographs) colours, drawing techniques, cartography, etc.
- **Formal criteria:** document, map, photo, moving images/video, audio, charter, seal, register, poster, etc.
- **Content related criteria:** church register, census records, etc.

Whereas technical criteria seem to be too detailed especially for a central portal, formal criteria could be an interesting facet to filter the data. Also the content related criteria are probably not necessary, especially when other facets like topics or keywords are available. Additionally, the formal criteria facet can easily be based on the ead//materialspect element which makes a first implementation rather simple. However, to make the facet on formal criteria possible solutions to deal with the variety of values input by content providers and their multilingualism needs to be worked out.

Hierarchical level of the archival description: This facet displays the level within the hierarchy of the archival description in the finding aid. So user can refine if they want to see results on the item level, the file level or the classification level as well. Even the fonds level or the finding aids title could be displayed here. By choosing the “type of document” before entering the advanced search and with “type of finding aid” within the current list view in Archives Portal Europe the user can already now filter between finding aid, holdings guide and source guide.

The hierarchical level facet can be implemented by using the EAD elements ead//archdesc@level and ead//c@level.

Digital object: This facet allows users to search only for archival descriptions with digital objects or to refine search results for this property.

Digital object type: Within the search results the facet for digital object type allows filtering for image, sound, text, video, and unspecified.

8.4.2. Dates

Dates as numbers: The facet on dates and periods is probably one of the most important ones for archival portals. The simplest solution is to just have a built in calendar or an input box for inserting start and end dates as a query argument. This is also the solution which is currently implemented in Archives Portal Europe.

In addition to inputting dates as query arguments there is a possibility to group the dates according to predefined values (i.e. centuries or decades) or cluster more flexibly based on the dates and periods highlighted in the descriptions of search results. This would allow the user to further refine the larger set of results into smaller bunches.

Timeline: Alternatively or additionally a timeline with a slider to select a year or a time span could be implemented. While the functionality itself is similar to the input of dates it would provide possibilities for creating a cleaner, more attractive and useful search interface.

Epoch vocabulary: The third possibility is the creation of an epoch vocabulary. A selection of historical, cultural or social epochs could be defined and related to certain dates and periods. This would allow the users to more easily execute queries on certain periods without being afraid of defining a too extensive or restrictive time span. This option is rather attractive also because it would allow starting rather easily with a high-level vocabulary which is extended to be more specific and detailed in future.

Another possibility is also to take into account the different durations of epochs in different regions. This would mean that the APE_x project defines centrally general cultural periods like ancient history, middle age, modern age, contemporary history. Each country can specify the exact time span for each epoch regarding the specific circumstances in this country.

8.4.3. Content providers

If a user knows already the country, the name of the archival institution or the fond where the necessary information is located, it should be possible to select this from a structured list.

The structured list should include at least three levels (or different, hierarchical, facets):

- Country
- Name of archival institution
- Fonds

The possibility of selecting the appropriate level and value should be available by browsing (for specifying it before the query) or as filtering possibilities (i.e. a list of countries, institutions and fonds is presented next to the search results which the user can select and deselect).

8.4.4. Places, person names and subjects

Place and location, geosearch: Another attractive possibility for the users would be to search using place names and locations. However, the main question when applying such a facet is the availability and quality of the information inside archival descriptions.

One possibility to achieve the necessary quality could be to automatically check archival descriptions against a place name list. Such solutions are already available as an example the DBpedia Spotlight⁶, a similar tool has also been developed by the EU funded QVIZ project⁷.

Based on the output of these tools the data in ape-format (apeEAD, apeEAC-CPF etc.) could be enriched with the automatically indexed place names. As well there are place names encoded in ead//controlaccess/geogname.

When looking at EAC-CPF it does also include place codifications. However, the usage of these elements is rather problematic and has to be further discussed before implementation while searching for archival records. The main question is whether the encoded place can really be related to the described item or not. For instance a named birthplace of a person is not necessarily related to a specific manuscript.

Person Names: Especially for hobby historians (genealogists) the search for persons and other names which are indexed in the archival descriptions can be relevant. However, similarly to places and locations this facet needs also a person name index to be available.

Currently the index could be based on the names encoded in ead//controlaccess/name, ead//controlaccess/famname, ead//controlaccess/corpname and ead//controlaccess/persname. Additional indexing could be done semi-automatically using tools similar to the ones mentioned in the previous facet. However, the added difficulty with especially person names is that in some languages these are related to common words in the respective language (person names like River, Black etc) and the automated indexing is probably not too efficient.

Names of persons, corporate bodies and families are also encoded in EAC-CPF. As soon as the standard will be implemented in the portal and content in EAC-CPF format is ingested, names can be taken from these data. A range of additional functionalities can be developed once Work Package 4 of the APE_x project has defined a profile for the portal (D4.4 planned for 05/2013).

Provenance: The possibility to use provenance information (names of corporate bodies, persons or families which have created fonds) as a facet can be predicted as a rather attractive possibility for both hobby historians and academics. The actual implementation of provenance as a facet could be based on two variations:

- Free text field available for user input
- Clusters of creator types available for browsing (e.g. administration, parties, companies, families, single persons, public / private institutions, etc.).

⁶ <https://github.com/dbpedia-spotlight/dbpedia-spotlight>

⁷ <http://www.qviz.eu/>

While the clustering of creator types seems attractive and useful it has to be remembered that typical archival descriptions only includes three different types – person, family and corporate body – inside the `eac-cpf//identity/entityType` element. Thus for more specific and useful clustering additional efforts have to be undertaken on automatically or manually analysing available descriptions in order to achieve it.

For simple names of provenances these can be found and used from all child elements of `ead//origination` and in `eac-cpf//identity/nameEntry`.

Subject and keywords: One of the most widely used facets is the subject of records. However using this in the most common way – based on controlled subject headings – is probably also one of the most difficult items to be implemented for Archives Portal Europe. Namely, it is almost impossible to use automated tools to add controlled subjects to descriptions of archival records. Also in situations where subjects are already available inside descriptions these are based on a variety of languages and different vocabularies which are time consuming to be mapped with each other.

There is also an issue of selecting the subject headings or keyword classification. For example, the most common ones are DDC (Dewey Decimal Classification) and LCSH (Library of Congress Subject Headings), but they come from the library domain and may not be most applicable for the archival records.

However, some partial solutions are possible. Especially the possibility of exploiting higher level descriptions (i.e. describing subjects mainly on fonds / series level) could be considered for assigning new facets. For already available subject headings also partial mapping of only most relevant terms into a central listing is possible. Still, even these partial solutions demand a high level of manual work and the quality and relevance of search results might still not be sufficient for the user.

Another growing trend is that instead of subjects (or additionally to these) users are able to browse the keywords created by other users. An especially attractive solution is to visualise keywords available in the preliminary result set as a “tag cloud” which presents more often used keywords with a larger font than others. Still it is quite obvious that this approach has multiple major problems:

- The scenario needs a reasonable number of keywords to be available;
- Additional (automated) quality checks might be needed before user-created keywords are possible to be used as query facets;
- Multilingualism still remains an issue.

