

Deliverable:

D4.3. Final report on the quantity and quality of content

Project URL:	http://www.apenet.eu /	
Portal URL	http://www.archivesportaleurope.eu	
Deliverable number/name:	D.4. 3. Final report on the quantity and quality of content	
Project runtime	15.01.2009 – 15.01.2012	
Document ID:		
Covered period:	15.01.2009 – 15.01.2012	
Dissemination level	PP (Partners)	
Due date of deliverable:	15th of January 2012	
Actual submission date:		
WP allocation:	WP4	
Status:	Final	
Authors:	Lucile Grand	
Reviewers	Kerstin Arnold, Wim van Dongen	

This project is funded under the eContent*plus* programme¹,(Best Practice Networks_ECP 2007 DILI 517011) a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.



eContent*plus*



Table of Contents

1 Introduction.	3
2 Quantity of data	
The descriptive units:	
The digitised documents.	
3 Quality of the data.	7
3.1 The portal in itself ensures a general level of quality.	7
3.2and reveals at the same time some issues that have to be solved	13
4. Conclusion.	16



1 Introduction

This final report intends to provide some figures and some statements about the quantity and the quality of the contents uploaded in the Archives Portal Europe. The partners are still uploading and indexing data, therefore, the amounts of data mentioned are not definitive.

2 Quantity of data

In the document of work of the project, the partners had indicated the amounts of descriptive units and linked images that they intended to upload in the portal within the lifetime of the project.

Different indicators had been included in the same document, in order to monitor the project with regard to its contents:

- 12 countries
- 50 000 institutions accessible via the portal

and

- 17 millions descriptive units
- 31 millions digitised documents accessible via the portal

At the date of the 15th January, the Archives Portal Europe offered the following:

- 14 countries representing 62 archival institutions uploaded content: the 12 partners plus Ireland and Belgium that joined the project shortly after its launch;
- 34 283 institutions are accessible via the directory of the portal.

The archival content itself, as available by the 15th of January, provided access to :

- 14 501 301 descriptive units
- 63 781 468 digitised documents.

So, basically, the target for the descriptive units has not yet been reached (but the institutions are still uploading content and no doubt that the amount of descriptive units will increase) whereas the second one has already been doubled.



The descriptive units:

This corresponds to the way archival material is described by the archivists, following the international standards (ISAD(G) and EAD). Depending on the software used by the participants, the figures provided in the APEnet description of work more than four years ago could only have been indicative because:

- some countries experienced internal delays in their own project for different reasons, and could not provide the expected amounts in time, but will do so in the near future. For instance, the project for the National Archives of France, including a new building, a new website and a new data management system was supposed to be finished and open to the public at the beginning of 2012, and it will be only in 2013;
- some countries manage their archives with databases and not with files; therefore, they had, during the project, to come up with specific exports, a task that was more or less difficult or expensive to do than foreseen, depending on the local situations, software used etc. The figures given at the beginning of the project were only estimations and/or the partners concerned could manage to handle different amount of data than expected: for two countries, the final amount of descriptive units is much more important, for one it is much less important.

Nevertheless, a solid basis for the Archives Portal Europe has been achieved jointly by all partners, and the amount of descriptive units will only increase.

The digitised documents

The amount of digitised documents made accessible via the Archives Portal Europe is very impressive. The extra-amount compared to the foreseen amount is due to the huge contributions of mainly three countries: Sweden, Finland and Spain.

The delicacy of the archival material lies in the fact that you can link one or several digitised images to one description, like books that have a different amount of pages, but in order to do this, you can provide one or several links, at different levels of the description. Besides, a link by itself does not indicate the final number of images that you will see on the original website of the partner.

For instance:

- Collections of maps or posters are usually in a 1:1 relation: one descriptive unit equals one image, therefore only one link leading to only one image is needed (like the maps of the Nationaal Archief, see figure 1).
- Most other types of archival documents are usually in a 1:n relation: one descriptive unit equals several images (up to several hundreds), but for each descriptive unit, you can either have several links (like the church records of Finland, see figure 2), or only one link (like the records of the Archivo de la corona de Aragon, Spain, see figure 3).

