

ECP-2008-DILI-518001

BHL-Europe

**Technical architecture status and progress
report with particular focus on the development
of the German prototype**

Deliverable number	<i>D3.5</i>
Dissemination level	<i>Public</i>
Delivery date	<i>30 April 2010</i>
Status	<i>Version 1.0</i>
Author(s)	<i>AIT, ATOS, MfN, NHM, NHMW</i>



eContentplus

This project is funded under the *eContentplus* programme¹,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

¹ OJ L 79, 24.3.2005, p. 1.

Table of contents

1	DOCUMENT HISTORY	3
1.1	CONTRIBUTORS.....	3
1.2	REVISION HISTORY.....	3
1.3	REVIEWERS AND APPROVALS.....	3
1.4	DISTRIBUTION.....	4
2	PURPOSE AND DOCUMENT STRUCTURE	5
3	TECHNICAL ARCHITECTURE STATUS	6
4	PROGRESS REPORT ON THE DEVELOPMENT OF THE GERMAN PROTOTYPE	9
4.1	PRE-INGEST	9
4.2	INGEST.....	11
4.3	DATA MANAGEMENT.....	11
4.4	ARCHIVAL STORAGE.....	12
4.5	ACCESS.....	12
4.6	PORTAL.....	13
5	ROADMAP	15
5.1	METHODOLOGY	15
5.2	PROJECT BACKLOG OVERVIEW.....	15
5.3	SPRINTS OVERVIEW.....	16
6	ACRONYMS AND ABBREVIATIONS	17
7	FIGURES	19

1 Document History

This chapter describes the document's creation events and contributors.

1.1 Contributors

This document is based on the meetings of the technical team in Leiden, Graz and Berlin with the following members contributing to this document.

Person	Partner
Alexander Herzog	AIT
Gerda Koch	AIT
Lee Namba	ATOS
Henning Scholz	MfN
Adrian Smales	NHM
Wolfgang Koller	NHMW

1.2 Revision History

Revision Date	Author	Version	Change Reference & Summary
15 April 2010	A. Herzog	0.1	1 st draft
21 April 2010	A. Herzog	0.2	Draft taking comments of the BHLE-tech group into account
22 April 2010	W. Koller	0.3	Draft adding information on SMT
27 April 2010	H. Scholz	0.4	Draft incorporating partner feedback
30 April 2010	H. Scholz	1.0	Final version

1.3 Reviewers and Approvals

This document requires the following reviews and approvals.

Name	Position	Date	Version
Adrian Smales	Work Package Leader	30 April 2010	1.0
BHL-Europe tech group	TMB	30 April 2010	1.0
Henning Scholz	Project Coordinator	30 April 2010	1.0

1.4 Distribution

This document has been distributed to:

Group	Date of issue	Version
BHL-Europe tech group	15 April 2010	0.1
BHL-Europe tech group and consortium	21 April 2010	0.2
BHL-Europe consortium	22 April 2010	0.3
BHL-Europe tech group	29 April 2010	0.4
BHL-Europe consortium	3 May 2010	1.0

2 Purpose and document structure

This document is publicly available and describes the concepts of the BHL-Europe system. It will introduce the components needed within the system without technical background or details. It will cover what has been done so far within the project and will provide a milestone plan for the final development.

The document will discuss the current technical architecture status by explaining the available components and their key aspects. Each component serves a specific need within the BHL-Europe system. These needs are either serving user and system interaction or scalability and preservation.

Chapter 4 is describing the development on the German prototype. All components are discussed in relation to the German prototype to be delivered in Month 18. This prototype is not scalable nor does it provide a preservation or archival storage component. The prototype will be used to collect feedback for the final implementation.

The document finishes with a milestone plan for the final development and will outline the availability of functionality through the following 24 month and a certain period after.

3 Technical architecture status

In addition to the BHL Europe System architecture described in D3.4 this chapter describes the extension of the BHL-Europe System architecture to a global scale taking into account other regional BHL repository systems, global identifiers between systems, and replication procedures between them. The current diagram is shown in Figure 3-1.

