

DELIVERABLE REPORT

Project Acronym: OpenUp!

Grant Agreement No: 270890

Project Title: Opening up the Natural History Heritage for Europeana

D02/D2.2.2 Harvesting and Transformation Component production version Part 1: Introduction

Revision: Doc02a – Final

Authors:

Gerda Koch AIT Forschungsgesellschaft mbH

Walter Koch AIT Forschungsgesellschaft mbH

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	x
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description
01	13.5.2011	Walter Koch	AIT	The OpenUp! harvesting prototype component installed with the help of: Odo Benda (AIT), Gerda Koch (AIT), Bernd Sproger (AIT)
01a	20.5.2011	Walter Koch	AIT	Start of Harvesting tests with BGBM data
01b	8.6.2011	Walter Koch	AIT	Start of Harvesting tests with NHMW data
01c	30.6.2011	Walter Koch	AIT	Start of Harvesting tests with MfN data
02	2.8.2011	Walter Koch	AIT	Virtual machine with harvesting prototype for test established and provided for the Work package leader (test111.ait.co.at with GBIF-Harvester and metadata manager)
02a	19.8.2011	Walter Koch	AIT	Start of Harvesting tests with BGBM Herbar data (test111.ait.co.at:8080/hit). Start of tests for Europeana thumbnail generation
02b	24.9.2011		UT-NHM	Harvesting test by UT-NHM
03	20.11.2011	Walter Koch	AIT	Virtual Machine with: GBIF-Harvester (http://test117.ait.co.at:8080/hit/) OAI-Provider http://test117.ait.co.at/oai-provider/index.php?form=index&db=0 Pentaho with jobs and transformation is also installed. On this machine we harvested appr. 200.000 ABCD-units which have been transformed (1st draft) to ESE 3.4 (can be searched and inspected via the OAI-Provider)
Doc01	17.11.2011	Walter Koch, Astrid Höller, Odo Benda, Gerda Koch	AIT	Draft of documentation (Part 1 to 3)
Doc02	18.11.2011	Gerda Koch	AIT	Version 1 of documentation (Part 1 to 3), formatting for distribution
Doc02a	21.11.2011	Coordination Team	BGBM	Minor editing

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

Distribution

Recipient	Date	Version	Accepted YES/NO
TMG	Continuous	WebSite: http://open-up.eu/content/harvesting-and-transformation-component	
Work Package Leader	2.8.2011	Virtual machine with harvesting prototype	
BGBM	19.8.2011	Feedback and communication on Harvesting Tests with HERBAR	
Work Package Leader	20.11.2011	Virtual Machine with: GBIF-Harvester (http://test117.ait.co.at:8080/hit/) OAI-Provider http://test117.ait.co.at/oai-provider/index.php?form=index&db=0 Pentaho with jobs and transformation is also installed. On this machine we harvested appr. 200.000 ABCD-units which have been transformed (1st draft) to ESE 3.4 (can be searched and inspected via the OAI-Provider)	
Work Package Leader	18.11.2011	Doc02 (Part 1 to 3)	WP leader G. Malarky and the Technology Management Group have been constantly following the progress of this Deliverable. Unfortunately, Mr. Malarky is not available until the end of November for providing his final acceptance.
Project Coordinator	18.11.2011	Doc02a (Part 1 to 3)	Yes

Table of Contents

Management Summary	5
1. Introduction	6
1.1. The Harvesting and Transformation component in general	6
1.2. The Hit Harvester	6
1.3. Data Integration via Pentaho	7
1.4. Access to Biological Collection Data (ABCD)	9
1.5. Ingesting data into Europeana	9
1.5.1 <i>The Europeana Semantic Elements v3.4</i>	10
2. Mapping ABCD v2.06 to ESE v3.4	11
Glossar	21
I. List of Figures	22
II. List of Tables	23

Management Summary

The Harvesting and Transformation component is a software that realizes the harvesting of ABCD metadata and the transforming of this metadata into the ESE format of the European Digital Library, Europeana¹. This software is accompanied by the following documentation:

Part 1: Introduction (with basic mapping information)

Part 2: Stylesheet

Part 3: The Pentaho Transformation Procedure

The present document “Part 1: Introduction” introduces to the Deliverable D02/D2.2.2 of Task 2.2 “The Harvesting and Transformation component” and shows the mapping that builds the baseline for the transformation process from ABCD metadata to Europeana conform ESE metadata. “Part 2: Stylesheet” describes the stylesheet that is used within the Pentaho procedure. In “Part 3: The Pentaho Transformation Procedure” the transformation procedure from the native ABCD record to the standard Europeana record is described.

