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:

Periodic report: 1ˢᵗ   2ⁿᵈ   3ʳᵈ   4ᵗʰ

Period covered: from 1 Dec 11   to 30 Nov 12

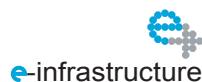**Name, title and organisation of the scientific representative of the project's coordinator:**

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) [3]:

  ☐ 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:

Date:

For most of the projects, the signature of this declaration could be done directly via the IT reporting tool through an adapted IT mechanism and in that case, no signed paper form needs to be sent

---

[3] 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. As part of this broad effort ViBRANT and BioVeL (EU Project 283359) have coordinated a community consultation to establish a collective view of priorities in biodiversity informatics over the coming decade. That exercise resulted in a White Paper that has been submitted for publication to *BMC Ecology*[1].

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 during the first half of the project year as far as users are concerned.

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 (see below) which will include making identification keys, calculating ecological impact factors and phylogenetic relationships. Another tool that has been launched this year is GeoCAT (http://geocat.kew.org) that performs rapid geospatial analysis to facilitate the process of Red Listing taxa (the international assessment of a species' risk of extinction). Other services will follow in the coming years.

A major part of the ViBRANT project effort is dedicated to training and support services to the growing ViBRANT and Scratchpads community with more than 6.400 active users at the end of the second project year. With the launch of the new version Scratchpads 2.0 in March 2012 a strong focus went into updating and improving training material notably the training courses and the online help wiki.

In order to promote Scratchpads use and to foster long-term sustainability of the Scratchpads community, the Ambassador programme was launched in the first project year. The network has now representatives from all over the world supporting their local community and linking the Scratchpads team with the growing user base to better understand the needs of users, so that the developers can keep improving Scratchpads (http://scratchpads.eu/support/ambassadors).

London's Natural History Museum, the lead partner in ViBRANT, has chosen to adopt the Scratchpads as part of its core business. This has resulted in the Natural History Museum creating a new post for the management of the project and assures the long-term stability of both the data entered into Scratchpads and the associated ICT infrastructure.

Central to these efforts is the inclusion of sociological studies of ViBRANT's user-base to get a better view of user needs and to influence the further development of Scratchpads and other ViBRANT tools as well as implementation priorities. Study results show that Scratchpads attract different audiences, especially the research and education category. Evaluation of the Scratchpads support services indicated that all services are highly appreciated by the user community. User satisfaction will further increase when all users are experiencing the many advantages of the Scratchpads 2.0 version.

---

1 Hardisty A. & Roberts D. (2013) A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecology* (submitted)

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), nomenclatorial acts to ZooBank, IPNI and Mycobank, all of which enhances their discoverability. 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/).

Pensoft (http://www.pensoft.net/), have launched a radical new type of publication in the Biodiversity Data Journal (http://biodiversitydatajournal.com/). This innovative journal is the first to introduce a full-cycle workflow for publication, from manuscript planning and creation, through submission, publication and archiving in relevant repositories. Authors prepare their manuscripts in templates through Scratchpads or Pensoft's own Pensoft Writing Tool, both of which are social media designed for collaborative working. The benefit this delivers is a very low-cost open-access journal that will accept a broad range of manuscripts, including traditional scientific papers, simple species descriptions or plain data sets. All will be peer reviewed and published through Pensoft's parallel output channels, as PDF, HTML and XML. Furthermore, key information types are automatically transmitted to appropriate central registers, including ZooBank, IPNI, Encyclopedia of Life, GBIF and others.

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. Recreational diving is most active over the summer season (July–August) and 5 diving clubs were involved in the COMBER activities over this period. They offered attractive discounts (20-25%) to divers who'd like to follow COMBER's underwater activities and also a discount for a ticket to the Cretaquarium and to an exhibition on Climate Change. Visitors to the Cretaquarium, on the other hand, had the option of a free tour and training on the fish species living in the tanks of the facility and illustrated in the FishCard (http://www.bio-watch.com/index.php?option=com_virtuemart&Itemid=95). The discount rate offered by the Cretaquarium was 50% of the ticket for both the aquarium and the Climate Change exhibition. More than 1,700 new records have been added to the COMBER web site of the pilot project in year 2, which will soon approach 4,000 records over two seasons. Professional divers from HCMR are also interested and involved in COMBER's activities. In summary, COMBER website: 273 dives submitted by 93 users - that is a total of 3,713 species.

