



**ViBRANT**  
*Virtual Biodiversity*

## *FRONT PAGE*

# PROJECT PERIODIC REPORT

**Grant Agreement number:** 261532

**Project acronym:** ViBRANT

**Project title:** Virtual Biodiversity Research and Access Network for Taxonomy

**Funding Scheme:** Research Infrastructures

**Date of latest version of Annex I against which the assessment will be made:**

<b>Periodic report:</b>	<b>1<sup>st</sup></b>	<b>2<sup>nd</sup></b>	<b>3<sup>rd</sup></b>	<b>4<sup>th</sup></b>
<b>Period covered:</b>	from 1 Dec 10		to 30 Nov 11	

**Name, title and organisation of the scientific representative of the project's coordinator:**

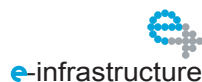
Dr Vince Smith,  
Natural History Museum,  
Cromwell Road,  
London SW7 5BD  
Great Britain

**Tel:** +44 (0) 207 942 5127

**Fax:**

**E-mail:** vince@vsmith.info

**Project website address:** <http://vbrant.eu>



## Declaration by the scientific representative of the project coordinator

I, as scientific representative of the coordinator of this project and in line with the obligations as stated in Article II.2.3 of the Grant Agreement declare that:

- The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;
- The project (tick as appropriate)<sup>1</sup>:
  - ☐ has fully achieved its objectives and technical goals for the period;
  - ☐ has achieved most of its objectives and technical goals for the period with relatively minor deviations.
  - ☐ has failed to achieve critical objectives and/or is not at all on schedule.
- The public website, if applicable
  - ☐ is up to date
  - ☐ is not up to date
- To my best knowledge, the financial statements which are being submitted as part of this report are in line with the actual work carried out and are consistent with the report on the resources used for the project (section 3.4) and if applicable with the certificate on financial statement.
- All beneficiaries, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SMEs, have declared to have verified their legal status. Any changes have been reported under section 3.2.3 (Project Management) in accordance with Article II.3.f of the Grant Agreement.

Name of scientific representative of the Coordinator: Dr V. Smith

Date:

Signed by Vince Smith

For most of the projects, the signature of this declaration could be done directly via the IT reporting tool through an adapted IT mechanism.

---

<sup>1</sup> If either of these boxes below is ticked, the report should reflect these and any remedial actions taken.

### 3.1. Publishable summary

The goal of the ViBRANT project (<http://vbrant.eu>) is to facilitate the mobilisation, sharing, reuse and publication of biodiversity data, multiplying the investments already spent by society in the collection of those data and helping to focus the output from the European and global biodiversity research community. Our ambition is to increase massively the number of people actively using e-Infrastructures to publish their biodiversity data for use by all.

The basic mechanism chosen to achieve this goal is to consider the end-to-end workflow in the generation of biodiversity data. The central system we use is Scratchpads (<http://scratchpads.eu>), which provide a social network and a primary user interface. We have taken the decision to upgrade the core software from Drupal 6 to Drupal 7 which means that there has been little change in the Scratchpads this year as far as users are concerned. The new software environment will be launched in early 2012.

Scratchpads are a system by which individuals can apply for a web site in any domain related to natural history. We supply an environment that supports natural history data and is, by default, based around a biological taxonomy. Although an individual applies, and takes responsibility, for a site, the intention was that the sites be used to support a community working to the same aims. Data entered into a Scratchpad remain the property of the data creator and, by default, are published under a Creative Commons licence. The data are stored in a way that makes exchange of data straightforward, for instance with the Encyclopedia of Life and GBIF. Under ViBRANT we will be offering tools that allow operations on data to enhance their presentation. Generally these tools will operate through a standard interface called OBOE (<https://vibrant.oerc.ox.ac.uk/>) which will include making identification keys (<http://identificationkey.fr>). Another tool that has been launched this year is GeoCAT (<http://geocat.kew.org>) that performs rapid geospatial analysis to ease the process of Red Listing taxa. Other services will follow in the coming years.

The primary unit of production on which scientists are assessed is the publication. ViBRANT, through Pensoft (<http://www.pensoft.net/>), have improved the overall workflow that makes preparation of manuscripts easier and their publication both faster and cheaper. The process also includes output in 4 distinct forms: full-colour print version; PDF; HTML; and XML. Given the volume of scientific information now being published, the primary readership is actually computers which select relevant articles for human readers. Machine-readable versions of output make the intelligent association of concepts possible, so in addition to search systems finding publications that are relevant to a particular question, the data can be extracted directly for re-use. Papers in Pensoft journals are now delivering information on new taxa to the Encyclopedia of Life; keys metadata to KeyCentral; taxon treatments to Species ID; all articles are archived in PubMed Central and articles are indexed in the Directory of Open Access Journals (DOAJ), all of which enhances their discoverability. Pensoft have also planned the launch next year of a data journal, i.e. dedicated to description of data sets that give the author a peer-reviewed, citeable publication for their career progression. Finally Pensoft published a special volume ( ZooKeys **150** Special Issue *e-Infrastructures for data publishing in biodiversity science*) featuring many ViBRANT products (<http://www.pensoft.net/journals/zookeys/issue/150/>).

The inclusion of marine systems, especially those within reach of recreational divers, have been under-represented in citizen-scientist initiatives. ViBRANT has developed a project and web site (<http://www.comber.hcmr.gr/>) to address that shortcoming. With outreach activities, diver training and active participation from recreational-diving businesses, primarily over the two months of the summer season for which the project was available (July–August 2011), 56 users (excluding the four supervising scientists) took part. Twenty of the users contributed data from more than one dive or snorkelling trip and thus expanded the sampling area beyond the training sites to several other locations in Greece. 2,030 species observations were recorded during 147 dives and 39 snorkelling trips. We look forward to this year's diving season.

## 3.2. Core of the report

### Project Partners

Partner	Name	Acronym
1	The Natural History Museum, London	NHM
2	Hellenic Center For Marine Research, Crete	HCMR
3	Royal Belgian Institute of Natural Sciences	RBINS
4	Oxford e-Research Centre	UOXF.E9
5	Vrije Universiteit Amsterdam	VU
6	Julius Kühn-Institute	JKI
7	Museum für Naturkunde, Berlin	MfN
8	University of Amsterdam	UvA
9	The Open University	OU
10	Karlsruher Institut für Technologie	KIT
11	Vizzuality	Vizz
12	PENSOFT Publisher	PENSOFT
13	Université Pierre et Marie Curie-Paris 6	UPMC
14	Global Biodiversity Information Facility	GBIFS
15	Freie Universität Berlin	FUB-BGBM
16	Universite de la Reunion	UdIR
17	Università di Trieste	UNITS

### 3.2.1. Project objectives for the period

ViBRANT's objectives are set out below, as defined in Annex 1 to the Grant Agreement. The objectives are not sequentially dependent and are therefore not specific to particular years. This is particularly important in the field of biodiversity informatics which benefits from several other large projects, such as BHL and EoL, making the landscape rather dynamic. ViBRANT's overall goal is to improve linkage between on-line resources with the ambition that data need be entered only once.

### Goal

ViBRANT will facilitate the mobilisation, sharing, reuse and publication of biodiversity data, multiplying the investments already spent by society in the collection of biodiversity data and helping to focus the output from the European and global biodiversity research community. Our system will massively increase the number of people actively using e-Infrastructures to publish their biodiversity data for use by all.

### Context

ViBRANT will deploy end-to-end e-infrastructure services and tools, capturing the lifecycle of biodiversity research from inception to publication in an open digital environment. ViBRANT is user orientated, enabling multidisciplinary groups to create their own virtual research communities supporting biodiversity science.

## Approach

ViBRANT will bring the operational capability of existing, EU funded infrastructures and services to a new level. The consortium will integrate and extend proven systems that are already delivering value to hundreds of virtual research communities worldwide. The increased quality and attractiveness of services provided by this e-Infrastructure in a key area is aimed at both reinforcing existing research communities

## Impacts

1. The Virtual Biodiversity Research and Access Network for Taxonomy will provide a direct route by which a wide range of stakeholders can access multi-level information and data.
2. ViBRANT's integration will dramatically increase the efficiency and capacity of European stakeholders to monitor and manage information about ecosystems, biodiversity and natural resources. These natural systems underpin Europe's economic, societal and individual wellbeing.
3. The network will support the emergence of virtual research communities of European and international dimension through the establishment of this virtual access network and by providing a framework for uniting national initiatives across the ERA.
4. ViBRANT will provide data management, analysis and publication tools in a self-governed, self-directed environment that, through a standards based framework and data exchange format, ensures that data can be integrated into the biodiversity information and support services needed by society. These actions will widen access by promoting the use of ViBRANT's services and tools by stakeholders from new disciplines and by seeking to engage with and support newly emerging scientific communities.
5. ViBRANT's enhanced infrastructure has the potential to deliver social change that goes far beyond earlier "name and fact recording" initiatives (which are the launch platform of this virtual infrastructure).

### 3.2.2. *Work progress and achievements during the period*

## Impacts

In the first year of the projects several services and products have been made public. At such an early stage it is unsurprising that use rates are not very high and they do not yet have baseline data by which to judge performance. We anticipate that some of these products will report increased growth in the project Year 2 and that they will become better integrated. Other elements, such as OBOE (WP5) represent infrastructural components that are ready for integration into the landscape but at this point have no complete routes for use. Note that ViBRANT was designed to be managed in an agile manner, so we launch components as early as possible then refine their performance in the light of user experience.

### **Scratchpads<sup>1</sup>**

Scratchpad usage has continued to grow in a satisfactory manner following the transition to ViBRANT. It has been a significant concern that development in Year one, from a user's perspective, has been restricted to bug-fixing. The efforts of WP3 in helping individual users has, we believe, contributed to this continued growth in the absence of obvious innovation. WP2 meanwhile have been developing Scratchpads 2, which is a full re-write of the code-base to take advantage of the changes in Drupal's core on the release of Drupal 7 (D7). All current sites, shown in the graph below, are running under Drupal 6 (D6, Scratchpads 1). Current access statistics can be found at <<http://scratchpads.eu/scratchpads/stats>>.

---

<sup>1</sup> <http://scratchpads.eu/>

Five modules, written by project staff and expected to have broader application than Scratchpads themselves, have been placed in the Drupal central repository. The code can be found by forming a URL as <<http://drupal.org/project/>> followed by the module name. Usage statistics can be found by forming a URL as <<http://drupal.org/project/usage/>> followed by the module name. Note that there are no Version 7 Scratchpads as yet, so use in D7 is outreach beyond Scratchpads themselves.

#### **slickgrid**

A module that provides spreadsheet functionality within a web-site. On 4th December there were 93 sites using the D6 version and 152 using the D7 version (total 245).

#### **hashcash**

A module to help protect sites from spam, originally written during the EDIT project. On 4th December there was one site using the D5 version, 420 sites using the D6 version and 44 using the D7 version (total 465).

#### **batax**

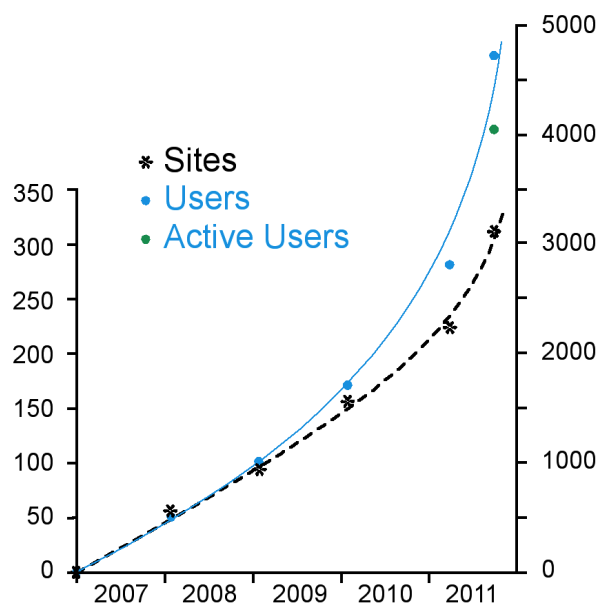
Provides an autocomplete form field instead of a drop down/select box if a taxonomy is bigger than a predefined number of terms. On 4th December there were 3 sites using the D5 version, 76 sites using the D6 version and 19 using the D7 version (total 98).

#### **femail**

This module enables Drupal to receive email messages and post them directly into a forum. On 4th December there were 364 sites using the D6 version and 10 using the D7 version (total 374).

#### **tui**

A module that provides a simple tree interface to allow quick and easy editing of a vocabulary and its terms. On 4th December there were 380 sites using the D6 version. The D7 version of this module has not yet been released.



### **GeoCat<sup>2</sup>**

GeoCAT is an open source, browser based tool that performs rapid geospatial analysis to ease the process of Red Listing taxa. Developed to utilise spatially referenced primary occurrence data, the analysis focuses on two aspects of the geographic range of a taxon: the extent of occurrence (EOO) and the area of occupancy (AOO). These metrics form part of the IUCN Red List categories and criteria and have often proved challenging to obtain in an accurate, consistent and repeatable way. Within a familiar Google Maps environment, GeoCAT users can quickly and easily combine data from multiple sources such as GBIF, Flickr and Scratchpads as well as user generated occurrence data. Analysis is done with the click of a button and is visualised instantly, providing an indication of the Red List threat rating, subject to meeting the full requirements of the criteria. Outputs including the results, data and parameters used

GeoCAT Statistics from March 2011	
Visits	3,580
Unique Visitors	1,494
Page views	10,780
Pages/Visit	3.01
Avg. Time on Site	00:05:56
Bounce Rate	40.11%
% New Visits	40.22%
Countries	91

2 <http://geocat.kew.org>

for analysis are stored in a GeoCAT file that can be easily reloaded or shared with collaborators. GeoCAT is a first step toward automating the data handling process of Red List assessing and provides a valuable hub from which further developments and enhancements can be spawned. GeoCAT will be open to a broader community from January 2012.

### **COMBER<sup>3</sup>**

COMBER (Citizens' Network for the Observation of Marine BiodivERsity), has been initiated under the ViBRANT EU e-infrastructure. It is designed and implemented for divers and snorkelers who are interested in participating in marine biodiversity citizen science projects. It shows the necessity of engaging the broader community in marine biodiversity monitoring and research projects, networks and initiatives. It analyses the stakeholders, the industry and the relevant markets involved in diving activities and their potential to sustain these activities. Future plans include promotion, links with citizen science web developments, data publishing tools, and development of new scientific hypotheses to be tested by the data collected so far.

During the two months of the project (July–August 2011), 48 users (excluding the four supervising scientists) participated. Twenty of the users contributed data from more than one dive or snorkelling trip and thus expanded the sampling area to several other locations in Greece. In total, 1,879 species observations were recorded during 95 dives and 39 snorkelling trips.

Participants came from ten countries, with the majority (42%) coming from Greece, followed by the United Kingdom and the Netherlands (12% each). The majority (70%) of the participants held a basic-level diving certificate (PADI Open Water / Advanced Open Water, CMAS \*), 12% held an advanced certificate (PADI Rescue Diver, CMAS \*\*) and 16% held a professional diving license. However, half of the divers had an advanced diving experience (>30 dives), independent of their certificate. Most of the participants already had a level of knowledge about marine organisms (72% declared they had advanced (36%) or basic (36%) knowledge about marine organisms, while 28% declared they had no knowledge at all). The genders were unevenly distributed (64% male, 36% female), but all age groups were present (21–55 years), with a slight dominance of 20–30 year old (39%) and 40–50 year old (30%), the other age ranges were 30–40 years old (18%) and >50 years old (13%).

The web site in the first 2 weeks of January 2012, attracted 72 unique visitors who visited the site on 100 occasions and viewed 431 pages, thus generating approximately 1650 hits per week in the closed season for diving. These figures will provide a baseline for comparison with the summer season.

### **Pensoft**

The activity of Pensoft journals during the period 30 June 2010 - 13 Jan 2012 is detailed below.

	ZooKeys	Phyto-Keys	IJM	CCG	JHR	NeoBiota	Myc-Keys	BioRisk	SubtBiol	TOTAL	Using PMT %
Submissions	803	65	20	129	60	37	15	24	19	1172	
Articles published	587	50	15	54	29	15	8	49	19	826	91.76
Issues	116	8	3	7	5	3	1	3	2	148	96.62
Pages	13781	574	256	628	467	215	91	1361	217	17590	91.02
Page ranges:											
1-10	164	30	6	27	9	4	6	8	12	266	
10-20	245	16	3	25	12	9	1	18	6	335	

<sup>3</sup> <http://www.comber.hcmr.gr>



20-30	75	2	5	2	6	2	0	11	0	103	
10-20	79	2	1	0	1	0	1	8	0	92	
60-100	12	0	0	0	1	0	0	2	1	16	
> 100	11	0	0	0	0	0	0	2	0	13	

Significant steps in changing the nature of the publication environment were:

- Realising the importance of Wiki environment for popularisation and dissemination of biodiversity data, in April 2011 Pensoft demonstrated the automated integration of species descriptions at the day of publication to Species-ID, achieved by programming a special tool that transforms the XML versions of the papers into MediaWiki-based pages.
- In October 2011, ZooKeys launched its multiple-choice model for publishing biodiversity data that employs the Darwin Core Archive, that encapsulates both data and metadata in machine-readable form. This allows the production of "Data Paper" manuscripts that formally describe a dataset's metadata as a peer-reviewed and citable scholarly publication.
- An important element of the multiple-choice data publishing model was the integration of its data publishing workflow with the Dryad Digital Repository, thus providing an option to its authors to archive data files of different kinds and complexity (e.g., phylogenetic, morphometric, ecological, environmental, etc.).
- Archiving of TaxPub-based articles in the PubMedCentral repository of the National Library of Medicine of the USA. This is the first case in the history of PubMedCentral, where a domain-specific XML schema (taxPub) is used, in the form of an extension to the NLM DTD, for archiving purposes and visualisation of some elements within the text, such as taxon treatments. Currently the whole content of [ZooKeys](#) and [PhytoKeys](#) is being exported and archived in PubmedCentral.
- Pensoft's latest innovation was announced on the 22nd of November 2011 with the launch of an automated export and indexing of identification keys metadata published in the journals in KeyCentral – a global database of keys and other identification resources for living organisms (<http://keycentral.identifylife.org/Secure/KeyStore/List.aspx>)

Data exported to external reference sources:

EOL - new taxa	1087	53	21		52					1213	
Keys metadata in KeyCentral	410	19	2		24					455	
Taxon treatments in Species ID										ca. 2500	
Articles archived in PubMed Central	587	50								637	
Articles in Directory of Open Access Journals (DOAJ)	587	50								637	

Some of these processes began life before ViBRANT kick-off (1 Dec 2010), during the project's preparation phase. These data represent baseline information against which future performance can be judged. The tight integration of Pensoft's corporate aims with ViBRANT's goals renders impractical the separation of those elements driven by ViBRANT funding from their overall activity.



### MediaWiki

The biowikifarm presently consists of 23 Wiki instances. Statistics are given below for wiki-internal measures (first table) and aggregated as Web Statistics (several wikis are grouped into a web statistics group).

ViBRANT biowikifarm statistic, 17.Jan.2012		Page statistics			Edit statistics		Google
		Content pages <sup>4</sup>	All Pages <sup>5</sup>	Up-loaded files	Page edits	Average edits per page	Analytics Group
1	Open Media shared repository	89,628	113,512	110,638	133,752	1.18	Species-ID
2	Pest Information Wiki	42,877	54,681	0	66,712	1.22	ISPI
3	Species-ID	5,590	9,708	2215	17,352	1.79	Species-ID
4	LIAS Glossary	2,624	3,335	3	6,830	2.05	OTHER
5	Offene-Naturführer.de	2,196	4,490	26	34,036	7.58	ON
6	orowiki.org	2,040	4,592	1,147	14,459	3.15	Orowiki
7	phytomedizin wiki	951	1,256	23	1,475	1.17	OTHER
8	KeyToNature	824	2,981	1,244	18,802	6.31	(none)
9	floramalesiana.org	509	2,097	0	2,864	1.37	OTHER
10	French Plantnet uses wiki	295	672	0	5,643	8.40	OTHER
11	English Plantnet uses wiki	279	15,584	0	123,357	7.92	OTHER
12	nnvm.eu	248	608	1	2,658	4.37	OTHER
13	KeyToNature Handbook	129	225	2	1,087	4.83	(none)
14	diversityworkbench	45	187	79	987	5.28	OTHER
15	phytopathology.net	8	172	115	539	3.13	OTHER
16	PMSL Wiki	7	42	27	114	2.71	(none)
17	zsm-entomology.de	7	36	13	331	9.19	OTHER
18	Studienstiftung Mykologie	3	20	11	36	1.80	OTHER
19	English Plantnet terms wiki	2	3	0	9	3.00	OTHER
20	French Plantnet terms wiki	1	2	0	19	9.50	OTHER
21	wiki.gbif.de	1	2	0	5	2.50	OTHER
22	FotoFlora BB	1	8	0	27	3.38	OTHER
23	French species-id instance	1	5	0	52	10.40	Species-ID
	Total	148,266	214,218	115,544	431,146	2.01	---

<sup>4</sup> article and media metadata namespace

<sup>5</sup> including talk pages, templates, redirects, etc.

Google Analytics, 1.Dec. 2011 to 31.12.2011					
	OTHER sites	ISPI	ON	Orowiki	Species-ID
Visits	1,591	3,044	4,238	517	1,528
Unique Visitors	1,287	2,614	3,298	324	1,127
Pageviews	4,842	7,696	10,444	3,265	4,436
Pages/Visit	3.14	2.53	2.46	6.32	2.9
Avg. Time on Site	00:02:45	00:01:28	00:02:15	00:05:42	00:02:35
Bounce Rate	20.46%	21.12%	15.64%	21.66%	17.60%
% New Visits	76.27%	82.06%	76.29%	59.96%	71.07%

Other biowikifarm sites = are all wikis without those analysed separately and without the KeyToNature sites (since that project is terminated and is currently not continued)

Total are not available to Google Analytics visitor statistics, Visitors may overlap

Biowikifarm users, 17.1.2012: 422

## WP2 - Technical architecture

Lead: Simon Rycroft, Natural History Museum

Activity type: Research

Partner	Name	Acronym	Effort (PMs Year 1)
1	The Natural History Museum, London	NHM	15
15	Freie Universität Berlin	FUB-BGBM	0.3

### Objectives

To develop and deliver the technical architecture required to host, integrate and sustain the Scratchpad framework within the ViBRANT consortium. This moves the Scratchpad prototype into a sustainable open-source, enterprise-level system.

1. Develop and deliver an enhanced technical framework for hosting the Scratchpad infrastructure.
2. Distribute the Scratchpad server so that other institutions can independently host Scratchpads.
3. Develop new functionality for Scratchpads, including human interfaces and APIs.
4. Develop a financial model for sustainable delivery of service.

### Description of work and role of partners

#### Task 1: Hosting architecture (NHM, BGBM)

1. Move the main Scratchpad host infrastructure to provide failover by load balancing over multiple servers.
2. Develop the capacity to distribute Scratchpad node servers to other institutions.
3. Provide technical installation support for hosting institutions and communities (distinct from Scratchpad user support under WP3).

#### Progress in Year 1

A test version of a mirror of the main Scratchpad server (Quartz) has been developed using the same operating system and software as Quartz, enabling it to be quickly and easily configured for use. Implementing this test server has enabled us to develop a script that will enable us to automate much of the process of creating a Scratchpad mirror.

Three key technologies were investigated to ensure that the files and data are kept up-to-date with Quartz, and that the servers are easy to maintain. The technologies were:

**rsync.** rsync is a software application for Unix and Windows systems which synchronises files and directories from one location to another while minimising data transfer using delta encoding when appropriate<sup>6</sup>. This means that we can ensure that we are able to update files as frequently as once an hour, so were Quartz to become unavailable, files should be present on the mirror.

**MySQL replication.** Replication enables data from one MySQL database server (the master) to be echoed to one or more MySQL database servers (the slaves). Replication is asynchronous - slaves need not be connected permanently to receive updates from the master<sup>7</sup>. This means that data are present on a mirror as soon as they are present on the master. The master in our case would be Quartz.

**Aegir.** The Aegir hosting system allows developers and site administrators to automate many of the common tasks associated with deploying and managing large websites. Aegir makes it easy to install, upgrade,

<sup>6</sup> <http://en.wikipedia.org/wiki/Rsync>

<sup>7</sup> <http://dev.mysql.com/doc/refman/5.0/en/replication.html>

deploy, and backup an entire network of Drupal sites<sup>8</sup>. Aegir will hopefully allow us to create a web of masters and mirrors, with each server in the web acting as the master for sites locally created, and a mirror for all other sites in the system. Further investigation will be carried out on this as part of the ViBRANT deliverable “D2.1 – Distributing servers”<sup>9</sup>.