However, these and other problems are discussed in more detail in Chapter 10.

9. Feedback tools

9.1. Short outline

Feedback tools are specific functionalities which are developed to simplify the provision of feedback from users to service providers. In case of archival portals the feedback can in general concern the following items:

- **The service itself:** General feedback which can include recommendations for updating the functionality and usability of the service.
- **Visualisation and presentation of archival data:** Specific feedback possibilities could be developed to address the visualisation and presentation of archival data.
- **Archival descriptions and objects:** Users might be allowed to report errors in or additions to archival descriptions and objects.

Taken the usual low quality and availability of archival descriptions especially on record level and the potential number of users in Archives Portal Europe especially the last item should be regarded as being of high relevance. Therefore also the rest of this chapter concentrates mainly on the feedback on archival descriptions and objects.

The area of feedback on archival descriptions and objects has also some overlap with the general area of crowdsourcing. However, in the current report the usual aspects of crowdsourcing like user motivation have been left unattended and will be looked into during the next phase of the APE_x project.

It also has to be added that according to the survey of APE_x partner archives the feedback tools are the most wished for category among all user-related tools (see Chapter 5).

9.2. Good examples

- **The National Archives, The new Catalogue (UK):**
<http://discovery.nationalarchives.gov.uk/SearchUI/> (Figure 4)

Reporting errors is possible on all fields of the respective archival description. When looking at a description the user is presented with a “Suggest a correction” button which opens a separate input form. The user is then able to select the specific description element where the error is, highlight the possible error and propose a correction. If there are errors in more than one field then a separate form has to be filled out and submitted for each of the errors.

It is also interesting to see that the form is available for all users (even those not logged in) and does not include any kind of automated input prevention tools.

Legal status: Public Record

Found an error? [Suggest a correction](#) to help improve our descriptions.

Suggest a correction for this record

Fields marked with * are mandatory.

Which field contains the error? *

What is the error? *

What is the correct information? *

Your details

Provide details if you would like to be contacted regarding your correction.

Name

Email address

Figure 4: Feedback form of the National Archives of UK

- **Photo Documentary Database of the 1956 Institute (HU):** <http://www.rev.hu/fotodb/> (log in required for feedback), Figure 5

The feedback tool of the database allows providing comments, corrections as well as adding new information to each record (photo). Therefore the aim is not only to correct the available descriptions but also extend these with user generated input.

The feedback form is only available for logged in users. Based on the information input into the form an e-mail is generated and sent to appropriate recipients. The comments will thereafter be manually processed by archivists and (if approved) the added data becomes available at the database.

Comments can only be given as open text, which means that there is no (technical) possibility for more detailed or specific feedback (as an example for a single description element).



Figure 5: Feedback form of the Photo Documentary Database

- **Prometheus Bildarchiv (DE):** <http://prometheus-bildarchiv.de/>

Prometheus Bildarchiv allows the registered users to search its content (and only the first week is free for use of personal account).

The Prometheus Bildarchiv enables the user to send a simple comment to each image. The comments can be published.

- **ICPSR - Inter-university Consortium for Political and Social Research (USA):** <http://www.icpsr.umich.edu>, Figure 6

ICPSR provides leadership and training in data access, curation, and methods of analysis for a diverse and expanding social science research community.

The feedback forms of the site are not data specific but rather “page specific”, which means that there is a feedback button on each of the separate web pages. Therefore the feedback form is not only about reporting errors in data but also about reporting other (more general issues like broken links etc).

The feedback function on all pages calls a form which allows selecting the type of feedback and adding a problem description. In addition the users might opt for leaving their e-mail address and the tool itself records the URL of the page (possible to be changed by the user).

In addition the CAPTCHA anti-automation function is used next to the feedback itself.

Report Problem

Please use the form below to report problems, such as broken links or out-of-date content. We need a clear description of the problem in order to correct it. We appreciate your feedback!

Bad Request: "The page is broken."

Good Request: "The browse by funding agency link takes me to a list of subject terms."

Email: (if you want a reply)

Type of Problem*:

URL:

Description*:



Figure 6: ICPSR feedback form

9.3. Main problems and research questions

Some questions and decision points can be identified based on the mentioned examples. Probably the first question to think about when developing feedback tools is the general scope of the feedback. In other words – it has to be decided whether the feedback we intend to gather should be more generic (which might be difficult for users to fill) or more specific (which might rule out some potential feedback areas). However, as already mentioned in Chapter 9.1 we might assume that the most important piece of feedback in archival applications would be specific error reporting about the archival descriptions. Additional, more generic feedback forms addressing the functionality, usability and visualisation of the portal should rather be implemented separately.

Another aspect is whether to allow everybody to deliver feedback or limit this possibility to only registered users. This is probably a question which cannot be answered without testing it in real life situations and it also might differ among different sets of data and portals. The limitation of allowing feedback only for the registered users might limit the amount of reasonable feedback while an open feedback form is more prone to abuse. Another problem with open feedback is that when the amount of requests increases, it may take a considerable time to process these and therefore the users may get disappointed with the portal in general (i.e. "I provided feedback but nothing has happened after this, they do not care about their users!"). Therefore it may be safer for archival institutions to first deliver feedback forms only to the registered users until the archive's team is sure that they can handle the incoming requests.

Related to the previous issue are also the questions regarding the workflow and technical possibilities to manage the feedback delivered by the users. The easiest solution is to define an e-mail address to which the filled form is delivered, but this is not the most effective solution. As an example in collaborative portals like Archives Portal Europe there is no possibility to define a single responsible but the feedback has to be delivered to the actual owners of the single descriptions (i.e. archival institutions). In addition e-mails might easily be overseen especially during/after holidays, exchange of staff or when there is a larger amount of feedback delivered by users. Therefore it seems reasonable to also consider other possibilities like developing a specialised feedback management and data correction interface for the data owners. This interface should also be specially designed to make error resolution as efficient as possible (as an example by providing simple Agree / Decline buttons next to a comparison of the original and improved description). Of course also e-mail notifications could be supported so that the data owners would first learn about the newly delivered feedback.

Especially in the case of the collaborative portals there is also a question of how to guarantee the integrity of the corrected data? While it is rather straightforward to build functionality for updating the descriptions on the portal the corrections also have to be delivered to the original systems of the data providers. In order to achieve this automatically a good level of interoperability needs to be there in both technical and semantic sense. However, as this is not yet the case among archival information systems the proposal of the APE_x team is to agree on a simple XML output format which could be exploited already in the feedback form (i.e. the feedback is sent as XML instead of simple text) or after it has been dealt with in a separate feedback management interface.

9.4. Possible components

As discussed above the feedback tools might consist of a range of separate functionalities. A summary of these is available in Table 5.