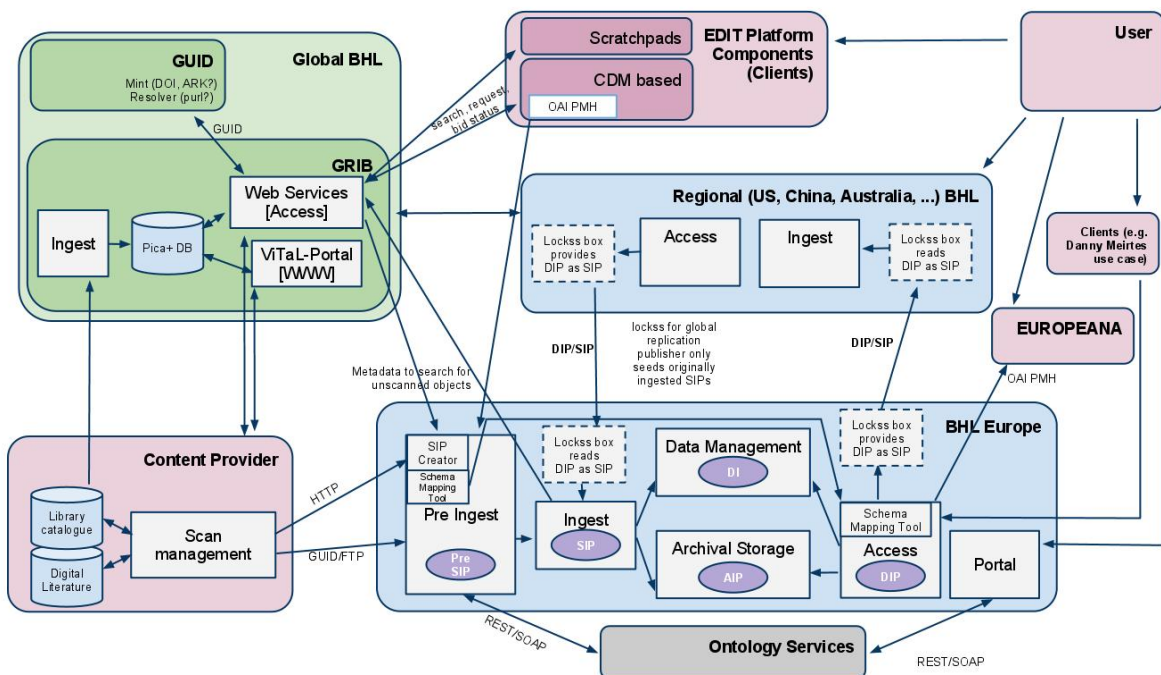


Figure 3-1 Architecture diagram

The technical architecture is based on the Open Archival Information System (OAIS) reference model. In addition to the core elements, BHL-Europe developed two modules that fulfil the roles Provider and Consumer. The provider role is covered by a Pre-Ingest module and the Consumer is represented by a Portal module.

The **Portal** is the user interface for all BHL-Europe users. Search and retrieval will be done within this component. The user will be able to search for books via a simple, one field search box and an advanced search based on various metadata fields like author, publication date or title. Details of the search functionality will also come from the use cases we currently develop and the results of the user requirement survey open from 15 March to 30 April 2010. Additional ontology services will help the users to find books by browsing multilingual taxons. The retrieval will provide the possibility to browse through the books of the result set

with a book reader application and to download whole books, serials or just parts of a book like an article. The portal will be based on Drupal².

The **Access** component is the API that the portal and 3rd party applications use to pull information from the BHL-Europe system. It handles the creation of result sets and the provision of dissemination information packages (DIPs). DIPs are the downloadable items displayed at the Portal. Access will be used to distribute the system's stored items throughout the web to other OAIS systems. This mechanism is used to support distributed storage and replication. One distribution mechanism can be Lockss³. The DIPs can be created using the Schema Mapping Tool⁴. The evaluation will be carried out in M12.

Data Management is used to store search and retrieval information about the items stored in Archival Storage. Data Management can be seen as the index Access will work with, when a search request is conducted. The information stored in Data Management is called Descriptive Information (DI). The main search index will be based on Solr⁵.

Archival Storage is the component where all digital representations of the books are stored. This includes the metadata and all images of the scanned pages. The data stored in Archival Storage is called Archival Information Package (AIP). By using only AIPs it must be possible to recreate the whole system. Existing projects and software packages like Fedora⁶, DuraSpace⁷, ArchiveMatica⁸, iRods⁹ or HoneyCom¹⁰ may be used. The evaluation will be finished in M16.

The **Ingest** component is responsible for taking Submission Information Packages (SIP), extracting the DIs and creating AIPs from the SIPs and passing them to Data Management and Archival Storage, respectively. It is also responsible for validating the SIP and monitoring the status of the Ingest procedures.

² Drupal: <http://drupal.org/> [2010-04-14]

³ Locks: <http://lockss.stanford.edu/lockss/Home> [2010-04-14]

⁴ Schema Mapping Tool: <http://code.google.com/p/schema-mapping-tool/> [2010-04-22]

⁵ Solr: <http://lucene.apache.org/solr/> [2010-04-14]

⁶ Fedora: <http://www.fedora-commons.org/> [2010-04-14]

⁷ DuraSpace: <http://duraspace.org/index.php> [2010-04-14]

⁸ ArchiveMatica: http://archivematica.org/wiki/index.php?title=Main_Page [2010-04-14]

⁹ iRods: <https://www.irods.org/> [2010-04-14]

¹⁰ HoneyCom: <http://www.honeycom.com.au/> [2010-04-14]

Pre-Ingest represents a provider of the Open Archival Information System (OAIS). It will help the content providers (libraries, digitization centres, etc.) to harmonize the data before it will be ingested. This will be done on a directory basis and the usage of ontology services to enrich the metadata. Pre-Ingest will push the SIP into Ingest once it is created. Within the SIP Creator the Schema Mapping Tool can be used to map the local library data to BHL Europe. The entire Pre-Ingest process will be managed by experienced BHL-Europe team members to assist content providers to understand the process and to successfully finish the harmonization process.

