



Technical Report: Archival Digital Object Ingestion into Europeana (ESE-EAD harmonisation)

Version 1.0, 07/08/2009

Europeana v1.0

Author(s)

Go Sugimoto (Europeana)

*Wim van Dongen (The National Archives of the
Netherlands, APEnet)*



co-funded by the European Union

The project is co-funded by the European Union, through the **eContentplus** programme
<http://ec.europa.eu/econtentplus>

Table of Contents

1. Introduction.....	3
2. Background	3
2.1 EAD and Digital Objects	3
2.2 Data sharing and central indexing in Europeana.....	4
3. EAD formalisation and mapping.....	6
3.1 EAD and Europeana interaction from technical point of view	6
3.2 EAD and Europeana from user's point of view	7
4. Mapping practice	10
4.1 EAD elements and the mapping principles.....	10
4.2 Example.....	16
5. Conclusions.....	19
6. Document History	19
Acknowledgement	19
Appendix 1. Relation between Europeana and APENet (Archival Portal Europe).....	20

1. Introduction

Europeana (<http://www.europeana.eu/>) has worked closely with archivists across Europe in order to try to integrate EAD (Encoded Archival Description)¹ data with ESE (Europeana Semantic Elements) Ver3.1 which is used for the Europeana Prototype launched in November 2008. This technical report is an evaluation of the EAD-ESE experiments based on this prototype and is particularly designed for those who use EAD as a metadata set of digital objects that will be contributed to Europeana (i.e. archivists) to review. It describes a methodology of formalising EAD and mapping it to ESE, therefore, ensuring full compatibility with ESE Ver3.1 (Now Ver3.2). This document is a result of a-year-collaboration between the Europeana team and several European archivists. In particular, the collaboration has started since a meeting was held at the Bundesarchiv in Berlin (2008-04-18) with the archivists from major national archives in Europe. In terms of the relationship between Europeana and the APENet project (<http://www.apenet.eu>), please see Appendix 1.

2. Background

The inclusion of EAD metadata in Europeana is challenging due to the particular characteristics of the schema. While the ESE is based on Dublin Core and is an item-centric model focusing on the description of a digital item/object, EAD is a hierarchical model containing descriptions of digital and non-digital items/objects (archival material), as well as their contextual information. In order to harmonise these two different models, it is necessary to find the linkage point.

2.1 EAD and Digital Objects

Basically there are two ways for linking digital objects to EAD files: in case only one object must be linked to a descriptive unit the EAD element <dao/> is used or <daoloc/> within <daogrp/> and in case digitising a descriptive unit results in a lot of digital objects, another schema (METS² etc) is used.

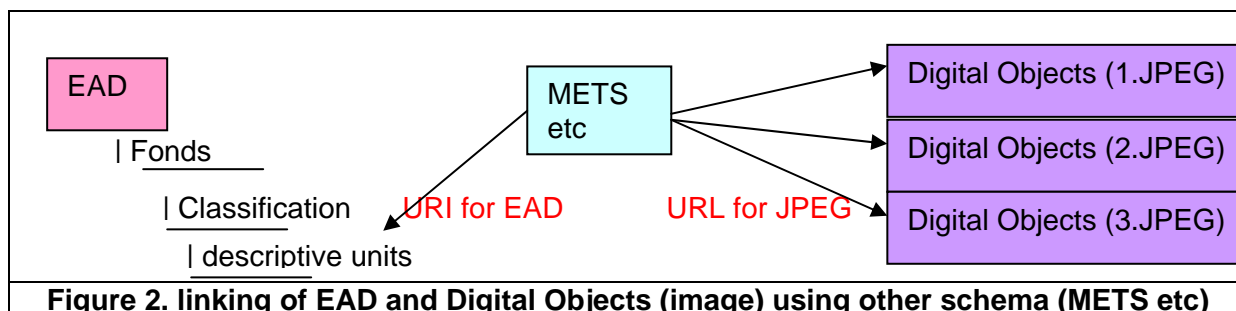
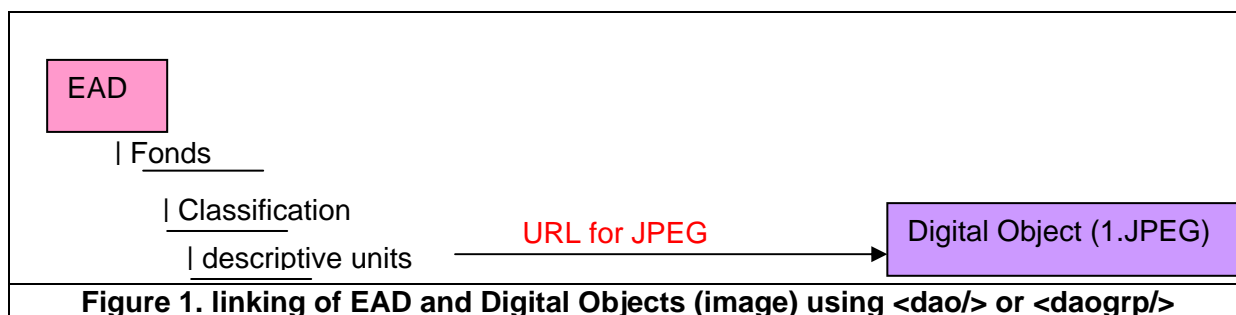
Figure 1 shows the linking situation in case <dao/> or <daoloc/> within <daogrp/> is used and Figure 2 shows the linking situation between EAD and digital objects using other schema. The use of METS is still quite limited to some archives.

The dilemma of using EAD as a metadata set for ingesting archival material into Europeana basically originates from Europeana's policy to focus on digital objects and the present state of digitisation within the archives sector. Whilst Europeana deals only with digital objects which are playable or viewable as computer files (JPEG, PDF, MP3, AVI etc), most archives do not hold many digital objects yet, but do hold a lot of contextual information (EAD) about archival resources. As such, dumping EAD in Europeana's repository with ESE is not possible because the users may find EAD information without a digital object when querying. For some national archives the percentage of digital objects in EAD is fairly low, resulting in the disappointment of the users who expect the same amount of digital objects stored in libraries and museums.