Therefore, the number of images indicated cannot be calculated automatically and is provided by the partners themselves who host the digitised documents.



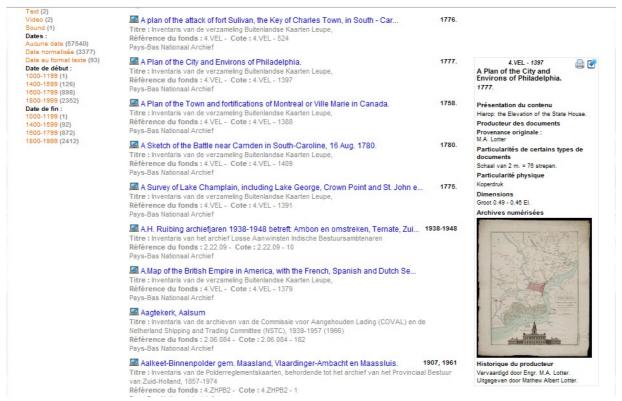


Figure 1. One descriptive unit links to one image – view in the list results (National Archives of the Netherlands)

CRITÈRES SÉLECTIONNÉS	: Avec images numérisées =		
Date de début : Titre : Pālk; 1800-1799 (20) Référence 1800-1999 (67) Finlande Na 1800-1999 (16) ☑ Pāā-jā 1800-1999 (62) Titre : Pālk; Référence Finlande Na ☑ Pāā-jā Titre : Pālk; Référence Finlande Na	Pää- ja rippikirjat Titre: Pälijärven seurakunnan arkisto Référence du fonds: FUFLHeNAF/223495.KA - Cote: FUFLHeNAF/I Aa:1 Finlande National Archives of Finland	1851-1860	FI/FI-HeNAF/I Aa:2 🚔 🕻 Pää- ja rippikirjat 1860-1870
	■ Pää- ja rippikirjat Titre: Päiljärven seurakunnan arkisto Référence du fonds: FVFI-HeNAF/223495.KA - Cote: FVFI-HeNAF/I Aa:2 Finlande National Archives of Finland	1860-1870	Archives numérisées
	☑ Pää- ja rippikirjat Titre: Pälkjärven seurakunnan arkisto Référence du fonds: FVFI-HeNAF/223495.KA - Cote: FVFI-HeNAF/I Aa:3 Finlande National Archives of Finland	1871-1879	
	■ Pää- ja rippikirjat Titre: Päiljärven seurakunnan arkisto Référence du fonds: FVFI-HeNAF/223495.KA - Cote: FVFI-HeNAF/I Aa:4 Finlande National Archives of Finland	1880-1890	
	☑ Pää- ja rippikirjat Titre: Pälkjärven seurakunnan arkisto Référence du fonds: FVFI-HeNAF/223495.KA - Cote: FVFI-HeNAF/I Aa:5 Finlande National Archives of Finland	1880-1890	
	■ Pää- ja rippikirjat Titre: Päikjärven seurakunnan arkisto Référence du fonds: FVFI-HeNAF/223495.KA - Cote: FVFI-HeNAF/I Aa:6 Finlande National Archives of Finland	1891-1900	
	Pää- ja rippikirjat Titre: Pälkjärven seurakunnan arkisto Référence du fonds: FVFI-HeNAF/223495.KA - Cote: FVFI-HeNAF/I Aa:7 Finlande National Archives of Finland	1891-1900	

Figure 2. One descriptive unit links to several images through several links – view in the list results (National Archives of Finland)



