

Ingestion of Metadata into Europeana

Update May 2011 for BHL-Europe
Gerda Koch (AIT)

1 The ingestion process from BHL-Europe to Europeana

Once the BHL-Europe content partners have uploaded their metadata and image files into the FTP directories of the London Natural History Museum's server the ingestions process to BHL-Europe and also to Europeana starts.

The content providers acknowledge their finalization of upload based on the Pre-Ingest guidelines to the WP2 lead. Next a schema mapping tool converts all metadata supplied in the MARC format (MACHINE-Readable Cataloging, <http://www.loc.gov/marc/>) into the BHL-Europe METS (Metadata Encoding and Transmission Standard, <http://www.loc.gov/standards/mets/>) format. The descriptive metadata of the BHL-Europe METS file is transformed into MODS (Metadata Object Description standard, <http://www.loc.gov/standards/mods/>).

Once these mapped files are available on the NHM FTP server the ingest process to Europeana is initiated.

In this phase of the project whilst the final BHL-Europe portal is still under development the established OAI-provider services of the BHL-Europe prototype are used for the direct provision of data to Europeana.

The METS files are mapped to the Europeana ESE standard (www.europeana.eu) which is the current data standard for Europeana and which is also integral part of the new Europeana Data Model (EDM). For the mapping process the well-established open source business intelligence suite Pentaho (<http://www.pentaho.com/>) is applied.

The toolset transforms the METS files into valid ESE files which may look like this:

```
<europeana:record xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xsi:schemaLocation="http://www.europeana.eu/schemas/ese/
http://www.europeana.eu/schemas/ese/ESE-V3.3.xsd">
  <dc:title>Notes from the Royal Botanic Garden Edinburgh: Volume 6, addition to numbers
  XXIX-XXX (January 1917) (part name)</dc:title>
  <dc:publisher>HMSO (Edinburgh)</dc:publisher>
  <dc:type>Text</dc:type>
  <dc:format>print</dc:format>
  <dc:identifier>d2bd6a0e-d46a-4f0c-9486-c6060235b201</dc:identifier>
  <dc:identifier>0080-4274</dc:identifier>
  <dc:identifier>26774</dc:identifier>
  <dc:language>eng</dc:language>
  <dc:relation>Edinburgh Journal of Botany</dc:relation>
  <dcterms:tableOfContents>Plates I-XXXVII illustrating Mr Takeda's paper on the Genus
  Mahonia.</dcterms:tableOfContents>
  <dcterms:issued>1921</dcterms:issued>
  <europeana:object>http://bhle-dev-1.nhm.ac.uk/aip-preview/uk-rbge/NRBGEv1-
  23/Volume06/NRBGE_0006/NRBGE_0006_1917_0029-30_PLATES/thumbnail.jpg</europeana:object>
  <europeana:provider>BHL-Europe </europeana:provider>
  <europeana:type>TEXT</europeana:type>
  <europeana:rights>http://www.europeana.eu/rights/rr-f/</europeana:rights>
  <europeana:dataProvider>Royal Botanic Garden Edinburgh; United
  Kingdom</europeana:dataProvider>
  <europeana:isShownBy>http://bhle-dev-1.nhm.ac.uk/aip-preview/uk-rbge/NRBGEv1-
  23/Volume06/NRBGE_0006/NRBGE_0006_1917_0029-30_PLATES</europeana:isShownBy>
  <europeana:isShownAt>http://bhl.ait.co.at/index.php?form=display&amp;oiid=RBGE/D2BD6A0ED46A4F
  0C9486C6060235B201</europeana:isShownAt>
</europeana:record>
```

Figure 1-1: Europeana ESE data

1.1 The Europeana fields

<europeana : isShownBy>

This field must contain the link to the digital object.

At this stage of the project a first book viewer version was established at the NHM server using the Open Library BookReader developed by InternetArchive. The link to this online viewer is added in the respective Europeana field.

