



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"

C6.07 To develop a migration path for the contents of the Scratchpads to the CDM

Due date of component: Month 29

Actual submission date: Month 31

Start date of project: 01/03/2006

Duration: 5 years

Organisation name of lead contractor for this component: 10 NHML

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)	

C6.07 To develop a migration path for the contents of the Scratchpads to the CDM

This document describes the migration path for data to be exported from the Scratchpads to the CDM, and to other sources.

A number of possible routes have been identified that allow data to be exported from the Scratchpads. These mainly build upon known standards, rather than using standards unique to the EDIT project. By using known standards, other consumers of biodiversity data should already be able to utilise the data that we are providing.

The following export data formats have been recognised as being potential candidates for exporting of data from the Scratchpads:

- DarwinCore (<http://wiki.tdwg.org/twiki/bin/view/DarwinCore/>). Although this data interchange format has not been fully ratified, it is considered as the *de facto* standard for exchange of specimen related data because of its use by GBIF.
- Taxonomic Concept Transfer Schema (<http://www.tdwg.org/standards/117/>). Although intrinsically quite complicated, the Taxonomic Concept Transfer Schema (TCS) can easily be simplified and used for the interchange of basic taxonomic data.
- Common Data Model XML exchange format. The Common Data Model (CDM), as developed by WP5 of the EDIT project, is expected to have a data interchange format based on XML for use by clients accessing the CDM. It has been proposed that the Scratchpads could provide data to the CDM using this interchange format.
- Tab separated text files. Exporting of data in a tab delimited file has been proposed as a simple, quick and easy way of exporting data to an external source. Use of tab delimited files is specific to the data being exported, and the source from which it is being exported. For these reasons, it is perhaps not advisable for widespread use.
- BibTeX and EndNote XML. Both of these data formats support the exporting of bibliographic data from the Scratchpads to external sources. Although EndNote is proprietary software, EndNote XML is an open standard and so can be easily interpreted by external sources.

As the Scratchpads develop, and their ability to handle different formats of data improves, other data interchange formats may well become suitable for the Scratchpads to use. These could include proposed standards like NeXML, or they could be already well used standards like FASTA.

Highlighted above are the interchange formats that we feel are important for use by the Scratchpads. Some of these are already fully implemented and in use, whilst the others are currently in development, or at the proposal stage. At present data are exported from the Scratchpads by:

- Export of bibliographic data from the Scratchpads to Vital via the Falx bibliographic data harvester. The Scratchpad bibliographies can be searched using the ViTaL search tool (<http://taxonlib.org/bibsearch/>).
- Export of specimen records from the Scratchpads to GBIF via TAPIR using the TapirLink tool. This is currently set-up for only one of our Scratchpads, however, set-up for additional Scratchpads can be done when requested.

- In partnership with the developers of LifeDesks (<http://lifedesk.eol.org>), code for exporting taxonomic data in the TCS format has been developed. The code for this has not been implemented on the Scratchpads, although we hope to have it there as soon as we updated components that the code requires.

It must be emphasised that data in the Scratchpads is owned by the community that created it. Tools to expose those data to external clients must be activated by the site Maintainers and cannot be centrally administered.