RefBank, a bibliography of life, has been established and has grown to over 160,000 references available from several mirror server site, e.g. at http://vbrant.ipd.uka.de/RefBank/search. The user interface is currently simplistic, but work is in hand to embed the search facility into Scratchpads and the Pensoft Writing tool. Currently under development are routines to import and export free-text bibliographies to accommodate the large number of people who do choose not to use reference management software. RefBank is collaborating with other large aggregators in the field, including CiteBank and is investigating data exchange mechanisms with projects like Mendeley, Bibserver and CORE.

OBOE (https://vibrant.oerc.ox.ac.uk/) is, as far as we are aware, a unique service that offers a drop-and-compute facility through a single API. The single API is the unique factor and is likely to be instrumental in opening access to a much wider range of potential users. OBOE now supports the following tools:

- LEFT (http://www.biodiversity.ox.ac.uk/LEFT): a paid-for service, the Local Ecological Footprinting Tool produces an ecological valuation report for a specified area. It is intended as a pre-planning tool to use before undertaking further field-based investigations and quickly highlights areas of high ecological value to avoid when siting facilities with an environmentally damaging footprint

- LCC : Land Cover Change

- BICT: Biological Index Calculation Tool, to calculate a range of biodiversity indicators

- Mr Bayes (http://mrbayes.sourceforge.net/): a phylogenetic analysis tool

- BEAST (http://beast.bio.ed.ac.uk/Main_Page): a phylogenetic analysis tool

- MUSCLE (http://www.drive5.com/muscle/): a sequence alignment tool

- iKey+ (http://identificationkey.fr): creation of identification keys from SDD files

- Durden (https://git.scratchpads.eu/v/durden.git): Durden, created in WP2, will take a large image and create the tiles necessary for the image to be viewed using various client software such as the Bigimage Google map image viewer.

- GoldenGATE (http://idaho.ipd.uni-karlsruhe.de/GoldenGATE/): Golden Gate offers various web services to parse bibliographic references, dates, coordinates and quantities.

Significant progress on solving barriers to adoption and sharing biodiversity data infrastructures, tools, data and services has been made. Specifically:

1. the development of a web-based vocabulary management service (so called 'Ontology Platform') to discover, use, create and publish controlled vocabularies, now also including MediaWikis as a projected vocabulary management system;

and

2. the design and implementation of routines to facilitate interoperability and common access to the different platforms, including the Scratchpads platform (http://scratchpads.eu), EDIT CDM platform (http://wp5.e-taxonomy.eu/), Xper2 platform (http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2), Ontology Platform (http://terms.gbif.org). and the Pensoft publishing tools, specifically Pensoft Journal Sysytem (PJS) (http://www.pensoft.net/journals), Pensoft Taxon Profile (PTP) (http://www.ptp.pensoft.eu/), Pensoft Writing Tool (http://www.pwt.pensoft.net/), Biodiversity Data Journal (http://www.pensoft.net/journals/bdj) and Pensoft's TRIADA book and journal publishing platform (http://www.pensoft.net/)

Further details are available from the project web site at http://vbrant.eu

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	Biodiversity Overlay sl.	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	UdlR
17	Università di Trieste	UNITS
18	Vizzuality	Vizz
19	VU-VUmc	VU

3.2.1. Project objectives for the period

ViBRANT's objectives are set out below, as defined in Annex 1 to the Grant Agreement and as updated on the project web site[2]. 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.