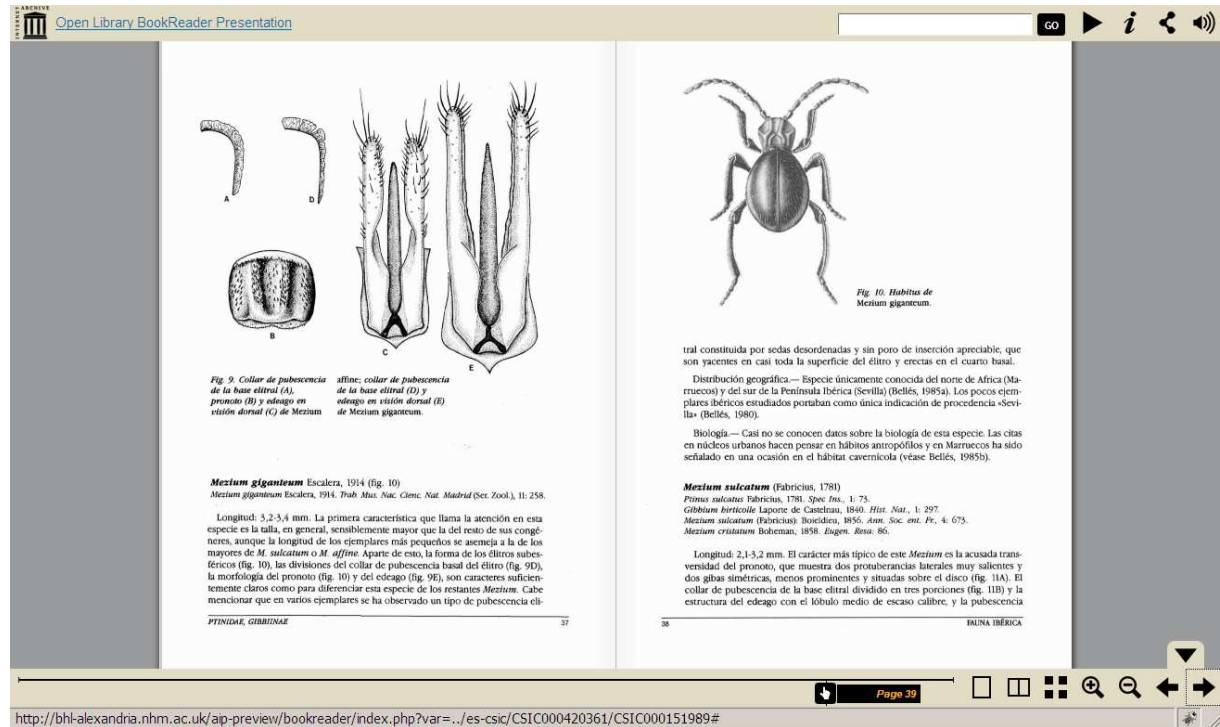


Figure 1-2: BookReader View of BHL-Europe content

<europeana : isShownAt>

This field will in future link directly to Website of the final BHL-Europe portal where the book is described.

The portal is still under development, but this fact was no obstacle to the 2010/11 Europeana ingests as in the very first ingestion phase partners with own online representations of their content (BHL-US, NAT, LANDOE, UBER, UBBI, UCPH) were selected. And for the partners of the second major ingestion phase (RMCA, RBGE, UHVIKKI, CSIC) the display of the BHL-Europe Prototype where the OAI provider for Europeana is located was used.