Table 5: Possible component functions of feedback tools

Function	Description	Comments
<i>User management and misuse prevention</i>		
Registration / Login	Only registered users are able to provide feedback	Probably one of the two possibilities should be used, both in parallel do not seem necessary
Prevention against automated responses (CAPTCHA)	Tools like CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) could be used to prevent automated filling of the feedback form	
<i>Scope of feedback</i>		
Software error reporting	Feedback about possible software bugs like search failure, incorrect search	In the case of archival portals it seems reasonable

Function	Description	Comments
	results, etc.	to provide a separate general form for these types of feedback
Usability and functionality feedback	Feedback about possible updates and changes in the functionality, usability and design of the portal	
Feedback about an archival description	General feedback about an archival description	
Feedback about a specific description element	Feedback which concerns a specific field of the archival description (i.e. title or date)	
Providing new information	Feedback about missing parts in the archival description	
<i>Managing feedback</i>		
Compiling e-mail to pre-defined address(es)	The feedback form is formatted into a simple e-mail and sent to a pre-defined e-mail address	This solution is possible for the Archives Portal Europe only for feedback about the portal itself
Compiling e-mail to the owner of the archival description	The feedback form is formatted into a simple e-mail and sent to the owner of the archival description	This functionality needs a connection between the description and data owner, especially that each data owner has a specific e-mail address declared
Feedback management interface	<p>The feedback is gathered into a separate interface which might include the following sub-functionalities:</p> <ul style="list-style-type: none"> • Simplified error review and description correction • Management of deadlines (i.e. feedback has to be reviewed within 3 working days) • Management of notifications • Contact user (for notifying about the correction, further 	

Function	Description	Comments
	questions etc)	
Feedback export		
Export as simple text	The feedback can be exported as simple text or PDF	These export possibilities might be supported by the feedback form and/or the separate feedback management interface
Export as XML	A XML structure has been predefined for exporting the feedback	
OAI-PMH support	The OAI-PMH protocol is used to export the feedback	

9.5. Possible implementation

As mentioned above the possible implementation of feedback tools inside an archival portal differs greatly depending on the exact scope of the portal and its contents. In this chapter you will find a recommendation for the implementation of feedback functionality inside the Archives Portal Europe.

9.5.1. User management and misuse prevention

In the case of Archives Portal Europe the potential number of users and amount of feedback is largely unpredictable. Also the amount of persons involved in the management and review of the feedback is rather high (including at least one representative from each data provider). Therefore it is recommended to first provide feedback functionality on archival description only for registered users which hopefully limits the amount of feedback and at the same time provides more quality.

In the long term the same feedback functionality could become available for other users but only after a validation of the first results. In addition the feedback functionality for public users should include tools like CAPTCHA to prevent automated misuse of the feedback form.

A possibility for delivering general feedback about the portal is available already now (see www.archivesportaleurope.net → Feedback). However, the current simple feedback form could be amended to include more detailed subject types (currently only three categories are available: Technical issues, How to join and contribute, Suggestions and other issues). In case of technical feedback also the address of the page where the issue was found could be automatically recorded as part of the feedback (similarly to the ICPSR feedback form described in Chapter 9.2).

9.5.2. Scope of feedback and structure of feedback form

As already mentioned above the most important type of information that should be gathered through the feedback functionality are the corrections to and amendments of archival descriptions. Therefore the feedback functionality should be available within all pages where archival descriptions are presented as a simple “comment” or “feedback” button which opens a separate

feedback form. The following list highlights the fields which could be available on the feedback form:

- **Type of feedback** (mandatory, implemented as a drop down list with possible values “report error”, “comment” or “extending description”);
- **Description element** (optional, should provide a drop-down list of possible description elements. In case of “report error” is selected as the type of feedback the list could be further limited to only include the names of filled description elements);
- **Description of error** (optional, might be automatically filled with the current value of the description element when the previous field has been selected);
- **Comment or correction** (mandatory);
- **Reference code** (mandatory, automatically filled. Alternatively the URL of the page could be recorded when persistent URL’s are in place);
- **Feedback to user** (mandatory, a “yes/no” checkbox where the user can highlight whether (s)he would like to be contacted on the comment or correction. In case the user would like to be contacted (s)he might also be prompted to include an e-mail address)

9.5.3. Managing feedback

The feedback form regarding additions or corrections to archival descriptions should be controlled by a specific management interface next to the data upload and management functionalities for data providers at the Archives Portal Europe “Dashboard”.

The data providers should be allowed to set up their feedback options and manage it using the following functions:

- **Review feedback.** Data providers should be allowed to review all feedback using a simple web form which allows for filtering and ordering of received feedback (as an example by selecting only reported errors or comments). Single feedback items should be able to be opened in a separate view which allows data providers to easily gain an overview of the proposed change or comment, if necessary open the full description, make further modifications and finally approve or decline the feedback. In case the user has also left his / her contact details it should be possible to deliver a notification or send additional comments to him / her.
- **Notification settings.** By default all feedback is delivered to the Dashboard accounts of the respective data providers. In addition it should be possible to further manage the notification settings by setting (an) e-mail address(es) where notifications about newly received feedback are delivered as well. In addition it might be reasonable to set a specific best effort deadline, as an example three or five working days for reviewing a certain piece of feedback. Further notifications could then be sent when some items are past their processing deadline.
- **Export settings.** The data provider should be able to export the original or reviewed feedback for importing it into the original description catalogue. The export functionality should support at least export in simple text, PDF and a predefined XML format (possibly

based on apeEAD). In addition OAI-PMH could be supported for automated discovery of changed archival descriptions.

10. Tagging

10.1. Short outline

The term *Tagging* refers commonly to the process of enriching any kind of resources with concise index terms (*tags*) to enhance the usability of these in a broader concept. Examples of tagging include the practice of adding tags to books in library catalogues or providing index terms to material in online databases like the photo sharing platform Flickr⁸, where tags are being used to describe the depicted images.

The main aim of tagging is to make it easier for users to identify material that is either related to their stated interest (for example in a web-search) or related to material they currently use (to provide “see also” recommendations). When looking at the other technologies described in this report we can also see that tagging is tightly related to some of the other areas. As an example users could use tagging to manage their link books inside their personal research space (see chapter 7), tags could be used to extend faceted search possibilities (see chapter 8) and tags could be reused for creating additional links for archival Linked Open Data (see chapter 11).

While improvement of usability is the common objective in tagging the source of tags can be varying to some degree. In very broad terms there are three different possibilities for tagging:

- **By owners of the data resources:** Based on internal rules data owners (in archival applications archivists) are provided with tagging possibilities next to usual (archival) description features;
- **By users of the data resources:** data users are provided with a possibility to provide tags while browsing and using content. Please also note the relation of this kind of tagging to two more generic Web 2.0 terms: user tagging is a *crowdsourcing* activity and the tags can be seen as *user-generated content*.
- **By artificial means:** software is used that is able to analyse certain aspects of the data and create meaningful tags without human interaction.

As this document is mainly concerned with Web 2.0 applications for the Archives Portal Europe and its data, tagging by users is the main objective of discussions in this chapter. In other words we look at tagging as a way of crowdsourcing which has the main objective to create additional metadata next to the archival descriptions created by professionals.