---

2   http://vbrant.eu

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

The defining feature of ViBRANT's second year has been the integration between partners as prototype services have been developed to link together. The Scratchpads, while acting as a hub, have not been as fully integrated as we might like because the work on upgrading from Drupal 6 to Drupal 7 has been far more involved than anticipated. In consequence, other prototype services have developed network linkages between themselves. Such linkages were expected to develop and are desirable because they will provide a far more robust environment.

Of particular note was an initiative to consult the wider biodiversity community to build a decadal vision for biodiversity informatics[3] that has been submitted for publication[4] and an invitation to contribute to the the GBIC conference[5] aiming to unite diverse communities to address key policy and science challenges. These

---

3   http://is.gd/WhitePaperChapters
4   Hardisty A. & Roberts D. (2013) A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecology* (submitted)
5   http://www.gbif.org/orc/?doc_id=4695

independent initiatives have come to similar conclusions that highlight the broad priorities for the field, particularly the nature of the steps that are needed to build a generic infrastructure.

OBOE[6]

Typical of the collaborative spirit, OBOE's user numbers are growing exponentially. OBOE is, as far as we are aware, a unique service that offers a drop-and-compute facility through a single API. The single API is the unique factor and is likely to be instrumental in opening access to a much wider range of potential users. OBOE now supports the following tools:

- LEFT (http://www.biodiversity.ox.ac.uk/LEFT): a paid-for service, the Local Ecological Footprinting Tool produces an ecological valuation report for a specified area. It is intended as a pre-planning tool to use before undertaking further field-based investigations and quickly highlights areas of high ecological value to avoid in the location of facilities with an environmentally damaging footprint

- LCC : Land Cover Change

- BICT: Biological Index Calculation Tool (see WP8, page 69)

- Mr Bayes (http://mrbayes.sourceforge.net/): a phylogenetic analysis tool

- BEAST (http://beast.bio.ed.ac.uk/Main_Page): a phylogenetic analysis tool

- MUSCLE (http://www.drive5.com/muscle/): sequence alignment tool

- iKey+ (http://identificationkey.fr): creation of identification keys from SDD files

- Durden (https://git.scratchpads.eu/v/durden.git): Durden, created in WP2, will take a large image and create the tiles necessary for the image to be viewed using various client software such as the Bigimage Google map image viewer.

- GoldenGATE (http://idaho.ipd.uni-karlsruhe.de/GoldenGATE/): Golden Gate offers various web services to parse bibliographic references, dates, coordinates and quantities.

Note that LCC has been written by UOXF.E9 and iKey by UPMC within WP5, BICT by HMCR in WP8, Durden by NHM in WP2 and GoldenGATE by KIT in WP7.

RefBank[7]

RefBank is another collaborative success, having been developed through close cooperation between OU and KIT and then incorporating Pensoft in WP6. It has grown to over 160,000 references available from several mirror server sites. The user interface is currently simplistic, but work is in hand to embed the search facility into Scratchpads and the Pensoft Writing Tool (see below, WP6). Currently under development are routines to import and export free-text bibliographies to accommodate the large number of people who do choose not to use reference management software. RefBank is collaborating with other large aggregators in the field, including CiteBank and is investigating data exchange mechanisms with projects like Mendeley, Bibserver and CORE.

Scratchpads[8]

Scratchpad usage has continued to grow (Fig 1). Current access statistics can be found at http://scratchpads.eu/explore. The growth rate of registered users has slowed during 2012, but it is unclear whether this is market saturation or a reduction in the marketing effort compared to Year 1. The continued logarithmic rise in the number of Scratchpad sites suggest that we may be looking at the transition to a more mature market within the target systematics community.

---

6   https://vibrant.oerc.ox.ac.uk/
7   http://vbrant.ipd.uka.de/RefBank/search
8   http://scratchpads.eu/

Use of information in the Scratchpads themselves (Fig. 2), continues to rise. Note that the predominant use of US English does not reflect geographical origin of the user, but reflects the fact that almost all Scratchpad information is created in English. Work is in hand in WP3 to translate the user interface into other languages, initially targeting French, German, Spanish and Portuguese.