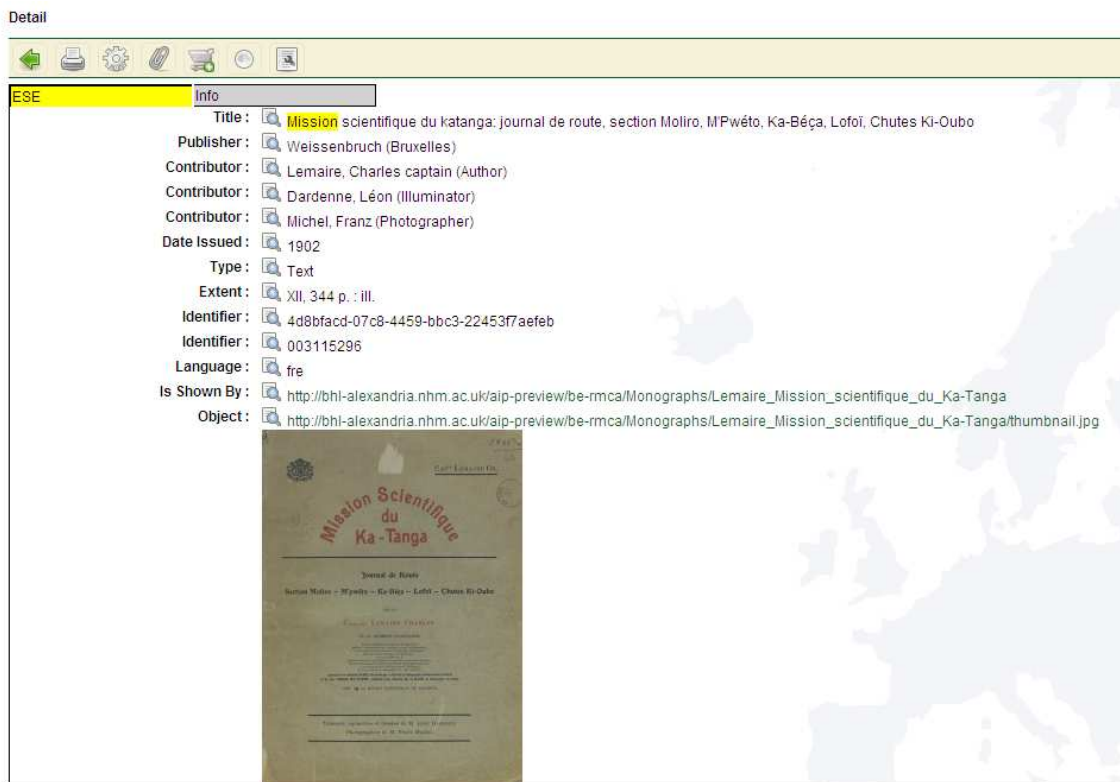


Figure 1-3: Detail view in the BHL-Europe prototype

<europeana:dataProvider>

The value of this field displays the official name of the content delivering institution and is depicted in the Europeana result list display above the name of the Europeana aggregator. The institutions need to acknowledge to WP3 which version of their name they want to use for Europeana. The value of this field is automatically inserted during the ingest to the Europeana OAI provider.

!!! This field is mandatory in ESEv3.4 !!!

<europeana:provider>

The value of this field shows the name of the institution, project or aggregator that directly supplies data to Europeana. The value "BHL-Europe" is added automatically.




Figure 1-4: Display of the europeana:dataProvider field and the europeana:provider field


<europeana:rights>

With the release of the ESE standard version 3.3 (autumn 2010) the data field europeana:rights was introduced. This field includes a statement about the rights status of the digital objects described in the metadata submitted to Europeana. The value of the field must be an URL linking to a license statement. The license is added to the content automatically if it applies to all content/records of the partner. If the partner wants to submit records with different license URLs the information has to be within the metadata.





Item details

[Return to results](#) share

 **Flora of Bhutan including a record of plants from Sikkim and Darjeeling: Volume 2, Part 1 (part number)**


Title:  Flora of Bhutan including a record of plants from Sikkim and Darjeeling: Volume 2, Part 1 (part number)


Date: 1991

Creator:  Bates, M. (Illus) ;  Grierson, Andrew John Charles 1929-1990 ;  Long, David Geoffrey ;  Rodrigues, G. (Illus)

Format: 1-426p. ill., figs., map, bibl, indices, PB, 20cm


Europeana rights: [Rights Reserved - Free Access](#)



Data provider:  Royal Botanic Garden Edinburgh; United Kingdom

Provider:  BHL Europe


[Less](#)

Identifier: 9b30c3b7-e6c1-4fca-82e4-66b3e62024fb ; 1872291023 ; 51159

Publisher:  Royal Botanic Garden, Edinburgh (Edinburgh)

Subject:  FH\qrh ;  Floras

Geographic coverage: Bhutan
Sikkim
Darjeeling

Type:  Text

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

Figure 1-5: Display of the europeana:rights field together with the license icon

Enter here the URL either for the Europeana rights or the Creative Commons statement:

EUROPEANA:

Public Domain > <http://creativecommons.org/publicdomain/mark/1.0/>

Rights Reserved - Free Access enter > <http://www.europeana.eu/rights/rr-f/>

Rights Reserved - Paid Access enter > <http://www.europeana.eu/rights/rr-p/>

Rights Reserved - Restricted Access enter > <http://www.europeana.eu/rights/rr-r/>

Unknown enter > <http://www.europeana.eu/rights/unknown/>

CC EXAMPLES:

CC – Zero (universal) enter > <http://creativecommons.org/publicdomain/zero/1.0/>

CC BY (v3.0 Unsorted) enter > <http://creativecommons.org/licenses/by/3.0/>

CC BY-SA (v3.0 Netherlands) enter > <http://creativecommons.org/licenses/by-sa/3.0/nl/>