All these listed components are part of the OAIS. To support the work of the content providers, an international book index, the **Global References Index to Biodiversity** (GRIB) was introduced. It will assign a **Global Unique Identifier** (GUID) to each item (e.g. book) that is already scanned or will be scanned. The communication between the OAIS components is defined within the OAIS reference specification however the GRIB provides additional information and systems that are detailed here.

Content providers transfer all their metadata to the GRIB for deduplication, digitisation management and to receive a GUID for each item that they intend to or have already scanned. BHL-Europe will then ingest all GRIB records to create a Functional Requirements for Bibliographic Records (FRBR) structure within its Data Management component. This could also be done manually, but because of the existing features the GRIB can provide, BHL-Europe will take it from there.

As soon as the providers are finished with scanning, they will upload their data to the Pre-Ingest component via FTP and use the SIP creator module to check for errors within the uploaded Pre-SIP via a web browser.

When the item is ingested, the Ingest module will update the GRIB and mark the record with the supplied GUID as already scanned.

4 Progress report on the development of the German prototype

This chapter describes the current status of the prototype and some features already available within it. The prototype is currently available at <http://bhl.ait.co.at> but will be moved to <http://prototype.bhl-europe.eu/> as soon as the servers are available. Figure 4-1 shows the home page of the prototype.

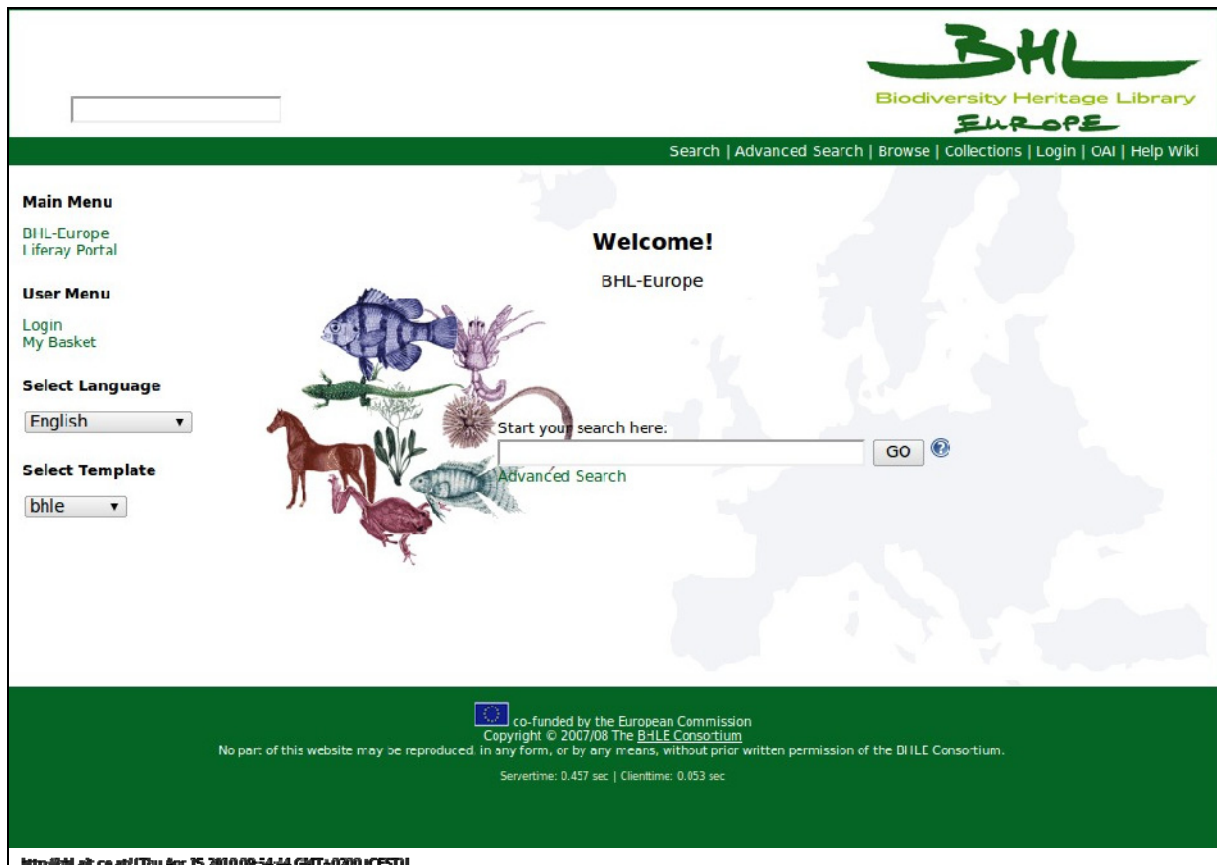


Figure 4-1 Home

4.1 Pre-Ingest

Pre-Ingest is used to harmonize data. Therefore a metadata mapping will be done prior to ingest. The currently manual procedure lets providers map their metadata fields to the predefined fields of BHL-Europe as shown in Figure 4-2 and Figure 4-4. These mappings are stored and configured within the system as part of Administration which can be seen at Figure 4-3. The current import is metadata only. This import will be shifted to GRIB for the final solution.

