



Indexing for Life

Deliverable D12.1

CoL Placement Service (demonstrator report)

Workpackage 12

Viktoras Didziulis, Kwok Yin Cheung

University of Reading

31-10-2012

Capacities Programme of Framework 7: EC e-Infrastructure Programme – Virtual Research
Communities - INFRA-2010-2

Grant Agreement No:	261555
Project Co-ordinator:	Dr Alastair Culham
Project Homepage:	http://www.i4Life.eu
Duration of Project:	36 months
Start Date:	November 2010
End Date:	November 2013



Contents

Introduction	3
The Supply Side Pipeline	4
The Distribution Side Pipeline.....	5
Process Monitor	6
Conclusion.....	6

Introduction

WP 12 has prototyped, made available for test, and implemented the Taxon & Name Placement Service components for ongoing usage in collaboration with WP 2. They are:

- i) Taxon & Name Supply-side Pipeline. (Task 12.1) - a protocol and procedure for pulling a data set containing particular batch of supplied name-records and taxon-records thought to be absent from the CoL, with their associated higher taxon tags, into the reception buffer for the GSD-piping tool.
- ii) Taxon & Name Distribution-side Pipeline. (Task 12.2) - a tool set for serving supplied Names & Taxa from the CoL Reception Buffer to the appropriate GSD system, and managing this process.
- iii) Processing Monitor. (Task 12.3) - a protocol and service for a) partitioning and distributing incoming data elements between the GSD-sectors maintained taxonomically by different GSD organisations within the Sp2000 GSD Network, and b) making the records of successful and rejected name and taxon positioning tasks available for collection by the instigator.

Access details are provided to the Commission in a separate document.

The Supply Side Pipeline

The CoL Piping Tools Supply Side interface is available at:
<http://www.catalogueoflife.org/piping/> (password protected).

User instructions are given on the front page of the service.

Note: name lists in user accounts have been frozen and cannot be changed because they are in use by pilot projects.

Piping Tools - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Catalogue of ... i4Life WP4 En... Piping Tools x Cross-mappin... Documents a... DOW i4Life (2... Index of /pipi... +

www.catalogueoflife.org/piping/help.php

i4Life Piping Tools Help

Login: *admin*

Help

- [Manually Uploading](#)
- [DwC-A Uploading](#)
- [BOLD Uploading](#)
- [X-map Uploading](#)
- [Uploaded Data](#)
- [Data Summary](#)
- [User List](#)
- [Change Password](#)
- [Piping Tools Settings](#)
- [Sign Out](#)

Introduction
 This is the supply side Piping Tools interface. The menus of the tools interface are located on the left of the page. To select a menu, just click on them.

Manually Uploading
 This page is for uploading cvs data file directly into taxa table. Use "Data Columns Setting" to select fields which match your taxa table for upload. The order of the field selected should be the same as your table. The "Default Setting" would preset the following fields: "Taxon ID, Genus, Specific epithet, Scientific name authorship, Intraspecific epithet, Verbatim taxon rank, Family, Order, Class, Phylum, Kingdom". Use the "Clear" button to clear all selected fields. Use the radio buttons to select source of the upload data: "File" for local file on your PC; "URL" for data which can be read on a remote server; "File on server" this option is for "admin" of the Piping Tools only. In the "Other Options" section, you can choose to ignore the first row of data for you upload as well as select different field delimited characters. The "Tag for this upload" is an optional input which will tag this particular upload with the given tag name, if no tag name is given, the program will generate a unique tag automatically which is comprised of year, month, data and the starting internal ID for the uploading records. After you have set all the correct options, you then press the "Upload" button to start the upload process. After the data been uploaded, you would be automatically re-directed to Uploaded Data page.

DwC-A Uploading (Admin Only)
 This page has four buttons for uploading data directly from "CBOL, ENA, GBIF and IUCN" via DwC-A format. URLs and data format transformation information are already stored on the systems database. To upload data from any of the four data providers, only need to press the corresponding button. If the data file is big, e.g. has more than a few hundreds of thousand records, it might take sometime for the tools to process the data.

