

*Not yet public EDIT C5.146,
also confidential BHL-Europe Technical Note Sprint-Story 211:
“Import BHL data into GRIB” (Boris Jacob, MfN, AUG2010)*



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

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	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) prototype¹ by the *Head office of the Common Library Network GBV* (VZG) has been enhanced by the bibliographic records from the *Biodiversity Heritage Library*² (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 portal³ – on 15th of May 2010. Metadata including hyperlinks from NNM⁴ had than been added and all data had been transformed to the Europeana-Schema *Europeana Semantic Elements* (ESE) and provided to Europeana⁵ in June.

Welcome to the GBV demonstrator for the EDIT/BHL-Europe GRIB - 1.83 - Mozilla Firefox

Search | Results | Advanced | My shelf | My profile | Help

search [and] [ALL] all words sort by year of publication

PPN = 001693719 search approximate search

User id: VZG | logout

search history | shortlist | full title Copyright © 2010 OCLC

results search [and] PPN = 001693719 | 1 hits loan request | order copy | analyse set | save/print info

[TEXT]

Title: A guide and key to the aquatic plants of the southeastern United States, &c

Author: Robertson, J. Lynne ; Eyles, Don E. ; National Institutes of Health U.S.

Published: Washington : U. S. Govt. Print. Off., 1963

Extent: iv, 151p. : text illust. ; 8°

Series: United States. Fish and Wildlife Service. Circular 158

Note: Reprint of Public health bulletin 286, 1944
MBLWHOI Library

Subject heading: *Marine plants / Southern States / 20th century CE/1960s

In: MBLWHOI Library

Electron. Reference: <http://www.biodiversitylibrary.org/item/23988>
<http://www.biodiversitylibrary.org/bibliography/4798>

Further information: object in context
Cover

Digitisation Status

Not digitised
Should be digitised
Will be digitised
Digitisation in progress
Digitisation completed
Document available

bhleurope.ppn:001693719
Last changed by EDIT at 2010-07-23 00:00

login

Institute(s): Please click on a institute name or on > detail layout

London, Natural History Museum Library Information Services <4793>

Please note that not all material is available for loan.

Fertig

Fig. 1 Example set PPN = 001693719

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

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

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

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

⁵ BHL content in Europeana http://europeana.eu/portal/brief-doc.html?query=europeana_collectionName:087*

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 MBLWHOI⁶ Library.

The same data has been harvested from AIT by VZG at July 30th 2010. It has than been deduplicated with and merged to the existing data within the GRIB prototype at <http://grib.gbv.de/> by August 10th. 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 NHM⁷; 1.316 from NNM; 320 from FUB BGBM⁸).

Altogether the GRIB now holds 325.805⁹ 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 MARC¹⁰ 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.

⁶ Marine Biological Laboratory and Woods Hole Oceanographic Institution

⁷ Natural History Museum, London

⁸ Freie Universitaet Berlin - Botanical Garden and Botanical Museum

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

¹⁰ Machine-Readable Cataloging