¹ <http://www.europeana.eu>

1. Introduction

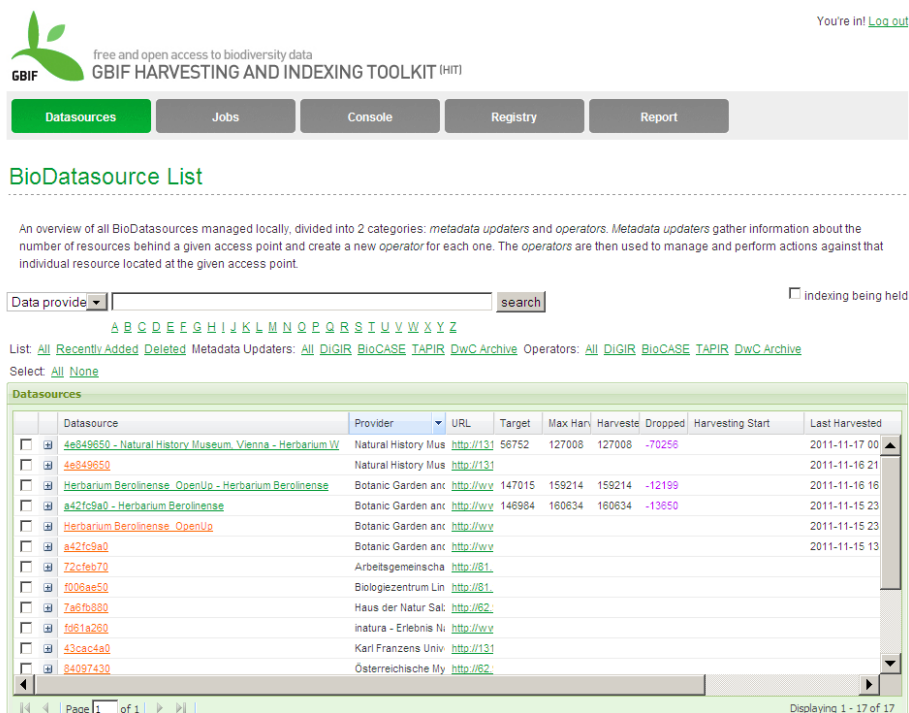
1.1. The Harvesting and Transformation component in general

The OpenUp! harvesting and transformation component will be set up to harvest data on request from the data repositories of the content providers and ingest the data to the OpenUp! OAI provider. D2.2.2 the OpenUp! harvesting and transformation component is delivered in a virtual machine to the WP2 lead, the Natural History Museum in London. It will be set productive at the NHM in London. D2.2.2 contains all components of the harvesting platform and the transformation component from ABCD format to ESE.

1.2. The Hit Harvester

For harvesting the OpenUp! Content providers' data the GBIF HIT (Harvesting Index Toolkit) Harvester has been used. "The Harvesting Index Toolkit is an open-source, Java-based web application that simplifies the process of harvesting biodiversity data from a distributed network of data publishers."²

The harvested native data (approx. 200.000 records) which was used for testing and developing the transformation component was put onto the virtual machine for NHM London, too. The Pentaho transformation procedures is taking the native data from there.



The screenshot shows the GBIF HIT web application interface. At the top, there is a navigation bar with buttons for 'Datasources', 'Jobs', 'Console', 'Registry', and 'Report'. Below this is a 'BioDatasource List' section. A search bar is present with a 'search' button and a 'Data provide' dropdown. A list of datasources is displayed in a table format. The table has columns for 'Datasource', 'Provider', 'URL', 'Target', 'Max Harv', 'Harveste', 'Dropped', 'Harvesting Start', and 'Last Harvested'. The first few rows are visible, showing various herbarium and botanical garden entries.

Datasource	Provider	URL	Target	Max Harv	Harveste	Dropped	Harvesting Start	Last Harvested
4e849650 - Natural History Museum, Vienna - Herbarium W	Natural History Mus	http://131	56752	127008	127008	-70256		2011-11-17 00
4e849650	Natural History Mus	http://131						2011-11-16 21
Herbarium Berolinense - OpenUp - Herbarium Berolinense	Botanic Garden anc	http://wv	147015	159214	159214	-12199		2011-11-16 16
a42fc9a0 - Herbarium Berolinense	Botanic Garden anc	http://wv	146984	160634	160634	-13850		2011-11-15 23
Herbarium Berolinense - OpenUp	Botanic Garden anc	http://wv						2011-11-15 23
a42fc9a0	Botanic Garden anc	http://wv						2011-11-15 13
72cfeb79	Arbeitsgemeinscha	http://81						
f006aae0	Biologiezentrum Lin	http://81						
7a6fb880	Haus der Natur Sal	http://62						
f961a260	inatura - Erlebnis Ni	http://wv						
43cac4a0	Karl Franzens Univ	http://131						
84097430	Osterreichische My	http://62						

Figure 1 Local GBIF harvester at the D2.2.2 component

² <http://www.gbif.org/informatics/infrastructure/indexing/>

Figure 1 shows a screenshot of the local HIT harvester. Only metadata that is connected to digital objects (images etc.) has been included in the harvest, plain metadata has been dropped.