### **Distributing Scratchpad servers**

The work, carried out by both the BGBM and NHM, will help ensure the reliability of the Scratchpads as a whole and seeks to make them more sustainable. Until now the only multiple site Scratchpad installation was the one hosted at the NHM. The work described below will play a key role in creating a Scratchpad distribution that other users/institutions can download and install on their own servers.

### **Creation of second server**

We diverged from our original concept, deciding to use only Aegir for keeping a site mirror up-to-date with its origin site. This decision was made to keep things as simple as possible. Using MySQL replication was unnecessary and could potentially cause undue strain on the primary server. The system we have created is capable of updating a site mirror at regular intervals. We have currently chosen a time period of one week to update a site mirror, although this could be reduced or increased depending on site activity. Constraints in the DNS configuration mean that we are currently only able to automate creation of accessible site mirrors with domain names that we control. This means that while a site such as [www.cicadina.eu](http://www.cicadina.eu) can have a mirror created using a domain name like [mirror.cicadina.eu](http://mirror.cicadina.eu), it would have to be done manually which imposes a unscalable and unsustainable workload on the Scratchpad team. It is our intention to start hosting the DNS records for all of the Scratchpad domains including those purchased specifically for a Scratchpad. This change to the way we are handling DNS records will enable us to host a Scratchpad on any Scratchpad server without the user being aware of the difference and also allow us to create a mirror with a domain name like [mirror.cicadina.eu](http://mirror.cicadina.eu).

The server at the BGBM<sup>10</sup> has been configured with almost exactly the same resources as the primary server at the NHM. It has four CPU cores available to it, and 16GB of RAM. This means that it should be more than capable of handling exactly the same quantity of traffic as the primary server. In the event of a catastrophic event at the NHM, we would be able to set the BGBM server as the primary Scratchpad source.

### **Primary/Mirror Interaction**

The primary Scratchpad server talks to the Scratchpad mirror using the SSH protocol. The Aegir instance, hosted on the NHM machine, logs into the BGBM mirror and issues commands. For example, to create a mirror of an NHM hosted Scratchpad on the BGBM server, the following operations are carried out:

1. The site folder on the NHM machine is compressed and added to a tar file.
2. The tar file is copied to the BGBM machine, and extracted in the same location as on the NHM machine.
3. Settings files are created for the new site.
4. The site database is copied from the NHM machine to the BGBM machine.
5. The service is reloaded, making the new site visible.
6. The process is verified, with it rolling back to a clean state if the process of creating the mirror failed.

This process can be automated so that any new sites created on the primary NHM server are automatically mirrored to the BGBM mirror, and also to ensure that the site mirrors are kept up to date.

### **Code updates**

Currently the code on all the servers are kept in synchronisation with a simple bash script. The script is executed by the lead Scratchpad developer when a new code release has been produced. The script logs

<sup>8</sup> <http://www.aegirproject.org/>

<sup>9</sup> <http://vibrant.eu/content/d21-distributing-servers>

<sup>10</sup> <http://e090.bgbm.fu-berlin.de/>

into each Scratchpad server in turn, and updates the code using standard Git<sup>11</sup> commands. At the moment, this more than satisfies our requirements, however in the future this may be changed to better integrate with the Aegir installation, and also with the output of the future deliverable D2.4.

### Further work

It is our intention to start creating Scratchpads whose primary server is the BGBM machine. These Scratchpads would then be mirrored on the NHM server. This process cannot be started, however, until we have made significant changes to the way we currently handle our DNS. Once the DNS changes have been made, there will be no restriction to the number of Scratchpad nodes we can have running across the ViBRANT partners, and therefore no limit to the number of Scratchpad users and usage.

### Task 2: Multi-site integration (NHM)

1. Build a dynamic registry of Scratchpad sites and associated data services (APIs).
2. Create a portal, using the registry, for cross-site searching.
3. Develop measures of data usage across all sites and a centralised method of publishing these statistics.
4. Research options for a data citation metric for biodiversity data published through the ViBRANT network.

### Progress in Year 1

Usage statistics help to prioritise further development in a project like ViBRANT. For this reason, we have put a lot of emphasis on collecting and analysing in fine detail as many user statistics as possible. The progress summarised here is more fully available in the paper *Who Learn from Whom*<sup>12</sup>, written for the ViBRANT special issue of ZooKeys.

Scratchpad usage statistics will be collected using the Scratchpad registry. The registry is a ViBRANT deliverable, due at the end of month 18. The registry is intended to run as a service at the NHM and will allow Scratchpads hosted by both the NHM and other institutions or individuals, to report their presence and statistics. This process will be fully automated and will be enabled by default on all Scratchpads. The exact configuration of the Scratchpad registry is yet to be finalised, although it is likely to be very similar to the GBIF IPT<sup>13</sup> registry system<sup>14</sup>.

We will also continue to collect usage statistics using the Google Analytics tool<sup>15</sup> that currently aggregates statistics across all Scratchpads, and is therefore of much benefit to us.

There are privacy concerns regarding the collection of usage information, although users will be advised in the terms and conditions of the Scratchpad that site usage statistics will be collected, and every effort will be taken to ensure that statistics are anonymised before publication or distribution, in compliance with relevant codes for data protection. We note that it may not be possible to collect such data from servers in some countries because of different legal frameworks.

It is our intention to produce a module registry system that is capable of storing any data sent to it from a Scratchpad. This will enable us to quickly and easily enable the collection of additional statistics as and when they are required. A default system will should be expected to collect, at the very least, the following statistics:

- Number of users, and statistics regarding the login of those users.

<sup>11</sup> <http://git-scm.com/>

<sup>12</sup> Brake, I., Duin, D., Van de Velde, I., Smith, V. & Rycroft, S. (2011). Who learns from whom? Supporting users and developers of a major biodiversity e-infrastructure. *ZooKeys* **150**, 177-192.

<sup>13</sup> Integrated Publishing Toolkit - <http://ipt.gbif.org/>

<sup>14</sup> <http://code.google.com/p/gbif-registry/>

<sup>15</sup> <http://www.google.com/analytics/>

- Quantity of data entered into the site, including statistics regarding the types of data entered and by whom.
- Statistics regarding the use of the sites by anonymous users (those not logged in to the site and providing content to it) will also be collected. These will be additional to the data collected from Google Analytics or a similar tool.
- Metadata about the site, helping to group analysis of multiple Scratchpads together (e.g. are bird Scratchpads more popular than banana Scratchpads?).

### **Task 3: Scratchpad Interface/functionality extensions (NHM)**

1. Lead in overall software development and project integration (point of contact management).
2. Develop a mechanism for compatibility testing of modules submitted from all contributors within a code repository for Scratchpad developers.
3. Prioritise development and user supplied feature requests (expected through WP3) and implement the most requested functions.
4. Manage the training resources, including the sandbox server, as defined by WP3.

#### **Progress in Year 1**

The decision was taken in Month 1 of ViBRANT to upgrade the code base for the Scratchpads from Drupal 6 to Drupal 7. This decision was made based on the average life cycle of a single Drupal release and also because a number of features that we had developed for Drupal 6 were now included in the core of Drupal 7. There will be a number of other benefits from upgrading the Scratchpad code base, including a performance increase, ease of maintenance and improved reliability. For this reason, a number of the feature requests that have come to us from WP3 and also via the Scratchpad issues queue<sup>16</sup>, have had to be delayed. These feature requests will, assuming they are reasonable, be added to the queue of issues for Scratchpads 2.0 once it is released in early 2012. A more complete description of Scratchpads 2.0 is available in the ZooKeys special issue<sup>17</sup>.

The upgrade work has taken longer than anticipated because some contributed modules have not been upgraded to D7, or have been upgraded in a way that does not take advantage of the new D7 architecture, so that the development team have had more work than forecast.

### **Task 4: Sustainability (NHM)**

1. Develop a financial model for sustainable delivery of service. This would include development of tiered services offering free and paid accounts for communities and institutions with different levels of service.
2. Provide a Service Level Agreement (SLA) to users.

#### **Progress in Year 1**

Work on this task is scheduled to begin in Month 27, forming the main body of D2.3 - *Financial sustainability*

<sup>16</sup> <http://dev.scratchpads.eu/>

<sup>17</sup> Smith, V., Rycroft, S., Brake, I., Scott, B., Baker, E., Livermore, L., Blagoderov, V. & Roberts, D. (2011). Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science. *ZooKeys* **150**, 53 - 70

### WP3 - Training, outreach and community support

Lead: Thierry Backeljau, Royal Belgian Institute of Natural Sciences      Activity type: Co-ordination

Partner	Name	Acronym	Effort (PMs Year 1)
1	The Natural History Museum, London	NHM	1
3	Royal Belgian Institute of Natural Sciences	RBINS	11.8
5	Vrije Universiteit Amsterdam	VU	10.3
17	Università di Trieste	UNITS	0.6

#### Objectives

Empower and enable biodiversity scientists to use ViBRANT tools to strengthen community collaboration by means of:

1. engaging with communities through using existing consortium services to develop an extended network of partners that will deliver training, outreach and community support;
2. provide input to consortium members on user needs and practices to feed into ViBRANT development, training and outreach priorities;
3. providing site development expertise and delivery of software documentation, training courses and service desk user support;
4. promoting and facilitating peer production support infrastructures.

Specific activities include:

1. Organise and deliver ViBRANT training resources in order to facilitate and improve production, management and publication of biodiversity data.
2. Undertake networking activities to enhance use of ViBRANT services and to develop a network to foster long-term sustainability of the ViBRANT community.
3. Undertake sociological studies of our user-base that will underpin development priorities and maximise engagement in the ViBRANT community.
4. Manage a user-feedback mechanism enabling users to make feature requests and bug reports.

#### Description of work and role of partners

##### Task 1 Organisation and delivery of ViBRANT training courses in association with WP2 (RBINS)

1. Provide intra-community training resources that facilitate experienced users to train their colleagues.
2. Deliver training courses and workshops to the IT managers within communities
3. Offer custom-made training modules for taxonomic research communities addressing their specific needs.
4. Provide specific training on the generic mechanical Turk module (built by WP2) that enables citizen scientists to mark-up and transcribe textual content from images of specimens and literature. The design will include game based (competitive) elements will provide incentive to citizen scientists to contribute. The module will be made available to all Scratchpad users and be developed in co-ordination with WP7 for mark-up of biodiversity materials.
5. Apply online learning (courses, video tutorials, etc.) and on-site training e.g. workshops including hands-on practice, organised at major scientific conferences, as effective training methods.



6. Optimise skill building and transfer of knowledge by using peer-based learning and cascade training (train-the-trainer). The network of associated member institutions (see Task 2) will be used to mobilise training providers willing to provide ViBRANT training.
7. Provide feedback for training content development and planning. A sociology survey (see Task 3) will evaluate the offered trainings and identify specific needs of the ViBRANT community.

### **Progress in Year 1**

Training is part of the outreach activities of the ViBRANT project. The major objective is to enhance the use of ViBRANT tools and to support and extend the user communities working with those tools that will lead to an increased production of biodiversity information. Central to these efforts is the inclusion of sociological studies of ViBRANT's user-base. This work will underpin Scratchpads' development priorities and maximise engagement in the user community. Hence, a significant emphasis is placed on training and services such as the help desk to support the users and on social science studies to understand the barriers to adoption and use of tools by our user community.

It is clear that the delivery of training to small groups is not scaleable within the available project resources, so two strategies are combined to deliver more effective training to a greater number of participants. First, the on-line help system has been developed to deliver context-sensitive training materials (i.e. elements of the normal training courses). Second a system of Ambassadors is being developed (see Task 2, below).

Training is organised by the Scratchpad support team. At the time of writing (November 2011) the Scratchpad team consists of the project leader, three developers, and three user support staff (some part time). At least 5 team members (developers and user support staff) take part in the delivery of training courses. In addition, it is expected that the Ambassadors network will engage in delivering demos and training in their research community.

WP3, in cooperation with WP2, will organise and deliver training resources in order to support and extend the user communities working with ViBRANT products. This includes training courses organised at regular intervals mainly at NHM London and at workshops associated with conferences, bringing ViBRANT products to meetings organised by user communities. In addition to training courses, electronic resources are accessible to users at their home institution via the website. Training manuals and other help files are available on each Scratchpad (.../admin/advanced\_help/scratchpadify\_help) for self-training, if no opportunity to participate in a training course arises. Self-training can also be done on a temporary site, the 'Sandbox', that is wiped every 6 hours. Alternatively, the Scratchpad team can provide a home-training site where users can practice for a longer period.

The Scratchpad platform also offers support systems (see Task 4, below).

Training includes basic and advanced courses. In addition, tailor-made courses can be delivered on demand. The one-day courses are intended to help current and prospective Scratchpad users to develop their site building skills. They are given the opportunity to learn best practice and gain a better understanding of what the Scratchpads can do for them and for their research community. The basic course focuses on adding various kinds of data, on generating taxon pages and on communication with other users. The goal is to provide a taste of what the sites can do, and allow independent exploration of their site after the training event. The focus of the advanced course lies on the import of data, creating of custom content types and views, and managing of projects (groups). Training takes place in small groups (usually 12 participants) in order to give the trainees sufficient individual attention. No course registration fee is asked but participants are responsible for their own travel and subsistence expenses. A budget has been anticipated for certain trainees to attend the courses, if needed, or to finance the trainers to deliver a course in other locations.

### **Training Delivery**

A new version, Scratchpads 2.0, is planned for release in January 2012 and will include many enhancements. Hence, it was decided to put priority on the development of the Ambassadors network during the second half of Year 1 instead of delivery of training courses. Nevertheless, for continuity and to meet demand, the following training courses were delivered during Year 1.

Date	Location	Participants	Course level
9 December 2010	NHM London	11	Basic
19 January 2011	MNHN Paris*	12	Basic
16 February 2011	NHM London	10	Basic
28 February 2011	NHM London**	14	Basic
15 August 2011	NHM London***	11	Basic
25 August 2011	Heriot Watt Univ. Edinburgh****	12	Basic

\* EDIT closing conference \*\* Throughflow Workshop \*\*\* Convolvulaceae Unlimited Workshop \*\*\*\* MASTS BEF JRT Workshop

Forthcoming training courses are advertised on the web<sup>18</sup>. In February 2012 a Scratchpad course is being run as part of the NHM/Imperial College MSc course in Taxonomy<sup>19</sup>.

## **Task 2 Implementation of networking activities to extend the ViBRANT community and to foster its long-term sustainability (RBINS)**

1. Build upon established networks and organisations (EDIT, GBIF, Scratchpad network) to create a network of associated member institutions each with a contact person in charge of informing on existing and new services generated by the ViBRANT consortium and delivering training, coaching and mentoring.
2. Make use of conferences, meetings, workshops organised by taxonomic, biodiversity research and conservation communities to promote the use of ViBRANT tools and to inform on training possibilities (info shops).
3. Encourage nascent communities to cooperate through targeted workshops and meetings.
4. Undertake outreach and advocacy work to stimulate and improve networking, to get involvement and commitment of organisations across Europe and beyond and to foster long- term sustainability of the ViBRANT community

### **Progress in Year 1**

In order to promote Scratchpad use and to foster long-term sustainability of the Scratchpad community, the ViBRANT project launched the Scratchpad Ambassador programme<sup>20</sup>. WP3 is continuing to recruit a select group of enthusiastic and experienced Scratchpad users to be the official local representative of the Scratchpad community, linking the Scratchpad team with Scratchpads' growing user base. Ambassadors spread the word about Scratchpads, promote the use of Scratchpads and arrange or give training in their local Scratchpad community. Ambassadors are our point of contact person for Scratchpad users in their local taxonomic community and in that way they help the Scratchpad team to better understand the needs of users, so that the Scratchpad developers can keep improving Scratchpads.

The process for planning and organisation of training courses and the ambassadors network is shown in D3.1<sup>21</sup> (p.7).

A promotional strategy for ViBRANT services<sup>22</sup> has been developed that describes ViBRANT's overall outreach efforts aiming to optimise the results and output of the project as well as the dissemination of these results to all relevant stakeholders in order to encourage use of ViBRANT tools and services. It encompasses a structured set of dissemination tools and planned outreach activities, that will effectively

<sup>18</sup> <http://scratchpads.eu/scratchpad-training-courses>

<sup>19</sup> <http://www.nhm.ac.uk/research-curation/postgraduate/msc-taxonomy/index.html>

<sup>20</sup> <http://scratchpads.eu/ambassadors-programme>

<sup>21</sup> <http://vibrant.eu/sites/vibrant.eu/files/D3.1-Training%20strategy.pdf>

<sup>22</sup> [http://vibrant.eu/sites/vibrant.eu/files/M.3.10\\_Delivery%20of%20a%20promotional%20strategy%20for%20ViBRANT%20services\\_0.pdf](http://vibrant.eu/sites/vibrant.eu/files/M.3.10_Delivery%20of%20a%20promotional%20strategy%20for%20ViBRANT%20services_0.pdf)

send the intended messages about the operational objectives to the existing and potential stakeholders. The approach proposed will enable the project to obtain maximum visibility and coverage. The outlined promotional strategy is not a static but a dynamic document that will be updated when the project results are moving forward and feedback is received from the stakeholders.

The University of Trieste had several formal and informal communications with the scientific community in Italy and more broadly in Europe, in which they described ViBRANT and Scratchpads. They proposed the adoption of Scratchpads in two Italian national initiatives:

1. commented lichen flora of Italy;
2. as a taxonomic framework for the forthcoming Italian National Biodiversity Node.

Both initiatives are currently in discussion by the relative scientific communities.

The reception by the Italian scientific community in particular was mixed: while there is an increasing interest in a new e-way to manage and share taxonomic information, few taxonomists are really ready to adopt a new community approach for managing information and producing new content. A certain interest was expressed by the members of the forthcoming Italian National Biodiversity Node, who are exploring new possibilities in digitalising and sharing biological data.

### **Task 3 User studies (VU)**

Success of infrastructure development and adoption requires networking and meeting needs and practices in scholarly communication of relevant (current and potential) users. In this task WP3 will study this from five perspectives:

1. Analysis of previous studies on use of ViBRANT and related tools in relation to needs, barriers and practices of the research communities involved.
2. Study of existing barriers and enabling factors for the diffusion and uptake of the ViBRANT tools, at the level of individuals and organizations - among others the current network of user communities: composition, motivation and social structure.
3. Study of the scholarly communication practices and output of potential user communities to identify trends, levels and holes in (interdisciplinary and multidisciplinary) practices of collaboration and networking.
4. Study of organisational forms for ViBRANT, that support the adoption and use, among others, investigation of the uptake of the open science approach.
5. Study of the scholarly communication practices and output of relevant user communities in order to assess the effects of the ViBRANT tools on the progress of sciences involved. Qualitative and quantitative data on these issues will be delivered in order to inform and improve development and organisation of the ViBRANT tools, and in order to support network building.

### **Progress in Year 1**

Several paths were followed in the first year: 1) collection of electronic user and use data of the Scratchpads; 2) refinement of the bibliometric map taxonomy and biodiversity research; 3) a scan developments of new forms of publication and research communication, and 4) a pilot study to test the feasibility of electronic data for studying users of Scratchpads.

1. The access logs of information systems such as Scratchpads hold interesting behavioural information on its users. Currently the Scratchpad access logs are stored on the NHM server, separately for each site. The integration of this information in a single database was one of the activities VU addressed in Year 1. With the integration of the access logs in one single database it will be possible to study more complex network activities of users. For instance their activities across Scratchpads, or to collect information on Scratchpad activity per institution or per country. The database will help, among other things, to identify power users who could then be approached

for the Ambassadors programme. This work has been carried out in collaboration with WP2. More information on the database can be found in the progress report<sup>23</sup>.

2. In order to map the potential users community of the ViBRANT infrastructure, an initial bibliometric map was made of the fields of taxonomy and biodiversity research. Although the used Web of Knowledge data do not cover the total output of the field, it may provide us with a reliable estimate of the size and composition of the relevant community. Knowledgeable representatives of the field were consulted and we are currently working on the second, refined version, taking into account feedback received.
3. A scan of broader science policy issues around e-infrastructures was initiated that will lead to a report next year. Among others, the adoption of e-infrastructures crucially depends on the way (i) data sharing, tool building and open access publishing are taken into account in research evaluation, and (ii) e-infrastructures and open access publishing can develop sustainable business models.
4. A pilot study was carried out that made use of the electronic available user and use data of Scratchpad users. Because the research environment of scholars is increasingly web-based it has become urgent to study the effects on research practices, scholarly output and innovation of moving science to the Web. A theoretical framework and a methodology to study these effects have been proposed<sup>24</sup>. Theory and method were applied on an online community in biodiversity research, to demonstrate the feasibility of the approach. Also the practical relevance of this kind of analysis for improving the quality of virtual research environments has been highlighted.

In the pilot study two research questions were addressed by asking whether Scratchpad membership: i) connects people that were otherwise not connected; ii) provide network conditions that are beneficial for the creation of new knowledge and conditions for stability, that is, does the Scratchpad link researchers from different but not too different fields?

Initial results suggest that through computing of relatively simple graphs we get a better understanding of the effects of Scratchpad membership on scholarly networks. Thus we can compare characteristics of Scratchpad networks with e.g. co-authors. Our analysis suggests that Scratchpads do create links between researchers that do not exist in the co-author network and therefore fill structural holes in the network. This is one of the enabling conditions for the creation of new knowledge.

Analysing the number of shared co-authors among members of typical Scratchpad indicates that the members form a loosely collaborating group of researchers. However, if we also take into consideration the collaboration between the members, the network seems denser. Some Scratchpad members were already collaborating before joining the Scratchpad, however most co-author relations are from outside the Scratchpad community. In other words, the Scratchpad partly reinforce extant relationships, but also creates new links for those members that were not included in the co-author network.

Our initial study shows that the selected approach is promising. In the next phase we will extend the study in several ways. Firstly, we used data on only one Scratchpad. We plan to repeat the analysis for a large set of Scratchpads, which will enable us to test whether the level of variety correlates with knowledge production and innovation, as the theory suggests. Secondly, in the current pilot study we treated all co-author relations as having the same importance, which does not reflect real world relationships particularly well. This is also something to take into account in future research. Thirdly, we used only two different networks of the researchers (Scratchpads; co-authorships) while neglecting many others, such as organisational proximity, professor-student relations, project membership and scientific speciality. In order to get the full picture of the role of Scratchpads in scholarly networks, the analysis should be extended with the kind of networks mentioned. Fourthly, it is crucial to compare Scratchpad members with non-members, in order to test whether changes in research practice and performance of members are different from eventual changes in the field at large. Finally, the Scratchpad we studied in this paper was launched in 2011. In order to assess the effects of the deployment of virtual environments, we suggest using a longitudinal research approach:

<sup>23</sup> <http://vbrant.eu/content/44/progress-report-integrated-scratchpad-user-database-year-1>

<sup>24</sup> Duin, D. & van den Besselaar, P. (2011). Studying the effects of virtual biodiversity research infrastructures. *ZooKeys* **150**, 193-210.

have co-author networks and the thematic orientation of Scratchpads users changed over time, and is this change different from other researchers in the field?

These questions are not only theoretical relevant, but may also be useful in the practice of organising Scratchpads and other virtual research environments. It may also help to identify potential interesting new Scratchpad members that might be actively invited to participate. Moreover, these lines of research could contribute to sustain user engagement and to general research infrastructure policy. More generally, the results of the activities in this task aim to inform design, management and evaluation of e-infrastructures.

#### **Task 4 Provision of community support for the use of ViBRANT services in association with WP2 (RBINS)**

1. Manage and update the Scratchpad self-help screen-casts and tutorials (currently circa 100 videos) incorporating details of other ViBRANT services.
2. Assist communities with site development and content migration to the Scratchpad framework through a service desk providing real-time telephone and instant messaging support.
3. Prioritise requests for short-term developer assistance (from WP2) based on potential impact, potential risk of data loss and maintenance of user engagement, in the absence of developer intervention.
4. Manage and respond to comments on the issue tracker that seamlessly integrates with the Scratchpad framework to provide single-click feedback on feature requests, queries and bug reports.

WP3 will act as a liaison between the user community and WP2 developers.

#### **Progress in Year 1**

Within ViBRANT, Scratchpads<sup>25</sup> are a data-publishing framework for people to create their own thematic virtual research communities supporting biodiversity science. The Scratchpad platform provides at present the following support systems for users:

- the help desk dealing with all emails, issues tracker, calls and meetings relating to user support;
- a help system with help pages integrated into the individual Scratchpads;
- the sandbox and home-training sites for self-training;
- training courses with training manuals (manuals available online);
- a blog to inform the community of new features, fixed bugs, training courses etc.