When looking at the usual applications of user tagging we can see that the available solutions are mainly based around digitised resources (see also Chapter 10.2). Due to the fact that the APE only stores finding aids (archival descriptions) and does not store digitised resources directly these examples will be hard to implement. For this reason this part of the document highlights a first collection of thoughts and some initial research on this topic which will be continued and enhanced during the course of the project and presented in more detailed in the follow-up deliverable early 2014.

⁸ <http://www.flickr.com>

A final introductory notice: It is difficult to differentiate tagging from other crowdsourcing techniques, like adding geo-location information to photos or other descriptive metadata to archival material by users. Especially the technical implementation (from a database perspective), PR and motivational strategies described in the next chapters will be possible to be reused also for other types of crowdsourcing besides user tagging.

10.2. Good examples

- **Tagging missions of the NARA (US):** <http://www.archives.gov/citizen-archivist/tag/>

The US National Archives and Records Administration (NARA) has recently intensified its crowdsourcing efforts with its Citizen Archivist Dashboard and has delivered various different approaches to the topic of crowdsourcing (Tagging, Transcription and Indexing). One of these efforts is the so called *Tagging Missions* which aims at enhancing the user experience through easily understandable goals and at the same time directing users to interesting topics relevant to everyday life events and history. The tags entered by registered users are reviewed by the archives personnel according to a fixed policy.⁹ If any tag is deemed appropriate by an archivist, it appears online next to the object that was tagged with this term. The tag is displayed so that it is clearly separated from other descriptive information.¹⁰ This approach has the advantage that the tags are trustworthy and user generated content cannot be mistaken for the descriptions created by the archive itself. The disadvantage is that every tag has to be reviewed at least once and that the user has to be logged in. This is a small barrier that may already hinder some people to use the system.

- **Discovery catalogue of the National Archives (UK):**
<http://discovery.nationalarchives.gov.uk/SearchUI>

The UK National Archives uses tagging to enhance the usability of their material and to “modernize the language”¹¹ of the finding aids. Authorised users can add new tags to archival material or reuse already existing ones. The National Archives markets the tagging function as a way for users to bookmark interesting resources – tagged items are listed in the user space of the user that provided this tag. Also, the National Archives do not seem to moderate the tags directly but have also *crowdsourced* this function – other users can “flag this tag as inappropriate”¹². This has the advantage that much less work is necessary to be done by

⁹ <http://www.archives.gov/social-media/policies/tagging-policy.html>

¹⁰ See <http://research.archives.gov/description/278340>. The tags are labelled “Tags Contributed by the Public” on the left border of the window.

¹¹ <http://www.nationalarchives.gov.uk/about/discovery-faqs.htm#Tagging>

¹² For instance on this search result: <http://discovery.nationalarchives.gov.uk/SearchUI/details/C210739-g-mepo-4.-no.-4-details>

archival staff. At the same time still the risk of inappropriate tags is there at least for the period until a tag is reported as inappropriate.

- **Flickr Photostreams**, for example from the NARA (US):
<http://www.flickr.com/photos/usnationalarchives>

The photo sharing platform Flickr is used by a number of archival institutions for sharing digitised photographs. NARA encourages users to post resources they digitised themselves in a research room into Flickr and therefore make these available for all other users. Flickr also allows adding tags and comment the images to provide additional context to the images. This user generated data has also good potential for extending available archival descriptions, however currently there is no easy way to export this user generated content. Instead images put online have to be monitored and interesting tag or other information has to be input manually into the archival system.

- **Waisda? Project of the Netherlands Instituut voor Beeld en Geluid (NL):**
<http://woordentikkertje.manbijthond.nl/>

The Instituut vor Beeld en Geluid (The Netherlands Institute for Sound and Vision) uses a kind of tagging game to make their TV database accessible. The principle is simple - registered users can look at a video and enter tags for things they hear and see. If the tag matches another one entered by a different user approximately at the same time, the user scores a point. This way, more relevant terms can be created because, if both tags match, chances are high that the tag is really relevant. This approach therefore creates good results and includes a high motivational factor for users.

10.3. Main problems and research questions

While showing good potential for extending archival descriptions such user tagging has also multiple problems which have to be solved first. The key areas and questions which have to be looked into are:

- *Scope of user tagging:* Which resources should be available for tagging?
- *Quality control and publishing:* Should the tagging function be only available for authorised users or the general public? Which methods should be applied to discover inappropriate tags? Should the user-generated tags be clearly differentiated when looking at archival descriptions
- *Use of controlled vocabularies:* Should the tagging function be based on certain controlled vocabularies or is it allowed to post free-form tags?
- *Reuse of user generated tags:* What has to be taken into account when reusing the tags for resource discovery and description actions?
- *Motivating users:* How to achieve a reasonable number of (quality) input?

Scope of user tagging

As already mentioned above, most tagging tools rely on direct access to digitised content. In the case of Archives Portal Europe this is not possible as all digital content is not stored at the portal but only linked to the original data providers' sites. Of course, theoretically it would be possible to demand users to look at a digital object at the external system and then come back to add tags to it inside the central portal. However, it is highly unlikely that this approach will be successful as it makes it much more difficult and unintuitive for users to take part. Therefore one of the main research tasks for the Archives Portal Europe is to decide on the actual scope of in a way that it is possible to be implemented, provides added value and is sufficiently user friendly.

One of the technical possibilities would be to first develop more sophisticated access tools for Archives Portal Europe which would allow for viewing external content inline (inside the Archives Portal Europe archival description view). However, this solution is technically hard to develop and would require certain technical developments at each of the data providers' systems.

Other more immediate possibilities would be to ask users to provide parallel terms to the available descriptions, highlight certain keywords in a more extensive description or ask for translations to already available tags or descriptions. Especially the latter two possibilities could be merged with automated tagging features – first tags could be extracted from descriptions automatically based on some available vocabularies, users would thereafter be asked to verify, modify or translate the pre-filled tags.

When looking at crowdsourcing projects in memory institutions we can also see that most of these recommend keeping the scope of activities focussed. Therefore it might also be relevant to consider limitations in tag categories (as an example focus only on tags around locations, persons and topics).

Taken the considerations above the current recommendation is to “start small” – the preliminary tagging functionalities should be rather focussed in regard to source data, user functionality and tag categories. The use of automated tagging solutions should be considered to create a preliminary tag population which the users are able to improve and extend.

Quality control and publishing

The potential of the *crowd* for extending archival descriptions can be quite high. However, one of the main problems associated to it is that most usually archival descriptions are created by professional archivists and the institutions do not have much experience regarding the quality of *crowdsourced* metadata. Taken this lack of experience it is also quite reasonable to assume that some means for providing additional misuse prevention and quality assurance have to be applied.