However, it should be noted that Europeana Ver1.0 project (February 2009-July 2011) and the APENet (January 2009-December 2011) currently discuss the possibility to include EAD files as digital objects in their own rights. Therefore, the integration methodology described in this report is specific to the current application.

¹ <http://www.loc.gov/ead/>

² <http://www.loc.gov/standards/mets/>



2.2 Data sharing and central indexing in Europeana

Europeana development team harvests metadata from content providers preferably via OAI-PMH³ in XML format. Data sharing in Europeana is driven by mapping local schemas to the ESE in the central repository. All metadata in the repository are indexed and searchable. Figure 3 represents the basic mechanism of the Europeana Prototype for ingesting, mapping, indexing and searching. There are five steps from data ingestion to querying.

Steps:

- 1) Content provider delivers metadata in original schemas
- 2) Europeana development team or the content provider maps the metadata to ESE
- 3) Europeana development team converts the ESE to indexing format
- 4) Europeana indexes all ESE fields in the index engine
- 5) The users search all ESE fields by simple search and selected ESE fields by advanced search.

This is a pragmatic method in order for us not to force the content providers to change their metadata formats. The ESE should be generic to encompass various types of metadata from archives, libraries and museums. Such uniform schema ensures the interoperability across different domains.

³ <http://www.openarchives.org/pmh/>

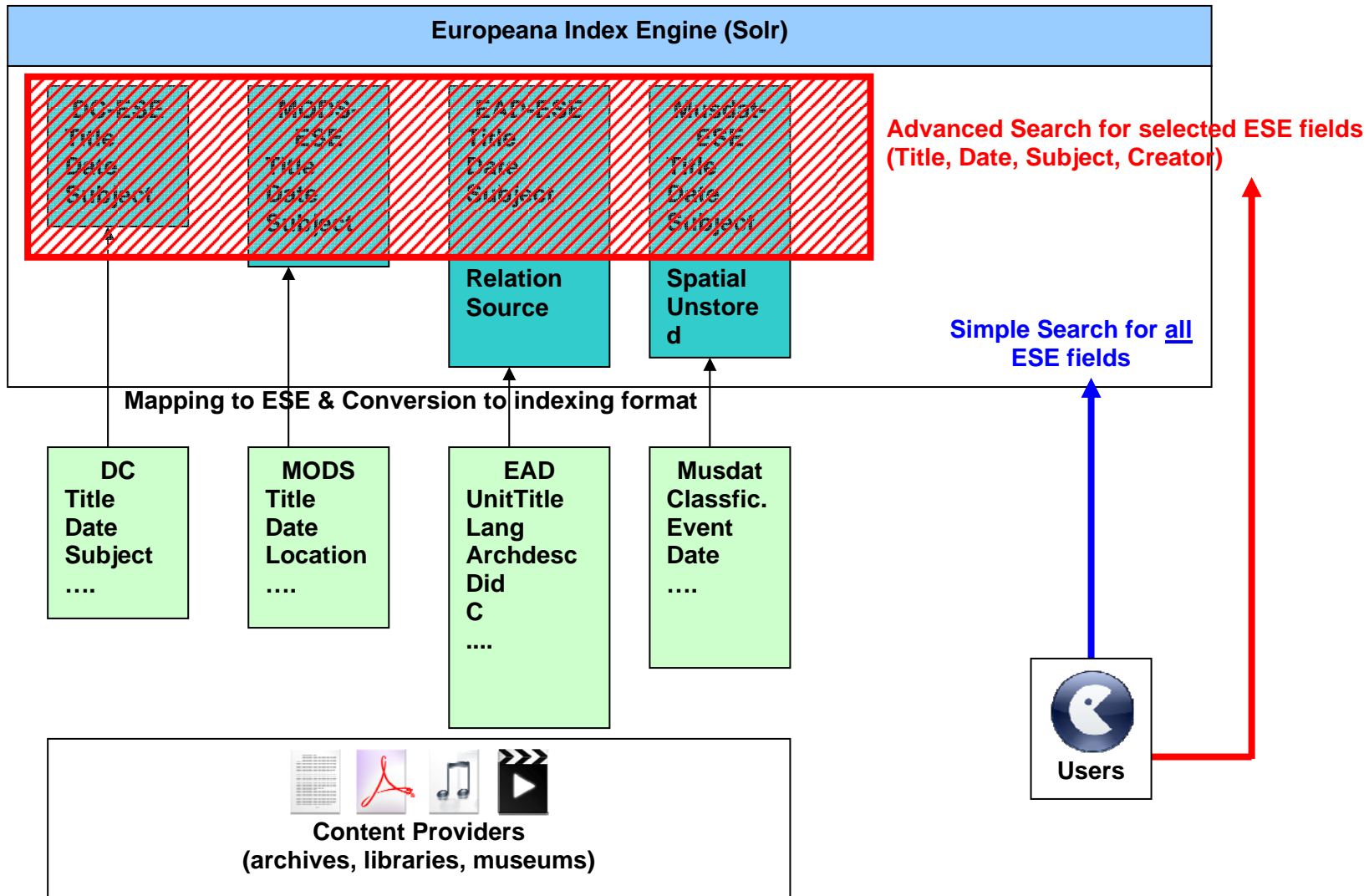


Figure3. Europeana indexing and searching scenario

3. EAD formalisation and mapping

3.1 EAD and Europeana interaction from technical point of view

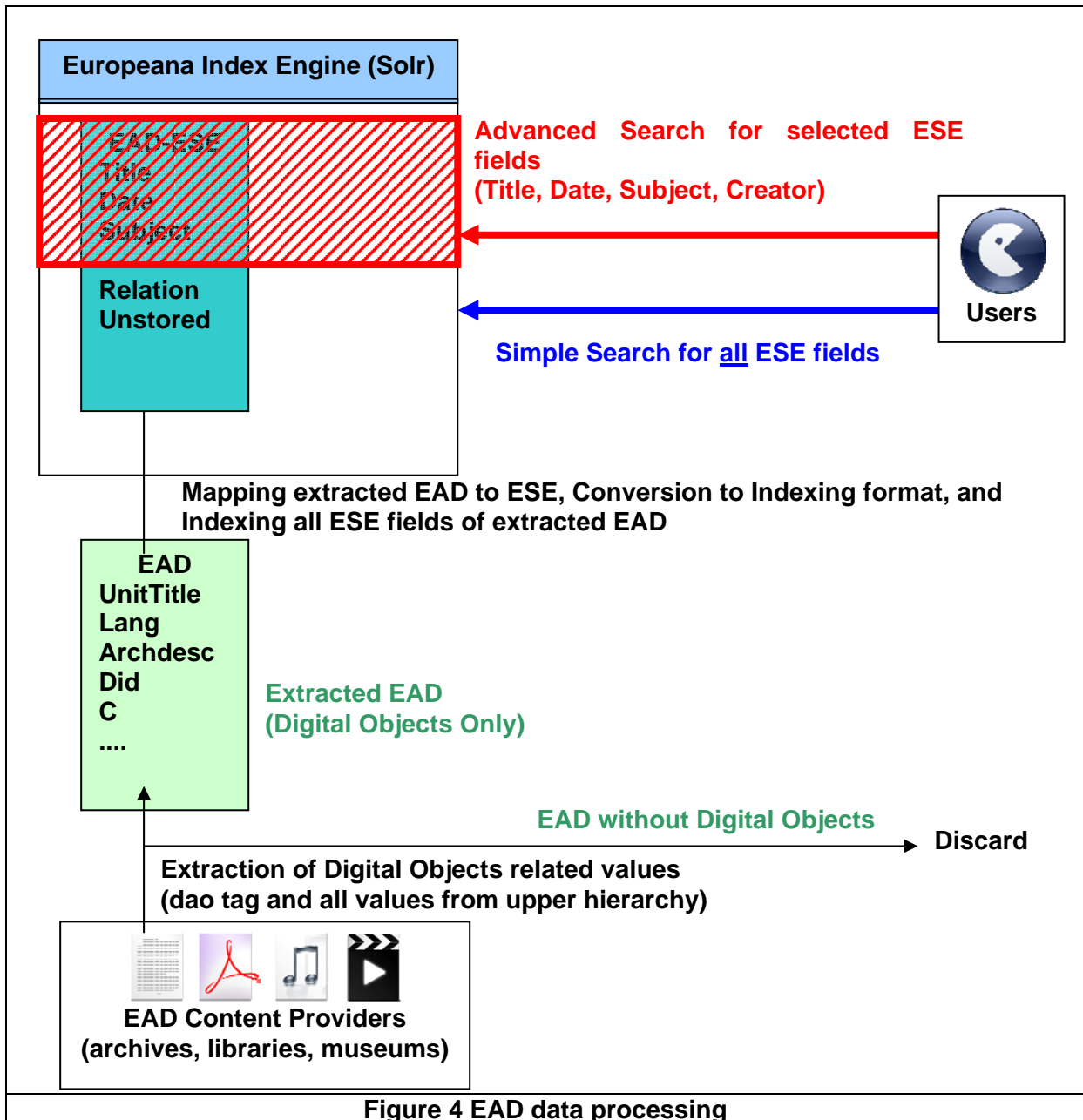


Figure 4 EAD data processing

As for EAD, there are six steps.

Steps:

- 1) Content provider delivers EAD in original (including information on non-digital objects) preferably via OAI-PMH.
- 2) Content provider or Europeana development team extracts the values only related to digital objects. (Extract all values from lowest hierarchical level of <dao/> or <daoloc/> within <daogrp/> up till upper hierarchy using XSLT etc).
- 3) Europeana development team or content provider maps the extracted EAD to ESE.
- 4) Europeana development team indexes ESE in the index engine
