



Milestone M4.25

Pilot implementation of single access key webservice at BGBM

Leading partner: BGBM and UPMC

Compiled by: Lorna Morris

Date: September 2012

Pilot deployment of the ViBRANT Single Access Key Service in the EDIT Platform

Introduction

Xper2 is a platform that utilises structured descriptive data to annotate taxonomic data, for the purpose of taxonomic identification. It is developed by UPMC. UPMC have implemented a web service for producing single access keys from the exchange format for structured descriptive data SDD. The aim of this workshop and hackathon was to specify and develop a pilot implementation of the web service in the EDIT platform.

Minutes of the 2 day developer meeting at the BGBM

Andreas Müller, Andreas Kohlbecker, Lorna Morris, Alexander Oppermann (BGBM)
Thomas Burguiere, Nils Paule (UPMC)

AK and AM gave an overview of the EDIT platform. One of the main aims of the EDIT platform is to promote interoperability and enable software tools used to support the taxonomic workflow to be combined. The main features and components of the EDIT platform were outlined:

- Common Data Model (CDM) Library
Consisting of several layers: Model for the entire taxonomic domain, persistence, service layer and web service layer, import and export functionality
- CDM Server
Offers the web service functionality
- Taxonomic Editor
- Web applications - Drupal based web portals e.g. Cichorieae portal (<http://wp6-cichorieae.e-taxonomy.eu/portal/>)
- Print publishing – a pipeline to automate the generation of a print version of the taxonomy of different species in a variety of formats e.g. PDF, ODF, TaxPub.

We discussed CDM import and export functionality. Currently SDD can be imported into the CDM. It is easier to run this from a console than via the Taxonomic Editor as decisions on how the data is mapped to existing taxa must be made.

TB and NP gave an overview of Xper2 and Xper3. Xper3 is a web version of Xper2. They demonstrated the user interface and their proposed plans for further development.

There are 3 main views in the Xper3 interface:

- Taxa (Items)
- Descriptors (Characters)
- Description *i.e.* the view which allows the user to display and edit the state of a Descriptor for a given Item.

Two types of Descriptors are supported: Categorical (states e.g. large stripes as in Xper2)

Numerical (these have a min, max, standard deviation and mean)

The user can select which descriptors to associate to a Taxon.

We discussed the handling of statistical data for numerical characters and making use of a good unit system. Currently the units are stored as a text string and it is the user's responsibility to enter sensible data. AM and AK suggested making use of an ontology for units. QUDT (<http://www.qudt.org/>) may be a good choice.

A Taxon or a Character can have one or multiple images associated to it via a URL. It is proposed to have a separate database for sharing images and metadata (e.g. the licence). Currently a thumbnail is created and this is stored in the database. Possibilities to use Jakota open source image server and BioMediaWiki for permanent URLs were discussed.

The single access key web service developed by UPMC generates a key from an SDD file. Several parameters are available, in order to generate alternative keys in different output formats e.g. HTML.

Documentation and access to binaries and source code can be found here:

<http://www.identificationkey.fr/index.php/aboutTheWebservice>

We discussed the import of SDD into the CDM and generation of a key from this import. We discussed a more granular import where the user could choose to import a sub-set of items. Importing a sub-set of descriptors is more complex, descriptors could be grouped (e.g. by body parts into head, body) but descriptors can belong to multiple groups so this could be difficult to handle.

It is planned that Xper2 will still be maintained alongside Xper3. SDD can be exported from Xper2 and imported into Xper3. We discussed synchronisation problems: e.g. if a descriptive element is edited and another user moves a Taxon, the descriptive element may then be incorrect. AM suggested a refresh mechanism – if the taxa being viewed are updated in the database.

Xper3 and CDM used the same technologies – a service orientated architecture built on Java, Maven and Hibernate. However Xper3 is deployed in Tomcat and the EDIT platform use Jetty for deploying web services.

Specification of single access key web service in the EDIT Platform

We discussed whether to install the web service at the BGBM. It was decided that the easiest approach would be to use the Java based single access key API directly which is the base for the single access key web service. Therefore there is no need to build a web service ourselves, just deploy the jar file in our maven repository and include it into the CDM import/export layer. We want to integrate the service in the EDIT portals and in the Taxonomic Editor. AM is currently working on SDD export from the CDM but this isn't completed. Therefore to get data into the CDM we first tested the pilot implementation with an external SDD file.