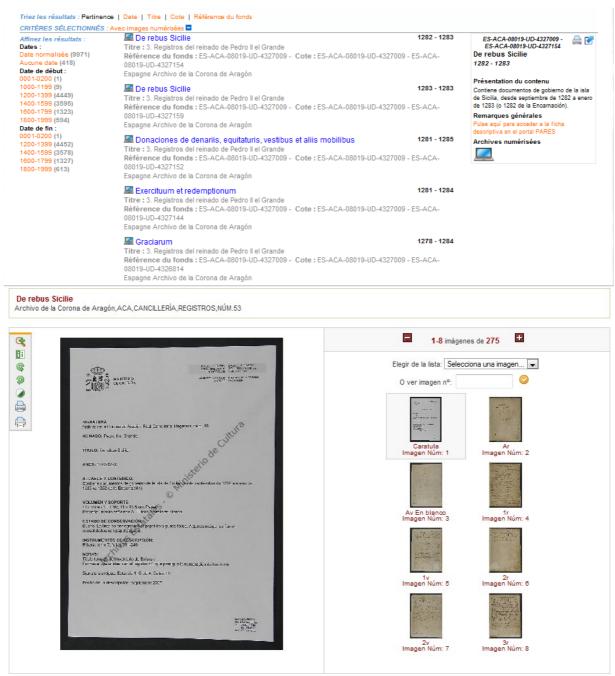


Figure 3. One descriptive unit links to several images through one link – view in the list results and view on the original website (Spain)



3 Quality of the data

3.1 The portal in itself ensures a general level of quality...

In order to index and publish content in the Archives Portal Europe, all partners have to provide data that are compliant to a certain number of standards and rules. If the data are not compliant to these standards and rules, they are rejected and will need to be converted to a common EAD XML profile, that has been defined by the project for finding aids and holdings guides and can be seen as a pivot format. The same goes for the EAG files used to describe the archival institutions.

The portal ensures thus a basic general data quality for all finding aids and holdings guides published that allows having a common search and harmonised presentation of the data. This point is very important for the final user of the portal.

Each partner can use the tools provided by the project in order to first check its own data (for instance check if their EAD is valid, if they use UTF-8 encoding etc.), and/or prepare and improve them by converting, adding or completing them, before uploading them in the Archives Portal Europe.

As a result of this efforts to ensure the quality of the data within the Archives portal Europe, the user can benefit from a joint search and common display of the search results, as well as a unique and harmonised display of the data (see figure 4). The major interest of the chosen display is to allow the user to navigate within the finding aids from any particular place in it.



Unique and harmonised display of the data in the Archives Portal Europe (see figure below):. the major interest of the chosen display is to allow the user to navigate within the finding aids from any particular place in it.



Figure 4. Display of the data in the Archives Portal Europe



Some examples of the different ways of displaying the data in the archival institutions' own webservices throughout Europe :