- 5) The users search all ESE fields by simple search and the selected ESE fields by advanced search.

As for the connection between ESE and Digital Objects (and EAD), the hyperlink from ESE to Digital Objects will be conducted by URL described in <dao/> or <daoloc/> within <daogrp/>. Otherwise the URL will be re-constructed by external files (METS etc). Similarly, the link to a particular position of EAD will be carried out by the <unitid/> connected to <dao/>. Therefore, content providers are requested to inform the Europeana development team how to re-construct the complete path of URL from fragments of URIs. For instance, the URL of a particular position of EAD (a <dao/> element) is <http://europearchive.org/collection/ead/eadid=123/> in front of <unitid/> in <dao/>.

Europeana offers the content providers to create two types of hyperlinks by using two elements in ESE; *europeana:isShownBy* and *europeana:isShownAt*. The former is a direct link to a best-quality digital object on the content provider's website where they can view or play it, whilst the latter is the link to the digital object on the content provider's website in its full information context. This distinction allows the content provider to offer two channels of presenting digital object: 1) quick link to a digital object (*isShownBy*) and 2) link to EAD contextual information of the digital object (*isShownAt*).

Some may think it is very likely that some EAD tags cannot be mapped to Dublin Core-based schema, however, ESE Ver3.1 (Now Ver3.2) has an element called *europeana:unstored* which is used for any valuable information not possible to map to any other ESE elements. As the Europeana result page does not show this element, it is only useful for indexing purposes.

3.2 EAD and Europeana from user's point of view

If a user makes a query in the Europeana Prototype, the used keywords will appear in Europeana's metadata display page (ESE) (Figure 5). The only exception is the values in *europeana:unstored* which are not shown. Then the user can move on to the best quality digital object via the hyperlink provided (*europeana:isShownBy*) with one click (Figure 6). At the same time, if contextual information is available, the user can follow the link ("View in original context") to see the information associated with the digital objects (*europeana:isShownAt*). This can be an EAD finding aid display page on the content provider's websites (Figure 7).