Search | Advanced Search | Browse | Collections | Logout | Admin Area | Thesauri WS | OAI | Help V

Import Mapping File No file chosen

Contact information

Contributing Institution

Name of Responsible Person

Email / Post Address

Phone Number

Note

ESE

BHL Field	Usage	Obligation	Your Field	Your Comment
Title [ese-Title]	Titel des Originalobjektes. (Buch, Video, Bild, Museumsobjekt etc.)	M	<input type="text"/>	<input type="text"/>
Title Alternative [ese-Alternative]	Alternativer Titel des Originalobjektes. (Buch, Video, Bild, Museumsobjekt etc.) Kann auch eine Abkürzung, oder Übersetzung sein.	O	<input type="text"/>	<input type="text"/>
Creator [ese-Creator]	Name des Verfassers, Urhebers oder Herstellers des Objektes in seiner aktuellen Fassung.	R	<input type="text"/>	<input type="text"/>
Subject [ese-Subject]	Thema des Inhalts in suchtauglichen Schlagwörtern.	R	<input type="text"/>	<input type="text"/>
Description [ese-Description]	Kurzzusammenfassung des Inhalts in Freitext. Das Objekt und seine Teile können beschrieben werden.	O	<input type="text"/>	<input type="text"/>
Table of	List of subunits of the original	O	<input type="text"/>	<input type="text"/>

Figure 4-2 Mapping

Search | Advanced Search | Browse | Collections | Logout | Admin Area | Thesauri WS | OAI | Help Wiki

Importer > Config

#	ID	Name of Responsible Person	Archive	Active
1	OAI_DC_IMPORTER	OAI Dublin Core Importer	*	No
2	OAI_GALLICA	OAI Dublin Core Gallica Importer	*	No
3	OAI_NATURALIS	OAI Dublin Core Naturalis Importer	*	No
4	MODS_IMPORTER	MODS XML Generic Importer	*	No
5	OAI_GDZ	OAI Dublin Core GDZ Importer	*	No
6	OAI_BHL_CITEBANK	OAI Dublin Core Citebank Importer	*	No
7	MARCYAZ_IMPORTER	MARC21 YAZ Importer	*	No
8	MARC_TEXT_IMPORTER	MARC Text Importer	*	No
9	INTERNETARCHIVE	BHL-US via IA Importer	*	No
10	NATURALIS_DIDL	Naturalis DIDL and PDF	*	No
11	NATURALIS_DIDL_XML	Naturalis DIDL and PDF from XML	*	No
12	BHLUS	BHL-US via item.txt	*	No
13	BHLUSMARC	BHL-US via item.txt in MARC	*	No
14	LBBXML	LBB importer	*	No