CC BY-NC (v2.5 Germany) enter > <http://creativecommons.org/licenses/by-nc/2.5/de/>

CC BY-NC-SA (v2.0 France) enter > <http://creativecommons.org/licenses/by-nc-sa/2.0/fr/>

CC BY-ND (v2.0 in Spain) enter > <http://creativecommons.org/licenses/by-nd/2.0/es/>

CC BY-NC-ND (v1.0 in Finland) enter > <http://creativecommons.org/licenses/by-nc-nd/1.0/fi/>

!!! This field is mandatory in ESEv3.4 !!!

<europeana:object>

This field provides for Europeana the link to the thumbnail for the record. It was suggested during a project meeting to use the image qualified as "title" for that purpose. Whenever there was no image with "title" in the filename the image with "cover" in the filename was used as thumbnail image.

During the Pre-Ingest process the thumbnails for the records are generated.



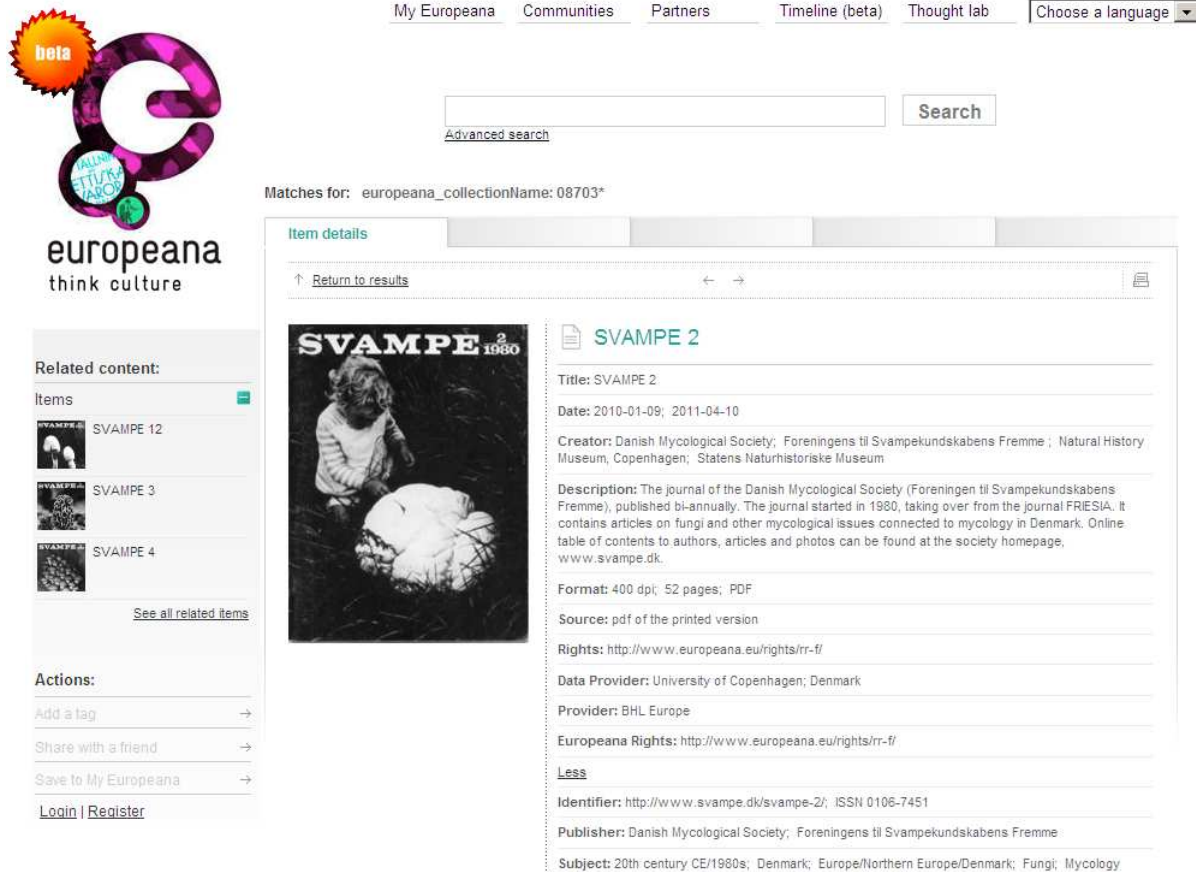
Figure 1-6: Thumbnails for Europeana

<europeana:type>

The value of this field must be one of: TEXT, IMAGE, VIDEO, SOUND. For BHL-Europe content the value "Text" is added automatically to each record as all of the BHL-Europe content is library content/books. We do not ingest separate images out of books into Europeana but always whole books or whole articles.