Europeana - Search results - Mozilla Firefox

http://www.europeana.eu/portal/full-doc.html?query=europeana_colli

Search

Advanced search

Matches for: europeana_collectionName:007^

Item details

Return to results

Brech : Pfahlbronn, Alfdorf WN

Title: Brech : Pfahlbronn, Alfdorf WN

Description: Kontext: Kieser-Ortsansichten (Inventar)>> 1. Ortsansichten >> 1.2. Anfangsbuchstabe B

Source: http://www.landesarchiv-bw.de/ntas Landesarchiv Baden-Württemberg: Hauptstaatsarchiv Stuttgart

Provider: Landesarchiv Baden-Württemberg ; germany

Identifier: DE_ArchLABW_1_513102; H 107/15 Bd 7 Bl. 27

Subject: 1685; Brech; Pfahlbronn, Alfdorf, WN

Relation: http://www.landesarchiv-bw.de/plink/?f=1-509056 Kieser-Ortsansichten (Inventar)

Link to Digital Object (isShownBy)

Link to Digital Object in context (isShownAt)

[View in original context](#)
Opens in a new window

Figure 5 Europeana's metadata display page

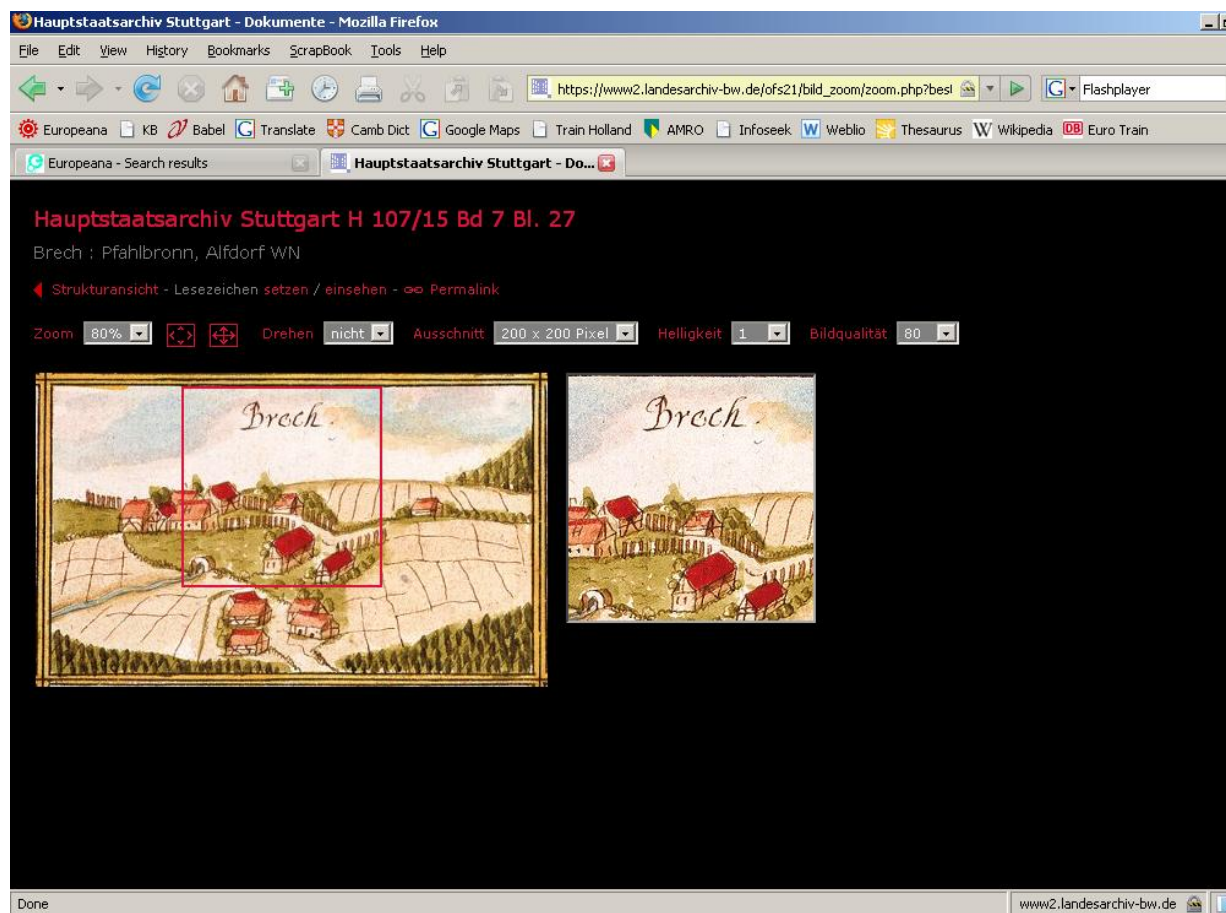


Figure 6. Digital object at content provider's website (when click the thumbnail (isShownBy) in Figure 5)