Total 14 rows: 1 | Items to display: 20

Figure 4-3 Mapping configuration

Importer > Edit

General | Reader | Mapper

@id OAI_GALLICA

@description OAI Dublin Core Gallica Importer

@active No

@archive *

@db OBJECTS

@indexerSave Yes

DefaultEncoding utf-8

ContactMail

Logfile oai_log isRelative: Yes

NewFilesDir pattern: isRelative: No

Repository isRelative: No

Importer > Edit

General | Reader | Mapper

Classname importparser.recordreaders.OaiRecordRea

verb ListRecords

metadataNamespace http://www.openarchives.org/OAI/2.0/oai_dc/

collection gallica.5

verb

Importer > Edit

General | Reader | Mapper

Classname importparser.mapper.bhle.OaiDc2BhleMap

lang fra

defaultCollection BNF

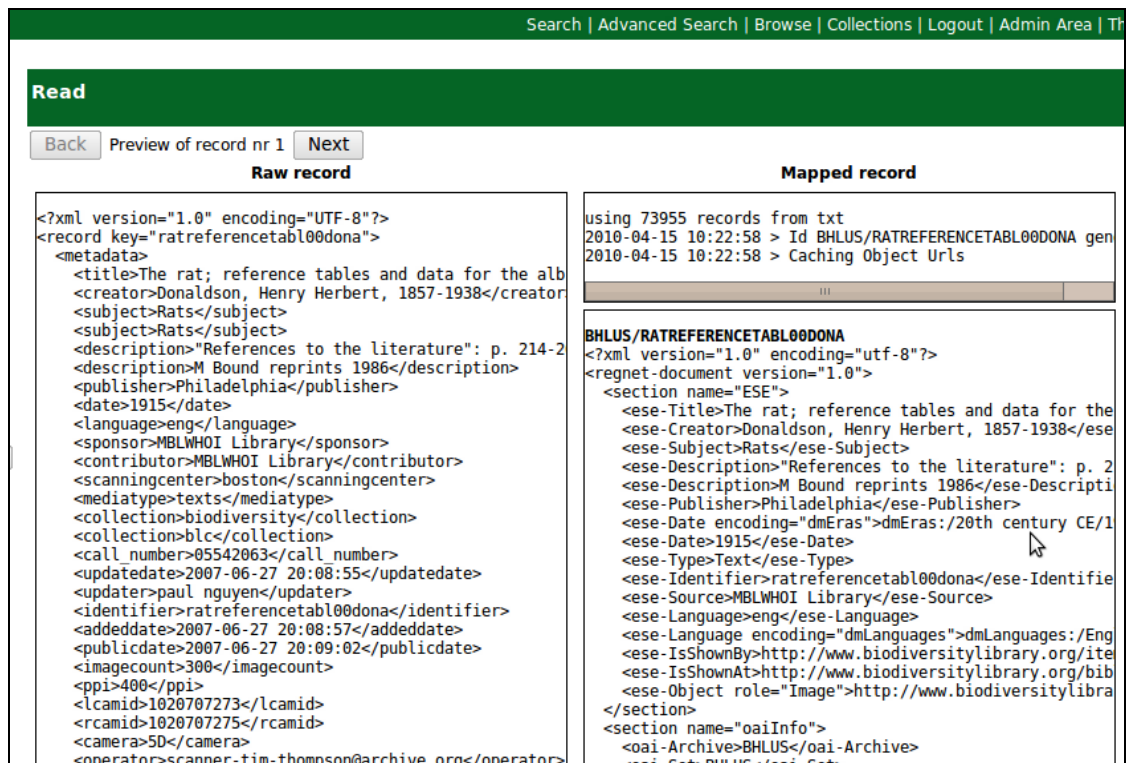
additionalSubFile

Figure 4-4 Edit configuration

4.2 Ingest

Ingest and Pre-Ingest are currently not separated. The conversion is done on demand which means there is no Pre-SIP in the prototype. Native data will be sequentially read and converted for Data Management.

The conversion can be checked before importing as shown in Figure 4-5.



The screenshot displays the 'Read' interface for a record. It features a navigation bar with 'Back', 'Preview of record nr 1', and 'Next' buttons. Below this, there are two main panels: 'Raw record' and 'Mapped record'.

Raw record: Shows the original XML metadata for a record with key 'ratreferencetabl00dona'. The metadata includes fields for title, creator (Donaldson, Henry Herbert, 1857-1938), subject (Rats), description (References to the literature: p. 214-2), publisher (Philadelphia), date (1915), language (eng), sponsor (MBLWHOI Library), contributor (MBLWHOI Library), scanning center (boston), mediatype (texts), collection (biodiversity), call number (05542063), update date (2007-06-27), identifier (ratreferencetabl00dona), public date (2007-06-27), image count (300), ppi (400), and camera (5D).

Mapped record: Shows the record after processing. It includes a header with 'using 73955 records from txt' and '2010-04-15 10:22:58 > Id BHLUS/RATREFERENCETABL00DONA gen'. Below this, it shows the XML metadata with enriched fields, including a section name 'ESE', an identifier 'BHLUS/RATREFERENCETABL00DONA', and various identifiers and dates.

Figure 4-5 Check SIPs prior to ingest

During the ingest, the data is enriched with multilingual thesauri terms. In the previous example, a language and a date thesaurus is used to unify records and make them searchable in multiple languages.

4.3 Data Management

Currently, the data management component can use either Zebra or Solr. The Zebra Index cannot be used due to the massive amount of records that BHL-Europe will face with the indexation of individual pages.

The Solr configuration will be tailored to fit BHL-Europe's needs.


4.4 Archival Storage

For the prototype, the Archival Storage will only cache one thumbnail for each item and store the raw and transformed metadata on a file server. This is neither safe nor meets the requirements of a backup and preservation system for the provider and therefore will be changed in the future.

4.5 Access

One of the Access methods for BHL-Europe is the OAI PMH provider for Europeana. An already working sample is shown in Figure 4-6. It uses a stylesheet transformation to present a user interface for the underlying XML.

OAI-PMH Response



[Identify](#) | [ListIdentifiers](#) | [ListMetadataFormats](#) | [ListRecords](#) | [ListSets](#)

You are viewing this page because you are using a browser that supports XSLT (XML stylesheet transforms), and the repository has indicated this page is to be displayed using the [Celestial](#) based XSL by [AIT](#)

responseDate	2010-04-15T08:58:05Z
request	http://bhl.ait.co.at/oai/index.php?verb=ListRecords&metadataPrefix=ese (validate)

ListRecords

The ListRecords verb provides metadata (e.g. Dublin Core) records for items that can be disseminated for the requested format, and optionally set membership and date restriction.

