



Project no. 018340

**Project acronym: EDIT**

**Project title: Toward the European Distributed Institute of Taxonomy**

Instrument: Network of Excellence

Thematic Priority: Sub-Priority 1.1.6.3: "Global Change and Ecosystems"

## **C5.146 Import of freely available bibliographic metadata of digital resources from BHL to allow linking**

Due date of component: Month 52

Actual submission date: Month 53

Start date of project: 01/03/2006

Duration: 5 years

Organisation name of lead contractor for this component: 30 MFN

Revision: Final

<b>Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)</b>		
<b>Dissemination Level</b>		
<b>PU</b>	Public	X
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

## C5.146 Import of freely available bibliographic metadata of digital resources from BHL to allow linking.

The *Global References Index to Biodiversity* (GRIB) prototype<sup>1</sup> by the *Head office of the Common Library Network GBV* (VZG) is the joint literature index of EDIT's Virtual Taxonomic Library (ViTaL) and the Biodiversity Heritage Library for Europe (BHL-Europe). The GRIB has now been greatly enhanced by adding the bibliographic records from the *Biodiversity Heritage Library*<sup>2</sup> (BHL), including hyperlinks to the digitised items.

BHL provided its bibliographic data via their API in Dublin Core (DC). The data has been harvested by *Angewandte Informationstechnik Forschungsgesellschaft mbH* (AIT) – the BHL-Europe partner responsible for building the portal<sup>3</sup> – on 15<sup>th</sup> of May 2010. Metadata including hyperlinks from NNM<sup>4</sup> had than been added and all data had been transformed to the Europeana-Schema *Europeana Semantic Elements* (ESE) and provided to Europeana<sup>5</sup> in June.

For Example (See Fig. 1) “Robertson, J. Lynne & Eyles, Don E. : A guide and key to the aquatic plants of the southeastern United States. (PPN = 001693719)”. The first set of this metadata in the GRIB was provided by NHM and it was deduplicated and merged with a set of metadata including hyperlinks to digital item and hyperlink to BHL metadata, both provided by the MBLWHOI<sup>6</sup> Library.

The same data has been harvested from AIT by VZG at July 30<sup>th</sup> 2010. It has than been deduplicated with and merged to the existing data within the GRIB prototype at <http://grib.gbv.de/> by August 10<sup>th</sup>. Altogether 85.115 title records have been imported, 4.832 of those have been identified as duplicates (3.196 are duplicates with titles from the NHM<sup>7</sup>; 1.316 from NNM; 320 from FUB BGBM<sup>8</sup>).

Altogether the GRIB now holds 325.805<sup>9</sup> Titles and Articles, In comparison to the last update (156.370 objects) it shows an increase of 169.435 objects, which is more than a doubling of the data amount.

Some difficulties occurred while deduplicating and merging of the provided BHL data. This is due to the loss of original MARC<sup>10</sup> data when transformed to DC for harvesting. The original MARC data shown in the BHL portal is much more accurate. This problem has also been identified by BHL-Europe and Europeana, and the import of the original BHL MARC data will be considered and discussed as a possible solution.

---

<sup>1</sup> GRIB prototype Homepage <http://grib.gbv.de/>

<sup>2</sup> BHL Homepage: <http://www.biodiversitylibrary.org/>

<sup>3</sup> Prototype accessible at: <http://prototype.bhl-europe.eu/>

<sup>4</sup> Netherlands Centre for Biodiversity Naturalis (NCB Naturalis, Formerly: Nationaal Natuurhistorisch Museum Naturalis / Stichting Nationaal Natuurhistorisch Museum Naturalis)

<sup>5</sup> BHL content in Europeane [http://europeana.eu/portal/brief-doc.html?query=europeana\\_collectionName:087\\*](http://europeana.eu/portal/brief-doc.html?query=europeana_collectionName:087*)

<sup>6</sup> Marine Biological Laboratory and Woods Hole Oceanographic Institution

<sup>7</sup> Natural History Museum, London

<sup>8</sup> Freie Universitaet Berlin - Botanical Garden and Botanical Museum

<sup>9</sup> Search for all bibliographic data sets after BHL content import (ppn = 0?) results in 325.805 hits.