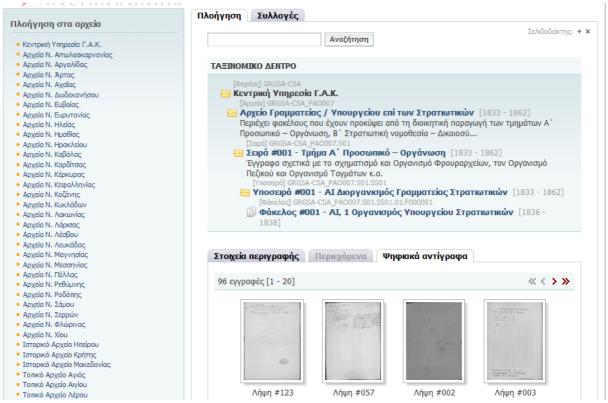


Figure 5. Display for Greece



Figure 6. Display for Finland





Figure 7. Display for France





Figure 8. Display for Slovenia

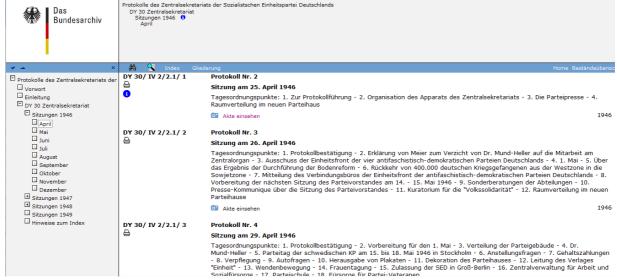


Figure 9. Display for Germany



Koloniën / Stamkaarten Oost-Indische Ambtenaren Zoeken

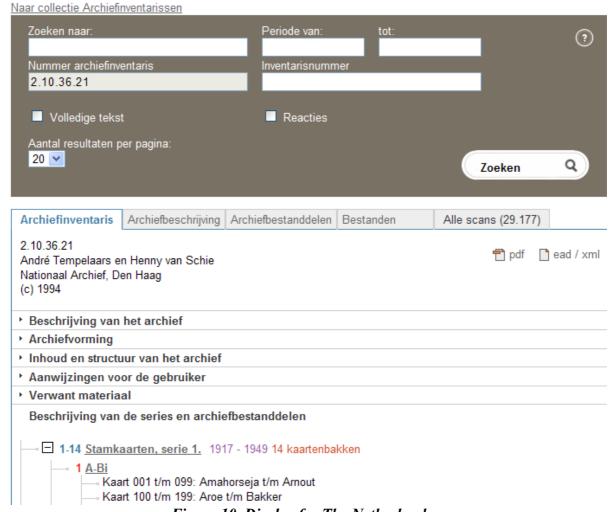


Figure 10. Display for The Netherlands

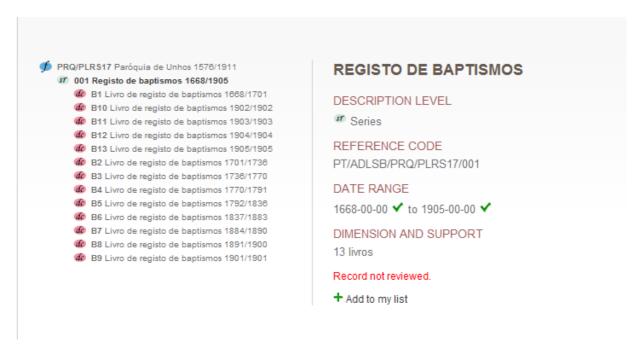




Figure 11. Display for Portugal

3.2 ...and reveals at the same time some issues that have to be solved.

Depending on regional or national traditions in the archival description, depending on the compliance to standards, and depending also on the software used, archival content is used and displayed quite differently in the content providers' own systems. When bringing the content to one joint publication and research platform – *i.e.* the Archives Portal Europe – however, this action might reveal gaps in the original data,, that are due to specific information necessary in a central system, but probably not needed that much in the system of only one institution, or due to the local systems specifically taking care of certain aspects in the set-up of the data, which might not be recommendable in the same way when gathering data from a lot of different sources in one joint access. These gaps and differences are not major ones, *i.e.* they are not hampering the joint presentation and overall search at the Archives Portal Europe in general, but they reveal aspects for possible future enhancements. They can be divided in four main groups: dates, titles, thumbnails and links.

3.3.1 The dates:

Normalised dates following ISO standard 8601 are the only ones that currently can be used for faceting and searching the data on the Archives Portal Europe. The current content has normalised dates for more than 75% of the total, representing a high proportion. But – due to various reasons _- some documents either might not have any information on dates, or might only contain textual information on dates, that would need human input in order to be normalised.

Therefore, the Archives Portal Europe offers different date-related facets for the refinement of search results in the "List view", covering all different types of dates. The refinement of search results is a good option to detect possible gaps in the date information.

```
Start timespan:
0001-0200 (5198)
0400-0599 (4)
0600-0799 (2)
0800-0999 (10)
1000-1199 (820)
1200-1399 (5654)
1400-1599 (166396)
1600-1799 (476115)
1800-1999 (6617183)
2000-2199 (38971)
End timespan:
0001-0200 (11)
0200-0399 (1)
0400-0599 (2)
0600-0799 (6)
0800-0999 (13)
1000-1199 (611)
1200-1399 (4733)
1400-1599 (160419)
1600-1799 (417263)
1800-1999 (6664399)
2000-2199 (57138)
```

Figure 12. We can see here that the amounts of "suspect" dates – in the archivistical European world, this would generally means prior to 800 – is really small compared to the



total amount of dates.

Besides, the dates are sometimes wrongly built, because of a typo in the date information, and among millions of data, it is difficult to isolate these errors, whereas the conversion into APEnet EAD (which includes a normalisation of dates) combined with the functionality of refinement developed for the "List view" at the Archives Portal Europe allow to do so.