1.3. Data Integration via Pentaho

The OpenUp! Project is a best practice project therefore it was decided to use the state-of-the-art open source Pentaho³ Data Integration system to support the data transformation process. This process is described in detail in Part 3 of the D02 documentation “The Pentaho Transformation Procedure”.

“Pentaho Data Integration, codenamed Kettle, consists of a core data integration (ETL) engine, and GUI applications that allow the user to define data integration jobs and transformations. It supports deployment on single node computers as well as on a cloud, or cluster.”⁴

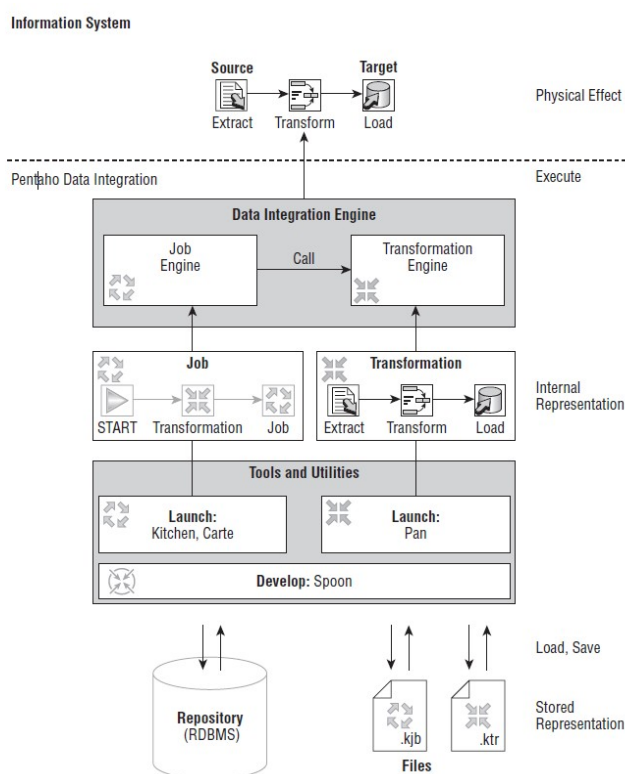


Figure 9-1: Pentaho data integration tools and components

Figure 2 Pentaho data integration tools and components⁵

³ <http://www.pentaho.com/>

⁴ <http://en.wikipedia.org/wiki/Pentaho>

⁵ figure taken from: “Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL”, published by Wiley Publishing, Inc., copyright 2009 Wiley Publishing, Inc., Indianapolis, Indiana

The next figure depicts a screenshot of the Pentaho user interface where the steps, hops and record streams for the transformation are designed.

The data of the OpenUp! Data providers is stored in the ABCDv2.06 format. The main target repository of the OpenUp! Project is the European Digital Library, called Europeana. Europeana uses the Europeana Semantic Elements (ESE) as a metadata standard. Therefore the data is transformed according to the requirements of the harvesting routines of Europeana.