1.2 Europeana content checking

Once the content of the partners is in the OAI repository the partners receive a link to their content in the Europeana content checker tool where they can preview their records. In general about 100 test items per partner are uploaded to the content checker as the tool is not suited for mass uploads but rather for a “look-and-feel” preview checking.



The screenshot displays the Europeana content checker interface. At the top, there are navigation tabs: "My Europeana", "Communities", "Partners", "Timeline (beta)", and "Thought lab", along with a language selection dropdown. A search bar with a "Search" button and a link to "Advanced search" is present. The main content area shows a search result for "europeana_collectionName: 08703*". The "Item details" section for "SVAMPE 2" includes a thumbnail of the journal cover, a "Return to results" link, and a list of metadata fields: Title, Date, Creator, Description, Format, Source, Rights, Data Provider, Provider, Europeana Rights, Identifier, Publisher, and Subject. On the left, there is a sidebar with the Europeana logo, a "Related content" section listing other "SVAMPE" items, and an "Actions" section with options like "Add a tag", "Share with a friend", and "Save to My Europeana".

Figure 1-7: Detail view in the content checker tool

Partners could give WP3 feedback on their content presentation in the content checker tool. Their feedback is incorporated before the data is finally provided to Europeana.

At this time Europeana office also receives the link to the new content in the content checker and provides feedback.

Europeana Harvesting procedure:

Ingestion to Europeana is still carried out on demand. WP3 is informing Europeana in regular time intervals when new data is available. We plan to have the next harvests in summer/autumn 2011.

All Texts (60) Images (0) Videos (0) Sounds (0)

Results 1 - 12 of 60 Page: 1 2 3 4 5 →

Mit Emin Pascha ins Herz von Afrika Stuhmann, Franz BHL Europe	Aus den Tiefen des Weltmeeres Chun, Carl BHL Europe	Die Antike Tierwelt Keller, Otto BHL Europe	Die Antike Tierwelt Keller, Otto BHL Europe
Forschungsreise S.M.S. Planet 1906/07 BHL Europe	Forschungsreise S.M.S. Planet 1906/07 Hergesell, Hugo BHL Europe	Forschungsreise S.M.S. Planet 1906/07 Brennecke, W. BHL Europe	Forschungsreise S.M.S. Planet 1906/07 BHL Europe

Figure 1-8: Books for Europeana (UBER)

All Texts (100) Images (0) Videos (0) Sounds (0)

Results 1 - 12 of 100 Page: 1 2 3 4 5 6 7 8 9 →

Vorbericht. Martini, F.H.W. BHL Europe	Entstehungsgeschichte der Gesellschaft N... Martini, F.H.W. BHL Europe	Gesetze der hiesigen Privatgesellschaft ... [Anonym] BHL Europe	Verzeichniß der ordentlichen, wie auch d... [Anonym] BHL Europe
Anrede des beständigen Sekretairs der Ge... Martini, F.H.W. BHL Europe	Beschreibung eines bequemen Instruments, ... Bode, J.E. BHL Europe	Abhandlung über die Kraft der Elektricit... Achard, F.K. BHL Europe	Abhandlung über die durch das Verdünsten... Achard, F.K. BHL Europe

Figure 1-9: Articles for Europeana (UBBI)

1.3 Important changes in the current version of the Europeana Semantic Elements (ESE3.4, May 2011)

In May 2011 Europeana published a next version of the Europeana Semantic Elements (ESE). With this version 3.4 new **mandatory elements** have been introduced.

Among these count:

- **europaana:rights**
- **europaana:dataProvider**
- **dc:language**
Mandatory for objects with type TEXT
- **dc:title** and **dc:description**
Mandatory to provide one of these two
- **dc:subject, dc:type, dc:coverage; dcterms:spatial**
Mandatory to provide one of these four

The Europeana Semantic Elements (the ESE) listed below, consist of the 15 original Dublin Core (DC) metadata elements¹, a subset of the DC terms² and a set of thirteen elements which were created to meet Europeana's needs.