Header: [ListMetadataFormatsoai:eu.dismarc:BHLUS/RATREFERENCETABL00DONA](#)

timestamp	2010-02-09T07:47:43Z
setSpec	BHLUS

Metadata

```

<europeana:record xsi:schemaLocation="http://www.europeana.eu/schemas/ese/
http://www.europeana.eu/schemas/ese/ESE-V3.2.xsd">
  <dc:title>The rat; reference tables and data for the albino rat (Mus norvegicus albinus) and the Norway rat (Mus
  norvegicus)</dc:title>
  <dc:creator>Donaldson, Henry Herbert, 1857-1938</dc:creator>
  <dc:subject>Rats</dc:subject>
  <dc:description>"References to the literature": p. 214-267; other bibliographies
  interspersed</dc:description>
  <dc:description>M Bound reprints 1986</dc:description>
  <dc:publisher>Philadelphia</dc:publisher>
  <dc:date>20th century CE;1910s</dc:date>
  <dc:date>1915</dc:date>
  <dc:type>Text</dc:type>
  <dc:identifier>ratreferencetabl00dona</dc:identifier>
  <dc:source>MBLWHOI Library</dc:source>
  <dc:language>eng</dc:language>
  <dc:language>English</dc:language>
  <europeana:object>http://bhl.ait.co.at/templates/bhl/sets/sampledata/cachedimage.php?
  maxSize=200&filename=http://www.biodiversitylibrary.org/page/thumb/870086</europeana:object>
  <europeana:provider>BHL Europe</europeana:provider>
  <europeana:type>TEXT</europeana:type>
  <europeana:isShownBy>http://www.biodiversitylibrary.org/item/1568</europeana:isShownBy>
  <europeana:isShownAt>http://www.biodiversitylibrary.org/bibliography/1036</europeana:isShownAt>
</europeana:record>
          
```

Header: [ListMetadataFormatsoai:eu.dismarc:BHLUS/DIECOCCIDIENKRAN00PFEI](#)

timestamp	2010-02-09T07:47:46Z
setSpec	BHLUS

Metadata

```

<europeana:record xsi:schemaLocation="http://www.europeana.eu/schemas/ese/
          
```

Figure 4-6 OAI-PMH for Europeana Access

Another DIP is the RSS feed as shown in Figure 4-7. It can be used to stay up to date on certain queries.



Feed for BHL-Europe: Search result for ("BHL")
 Subscribe to this feed using: [Google Reader](#) [Subscribe Now](#)
 Always use this reader to subscribe to feeds.

Feed preview [\[Feed \]](#)

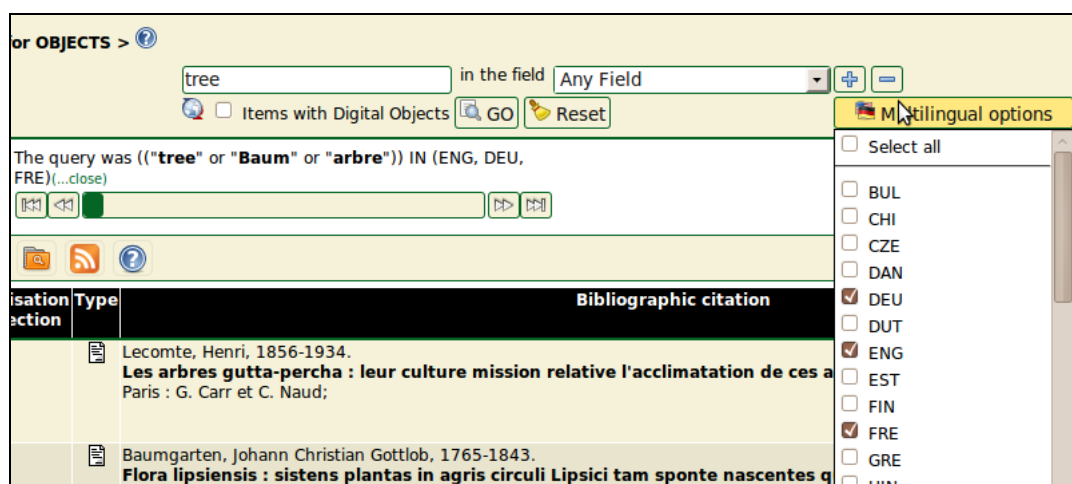
Rivierkreeften : proeftabel (versie 20 juni 2008)
 Title: Rivierkreeften : proeftabel (versie 20 juni 2008)
 Creator: Koese B.
 Subject: Arthropoda
 Subject: rivierkreeften
 Publisher: Stichting EIS-Nederland en Bureau Waardenburg
 Date Issued: 2008
 Date Issued: dmEras:/21st century CE
 Type: Text
 Extent: 17
 Identifier: oai:naturalis.nl:273798
 Source: Naturalis
 Is Shown By: <http://www.repository.naturalis.nl/document/107476>
 Is Shown At: <http://www.repository.naturalis.nl/record/273798>
 Rights: <http://purl.org/eprint/accessRights/OpenAccess>
 Object: <http://bhl.ait.co.at/data/naturalis/273798.jpg>