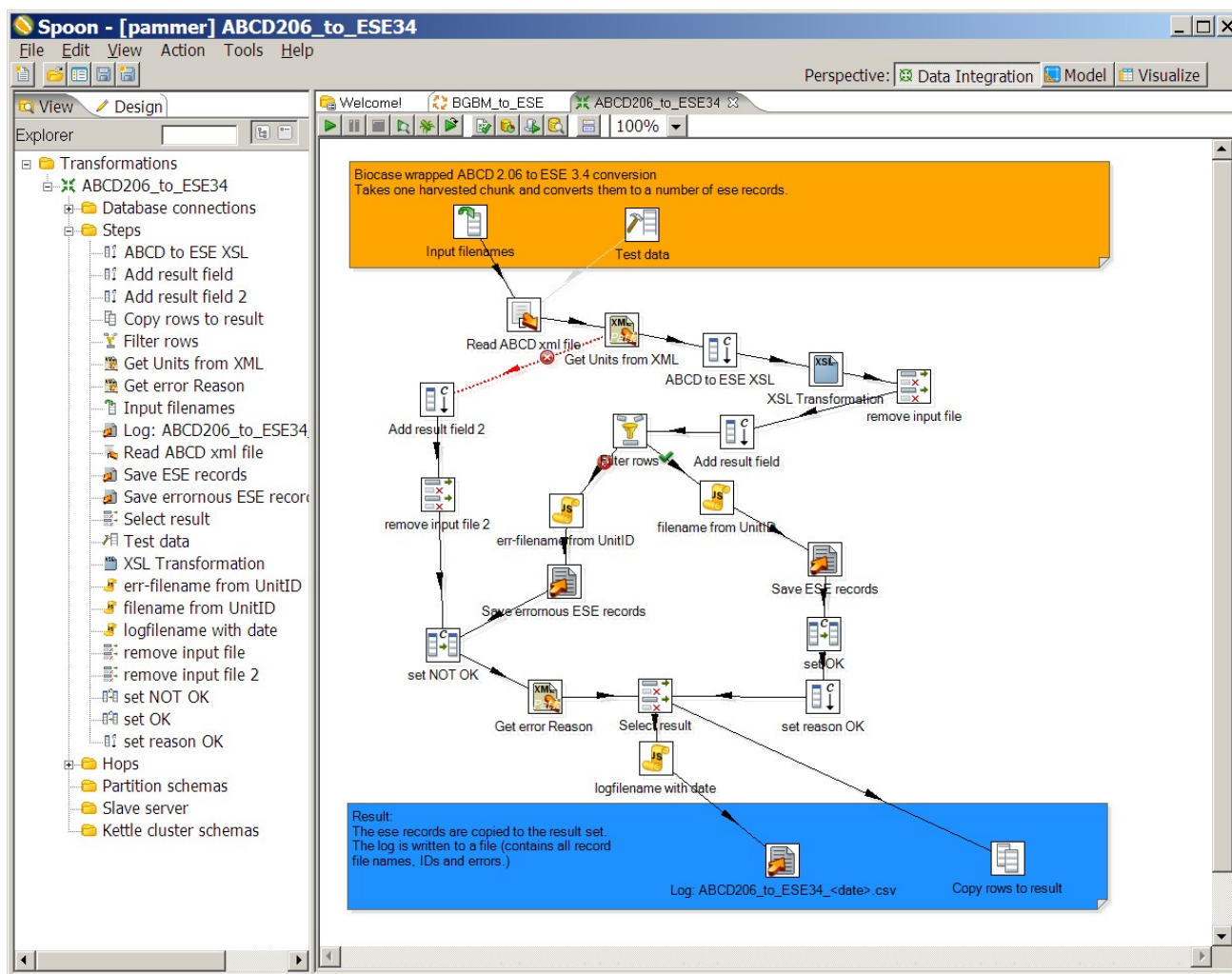


Figure 3 ABCD to ESE Pentaho transformation diagram

1.4. Access to Biological Collection Data (ABCD)

The data of the OpenUp! Content providers is primarily documented using the ABCD data schema. The Access to Biological Collections Data (ABCD)⁶ Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data).

The ABCD Schema⁷ attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomized data and free-text can be accommodated. Version 2.06 is currently in use with the GBIF (Global Biodiversity Information Facility⁸) and BioCASE (Biological Collection Access Service for Europe⁹) networks.

1.5. Ingesting data into Europeana

The objectives of Europeana, the European Digital Library are:

- providing access to Europe's cultural and scientific heritage by way of a cross-domain portal
- facilitating formal agreement across museums, archives, audiovisual archives and libraries on how to co-operate in the delivery and sustainability of a joint portal
- stimulating and facilitating initiatives to bring together existing digital content
- supporting and facilitating digitization of Europe's cultural and scientific heritage¹⁰

The aim of providing a single entry point to the diverse cultural and scientific data bases and content repositories spread over Europe led to the need to develop a quite simple and straight forward data scheme at the beginning of Europeana (late 2008 onwards). This scheme is the Europeana Semantic Elements scheme (ESE) which builds on the international Dublin Core Qualified Standard. In agreement with Europeana office the first data transfer from OpenUp! to the Europeana portal in February 2012 will make use of this ESE format.

In October 2011 the first official guidelines for mapping data to the EDM, the Europeana Data model¹¹ have been published. EDM is a more advanced data model allowing for a semantic approach and the integration of rich information while supporting a better linkage between the different information units. In addition to its own elements EDM reuses elements from existing namespaces like the Resource Description

⁶ <http://wiki.tdwg.org/twiki/bin/view/ABCD/AbcdPrimer>

⁷ <http://www.tdwg.org/standards/115/>

⁸ <http://www.gbif.org/>

⁹ <http://www.biocase.org/>

¹⁰ <http://version1.europeana.eu/web/europeana-project/home>

¹¹ <http://version1.europeana.eu/web/europeana-project/technicaldocuments/>

Framework (RDF) and the RDF Schema (RDFS) namespaces (<http://www.w3.org/TR/rdf-concepts/>), the OAI Object Reuse and Exchange (ORE) namespace (<http://www.openarchives.org/ore>), the Simple Knowledge Organization System (SKOS) namespace (<http://www.w3.org/TR/skos-reference/>) and the Dublin Core namespaces for elements (<http://purl.org/dc/elements/1.1/>, abbreviated as DC), terms (<http://purl.org/dc/terms/>, abbreviated as DCTERMS) and types (<http://purl.org/dc/dcmitype/>, abbreviated as DCMITYPE).

It is planned that this more comprehensive data scheme will be evaluated and used for data transfers after February 2012.