Support systems are vital in maintaining user engagement in the early stages of use and they play an important role for the communication between users and developers of software. In a recent paper<sup>26</sup> two Scratchpad support systems, the issues tracker and the email service, were studied in detail. The aim was to identify co-learning opportunities between users and developers of the Scratchpad system by asking which support system was used by whom and for what type of questions. The results show that the issues tracker and e-mails cater to different user mentalities as well as different kind of questions and suggest ways to improve the support system as part of the Scratchpad development.

Getting to know the specific training needs of the Scratchpad users and of potential users is important. Within the framework of the VU user studies (cf Task 2), the Scratchpad team and WP3 have set up a user survey<sup>27</sup> in order to get a better view of user needs and to influence the further development of Scratchpads 2.0 as well as implementation priorities. The survey covers the following issues: Scratchpad community characteristics, system improvements, training courses, and technical support services. The focal maintainers ( $\pm 350$ ) of all Scratchpad sites were asked to fill in the online survey. The request was sent out

<sup>25</sup> <http://scratchpads.eu>

<sup>26</sup> Brake, I., Duin, D., Van de Velde, I., Smith, V. & Rycroft, S. (2011). Who learns from whom? Supporting users and developers of a major biodiversity e-infrastructure. *ZooKeys* **150**, 177-192.

<sup>27</sup> <http://www.qualtrics.com>

on 9 September 2011 with a reminder on 6 October 2011. At the time of writing this report, 53 surveys were fully completed. Most of the responses received are very positive with valuable suggestions and comments on the support systems and for Scratchpads software development. An evaluation of the results of the user survey will be dealt with in the subsequent milestone: M3.14 'Assessment of user support systems and promotional activities' due for M24 (November 2012).

Evaluation of delivered training is essential if maximum impact is to be achieved. Hence, all course attendees are asked to complete an online feedback form<sup>28</sup> (which can be done anonymously if desired). Most of the responses received are very positive and a detailed evaluation will be given in a forthcoming report (M3.14 'Assessment of user support services and promotional activities'). The feedback given by the participants after training influences the delivery and content of training as well as the future design of the Scratchpad software.

---

<sup>28</sup> <http://scratchpads.eu/feedback-form-training-courses>



## WP4 - Standardisation

Lead: Wouter Los, University of Amsterdam

Activity type: Co-ordination

Partner	Name	Acronym	Effort (PMs Year 1)
6	Julius Kühn-Institute	JKI	10
8	University of Amsterdam	UvA	4
14	Global Biodiversity Information Facility	GBIFS	6
15	Freie Universität Berlin	FUB-BGBM	8.1

### Objectives

This workpackage ensures that all data within ViBRANT are compatible with and available to other research and publishing infrastructures and services. Where applicable, it enforces or facilitates the use of externally standardised ontologies. This is especially applicable to TDWG, bibliographic, geographic, media metadata data exchange standards, but also to geographic and taxonomic content standards (taxon names, authorities, standard literature abbreviations, geographic entities, etc.).

Where no adequate standards exist, it enables ViBRANT users to develop, document, and voluntarily share new ontologies. The latter case is especially relevant with respect to biological character/feature ontologies, the associated categorical values, observation and measurement methods, organism-parts, and development stages. Although standardisation is in progress here as well, acceptable standards are generally limited to small taxonomic groups and specific use-cases (e.g. genomic analysis).

1. Facilitate integration and harmonisation of distributed data-sets.
2. Provide management and dissemination facilities for the necessary ontologies by:
  - a. Providing user-friendly access to external services based on these ontologies and APIs.
  - b. Developing APIs building on standard ontologies and protocols.
3. Provide biologists with the necessary flexibility to express their knowledge regardless of whether the terminology has been standardised yet or not.

### Description of work and role of partners

#### Task 1 Ontology platform (GBIF, JKI)

ViBRANT needs a flexible, user-friendly ontology management environment, enabling users to create, define, extent and share their own terms and concepts where needed, providing options for discussions and annotation, while supporting re-use of terms from standardised ontologies wherever possible (via Task 4.2). For this purpose ViBRANT will extend the functionalities of both the ontology managers of existing vocabulary services (like GBIF) and will develop a collaborative community interface (JKI) for users and user-networks to facilitate the (bottom-up) definition and sharing of their ontologies in a user-friendly (non-technical) way.

#### Progress in Year 1

The GBIF task group report on metadata implementation recommended the use of controlled vocabularies where the terms would be identified by persistent identifiers (Jones et al., 2010<sup>29</sup>, page 22-23, recommendation R53-R57). “*The identifier of the vocabulary should use existing identifiers from other*

<sup>29</sup> Jones, M.B., N. Bertrand, J. Holetschek, V. Hutchison, B.C.-J. Ko, A., Suarez-Mayorga, M. Meaux, W. Ulate, D. Watts, T. Robertson, and E. O Tuama (2010). Report of the GBIF metadata implementation framework task group (MIFTG). Global Biodiversity Information Facility (GBIF), Copenhagen. Available at [http://imgbif.gbif.org/CMS\\_NEW/get\\_file.php?FILE=2d85d0e8c76408129024c09aa\\_072d6](http://imgbif.gbif.org/CMS_NEW/get_file.php?FILE=2d85d0e8c76408129024c09aa_072d6)



*registries where possible. If one does not exist, then GBIF should construct and publish the identifier.” (...) “At present there is no well-defined and consistent means of referencing an identifier of a vocabulary or a vocabulary term. The proposed GBIF registry should provide an unambiguous citation method for each vocabulary and the terms they contain.” (...) “Some vocabularies will be global in use but some will be domain specific. To ensure compatibility across all metadata records, it is important that users use the appropriate and community agreed vocabularies”.*

The GBIF task group report on persistent identifiers provided a recommendation for GBIF to “take a leadership role in encouraging the use of metadata vocabularies for information in the GBIF data portal and extending the role of the data portal by hosting resources related to the use of identifiers, such as the TDWG vocabularies” (Cryer et al., 2010<sup>30</sup>, page 14, recommendation 12).

The GBIF beginner’s guide to persistent identifiers stated the importance “to reuse, where appropriate, the vocabularies and schemas that other communities have developed, to aid interoperability and save reinventing the wheel.” On the same page this guide also commented: “Because biodiversity informatics is a fairly specialized area of expertise, it is likely that a large proportion of the vocabularies and ontologies required for this domain will need to be developed within this community” (Richards et al., 2011<sup>31</sup>, page 20).

The GBIF task group for Knowledge Organisation Systems (KOS) made recommendations in a white paper on the use of vocabularies and ontologies for biodiversity informatics (Catapano et al., 2011<sup>32</sup>).

#### *DCMI 2011 pre-conference on vocabulary management*

In September 2011 GBIF represented the biodiversity informatics community at the Dublin Core annual conference for metadata and vocabulary management in Hague, Netherlands. The pre-conference that we attended addressed the best practices for maintaining a federated KOS with a common vocabulary of terms used by a decentralised network. The most important lesson learnt was that communities of domains other than ours were also just starting to work with very many of the same issues we face. The best practices for maintaining a federated KOS are under development in the Dublin Core (DCMI) community and the early experiences from the biodiversity informatics community will provide important input to this process.

#### *TDWG 2011 KOS Symposium*

GBIF organised a special symposium at the BIS (TDWG) 2011 conference<sup>33</sup>. A series of presentations introduced recent activities related to KOS in the GBIF work programme, the current status and history of the TDWG vocabularies and the management of the Darwin Core set of terms. The presentations left little time for discussion around the central issue of identifying key actors and activities that would enable TDWG, GBIF and related communities of practice to develop a road map on how to engage all players in creating a global infrastructure for the development, maintenance and governance of such vocabularies. However some of the issues were considered in the subsequent TDWG Technical Architecture Group (TAG) meeting and, as an outcome, GBIF proposes to create a task group within the TAG to take this work forward. This will entail submitting a charter outlining the purpose of the task group with expected deliverables and timelines.

30 Cryer, P., R. Hyam, C. Miller, N. Nicolson, E. O Tuama, R. Page, J. Rees, G. Riccardi, K. Richards, and R. White (2010). Adoption of persistent identifiers for biodiversity informatics: Recommendations of the GBIF LSID GUID task group, 6. November 2009. Global Biodiversity Information Facility (GBIF), Copenhagen. Available at [http://www.gbif.org/orc/?doc\\_id=2956&l=en](http://www.gbif.org/orc/?doc_id=2956&l=en)

31 Richards, K., R. White, N. Nicolson, R. Pyle (2011). A beginner’s guide to persistent identifiers, version 1.0. Released on 9 February 2011. Global Biodiversity Information Facility (GBIF) Copenhagen. Available at [http://www.gbif.org/orc/?doc\\_id=2428](http://www.gbif.org/orc/?doc_id=2428)

32 Catapano, T., D. Hobern, H. Lapp, R.A. Morris, N. Morrison, N. Noy, M. Schildhauer, and D. Thau (2011). Recommendations for the use of knowledge organization systems by GBIF. Released on 4 February 2011. Global Biodiversity Information Facility (GBIF), Copenhagen. Available at [http://www.gbif.org/orc/?doc\\_id=2942&l=en](http://www.gbif.org/orc/?doc_id=2942&l=en)

33 Eamonn O Tuama, Dag Terje Filip Endresen, David Remsen (2011) Establishing a support infrastructure for Knowledge Organisation Systems (KOS) in biodiversity informatics. Available at <https://mbgserv18.mobot.org/ocs/index.php/tdwg/2011/paper/view/162>

GBIF, in consultation with the community, will develop a charter for a “terms and vocabularies” management interest group; an initial draft charter to be circulated by Dec 31, 2011.

The increasing ease of creating media objects (still and moving images, sound, etc.) and distributing them in digital form through online channels is affecting the scientific workflow. Multimedia files become increasingly important as vouchers of scientific information. The Audubon Core, developed by the joint GBIF and TDWG multimedia taskgroup is a set of vocabularies designed to represent metadata for biodiversity multimedia resources and collections. The standardisation process is currently nearing completion. Since standardisation of this information is important for the purposes of ViBRANT, JKI and GBIF have invested resources in bringing the Audubon Core towards the present public review stage. The ViBRANT site “Species-ID” is hosting the vocabulary<sup>34</sup>.

## Task 2 Development of taxonomic reference lists (UvA, BGBM)

ViBRANT will add internationalisation and localisation to the ontology editor to facilitate the production of regional checklists and catalogues in co-ordination with PESI<sup>35</sup> – a Pan-European Species directories Infrastructure.

### Progress in Year 1

Tasks in WP4 have been clarified in intense discussions within the workpackage and with the project co-ordinator. This task (originally named “Development of taxonomic reference lists” and thus rather restricted in scope) has been considerably widened and has been renamed “**Standardised cross-platform integration of taxonomic data**”. The objective of facilitating the production of regional checklists and taxonomic catalogues in co-ordination with PESI – a Pan-European Species directories Infrastructure remains included, but will be based on an improved interoperability infrastructure created by ViBRANT.

Two workshops were held to solidify the somewhat diffuse focus of the entire workpackage (in February) and of this task (in June 2011), with follow-up discussions during the rest of the project year. In spite of the problems introduced by the change of leadership of the WP, the aims of this task are now clearly defined<sup>36</sup> and work in progress is fully in line with expectations.

To enable the exchange of data between Scratchpads and the Common Data Model (CDM)<sup>37</sup>, a suitable data format had to be chosen. BGBM and NHM agreed to use the extensible Darwin Core Archive (DwC-A) format developed by GBIF that aims to provide a stable, standard reference for sharing information on biological diversity. It is widely accepted in the taxonomic data-processing community and will leverage the usage of such a common access point to other applications.

As subtask 2.1, a module for Drupal 7, i.e. compatible with Scratchpads 2.0, has been written and enables a user to configure the DwC-A output generated by the export.

1. The module comes with a set of preconfigured views.
2. Pressing the button “Export to DarwinCore Archive” will execute the export routine and return the generated archive.

The routine can also be executed by requesting this URL:

[http://<scratchpads-base-url>/dwca\\_export](http://<scratchpads-base-url>/dwca_export)

### Future work.

We foresee the implementation of persistent identifiers in Scratchpads-2.0 for taxon centric and specimen centric data. Once this is in place we will be able to implement strategies for shared objects. The CDM already implements these identifiers in the form of UUIDs.

<sup>34</sup> <http://species-id.net/wiki/ViBRANT>

<sup>35</sup> <http://www.eu-nomen.eu/pesi/>

<sup>36</sup> <http://vbrant.eu/sites/vbrant.eu/files/2011-06-07-Berlin-ViBRANT-Paris-gh-wgb-gh-am-vu-rvl-fc-wgb.doc>

<sup>37</sup> <http://wp5.e-taxonomy.eu/blog/cdm>

Subtask 1, the functionality of the DarwinCore Archive module will be further improved, to allow flexible mapping of new Scratchpad data items and allow the integration of new DwC extensions.

Subtask 2: Implement a CDM datastore as a ViBRANT index (BGBM).

Subtask 3: Implementation of a CDM->Scratchpad and CDM->MediaWiki pipeline beyond simple DwC-A data items (BGBM). This also includes the PESI pipeline, to give integrated access to taxonomic resources.

Subtask 4: Descriptive features across ViBRANT environments (BGBM, UPMC, JKI). This will integrate the data-stores, software and services used for descriptive data based on the TDWG SDD standard (Structure of Descriptive Data).

### **Task 3 Standard data interfaces and APIs (JKI)**

ViBRANT will prioritise the development of Application Program Interface (API) data delivery services to support an open data strategy to enable ViBRANT users to deliver and expose their data to external users and services, including researchers and major global biodiversity informatics initiatives (specifically PESI, 4D4Life/i4Life, LifeWatch, GBIF, EoL, CDM and BHL; see letters of support).

#### **Progress in Year 1**

Ontology vocabularies are important, as discussed under Task 1 above. The possibility of using a MediaWiki system with its built-in collaboration and trust-building mechanisms, including full revision-control has been explored. MediaWiki provides semantic web extensions that are presently planned to be deployed in the near future on Wikipedia itself. The major extensions are SemanticMediaWiki and SemanticForms, that provide a markup-mechanism to enrich simple text data with defined semantic properties and combine human-readable text content with semantic markup. Semantic markup itself is used only to a limited extent for reasoning purposes within the wiki, but is exposed as RDF, allowing third parties to explore the full power of semantic machine reasoning.

The following progress has been made:

- The MediaWiki and server setup have been optimised for ViBRANT use.
- The Semantic MediaWiki extension have been installed and extensively tested.
- A prototype interface was developed on <http://species-id.net/wiki/> to illustrate the possibilities of a collaborative community interface for an ontology standardisation.
- An example vocabulary case using the taxpub vocabulary (together with the ViBRANT partner Pensoft) has been imported and annotated on the ViBRANT platform.
- As an external vocabulary to be used in the definition of new terms, the mapping relation definitions of the Simple Knowledge Organisation System (SKOS) have been imported. In SemanticMediaWiki it is possible to reuse external ontology vocabularies by creating a special import definition. Terms can then be related internally to each other by setting up sub-property relationships. Local term definitions within the Wiki system can be exported using the Resource Description Framework (RDF) export function of SemanticMediaWiki and thus they can be read for instance by RDF/ontology browsers.
- To facilitate data inputs by biologists, web forms are provided, helping users to fill in appropriate data without having to know the technical background or the syntax of semantic properties. Appropriate form data can be provided as selectable options or by saving input data internally in semantic properties and let them be proposed using automatic word completion while typing in words.

#### **Future work.**

We need to find a community of interested users with whom to collaborate to integrate this approach with the Drupal-based ViBRANT Scratchpads and the identification tools. Ideal would be a richly illustrated (or to be illustrated) glossary-like vocabulary that is available as open content (Creative Commons CC by, CC by-sa, or CC0).

## WP5 - Interaction and services

Lead: Neil Caithness, Oxford e-Research Centre

Activity type: Service

Partner	Name	Acronym	Effort (PMs Year 1)
4	Oxford e-Research Centre	UOXF.E9	12.8
6	Julius Kühn-Institute	JKI	6.5
11	Vizzuality	Vizz	4
13	Université Pierre et Marie Curie-Paris 6	UPMC	5
16	Universite de la Reunion	UdIR	1

### Objectives

As a workbench for taxonomic and biodiversity researchers, ViBRANT will integrate the most important services to deliver taxon identification and data analysis tools and facilities directly to Scratchpad users. Prioritisation is based on research of previous user requirements obtained from the EDIT pilot research, and will be adjusted based on the ongoing user requirement research pursued in WP3.

### Description of work and role of partners

#### Task 1 API for tools integration: design, specification and documentation (UOXF.E9)

The sustainable integration of external tools and services into the Scratchpad environment requires a well designed and documented API. In this task we will work closely with WP4, specifically their Task 3 (Standard data interfaces and APIs) to ensure that the tools API is consistent with development of a data API. The tools API specification and documentation will to be maintained under strict change control.

#### Progress in Year 1

A service has been mounted called OBOE<sup>38</sup> (Oxford Batch Operation Engine) that realises Task 1, although it will continue to evolve as services become available. The purpose is to run analyses that might normally be performed by a self-contained command-line application and to make them available via the web.

At the start of the project it seemed likely that there would be an external user interface for computing services. The final design and implementation of OBOE and the OBOE API has made this requirement obsolete. No external user interface is required and all user interaction is done via simple web-forms implemented in Scratchpads using information retrieved from the OBOE API queries for specific service types. At this stage there has been no need to collaborate with WP4 on standard API definitions.

#### Workflow

1. A Scratchpad (at NHM) posts a job request to the Job Scheduler (at Oxford). Authentication is simple as there is currently only one authentic user (i.e. the Scratchpad server at the NHM, but a small number of such clients when we begin to support external servers).
2. The Job Scheduler enters the job in its Job Table and retrieves data files based on information in the post. The job is sent to the appropriate service (at Oxford or elsewhere). On completion of the job the Job Scheduler updates the Job Table with a job completed status flag and an output url.
3. The Scratchpad server polls the Job Table until the completed status flag is returned and then fetches the output from the supplied URL.

<sup>38</sup> <https://vibrant.oerc.ox.ac.uk/>

### Implementation

1. Developed with Ruby on Rails with MongoDB as the database. This choice of database is anticipated as allowing more flexibility for dealing with job parameters than a standard relational database.
2. Scratchpad to post data to the Job Scheduler, comprising of job type, user id and a URL for the data files. The system will then wget the files. Post to be on port 443 using certificate authentication.
3. Information to be stored in a "jobs" table using a "job" model. This will contain:
  1. Rails job id
  2. NHM supplied user id
  3. Job type
  4. Location of data file(s) (URL)
  5. Job status (an integer, pending, sent, processing, finished, error etc.)
  6. An array of any parameters required for the job.
4. Whilst the job is running the Scratchpad should poll a URL using their job ID which would find\_by\_id and return the info. by rendering JSON.
5. Submission to the remote service would presumably have to be by means of a Rake task run from cron, to process any pending jobs and attempt to submit them to the processing service.
6. An implication for the Scratchpad server(s) is that the Scratchpad code would require forms for job submission to be added, and certificates would be needed on the server(s) for authentication against one held at Oxford.
7. Details of the data processing service still undetermined but we hope that it will be able to post queries back (see the separate data\_service\_post controller above).

OBOE documentation is attached to this report as Appendix 1.

### Task 2 Computing platforms deployment infrastructure (UOXF.E9)

In this task we will develop the software infrastructure, or middleware, required for hosting external tools on a variety of computing platforms: from dedicated high performance clusters (e.g. Oxford Supercomputing Centre - OSC), to distributed high-throughput grids (e.g. NGS and Campus Grids), to volunteer computing using donated desktop cycles (e.g. BOINC), to distributed cloud computing (e.g. Amazon Cloud-like services at the OeRC). The variety of computing tools that will be integrated requires a careful analysis of the computing platforms; their specific API requirements need to be addressed in Task 1.

#### Progress in Year 1

The platform requirements review has been superseded by full implementation for services on both Unix and Windows systems.

The OBOE system now consists of at least nine machines:

- Unix physical machines (Vibrant, Bonnacon, Cockatrice, Karakal)
- Unix virtual machines (Goblin, Voonith)
- Windows physical machines (Gecko)
- Windows virtual machines (Nyogtha, Syncerus)

We have in principle agreements and a free allocation of startup CPU hours to run services on the OSC clusters and on NGS resources.

A simple service API is now in place that will allow any authenticated Windows or Unix machine, located anywhere on the internet, to run OBOE services.

### **Task 3 Job description and metadata repository (UOXF.E9)**

In order to hold the specifications for tasks to be submitted to external computing services we will design a metadata repository to be integrated with Scratchpad instances. This repository will hold job specifications, and the history of results. It will be exposed to the Scratchpad user via the enhanced user interface developed in Task 8.

#### **Progress in Year 1**

The metadata repository has already been put in place as an array of replicated MongoDB databases. This is an integral part of the OBOE platform.

We should at some stage, though, review the job metadata that is being stored. This would be most appropriate with the review of user feedback.

### **Task 4 Phylogenetic services (UOXF.E9)**

As far as possible existing software for phylogenetic analysis (GARLI, PAUP, TNT, Mr Bayes) will be wrapped for deployment on the various platforms developed in Task 2. The metadata for job specifications will be encoded and stored in Scratchpad instances as defined in Task 3. These will specify the taxa and character data to be delivered to the phylogenetic service, and maintain the history of results returned.

#### **Progress in Year 1**

This task has made unsatisfactory progress thus far and is clearly going to require more time than was anticipated.

1. Many attempts to contact the Lattice project (<http://lattice.umiacs.umd.edu/>) have met with no response.
2. Likewise, attempts to contact LIRMM ([http://www.phylogeny.fr/version2\\_cgi/simple\\_phylogeny.cgi](http://www.phylogeny.fr/version2_cgi/simple_phylogeny.cgi)) has also been unsuccessful.
3. The Lisbeth software (<http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/lisbeth>) is not yet available for Unix platforms, though as OBOE now includes Windows VMs as available platforms, running Lisbeth as a service can be reinvestigated.

We have been investigating a one-click process from data retrieval (molecular) to alignment and phylogenetic analysis along the lines of 2 above. For data retrieval we have experimented with Genbank, for alignment with Clustalw and Muscle, and for phylogenies, Phylip. Automating a seamless process is not straightforward and without risks but we believe it is achievable and potentially useful within the OBOE framework.

### **Task 5 Bioclimatic modelling services (UOXF.E9)**

This task will build on existing algorithms (GAPR, MaxEnt, CSM) for bioclimatic-niche modelling of species distributions, and wrap them for deployment on the various platforms developed in Task 2. As before, this task builds on the infrastructure (tools API, computing platforms, metadata repository) developed in tasks 1, 2 and 3.

#### **Progress in Year 1**

This task is scheduled to begin in Month 13.

### **Task 6 Identification services for molecular data (UOXF.E9)**

In this task we will develop a service that integrates Scratchpad instances with NCBI BLAST and the Barcode of Life Database (BOLD) identification service. This will provide taxon identifications of DNA



sequences using the EMBL and BOLD reference databases that contain validated reference sequences of known taxa.

### **Progress in Year 1**

This task is closely associated with Task 4 and will follow when Task 4 is more mature.

### **Task 7 Identification services for morphological description (UPMC)**

This task will develop a service for the automated construction of taxonomic identification keys and natural language textual descriptions based on existing Xper2 key construction software. Morphological data for input to the service will be entered via the Scratchpad matrix editor, or derived from the descriptive ontologies developed in WP4. Results from the service (keys) will be returned to the Scratchpad instance for user access, publishing, and further analysis.

### **Progress in Year 1**

The IdentificationKey webservice<sup>39</sup> provides a service to create single-access keys on demand from a descriptive data stored as SDD format.

There are two main usage scenarios of this WebService :

1. A client component is integrated within the Scratchpads biodiversity networking tool, making the WebService available to Scratchpads user transparently.
2. The WebService can be called directly from a WebService client. We provide barebones clients written in Java and PHP, for both communication protocols supported by our WebService; SOAP and REST.

The new Scratchpads will be released at the end of January as a sandbox for people to test services.

### **Key generation algorithm:**

The method of creating the graph (the identification key) is a heuristic algorithm that selects, step by step, the "best" character to form a node of the graph according to the criteria to optimise, and then create branches from this node and continues the process in each branch until a stop criterion is reached.

Several parameter are available, in order to generate alternative keys in different formats. Details are in the online userguide documentation.

### **WebService installation:**

The WebService is coded in the Java programming language, using the J2EE framework, thus the host machine needs to have a recent JDK (v. ≥ 6) installed, as well as a J2EE web application server (e.g. Apache Tomcat, WebSphere or WebLogic).

The identification keys generated by the WebService are stored inside the web application server directory (in webapps/generatedKeyFiles by default). To avoid hard drive saturation, there is an automated deletion mechanism implemented in the WebService : every night the WebService deletes every file in webapps/generatedKeyFiles that is more than 30 days old.