In very broad terms we can identify two larger areas of threats:

- Deliberate misuse and “*spamming*”
- Imprecise or misleading tags due to poor language skills, content or background knowledge

When looking at the threat of deliberate misuse the main question is whether to limit the availability of tagging functions to certain user groups or apply appropriate tools to prevent the use of automated spamming tools. As the threat is similar to the issues discussed in chapter 9 (Feedback Tools) also the arguments and possible solutions are the same. In short – due to the lack of experiences it is recommended to first test the tagging functionality with a limited number of

users, e.g. make it only available for authorised users. This might discipline the users to not deliberately misuse the functionality as there is the possibility to identify and, if needed, block them. Especially in the case when the test period reveals that the limitation has a negative influence on the amount of users exploiting the functionality (and therefore also the amount of tags created) it should be considered to open the tagging functions also for unauthorised users. Still, in this case the use of tools like CAPTCHA should be considered to prevent automated spamming.

Another (and more crucial) issue is the intellectual quality of the user-generated tags. Potentially some users might be “too eager” to use the tagging functions even without sufficient language and/or content skills which might end up in the availability of inappropriate and misleading descriptions. There are numerous possibilities to deal with such a threat:

- **Archival moderation:** The tags are not published until these have been double-checked by archival staff (either by APE or data provider staff). While this solution does provide quality which is suitable for the content owners themselves the main threat here is the potential large amount of user-generated tags which makes the application of archival moderation questionable.
- **Expert user community:** The tags are not published until these are reviewed by a special group of expert users which present a range of different cultures, regions and professions. While this solution does possibly provide even better quality than archival moderation it is quite difficult to acquire and maintain a sufficient and sustainable group of expert users which would be willing to take the task. One possibility for identifying the experts is to first monitor the actual tagging activity and identify the most active and appropriate users which are then approached with the proposal for tag moderation. However, this means that the group takes probably quite a lot of time to establish and thus the possibility can only be applied after the general tagging functionality has been available for some time. Until then other methods have to be used to ensure the quality of the tags. As well, after establishing the group the users also have to be motivated to keep fulfilling the task.
- **User cross-moderation:** The tags are published immediately; all users have the possibility to report inappropriate tags to a (team of) central moderator(s). This possibility is probably the least resource intensive for the content providers themselves. However, there still remains the threat that especially for rarely used resources the inappropriate tags are not discovered and therefore some might remain to be available for considerable periods. Another variation of this method is to use “blind cross-moderation”. This would mean that the tags created by users are not published immediately but only after the same or similar tag is provided by another user. Of course this solution needs some consideration in regard to assuring that a resource is really accessed by multiple “taggers”. Therefore it is only possible to use this method within specific applications like the Tagging Missions described above.
- **Combined approaches:** Next to the possibilities described above it is also possible to combine these for achieve more reasonable and efficient results. As an example the tags which are reported to be inappropriate or appear next to resources which are not accessed for a considerable timeframe might be reported for further review by the expert user community. In case the expert community does not review these tags in a reasonable timeframe or is not able to decide these might be forwarded for (additional) review to the data providers.

In general we recommend to start with using at least one of the quality control tools mentioned above. However, we also recommend establishing a regular review mechanism which would

measure the efficiency of the applied method and, if necessary, modify or change the selected approach and its exact implementation. Inside the Archives Portal Europe it might also prove necessary to implement multiple quality control mechanism which single content providers might opt to use or not.

Related to the quality control issue is also the question on how to publish and present the user created tags, i.e. whether the tags should become a part of the original archival description or highlighted separately, possibly including a disclaimer on the user involvement.

Until more experience is gathered the separate display of user generated content in Archives Portal Europe can be recommended. However, content providers should at the same time be allowed to individually decide whether to integrate these tags also into the original description. In addition content providers might also be allowed to decide whether the tags could be used for certain use scenarios (as an example faceted search functions and linked open data). We recommend that by default such reuse of tags is only allowed for these which have passed the expert or archival review.

Use of controlled vocabularies

While developing the actual tagging tools and applications one of the main questions is whether to also limit the amount of allowed keywords or keep it fully open – i.e. users are allowed to tag with whatever keyword they come up with.

The issue is tightly related to several other topics like the quality of tags, reuse of these and user motivation. While open tagging does potentially provide more quantity (users are more willing to tag if they are not bound to a list of terms which they first have to learn to use), the limitation in keywords would provide more technical quality (no space for spelling errors and misuse) as well as some more potential for reuse (more resources being tagged with the same keyword, resulting in a more extensive list of related resources for the users accessing a certain item).

Another argument when deciding on the use of controlled vocabularies is also the availability of these. Especially in the case of multilingual portals like Archives Portal Europe it can be argued that extensive application of controlled vocabularies is not possible due to the fact that a sufficient number of these is not available in all languages.

Based on these arguments we recommend to apply a combined approach – while the users should be able to use a simple text input box for tagging recommendations are provided based on available vocabularies. This approach should be reasonably convenient for the users with the recommendations providing additional quality and reuse potential.

In addition we also recommend the limitation of tag categories (topic, date, location, person, event, etc.). This would enhance the reuse potential as well as add new possibilities for the tagging itself. Especially in the case of dates and locations such categorisation provides the possibilities to develop more visual solutions like calendar or slider based tagging (for dates) or geotagging (for locations).

Reuse of user generated tags

When looking at possible reuse scenarios the majority of these include explicit use of tags to facilitate better resource discovery results either as part of the already established search tools or separate browsing possibility.

When using tags as a part of available search functions the question is whether to highlight user generated information as separate facet(s) or simply include the tags into already established facets and simple search tools. The decision should be taken based on the quality of the user generated tags.

The most common example of a separate tag browsing functionality is the so-called “tag cloud” which visualises most used tags and allows accessing the resources which have been tagged with these. If controlled vocabularies are used also possibilities for browsing vocabulary terms can be made available, in the case of geotagging also map-based solutions similar to faceted search cases can be developed. However, in all of these examples the user has to be indicated about the amount of resources which have been tagged. In other words – the user has to know that when browsing the tag cloud, map-based applications or any other tools the amount of resources (s)he can access is rather limited as not all of the records have been tagged.

For Archives Portal Europe also principles have to be in place for the reuse of user generated content in donor systems as well as other portals receiving information from Archives Portal Europe. While a central reuse policy would simplify the development of necessary functionality and also allow other portals to get a more simple understanding of the content it is not reasonable to assume that all (or most) content providers would be able to agree on this. It is instead recommended to develop functionality which allows individual content providers to consider whether user generate content is of sufficient quality, integrating tags into original archival descriptions and/or allow reusing these for export into Europeana and other systems which reuse content from Archives Portal Europe.

Motivating users

Last it also necessary to discuss the motivation aspect – how to attract a reasonable number of users to take part in the tagging?

One possibility for this would be to include some gaming and competitive aspect into the actual task. This means that tagging is “hidden” into some interactive games and/or best taggers are highlighted on leaderboards. However, a number of possible problems can be identified next to such gaming solutions:

- While participating in games might be attractive to some users it might also be disturbing to others
- Some users might just join to get on the leaderboard and not really bother about the quality of the data they deliver
- In long term users might also get bored from a certain game and therefore these have to be constantly updated and new concepts developed.

To summarise – while such games might help to attract a reasonable amount of publicity and users in the start, the solution needs constant rework and attention in long term.