1.5.1 The Europeana Semantic Elements v3.4

The ESE v3.4 Schema¹² is the currently used data scheme for data integration into the Europeana portal (www.europeana.eu). Version 3.4 of this metadata scheme introduced a number of mandatory elements that must be considered as well in the transformation process from ABCD to ESE.

ESE v3.4 comprises the following data elements:

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	

Figure 4 ESE v3.4 data elements¹³

¹² <http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd>

2. Mapping ABCD v2.06 to ESE v3.4¹⁴

The following table depicts the first version of the mapping of the ABCD data fields to the ESE data scheme. It is still very comprehensive taking into account that the content providing partners have used the ABCD elements very differently (meaning that some partners have used lots of fields while others might have used only a few etc..., depending as well on the data they have been describing).

	ESE	ABCD 2.06	Comment	Obligation
1	dc:title	/DataSets/DataSet/Units/Unit/Identifications/Identification/Result/TaxonIdentified/ScientificName/FullScientificNameString	take the value with: PreferredFlag	Mandatory to provide one of these two (dc:title or dc:description)
	dc:title	/DataSets/DataSet/Units/Unit/SpecimenUnit/NomenclaturalTypeDesignations/NomenclaturalTypeDesignation/TypifiedName/FullScientificNameString	Append: (Typus) repeatable	
2	dc:description	/DataSets/DataSet/Units/Unit/Notes		Mandatory to provide one of these two (dc:type or dcterms:spatial)
	dc:description	/DataSets/DataSet/Units/Unit/SpecimenUnit/Preparations/Preparation/PreparationType		
	dc:description	/DataSets/DataSet/Units/Unit/Sex	evt. Append (sex)	
	dc:description	/DataSets/DataSet/Units/Unit/KindOfUnit		
	dc:description	/DataSets/DataSet/Units/Unit/Gathering/Notes	check what is usually used in this field!	
	dc:description	/DataSets/DataSet/Units/Unit/SpecimenUnit/Marks/Mark/MarkText		
3	dc:type	/DataSets/DataSet/Units/Unit/RecordBasis	(Preserved Specimen, Fossil Specimen, Living Specimen, Human Observation, Machine Observation)	Mandatory to provide one of these two (dc:type or dcterms:spatial)
4	dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/LocalityText		
	dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/SiteCoordinateSets/SiteCoordinates/CoordinatesLatLong/LatitudeDecimal	EDM Preferred display sequence: date, method, coordinates (lat+long), error, distance, accuracy statement (alternative if no coordinates used: coordinate text)	
	dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/SiteCoordinateSets/SiteCoordinates/CoordinatesLatLong/LongitudeDecimal	EDM	
	dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Country/Name		

¹³ Source 'DC ...': This element is part of the Dublin Core Standard; Source 'Europeana': This element is introduced by Europeana.

¹⁴ The baseline for this mapping was developed during a workshop at BGBM in June 2011. For more information see: <http://open-up.eu/content/wp3-abcd-ese-mapping-workshop-june-21-22-bgbm>

ESE	ABCD 2.06	Comment	Obligation
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Altitude/MeasurementOrFactAtomised/LowerValue		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/NamedAreas/NamedArea/AreaName		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/SiteCoordinateSets/SiteCoordinates/CoordinatesLatLong/CoordinateErrorDistanceInMeters	put it after longitude	
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Altitude/MeasurementOrFactAtomised/UpperValue		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/NamedAreas/NamedArea/AreaClass	qualifies Area (put it in brackets after the Area Name)	
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Altitude/MeasurementOrFactAtomised/UnitOfMeasurement	if not there put m as default value	
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/AreaDetail		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Biotope/Text		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Depth/MeasurementOrFactAtomised/LowerValue		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/SiteCoordinateSets/SiteCoordinates/CoordinateMethod		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/SiteCoordinateSets/SiteCoordinates/CoordinatesLatLong/SpatialDatum	put it in brackets after the longitude	
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Biotope/Name		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Stratigraphy/ChronostratigraphicTerms/ChronostratigraphicTerm/Term		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Depth/MeasurementOrFactAtomised/UnitOfMeasurement	default m (meter) if no value here	
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Aspect/Ordination		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Aspect/Text		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Height/MeasurementOrFactAtomised/UnitOfMeasurement	default m (meter) - if not used	
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Height/MeasurementOrFactAtomised/UpperValue		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/NearNamedPlaces/NamedPlaceRelation/NearNamedPlace		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Height/MeasurementOrFactAtomised/LowerValue		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Altitude/MeasurementOrFactAtomised/LowerValue/accuracy	not yet in use	
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Altitude/MeasurementOrFactAtomised/UpperValue/accuracy	not yet in use	
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Stratigraphy/LithostratigraphicTerms/LithostratigraphicTerm/Term		
dcterms:spatial	/DataSets/DataSet/Units/Unit/Gathering/Stratigraphy/BiostratigraphicTerms/BiostratigraphicTerm/Term		