<sup>10</sup> Machine-Readable Cataloging

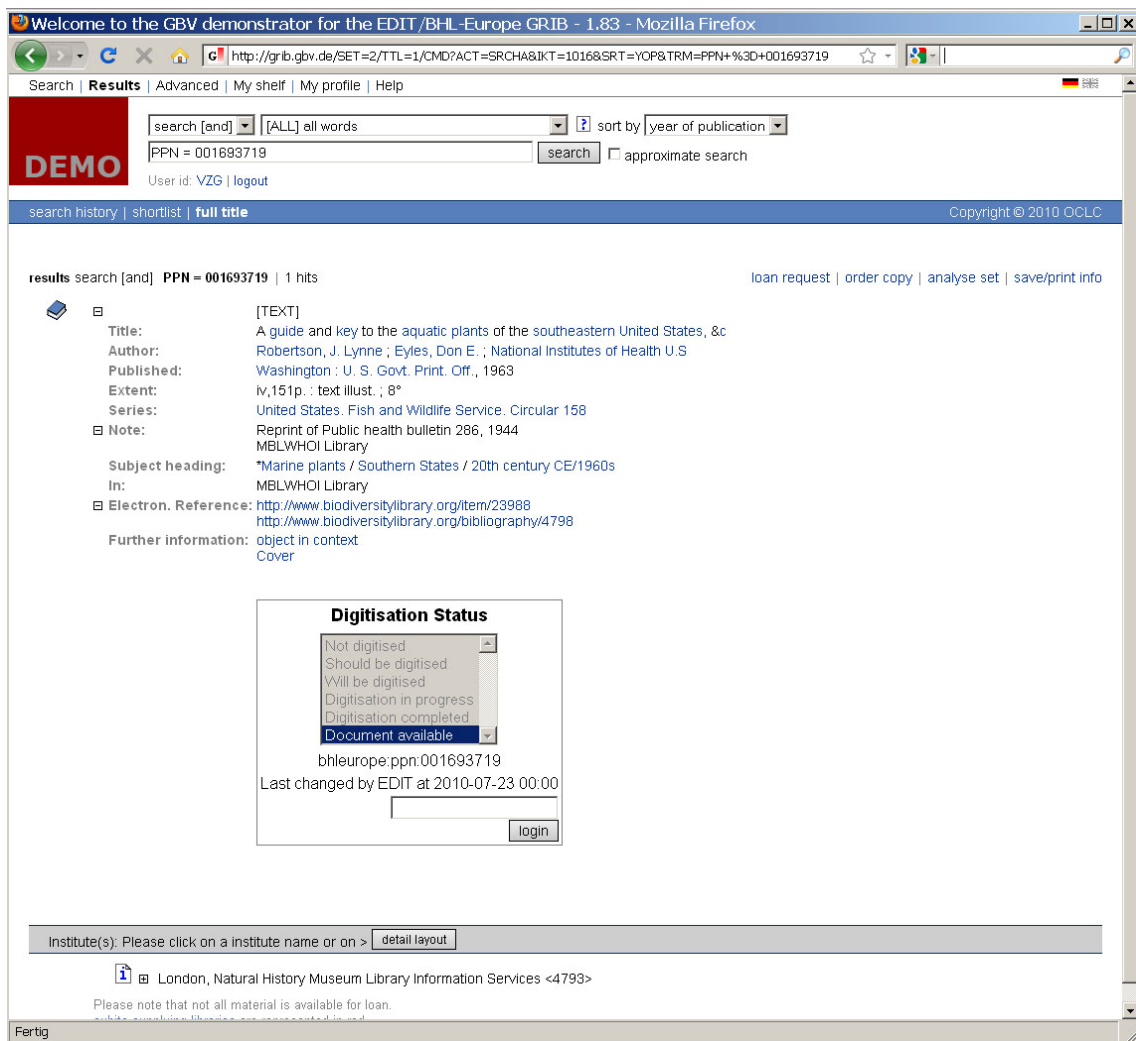


Fig. 1 Example set PPN = 001693719