Het vliegend hert in Utrecht resultaten 2003
 Title: Het vliegend hert in Utrecht resultaten 2003
 Creator: Huijbregts J.
 Subject: Insecta
 Subject: Nederland
 Subject: Verspreiding
 Subject: Lucanus
 Subject: cervus
 Publisher: European Invertebrate Survey - Nederland
 Date Issued: 2003
 Date Issued: dmEras:/21st century CE
 Type: Text
 Extent: 7
 Identifier: oai:naturalis.nl:220036
 Source: Naturalis
 Is Shown By: <http://www.repository.naturalis.nl/document/46538>
 Is Shown At: <http://www.repository.naturalis.nl/record/220036>
 Rights: <http://purl.org/eprint/accessRights/OpenAccess>
 Object: <http://bhl.ait.co.at/data/naturalis/220036.jpg>

Figure 4-7 Syndication for ingested items

4.6 Portal

The Portal is the multilingual user interface for the Access component. One of the multilingual functions is a dictionary based query expansion as shown in Figure 4-8. It will look up the search term in a predefined dictionary and return alternative search terms in the desired languages.



or OBJECTS > ?

tree in the field Any Field + -

Items with Digital Objects [GO](#) [Reset](#) [Multilingual options](#)

The query was (("tree" or "Baum" or "arbre")) IN (ENG, DEU, FRE)(...close)

[Bibliographic citation](#)

Author	Type	Bibliographic citation
Lecomte, Henri, 1856-1934.		Les arbres gutta-percha : leur culture mission relative l'acclimatation de ces a Paris : G. Carr et C. Naud;
Baumgarten, Johann Christian Gottlob, 1765-1843.		Flora lipsiensis : sistens plantas in agris circuli Lipsici tam sponte nascentes q

Select all

BUL

CHI

CZE

DAN

DEU

DUT

ENG

EST

FIN

FRE

GRE

HIN

Figure 4-8 Query expansion with dictionary

Using this query, a multilingual result set is created at query time. An alternative is a multilingual result set based on the index. Therefore, the ontology services used to enrich the data during ingest can be used to retrieve a multilingual record set. This predefined ontology can be used as query parameter like shown in Figure 4-9.

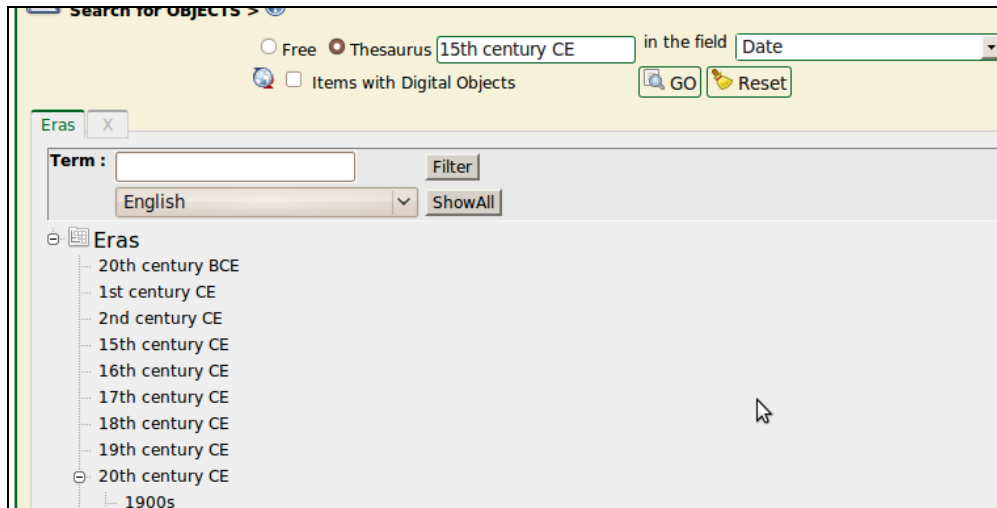


Figure 4-9 Ontology based search

The records that are enriched with ontology entries provide multilingual feedback on the search as shown in Figure 4-10.

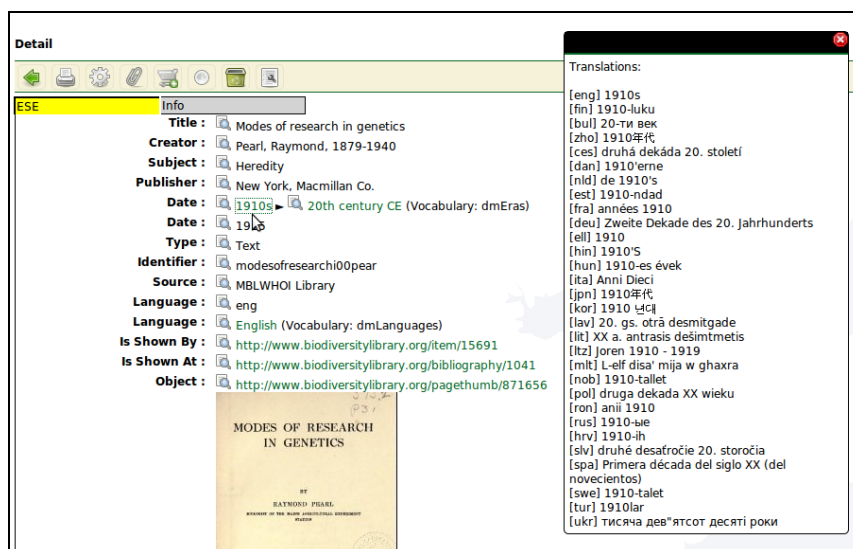


Figure 4-10 Ontology augmentation of records

Several other features like favourites, saved searches and a shopping basket are available as well. They are not described here in detail since this document's aim is to provide an overview on the core functionalities of the system.