BOLD Uploading (Admin Only)

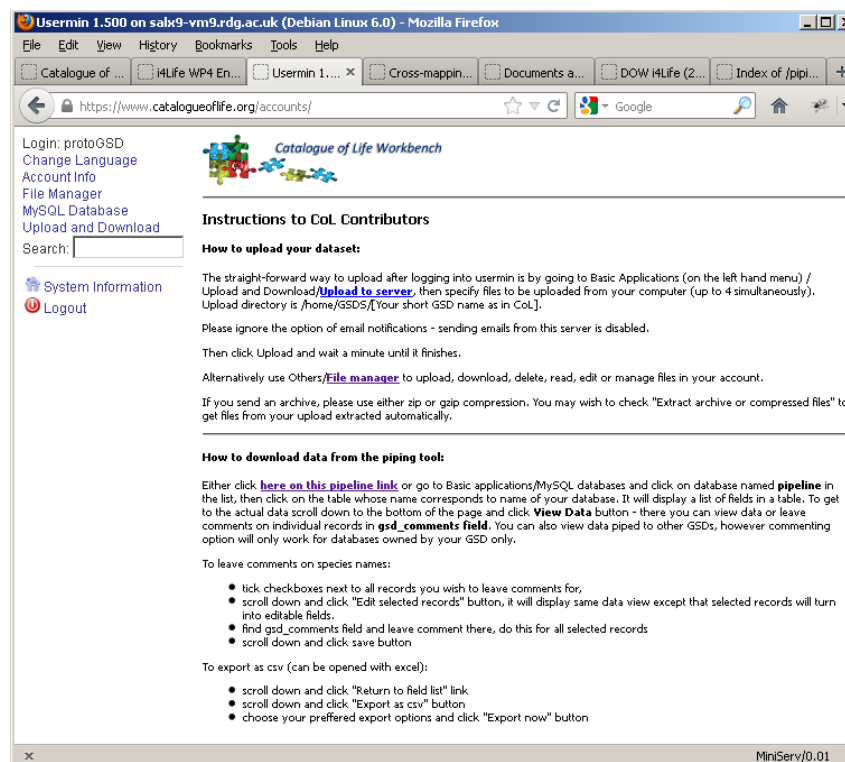
Apache/2.2.16

A document with WP12 requirements for data input was distributed to Global Partners in August 2011 (a few minor updates followed later) and is available online at http://i4life.eu/i4life/uploaded/Directives%20for%20implementation%20of%20the%20supply%20side%20piping_v2_2.pdf.

The Distribution Side Pipeline

The CoL Piping Tools Distribution Side download service is available at <https://www.catalogueoflife.org/accounts> as the system is integrated with the CoL Workbench 4D4Life. Please ignore browser warnings when accessing the resource for the first time - we use a self signed SSL certificate for encryption, not identification, purposes only. There is a password protected account which does not belong to any particular GSD and can be used to explore what is inside. Information on usage is provided on the “Account Info” (front) page. GSD can download an entire dataset by following a link to pipeline, selecting his table and clicking “Export as CSV” button (scroll down to see it). An alternative way to download datasets without logging into user’s accounts is provided by the Process Monitor webservice described further below.