### **Task 8 Enhanced Scratchpad user interface and matrix editor (JKI)**

In this task we will enhance the existing Scratchpad matrix editor so that it integrates with the metadata repository and other services defined in tasks 3-7. The editor will be a fully featured service to define morphological ontologies in a way appropriate for both taxonomists and applied users. It will also be appropriate for use with the machine-reasoning services to be developed in WP4.

---

<sup>39</sup> <http://identificationkey.fr>



### **Progress in Year 1**

This task is essentially a collaboration between two workpackages. The decision to upgrade Scratchpads from Drupal 6 to Drupal 7 meant that resources for this task were not available in WP2 in the first year. Resources at JKI were used to develop a prerequisite: a simple but powerful relational information model for matrix data, based on earlier DELTA and DiversityDescriptions models, but improved to capture essential newer developments and compatible with the SDD standard was developed and implemented in mysql. For this a simple user interface primarily focussed on quick by-taxon data capture was developed. First steps towards a full, js-based matrix editor, were evaluated, but some problems encountered.

### **Task 9 Visualization services (Vizz)**

In this task we develop tools for the visualization of geospatial and temporal data from Scratchpad instances. These tools will be integrated with the results returned by the analytical services developed in other tasks.

### **Progress in Year 1**

This task is closely associated with Task 10 , so both are reported together (below).

### **Task 10 Biodiversity patterns and indices (Vizz)**

Working closely with Task 9, and based on the standardised ontologies and concepts developed in WP4, this task will develop a series of taxonomic, phylogenetic and functional indices to suit the needs of the conservationists and spatial managers in collaboration with WP8. The service will provide a tool for addressing questions on the patterns biodiversity change and the consequences for the functioning of the ecosystems.

### **Progress in Year 1**

Vizzuality's participation in ViBRANT mostly revolves around the development of the GeoCAT Tool<sup>40</sup>, designed to produce rapid species level conservation assessments based on IUCN Red List Categories and Criteria<sup>41</sup>. Through an easy to use interface with a familiar Google Map underlay you can upload primary occurrence data for a species and at the click of a button calculate values relating to the geographic range of a species.

At the beginning of February 2011 we organised a meeting with different potential users of the tool to discuss what the tool could do and plan the next stages of development. A list of requirements was circulated before the meeting and another list collected during the meeting. The refined list is as follows:

- Read Primary data from scratchpads
- Visualise other layers like Protected Areas, Range maps...
- Overlay Niche models
- Visualization of the temporal component
- Be able to visualise results from other tasks within WP5

<sup>40</sup> <http://vbrant.eu/sites/vbrant.eu/files/M5-12.pdf>

<sup>41</sup> Bachman, S., Moat, J., Hill, A., de la Torre, J. & Scott, B. (2011). Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool. *ZooKeys* **150**, 117-126

## WP6 - Scholarly publishing

Lead: Lyubomir Penev, Pensoft Publisher

Activity type: Service

Partner	Name	Acronym	Effort (PMs Year 1)
1	The Natural History Museum, London	NHM	1
12	PENSOFT Publisher	PENSOFT	12

### Objectives

Achieve automatic submission, review and publication of species descriptions and taxonomic acts from Scratchpads and GBIF data sources to scholarly publishers. This builds on an existing prototype and will be compliant with the Zoological and Botanical rules of nomenclature that currently require taxonomic acts be published on paper and deposited in a minimum of five libraries. Develop new methods of XML tagging and semantic enhancements of biodiversity publications.

1. Streamline the scholarly publication process through the development of workflows within Scratchpads that reuse content present within the ViBRANT consortium infrastructure.
2. Facilitate the push-button submission of formatted manuscripts to paper publishers (initially Pensoft – publisher of the Open Access journals ZooKeys and PhytoKeys for animal and plant descriptions). This automation will significantly reduce the publishing costs for all stakeholders and will be used by other publishers in the future.
3. Simultaneously publish peer reviewed content in print and electronic form, with automated deposition of taxon descriptions to the Encyclopedia of Life framework and taxon names to PESI, 4d4Life, Zoobank and IPNI as appropriate.
4. Provide a highly automated workflow of publishing open peer-reviewed data papers from GBIF infrastructure, Scratchpads and other data networks through a newly established, open-access Biodiversity Data Journal, the first specialised data journal in biodiversity science.
5. Review and implement various semantic Web enhancements to taxonomic papers (e. g., data publication, cross-linking, data integration between papers, facilitated automated discovery, indexing and aggregation, online mapping, and others). Semantic enhancements increase the intrinsic value of taxonomic papers and multiply use/re-use of data and results from publicly funded research.

### Description of work and role of partners

#### Task 1 Workflow development and implementation (PENSOFT)

1. Develop and streamline a workflow that facilitates the automatic construction of taxon descriptions from content held within the ViBRANT consortium.
2. Develop and streamline the workflow that facilitates the automatic submission of data papers from content held within the ViBRANT consortium and external network systems.
3. Modularise these workflows to produce an XML document that can be submitted to PENSOFT and other scholarly publishers.
4. Work with other paper publishers (notably Zootaxa) to establish a competitive, low cost, Open Access market for the paper publication of descriptive taxonomic works. Because everything except the external peer review process will be completely automated we expect the article publication costs will be exceptionally low. Source XML documents will be available to any potential publisher in an attempt to create a pay-to-publish market amongst publishers of descriptive taxonomy.

### Progress in Year 1

The workflow extends the prototype developed under EDIT (grant No 018340) and illustrated in Bladederoov et al (2010)<sup>42</sup> Fig 5, has been refined from that experience and realised in Drupal 7 during this year.

A single Drupal module (called "Publication") has been prototyped to support the technical implementation of this workflow within the Scratchpads. This is available for Drupal 6 from the Scratchpad Git repository (<http://scratchpads.eu/develop>) along with other Scratchpad project written dependencies. Software dependencies include the Drupal community's Organic Groups module (<http://drupal.org/project/og>) and Content Construction Kit (<http://drupal.org/project/cck>) modules, in addition to the Scratchpad project's Species Profile Module (SPM) and Taxonomy Tree modules.

In summary the Publication module provides an intuitive interface that allows users to select and order content from their site and associate this with the publication, providing a many-to-many link between publication objects and other content types (e.g. Image, Bibliography). Thus for example, a single image can be used in many publications, and a single publication can have many images. The module also supports the communication between the user's Scratchpad and the publisher transferring the TaxPub XML representation of the manuscript to ZooKeys during submission, revision and final acceptance. TaxPub is an extension of the National Library of Medicine (NLM) / National Center for Biotechnology Information (NCBI) Journal Archiving Document Type Definition (DTD) for the markup of taxonomic treatments.

The workflow is illustrated in the example of a new polychaete species description by Faulwetter et al. (2011)<sup>43</sup>.

### Description of the workflow:

1. An author creates a Publication project within a Scratchpad to which only a restricted set of users have access. The author(s) also provide additional information required by the article (e.g., title, author's details etc.).
2. The author(s) prepare species pages (including descriptions, images, specimens etc.) within the Scratchpad. In case of a new taxon description author(s) use a temporary name (a placeholder). This placeholder acts as a surrogate for the final taxon name to ensure that the new name is not disclosed until the description has been accepted by the journal. The placeholder is linked (tagged) to data on their site, and the placeholder taxon name is linked to the final name. The author(s) select data to be included in the manuscript. Additional sections are added to the manuscript using a structure that will accommodate most taxonomic descriptions and images uploaded. When the preparation stage is complete, the author(s) preview the manuscript to make sure it is satisfactory.
3. Author(s) submit the manuscript, which creates an archive of the manuscript components. The submission process automatically generates an XML representation of the document according to the TaxPub DTD (<http://sourceforge.net/projects/taxpub/>). This document is then automatically sent to the journal ZooKeys. Other destinations will be possible when other journals accept TaxPub submissions.
4. ZooKeys organises the peer review. The reviewed paper, including reviewer's comments, is sent by e-mail back to the corresponding author.
5. Author(s) revise their manuscript and supporting data on their Scratchpad in response to the reviewers' comments (if necessary).
6. Author(s) re-submit the manuscript, which generates an updated XML file that is automatically sent back to ZooKeys. The publisher parses the final accepted XML document, adding additional XML

42 Blagoderov, V., Brake, I., Georgiev, T., Penev, L., Roberts, D., Rycroft, S., Scott, B., Agosti, D., Catapano, T. & Smith, V.S. (2010). Streamlining taxonomic publication: a working example with Scratchpads and ZooKeys. *ZooKeys*, **50**, 17-28.

43 Faulwetter, S., Chatzigeorgiou, G., Galil, B.S., Nicolaidou, A. & Arvanitidis, C. (2011). *Sphaerosyllis levantina* sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in *Sphaerosyllis hystrix* Claparède, 1863. *ZooKeys* **150**, 327-345.

mark-up for nomenclatural acts required by ZooBank registration, in addition to other semantic enhancements.

7. ZooKeys publishes the paper adding DOIs for the paper and supplementary material. The printed published paper includes a link back to the accepted manuscript on the Scratchpad. The Scratchpad version of this article also includes link(s) to the dynamic descriptions of each taxon page showing versions of updated descriptions if they have been edited after publication. New taxa descriptions are registered online by the journal's editorial office. In the future, ZooBank will provide receipt of an XML file from ZooKeys and create new records for published nomenclatural acts. The manuscript is submitted to PubMed /PubMedCentral for optimal distribution archival purposes.
8. The manuscript and all supplementary data are unlocked on the Scratchpad and made public on the day of printed publication. At this time the placeholder taxon names are automatically substituted by the final published taxon name.

By default all Scratchpad data concerning the ZooKeys publication are kept private for steps 1 to 8 and made public as step 9, although the original taxon pages are normally public. However, the author(s) have the capacity to make all these data public from the outset.

## Task 2 Extra-consortium data publishing (PENSOFT)

1. Develop a system to assign automatically Globally Unique Identifiers within the article XML document to each new taxon and / or nomenclatural act referenced in the text. These will be automatically submitted to the ZooBank database (for animals) and IPNI (for plants) on acceptance of the manuscript after peer review.
2. Automatically publish primary biodiversity data records identified within the article XML document through GBIF on acceptance of the manuscript after peer review.
3. Automatically submit new taxon descriptions to the Encyclopedia of Life with bibliographic links to the source article on acceptance of the manuscript after peer review.
4. Develop an automated registry site that announces new taxon descriptions (via RSS) published through the ViBRANT infrastructure. This site will include data visualisations showing geolocation of described taxa, the authors, specimens, and include links to the underlying source data.
5. Explore and implement mechanisms to institutionalise automated workflows with GBIF infrastructure, Scratchpads and other data networks for publishing data papers through the first ever, open access, open peer review Biodiversity Data Journal.

## Progress in Year 1

On 30th June 2010, anticipating the start of ViBRANT, ZooKeys published a special issue<sup>44</sup> '*Taxonomy shifts up a gear: New publishing tools to accelerate biodiversity research*' which marked the innovative publishing model, based on XML editorial workflow and on the TaxPub XML schema. From that time on, ZooKeys has been published in four formats – full-colour print version, PDF, HTML, and XML. This happened simultaneously with the implementation in the editorial process of the Pensoft Mark Up Tool (PMT)<sup>45</sup>, a program specially designed for XML tagging and semantic enhancements. Papers by Stoev et al. 2010<sup>46</sup>;

44 Penev, L., Roberts, D., Smith, V., Agosti, D. & Erwin, T. (2010) Taxonomy shifts up a gear: New publishing tools to accelerate biodiversity research. *ZooKeys* **50**, i-iv.

45 Penev, L., Agosti, D., Georgiev, T., Catapano, T., Miller, J., Blagoderov, V., Roberts, D., Smith, V. S., Brake, I., et al. (2010). Semantic tagging of and semantic enhancements to systematics papers: ZooKeys working examples. *ZooKeys* **50**, 1-16.

46 Stoev, P., Akkari, N., Zapparoli, M., Porco, D., Enghoff, H., Edgecombe, G., Georgiev, T. & Penev, L. (2010). The centipede genus *Eupolybothrus* Verhoeff, 1907 (Chilopoda: Lithobiomorpha: Lithobiidae) in North Africa, a cybertaxonomic revision, with a key to all species in the genus and the first use of DNA barcoding for the group. *ZooKeys* **50**, 29-77.

Blagoderov et al. 2010<sup>1</sup>; Brake and Tschirnhaus 2010<sup>47</sup>; Taekul et al. 2010<sup>48</sup> that used three different types of manuscript submission, exemplified the process.

Realising the importance of Wiki environment for popularisation and dissemination of biodiversity data, in April 2011 Pensoft took another major step towards modernisation of its journals. Species descriptions from 3 papers (Hendrich and Balke 2011<sup>49</sup>; Stoev and Enghoff 2011<sup>50</sup>; Bantaowong et al. 2011<sup>51</sup>) were integrated automatically on the day of publication to Species-ID, an open access Wiki-based resource for biodiversity information. This was achieved by programming a special tool, named Pensoft Wiki Converter (PWC)<sup>52</sup>, that transforms the XML versions of the papers into MediaWiki-based pages.

In October 2011, ZooKeys launched its multiple-choice model for publishing biodiversity data that provides a non-exclusive choice of mechanisms for the publication of data of different kinds and complexity, in co-operation with specialised data repositories and data aggregators, based on the previously published Pensoft Data Publishing Policies and Guidelines for Biodiversity Data<sup>53</sup>. One of the most important steps in this direction was the launch of an innovative route for publishing occurrence data and taxon checklists using an approved TDWG standard (Darwin Core), enriched metadata descriptions for published datasets, and the possibility of downloading both data and metadata in a machine-readable form, the Darwin Core Archive. This is supported by a specialised GBIF tool<sup>54</sup>, the Integrated Publishing Toolkit (IPT) and Scratchpads. Use of this tool allows the production of "Data Paper" manuscripts that formally describe a dataset's metadata as a peer-reviewed and citable scholarly publication<sup>55</sup>.

A second important element of the multiple-choice data publishing model of Pensoft's journals was the integration of its data publishing workflow with the Dryad Digital Repository, thus providing an option to its authors to archive data files of different kinds and complexity (e.g., phylogenetic, morphometric, ecological, environmental, etc.). The workflow of deposition of data in Dryad was illustrated by three exemplar papers published in ZooKeys 150 special issue (<http://www.pensoft.net/journals/zookeys/issue/150/>).

Pensoft's latest innovation was announced on the 22nd of November 2011 with the launch of an automated export and indexing of identification keys metadata published in the journals in KeyCentral – a global database of keys and other identification resources for living organisms (<http://keycentral.identifylife.org/Secure/KeyStore/List.aspx>)

- 
- 47 Brake, I. & von Tschirnhaus, M. (2010). *Stomosis arachnophila* sp. n., a new kleptoparasitic species of freeloader flies (Diptera, Milichiidae). *ZooKeys* **50**, 91-96.
- 48 Taekul, C., Johnson, N. F., Masner, L., Polaszek, A. & Rajmohana K. (2010). World species of the genus *Platyscelio* Kieffer (Hymenoptera: Platygasteridae). *ZooKeys* **50**, 97-126.
- 49 Hendrich, L. & Balke, M. (2011). A simultaneous journal / wiki publication and dissemination of a new species description: *Neobidessodes darwiniensis* sp. n. from northern Australia (Coleoptera, Dytiscidae, Bidessini). *ZooKeys* **79**, 11-20.
- 50 Stoev, P. & Enghoff, H. (2011). A review of the millipede genus *Sinocallipus* Zhang, 1993 (Diplopoda: Callipodida: Sinocallipodidae), with notes on gonopods monotony vs. peripheral diversity in millipedes. *ZooKeys* **90**, 13-34.
- 51 Bantaowong, U., Chanabun, R., Tongkerd, P., Sutcharit, C., James, S. & Panha, S. (2011). New earthworm species of the genus *Amyntas* Kinberg, 1867 from Thailand (Clitellata: Megascolecidae). *ZooKeys* **90**, 35-62.
- 52 Penev, L., Hagedorn, G., Mitchen, D., Georgiev, T., Stoev, P., Sautter, G., Agosti, D., Plank, A., Balke, M., et al. (2011). Interlinking journal and wiki publications through joint citation: Working examples from ZooKeys and Plazi on Species-ID. *ZooKeys* **90**, 1-12.
- 53 Penev L, Mitchen D, Chavan V, Hagedorn G, Remsen D, Smith V, Shotton D (2011). Pensoft Data Publishing Policies and Guidelines for Biodiversity Data. Pensoft Publishers, [http://www.pensoft.net/J\\_FILES/Pensoft\\_Data\\_Publishing\\_Policies\\_and\\_Guidelines.pdf](http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf)
- 54 <http://code.google.com/p/gbif-providertoolkit/>
- 55 Chavan, V. & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* **12**, S2.



### Task 3 Improve accessibility of taxonomic papers (PENSOFT)

1. Develop a mark-up schema optimised for semantic searching (in collaboration with WP7)
2. Develop and implement a workflow for public peer review to complement the existing form of conventional peer review.
3. Integrate the whole process with mark up of legacy literature (WP7) and ecological data mobilisation (WP8) on the basis of common standards and XML schemas.
4. ViBRANT needs a flexible, user-friendly ontology management environment, enabling users to create, define, extend and share their own terms and concepts where needed, providing options for discussions and annotation, while supporting re-use of terms from standardised ontologies wherever possible (via WP4 Task 2). For this purpose ViBRANT will extend the functionalities of both the ontology managers of existing vocabulary services (like GBIF) and will develop a collaborative community interface (JKI) for users and user-networks to facilitate the (bottom-up) definition and sharing of their ontologies in a user-friendly (non-technical) way.

#### Progress in Year 1

A report<sup>56</sup> and an open access publication<sup>57</sup> have been produced to serve as a basis for choosing the appropriate strategy in implementation of different XML schemas for mark up of taxonomic texts within the ViBRANT project and beyond. The three reviewed schemas – taxonX, TaxPub and taXMLit – cover the main tasks of taxonomy mark up rather well. TaxonX is a lightweight, object-centred schema focusing on taxonomic treatments extracted from legacy literature; taXMLit is a document-centred, very detailed schema covering mark up of legacy literature; TaxPub is an extension to the NLM journal publishing DTD and has been created to support prospective publishing in taxonomy. All three schemas have advantages and shortcomings outlined in the text; they also have passed the stages of creation, testing and implementing through a number of use cases.

The report does not recommend choosing one or some of the schemas. Rather, it proposes several cross-points that can be used to match common elements present in both legacy and present-day taxonomic literature. A common output, when needed, from documents marked up in the different schemas could be achieved through XSLIT conversions. Most important common elements in differently tagged text are: taxonomic names, taxon treatments, nomenclatural acts, literature references, as well as the overall structure of the published document and its bibliographic metadata. The report also outlines several questions to be answered when evaluating a certain schema to be used for the different goals of ViBRANT, Scratchpads and beyond.

The process of improving TaxPub was completed in 2011, when Pensoft finalised the testing and implementation of archiving of TaxPub-based articles in the PubMedCentral repository of the National Library of Medicine of the USA. This is the first case in the history of PubMedCentral, where a domain-specific XML schema (taxPub) is used, in the form of an extension to the NLM DTD, for archiving purposes and visualisation of some elements within the text, such as taxon treatments. Currently the whole content of [ZooKeys](#) and [PhytoKeys](#) is being exported and archived in PubmedCentral.

Automated export and harvesting mechanisms of bibliographical metadata is crucially important for increasing accessibility of taxonomic literature. In 2011, Pensoft successfully implemented harvesting mechanisms, based on OAI-PMH and POAI-MODS protocols to the CiteBank of the Biodiversity Heritage Library, Vifabio (an indexing service provided by a consortium of German libraries), Mendeley and others.

The aforementioned innovations and achievements during the first year of ViBRANT were published in a special issue of ZooKeys No 150 (Smith and Penev 2011<sup>58</sup>).

56 <http://vibrant.eu/sites/vibrant.eu/files/Milestone%206%2010-Review%20of%20mark%20up%20and%20tagging%20tools.pdf>

57 Penev, L., Lyal, C., Weitzman, A., Morse, D., King, D., Sautter, G., Georgiev, T., Morris, R., Catapano, T. & Agosti, D. (2011). XML schemas and mark-up practices of taxonomic literature. *ZooKeys* **150**, 89-116.

58 Smith V, Penev L (2011) E-Infrastructures for data publishing in biodiversity science, *ZooKeys* 150: 1-417.



## WP7 - Biodiversity literature access and data mining

Lead: David Morse, The Open University

Activity type: Research

Partner	Name	Acronym	Effort (PMs Year 1)
1	The Natural History Museum, London	NHM	1
9	The Open University	OU	13
10	Karlsruher Institut für Technologie	KIT	18.5
12	PENSOFT Publisher	PENSOFT	2.5

### Objectives

To facilitate the recovery of key content types (e.g. taxonomic treatments) and underpinning data from published documents, and the integration of those documents into the Scratchpad framework developed by ViBRANT WP2 to meet the needs of Scratchpad users. Workpackage 7 will make use of semantic search, data-mining and mark-up techniques to identify data elements within digitally readable text blocks, including OCR scans.

1. Develop the infrastructure to support the creation and ongoing maintenance of community constructed digital bibliographies within the Scratchpad virtual research environment.
2. Develop a robust, federated search mechanism and context-sensitive ranking of search results for biodiversity literature that is suitable for future application to other digitised literature resources.
3. Develop a web service to recover certain content elements, such as taxonomic names, author names and locality, from within text blocks.
4. Develop the means to identify structural elements (text blocks) of different types within published documents.
5. Develop the infrastructure to support annotation and correction of documents by citizen scientists and others.

### Description of work and role of partners

The biodiversity literature has a long and distinguished heritage and, unlike many disciplines, the old literature is still extremely relevant. Overall, the historical literature can be used to place new data in its proper context and inform management practices in modern concerns, especially biodiversity loss, land-use patterns, sustainability and climate change. For the information in the old (pre-digital) literature to be useful to biodiversity researchers, it needs to be made available in searchable, marked-up electronic formats. This provides the overall rationale for WP7.

### Task 1 Community constructed digital bibliographies (OU)

1. Develop aggregators to harvest bibliographic metadata from publicly accessible repositories of biodiversity science literature.
2. Develop software to identify and remove duplicate citations from the harvested metadata. In other words, de-duplicate reference lists whilst retaining links to external sources. Add stable, persistent identifiers where no identifier previously exists.
3. Identify and create linkages between a citation and a publicly accessible digital copy of the source material, resolving pagination where possible.
4. Provide Scratchpad users with access to the aggregated bibliography and to external initiatives as appropriate (e.g. CiteBank).

5. Solicit community help in developing a suite of test cases to test the de-duplication software by contributing examples of duplicates that they encounter.
6. Work with WP3 in producing the necessary training materials so that individual researchers can contribute their own bibliographies to the Scratchpad-hosted digital library. Work with WP4 and other relevant organisations such as TDWG to ensure that appropriate standards are adopted thereby ensuring interoperability between this and other digital libraries.

### Progress in Year 1

A review of publicly available data formats that are widely used for storing and exchanging bibliographic information has been completed<sup>59</sup>. The focus of the report is on taxonomic bibliographic data and it is not intended to be comprehensive, but it includes BibTeX, DublinCore, EndNote/RIS, MODS, Zotero, Citebank/GNUB information model.

The comparison focussed on the different fields and sub-structure of those fields (if any): for example, author fields may be decomposable in some formats but not others. Particular attention will be paid to any loss of information when converting between different formats. A sample document is encoded in each format. The sample was a letter submitted to Nature by Dave Roberts and Vishwas Chavan about the use of standard identifiers to mobilise data.

The review was used to select the output standards to be used in WP7, documenting relevant bibliographic formats and supporting services. A key service will be a comprehensive database of taxonomic literature, a bibliography of life. The review also includes a list of various bibliographic services, their capabilities and available functionality.

The review also formed the basis for one of our contributions to the ViBRANT special issue of Zookeys, *Towards the Bibliography of Life*<sup>60</sup>, which describes the issues and our proposed resolutions to provide a comprehensive bibliographic reference database.

The initial vision of the ViBRANT project hosting our own bibliography, with all the long term issues of supporting and running hardware and software that this would entail, was superseded when further investigating hosting of the bibliography. CiteBank<sup>61</sup>, the bibliographic offshoot of the Biodiversity Heritage Library<sup>62</sup> (BHL), was identified as a strong candidate with its strap line '*an open access repository for biodiversity publications*'. It matched the sustainable model for the comprehensive bibliography service provided by the Digital Bibliography & Library Project<sup>63</sup> in the domain of computer science. Work investigating the alternatives (such as using a commercially provided service such as Connotea <http://www.connotea.org/>, Zotero <http://www.zotero.org/> or Mendeley <http://www.mendeley.com/>) identified flaws of one sort or another and in the Summer of 2011 the choice of CiteBank as the repository of bibliographic information for ViBRANT was confirmed. Work began on automatically populating CiteBank with data from Scratchpads and other sources via the OAI-PMH protocol.