5 Roadmap

This chapter drafts an outline for the tasks that need to be fulfilled in order to provide all desired future functionalities of the system.

5.1 Methodology

Since many functional requirements will be defined incrementally, the agile software development approach SCRUM was chosen for the distributed development. Scrum can be used within the PRINCE 2 methodology BHL-Europe uses.

5.2 Project backlog overview

This section provides an overview of the currently identified use cases for content providers

1. Description of the different processes
 - a. Catalogue building
 - b. Scanning planning
 - c. GUID - Mint Report - Project Template Generation
 - d. Scanning
 - e. Pre-Ingest
2. Description of use cases within each process:
 - a. Connecting to the GRIB
 - b. (Bulk) Search within the GRIB
 - c. (Bulk) Selection of literature to be digitised
 - d. Indicating scanning intention
 - e. GUID - Mint Report
3. Description of scenarios within each use case to define wire frames
 - a. Web access
 - b. Desktop application access

5.3 Sprints overview

The remaining 24 months have been split into 12 sprints each targeting a specific goal. The last two sprints are used to finalize the system without adding new features.

- M10 Sprint 0 Establish Development Team and Platform
- M12 Sprint 1 First mock-up integration GRIB + GUID
- M14 Sprint 2 Content provider - Pre-Ingest (upload files)
- M16 Sprint 3 PI – ING – AS – DM (1st ingest) + basic security
- M18 Sprint 4 DM-ACC (1st search – with a web service call not via the portal)
- M20 Sprint 5 DM - AS - ACC (1st dissemination DIP – ask for pdf / zip / ...)
- M22 Sprint 6 ACC - PORTAL (basic search retrieval)
- M24 Sprint 7 All module feature improvement + integration (alpha release)
- M26 Sprint 8 Administration + Finish Security
- M28 Sprint 9 Continue development of all modules
- M30 Sprint 10 Finish all modules (Beta version)
- M32 Sprint 11 Beta Testing and Bug Fixes (Feature Freeze)
- M34 Sprint 12 Final Testing and Bug Fixes
- M36 Go live Launch Release

6 Acronyms and Abbreviations

ACC	Access
AIP	Archival Information Package
API	Application Programming Interface
AS	Archival Storage
BHL	Biodiversity Heritage Library
CRC	Cyclical Redundancy Check
DI	Descriptive Information
DIP	Dissemination Information Package
DM	Data Management
ESE	Europeana Semantic Elements
ETL	Extract, Transform, Load
Europeana	European Digital Library
GRIB	Global References Index to Biodiversity
GUID	Global Unique Identifier
HTML	Hyper Text Markup Language
HTTP	Hyper Text Transfer Protocol
ING	Ingest
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
LOCKSS	LOCKSS (Lots of Copies Keep Stuff Safe), based at Stanford University Libraries, is an international community initiative that provides libraries with digital preservation tools and support so that they can easily and inexpensively collect and preserve their own copies of authorized e-content
MARC	MACHINE-Readable Cataloging
METS	Metadata Encoding and Transmission Standard
MIME type	Internet Media Type
OAI	Open Archives Initiative

OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OCR	Optical Character Recognition
ODBC	Open Database Connectivity
PDI	Preservation Description Information
PI	Pre-Ingest
RDF	Resource Description Framework
REST	Representational state transfer
RSS	Really Simple Syndication (RSS) is a lightweight XML format designed for sharing headlines and other Web content.
SIP	Submission Information Package
SOAP	Simple Object Access Protocol providing a simple and lightweight mechanism for exchanging structured and typed information between peers in a decentralised, distributed environment using XML proposed under the W3C.
SWORD	Simple Webservice Offering Repository Deposit
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
XML	EXtensible Markup Language
XSLT	XSLT (XSL Transformations) is a declarative, XML-based language used for the transformation of XML documents into other XML documents.

7 Figures

Figure 3-1 Architecture diagram.....	6
Figure 4-1 Home	9
Figure 4-2 Mapping.....	10
Figure 4-3 Mapping configuration	10
Figure 4-4 Edit configuration.....	10
Figure 4-5 Check SIPs prior to ingest.....	11
Figure 4-6 OAI-PMH for Europeana Access.....	12
Figure 4-7 Syndication for ingested items	13
Figure 4-8 Query expansion with dictionary	13
Figure 4-9 Ontology based search.....	14
Figure 4-10 Ontology augmentation of records	14