The Scratchpad system has been adopted by the NHM as part of its core informatics systems, which means that the Scratchpads' content and infrastructure are secure for the foreseeable future. This, of course, goes a long way to achieving the sustainability of the infrastructure.
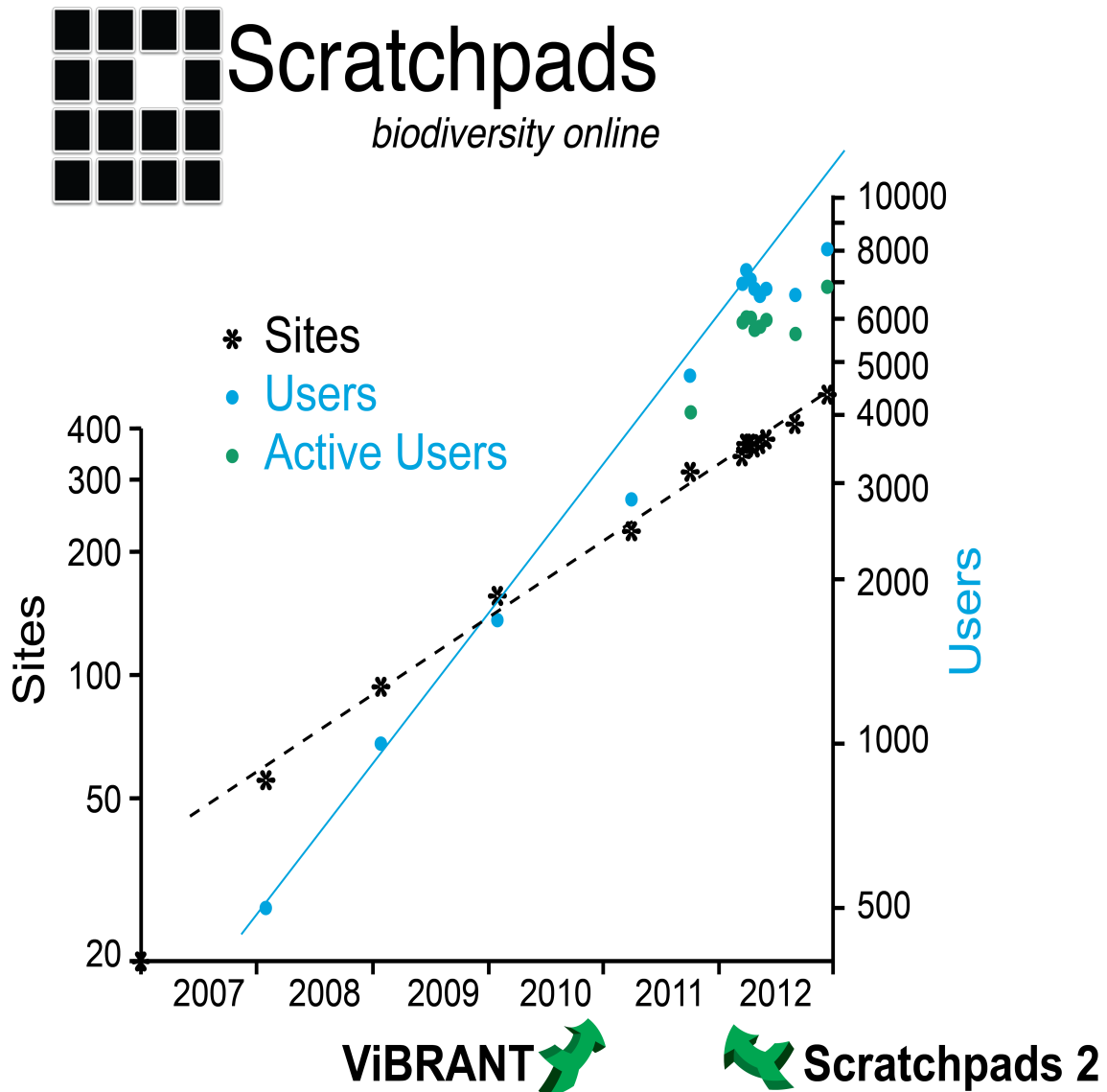
Figure 1 - Scratchpad uptake