	ESE	ABCD 2.06	Comment	Obligation
5	europa:dataProvider	/DataSets/DataSet/Metadata/Owners/Owner/Organisation/Name/Representation/Text	if no unit owner is given	mandatory
	europa:dataProvider	/DataSets/DataSet/Units/Unit/Owner/Organisation/Name/Representation/Text	this is the preferred ABCD field for dataprovider if not used than take the value from /DataSets/DataSet/Metadata/Owners/Owner/Organisation/Name/Representation/Text	
6	europa:isShownAt	/DataSets/DataSet/Units/Unit/MultiMediaObjects/MultiMediaObject/ProductURI		mandatory
7	europa:isShownBy	/DataSets/DataSet/Units/Unit/MultiMediaObjects/MultiMediaObject/FileURI		mandatory
8	europa:rights	/DataSets/DataSet/Units/Unit/MultiMediaObjects/MultiMediaObject/IPR/IPRStatements/Licenses/License/URI	License.URI [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	mandatory
9	dc:identifier	/DataSets/DataSet/Units/Unit/SourceID	all three together: unique identifier	Mandatory
	dc:identifier	/DataSets/DataSet/Units/Unit/SourceInstitutionID	all three together: unique identifier	
	dc:identifier	/DataSets/DataSet/Units/Unit/UnitID	all three together: unique identifier	
10	dc:contributor	/DataSets/DataSet/Units/Unit/Gathering/Agents/GatheringAgent/AgentText		Optional
	dc:contributor (role:identifier)	(role:/DataSets/DataSet/Units/Unit/Identifications/Identification/Identifiers/IdentifiersText	preferred flag	
	dc:contributor (role:collector)	/DataSets/DataSet/Units/Unit/Gathering/Agents/GatheringAgent/Person/FullName	Append: (/DataSets/DataSet/Units/Unit/CollectorsFieldNumber), (/DataSets/DataSet/Units/Unit/Gathering/Code)	
	dc:contributor (role:collector)	/DataSets/DataSet/Units/Unit/Gathering/Agents/GatheringAgentsText	attribute: primary etc. most important	
	dc:contributor (role:identifier)	/DataSets/DataSet/Units/Unit/Identifications/Identification/Identifiers/Identifier/PersonName/FullName	only for identification with preferred flag after (role:identifier) put: (/DataSets/DataSet/Units/Unit/Identifications/Identification/Date/ISODateTimeBegin(if not available put: (/DataSets/DataSet/Units/Unit/Identifications/Identification/Date/DateText)	
	dc:contributor	/DataSets/DataSet/Units/Unit/CollectorsFieldNumber	Append after: dc:contributor field number(collector)	
	dc:contributor	/DataSets/DataSet/Units/Unit/Identifications/Identification/Date/ISODateTimeBegin	dc:contributor (role:identifier)	
	dc:contributor	/DataSets/DataSet/Units/Unit/Gathering/Code	dc:contributor field number(collector)	
11	dc:date	/DataSets/DataSet/Units/Unit/Gathering/DateTime/DateText	EDM if the recordbasis is NOT FossilSpecimen use> dcterms:temporal	Optional
	dc:date	/DataSets/DataSet/Units/Unit/Gathering/DateTime/ISODateTimeBegin	EDM if the recordbasis is NOT FossilSpecimen use> dcterms:temporal	

ESE	ABCD 2.06	Comment	Obligation
dc:date	/DataSets/DataSet/Units/Unit/Gathering/DateTime/ISODateTimeEnd	EDM if only timebegin: put timebegin (just one date); if timeend available put: timebegin - timeend if the recordbasis is NOT FossilSpecimen than use> dcterms:temporal	
	/DataSets/DataSet/Units/Unit/Identifications/Identification/Date/DateText	if timebegin not used	
12	dc:identifier /DataSets/DataSet/DatasetGUID		
	dc:identifier /DataSets/DataSet/Units/Unit/UnitGUID		
13	dc:relation /DataSets/DataSet/Units/Unit/Associations/UnitAssociation/AssociatedUnitID	EDM	Optional
	dc:relation /DataSets/DataSet/Units/Unit/Associations/UnitAssociation/AssociatedUnitSourceInstitutionCode	EDM	
	dc:relation /DataSets/DataSet/Units/Unit/Associations/UnitAssociation/AssociatedUnitSourceName	EDM	
	dc:relation /DataSets/DataSet/Units/Unit/Associations/UnitAssociation/AssociationType	EDM	
	dc:relation /DataSets/DataSet/Units/Unit/Associations/UnitAssociation/Comment	EDM	
	dc:relation /DataSets/DataSet/Units/Unit/Assemblage/UnitAssemblage/AssemblageID	EDM	
	dc:relation /DataSets/DataSet/Units/Unit/Assemblage/UnitAssemblage/AssemblageName	EDM	
14	dc:rights /DataSets/DataSet/Metadata/IPRStatements/TermsOfUseStatements/TermsOfUse/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	Optional
	dc:rights /DataSets/DataSet/Metadata/IPRStatements/Copyrights/Copyright/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
	dc:rights /DataSets/DataSet/Metadata/IPRStatements/Disclaimers/Disclaimer/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	