Another possibility is to prepare “*microtasks*” like the tagging missions in NARA. The benefit of this solution is that a user can see a clear scope and also the benefit of the tagging mission. Also the archives are benefitting from this as they can guide users to certain materials which they prefer to have tagged and are not only reliant on the users “accidentally” accessing the material. While also the preparation of such missions requires some effort it might be assumed that users are attracted to such environments for longer periods than for games and especially when the preparation of

these missions is done semi-automatically the amount of work needed for the maintenance is not too extensive.

Next to the development of attractive technical solutions also a “reward program” is possible. As an example users might gather points when doing tagging which can then be used to get invitations to special events, branded cups, pencils or free copies of archival records. However, again because of the multitude of archives participating in the Archives Portal Europe it might be rather time consuming to reach an agreement on the principles and rewards in the program.

It is rather clear that each of the different solutions can be attractive to certain user groups. To fully exploit the potential of all users we therefore recommend a combined approach – while some tagging solutions (like tagging missions or games) might be available in a central area users should also be able to add tags when casually browsing Archives Portal Europe. Despite of the location of the tagging action also some limited and mostly non-monetary rewards should be available for the most active participants.

10.4. Proposed components and architecture

The concepts discussed above are in a rather early phase and therefore it is not possible to deliver detailed descriptions of a tagging solution. However, we can identify the main building blocks which need to be available for a holistic tagging solution:

End-User (Portal, public and authorised users):

- **Tagging area:** Separate section where users are allowed to participate in specialised tagging missions and games. It should also be possible to see current “highscore tables” on best contributors
- **Tag access applications:** Specialised area to include applications which allow the browsing of tags and accessing tagged resources. Authorised users should be able to change or delete the tags they have created themselves. This area should also be tightly connected to the gaming area
- **Single item tag area:** When accessing archival descriptions users should be able to view tags delivered to the item and provide new tags in an separated area next to the original archival description
- **Search extension:** The current extended search needs to be extended so that it allows users to select whether they want to search among the tags created by other users or not.

Moderator (Portal, selected authorised users):

- **Tag search:** The moderator is allowed to search and browse for new tags, sort and filter these.
- **Tag review:** The moderator is able to view the tag in context (i.e. together with the original description on Archives Portal Europe), change, accept or decline these. If appropriate quality control method has been chosen it might additionally be possible to forward the tag for additional review or decline reviewing a certain tag.
- **Tagging management:** Some or all moderators could have an area to view statistics on tagging and if necessary “ban” or “promote” certain users. Alternatively these functions could be available for the general portal administrators.

Content provider (Dashboard):

- **Tagging settings:** In case multiple options are available the content provider should be able to select which control method to use and how/if tags are published on the portal.
- **Reuse area:** The content provider should be able to decide on the scenarios where the tags are allowed to be reused (as an example linked open data). It should also be able to integrate tags into archival descriptions and export the resulting ead files.

11. *Linked Open Data*

11.1. Short outline

The terms Open Data (OD) and Linked Open Data (LOD) are in the current document understood as respectively the four or five star open data defined in the “five star open data” model¹³. This means that:

- **Four star open data** (Open Data) is data which is made available on the web as structured data and is using open formats. Each of the data entities is identified by an URI so that it can be referenced by other people or applications. This makes it possible for external actors to reuse the data to create novel mash-up applications.
- **Five star open data** (Linked Open Data) has the same characteristics as four star open data but in addition links the data to other resources, especially to global or local ontologies. This makes it possible to automatically build connections to other data sets and therefore add further potential for developing external mash-up applications.

In general open data and linked open data should be seen as methodologies rather than certain tools or technologies. In a very broad sense the aim of (Linked) Open Data is to publish information on the internet in a way which would be understandable for computers. As such, it is also tightly related to the concept of the Semantic Web as it is expected that the *Data* is semantically enriched.

The benefit of the (Linked) Open Data varies for the different stakeholders. For the general public (and the users of the Archives Portal Europe), open access to more data of public archives previously unavailable is simply a big step forward. It is also important to notice that this would be the first time in Europe that a multitude of data from a large number of European archival institutions will be aggregated and accessible in one place. It is also apparent that Open Data stimulates the re-use data. With the Linked Open Data, archival data can be connected to a wider data of cloud where new research questions and discoveries may happen. As for the content providers of the Archives Portal Europe, this is an opportunity for them to publish their data with ease. Some small size archives simply may not have time and resources to work on Open Data, therefore, the Archives Portal Europe is a great help to gain access to the tool with Open Data functionalities and support from experts in Europe. It is also worth stressing that Linked Open Data in Archives Portal Europe will be one of the long-term investments of the infrastructure, therefore, being able to encourage the stakeholders and sponsors to promote the project into the sustainable service.

When comparing the different tools or technologies described in the current report we can also see that the (Linked) Open Data solutions are rather less mature than in other areas. Therefore this chapter does not deliver specific recommendations or solutions but rather discusses the problems and possibilities of Linked Open Data in archival applications. More specific solutions should be available in the next version of this report in early 2014.

¹³ See also <http://5stardata.info/>

11.2. Good examples

- **Europeana LOD:** <http://pro.europeana.eu/linked-open-data>

Europeana is actively looking into applying LOD and has also some data already available for reuse. The project has defined a Europeana LOD profile which builds on the Europeana Data Model (EDM) and supports a variety of international ontologies and vocabularies. Also the technical documentation and best practice recommendations available on the referred site is good for getting a first insight into applying LOD principles in memory institutions.

- **Amsterdam City Museum** <http://semanticweb.cs.vu.nl/lod/am/>

As a partner in EuropeanaConnect the Amsterdam City Museum has provided parts of their collections based on the Europeana LOD model. The page also includes more specific descriptions on the conversion of data into the LOD format and enriching the data with links to global ontologies

- **Linking Lives:** <http://archiveshub.ac.uk/linkinglives/>

Linking Lives is a project by the UK-based Archives Hub – a consortium of academic archives in UK. The project has specified a RDF application profile for archival descriptions based on the EAD schema and XML Schemas to map archival descriptions from EAD into Linked Open Data (RDF). The process and background of the project are well documented and the project page provides a wealth of information around the issues.

- **SNAC: The Social Networks and Archival Context Project:**
<http://socialarchive.iath.virginia.edu/prototype.html>

The project is not directly about Linked Open Data, but its scope is highly relevant to the LOD movement in the archival domain. The main goal of the project is to derive EAC-CPF description from EAD finding aids. The named entities of creators and people referred in the archival descriptions are extracted and are matched against each other, as well as combing them with known authority records such as OCLC's VIAF (The Virtual International Authority Files) and Paul Getty's ULAN (Union List of Artist Names).

- **British Museum:** <http://collection.britishmuseum.org/>

The Collection Online from the British Museum is mapped to CIDOC-CRM and the website provides a Linked Open Data and SPARQL endpoint. The data itself is available in various formats such as RDF/XML, Turtle, Notation-3, and JSON.