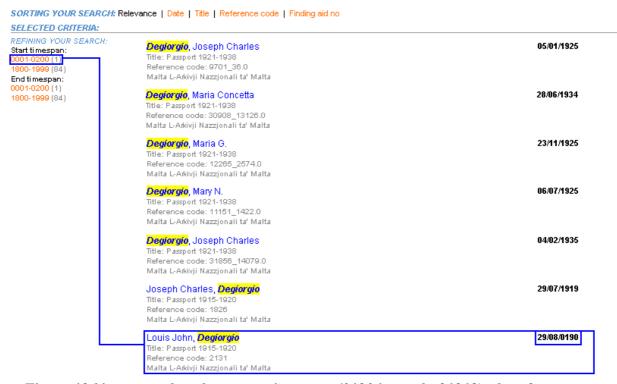


Figure 12 bis. wrong date due to a typing error (0190 instead of 1919), therefore easy to correct in the original data.

Sometimes the normalised dates can be misinterpreted when not in their original website context.





Figure 12 ter. Wrongly normalised dates due to the first part of the date being interpreted as the year instead of the day and month during normalisation. Solving the issue might take more time since it implies changes in the software used locally.

The dashboard and the Data Preparation Tool already offer some facility to improve this. But one challenge for the future will be to insert a conversion tool from textual dates to normalised dates, whenever it is possible to do so: for instance, 14th century could be turned into 1301/01-1400/12/31. But this means a rather huge work, that cannot completely be automated.

3.2.2. The title of the low levels of description:

As the archives are described hierarchically, they usually are following the principle of internally inheriting information from higher to lower levels. Therefore it often happens, for repetitive or serial documents, such as correspondence, church records, atlas, minutes etc., that the title is not repeated in the lowest level of description, because only the date and/or the place for instance are the necessary information for the user allowing him to select and understand the document.

Generally, the descriptions fit a particular web environment developed on purpose. Once in the Archives Portal Europe, with other rules to display the archival description, this becomes a challenge: the choice has been made to display the default text "No title" where there is no title, in order to avoid blank clickable lists, and following this something usual seems to appear as an error.



Figure 13. Death and burial records of Sweden - the only interesting information at the low level is the date that changes from one descriptive unit to the next and there is no need to repeat everything of the upper levels (place and type of document that can be seen on the left)



This issue concerns all partners, and we could not find any completely satisfactory solution during the lifetime of the project. This has to be solved in the future at the lowest possible cost: we cannot ask our partners to redo for instance their database simply for the sake of a central presentation.

3.2.3. The thumbnails:

Some institutions can provide thumbnails, while others cannot, for technical reasons, or do not like to due to the type of digitised documents they have. Therefore, the display for indicating digitised documents in the Archives Portal Europe can be seen as not harmonised with this regard (see figures 1 to 3 above).

The project has to find an easy way to create/store thumbnails for the partners that do not have them themselves, but would like to include them in the Archives Portal Europe. Some ideas have already been explored, but were discarded so far because their implementation would have been too time-consuming in the course of this project.

3.2.4. The links:

Lots of different links can be found within the finding aids and holdings guides. The most important ones are the ones linking back from a search results to the original display in the websites of the partner. The difficulty is that the links lead to different places: either a finding aid, either a record, either a digitised document... It is not always possible for the partners to provide direct links to the precise item needed, because of security rules of their host for instance. Besides, links have a tendency to change and very few of the partners at the moment are using permanent links. Therefore, it might happen that the user comes across broken links and cannot access, for instance, to the digitised images he wanted to see. This situation is a rather common one within big portals dealing with millions of items. No doubt that in the future, the use of the permanent links will be the rule and will help to solve these problems.

4. Conclusion

In conclusion, all the raised issues will be addressed in the future, either by improving the original data or the export of it, or by improving the conversion to the common European format, or by improving the display and layout of the Archives Portal Europe. Most probably, a combination of these different possibilities will be the solution, adapted to the different questions, and in close cooperation with each institution thanks to a permanent and fruitful dialogue that has permitted during this project to solve all major issues.