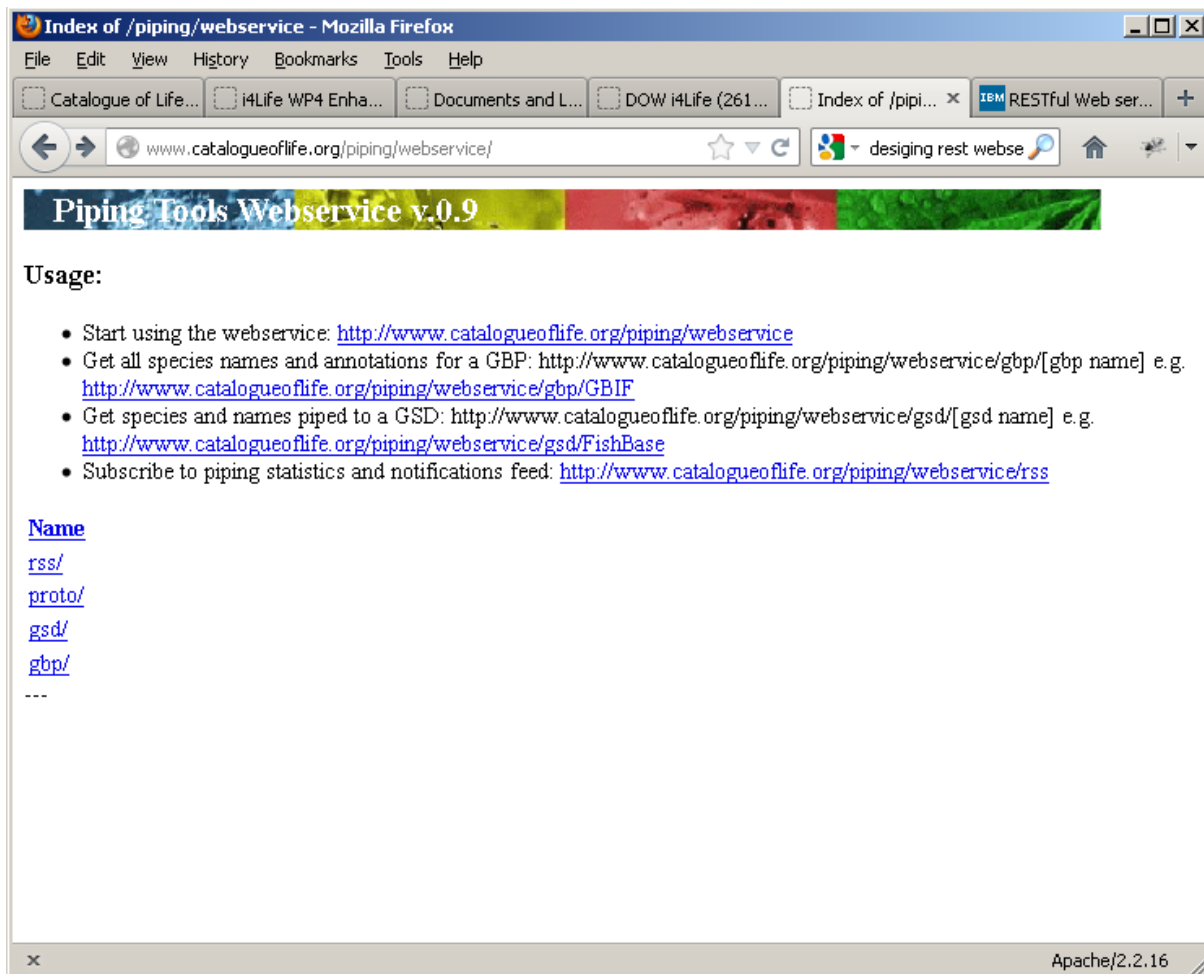
Datasets are editable by GSDs to enable them to leave annotations by choosing their dataset, clicking “View Data” button at the bottom of the page, checking rows he wants to annotate, clicking “Edit selected rows” and writing his opinions into a `gsd_comments` field and name acceptance status in “`gsd_status`” field and clicking “Save” button. GSDs can only annotate own tables e.g. protoGSD can only annotate protoGSD table, otherwise the system will throw an error. However if GSDs wish to send us annotated lists done on their side, then they would need to replace the old table with an annotated table and loading it by following the link to the pipeline, clicking “Execute SQL” button (please scroll down if not visible), choosing “Import text file” tab, specifying Table to import data into, checking “Yes” next to Delete data in table first, choosing file format (csv with quotes). The GSDs can also ask data managers in Reading to help with this.



A document explaining the system has been distributed to GSDs and is available online at http://i4life.eu/uploaded/Roadmap_for_i4Life_pilot_projects.pdf.

Process Monitor

The Process Monitor provides access to rss notifications with some statistics and to the Distribution Side Pipeline datasets at <http://www.catalogueoflife.org/piping/webservice/>. It also provides a minimal GUI so that end user can explore what is inside before configuration of their software to access specific resources. Usage instructions are provided on the front page of the webservice.



Note: there is an additional “wish lists” service used to provide lists of names for taxonomic groups that are not currently covered by the CoL, to enable applications for new GSD sectors via the pilot projects at <http://www.catalogueoflife.org/piping/webservice/proto/>. Although the wish-lists are presented via the piping tools and have been generated using piping tools infrastructure.

Conclusion

The piping tools have been prototyped, tested and used to produce species name lists for the pilot projects. The tools have been deposited into our software repository and subversion (SVN) at ETI and are being prepared for production in WP 4.