ESE	ABCD 2.06	Comment	Obligation
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Acknowledgements/Acknowledgement/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Copyrights/Copyright/URI	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Citations/Citation/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Disclaimers/Disclaimer/URI	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/IPRDeclarations/IPRDeclaration/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Licenses/License/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Acknowledgements/Acknowledgement/Details	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	

ESE	ABCD 2.06	Comment	Obligation
dc:rights	/DataSets/DataSet/Units/Unit/MultiMediaObjects/MultiMediaObject/IPR/Copyrights/Copyright/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Units/Unit/MultiMediaObjects/MultiMediaObject/IPR/TermsOfUseStatements/TermsOfUse/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Acknowledgements/Acknowledgement/URI	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/IPRDeclarations/IPRDeclaration/URI	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/TermsOfUseStatements/TermsOfUse/Details	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Disclaimers/Disclaimer/Details	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/TermsOfUseStatements/TermsOfUse/URI	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	

ESE	ABCD 2.06	Comment	Obligation
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Citations/Citation/URI	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Copyrights/Copyright/Details	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/IPRDeclarations/IPRDeclaration/Details	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Licenses/License/Details	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Metadata/IPRStatements/Citations/Citation/Details	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Units/Unit/IPRStatements/Copyrights/Copyright/Text	overrides metadata rights Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements	
dc:rights	/DataSets/DataSet/Units/Unit/IPRStatements/Citations/Citation/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	

ESE	ABCD 2.06	Comment	Obligation
dc:rights	/DataSets/DataSet/Units/Unit/IPRStatements/Licenses	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Units/Unit/IPRStatements/Licenses/License/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
dc:rights	/DataSets/DataSet/Units/Unit/IPRStatements/TermsOfUseStatements/TermsOfUse/Text	Concatenate Statement.Text + Statement.Details + Statement.URI for IPR-Declaration, Copyrights, Licenses, TermsOfUse, Disclaimer, Acknowledgements, and Citations [Priority: 1.MultimediaObject.IPRStatements, 2. Unit.IPRStatements and 3. Dataset.Metadata.IPRStatements)	
15 dc:source	/DataSets/DataSet/Metadata/Description/Representation/Title		Optional
16 dcterms:provenance	/DataSets/DataSet/Units/Unit/SpecimenUnit/History/PreviousUnitsText		Optional
17 europeana:object		used for thumbnail creation possible solutions: extension for ABCD Europeana:object, or encryption within unit:notes	Optional
18 europeana:unstored			Optional

Table 1 Mapping ABCD to ESE (Version 1)

The following figures depict screenshots of mapped ABCD data to ESE.

XML		
version	1.0	
encoding	UTF-8	
europeana:record		
xmlns:xsi	http://www.w3.org/2001/XMLSchema-instance	
xmlns:europeana	http://www.europeana.eu/schemas/e/e/	
xmlns:dc	http://purl.org/dc/elements/1.1/	
xmlns:dcterms	http://purl.org/dc/terms/	
_record_invalid	true	
_record_invalid_reason	Missing license uri for rights.	
dc:title (3)		
	Rbc Text	
1	Lycopodium capillaceum (Spring) Hieron.	
2	Lycopodiella cernua (L.) Pic.Serm.	
3	Lycopodium cernuum L.	
dc:description (3)		
	Rbc Text	
1	another specimen of Bonpland & Humboldt 473, see Lycopodium capillaceum (B -W 19429 -01 0)	
2	herbarium sheet	
3	another specimen of Bonpland & Humboldt 473, see Lycopodium capillaceum (B -W 19429 -01 0)	
dc:type	PreservedSpecimen	
dcterms:spatial (3)		
	Rbc Text	
1	La Cuchila prope Guanaguana.	
2	Venezuela	
3	Middle and South America	
europeana:dataprovider	Botanic Garden and Botanical Museum Berlin-Dahlem	
europeana:is ShownAt	http://www2.bgbm.org/herbarium/view_biocase.cfm?SpecimenPK=55375	
europeana:is ShownBy	http://www2.bgbm.org/herbarium/images/B/-W/19/36/thumbs/B_-W_19368%20-05%200.jpg	
dc:identifier	Herbarium Berolinense - BGBM - B -W 19368 -05 0	
dc:contributor (2)		
	Comment	
1	role="collector"	A. J. A. Bonpland & F. W. H. A. von Humboldt
2	role="identifier"	G. Hieronymus
europeana:rights	http://www.europeana.eu/portal/rr-f.html	
dc:rights (2)		
	Rbc Text	
1	The use of the data is allowed only for non-profit scientific use and for non-profit nature conservation purpose. The data base or part of it may only be used or copied by the written permission from the legal owner.	
2	Röpert, D. (Ed.) 2000 - (continuously updated): Digital specimen images at the Herbarium Berolinense (B).	
dc:source	Herbarium Berolinense	