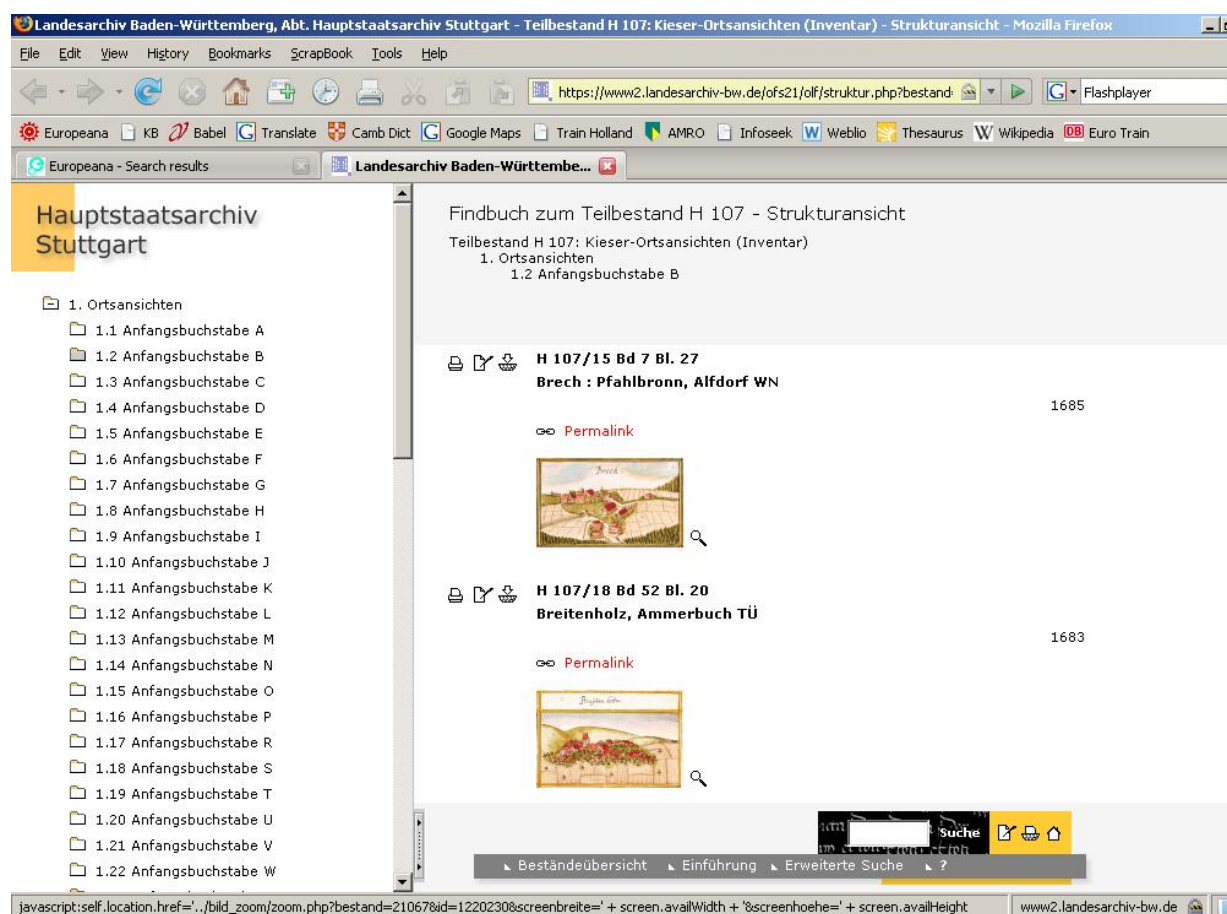


Figure 7. The exact position of EAD for the digital object at content provider’s website (when click “View in original context” (isShownAt) in Figure 5)

4. Mapping practice

4.1 EAD elements and the mapping principles

This chapter gives an instruction for the mapping practice in principle. We identified the most important EAD tags to be used for Europeana (Table 1). The following list of principles should be respected while mapping those tags.

- 1) The title or description of the item/digital object (<unittitle> of <daogrp>) should be mapped to **dc:title**
- 2) All <unittitle>(s) higher in the EAD hierarchy and related to the <unittitle>(s) as mentioned under 1) are mapped to **dcterms:isPartOf**
- 3) **dc:contributor** is used for the person/organization related to the archive (e.g. <corpname>, <famname>, <persname> etc within <origination>), while **europena:provider** is used for the name of the content provider (e.g. <publisher> etc).
- 4) EAD identifier (<eadid>) and <unitid> of <daogrp> are preserved as **dc:identifier**, whilst **europena:uri** is the unique identifier of the Europeana system assigned by the central office (based on the record identifiers of the record).

- 5) Item/digital object level <unitdate> is used for **europeana:year** (Date facet and TimeLine function⁴).
- 6) All other item/digital object related <unitdate>s (in the higher hierarchy) are mapped to **dc:date**, so the user can find these related dates, but not in the Date facet.
- 7) URL and/or URI described in <dao> should be mapped to **europeana:isShownBy** or **europeana:isShownAt** depending on the type of URL/URI target. If the URL/URI of europeana:isShownBy is actually an digital object which can generate thumbnails (such as JPEG and PDF), it should be also mapped to **europeana:object**
- 8) **dc:language** should be used for the language of the content (item/digital object) not the language in which the EAD finding aid is written (e.g. <language> under <langmaterial> etc).
- 9) <subject> should be mapped to **dc:subject**, whereas <geoname> should be **dcterms:spatial**
- 10) Any important information at any level which is associated with an item/digital object can be dumped into **dc:description**.
- 11) Any important information at any level which is associated with an item/digital object can be dumped into **europeana:unstored** only for indexing purposes. The values of this element will not be displayed in Europeana.

These are the basic principles in order to preserve the semantics of EAD as much as possible in the framework of ESE Ver3.1 (and Ver3.2), therefore, the mapping of each case should be more flexible. For example, an archive institution prefers to use dc:description for the mapping of higher level item/digital object related <unittitle>(s), while another institution prefers to dump those in dcterms:isPartOf. This is basically a matter of how (and where) a content provider wants this specific EAD context information to be displayed on the Europeana results page. However, **it is highly recommended to follow the principle as much as possible**, because it will improve the query performance of Europeana content across EAD-based metadata and it is obvious that data inconsistency will reduce the possibility to find information in a standardized way. Other major EAD elements not described in this mapping principle can be discussed and agreed upon in the archival community, for instance, via the APENet project

Presentation of hierarchy

As the ESE is an item/object-centric schema and Europeana displays the metadata of digital objects in a flat structure, at the moment the rich structural semantics of EAD cannot be presented as they are. However, Europeana recommends to display hierarchical structure by adding XML attributes to ESE elements (i.e. dcterms:isPartOf). For example, please look at the following part of an EAD finding aid with 'level'-attributes at <c>-level (if such XML attributes are not available, Europeana will use default prefixes):

⁴ <http://www.europeana.eu/portal/year-grid.html>

```

<c01 level="series">
  <did>
    <unittitle>Title A</unittitle>
  </did>
  <c02 level="subseries">
    <did>
      <unittitle>Title B</unittitle>
    </did>
    <c03 level="file">
      <did>
        <unittitle>Title C</unittitle>
      </did>
    </c03>
  </c02>
</c01>

```

This EAD finding aid part can be converted into ESE as follows:

```

<dc:title>Title C</dc:title>
<dcterms:isPartOf> subseries: Title B</dcterms:isPartOf>
<dcterms:isPartOf> series: Title A</dcterms:isPartOf>

```

The prefix “subseries” and “series” define the distinction between the different levels of hierarchy.

