



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"

C5.106 Implementation of an im-/export data interface to and from CDM

Due date of component: Month 53

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, FUB-BGBM

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)	

C5.106 Implementation of an im-/export data interface to and from CDM

The *Global References Index to Biodiversity (GRIB)* prototype (<http://grib.gbv.de/>) is being implemented by the Head office of the Common Library Network (VZG). For exporting data the GRIB offers a Search/Retrieval via URL interface (SRU) which allows CDM based EDIT platform components to find data in the GRIB and retrieve those in several standard formats. On the other hand the GRIB can harvest and import data from the CDM store via the Open Archives Initiative Protocol for Metadata Harvesting¹ (OAI-PMH).

The SRU interface is accessible at <http://kavia1.gbv.de/sru/DB=1.83>. Besides the standard search fields for Author, Title and Year – the specifications can be found at the Library of Congress² – there are some customized fields to be searched and retrieved:

- The Pica Production Number (PPN), a unique identifier for every set of bibliographic records within the GRIB.
- The Digitisation Status (DST), indicating the current stage of an item in the digitisation workflow (see Table 1 and C5.147 for specifications).

Digitisation Status	Status No.	Status Id
Not digitised	1	8300
Should be digitised	2	8301
Will be digitised	3	8302
Digitisation in progress	4	8303
Digitisation completed	5	8304
Document available	6	8305

Table 1 Values for the Digitisation Status

From a technical perspective the GRIB is functioning, and connection to CDM based EDIT platform components can be done. Data can already be imported from the GRIB in the following standards: MARC21, DC (Dublin Core), PICA, PICA short, UNIMARC and UNIMARC short.

The GRIB is still considered at a prototypical stage, in that the data consists of the test sets (see C5.145) together with data derived from BHL (see C5.146), and, that the data is mapped within the database to the Pica fields alone and not to DC fields yet (i.e. search queries should be done on the PICA fields, not on the DC fields). This will change as soon as more library catalogues and also data from the CDM store have been imported this autumn. The GRIB will import bibliographic and taxonomic data from the CDM store via the provided OAI-PMH interface (described in C5.143³).

The technical specification for the GRIB can be found at <http://bhleurope.gbv.de/>.

¹ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

² <http://www.loc.gov/standards/sru/>

³ C5.143 “Pilot implementation of integration with library catalogue and digital literature services“ (lead contractor for this component: 9 FUB-BGBM)