Figure 5 ESEv3.4 XML expression of an ABCD record (displayed in an XML viewer)

```

- <europeana:record_record_invalid="true" _record_invalid_reason="Missing license uri for rights. ">
  <dc:title>Lycopodium capillaceum (Spring) Hieron.</dc:title>
  <dc:title>Lycopodiella cernua (L.) Pic.Serm.</dc:title>
  <dc:title>Lycopodium cernuum L.</dc:title>
- <dc:description>
  another specimen of Bonpland & Humboldt 473, see Lycopodium capillaceum (B -W 19429 -01 0)
</dc:description>
<dc:description>herbarium sheet</dc:description>
- <dc:description>
  another specimen of Bonpland & Humboldt 473, see Lycopodium capillaceum (B -W 19429 -01 0)
</dc:description>
<dc:type>PreservedSpecimen</dc:type>
<dcterms:spatial>La Cuchila prope Guanaguana.</dcterms:spatial>
<dcterms:spatial>Venezuela</dcterms:spatial>
<dcterms:spatial>Middle and South America</dcterms:spatial>
<europeana:dataprovider>Botanic Garden and Botanical Museum Berlin-Dahlem</europeana:dataprovider>
- <europeana:isShownAt>
  http://ww2.bgbm.org/herbarium/view_biocase.cfm?SpecimenPK=55375
</europeana:isShownAt>
- <europeana:isShownBy>
  http://ww2.bgbm.org/herbarium/images/B/-W/19/36/thumbs/B_-W_19368%20-05%200.jpg
</europeana:isShownBy>
<dc:identifier>Herbarium Berolinense - BGBM - B -W 19368 -05 0</dc:identifier>
- <dc:contributor>
  <!--role="collector"-->
  A. J. A. Bonpland & F. W. H. A. von Humboldt
</dc:contributor>
- <dc:contributor>
  <!--role="identifier"-->
  G. Hieronymus
</dc:contributor>
<europeana:rights>http://www.europeana.eu/portal/rr-f.html</europeana:rights>
- <dc:rights>
  The use of the data is allowed only for non-profit scientific use and for non-profit nature conservation purpose.
  The data base or part of it may only be used or copied by the written permission from the legal owner.
</dc:rights>
- <dc:rights>
  Röpert, D. (Ed.) 2000 - (continuously updated): Digital specimen images at the Herbarium Berolinense (B).
</dc:rights>
<dc:source>Herbarium Berolinense</dc:source>
</europeana:record>

```

Figure 6 Plain ESEV3.4 XML expression of an ABCD record

Glossar

ABCD	Access to Biological Collection Data An evolving comprehensive standard for the access to and exchange of data about specimens and observations. http://wiki.tdwg.org/twiki/bin/view/ABCD/AbcdPrimer
BioCASE	Biological Collection Access Service for Europe A transnational network of biological collections of all kinds. http://www.biocase.org/
DC	Dublin Core A vocabulary of fifteen properties for use in resource description. http://www.dublincore.org/documents/
ESE v3.4	The Europeana Semantic Elements specification of the Europeana Portal (www.europeana.eu) http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd
GBIF	Global Biodiversity Information Facility aim: encourage free and open access to biodiversity data, via the Internet. http://www.gbif.org/
HIT	Harvesting Index Toolkit An open source Java-based web application developed by the GBIF Secretariat to manage biodiversity data harvesting and quickly build indexes of harvested data. http://code.google.com/p/gbif-indexingtoolkit/
ORE	OAI Object Reuse and Exchange namespace http://www.openarchives.org/ore
RDF/RDFS	Resource Description Framework and the RDF Schema http://www.w3.org/TR/rdf-concepts/ ,
SKOS	Simple Knowledge Organization System namespace http://www.w3.org/TR/skos-reference/

I. List of Figures

Figure 1 Local GBIF harvester at the C2.2.2 component	6
Figure 2 Pentaho data integration tools and components	7
Figure 3 ABCD to ESE Pentaho transformation diagram	8
Figure 4 ESE v3.4 data elements	10
Figure 5 ESEv3.4 XML expression of an ABCD record (displayed in an XML viewer)	19
Figure 6 Plain ESEv3.4 XML expression of an ABCD record.....	20

II. List of Tables

Table 1 Mapping ABCD to ESE (Version 1).....	18
--	----