Nesting tags

Sometimes tags are embedded in EAD elements and they have to be mapped in the intended way. The recommended mapping is to extract all relevant nesting tags and map them separately to target elements, while the wrapper tag should be also mapped as a whole by eliminating nesting tags. For example;

```

<p>Tafel der watergeteyen - in de <geogname>Maze</geogname> - in de
<geogname>Goedereede</geogname> / getekend en gegraveerd door <persname>J. C.
Philips</persname> [tussen1742 voor 1775]</p>

```

should become

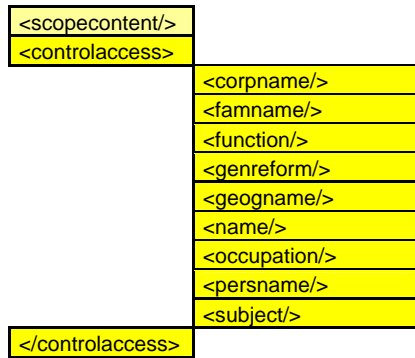
```

<dcterms:spatial>Maze</ dcterms:spatial>
<dcterms:spatial>Goedereede</ dcterms:spatial>
<dc:contributor> J. C. Philips <dc:contributor>
<dc:description>Tafel der watergeteyen - in de Maze - in de Goedereede / getekend en
gegraveerd door J. C. Philips [tussen1742 voor 1775]</dc:description>

```

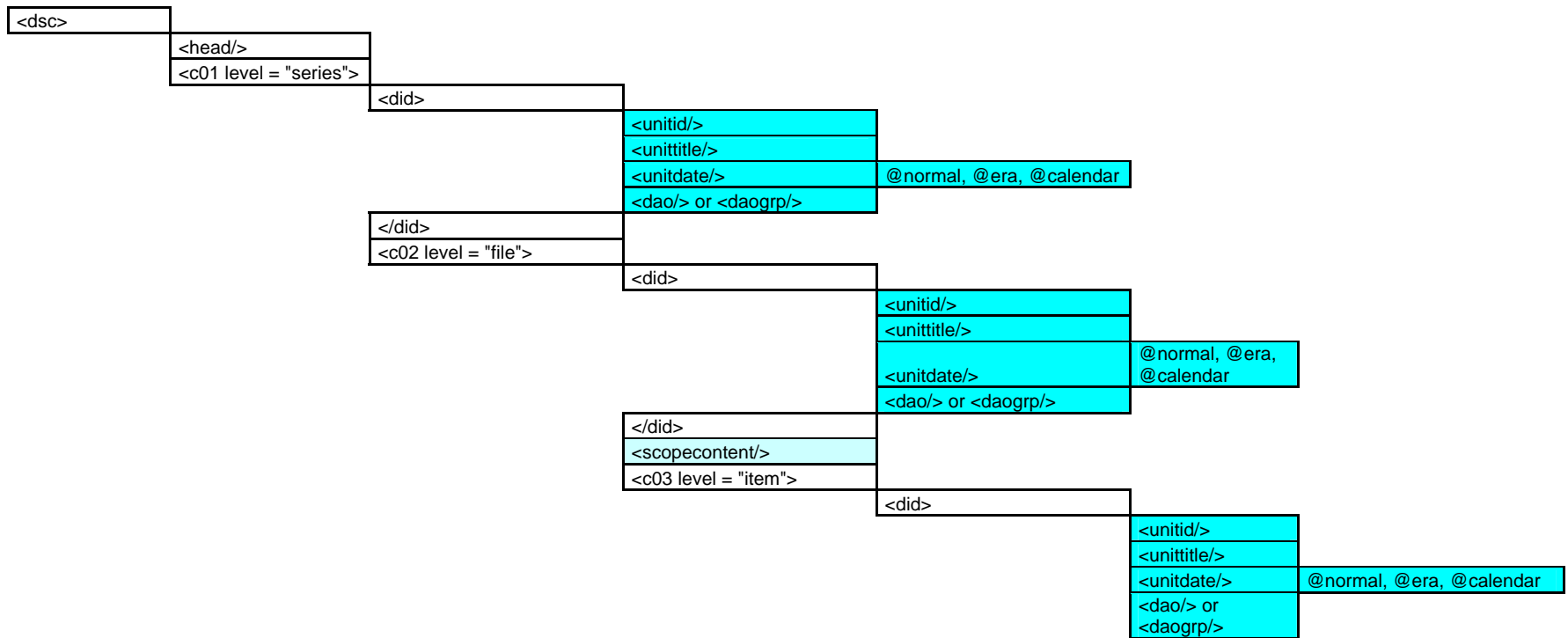
Table 1. EAD elements suitable for mapping to ESE