Source	Element	Element Refinement(s)
DC	title	alternative
DC	creator	
DC	subject	
DC	description	tableOfContents
DC	publisher	
DC	contributor	
DC	date	created ; issued
DC	type*	
DC	format	extent ; medium
DC	identifier	
DC	source	
DC	language*	
DC	relation	isVersionOf ; hasVersion ; isReplacedBy ; replaces ; isRequiredBy ; requires ; isPartOf ; hasPart ; isReferencedBy ; references ; isFormatOf ; hasFormat ; conformsTo
Europeana		isShownBy ; isShownAt
DC	coverage	spatial ; temporal
DC	rights*	
DC terms	provenance	
Europeana	country	
Europeana	dataProvider	
Europeana	language*	
Europeana	object	
Europeana	provider	
Europeana	rights*	
Europeana	type*	
Europeana	unstored	
Europeana	uri	
Europeana	userTag	
Europeana	year	

Source: <http://group.europeana.eu/web/guest/technical-requirements/>

The following pages provide an overview of the display of metadata fields in Version 1 of the BHL-Europe Portal (May 2011). The mandatory Europeana elements are highlighted and some examples on how they relate to the native data are provided.

DISPLAY FIELDS BHL-EUROPE		MODS	Namespace Dublin Core	Field	Namespace Europeana	Field	MARC21 (examples)	COMMENTS
Title		{modstitleInfo/nonSort} {modstitleInfo/title}; {modstitleInfo/subTitle}; {modstitleInfo/partNumber}; {modstitleInfo/partName}	dc:	title	dc:	title	245	
	Other Title	{modstitleInfo[type = '...']} {nonSort} {modstitleInfo[type = '...']/title}; {modstitleInfo[type = '...']/subTitle}; {modstitleInfo[type = '...']/partNumber}; {modstitleInfo[type = '...']/partName}	doterms:	alternative	doterms:	alternative		
Author		{modsfname/namePart} ({modsfname/displayForm}) <i>auv</i>	dc:	creator	dc:	creator		
Contributor		{modsfname/namePart} ({modsfname/displayForm}) ({modsfname/role/roleTerm})	dc:	contributor	dc:	contributor		
Date			dc:	date	dc:	date		
	Created	{modsforiginInfo/dateCreated}	doterms:	created	doterms:	created		
		{modsforiginInfo/dateCreated[point = 'start']} - {modsforiginInfo/dateCreated[point = 'end']}	doterms:	created	doterms:	created		
	Captured	{modsforiginInfo/dateCaptured}						
		{modsforiginInfo/dateCaptured[point = 'start']} - {modsforiginInfo/dateCaptured[point = 'end']}						
	Valid	{modsforiginInfo/dateValid}	doterms:	valid				
		{modsforiginInfo/dateValid[point = 'start']} - {modsforiginInfo/dateValid[point = 'end']}	doterms:	valid				
	Copyright	{modsforiginInfo/copyrightDate}	doterms:	copyrighted				
	Other	{modsforiginInfo/dateOther}						
		{modsforiginInfo/dateOther[point = 'start']} - {modsforiginInfo/dateOther[point = 'end']}						
Type		{modsfgenre}	dc:	type	dc:	type	655 \$a	
		{modsforiginInfo/issuance}	dc:	type	dc:	type	Leader/07	
		{modsftypeOfResource}	dc:	type	dc:	type	Leader/06	
Publication		{modsfname/namePart} ({modsfname/displayForm}) <i>pub</i>	dc:	publisher	dc:	publisher		
		{modsforiginInfo/publisher}; {modsforiginInfo/place/placeTerm}	dc:	publisher	dc:	publisher		
		{modsforiginInfo/dateIssued}	doterms:	issued	doterms:	issued		
		{modsforiginInfo/dateIssued[point = 'start']} - {modsforiginInfo/dateIssued[point = 'end']}	doterms:	issued	doterms:	issued		
		{modsfrequency}						
Language		{modsflanguage/languageTerm}; {modsflanguage/scriptTerm}	dc:	language	dc:	language	041; 008/35-37; 546\$b	mandatory for TEXT objects
Physical Description		{modsfphysicalDescription}	dc:	format	dc:	format		
		{modsfphysicalDescription/form}	dc:	format	dc:	format		
		{modsfphysicalDescription/internetMediaType}	dc:	format	dc:	format		
		{modsfphysicalDescription/extent}	doterms:	extent	doterms:	extent		
Abstract		{modsfabstract}	doterms:	abstract	dc:	description	520	
Table of Contents		{modsftableOfContents}	doterms:	tableOfContents	doterms:	tableOfContents		
Subject		{modsfclassification[authority = '...']} (...)	dc:	subject	dc:	subject	600, 610, 611, 630, 648, 650, 651	
		{modsfclassification}	dc:	subject	dc:	subject	050, 082, 080, 060, 086, 084	
		{modsfsubject/topic}; {modsfsubject/titleInfo/nonSort} {modsfsubject/titleInfo/title}; {modsfsubject/titleInfo/subTitle}; {modsfsubject/titleInfo/partNumber}; {modsfsubject/titleInfo/partName} / {modsfsubject/name/namePart} ({modsfsubject/name/displayForm}) / {modsfsubject/genre} / {modsfsubject/occupation} / {modsfsubject/geographic}	dc:	subject	dc:	subject	650 \$a\$b\$c\$d, 600,610,611,630,650,651 \$x, 6XX \$v, 656	
		{modsfsubject/hierarchicalGeographic/continent} - {modsfsubject/hierarchicalGeographic/country} - {modsfsubject/hierarchicalGeographic/region} - {modsfsubject/hierarchicalGeographic/state} - {modsfsubject/hierarchicalGeographic/territory} - {modsfsubject/hierarchicalGeographic/county} - {modsfsubject/hierarchicalGeographic/city} - {modsfsubject/hierarchicalGeographic/island} - {modsfsubject/hierarchicalGeographic/area}	doterms:	spatial	doterms:	spatial	651	
		{modsfsubject/cartographics}	doterms:	spatial	doterms:	spatial	255	
		{modsfsubject/temporal[point = 'start']} - {modsfsubject/temporal[point = 'end']}	doterms:	temporal	doterms:	temporal		
		{modsfsubject/temporal}	doterms:	temporal	doterms:	temporal	650 \$a	