- **German National Library:**

<https://wiki.dnb.de/display/LDS/Dokumentation+des+Linked+Data+Services+der+DNB>

As a part of knowledge base of the German National Library, Linked Open Data is available license through various interfaces such as SRU, OAI-PMH, and data dump. The data includes bibliographic records and authority records published as CC0, while the German Dewey Decimal Classification under CC-NC-BY.

- **LODAC project (Japanese):** <http://lod.ac/>

The project attempts to connect academic data by converting traditional data into LOD. As an example information on species and archaeological objects is available and linked with external data such as locations, events, and DBpedia. The project provides a SPARQL endpoint as well as various practical mobile and web applications (<http://lod.ac/apps>) to demonstrate the power of LOD and mash-up applications (multi-data geo application, data visualisation, etc.)

- **Bibliothèque nationale de France:** <http://data.bnf.fr/>

The French National Library has developed a rather interesting access tool based on their Open Data. The page includes possibilities to browse through the list of authors, works and topics all of which are based on open data provided by the library and connected to each other as well as some external resources like Wikipedia.

- **CKAN The Data Hub:** <http://datahub.io/>

The CKAN Data Hub is probably the largest open data catalogue currently available. The catalogue includes at the point of writing 5050 open data sets, among which are also hundreds of examples from memory institutions.

11.3. Main problems and research questions

As Linked Open Data is a rather new concept in the archival sector there are quite a few possibilities and discussion items which need to be addressed in order to be able to provide archival data as (Linked) Open Data.

This chapter provides just a first insight into some of the issues. If the reader is especially interested in the technical details of providing LOD we recommend to use the information available at

<http://linkeddata.org/> as well as from the book “Linked Data: Evolving the Web into a Global Data Space” by Tom Teath and Christian Bizer¹⁴.

Scope of data: When starting to think about the provision of LOD the first question is the scope of data which should be delivered as Open Data. In most of the current examples (Europeana, Archives Hub) the approach is to base the Open Data profile on the already available description standards like EDM or EAD and this seems also to be the reasonable approach for Archives Portal Europe. However, we have to take into account that in the case of Linked Open Data the hierarchical structure of archival descriptions is not easy to be reused in a reasonable way. In other words – the information which can be used to create links between data can be available only on higher description levels and not on single record level while still being relevant for these. This difficulty might probably be overcome with some limited data duplication – most crucial descriptions should not only be highlighted in the open data set of the original description unit but also repeated in lower aggregations.

In addition to the archival descriptions the possibility to publish authority records as LOD has to be looked upon. Authority records are usually described in more detail than archival records or aggregations and in most cases they also represent a single (holistic) entity rather than a complex relational or hierarchical network. Also the linking of authority records might produce good results (it is reasonable to assume that quite many archival creators are also present in other data sources). Therefore it might be argued that for archival institutions interested in doing first experiments in the field of LOD it is more reasonable to start with the authority records rather than archival descriptions to avoid some of the problems present with the latter.

Referencing archival descriptions: One of the crucial requirements for Open Data is that all relevant data entities (like archival descriptions or authority records) should be referenced by unique (and hopefully persistent) URIs. In addition it is preferred that the URIs would include both human and machine readable presentations. In the case of archival descriptions this would mean that when a description unit is identified by the URI <http://www.archivesportaleurope.net/data/ID1> a normal user accessing this page should be able to see a usual visual presentation of the description and associated computer files (digitised or born digital) while a semantic search engine or some other LOD application would be able to exploit the same description in machine readable form.

In addition the provision of unique identifiers requires the availability of Persistent Identifiers (PIDs). In other words it is not possible to deliver reasonable URIs when there is no persistent identifier available to uniquely identify the archival description or authority record. Especially the persistence of the identifiers is crucial in order to raise the level of trust and long term sustainability of the LOD. This is a question which is especially difficult to be answered in the case of central portals with multiple data providers. While original archival reference codes might be a reasonable first choice of a persistent and unique identifier in practice these tend to change due to rearrangement and classification processes. In addition there is currently no common agreement on the structure and management rules of reference codes among archival institutions which makes these unsuitable for long term identification. Another possibility is to provide identifiers

¹⁴ A web version of the book is available at <http://linkeddatatbook.com/editions/1.0/>

centrally at the portal (as an example during data upload) which would simplify the management of these greatly. However, in this case special solutions have to be provided which would allow managing and maintaining the integrity between the central portal and single data providers' systems. Also the possible re-uploading of data needs to be analysed and solutions found in order to not assign new identifiers to re-uploaded data.

Annotating and linking the data: Open data as such does not demand the data to be annotated and linked to any ontologies or vocabularies. Still, by doing so we largely increase the potential for data reuse and therefore also the added value delivered by mash-up applications.

When looking at the issues associated with the annotation of Open Data we can see that these are very much similar to the problems and issues discussed in the previous chapter (Chapter 10, Tagging). In other words, to publish data as Linked Open Data the information has to be tagged according to controlled vocabularies. In the case of archival information the following types of vocabularies and ontologies seem to be suitable for primary consideration:

- Persons, families and corporate bodies;
- Locations
- Events
- Dates and periods
- Subject keywords.

While the identification of most useful global vocabularies still has to be done it might be assumed that the provision of reasonable and quality annotations will be a major task inside the Archives Portal Europe Web 2.0 activities. Taken the amount of information included in the portal it does make most sense to start with the analyses of automated annotation tools like the DBpedia Spotlight¹⁵. However, as the quality of such tools is currently not sufficient for full automation the results still have to be manually checked possibly by the users of the Archives Portal Europe as part of the tagging functions. Also the availability and extent of vocabularies in different European languages varies significantly which can further reduce the applicability of automated solutions.

Technical format of open data: One of the last (technical) issues is the selection of the most suitable technical format for archival open data. Currently the most used format is RDF (Resource Description Framework)¹⁶. However, also other formats like N3, JSON or Turtle are used to some extent.

For more information on the possible formats and their application we recommend you to consult <http://linkeddata.org/> and look for examples available on the CKAN Data Hub (<http://datahub.io>).

Legal issues: Next to the technical issues Open Data includes also some questions and discussion items regarding the legal licensing of the data. Currently archival data is often only possible to be used inside closed archival information systems and according to strictly regulated use cases. When opening up your data especially the end-user application scenarios and use cases are rather

¹⁵ <https://github.com/dbpedia-spotlight/>

¹⁶ http://en.wikipedia.org/wiki/Resource_Description_Framework

unpredictable. While this can clearly lead to benefits for the archival researcher it also includes possible risks for archival data and institutions.

Probably the most discussed risk is the potential misuse of open data. This means that there is potential for applications which present and mash-up open data in a way which is harmful for persons or institutions. As an example open crime data can be used to create “heat maps” which visualise the amount of offenses on a mapping application like Google Maps¹⁷. While normal citizens can exploit this information to avoid certain areas it might be argued that persons with criminal intentions might be able to more easily select the areas of activity based on the visualisation.

Because of the usual old age of archival open data such harmful reuse is rather less obvious. However, especially when considering the rather poor quality of archival descriptions “unintentional misuse” might happen. This would mean that data is used and visualised in a way which leads to misinterpretations of it.