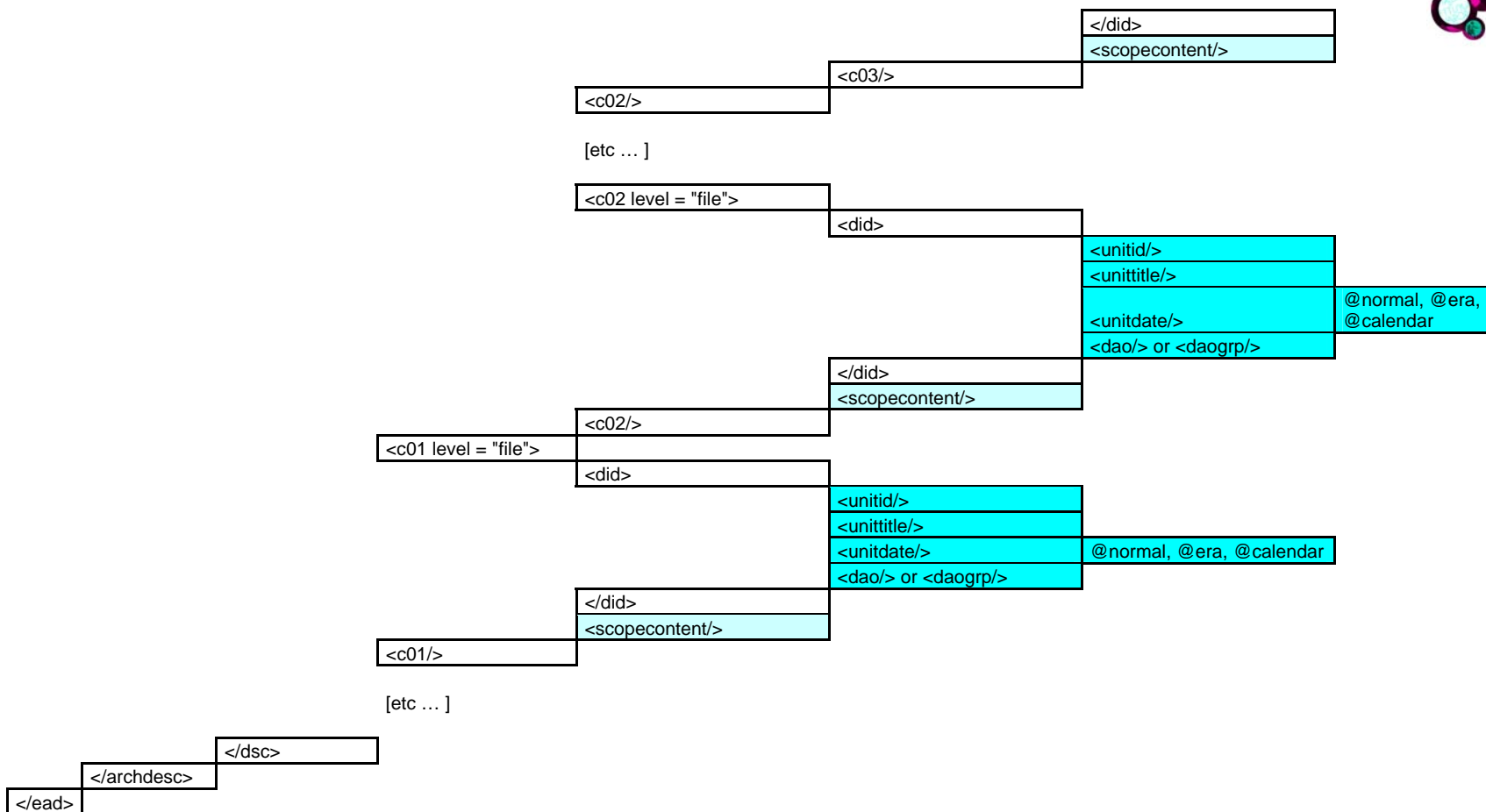
<ead>						
1	<eadheader>	@langencoding	@relatedencoding	@repositoryencoding	@countryencoding	@dateencoding
	<eadid/>	@countrycode	@mainagencycode	@url		
	<filedesc>					
	<titlestmt>					
		<titleproper/>				
		<author/>				
		</titlestmt>				
		<publicationstmt>				
		<publisher/>				
		<date/>	@normal, @era, @calendar			
		</publicationstmt>				
		</filedesc>				
		<profiledesc>				
		<language>				
	<language/>					
	</language>					
	</profiledesc>					
	</eadheader>					
2	<frontmatter>					
	</frontmatter>					
3	<archdesc>	@level	@type			
	<did>					
		<unittitle/>				
		<unitdate/>	@normal, @era, @calendar	@datechar	@type	@certainty
		<unitid/>	@countrycode	@repositorycode		
		<langmaterial				
		<language/>	@langcode	@scriptcode		
		</langmaterial>				
		<repository>				
		</repository>				
		<origination>				
		<corpname/>				
		<famname/>				
		<name/>				
	<persname/>					
	</origination>					
	<abstract/>					



high level EAD finding aid

low level EAD finding aid





Explanation of the use of the colors:
strong recommendation to deliver this element to Europeana (only useful once standardised)
to be discussed whether to deliver this element to Europeana (only useful once standardised)
basic set of low(est) level elements Europeana needs for describing digital archival objects, plus upper hierarchical context information

4.2 Example

An example of ESE data converted from EAD is given below.

```

<record>
<eadheader>
  <eadid>4.VTH</eadid>
  <filedesc>
    <titlestmt>
      <titleproper>Inventaris van de verzameling Binnenlandse Kaarten Hingman</titleproper>
      <author>J.H. Hingman</author>
    </titlestmt>
    <publicationstmt>
      <publisher>Nationaal Archief, Den Haag</publisher>
      <date normal="1871" era="ce" calendar="gregorian">(c) 1871</date>
    </publicationstmt>
  </filedesc>
  <profiledesc>
    <language>This finding aid is written in <language langcode="dut" scriptcode="Latn">Dutch</language>.</language>
    <describes audience="internal" />
  </profiledesc>
</eadheader>

<archdesc type="inventory" level="fonds">
<did>
  <head>Beschrijving van het archief</head>
  <unittitle label="Naam archiefblok: ">VTH Verzameling Binnenlandse Kaarten Hingman, 15e-19e eeuw</unittitle>
  <unittitle type="short">Kaarten Hingman</unittitle>
  <unitdate normal="1400/1900">merendeel 15e - 19e eeuw</unitdate>
  <unitid label="Archiefbloknnummer: " repositorycode="NL-HaNA" countrycode="NL">4129</unitid>
  <physdesc label="Omvang: ">
    <extent unit="files">5014 inventarisnummers</extent>
  </physdesc>
  <langmaterial label="Taal van het archiefmateriaal: ">Het merendeel der stukken is in het <language langcode="dut"
scriptcode="Latn">Nederlands</language>.</langmaterial>
  <repository label="Archiefbewaarpplaats: ">Nationaal Archief, Den Haag</repository>
  <physloc />

```



```

<origination label="Archiefvormers: ">
  <corpname>Agentschap van Inwendige Politie en Toezicht op de Staat van Dijken, Wegen en Wateren</corpname>
</origination>
  <abstract label="Samenvatting van de inhoud van het archief: ">De collectie bevat de unieke manuscript kaarten van Nederland
van de middeleeuwen tot halverwege de 19e eeuw..... </abstract>
</did>

<dsc type="combined">
<head>Beschrijving van de series en archiefbestanddelen</head>
<c01 level="series">
  <did>
    <unittitle>ATLASSEN EN KAARTBOEKEN</unittitle>
  </did>
  <c02 level="file">
    <did>
      <unitid>E</unitid>
      <unittitle>Kaartboek van den lande van <geogname>Voorne</geogname/>, gemaakt ingevolge resolutie van heeren breede
geerfden van 7 Juni 1695.</unittitle>
      <physdesc>Drie en dertig kaarten. Koperdruk. Gebonden</physdesc>
    </did>
    <daogrp linktype="extended">
      <resource linktype="resource" label="start" />
      <daoloc href="http://beeldbank.nationaalarchief.nl/na:col1:dat512752.jpg" linktype="locator" label="thumb" />
      <daoloc href="http://beeldbank.nationaalarchief.nl/na:col1:dat512752" linktype="locator" label="reference" />
      <arc linktype="arc" show="embed" actuate="onload" from="start" to="thumb" />
      <arc linktype="arc" show="new" actuate="onrequest" from="thumb" to="reference" />
    </daogrp>
    <odd>
      <p>Vervaardigd door <persname role="creator">Heyman van Dyck</persname>.</p>
    </odd>
  </c02>
</c01>
</dsc>
</archdesc>
</record>

