

An LSID Resolution Service for CATE

This document is a report describing the progress made on the contract between the Global Biodiversity Information Facility (GBIF), and the Department of Zoology, specifically the project entitled “Creating a Taxonomic e-Science (CATE)”.

The outputs of the project specified in the contract are as follows:

“Produce a working and documented implementation of an LSID Resolution Service, written in Java and integrated with the CATE web application.

[The service will implement the recommendations of the Life Sciences Identifiers Applicability Statement published by the Taxonomic Databases Working Group (TDWG). Each CATE web revision will resolve identifiers for taxon concepts published by the web revision. Metadata will be typed using the TDWG ontology. In particular, where an external LSID has been mapped onto the taxon name of a concept, the LSID of that taxon name will be returned as part of the metadata.]”

The deliverable can be divided into four areas: Software, Data, Documentation, and Implementation:

Software

This contract necessitated the creation of over 40 java classes, 31 of which can be found in the main modules of CATE, 7 of which are integration tests which verify that the components are working properly, and 6 of which form part of a standalone LSID demonstration web application.

Within the main modules, specific lsid-related classes are found within packages which follow the pattern `org.cateproject.*.lsid` or `org.cateproject.*.lsid.impl`. The software is written in Java and packaged as Maven 2 projects along with the rest of the CATE software. The software can be downloaded from:

The UK National e-Science Centre’s cvs repository (details here:

http://forge.nesc.ac.uk/scm/?group_id=51)

The CATE project site (here http://forge.nesc.ac.uk/project/showfiles.php?group_id=51)

or downloaded using the apache Maven software project management tool, by adding the following to your projects descriptor (pom.xml) file.

```
<project>
  ...
  <repositories>
    <repository>

      <id>cate-repository</id>
      <name>CATE Repository</name>
      <url>http://www.cate-project.org/repository</url>
    </repository>
```

```
</repositories>
...
</project>
```

The list of dependencies required to run each module can be found on the CATE project website (<http://www.cate-project.org>).

The software produced by the CATE team was based upon the reference LSID Java implementation version 1.1.2, available from Sourceforge (http://sourceforge.net/project/showfiles.php?group_id=130827&package_id=144604), and written by Ben Szekely of IBM. Changes were required to integrate this code with the CATE application, in particular because of the Spring Inversion-of-Control container used. This meant that:

- There was no need for application-specific plumbing code which can be found in the toolkit, since this is handled by the Spring container.
- Because CATE is structured into layers, some code was refactored, moving the code which is binding specific into the controller layer, and leaving the binding independent code in the service objects.
- In addition, authentication and authorization is handled as a cross-cutting aspect within CATE, following the philosophy that business objects should not know about such concerns. We replaced the LSIDContext object in the LSID Service method calls with an LSID, with the intention that security would be configured as necessary using AOP, removing the need for security context information to be passed to the LSID Services.

Data

In the course of this project we assigned LSID's to 11,160 Taxon Concepts across both exemplar web revisions. In addition, we mapped the names within CATE Araceae, which are derived from RBG Kew's Monocot Checklist to the names within the International Plant Names Index (IPNI), and return the IPNI lsid for the taxon name for 7429 of the names within CATE Araceae.

We return metadata for the objects identified by the LSID's we assigned in RDF following the schema published by TDWG for Taxon Concepts, including mappings for taxonomic relationships of the taxon.

Documentation

In addition to this document, documentation can be found on the CATE project website (<http://www.cate-project.org>), including generated API documentation (javadoc, <http://www.cate-project.org/apidocs/index.html>), and a developers guide (http://www.cate-project.org/Developers_Guide.html), which contains more information about the LSID Resolution Service implemented in CATE.

Implementation

One unexpected barrier to implementation is that the CATE project has been forced to move hosts from Imperial College London to the Natural History Museum London, mid

iteration, and it has consequently delayed the release by about seven weeks whilst server resource was found at the NHM, and the resources were configured. Currently CATE has deployed working final versions of the software (which can be seen at temporary urls <http://cate-dev.nhm.ac.uk> and <http://cate-live.nhm.ac.uk>). The only remaining task is the transfer of the domain name from IC to the NHM and the insertion of SRV records into the domain name server there. At this point we expect to have fulfilled the contract. This should take place over the next couple of weeks.