D5.3 - Sustainable software services

Date	29/11/2013
Reporter	Neil Caithness
Description	Refined and sustainable software services available for public use with mechanisms to measure usage rates.

1   Summary

The objective for WP5 was to provide seamless integration of relevant external computing services for biodiversity researchers and Scratchpad users. Success is to be measured against this final deliverable. In this report I list the overall achievements of WP5 as a set of functional and sustainable products.

The principle product is OBOE, The Oxford Batch Operation Engine, for which I provide a more detailed description and present usage statistics.

Whereas OBOE is explicitly for batch operation services, a number of other software applications were developed within WP5, but separately from OBOE as they fall into the category of interactive services. (See Xper², GeoCAT, and CartoDB below.)

2   Products

WP5 has delivered the following products:

a) **OBOE**, The Oxford Batch Operation Engine
https://oboe.oerc.ox.ac.uk/, see also https://oboe.oerc.ox.ac.uk/docs

In a paper submitted to CCGrid2014 we describe the architecture and design decisions we made in building OBOE:

Caithness, Neil, and Milo Thurston. *A simple drop and compute model for a SaaS cloud infrastructure for advanced research computing: the experiences of building OBOE, the Oxford Batch Operation Engine*. In CCGrid2014, the 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Edited by Esmond G. Ng. Chicago, Illinois, USA: IEEE Computer Science Press; IEEE Digital Library, 2014. (Submitted on 18 November, 2013; Notification expected 24 January, 2014)
http://vbrant.eu/sites/vbrant.eu/files/ccgrid2014_submission_177_2.pdf

OBOE is a general platform for research computing and currently makes the following batch services available:

Phylogenetics

i.   **Mr Bayes** (Community application for Bayesian Inference of Phylogeny, http://mrbayes.sourceforge.net/. Deployed on ARC, Oxford Advanced Research Computing facility.)

ii.   **BEAST** (Community application for Bayesian MCMC analysis of molecular sequences, http://beast.bio.ed.ac.uk/Main_Page. Deployed on ARC, Oxford Advanced Research Computing facility.)

Sequence Analysis

iii. **MUSCLE** (Community application for multiple alignment of molecular sequences, http://www.drive5.com/muscle/. Deployed on ARC, Oxford Advanced Research Computing facility.)

iv.   **Msatfinder** (Community application for the identification and characterization of microsatellites in a comparative genomics, http://web.archive.org/web/20071026090642/http://www.genomics.ceh.ac.uk/msatfinder/. Previously hosted by CEH, but now hosted by UOXF.)

Taxonomy

    v.   **IKey+** (Generates an identification key, http://www.identificationkey.fr/)

    vi.   **GoldenGate** (A service produced by WP7 to convert OCR'ed biosystematics documents into XML content marked up in the TaxonX XML schema, http://idaho.ipd.uni-karlsruhe.de/GoldenGATE/)

Ecological Informatics

    vii.   **LEFT – The Local Ecological Footprinting Tool** (Produces an ecological valuation report for a specified terrestrial area. It is intended as a planning tool and quickly highlights areas of high ecological value.)

    viii.   **LCC – Land Cover Change** (Produces a report for a specified area showing an analysis of land cover change between 2005 and 2009 using the Globcover products of the ESA (http://due.esrin.esa.int/globcover/)

    ix.   **CDE – Climate Data Extractor** (Provides data extraction and format conversion for the WorldClim data set http://www.worldclim.org/)

    x.   **SDM – Species Distribution Modelling** (A pipeline service that starts with a list of species names and produces stacked SDM models for the analysis of regional alpha and beta diversity.)

    xi.   **BICT – The Biological Index Calculation Tool** (An index calculation tool produced by WP8, see M8.20 - Ecological and conservation applications implemented as a service.)
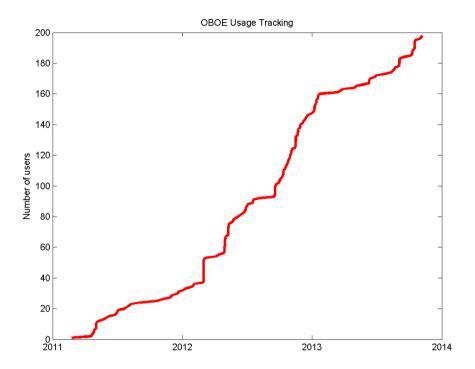
Miscellaneous

    xii.   **DURDEN** (Community application for large image tiling used by Scratchpads, https://git.scratchpads.eu/v/durden.git)

    xiii.   **ZooQuery** (An ad. hoc. service for the Zooniverse project allow researchers to investigate anonymised data about project users, https://www.zooniverse.org/project/planethunters.)

b) **GNRD**, Global Names Recognition and Discovery http://gnrd.oerc.ox.ac.uk/
(Note that GNRD itself was not developed by WP5, and is currently also available at its original site http://gnrd.globalnames.org/. However, ongoing funding for globalnames.org is not secure so, on the request of WP7, UOXF is hosting a duplicate of the service in order to ensure that it remains available in the future.)

c) **Xper²**, Tool for Managing and Editing Taxonomic Descriptions
http://www.infosyslab.fr/lis/?q=en/resources/software/cai/xper2/downloads/last

d) **GeoCAT**, Geospatial Conservation Assessment Tool http://geocat.kew.org/

e) **CartoDB**, Geospatial Visualization http://cartodb.com/
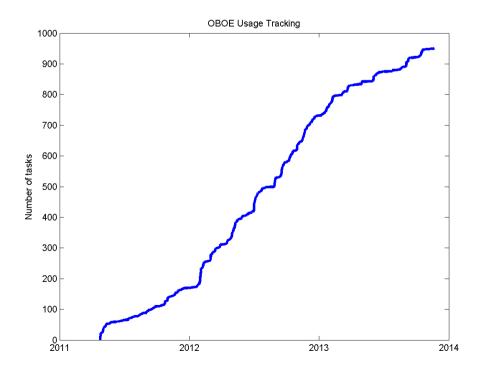
3   OBOE Usage Tracking

Use of OBOE has grown remarkably consistently over the three years of the ViBRANT project and we fully expect to continue building a solid community of users.

See below for cumulative plots of the number of users and tasks, and also of the global and European distribution of users. These plots are periodically updated and are available on the ViBRANT website http://vbrant.eu/content/wp5-data-interaction-and-services
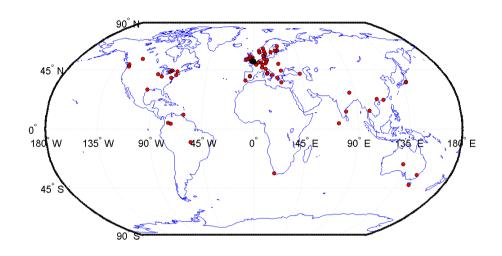
a) Number of Users

b) Number of Tasks

c) Global distribution of users

d) European distribution of users