Based on these considerations we can see that the quality and sensitivity of data need to be re-considered by each of the data providers when thinking about publishing archival data as Open Data. We can also see that there may not be a single “publish everything” solution in place in the case of central portals like Archives Portal Europe. Instead each of the data providers has to be made clear about the possible risks and possibilities of Open Data so that they can make individual informed decisions about whether to publish their data at all and under which conditions. Obviously also technical functionalities need to be available to support such individual decision-making and data publishing workflows.

In recent years also a range of formal licenses have been developed to simplify open data publishing from a legal perspective. A rather good overview of these is offered in “Licensing Open Data: A Practical Guide” by Naomi Korn and Prof. Charles Oppenheim¹⁸. A list of potential licenses is also available at [opendefinition.org \(http://opendefinition.org/licenses/\)](http://opendefinition.org/licenses/).

Encouraging reuse: Creating LOD is of course not the end of the story. Next to the provision of data in open formats it is also essential to encourage the public to get involved for the re-use of data. In recent years a number of possibilities have become available to do so. From the memory institutions perspective it is probably worth to mention the Europeana hackathon¹⁹ which provides a possibility for data owners and developers to come together with the goal of creating novel access and reuse tools. When looking at US the LOD-LAM (Linked Open Data in Libraries, Archives and Museums) Challenge²⁰ is worth mentioning as well. It is clear that exploiting and creating such possibilities also in Archives Portal Europe would be essential to avoid just creating “unused LOD” and to gain better visibility of archival data and institutions.

¹⁷ The example is based on a real application available at <http://crimeview.psi.enacting.org/>

¹⁸ Web version available at http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf

¹⁹ <http://pro.europeana.eu/web/guest/hackathons>

²⁰ <http://summit2013.lodlam.net/challenge/challenge-entries/>

12. *Summary and outlook*

The work on delivering user-centric functionalities inside Archives Portal Europe is still in its early phases. Especially the areas of tagging and linked open data along with related items like applying semantic tools and persistent identifiers will be more closely reviewed and researched in 2013. In addition we will continue to connect to both our users and content providers to look more closely into their needs and expectations.

We also encourage you to contact the APE_x project and its Work Package 6 in case you have any comments, additions or good hints in respect to the content of this deliverable. This would help us to provide even more value for you in the next version of this report in 2014!

Annex I: APE_x Web 2.0 questionnaire

This questionnaire is intended to gather feedback about social collaboration and contribution tools and their usage in APE_x partner institutions. The answers will form the basis of a public report "Web 2.0 in archival institutions" and will also be used to propose tools and principles to be implemented in the Archives Portal Europe.

While we are disseminating this questionnaire among APE_x partner institutions only, you are welcome to forward it to other archives which are actively implementing novel tools and principles!

General questions

Please provide contact information for your organisation so that we can contact you in the event of more specific enquiries!

1. Name of your institution	
2. Name and e-mail address of contact person	

Usage of Web 2.0 tools

This section is intended to gather information about your current experiences and future plans on social collaboration and contribution tools. Please provide answers about tools used by your researchers which are either developed by your own organisation or externally (i.e. by other memory institutions).

3. Have you implemented social networking tools (e.g. forums, blogs, wikis, etc) for your researchers? If yes, please provide a short description and web address	
4. Have you implemented feedback tools (e.g. reporting on errors in archival description)? If yes, please provide a short description and web address	

<p>5. Have you implemented personalised research tools (e.g. creation of personal databases, save and organise bookmarks, documenting research results) for your researchers? If yes, please provide a short description and web address</p>	
<p>6. Have you implemented crowdsourcing tools (e.g. transcription of digitised images, connecting collections, storytelling) for your researchers? If yes, please provide a short description and web address</p>	
<p>7. Have you implemented an environment for organising virtual exhibitions? If yes, please provide a short description and web address</p>	
<p>8. Have you implemented faceted search tools in your research environment? If yes, please provide a short description and web address</p>	
<p>9. Have you implemented tools to allow for multilingual access to archival records and / or their metadata? If yes, please provide a short description</p>	
<p>10. Have you implemented linked data principles to publish and / or connect your collections? If yes, please provide a short clarification</p>	
<p>11. Have you implemented tools for tagging archival records and descriptions (e.g. keywords, person names, geotagging)? If yes, please provide a short clarification</p>	

<p>12. If there are other Web 2.0 tools which you have implemented but not mentioned in earlier questions, please provide a short description here</p>	
<p>13. Which social collaboration and contribution tools do you intend to implement in the next two years? Please provide the name(s) and short descriptions</p>	

Web 2.0 and archival users

This section is intended to gather information about the relevance and importance of different Web 2.0 tools for your researchers and organisation.

We would also like to know your opinion about users in general - which are the main user groups of your archives, which could be the main user groups of the Archives Portal Europe and which user groups would potentially be the ones most interested in using social collaboration and contribution tools.

14. Please rate on a scale of 1–5 which of the listed tools would your researchers find most beneficial for their work? (1 – least relevant; 5 – highly relevant)

Social networking	
Feedback tools	
Personal research space	
Crowdsourcing	
Virtual exhibitions	
Multilingual access	
Linked data	
Faceted search	
Tagging archival records and descriptions	

15. Please rate on a scale of 1–5 which of the listed types of tools / technologies would your institution find most beneficial for its operations? (1 – least relevant; 5 – highly relevant)

Social networking	
Feedback tools	
Personal research space	
Crowdsourcing	
Virtual exhibitions	

Multilingual access	
Linked data	
Faceted search	
Tagging archival records and descriptions	

16. Please rate on a scale of 1–5 how important do you regard the following user groups **for your institution** (1 – least important; 5 – most important):

Academics (historians)	
Academics (other studies)	
Hobby historians (genealogists, local history, war history etc)	
Students (elementary, secondary and high school)	
Teachers (elementary, secondary and high school)	
Librarians and museologists	
Archaeologists	
Photographers	
Writers	
Journalists	
Lawyers	
Economists	
Architects	
Civil servants	

17. Please rate on a scale of 1–5 how important do you regard the following user groups **for the Archives Portal Europe** (1 – least important; 5 – most important):

Academics (historians)	
Academics (other studies)	
Hobby historians (genealogists, local history, war history etc)	

Students (university studies)	
Teachers (elementary, secondary and high school)	
Librarians and museologists	
Archaeologists	
Photographers	
Writers	
Journalists	
Lawyers	
Economists	
Architects	
Civil servants	

18. Please rate on a scale of 1–5 which user groups would in your opinion be most interested in using social collaboration and contribution tools (1 – least interested; 5 – most interested)

Academics (historians)	
Academics (other studies)	
Hobby historians (genealogists, local history, war history etc)	
Students (university studies)	
Teachers (elementary, secondary and high school)	
Librarians and museologists	
Archaeologists	
Photographers	
Writers	
Journalists	
Lawyers	
Economists	

Architects	
Civil servants	

General comments

19. If you have any additional thoughts or comments please use the box below!