### Issues hindering progress

An early concern was the changing landscape in which the bibliography was to be delivered. There were two aspects to this concern.

#### 1. ViBRANT technical architecture

The design of the Biblio module is problematic and various solutions have been proposed within the biodiversity community, as well as the Drupal community at large, including modifying or even replacing this module. Effort was expended investigating this issue, effort which has subsequently not contributed to delivery of the bibliography as now envisaged.

<sup>59</sup> <http://wiki.scratchpads.eu/w/M710report>

<sup>60</sup> King D, Morse D, Willis A, Dil A (2011) Towards the bibliography of life. ZooKeys 150: 151-166. doi: 10.3897/zookeys.150.2167

<sup>61</sup> <http://citebank.org/>

<sup>62</sup> <http://www.biodiversitylibrary.org/>

<sup>63</sup> <http://www.informatik.uni-trier.de/~ley/db/>

## **2. CiteBank**

BHL's decision in late October 2011 to repurpose CiteBank purely as an index to its own content and not for CiteBank to serve as a repository of bibliographic references for the larger community was unexpected. The change in the functionality of CiteBank necessitated a change in plan within the Work Package that impacted directly on Deliverable D7.1 due at the end of November 2011. A clone of CiteBank was established using the code that BHL has made publicly available (<http://code.google.com/p/bhl-bits/>).

## **3. Staffing**

The late recruitment of both Guido Sautter (KIT) and David King (OU) caused minor disruption to the project timeline, now resolved. David King needed to take considerable amounts of time off for personal reasons in the autumn, which delayed resolving issues such as those related to CiteBank.

## **4. Support arrangements within the OU**

The Open University re-organised technical support with effect from 1st December 2011, including communication policy. Establishing the CiteBank clone in this disrupted environment was inevitably delayed.

### **Current status**

Given the issues with CiteBank that arose during October, we attempted to implement two interim solutions to meet the deliverable: a clone of CiteBank and our own Drupal service. These services will be made public as soon as possible, pending approval of the OU's Communication team.

Guido Sautter, as part of his work for ViBRANT and his previous work on processing bibliographic references, has been developing a suitable storage system. Emmanuel Müller and Mathias Bracht contributed to the details of both the design and the implementation. The code forms the basis for RefBank<sup>64</sup>. At present, this node of RefBank only contains about 12,000 records. We are adopting this code as our base infrastructure for the community constructed bibliography. Currently, we have one implementation within the OU. In the meantime, we are amending the branding of the code and web front-end to fit in with ViBRANT and are populating the database with more records.

We believe that using the RefBank code base will be beneficial because we can replicate content across all servers. Replication is built into the architecture of RefBank and the system is already running on two servers (three if you count the one that is internal to the OU). The ability to replicate content builds in recoverability to the service and also exposes the service to more potential users whose data, wherever uploaded, will contribute to the bibliography of life. We will explore the integration of RefBank services with the needs of individual Scratchpads.

### **The way forward**

We expect the OU implementation of RefBank to be available with a substantial number of references in December 2011 through our additions from existing publicly available resources. We are fortunate that because of the server architecture required by RefBank, our technical services have agreed to expose the service to the web using a temporary URI. This URI will be superseded by the correct one once the Communications section of our central technical support have completed the formal work in January 2012.

In the short to medium term, we anticipate making the following technical and functional improvements to the RefBank service.

1. Connect RefBank to the Scratchpad 2.0 architecture so that users of Scratchpads are offered the full functionality of the RefBank system, from within Scratchpads.
2. Develop file upload facilities so that users can bulk-upload references in certain formats.
3. Enhance the facilities with a "parse / re-parse this reference" mechanism, to be displayed in the search result page.
4. Add support for plain text export for those users who prefer not to work with a reference manager.
5. Expose the content to Mendeley.

---

<sup>64</sup> <http://plazi2.cs.umb.edu:8080/RefBank/search>

In the medium to long term, we will

1. Integrate the reference parser that is built in to RefBank with the reference discovery engine that is also being developed by WP7 so that users can extract references from papers hosted by BHL and other online repositories.
2. Port the system to PHP. We also intend to implement a different database engine, possibly MySQL, if we decide to retain the relational model or MongoDB if we opt for the NoSQL model. This port will add resilience to the service, each node using a different technology to deliver its share of the service.
3. Investigate the use of Apache SOLR for indexing if performance becomes an issue.
4. Investigate and implement exposing the content as RDF triples for data linking, as anticipated in agINFRA.

## **Task 2 Identification and mark-up of elements within documents (OU, KIT)**

1. Research and develop the means to use typographical information and other contextual clues to identify and tag document content by type.
2. Identify structural elements (i.e. typographical constructs such as titles, authors, bibliographies, tables and paragraph types) within articles and automatically assign tags (isolating specific terms, such as taxon names, personal names and localities) to articles to improve search.
3. Extend and integrate the GoldenGATE interactive mark-up tool within the Scratchpad infrastructure. GoldenGATE is our tool of choice because it has the mechanisms for handling the stylised structures common in taxonomic literature. Should integration of the complete tool prove difficult, GoldenGATE's modular structure will permit it to be decomposed so that individual modules can be integrated into the Scratchpad infrastructure or deployed as web services.
4. Manual mark-up, disambiguation and interpretation will be provided under Task 3; the potential for teaching the automated system through this process will be investigated.
5. Work with WP6 so that output from the historical literature is in XML format the same as or capable of being parsed to XML used in WP6, and aligned with the workflow developed in Task 6.1 for new literature so that the different document types can still be processed by a common workflow to achieve the benefits identified in Task 6.2, with manual rather than automatic transmission of records to GBIF, ZooBank and EoL. This will follow the standards identified in WP4.

## **Progress in Year 1**

An assessment<sup>65</sup> was undertaken of the range of options available when adding interactive mark up tools to the Scratchpads environment to permit taxonomists to mark up literature. For maximum flexibility we prefer to implement web services, allowing us to exploit the facilities of GoldenGATE while providing a front end more in line with users' current expectations. The service will also be accessible outside of Scratchpads, so helping to establish a sustainable service through the potential for a larger user base. The web service will also allow us more easily to exploit other web resources such as BHL and Plazi for additional information.

This work informed the review of options for XML mark up described in *XML schemas and mark-up practices of taxonomic literature*<sup>66</sup>, published in the ViBRANT special issue of Zookeys.

## **Use of typographical information to support mark-up**

GoldenGATE<sup>67</sup>, developed at KIT, is the tool of choice for identifying typographical information and other contextual clues (such as font changes, layout information and headings). As reported in Sautter et al

<sup>65</sup> <http://wiki.scratchpads.eu/w/M711report>

<sup>66</sup> Penev L, Lyal C, Weitzman A, Morse D, King D, Sautter G, Georgiev T, Morris R, Catapano T, Agosti D (2011) XML schemas and mark-up practices of taxonomic literature. *ZooKeys* 150: 89-116. doi: 10.3897/zookeys.150.2213

<sup>67</sup> <http://idaho.ipd.uka.de/GoldenGATE/>

(2009)<sup>68</sup>, while an experienced taxonomist can mark up materials more quickly when using the current semi-automatic GoldenGATE, it remains a relatively slow process and it will not scale to the needs of ViBRANT. GoldenGATE can, however, form the basis for the development of an automated tool and is thought to be suitable for enhancement by making increased use of typographical information. This assumption will be revisited as ViBRANT progresses and if automation proves too costly, particularly if the latest developments in document analysis and information extraction prove difficult to integrate, then a new tool will be developed.

This work is an underpinning for several milestones and deliverables that are forthcoming in the second year of the ViBRANT project for WP7. These milestones and deliverables are as follows:

ViBRANT output	Short description	Month due
MS7.16	Mark-up modules delivering outline mark-up e.g. for article boundaries, treatment boundaries, headings and authors	18
MS7.17	Review of pilot mark up processes within the Scratchpad infrastructure	20
D7.2	Mark-up modules	24

In MS7.16 the team responsible for WP7 will be developing software to identify document structure elements such as article boundaries, treatment boundaries, headings and other significant document structures automatically. In other words, we will extend GoldenGATE's semi-automated approach to the identification of document structures in two ways.

1. We will extend the software so that it identifies a greater range and complexity of document structures than it does at present. This is particularly important for historic literature with its variety of styles.
2. We will enhance the software so that it identifies these document structures automatically, with a greater level of automation and degree of precision than GoldenGATE is able to achieve currently. This is important to achieve the scalability sought in ViBRANT.

### Task 3 Disambiguation and annotation of mark-up (OU, KIT)

1. Develop a manual review and annotation module by extending GoldenGATE, or other open source distributed proofreading tool, to emplace the necessary infrastructure within Scratchpads to improve content marked up through automated techniques (from Task 2).
2. Users of the module will be able to take content produced automatically and review documents, provide expert interpretation and disambiguation as required. The system will allow use by (i) Scratchpad users as part of their mark-up process, (ii) citizen scientists contributing to the document mark-up process, scientists and amateur naturalists working to professional standards.
3. Work with WP8 to integrate the document mark-up module into the processes and workflows developed by WP8. Work with WP3 in producing the necessary training materials so that citizen scientists can contribute to annotation and correction of content marked up through automated techniques.

### Progress in Year 1

#### *Investigating the scale of the problem*

A suite of test cases has been designed to support the development of reference de-duplication software. The objective is to remove duplicate citations from metadata harvested from different sources, to avoid explosive inflation of record numbers and to improve the likelihood that references returned will fulfil the

---

68 Sautter, G., Böhm, K., Agosti, D. & Klingenberg, C. (2009). Creating Digital Resources from Legacy Documents: An Experience Report from the Biosystematics Domain In: *The Semantic Web: Research and Applications*, pp. 738-752.

user's requirements. To that end, we adopted two approaches to identifying test cases for the de-duplication software:

1. Task 1.5 to 'Solicit community help ...' has not been successful and we have been unable to mobilise community support for this activity.
2. Identify duplicates in 'publicly accessible repositories of biodiversity science literature' (Task 1.1).

Having built a repository for the bibliographic references (RefBank, see above), it needed to be populated. We have been developing import routines and aggregators that will load references from other repositories. Some of these repositories already contain duplicates so our first task is to identify duplicates within a repository, before tackling the problem of finding references to the same source material that appear in different repositories.

For example, the bibliographic references in both ITIS (Integrated Taxonomic Information System) and Catalogue of Life contain duplicate references.

#### *Duplicate references in ITIS<sup>69</sup>*

This example is taken from the October 2011 dump of the ITIS database. It is a typical example of duplicate references where one or more fields in the database are almost, but not exactly, the same. In this example the publication title appears in both abbreviated and full forms:

Annot. Zool. Japan. 35                      vs                      Annotationes Zoologicae Japonenses 35 (3)

and there are different page numbers

162-165    vs                      162-169

in full the two listings are:

PUB; "2"; "Matsumoto K. 1962."; "Two new genera and a new subgenus of the family Asellidae of Japan."; "Annot. Zool. Japan. 35"; NULL; "1962-01-01"; NULL; NULL; NULL; NULL; "162-165"; NULL; "1998-04-02"

vs

"PUB; "3"; "Matsumoto K. 1962."; "Two new genera and a new subgenus of the family Asellidae of Japan"; "Annotationes Zoologicae Japonenses 35 (3)"; NULL; "1962-01-01"; NULL; NULL; NULL; NULL; "162-169"; NULL; "2000-08-01"

#### *Duplicate references in Catalogue of Life*

The Catalogue of Life is available from <http://www.catalogueoflife.org/> and the example is taken from the 2011 edition of Catalogue of Life.

The following two references differ in language: the first reference is in Russian and the second is in French. As such, this example represents an angle on the concept of duplicate references not normally considered.

""44445""; ""Uvarov""; ""1929""; """"; ""Ezhegodnik Muz. Akad. Nauk SSSR 31""; ""N"

""45300""; ""Uvarov""; ""1929""; """"; ""Annuaire du Musée Zoologique de l'Acad. des Sciences de l'URSS (Ann. Mus. Zool. Acad. Sci. USSR) 31""; ""N"

We have shown that two large, well-known databases that are widely used by the biodiversity community contain duplicate references. While this is not surprising, it demonstrates the scale of the problem of duplication and that de-duplication should be a two-stage process. First, to identify duplicates within a single bibliographic data source and second, to identify duplicates between data sources, i.e. to de-duplicate a source and second to identify references missing from our Bibliography of Life for merger.

---

<sup>69</sup> <http://www.itis.gov/>



The second example also demonstrate the challenge posed by different languages. One of the resources that we intend to develop in WP7 is a look-up service of journals, their abbreviations and variants. Ideally, as illustrated by the Catalogue of Life example, such a service should also provide language variants too.

#### **The way forward**

We will evaluate the cost-benefit ratio for resources invested in de-duplication software and examine approaches from other aggregators such as Mendeley.

#### **Task 4 Web service for search and information retrieval (OU)**

1. Develop the means to (semi-)automatically data-mine marked-up text to extract structural information on the location of biodiversity elements (predominately species), taxon names, person names, publication metadata etc.
2. Make the data available as a web service to the central Scratchpad infrastructure and to other applications such as those developed in WP8.
3. Work with WP2, WP4, WP6 and WP8 to ensure interoperability of the service with the Scratchpad infrastructure, with external digital libraries and applications wishing to use the web service.
4. Work with WP2 to ensure that extensions to scratchpad infrastructure are compatible to WP 6 and WP7.

#### **Progress in Year 1**

This task has been deprecated because it is being delivered within the re-structured Task 1 (see above) and Task 2. All elements of Task 4 will be delivered as part of the on-going work programme.

## WP8 - Ecological and conservation data mobilisation

Lead: Christos Arvanitidis, Hellenic Center For Marine Research, Crete Activity type: Co-ordination

Partner	Name	Acronym	Effort (PMs Year 1)
2	Hellenic Center For Marine Research, Crete	HCMR	18
7	Museum für Naturkunde, Berlin	MfN	1
12	PENSOFT Publisher	PENSOFT	0.5
14	Global Biodiversity Information Facility	GBIF	4

### Objectives

Extending the capacity of the Scratchpads (in collaboration with WP2, WP4, and WP5) into ecology and conservation science.

1. To facilitate naturalist citizen scientists to contribute their observational information to the e-infrastructure.
2. To improve data quality of information present in thematic and global biodiversity databases.
3. To boost GBIF network expansion at national and regional scales.
4. To integrate data from various biodiversity sources for ecological and biodiversity assessment.
5. To facilitate the Red List assessment processes for endangered species by IUCN and other stakeholders.

### Description of work and role of partners

#### Task 1 Environmental data recording module for general naturalists (GBIF)

Develop a module that enables naturalists to record and monitor observations of biological diversity and associated environmental variables (e.g. temperature, habitat). This will integrate with national biological recording programmes, for instance the UK's Biological Records Centre and Joint Nature Conservancy Council, and will be developed in co-ordination with the relevant European biodiversity monitoring projects and initiatives such as the ESF recommended "Citizens Monitoring Biodiversity" project, EDIT's WP7 (All Taxon Biodiversity Inventories + Monitoring) program and the EuMon project. A targeted pilot project will be developed and applied on the biodiversity data flow from diving clubs. This module will also be used to enhance the contributions made by GBIF's network of National Nodes. To encourage participation, APIs will be developed (linked to WP2.3 and WP4) in order to facilitate submissions via other web tools that are commonly used by citizen scientists (e.g. Flickr). The Scratchpads module itself will also contain an API to allow it to be integrated not only within a Scratchpads environment system on other websites as well. These tools will also be used to enhance the contributions made by GBIF's network of National Nodes.

#### Progress in Year 1

COMBER<sup>70</sup>: Citizens' Network for the Observation of Marine BiodivERsity was created.

#### The need for a citizen scientists approach in marine biodiversity:

Most biodiversity information and data are collected in the framework of comparatively short-term research and monitoring projects, producing datasets that are predominately discontinuous or unevenly spread, geographically, temporally or ecologically. The latter is most obvious in marine biodiversity data where the costs are many times higher than for terrestrial cases, due to the diverse and expensive access issues and the need for specific sampling gear. If the decreasing trend in the numbers of professional taxonomists is

<sup>70</sup> <http://www.comber.hcmr.gr/>

considered, then only the development and mobilisation of citizen scientist networks seems to offer a sound and sustainable solution to continued biodiversity data-gathering.

Although several international projects targeted at continuous data collection from specific habitats have been launched in the last couple of decades, such as the NaGISA project<sup>71</sup> (National Geography in Shore Areas), the implementation of citizen science in the marine environment currently faces two difficulties:

1. only the tidal zone can be approached by all citizens, and
2. the maximal depth safely reachable by recreational SCUBA divers is limited to 40 m.

#### **Basic concept and design of the COMBER (Scratchpads) modules:**

The COMBER pilot project attempts to address the problem by engaging the broader community in marine biodiversity research data and information collection. This pilot project taps into a suite of developments aimed at supporting virtual research communities in biodiversity science.

COMBER aims to engage citizen scientists, that is, those interested in coastal marine biodiversity and willing to participate in an observation network. It is currently operating in the Cretan Sea and Southern Aegean (Greece) with the potential to expand to the whole Mediterranean basin or any other recreational diving area. The basic characteristics of this pilot project are:

1. a web site that functions as the main communication and promotion vehicle of the network, offering data-entry tools for collecting information that, at a later stage, are channeled to large data aggregators (e.g. GBIF) and publication media (e.g. PENSOFT);
2. a well-defined scientific hypothesis which has been formulated to be tested with the collected data;
3. a focus on fish species;
4. a suite of tools, such as a waterproof identification guide, on-the-spot professional introductory lectures, underwater training, and demonstration of web site usage as well as data entry which are used to facilitate *in vivo* identifications by participating divers;
5. collaboration with two commercial diving centres in order to ensure operational safety and to explore the market development potential for the sustainable continuation of the initiative;
6. exploration of new services and tools to enhance the SCUBA diving and snorkelling services which are targeted towards the tourism industry.

Like Scratchpads, COMBER uses Drupal to perform all underlying functionality of the system. This allows full interoperability with Scratchpads and ViBRANT platform which are based on the same software. Many elements of the site, such as user management, profile creation, image galleries and discussion fora, have been created using built-in features or readily available Drupal modules.

#### **Functioning of the modules and their operation in the citizen science network:**

The web site has a simple and flexible structure. On the front page the user is prompted to log in, in order to make use of the special features and options of the site. Users can also log into the site with their Facebook account, a valuable feature to promote the registration process. On the front page, information on various aspects of the COMBER project, such as a brief message on the concept and operation of the project, is also published. The top menu provides additional information on the project, including:

- news;
- answers to the potential “what”, “where”, “how” and “why” type of questions;
- ways in which the user can participate to the project;
- photos of the most common fish species in the Mediterranean coastal habitats;
- links with the activities of other projects and initiatives;

---

<sup>71</sup> <http://www.nagisa.coml.org/>

- a forum page, where the user may find discussions on various topics of local interest.

As soon as the user logs into the web page a new menu, on the left, opens offering options for the user:

- to see and edit their profile ("My profile");
- an overview, with an "edit" option, to their dives ("My dives");
- a form to submit information on a new (recent) dive, not yet registered on the COMBER's system ("Submit a new observation");
- to provide information on their experience ("Comment your experience"), although this information is supposed to be submitted only once there is an "edit" option;
- to have an overview of the top contributors along with their accreditation (number of stars, see below) automatically provided by the COMBER system ("Top contributors").

Registered users can continue to contribute data after participation in the seminars, use the diving log to keep track of their dives and species observations, upload photos of fish species and discuss various topics in the fora. A competitive element is introduced by a five-star ranking system indicating the activity level of the user: the more dives with fish observations are contributed to the system, the higher the user ranks in a "Top contributors" list, thus providing a playful incentive.

Recreational divers that participated in the COMBER initiative uploaded more than two thousand fish species records over the functioning of the project of this first year.

A paper describing the design, current, and future implementation of the project has been published in Zookyes: doi: 10.3897/zookyes.150.2149.

#### **Future plans:**

GBIF (Global Biodiversity Information Facility) has initiated a community driven project called the 'Nodes Portal Toolkit' (NPT) that should enable communities to deploy, maintain, and extend biodiversity data portals. The platform will also allow community development of new modules with extended functionalities. The initial implementation of the NPT will be built around Scratchpads, linking well with developments in ViBRANT platform. A second version will have extended functionalities modelled on the OBIS data portal. COMBER is expected to be integrated with both ViBRANT and the NPT, offering interested parties a ready-made installation file allowing them to set up and deploy their own citizen-science portals without prior technical knowledge or an on-line application form for a site hosted by the ViBRANT consortium.

From a research perspective, after data cleaning, the data are available for use (and re-use) in testing the hypothesis that data collected by recreational divers are suitable for biodiversity monitoring. Recent biodiversity measures, based on species relatedness such as the taxonomic distinctness, provide the concept to formulate and test the hypothesis.

COMBER will run over the next year of the project, specifically over the high-activity summer months, after which its cost effectiveness and sustainability will be evaluated.

#### **Task 2 Data quality improvement module (GBIF<sup>72</sup>)**

A Scratchpad module will be developed to enable experts or expert communities to supplement and correct errors in data from GBIF and other sources such as OBIS, PESI and LifeWatch, based on WP4.1 and 4.2 standards. Data served by GBIF data are principally of high value for biodiversity and ecological research / applications, but the original resources are of variable quality. Through a dual approach employing both computational and human powers, experts will be enabled to annotate or correct the data. An algorithm will filter the data for "suspicious" records to be reviewed by an expert. On the other hand, corrections made by experts will be fed into a knowledge base that allows the algorithm to "learn" and to be automatically improved. This algorithm will be linked to the mechanical Turk (WP2). Corrections and annotations can then

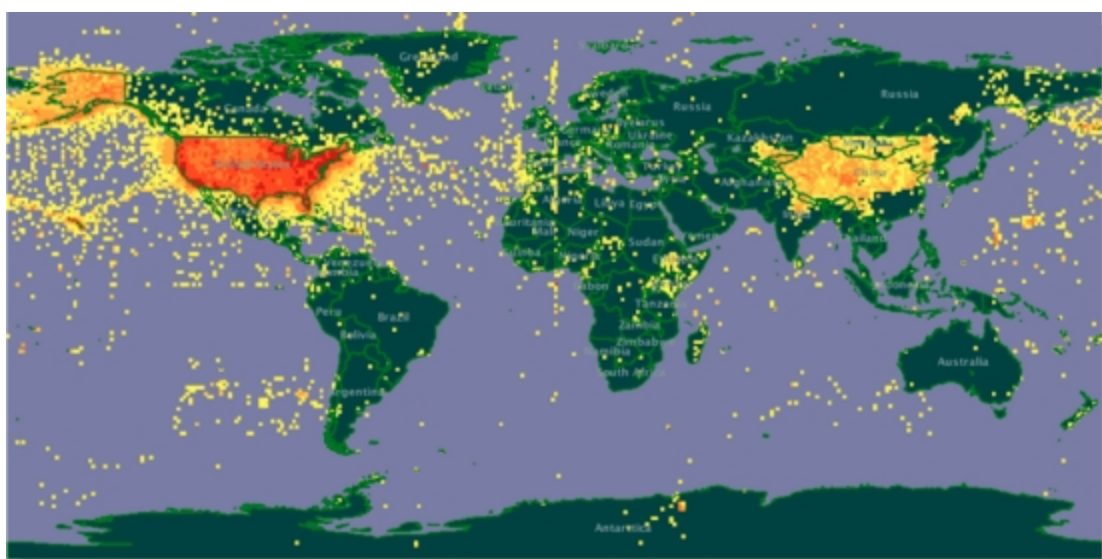
---

<sup>72</sup> In the technical annex this task is mistakenly assigned to HCMR but, in fact, it has carried out by GBIF. This had no impact on the timing, quality, and resources spent.

be fed back to the original provider. Finally, through Biodiversity Data Journal established within ViBRANT, taxonomic and ecological data could be submitted from Scratchpads through the metadata catalogue of GBIF and turned into peer-reviewed data papers, providing option to the authors and data owners to benefit from the credibility, priority registration and citability of their work (link to WP6, Fig. 3)

### Progress in Year 1

The approach described for this task was modified. GBIF have implemented data quality algorithms as illustrated by David Remsen's talk in TDWG 2011<sup>73</sup>. To illustrate what has been achieved, GBIF have improved geo-referencing processes that enable better matching of records to their intended country of origin. The map below shows raw data originating in the United States.



The large set of data appearing in central Asia have errors in the longitude relating to E-W notation rather than 360 degrees notation. After correction, the data are all located within the USA.



This map demonstrates the effect of data correction. We can now recognise international waters and offshore islands.