We discussed time taken for parsing the SDD file and key generation. We tested the workflow with an SDD file containing 300 taxa and the time taken was 2 seconds for

parsing and 15 seconds for key generation. We discussed the workflow for the CDM (export SDD, generate the key, transform it to a CDM key). Generating the key on the fly would take too long therefore it was suggested that the key could be updated every night on CDM server.

As we're using an external SDD file (for the pilot implementation) this does not contain UUIDs for the taxa so the resulting key file will not have links to each Taxon but for SDD files created from CDM based data it will.

Every time UPMC release a new jar file it needs to be made available to our maven repository. We worked on easy deployment.

In order to get the SDD data into the CDM we used the iKey-Plus API to generate a key and then we mapped between the objects in the key to CDM objects for polytomous keys.

We persisted the generated key and wrote a unit test to check if this was successful. We used a SDD file containing 300 Cichorieae taxa and visualised the results in the portal. Currently media is still missing in the generated key. In the iKey-Plus model images can be attached to taxa and character states but not to the characters themselves. TB said that this could be implemented in the model. The CDM allows images for all 3 data types.

Key generation was successful and could be viewed in the portal. It took approximately 15 seconds to open the HTML page displaying the key. Performance for large Single-Access-Key visualization needs to be improved in the CDM dataportals. A screenshot showing the generated key is shown in Figure 1.



Search taxa

Misapplied names

Search

[Advanced Search](#)

Taxon tree

Available classifications:

standard view

Amaranthaceae

Cichorieae

Identification Keys

Polytomous

- Project: Cichorieae
- Project: Cichorieae
- Project: Cichorieae
- Key to Crepis species

Key to Crepis species

1	rosette leaves : glabrous	2
1'	rosette leaves : pubescent	34
1''	rosette leaves : puberulous	69
1'''	rosette leaves : floccose	<i>Crepis hookeriana</i> Ball
1''''	rosette leaves : woolly	77
1'''''	rosette leaves : scabridulous	<i>Crepis urundica</i> Bab.
1''''''	rosette leaves : subglabrous	79
1'''''''	rosette leaves : glabrescent	82
1''''''''	rosette leaves : hispid	100
1'''''''''	rosette leaves : hispidulous	108
1''''''''''	rosette leaves : tomentose	120
1'''''''''''	rosette leaves : hirsute	<i>Crepis vesicaria</i> L.
1''''''''''''	rosette leaves : glandular	142
1'''''''''''''	rosette leaves : pilose	<i>Crepis fraasii</i> Sch. Bip.
2	growth form : herb	3
2'	growth form : rosette herb	30
3	involucre : campanulate	4
3'	involucre : cylindrical	12
3''	involucre : turbinate	14
3'''	involucre : cylindric-campanulate	15
4	pappus : persistent	5
4'	pappus : caducous	7
5	outer involucral bracts : acute	<i>Crepis viscidula</i> Froel.
5'	outer involucral bracts : acuminate	6
5''	outer involucral bracts : obtuse	<i>Crepis aurea</i> (L.) Cass.
6	anther tube apical appendages : acute	<i>Crepis chondrilloides</i> Jacq.
6'	anther tube apical appendages : obtuse	<i>Crepis baldaccii</i> subsp. <i>albanica</i> Jáv.

Figure 1: Screenshot of the Cichorieae test portal showing the key generated using the iKey-Plus single access key web service.

In the second afternoon of the meeting we discussed the upcoming WP4 ViBRANT milestones. UPMC have implemented an applet for Xper2. For the M4.29 milestone (due Nov 30th) CDM SDD output needs to be optimised for Xper2 Applet. We plan to complete the export of SDD from the CDM and import the SDD data into the Applet, then use the applet in our CDM portals.

Further discussion is required to decide which data would be exported, and where to store the resulting SDD file. There are several milestones due in the middle of next year (e.g. M4.34). We briefly discussed the aggregation of species level data to a higher Taxon level. This is complex e.g. how do we aggregate data for combined states (leaf colour red and length short, leaf colour yellow and length long). These will require further discussion and another hackathon could be useful.

Pilot implementation of the web service

The source code for the pilot implementation is available from EDIT's subversion repository:

<http://dev.e-taxonomy.eu/svn/branches/cdmlib/ikey-plus-import/>