DISPLAY FIELDS BHL-EUROPE		MODS	Namespace Dublin Core	Field	Namespace Europeana	Field	MARC21 (examples)	COMMENTS
Identification		{mods/identifier[invalid = 'yes']} (invalid) {mods/identifier[type = '...'][invalid = 'yes']} (invalid ...)	dc:	identifier	dc:	identifier		
		{mods/identifier[type = '...']} (...)	dc:	identifier	dc:	identifier		
		{mods/identifier}	dc:	identifier	dc:	identifier		
Holding Information		{mods/location/physicalLocation} - {mods/location/shelfLocator} {mods/location/holdingSimple/copyInformation/form} - {mods/location/holdingSimple/copyInformation/subLocation} - {mods/location/holdingSimple/copyInformation/shelfLocator} - {mods/location/holdingSimple/copyInformation/electronicLocator} - {mods/location/holdingSimple/copyInformation/note} - {mods/location/holdingSimple/copyInformation/enumerationAndChronology} {mods/location/url}			europaena:	dataProvider		YOUR institution name
					europaena:	isShownAt		URL to the record description at the BHL-Europe portal
Rights		{mods/accessCondition}	dc:	rights	dc:	rights		
					europaena:	rights		URL to licence
Related Items		{mods/relatedItem/titleInfo/nonSort} {mods/relatedItem/titleInfo/title}: {mods/relatedItem/titleInfo/subTitle}, {mods/relatedItem/titleInfo/partNumber}, {mods/relatedItem/titleInfo/partName} -- {mods/relatedItem/titleInfo[type = '...']/nonSort} {mods/relatedItem/titleInfo[type = '...']/title}: {mods/relatedItem/titleInfo[type = '...']/subTitle}, {mods/relatedItem/titleInfo[type = '...']/partNumber}, {mods/relatedItem/titleInfo[type = '...']/partName}	dc:	relation	dc:	relation		
Notes		{mods/note}	dc:	description	dc:	description		
Digital Object					europaena:	isShownBy		URL to the bookreader page at the BHL-Europe portal
N/A					europaena:	provider		BHL-Europe
N/A					europaena:	type		TEXT
Used for Thumbnail creation					europaena:	object		URL to the Page which was indicated as TITLE page of the book! > Europeana thumbnail!
					EXPLANATION: mandatory in Europeana			
					Mandatory to provide one of these			
					Mandatory to provide one of these			