<sup>73</sup> <http://www.slideshare.net/DavidRemsen/tdwg-1remsen>



Similar data cleaning algorithms have been applied to taxonomic hierarchies, which avoid taxonomic classification errors, such as a European bird species was mistakenly placed in the hummingbird family (restricted to the Americas). Similar processes are used to link latinate names to common names which makes display of information much more intelligible to the majority of users.

### **Task 3 Ecological and conservation data applications module (HCMR)**

A special module will be developed in the Scratchpads which will allow to make use of the available data resources (e.g. GBIF, OBIS, LifeWatch) and calculate various ecological quality / biodiversity assessment indices based on the services to be delivered by WP 5.4 and 5.6. The data used can primarily consist of either species or functional traits occurrences along with their higher classification levels or abundance/biomass data. A subset of the indices will be chosen to address the relevant EU legislation (e.g. Habitats Directive, Water Framework Directive) on ecological quality issues. This work will also be based to the results delivered by other EU funded projects (e.g. ALARM, ALTERNet) and linked to the activities of major environmental agencies such as EEA and ICES. Therefore, the module will be also able to be linked to other ecological quality index calculation web tools (e.g. M-AMBI).

#### **Progress in Year 1**

Although work on this task is scheduled to begin in Month 15 (February 2012), substantial progress has already been made in the first year of the project.

First, the type of indices to be used have been identified. Two broad categories of indices that can be applied to normal sampling data have been chosen: (a) sample-size, sample effort dependent (Species Richness (ad hoc code), Number of individuals (ad hoc code), Pielou's evenness (ad hoc code), Shannon Index (ad hoc code)), (b) sample-size, sample effort free indices (ES 50, taxonomic distinctness). These are widely used in the marine habitats.

The routines for calculating indices are written in C and can be executed on any machine with a C compiler.

Over the next year, routines for selected indices of biodiversity and ecological health assessment for terrestrial habitats will be developed. Issues to be solved are those of the processing demand and parallelising, which may be required (openParallel, GPU coding, MPI) to offer fast user service.

### **Task 4: Ecological and conservation visualisation and assessment module (Vizz)**

Based on the visualisation module developed in WP5.5, this task will extend its functionality for ecological and conservation purposes. The tool will enable the visualisation and cross-analysis of Scratchpad generated data with existing data sources, like the World Database on Protected Areas, Range maps of endangered species published in the IUCN Red List and biodiversity primary data available on the GBIF network. It will allow also the overlay of information taken from existing niche modelling projects, like Aquamaps or LifeMapper, for current and climate change projections. The displayed data will be possible to export as a set of reports, taking in consideration the geo temporal aspects of the data. Specifically one report will be developed to support the IUCN threat assessment criteria and facilitate decision making on conservation actions taken by formal peer review. Ideally the module will integrate with IUCN's Species Information System (SIS) used on the expert review process. The tool will access most of its data using OGC standards in order to allow integration with multiple geospatial services as model by the INSPIRE directive and also for LifeWatch.

#### **Progress in Year 1**

This task is effectively the same as the re-structured Task 9 in WP5 in the first year. See WP5 for progress report. In WP5 the functionality of the module is the prime concern whereas in WP8 the focus is on its application to ecological and conservation datasets. In the initial stages of development these two threads are inseparable.



## Task 5 Field recording applications module (MfN)

Targeting communities and stakeholder groups actively involved in biodiversity inventory and monitoring activities from an ecological and conservation perspective, a separate Scratchpad module will be developed to promote and test digital field recording devices, tools, and technologies. For scientific projects as well as for citizen science activities engaged in recording, documenting and monitoring biodiversity in the field, efficient and standardized digital data capture remains a major bottleneck in streamlining the work flow from primary data collection to subsequent analysis. The module to be developed will provide information about and access to available field recording tools and technologies, assess and review the usability and cost efficiency of these devices, and support the development and adaptation of specific software applications, especially for conservation-oriented biodiversity monitoring schemes. For developing pilot applications, this module will build on results from projects such as EDIT (All Taxa Biodiversity Inventory + Monitoring program), ALARM, and MARBEF. The task will be implemented in direct collaboration with WP3 (tasks 3.2, 3.3), and WP4 (tasks 4.2, 4.3, 4.4).

### Progress in Year 1

A review of mobile phone apps<sup>74</sup> suitable for field recording classified them into 4 groups:

- |                    |  |
|--------------------|--|
| guides             | - a pure informational purpose. The user can search for species names and get some information and multimedia material about the species found. Thus, the user has to know what he or she is looking for, like an encyclopaedia. These apps are essentially compact multimedia nature guides.          |
| determination apps | - helping the user to determine a certain taxon. A lot of apps for different groups of organisms can be found, but fewer apps that combines the identification of groups in a single key. Users interested in various groups find themselves having lots of apps each for a certain group of taxa.     |
| recording apps     | - it is convenient to record biodiversity data directly electronically. This can be done either via simple checklists, web forms optimised for mobiles or interactive and/or multimedia features. Such forms may also include images. Some advanced apps have the ability to develop customised forms. |
| allrounder         | - multi-functional tools and have no strong focus on one particular functionality, but on several. It is important to stress that the definition of this category refers only to the functionality and not to the taxonomical scope.   |

The classification of biodiversity apps has shown that a lot of tools exist and can be applied in different scopes. Most of them address citizen scientists and can be used in schools and from interest groups, but are also able to satisfy the demand of scientists. The linkage of different applications has to be encouraged to reinforce the usability of the high potential this mobile software section has as a whole.

For scientists smartphone apps can simplify the acquisition of data. For citizen scientists these apps are informative, funny and make the subject of biodiversity more on the spot in people's daily routine.

### Future work

Interfaces are necessary to connect apps and their data. So, if data transfer between different apps worked in a general way, the user would not have to worry about using different apps for different demands. Such an approach could be easily realised (especially with construction tools) by using an international standard like the ABCD-schema<sup>75</sup> or DarwinCore<sup>76</sup> as a kind of uniform "language" for biodiversity apps, provided the standards contain sufficient flexibility to handle the data being collected. The establishment of uniform import and export formats would also promote the cross-linkage as well as data migration for both, simple and complex biodiversity data. Networks like the Global Biodiversity Information Facility (GBIF) might promote such an approach.

<sup>74</sup> <http://vbrant.eu/content/ms811-review-relevant-field-recording-tools-and-applications>

<sup>75</sup> <http://www.tdwg.org/activities/abcd>

<sup>76</sup> <http://www.tdwg.org/activities/darwincore>

### 3.2.3. Project Management

#### **Consortium management tasks and achievements**

The Consortium Management Committee has met 4 times in the first year, as anticipated. Minutes are available to consortium members on the project web site (<http://vbrant.eu>).

The strategic board has met once, at the beginning of the project, to validate the overall direction of the project and establish immediate priorities. The board has not met since, because no substantive issues have been identified for discussion.

#### **Communication between beneficiaries**

The primary achievement of the management committee has been an integration of activities between workpackages. For example, WP6 (Scholarly publishing) and WP7 (Biodiversity literature access and data mining) have held joint workshops and refined a common mark-up standard in TaxPub. WP6 is feeding citation information back to WP7 to contribute to the Bibliography of Life.

WP2 and WP3 are cooperating closely over the development of Scratchpads 2, specifically in developing the user interface, the training materials and the on-line help system. They have also collaborated to create a database<sup>77</sup> of user access records combined with user registration that facilitates analysis of user activity, which in turn will provide information on workflows and is expected to guide strategy for developing the user base.

WP2 and WP4 are collaborating to establish standardised information flows between Scratchpads and other systems, particularly the EDIT Platform for Cybertaxonomy.

WP3 and WP7 are collaborating to apply data mining techniques to web traffic data to produce a greater understanding of the Scratchpad user base.

WP4 and WP2 are working closely to develop a Drupal module for Darwin Core Archive export of Scratchpad content.

WP4 and WP5 are collaborating on further development of the service environment for descriptive data based on the TDWG SDD standard (Structure of Descriptive Data).

WP4 is working with WP7 to apply vocabularies to semantic entity extraction and automatic mark-up of documents.

WP6 and WP8 have been collaborating on mechanisms to publish data gathered by the ecological and conservation communities.

WP6 is collaborating with WP4 over the development of the wikimedia platform SpeciesID by automating the transfer of information from newly published documents into the web site

WP8 and WP5 are collaborating over services relating to the calculation of biodiversity indices and the application of high-resolution 3-D imaging techniques as an aid to identification.

#### **Co-operation with other projects**

ViBRANT has been co-operating closely with the Biodiversity Heritage Library (BHL<sup>78</sup>) particularly through WP7 in the construction and population of CiteBank (see WP7 report above). Collaboration with BHL is on-going.

Plazi<sup>79</sup> is an open repository of structured, biodiversity-related publications. Donat Agosti, the project leader and is a member of ViBRANT's advisory board. Plazi is heavily engaged with both WP6 and WP7 in the development of TaxPub DTD, also used by WP2 in the Scratchpad Publication module.

<sup>77</sup> [http://vbrant.eu/sites/vbrant.eu/files/Report\\_year1\\_%20IntegratedUserDatabase\\_final\\_0.pdf](http://vbrant.eu/sites/vbrant.eu/files/Report_year1_%20IntegratedUserDatabase_final_0.pdf)

<sup>78</sup> <http://www.biodiversitylibrary.org/>

<sup>79</sup> <http://plazi.org/>

Encyclopedia of Life (EoL<sup>80</sup>) closely modelled their LifeDesks on Scratchpads as a means for users to contribute data to the encyclopaedia. Scratchpads now export structured information to EoL and we plan to migrate the active LifeDesks into the Scratchpad environment in the coming year, which will be achieved by WP2.

LifeWatch<sup>81</sup> is an ESFRI project that aims to be a world leading e-infrastructure to support all aspects of research on the protection, management and sustainable use of biodiversity. It began its construction phase in 2011 and has been heavily engaged in marketing to get the necessary financial support. The Scratchpads have been planned as a prototype for the primary user interface. Wouter Los, the LifeWatch project leader, is also the current leader of ViBRANT WP4.

BioVeL (Grant No. 283359) is intended to meet the needs of Europe's Biodiversity Science research community with tools for pipelining data and analysis into efficient workflows, urgently needed to understand biodiversity in a rapidly changing environment. It customises, deploys and supports the Taverna / myExperiment / BioCatalogue family of software to achieve this. Such workflows will be supported through WP5's OBOE service (see WP5 report) and Neil Caithness, WP5 leader, was invited to a BioVeL workshop in late 2011 focussed on bioclimatic modelling. A further meeting is planned in January 2012 between BioVeL, agINFRA and ViBRANT.

agINFRA (Grant No. 283770) is a new project that started in November 2011 intent on creating a data infrastructure to support agricultural communities. Scratchpads will be used to supplement agriDrupal by combining the special module written for agriDrupal into a Scratchpad instance. WP7 is engaged with agINFRA because the ViBRANT partner the OU is also a partner in agINFRA focussing on mobilising legacy information.

4D4Life & i4Life: the pipelines build in ViBRANT to export/import PESI data might also be used to transfer CoL data, since BGBM is in charge of i4Life's objective of connecting both (global & European) taxonomic hubs. So a generic use of both infrastructures broadening the functional scope. In 4D4Life we intended to discuss the potential use of Scratchpads (1) to establish proto-GSD for those group absent in CoL, gathering data from other taxonomic resources and building an ad-hoc expert network to sort out the taxonomy, and (2) providing an alternative data management environment for already existing GSDs. Unfortunately the untimely death of Prof. Bisby at the end of October has delayed such a discussion.

### **Problems which have occurred and how they were solved or envisaged solutions**

The most significant problem faced by the consortium was the loss of the WP4 leader during negotiation phase. His place was ably filled by Olaf Bánki, also from UvA and associated with the ESFRI LifeWatch project. Olaf, in turn, resigned from UvA to take a job in GBIF (also a ViBRANT partner) but with duties that were incompatible with WP4. Wouter Los then took over as WP4 leader. Shortly after that, Yde de Jong's position in UvA stabilised and he has re-engaged with the project. The change of leadership has inevitably meant that the primary focus of the workpackage has been somewhat diffuse. The involvement of the original architect of the workpackage brings a welcome return of focus.

BGBM have had serious problems over staff recruitment and retention (developers) in the first year of ViBRANT, although they have managed to meet all their commitments, they have not spent as much resource as allocated. For WP4, it was agreed to transfer part of the BGBM funding from BGBM primarily to UPMC, who were underfunded as a result of budget restrictions in the project negotiation phase, and to JKI.

### **Changes in the consortium**

The Royal Netherlands Academy of Arts and Sciences (In Dutch: *Koninklijke Nederlandse Akademie van Wetenschappen*, abbreviated: *KNAW*) was added to the consortium to correct an anomaly with the employment status of one project member, Prof. Peter van den Besslaar. The purpose of the change was to allow proper payment for Prof. Besslaar's time, which in the Grant Agreement was funded at VU. The change affects VU's nominal budget, but has no other impact on the consortium.

<sup>80</sup> <http://eol.org/>

<sup>81</sup> <http://www.lifewatch.eu/>

### List of project meetings, dates and venues

Meeting	from	to	location	url
Strategic Board meeting	19/01/11	19/01/11	Paris	<a href="http://vbrant.eu/content/strategic-board-meeting">http://vbrant.eu/content/strategic-board-meeting</a>
Management Committee 1	19/01/11	19/01/11	UPMC, Paris	<a href="http://vbrant.eu/content/management-committee-meeting">http://vbrant.eu/content/management-committee-meeting</a>
Scripting Life' ViBRANT kick-off meeting	20/01/11	21/01/11	MNHN, Paris	<a href="http://vbrant.eu/scripting_life">http://vbrant.eu/scripting_life</a>
WP4 workshop	18/02/11	18/02/11	BGBM, Berlin	<a href="http://vbrant.eu/sites/vbrant.eu/files/Minutes%20Wp4%20meeting%20Berlin%2018%20Feb%202011.doc">http://vbrant.eu/sites/vbrant.eu/files/Minutes%20Wp4%20meeting%20Berlin%2018%20Feb%202011.doc</a>
WP3 workshop	25/03/11	25/03/11	RBINS, Brussels	<a href="http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_Summary_meeting%20WP1-WP3_RBINS_250311_D.pdf">http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_Summary_meeting%20WP1-WP3_RBINS_250311_D.pdf</a>
Management Committee 2	20/04/11	20/04/11	virtual	<a href="http://vbrant.eu/content/man-comm-2">http://vbrant.eu/content/man-comm-2</a>
WP2/WP3 workshop	03/05/11	03/05/11	NHM, London	<a href="http://vbrant.eu/content/55/vibrant-notes-data-meeting-3511">http://vbrant.eu/content/55/vibrant-notes-data-meeting-3511</a>
WP8 workshop	12/05/11	13/05/11	Heraklion, Crete	<a href="http://vbrant.eu/content/60/wp8-crete-meeting-minutes-and-presentations">http://vbrant.eu/content/60/wp8-crete-meeting-minutes-and-presentations</a>
WP4 workshop on descriptive data	06/06/11	07/06/11	BGBM, Berlin	
WP6/7 workshop	07/06/11	07/06/11	NHM, London	<a href="http://vbrant.eu/sites/vbrant.eu/files/WP6-WP7%20meeting%20Agenda%20v4_0.doc">http://vbrant.eu/sites/vbrant.eu/files/WP6-WP7%20meeting%20Agenda%20v4_0.doc</a>
Management Committee 3	20/07/11	20/07/11	virtual	<a href="http://vbrant.eu/content/man-comm-3">http://vbrant.eu/content/man-comm-3</a>
WP6 workshop	24/10/11	25/10/11	NHM, London	<a href="http://vbrant.eu/sites/vbrant.eu/files/ViBRANT-PWT_meeting_London_24_October_2011.ppt">http://vbrant.eu/sites/vbrant.eu/files/ViBRANT-PWT_meeting_London_24_October_2011.ppt</a>
Management Committee 4	03/11/11	04/11/11	NHM, London	<a href="http://vbrant.eu/content/man-comm-4">http://vbrant.eu/content/man-comm-4</a>
WP2/WP5 workshop	07/11/11	07/11/11	NHM, London	

### Publications

- Arvanitidis, C., Faulwetter, S., Chatzigeorgiou, G., Penev, L., Bánki, O., Dailianis, T., Pafilis, E., Kouratoras, M., Chatzinikolaou, E., et al. (2011). Engaging the broader community in biodiversity research: the concept of the COMBER pilot project for divers in ViBRANT. *ZooKeys* **150**, 211-229.
- Bachman, S., Moat, J., Hill, A., de la Torre, J. & Scott, B. (2011). Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool. *ZooKeys* **150**, 117-126.
- Baker, E., Johnson, K. & Young, J. R. (2011). The future of the past in the present: biodiversity informatics and geological time. *ZooKeys* **150**, 397-405.

- Baker, E. & Michel, E. (2011). Data standards, sense and stability: Scratchpads, the ICZN and ZooBank. *ZooKeys* **150**, 167-176.
- Berendsohn, W., Güntsch, A., Hoffmann, N., Kohlbecker, A., Luther, K. & Müller, A. (2011). Biodiversity information platforms: From standards to interoperability. *ZooKeys* **150**, 71-87.
- Brake, I., Duin, D., Van de Velde, I., Smith, V. & Rycroft, S. (2011). Who learns from whom? Supporting users and developers of a major biodiversity e-infrastructure. *ZooKeys* **150**, 177-192.
- Duin, D. & van den Besselaar, P. (2011). Studying the effects of virtual biodiversity research infrastructures. *ZooKeys* **150**, 193-210.
- Erwin, T., Stoev, P., Georgiev, T. & Penev, L. (2011). ZooKeys 150: Three and a half years of innovative publishing and growth. *ZooKeys* **150**, 5-14.
- Faulwetter, S., Chatzigeorgiou, G., Galil, B. S. & Arvanitidis, C. (2011a). An account of the taxonomy and distribution of Syllidae (Annelida, Polychaetes) in the eastern Mediterranean, with notes on the genus Prosphaerosyllis San Martín, 1984 in the Mediterranean. *ZooKeys* **150**, 281-326.
- Faulwetter, S., Chatzigeorgiou, G., Galil, B. S., Nicolaidou, A. & Arvanitidis, C. (2011b). Sphaerosyllis levantina sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in Sphaerosyllis hystrix Claparède, 1863. *ZooKeys* **150**, 327-345.
- Hagedorn, G., Mietchen, D., Morris, R., Agosti, D., Penev, L., Berendsohn, W. & Hobern, D. (2011). Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. *ZooKeys* **150**, 127-149.
- Kerner, A., Debrenne, F. & Vignes-Lebbe, R. (2011). Cambrian archaeocyathan metazoans: revision of morphological characters and standardization of genus descriptions to establish an online identification tool. *ZooKeys* **150**, 381-395.
- King, D., Morse, D., Willis, A. & Dil, A. (2011). Towards the bibliography of life. *ZooKeys* **150**, 151-166.
- Lambkin, C. & Bartlett, J. (2011). Bush Blitz aids description of three new species and a new genus of Australian beeflies (Diptera, Bombyliidae, Exoprosopini). *ZooKeys* **150**, 231-280.
- Narwade, S., Kalra, M., Jagdish, R., Varier, D., Satpute, S., Khan, N., Talukdar, G., Mathur, V., Vasudevan, K., et al. (2011). Literature based species occurrence data of birds of northeast India. *ZooKeys* **150**, 407-417.
- Penev, L., Lyal, C., Weitzman, A., Morse, D., King, D., Sautter, G., Georgiev, T., Morris, R., Catapano, T. & Agosti, D. (2011). XML schemas and mark-up practices of taxonomic literature. *ZooKeys* **150**, 89-116.
- Prous, M., Heidemaa, M., Shinohara, A. & Soon, V. (2011). Review of the sawfly genus Empria (Hymenoptera, Tenthredinidae) in Japan. *ZooKeys* **150**, 347-380.
- Smith, V. & Penev, L. (2011). Collaborative electronic infrastructures to accelerate taxonomic research. *ZooKeys* **150**, 1-3.
- Smith, V., Rycroft, S., Brake, I., Scott, B., Baker, E., Livermore, L., Blagoderov, V. & Roberts, D. (2011). Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science. *ZooKeys* **150**, 53-70.
- Thessen, A. & Patterson, D. (2011). Data issues in the life sciences. *ZooKeys* **150**, 15-51.

## Dissemination activities

Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
F	RBINS	BHL-Europe Content Provider and Technical Review Meeting	BHL-Europe meeting	29/11/2011 - 02/12/2011	RBINS, Brussels	Scientific community	40	Europe	No	
W	VU	Virtual Research Environments: Catalysts of Change	Knowledge Exchange Workshop "Virtual Research Environments: Catalysts of Change"	17/11/2011 - 18/11/2011	Birmingham	Scientific communityPolicy makers	40	Europe, US	No	
W	HCMR		LifeWatch stakeholders meeting, Brussels	15/11/2011 - 16/11/2011	Brussels, Belgium	Policy makers	20	EU	No	
W	HCMR		BioDivMex (MISTRALS) Workshop	08/11/2011 - 09/11/2011	Marseille, France	Scientific communityMedia	50	EU, Mediterranean	No	<a href="http://www.marbigen.org/content/contribution-biodivmex-W">http://www.marbigen.org/content/contribution-biodivmex-W</a>
C	PENSO FT	Sherborn Meeting	Streamlining the registration-to-publication pipeline	28/10/2011	Natural History Museum, London, UK	Scientific community	80	international	Yes	
W	VU	Social Network Analysis and Organizations	Social networks in biodiversity research	27/10/2011	Amsterdam	Scientific communityIndustryPolicy makers	40	Netherlands, Germany, UK, France, Portugal, Greece, Romania, US	No	
W	HCMR		EMODnet Workshop	25/10/2011 - 27/10/2011	Heraklion, Crete, Greece	Scientific community	20	Global	No	
W	HCMR		EuroMarine data management meeting	18/10/2011	Athens, Greece	Scientific community	20	EU	No	
P	NHM	TDWG 2011	ViBRANT: linking communities and services	17/10/2011 - 21/10/2011	New Orleans, USA	Scientific community	250	Global	Yes	<a href="http://vbrant.eu/sites/vbrant.eu/files/Roberts_ViBRANT.pdf">http://vbrant.eu/sites/vbrant.eu/files/Roberts_ViBRANT.pdf</a>



Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
P	NHM	TDWG 2011	Thoughts on addressing data citation challenges: experiences of Vibrant project	17/10/2011 - 21/10/2011	New Orleans, USA	Scientific community	250	Global	Yes	<a href="http://vbrant.eu/sites/vbrant.eu/files/Roberts_DataPublication.pdf">http://vbrant.eu/sites/vbrant.eu/files/Roberts_DataPublication.pdf</a>
Pr	UPMC	TDWG 2011	A new open source identification key generation webservice	17/10/2011 - 21/10/2011	New Orleans, USA	Scientific community	250	Global	No	<a href="http://vbrant.eu/sites/vbrant.eu/files/XPER2posterTDWG181011.pdf">http://vbrant.eu/sites/vbrant.eu/files/XPER2posterTDWG181011.pdf</a>
P	Vizz	TDWG 2011	Supporting red list threat assessments with GeoCAT: Geospatial Conservation Assessment Tool	17/10/2011 - 21/10/2011	New Orleans, USA	Scientific community	250	Global	Yes	<a href="http://www.tdwg.org/fileadmin/2011C/slides/Javier_de_la_Torre_GeoCAT.pdf">http://www.tdwg.org/fileadmin/2011C/slides/Javier_de_la_Torre_GeoCAT.pdf</a>
P	PENSO FT	TDWG 2011	Data Citation from the perspective of Scholarly Publisher	17/10/2011 - 21/10/2011	New Orleans, USA	Scientific community	250	Global	Yes	
P	PENSO FT	TDWG 2011	Data publishing from the viewpoint of a biodiversity publisher	17/10/2011 - 21/10/2011	New Orleans, USA	Scientific community	250	Global	Yes	
P	UPMC	TDWG 2011	Xper2: new features for an improved flexibility	17/10/2011 - 21/10/2011	New Orleans, USA	Scientific community	250	Global	Yes	<a href="http://www.tdwg.org/fileadmin/2011C/slides/Ung_Xper2.ppt">http://www.tdwg.org/fileadmin/2011C/slides/Ung_Xper2.ppt</a>
W	Vizz	TDWG 2011	Easy mapping with CartoDB: how to make beautiful maps fast and easy	17/10/2011 - 21/10/2011	New Orleans, USA	Scientific community	250	global	Yes	
W	PENSO FT	TDWG Data Citation W	Data citation from the perspective of a scholarly publisher	16/10/2011 - 21/11/2011	New Orleans, USA	Scientific communityIn dustry	120	internati onal	Yes	
W	PENSO FT	TDWG Annual Meeting	Data publishing from the viewpoint of a biodiversity publisher	16/10/2011 - 21/10/2011	New Orleans, USA	Scientific communityIn dustry	120	internati onal	Yes	

Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
W	HCMR		EMBRC (European Marine Biological Resource Centre) Workshop	13/10/2011 - 14/10/2011	Plymouth, England, UK	Scientific communityPolicy makers	50	EU	No	
P	KIT	Third International C on Social Informatics (SocInfo'11)	High-Throughput Crowdsourcing Mechanisms for Complex Tasks	06/10/2011 - 08/10/2011	Singapore	Scientific community	100	International	Yes	
F	UNITS	Green Ideas 2011	ViBRANT	05/10/2011 - 07/10/2011	Mesta, Chias, Greece	Scientific communityIndustryPolicy makers	50	Worldwide	No	
E	UNITS	Annual Meeting of the Italian Lichen Society	Scratchpads as an instrument for the joint creation of a lichen flora of Italy	28/09/2011 - 30/09/2011	Terni, Umbria, Italy	Scientific community	100	Italy	Yes	
W	HCMR		World C on Marine Biodiversity (WCMB)	27/09/2011 - 30/09/2011	Aberdeen, Scotland, UK	Scientific communityIndustryPolicy makersMedia	1000	Globe	No	<a href="http://www.marbigen.org/content/world-C-marine-biodiversity">http://www.marbigen.org/content/world-C-marine-biodiversity</a>
W	HCMR		MAPMED Kick-off meeting Cagliari	19/09/2011 - 20/09/2011	Cagliari, Italy	Scientific communityIndustryCivil societyPolicy makers	30	EU, Mediterranean North African	No	
E	NHM		ViBRANT @ Science Uncovered	19/09/11	London	Civil Society	3000	UK	No	
P	Vizz	FOSS4G 2011	HTML5 for Rich Geospatial applications on the web	16/09/2011	Denver	Industry	1000	International	Yes	<a href="http://xavijam.github.com/html5-foss4g/">http://xavijam.github.com/html5-foss4g/</a>
P	VU	Atlanta C on Science and Innovation Policy 2011	Working Together on the Web, Working Well? Innovation of a Research Work Environment	15/09/2011 - 17/09/2011	Atlanta, GA USA	Scientific communityPolicy makers	300	International	Yes	
F	RBINS		CETAF 30th General Meeting	13/09/2011 - 14/09/2011	RJB Madrid	Scientific community	30	Europe	No	
C	VU		European Network of Indicator Designers / STI Indicators C Series 2011	07/09/2011 - 09/09/2011	Rome	Scientific community	80	Global	No	

Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
W	RBINS	MASTS BEF JRT Workshop	Scratchpads - Basic training course	25/08/11	Edinburgh	Scientific community	12	UK	Yes	
W	VU		FoRC: Future of Research Communication	15/08/2011 - 18/08/2011	Dagstuhl	Scientific community Policy makers	35	Global	No	
W	RBINS	Convolvulaceae workshop	Scratchpads - Basic training course	15/08/11	NHM London	Scientific community	11	Global	Yes	
C	PENSO FT		Link yourself or perish? ###PhytoKeys, the next generation journal in systematic botany	23/07/2011 - 30/07/2011	Melbourne, Australia	Scientific community	2300	international	Yes	
W	HCMR		EUROMARINE Workshop, Roscoff	11/07/2011 - 13/07/2011	Roscoff, France	Scientific community	50	EU	No	
P	JKI	Open Knowledge C (OKCon)	Collaborative platforms for streamlining workflows in Open Science	30/06/2011 - 01/07/2011	Berlin	Scientific community	50	European	Yes	
P	VU	altmetrics11: Tracking scholarly impact on the social Web	The search for alternative metrics for taxonomy	13/06/2011 - 14/06/2011	Koblenz (Germany)	Scientific community	100	Europe	Yes	
P	NHM	Kew eTaxonomy Day	Scratchpads: past, present and future	26/05/2011	Kew Gardens, London	Scientific community	50	UK	Yes	<a href="http://www.slideshare.net/vsmithuk/scratchpads-past-present-and-future">http://www.slideshare.net/vsmithuk/scratchpads-past-present-and-future</a>
P	BGBM	Biodiversity in the Silicon Age	Sustainable infrastructures for e-taxonomy in Europe	17/05/2011 - 18/05/2011	Copenhagen	Scientific community	100	European	Yes	
W	HCMR		ENSSA: Young Europeans Discuss Sustainable Development 9-14 of May 2011	11/05/2011	Athens, Greece	Civil society	100	EU	No	
Pr	VU	Network Research Symposium	The social design of a science (e)-infrastructure	10/05/2011	Amsterdam	Scientific community	100	National (Netherlands)	No	

Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
P	NHM	Open Access Coordination W	New ways to communicate in science: perspectives from biodiversity research	04/05/2011	Brussels	Scientific communityPolicy makers	100	European	Yes	<a href="http://www.slideshare.net/vsmithuk/new-ways-to-communicate-in-science-perspectives-from-biodiversity-research">http://www.slideshare.net/vsmithuk/new-ways-to-communicate-in-science-perspectives-from-biodiversity-research</a>
P	VU	The VU Network Institute Symposium	The social design of an (e)-infrastructure	03/05/2011	Amsterdam	Scientific community	50	Netherlands	Yes	
W	HCMR		LifeWatch stakeholders meeting, Budapest	10/04/2011 - 11/04/2011	Budapest, Hungary	Scientific communityPolicy makersMedia	30	EU	No	
P	NHM	GBIF European Nodes W	Community web sites: small pieces loosely joined	04/04/2011 - 06/04/2011	Paris	Scientific community	35	European	Yes	<a href="http://vsmith.info/files/2011GBIFNodes050411_ViBRANT.pdf">http://vsmith.info/files/2011GBIFNodes050411_ViBRANT.pdf</a>
W	HCMR		MARCOM+ meeting	04/04/2011 - 06/04/2011	Palma de Majorca, Spain	Scientific communityIndustryPolicy makersMedia	30	EU	No	
W	HCMR		EMBOS kick-off meeting	29/03/2011 - 30/03/2011	Brussels, Belgium	Scientific community	100	EU	No	
W	HCMR		EUROMARINE kick-off meeting, Gothenborg	21/03/2011 - 22/03/2011	Gothenborg, Sweeden	Scientific communityPolicy makers	100	EU	No	
W	HCMR		VECTORS project kick-off meeting	01/03/2011 - 03/03/2011	Olhao, Faro, Portugal	Scientific communityIndustry	150	global	No	
P	NHM	Indo-Pacific Ancient Ecosystems Group, at EU FP7 Marie Curie Initial Training Network "Throughflow" meeting	Scratchpads: an introduction	28/02/2011	London	Scientific community	15	International	Yes	<a href="http://www.slideshare.net/vsmithuk/scratchpad-training">http://www.slideshare.net/vsmithuk/scratchpad-training</a>
W	RBINS	Throughflow workshop	Scratchpads - Basic training course	28/02/11	NHM London	Scientific community	15	Global	Yes	

Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
P	NHM	BioSystematics Berlin 2011	Scratchpads - Social websites to mobilise taxonomic data	21/02/2011 - 27/02/2011	Berlin	Scientific community	100	International	Yes	<a href="http://scratchpads.eu/sites/scratchpads.eu/files/Brake_Scratchpads.ppt">http://scratchpads.eu/sites/scratchpads.eu/files/Brake_Scratchpads.ppt</a>
P	JKI	BioSystematics Berlin 2011	Open Nature Guides ("Offene Naturführer") – an open access publication and collaboration site for biodiversity education	21/02/2011 - 27/02/2011	Berlin	Scientific community	100	European	Yes	
W	RBINS		Scratchpads - Basic training course	16/02/11	NHM London	Scientific community	10	Global	Yes	
P	HCMR	VeSTIS Project meeting	Recent advances and challenges in the field of biodiversity data and data observatories in the EU: The Research Infrastructures LifeWatch and ViBRANT	01/02/2011	University of Cyprus, Nicosia	Scientific community	20	Greece, Italy	Yes	
C	PENSO FT	South African Society for Systematic Biology 9th Congress "Biodiversity Matters", Digital Taxonomy W	Cybertaxonomy, Semantic Markup and Semantic Enhancements: The PhytoKeys and ZooKeys Experience	19/01/2011 - 21/01/2011	Grahamstown, South Africa	Scientific community	70	South Africa	Yes	
W	RBINS	EDIT final meeting	Scratchpads - Basic training course	19/01/11	MNHN, Paris	Scientific community	12	EU	Yes	
C	PENSO FT	Academic Publishing in Europe	To be Published for free or to be Read for free: OA publishing from an Eastern European perspective	10/01/2011 - 13/01/2011	Berlin	Scientific communityIndustry	100	Europe	Yes	
W	RBINS		Scratchpads - Basic training course	09/12/10	NHM London	Scientific community	11	UK	Yes	

Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
P	NHM	8th e-Infrastructure Concertation Meeting	ViBRANT Overview	04/11/2010 - 05/11/2010	CERN - Geneva	Scientific community Policy makers	100	Europe	Yes	<a href="http://www.slideshare.net/vsmithuk/vibrant-8th-econcertation-meeting">http://www.slideshare.net/vsmithuk/vibrant-8th-econcertation-meeting</a>
P	NHM	Fourth Metadata and Semantics Research C (MTSR 2010)	Community web sites: small pieces loosely joined	20/10/2010 - 22/10/2010	Acala de Henares, Spain	Scientific community	150	Europe	Yes	<a href="http://www.slideshare.net/vsmithuk/community-web-sites-small-pieces-loosely-joined">http://www.slideshare.net/vsmithuk/community-web-sites-small-pieces-loosely-joined</a>
P	NHM	Biolentify2010	ViBRANT—Virtual Biodiversity Research and Access Network for Taxonomy	20/09/2010 - 22/09/2010	Paris	Scientific community	200	Europe	Yes	<a href="http://www.slideshare.net/vsmithuk/biolentify2010-vibrant">http://www.slideshare.net/vsmithuk/biolentify2010-vibrant</a>

C - Conference  
E- Exhibitions  
F - Flyers  
P - Presentations  
Pr - Posters  
W - Workshop

### Project planning and status

ViBRANT is functioning as planned and is currently on schedule and on budget.

ViBRANT was planned as an agile project, which means that milestones for year 2 are set towards the end of year 1. See Table 2 for details. Note that the definitive list of milestones is maintained on the project web site, not in Annex 1, which will be updated periodically. This was a project design intention from project conception.

### Impact of possible deviations from the planned milestones and deliverables

No significant deviations.

### Any changes to the legal status of any of the beneficiaries, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SMEs

No changes.



## Development of the Project website

The project web site (<http://vbrant.eu>) serves both for outreach and as a project management tool, being a mixture of public and private (to consortium members) pages. The management functions are to hold the master list of project deliverables and milestones, to run a mail forum and archive, a document repository, a data gathering site (publications and presentations) and a notification site for forthcoming meetings that may be of interest to our community.

The public face of the web site includes dissemination, both presentations and publications (see Sect. 3.3).

The project website also lists the functional sites receiving project funding and sites with project associations.

Project	website	Description
Sites receiving ViBRANT funding		
Scratchpads	<a href="http://scratchpads.eu/">http://scratchpads.eu/</a>	The social media component of ViBRANT. Born in the FP6 EDIT project ( <a href="http://www.e-taxonomy.eu">http://www.e-taxonomy.eu</a> ).
Species-ID	<a href="http://species-id.net/wiki/">http://species-id.net/wiki/</a>	Species-ID offers a platform for text, media, and data facilitating the identification of species on Earth. The project is supported by ViBRANT and presently in its initial phase.
biowikifarm	<a href="http://biowikifarm.net/">http://biowikifarm.net/</a>	The wiki component of ViBRANT, supporting open biodiversity content. Similar to the DRUPAL environments (Scratchpads), the ViBRANT project offers a comprehensive Mediawiki environment, the "biowikifarm". Participation in existing Wikis is welcome, but it is also possible to create a new wiki with a special goal. The benefits of using the biowikifarm over installing a wiki on any other server or homepage plan are the shared administration and the long-term availability and maintenance plan. Wikis on the biowikifarm can be maintained over significantly longer time periods. Custom extensions for biodiversity are developed or already available.
OBOE	<a href="https://oboe.oerc.ox.ac.uk">https://oboe.oerc.ox.ac.uk</a>	A front-end for non-scratchpad users to the database and scheduling services being developed as part of WP5.
Identification key webservice	<a href="http://www.identificationkey.fr">http://www.identificationkey.fr</a>	Identification keys are widely used by scientists to identify taxa. This new identification key generation WebService will be able to generate single-access keys on demand, for single users or research institutions. It will receive user input data (using the standard SDD format), accept several parameters for the key generation (impacting the key topology and representation), and will support several output formats. Furthermore, key generation automation will be possible thanks to the WebService architecture. This WebService will be integrated in the Scratchpads biodiversity networking tool, with an embedded client component. It will also be possible for anyone to develop his own client component in order to call the WebService directly. The whole WebService and its source code will be freely available, thus allowing large institutions to deploy the WebService on their own network and adapt it to their specific needs.
COMBER	<a href="http://www.comber.hcmr.gr/">http://www.comber.hcmr.gr/</a>	A citizen science portal for recreational divers and snorkelers to record observations of fish species, developed by WP8
GeoCAT	<a href="http://geocat.kew.org/">http://geocat.kew.org/</a>	GeoCAT is an open source, browser based tool that performs rapid geospatial analysis to ease the process of Red Listing taxa. Developed to utilise spatially referenced primary occurrence data, the analysis focuses on two aspects of the geographic range of a taxon: the extent of occurrence (EOO) and the area of occupancy (AOO).

Projects related to or using ViBRANT products		
agINFRA	<a href="http://www.aginfra-project.eu/">http://www.aginfra-project.eu/</a>	An FP7 project that kicked off in November 11. Will use Scratchpads as a social media tool in the agricultural domain (a development of agriDrupal). One funded post will work on data extraction at the OU with the WP7 team.
BioVeL	<a href="http://www.cs.cf.ac.uk/biovel/">http://www.cs.cf.ac.uk/biovel/</a>	BioVeL (Biodiversity Virtual e-Laboratory) will meet the needs of Europe's Biodiversity Science research community with tools for pipelining data and analysis into efficient workflows, urgently needed to understand biodiversity in a rapidly changing environment. ViBRANT provided a letter of support. The project started 1st Sept. 2011 and relates to ViBRANT through WP5 and WP4 (BGBM is a member of the BioVeL consortium).
EMBOS	<a href="http://www.cost.esf.org/domain_s_actions/essem/Actions/ES1003">http://www.cost.esf.org/domain_s_actions/essem/Actions/ES1003</a>	Development and implementation of a pan-European Marine Biodiversity Observatory System (EMBOS), is an EU project funded by the COST instrument.
eMonocot	<a href="http://e-monocot.org/">http://e-monocot.org/</a>	A UK-NERC funded project to create a global online resource for monocot plants. It will use some Scratchpad technology and Vince Smith (Project coordinator) is a PI on the project.
EoL	<a href="http://www.eol.org/">http://www.eol.org/</a>	Collaboration between Scratchpads and EoL's LifeDesks, moving to a common code base. Data exchange functions between individual Scratchpads and EoL
JEMU & FWO Research community 'Belgian Network for DNA Barcoding'	<a href="http://jemu.myspecies.info">http://jemu.myspecies.info</a>	Uses Scratchpads as website for its integrated research infrastructure JEMU (Joint Experimental Molecular Unit), for BeBOL, the Belgian Network for DNA barcoding and for the announcement and logistics of the congress ECBOL3. Project coordinator is Thierry Backeljau, WP leader in ViBRANT WP3.
LifeWatch	<a href="http://www.lifewatch.eu/">http://www.lifewatch.eu/</a>	An ESFRI project in the implementation phase in several EU countries. Scratchpads will provide a prototype service desk (user interface). GBIF will provide data. LifeWatch will provide analytical services. Wouter Los, lead of ViBRANT WP4 has been the coordinator of the LifeWatch preparatory projects. W. Berendsohn has led the LifeWatch WP on the LifeWatch implementation plan.
OpenUp!	<a href="http://open-up.eu/">http://open-up.eu/</a>	Opening up the Natural History Heritage for Europeana. Uses a Scratchpad for its home page. Project coordinator is Walter Berendsohn, FU-BGBM, partner in ViBRANT WP4 & WP2.
VECTORS	<a href="http://www.marine-vectors.eu/">http://www.marine-vectors.eu/</a>	Vectors of Change in Oceans and Seas Marine Life, Impact on Economic Sectors (VECTORS) is a European Commission Seventh Framework Programme (FP7) project.

Del.no.	Deliverable name	Version	WP no.	Lead beneficiary	Nature	Dissemination level	Delivery date from Annex I (proj month)	Actual / Forecast delivery date	Status No submitted/ Submitted	Contractual Yes/No	Comments
D1.1	Overall Management		1	NHM		PU	36	30/11/2013	No submitted		
D2.1	Distributing servers	Final	2	NHM	Report	PU	12	23/11/2011	Submitted		<a href="http://vbrant.eu/sites/vbrant.eu/files/d2.1.pdf">http://vbrant.eu/sites/vbrant.eu/files/d2.1.pdf</a> The Scratchpad Mirror server is: <a href="http://e090.bgbm.fu-berlin.de/">http://e090.bgbm.fu-berlin.de/</a> An example mirrored site is: <a href="http://mastigoteuthidae.myspecies.info/">http://mastigoteuthidae.myspecies.info/</a> => <a href="http://mastigoteuthidae.m1.myspecies.info/">http://mastigoteuthidae.m1.myspecies.info/</a>
D2.2	Scratchpad Registry		2	NHM		PU	18	31/05/2012	No submitted		
D2.3	Financial sustainability		2	NHM		PU	30	31/05/2013	No submitted		
D2.4	Unit testing		2	NHM		PU	36	30/11/2013	No submitted		
D3.1	Training strategy	Final	3	RBINS	Report	PU	12	29/11/2011	Submitted		<a href="http://vbrant.eu/sites/vbrant.eu/files/D3.1-Trainingstrategy.pdf">http://vbrant.eu/sites/vbrant.eu/files/D3.1-Trainingstrategy.pdf</a>
D3.2	Service delivery and evaluation		3	VU		PU	24	30/11/2012	No submitted		
D3.3	Community delivery and evaluation		3	RBINS		PU	36	30/11/2013	No submitted		
D4.1	Scratchpad common access point	Final	4	UvA	Report	PU	12	08/12/2011	Submitted		<a href="http://vbrant.eu/sites/vbrant.eu/files/VIBRANT-D4.1.doc">http://vbrant.eu/sites/vbrant.eu/files/VIBRANT-D4.1.doc</a>
D4.2	Ontology tools	Final	4	UvA	Report	PU	25	22/12/2011	Submitted		<a href="http://vbrant.eu/sites/vbrant.eu/files/D4.2_ProgressReport301111.pdf">http://vbrant.eu/sites/vbrant.eu/files/D4.2_ProgressReport301111.pdf</a>

D5.1	Prototype workflows and API	Final	5	UOXF.E9	Report	PU	12	30/11/2011	Submitted		<a href="http://vbrant.eu/content/d51-prototype-workflows-and-api">http://vbrant.eu/content/d51-prototype-workflows-and-api</a> <a href="https://obo.e.oerc.ox.ac.uk/docs">https://obo.e.oerc.ox.ac.uk/docs</a>
D5.2	Functional tools		5	UOXF.E9		PU	24	30/11/2012	No submitted		
D5.3	Sustainable software services		5	UOXF.E9		PU	36	30/11/2013	No submitted		
D6.1	XML mark up tool & service	Final	6	PENSOFT	Report	PU	12	11/11/2011	Submitted		<a href="http://vbrant.eu/sites/vbrant.eu/files/D6-1-PENSOFT-XML%20mark%20up%20to...">http://vbrant.eu/sites/vbrant.eu/files/D6-1-PENSOFT-XML%20mark%20up%20to...</a>
D6.2	Review, refine & evaluate services		6	PENSOFT		PU	18	31/05/2012	No submitted		
D6.3	Data publication workflow		6	PENSOFT		PU	36	30/11/2013	No submitted		
D7.1	Community contributed bibliography	Final	7	OU	Report	PU	12	02/12/2011	Submitted		Prototype Service: <a href="http://plazi2.cs.umb.edu:8080/RefBank/search">http://plazi2.cs.umb.edu:8080/RefBank/search</a> Description: <a href="http://vbrant.eu/sites/vbrant.eu/files/D7-1_report_final.pdf">http://vbrant.eu/sites/vbrant.eu/files/D7-1_report_final.pdf</a>
D7.2	Mark-up modules		7	KIT		PU	24	30/11/2012	No submitted		
D7.3	Literature search		7	OU		PU	35	31/10/2013	No submitted		
D8.1	Scratchpad modules engaging citizen scientists	Final	8	HCMR	Web site	PU	12	30/11/2011	Submitted		<a href="http://www.comber.hcmr.gr/">http://www.comber.hcmr.gr/</a> <a href="http://vbrant.eu/sites/vbrant.eu/files/Deliverable8.pdf">http://vbrant.eu/sites/vbrant.eu/files/Deliverable8.pdf</a>

**Table 2: Milestones**

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M1.10	Confirmation of year 1 meeting dates	WP1	NHM	31/12/2010	Yes	31/12/2010	<a href="http://vbrant.eu/content/management-committee">http://vbrant.eu/content/management-committee</a>
M1.11	Circulation of next meeting agenda	WP1	NHM	31/12/2010	Yes	31/12/2010	<a href="http://vbrant.eu/content/management-committee-meeting">http://vbrant.eu/content/management-committee-meeting</a>
M1.12	Organise kick-off meeting	WP1	NHM	31/12/2010	Yes	31/12/2010	<a href="http://vbrant.eu/scripting-life">http://vbrant.eu/scripting-life</a>
M1.13	Minutes of management committee meetings	WP1	NHM	28/02/2011	Yes	28/02/2011	<a href="http://vbrant.eu/content/management-committee-meeting">http://vbrant.eu/content/management-committee-meeting</a>
M1.14	Minutes of strategic board meetings	WP1	NHM	28/02/2011	Yes	28/02/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_StrategicBoardMinutes_Jan11.pdf">http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_StrategicBoardMinutes_Jan11.pdf</a>
M1.15	Circulation of next meeting agenda	WP1	NHM	30/04/2011	Yes	30/04/2011	<a href="http://vbrant.eu/content/man-comm-2">http://vbrant.eu/content/man-comm-2</a>
M1.16	Minutes of management committee meetings	WP1	NHM	30/06/2011	Yes	30/06/2011	<a href="http://vbrant.eu/content/man-comm-2">http://vbrant.eu/content/man-comm-2</a>
M1.17	Minutes of strategic board meetings	WP1	NHM	30/06/2011	Yes	30/06/2011	<a href="http://vbrant.eu/content/strategic-board-2">http://vbrant.eu/content/strategic-board-2</a>
M1.18	Circulation of next meeting agenda	WP1	NHM	30/09/2011	Yes	30/09/2011	<a href="http://vbrant.eu/content/man-comm-4">http://vbrant.eu/content/man-comm-4</a>
M1.19	Preparation of first JPA templates	WP1	NHM	31/08/2011	Yes	31/08/2011	<a href="http://vbrant.eu/content/m119-preparation-first-jpa-templates">http://vbrant.eu/content/m119-preparation-first-jpa-templates</a>
M1.20	Minutes of management committee meetings	WP1	NHM	18/11/2011	Yes	18/11/2011	<a href="http://vbrant.eu/content/man-comm-4">http://vbrant.eu/content/man-comm-4</a>

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M1.21	Confirmation of year 2 and 3 meeting dates	WP1	NHM	30/11/2011	Yes	30/11/2011	<a href="http://vbrant.eu/content/man-comm-3">http://vbrant.eu/content/man-comm-3</a>
M1.22	Circulation of next meeting agenda	WP1	NHM	31/12/2011	No		
M1.23	First year financial reports	WP1	NHM	31/12/2011	No		
M1.24	First year annual progress report	WP1	NHM	31/01/2012	No		
M1.25	Minutes of management committee meetings	WP1	NHM	29/02/2012	No		
M1.26	Minutes of strategic board meetings	WP1	NHM	29/02/2012	No		
M1.27	Organise mid-project meeting	WP1	NHM	31/03/2012	No		
M1.28	Circulation of next meeting agenda	WP1	NHM	30/04/2012	No		
M1.29	Minutes of management committee meetings	WP1	NHM	30/06/2012	No		
M1.30	Minutes of strategic board meetings	WP1	NHM	30/06/2012	No		
M1.31	Circulation of next meeting agenda	WP1	NHM	31/08/2012	No		
M1.32	Preparation of second JPA templates	WP1	NHM	31/08/2012	No		
M1.33	Minutes of management committee meetings	WP1	NHM	31/10/2012	No		
M1.34	Circulation of next meeting agenda	WP1	NHM	31/12/2012	No		



Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M1.35	Second year financial reports	WP1	NHM	31/12/2012	No		
M1.36	Second year annual progress report	WP1	NHM	31/01/2013	No		
M1.37	Minutes of management committee meetings	WP1	NHM	28/02/2013	No		
M1.38	Minutes of strategic board meetings	WP1	NHM	28/02/2013	No		
M1.39	Organise final meeting	WP1	NHM	31/03/2013	No		
M1.40	Circulation of next meeting agenda	WP1	NHM	30/04/2013	No		
M1.41	Minutes of management committee meetings	WP1	NHM	30/06/2013	No		
M1.42	Minutes of strategic board meetings	WP1	NHM	30/06/2013	No		
M1.43	Circulation of next meeting agenda	WP1	NHM	31/08/2013	No		
M1.44	Minutes of management committee meetings	WP1	NHM	31/10/2013	No		
M1.45	Third year annual progress report	WP1	NHM	30/11/2013	No		
M1.46	Third year financial reports	WP1	NHM	30/11/2013	No		
M2.10	Test version of distributed Scratchpad server	WP2	NHM	31/05/2011	Yes	01/06/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/m2.10.pdf">http://vbrant.eu/sites/vbrant.eu/files/m2.10.pdf</a>
M2.11	Report on Scratchpad usage statistics options	WP2	NHM	31/05/2011	Yes	21/11/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/m2.11.pdf">http://vbrant.eu/sites/vbrant.eu/files/m2.11.pdf</a>

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M2.12	Ongoing module development supporting prioritised features requested by users through WP3	WP2	NHM	30/11/2011	Yes	21/11/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/m2.12.pdf">http://vbrant.eu/sites/vbrant.eu/files/m2.12.pdf</a>
M2.13	Define further milestones in the light of usage and feedback	WP2	NHM	30/11/2011	Yes	21/11/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/m2.13.pdf">http://vbrant.eu/sites/vbrant.eu/files/m2.13.pdf</a>
M2.14	Report on the data search portal providing a single point of entry to all Scratchpad data completed	WP2	BGBM	31/05/2012	No		
M2.15	Biodiversity data citation metric	WP2	NHM	31/05/2012	No		
M2.16	Localised unit testing mechanism	WP2	NHM	30/11/2012	No		
M2.17	Ongoing module development supporting prioritised features requested by users through WP3	WP2	NHM	30/11/2012	No		
M2.18	Data search portal providing a single point of entry to all Scratchpad data completed	WP2	BGBM	31/05/2013	No		
M2.19	Report on the options for a biodiversity data citation metric for data published through ViBRANT	WP2	NHM	31/05/2013	No		

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M2.20	Ongoing module development supporting prioritised features requested by users through WP3	WP2	NHM	30/11/2013	No		
M3.10	Delivery of a promotional strategy for ViBRANT services	WP3	RBINS	30/09/2011	Yes	27/09/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M.3.10_Delivery%20of%20a%20promotional%20strategy%20for%20ViBRANT%20services_0.pdf">http://vbrant.eu/sites/vbrant.eu/files/M.3.10_Delivery%20of%20a%20promotional%20strategy%20for%20ViBRANT%20services_0.pdf</a>
M3.11	Potential Mechanisms to increase user uptake	WP3	VU	13/01/2012	Yes	13/01/2012	<a href="http://vbrant.eu/sites/vbrant.eu/files/Milestone_311_VU_130112.pdf">http://vbrant.eu/sites/vbrant.eu/files/Milestone_311_VU_130112.pdf</a>
M3.12	Assessment of user support services, with the refinement of subsequent milestones	WP3	RBINS	29/11/2011	Yes	29/11/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M3.12-Assessment">http://vbrant.eu/sites/vbrant.eu/files/M3.12-Assessment</a>
of user support services.p df	29/11/2011						
M3.13	Recommendations to modify products	WP3	VU	30/11/2012	No		
M3.14	Assessment of user support services and promotional activities	WP3	RBINS	30/11/2012	No		
M4.10	Define methods to use the Scratchpad service layer deliverable within the CDM store	WP4	GBIF	31/05/2011	Yes	25/07/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/ViBRANT%20M410%20Specification%20...">http://vbrant.eu/sites/vbrant.eu/files/ViBRANT%20M410%20Specification%20...</a>

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M4.11	Prototype of collaborative community interface	WP4	JKI	31/05/2011	Yes	31/05/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M4.11%E2%80%9494Prototype%20of%20co...">http://vbrant.eu/sites/vbrant.eu/files/M4.11%E2%80%9494Prototype%20of%20co...</a>
M4.12	Liaison and networking with ontology experts and existing ontology providers	WP4	GBIF	31/07/2011	Yes	02/12/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/Smith-Submitted-v2Revised_ed_fin...">http://vbrant.eu/sites/vbrant.eu/files/Smith-Submitted-v2Revised_ed_fin...</a>
<a href="http://vbrant.eu/sites/vbrant.eu/files/D4.2_Progress_Report30111.pdf">http://vbrant.eu/sites/vbrant.eu/files/D4.2_Progress_Report30111.pdf</a>	31/07/2011						
M4.13	Release an API on the catalogue of resources	WP4	JKI	31/08/2011	Yes	05/10/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M4.13%20-%20Release%20an%20API%20...">http://vbrant.eu/sites/vbrant.eu/files/M4.13%20-%20Release%20an%20API%20...</a>
M4.14	Define further milestones in the light of usage and feedback	WP4	UvA	30/11/2011	Yes	24/12/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/WP4_milestones_2.xls">http://vbrant.eu/sites/vbrant.eu/files/WP4_milestones_2.xls</a>
M4.15	Ontology Tools: migrate GBIF Vocabularies Service	WP4	UvA	31/12/2011	No		
M4.16	Ontology Tools: scoping document on a KOS architecture	WP4	UvA	31/12/2011	No		
M4.17	Extensions for DwC-A export functionality	WP4	UvA	28/02/2012	No		
M4.18	Human Interface for CDM-ViBRANT Index	WP4	UvA	28/03/2012	No		

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M4.19	Canvassing glossary-oriented potential contributors	WP4	UvA	31/03/2012	No		
M4.20	Ontology Tools: new prototype (GBIF) Glossary of Terms registry	WP4	UvA	30/04/2012	No		
M4.21	Ontology Tools: new prototype Biodiversity BioPortal	WP4	UvA	30/06/2012	No		
M4.22	BioFlor Data as semantic web	WP4	UvA	30/06/2012	No		
M4.23	Scratchpad to DwC-A Mapping Module	WP4	UvA	30/09/2012	No		
M4.24	Full text search CDM-ViBRANT Index	WP4	UvA	30/09/2012	No		
M4.25	Pilot implementation of webservice at BGBM	WP4	UvA	30/09/2012	No		
M4.26	Practical application milestone for media wiki	WP4	UvA	30/09/2012	No		
M4.27	Audobon Core standard on Mediawiki	WP4	UvA	30/09/2012	No		
M4.28	Statistical output provided for CDM-ViBRANT Index query Interface	WP4	UvA	30/11/2012	No		
M4.29	CDM SDD output optimised for Xper2	WP4	UvA	30/11/2012	No		
M4.30	Integrating the biowikifarm into ViBRANT authentication	WP4	UvA	30/11/2012	No		

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M4.31	Webservice-Layer adapted for ViBRANT	WP4	UvA	31/01/2013	No		
M4.32	Integrating a wiki-glossary into the scratchpad	WP4	UvA	31/01/2013	No		
M4.33	XML Transformations CDM export	WP4	UvA	30/04/2013	No		
M4.34	Prototype for taxonomic generalisation	WP4	UvA	31/05/2013	No		
M4.35	Filter functionality CDM export	WP4	UvA	31/05/2013	No		
M4.36	XPer2-CDM-relation for taxonomic hierarchy specification	WP4	UvA	30/06/2013	No		
M4.37	Xper2 and EDIT Platform publishing capability on Mediawiki	WP4	UvA	31/05/2013	No		
M5.10	API design plan	WP5	UOXF.E9	31/12/2010	Yes	31/12/2010	<a href="http://vbrant.eu/content/ms-510-api-design-plan">http://vbrant.eu/content/ms-510-api-design-plan</a>
M5.11	Review existing key construction software and workflow interaction with both Scratchpads and the CDM	WP5	UPMC	28/02/2011	Yes	28/02/2011	<a href="http://www.infosylab.fr/vibrant/M5_11_deliverable_UPMC_february2011.pdf">http://www.infosylab.fr/vibrant/M5_11_deliverable_UPMC_february2011.pdf</a>
M5.12	Review user requirements for the visualisation tool for Scratchpads	WP5	Vizz	28/02/2011	Yes	25/04/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M5-12.pdf">http://vbrant.eu/sites/vbrant.eu/files/M5-12.pdf</a>
M5.13	API baseline documentation	WP5	UOXF.E9	31/03/2011	Yes	19/04/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M5.13.api_baseline.txt">http://vbrant.eu/sites/vbrant.eu/files/M5.13.api_baseline.txt</a>



Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M5.14	Review of target applications (phylogenetics)	WP5	UOXF.E9	31/03/2011	Yes	31/03/2011	<a href="http://vbrant.eu/content/m514-review-target-applications-phylogenetics">http://vbrant.eu/content/m514-review-target-applications-phylogenetics</a>
M5.15	API version control	WP5	UOXF.E9	30/04/2011	Yes	30/04/2011	<a href="https://oboe.oerc.ox.ac.uk/docs">https://oboe.oerc.ox.ac.uk/docs</a>
M5.16	Review of target platforms and requirements	WP5	UOXF.E9	30/06/2011	Yes	30/06/2011	<a href="http://vbrant.eu/content/m516-review-target-platforms-and-requirements">http://vbrant.eu/content/m516-review-target-platforms-and-requirements</a>
M5.17	Define further milestones in the light of usage and feedback	WP5	UOXF.E9	31/03/2012	No		
M5.18	Review user requirements for enhanced user interface	WP5	UOXF.E9	30/11/2011	Yes	30/11/2011	<a href="http://vbrant.eu/content/m518-review-user-requirements-enhanced-user-interface">http://vbrant.eu/content/m518-review-user-requirements-enhanced-user-interface</a>
M5.19	Deliver prototype key-generating service through Scratchpads	WP5	UPMC	30/11/2011	Yes	22/12/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M5_19_report2.doc">http://vbrant.eu/sites/vbrant.eu/files/M5_19_report2.doc</a> <a href="http://identificationkey.fr">http://identificationkey.fr</a>
M5.20	Metadata repository design plan	WP5	UOXF.E9	31/12/2011	Yes	31/12/2011	<a href="http://vbrant.eu/content/m520-metadata-repository-design-plan">http://vbrant.eu/content/m520-metadata-repository-design-plan</a>
M5.21	Implement visualisation tool for Scratchpads	WP5	Vizz	29/02/2012	No		
M5.22	Review algorithms for biodiversity indices	WP5	Vizz	29/02/2012	No		
M5.23	Implementation of platforms specific middleware	WP5	UOXF.E9	31/03/2012	No		

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M5.24	Implement application specific wrappers (phylogenetics) that can be used by Scratchpads	WP5	UOXF.E9	31/03/2012	No		
M5.25	Implement custom wrapper for the identification service that can be used by Scratchpads	WP5	JKI	31/03/2012	No		
M5.26	Review of target applications (bioclimatic modelling)	WP5	UOXF.E9	30/06/2012	No		
M5.27	Metadata repository implementation	WP5	UOXF.E9	30/09/2012	No		
M5.28	Implement enhanced user interface tools for managing keys in Scratchpads	WP5	JKI	30/11/2012	No		
M5.29	Implement application specific wrappers (bioclimatic modelling)	WP5	UOXF.E9	31/12/2012	No		
M5.30	Implement biodiversity indices tool	WP5	Vizz	28/02/2013	No		
M5.31	Review of target applications and services (molecular identification)	WP5	UOXF.E9	31/03/2013	No		
M5.32	Tests and validation of the services in coordination with Workpackage 3	WP5	UOXF.E9	31/05/2013	No		

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M6.10	Use cases of existing standards of XML mark up tagging and semantic enhancement collected and review	WP6	PENSOFT	28/02/2011	Yes	04/04/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/Milestone%206%2010-Review%20of%20...">http://vbrant.eu/sites/vbrant.eu/files/Milestone %206%2010-Review%20of %20...</a>
M6.11	Prototype and beta-version of XML submission from Scratchpads to publishers	WP6	NHM	31/07/2011	Yes	11/11/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M6.11-Prototype%20and%20beta-vers...">http://vbrant.eu/sites/vbrant.eu/files/M6.11-Prototype %20and%20beta-vers...</a>
M6.12	Workshop on mark up tagging tools and implementations at TDWG	WP6	PENSOFT	31/10/2011	Yes	11/11/2011	The milestone report includes minutes and presentations from two meetings on XML mark up schemas and tagging tools held in Paris (during the kick-off) and in London (24-25 Oct 2011)
M6.12	Workshop on mark up tagging tools and implementations at TDWG	WP6	PENSOFT	31/10/2011	Yes	11/11/2011	The milestone report includes minutes and presentations from two meetings on XML mark up schemas and tagging tools held in Paris (during the kick-off) and in London (24-25 Oct 2011)

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M6.12	Workshop on mark up tagging tools and implementations at TDWG	WP6	PENSOFT	31/10/2011	Yes	11/11/2011	The milestone report includes minutes and presentations from two meetings on XML mark up schemas and tagging tools held in Paris (during the kick-off) and in London (24-25 Oct 2011)
M6.12	Workshop on mark up tagging tools and implementations at TDWG	WP6	PENSOFT	31/10/2011	Yes	11/11/2011	The milestone report includes minutes and presentations from two meetings on XML mark up schemas and tagging tools held in Paris (during the kick-off) and in London (24-25 Oct 2011)
M6.13	Define further milestones in the light of usage and feedback	WP6	PENSOFT	30/11/2011	Yes	11/11/2011	List of two additional milestones planned for the second period if ViBRANT.
M6.14	Review of existing standards of semantic enhancements to taxonomic papers	WP6	PENSOFT	31/12/2011	No		
M6.15	Sample papers testing the XML-based editorial workflow elaborated and published	WP6	PENSOFT	31/05/2012	No		

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M6.16	Refinements to XML-based workflow for peer-review based on user feedback	WP6	NHM	30/11/2012	No		
M6.17	New open access data publishing journal	WP6	PENSOFT	31/03/2013	No		
M6.18	Testing automated tools of dissemination of published results to aggregators and indexing services	WP6	PENSOFT	31/08/2013	No		
M6.19	XML export format of metadata on identification keys to Key Central (ALA)	WP6	PENSOFT	31/01/2012	No		
M6.20	Beta version of XML queries to IPNI and ZooBank	WP6	PENSOFT	31/03/2012	No		
M7.10	Agreement of standard format for community contributed bibliographies in conjunction with WP4	WP7	OU	17/06/2011	Yes	27/07/11	
M7.11	Review of options for interactive mark up tools within the Scratchpad infrastructure	WP7	OU	17/06/2011	Yes	04/08/11	<a href="http://wiki.scratchpads.eu/w/M711report">http://wiki.scratchpads.eu/w/M711report</a>
M7.12	A suite of test cases that will be used to test the de-duplication software	WP7	OU	31/07/2011	Yes	16/12/2011	

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
<a href="http://vbrant.eu/sites/vbrant.eu/files/M7-12_report.pdf">http://vbrant.eu/sites/vbrant.eu/files/M7-12_report.pdf</a>	31/07/2011						
M7.13	Review of options to use typographical information and other contextual clues	WP7	OU	31/07/2011	Yes	20/12/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M7-13_report.pdf">http://vbrant.eu/sites/vbrant.eu/files/M7-13_report.pdf</a>
<a href="http://vbrant.eu/sites/vbrant.eu/files/Report%20-%20Gold%20enGATE%20Modules...">http://vbrant.eu/sites/vbrant.eu/files/Report%20-%20Gold%20enGATE%20Modules...</a>	31/07/2011						
M7.13	Review of options to use typographical information and other contextual clues	WP7	OU	31/07/2011	Yes	20/12/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M7-13_report.pdf">http://vbrant.eu/sites/vbrant.eu/files/M7-13_report.pdf</a>
<a href="http://vbrant.eu/sites/vbrant.eu/files/Report%20-%20Gold%20enGATE%20Modules...">http://vbrant.eu/sites/vbrant.eu/files/Report%20-%20Gold%20enGATE%20Modules...</a>	31/07/2011						
M7.14	Collaboration with WP6 on mark-up tagging tools and implementation	WP7	OU	31/10/2011	Yes	16/12/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/M7-14_report.pdf">http://vbrant.eu/sites/vbrant.eu/files/M7-14_report.pdf</a>
M7.15	Define further milestones in the light of usage and feedback	WP7	OU	29/02/2012	No		



Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M7.16	Mark-up modules delivering outline mark-up	WP7	KIT	31/05/2012	No		
M7.17	Review of pilot mark up processes within the Scratchpad infrastructure	WP7	NHM	31/07/2012	No		
M7.18	First integration phase complete	WP7	OU	31/03/2013	No		
M7.19	Review of pilot of reference de-duplication software	WP7	NHM	31/07/2013	No		
M7.20	Workpackage software documentation produced	WP7	OU	31/10/2013	No		
M8.10	Prototype ecological and conservation applications	WP8	HCMR	31/05/2011	Yes	03/06/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/Milestone_8.10_Indices_C_code_v0....">http://vbrant.eu/sites/vbrant.eu/files/Milestone_8.10_Indices_C_code_v0....</a>
<a href="https://git.scratchpads.eu/view/git?p=indices.git;a=summary">https://git.scratchpads.eu/view/git?p=indices.git;a=summary</a>	31/05/2011						
M8.11	Review of relevant field recording tools and applications	WP8	MfN	25/11/2011	Yes	25/11/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/a%20review%20mobile%20apps_201111...">http://vbrant.eu/sites/vbrant.eu/files/a%20review%20mobile%20apps_201111...</a>
M8.12	Identify rules for an intelligent algorithm to identify suspicious data records in GBIF data	WP8	GBIF	30/11/2011	Yes	30/11/2011	<a href="http://www.slideshare.net/DavidRemsen/tdwg-1remsen">http://www.slideshare.net/DavidRemsen/tdwg-1remsen</a>

Milestone no.	Milestone name	WP no	Lead beneficiary	Delivery date from Annex I	Achieved Yes/No	Actual / Forecast achievement date	Comments
M8.13	Define further milestones in the light of usage and feedback	WP8	HCMR	30/11/2011	Yes	30/11/2011	<a href="http://vbrant.eu/sites/vbrant.eu/files/Milestones%20WP8Overview_FINAL_Ye...">http://vbrant.eu/sites/vbrant.eu/files/Milestones%20WP8Overview_FINAL_Ye...</a>
M8.14	Prototype demonstration project with non-professional marine naturalists	WP8	HCMR	30/11/2011	Yes	30/11/2011	<a href="http://www.pensoft.net/journals/zookeys/article/2149/abstract/engaging-t...">http://www.pensoft.net/journals/zookeys/article/2149/abstract/engaging-t...</a>
M8.15	Report on existing processing services and analytical services	WP8	HCMR	29/02/2012	No		
M8.16	Prototype analytical tools and reporting tools implemented on the visualization tool	WP8	Vizz	31/05/2012	No		
M8.17	APIs to integrate various data sources in support of analytical and reporting tools	WP8	HCMR	31/05/2012	No		

## Appendix 1

The documentation for OBOE available to on-line clients.

### /docs:

Data should be posted as json to various URLs on the site to create or update job requests, and various URLs can be accessed to get information. There's also a web interface. The base URL is <https://vibrant.oerc.ox.ac.uk>.

### GET:

/routes => list all available URLs.

/docs => show this information.

/search.json => show all available jobs.

/jobs.json => show all available jobs.

/download/\$id => download output for a job

/download/infile/\$id => download input file for a job

...more are available in the routes output.

### POST:

Post (and PUT) requests can be made with curl as below for testing purposes, although it is presumed that other applications (e.g. Scratchpads) will use other means (e.g. php) to do so. If posting JSON then it is necessary to also post an auth\_token, as shown below. Users using the web interface can upload files directly, but servers will have to send the link of an http accessible file instead; the system will download this.

N.B. --insecure is only required if there is no proper certificate; this requirement will eventually be removed for the production system.

### Create:

```
curl --insecure -i -X POST -H 'Content-Type: application/json' -H 'Accept: application/json' -d '{"job":
{"inputurl":"http://input/7","type":"test","user":"milo"},"auth_token":"ehioesihp382uah"}'
https://vibrant.oerc.ox.ac.uk:3000/jobs.json
```

### Search:

```
curl --insecure -i -X POST -H 'Content-Type: application/json' -H 'Accept: application/json' -d '{"job":
{"status":"finished"},"auth_token":"ehioesihp382uah"}' https://vibrant.oerc.ox.ac.uk/jobs/search.json
```

### PUT:

#### Update:

```
curl --insecure -i -X PUT -H 'Content-Type: application/json' -H 'Accept: application/json' -d '{"job":
{"user":"new_user"},"auth_token":"ehioesihp382uah"}'
https://vibrant.oerc.ox.ac.uk/jobs/4d27036135d3101c68000001.json
```

Timestamps will be updated automatically when jobs are created or modified.

### /routes:

jobs GET	/jobs(.:format)	{:action=>"index", :controller=>"jobs"}
POST	/jobs(.:format)	{:action=>"create", :controller=>"jobs"}
new_job GET	/jobs/new(.:format)	{:action=>"new", :controller=>"jobs"}
edit_job GET	/jobs/:id/edit(.:format)	{:action=>"edit", :controller=>"jobs"}

job GET	/jobs/:id(.:format)	{:action=>"show", :controller=>"jobs"}
PUT	/jobs/:id(.:format)	{:action=>"update", :controller=>"jobs"}
DELETE	/jobs/:id(.:format)	{:action=>"destroy", :controller=>"jobs"}
new_user_session GET	/users/sign_in(.:format)	{:action=>"new", :controller=>"devise/sessions"}
user_session POST	/users/sign_in(.:format)	{:action=>"create", :controller=>"devise/sessions"}
destroy_user_session GET	/users/sign_out(.:format)	{:action=>"destroy", :controller=>"devise/sessions"}
user_password POST	/users/password(.:format)	{:action=>"create", :controller=>"devise/passwords"}
new_user_password GET	/users/password/new(.:format)	{:action=>"new", :controller=>"devise/passwords"}
edit_user_password GET	/users/password/edit(.:format)	{:action=>"edit", :controller=>"devise/passwords"}
PUT	/users/password(.:format)	{:action=>"update", :controller=>"devise/passwords"}
user_registration POST	/users(.:format)	{:action=>"create", :controller=>"devise/registrations"}
new_user_registration GET	/users/sign_up(.:format)	{:action=>"new", :controller=>"devise/registrations"}
edit_user_registration GET	/users/edit(.:format)	{:action=>"edit", :controller=>"devise/registrations"}
PUT	/users(.:format)	{:action=>"update", :controller=>"devise/registrations"}
DELETE	/users(.:format)	{:action=>"destroy", :controller=>"devise/registrations"}
user_confirmation POST	/users/confirmation(.:format)	{:action=>"create", :controller=>"devise/confirmations"}
new_user_confirmation GET	/users/confirmation/new(.:format)	{:action=>"new", :controller=>"devise/confirmations"}
GET	/users/confirmation(.:format)	{:action=>"show", :controller=>"devise/confirmations"}
user_unlock POST	/users/unlock(.:format)	{:action=>"create", :controller=>"devise/unlocks"}
new_user_unlock GET	/users/unlock/new(.:format)	{:action=>"new", :controller=>"devise/unlocks"}
GET	/users/unlock(.:format)	{:action=>"show", :controller=>"devise/unlocks"}
root	/(:.format)	{:controller=>"info", :action=>"welcome"}
POST	/search(.:format)	{:controller=>"jobs", :action=>"search"}
jobs_new POST	/jobs/new(.:format)	{:controller=>"jobs", :action=>"new"}

GET	/search(..format)	{:controller=>"jobs", :action=>"index"}
GET	/download/:id(..format)	{:controller=>"jobs", :action=>"download"}
GET	/download/infile/:id(..format)	{:controller=>"jobs", :action=>"get_infile"}
show_all GET	/show_all(..format)	{:action=>"show_all", :controller=>"jobs"}
routes GET	/routes(..format)	{:action=>"routes", :controller=>"info"}
docs GET	/docs(..format)	{:action=>"docs", :controller=>"info"}
left GET	/left(..format)	{:action=>"welcome_left", :controller=>"info"}