```

EAD containing only Digital Object description (Simplified for display purpose)

```

<record>
<europeana:uri>http://www.europeana.eu/resolve/record/NL/DDFE163345D338193AC2BDC183F8E9DCFF904B43</europeana:uri>
<dc:identifïer>4.VTH</dc:identifïer>
<dc:identifïer>E</dc:identifïer>
<dc:source>Nationaal Archief, Den Haag</dc:source>
<dcterms:isPartOf>Inventaris van de verzameling Binnenlandse Kaarten Hingman</dcterms:isPartOf>
<dcterms:isPartOf>fonds: VTH Verzameling Binnenlandse Kaarten Hingman, 15e-19e eeuw</dcterms:isPartOf>
<dcterms:isPartOf>series: ATLASSEN EN KAARTBOEKEN</dcterms:isPartOf>
<dc:title>file: Kaartboek van den lande van Voorne, gemaakt ingevolge resolutie van heeren breede geerfdens van 7 Juni 1695.</dc:title>
<dcterms:spatial>Voorne</dcterms:spatial>
<dc:description>Vervaardigd door Heyman van Dyck.</dc:description>
<dc:contributor>Heyman van Dyck.</dc:contributor>
<europeana:hasObject>>true</europeana:hasObject>
<europeana:isShownBy>http://beeldbank.nationaalarchief.nl/na:col1:dat512752.jpg</europeana:isShownBy>
<europeana:object>http://beeldbank.nationaalarchief.nl/na:col1:dat512752.jpg</europeana:object>
<europeana:isShownAt>http://beeldbank.nationaalarchief.nl/na:col1:dat512752</europeana:isShownAt>
<europeana:language>nl</europeana:language>
<europeana:country>netherlands</europeana:country>
<europeana:provider>Nationaal Archief</europeana:provider>
<europeana:type>IMAGE</europeana:type>
</record>
    
```

ESE record to which EAD is mapped

5. Conclusions

The data integration methodology described in this document allows us to ingest only those parts of an EAD finding aid that are related to digital objects which Europeana would like to provide. The valuable semantics of an EAD finding aid can be preserved as much as possible by following the mapping principles. This report serves as a first consulting point for archival community to understand the basic principles of integrating archival objects in EAD with ESE in Europeana.

6. Document History

Version	Author	Updated	Comments
V0.5	Go Sugimoto	2009-02-13	Draft creation
V0.6	Wim van Dongen	2009-02-26	Revision
V0.7	Go Sugimoto	2009-04-08	Final draft for review
V0.8	Go Sugimoto	2009-04-23	Title change and minor changes
V0.9	Go Sugimoto	2009-05-19	Some feedback and questions
V1.0	Go Sugimoto Wim van Dongen	2009-08-07	Inclusion of nesting issues, change of example, and other minor changes.

Acknowledgement

Thanks are due to the following for the participation in the Berlin meeting and useful feedback on this report.

Andes Imhof (Bundesarchiv)
 Angelika Menne-Haritz (Bundesarchiv)
 Börje Justrell (Swedish National Archives)
 Catherine Lupovici (Europeana)
 Christina Wolf (Landesarchiv Baden-Württemberg)
 Gerald Maier (Landesarchiv Baden-Württemberg)
 Göran Kristiansson (Swedish National Archives)
 Henny van Schie (The National Archives of the Netherlands)
 Karin Bredenberg (Swedish National Archives)
 Lucile Grand (Direction des Archives de France)
 Luis Ensenat (Spanish Archives)
 Markus Merenmies (Finnish National Archives)
 Martin Bjersby (Swedish National Archives)
 Olivier de Solan (Archives départementales de la Somme)
 Peder Andrén (Swedish National Archives)
 Stefan Gradmann (Humboldt Universität zu Berlin)
 Thomas Fricke (Landesarchiv Baden-Württemberg)

If you have any enquiries on the report, please feel free to contact us.

Contact

Go Sugimoto (Europeana)

go.sugimoto@kb.nl

Wim van Dongen (The National Archives of the Netherlands / APEnet project)

wim.van.dongen@nationalearchief.nl

Appendix 1. Relation between Europeana and APENet (Archival Portal Europe)

It is important to understand the difference and relationship between the two projects. The APENet project intends to develop an ideal European archives portal which will present information on all archival objects, whether digitized or not, including the presentation of all digital objects available, whereas Europeana will demonstrate the potential of integrating digitized archival material with resources from other domains. The cooperation between Europeana and the APENet project will provide Europeana a facility to access and harvest the APENet portal resources

About APENet:

Archives are a society's memory but also its treasure chamber that can enhance the quality of life for any citizen, on a personal as well as a societal level. Providing all Europeans with easy online access to the content of the collective memory contained in archives, will be an important contribution to the realisation of European social and cultural objectives.

With the rapidly growing number of archives services using the Internet for the dissemination of their holdings and their growing competence in using the new technologies, the creation of an European Archives Internet Gateway ("**Internet Gateway for Documents and Archives in Europe**") which pools efforts and expertise at the national and European levels, has been defined as a top priority by European professionals, and endorsed by the European Council in its recommendation of 14 November 2005 (Official Journal of the European Union, 29.11.2005; 2005/535/EC).

The fundamental aim of the APENet project is to provide EU citizens, public authorities and companies with a common gateway, which will enable them to find out for themselves about Europe's archives and about archival material that is relevant to them, whether or not they live in the European Union. Archives across the European Union will be easily accessible to EU citizens for the first time from a single entry point, and should improve public understanding of European history and culture. Therefore the Archival Portal Europe will also have a strong relationship with the Europeana and other European portals or gateways which hold information on archives.

The gateway will connect disparate archival Internet initiatives and will act as a powerful catalyst in breaking down the remaining barriers to easy access to information about archives in the European Union and the important cultural resources which they hold, making them available through a virtual catalogue. The services of the gateway will be available in different languages.

More information about the APENet project can be found on <http://www.apenet.eu>