ViBRANT
Virtual Biodiversity

Google Analytics

**http://myspecies.info - http://myspeci...
myspecies.info [DEFAULT]**

Audience Overview

Nov 1, 2007 - Dec 17, 2012

● % of visits: 100.00%

Overview
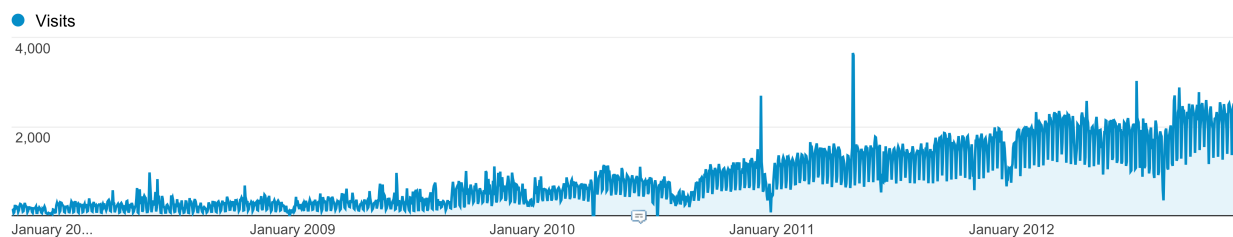
● Visits

1,049,747 people visited this site

Visits: **1,602,960**

Unique Visitors: **1,049,747**

Pageviews: **6,514,083**

Pages / Visit: **4.06**

Avg. Visit Duration: **00:03:26**

Bounce Rate: **57.71%**

% New Visits: **65.48%**

■ **65.47%** New Visitor
1,049,414 Visits

■ **34.53%** Returning Visitor
553,546 Visits

	Language	Visits	% Visits
1.	en-us	895,031	55.84%
2.	en-gb	132,516	8.27%
3.	fr	67,391	4.20%
4.	es	57,683	3.60%
5.	de	50,540	3.15%
6.	pt-br	49,440	3.08%
7.	nl	32,860	2.05%
8.	en	31,930	1.99%
9.	es-es	29,139	1.82%
10.	de-de	26,434	1.65%

view full report

© 2012 Google

Figure 2 - Access to Scratchpad information

GeoCat[9]

Launched at the end of Year 1, to date more than 1,000 assessments had been made and at least 5 assessments have officially been published by IUCN that made use of GeoCAT. IUCN is working on a method to implement a batch upload from GeoCAT to their system. GeoCAT has been cited 5 times already and has been presented in more than 10 workshops and conferences. It has become part of the standard toolkit for producing IUCN Conservation Threat Assessments.

From January 2012 to November 2012, we recorded 7,082 unique visitors to the application online. Around 40% of the users coming from UK, 20% Brazil, 10% US and the remainder countries around the world. The total amount of users, based on the visits and the participation in the forum is around 40 users worldwide with numbers increasing as more training/workshops are performed using the tool.

CartoDB[10]

An Open Source visualisation platform, built to make it easier for people to tell their stories by providing them with flexible and intuitive ways to create maps and design geospatial applications. It is now used by more than 7,000 users around the world (Fig. 3). Many users are through Biodiversity Informatics projects, like Map of Life, Vertnet or Canadensys, but the majority are currently not attached to the Biodiversity field. The developed platform is remarkably flexible with usages going from Carbon Calculators (UNEP-WCMC) up to visualising the results of US elections (Wall Street Journal). More than 100 applications have already been developed using CartoDB, and more than 20 Million visits had been made to visualisations using CartoDB. The current growth rate is of 1,500 new users per month. As part of Year 2 activities technical workshops have been held to explore the integration of CartoDB into OBOE services and the Scratchpads.
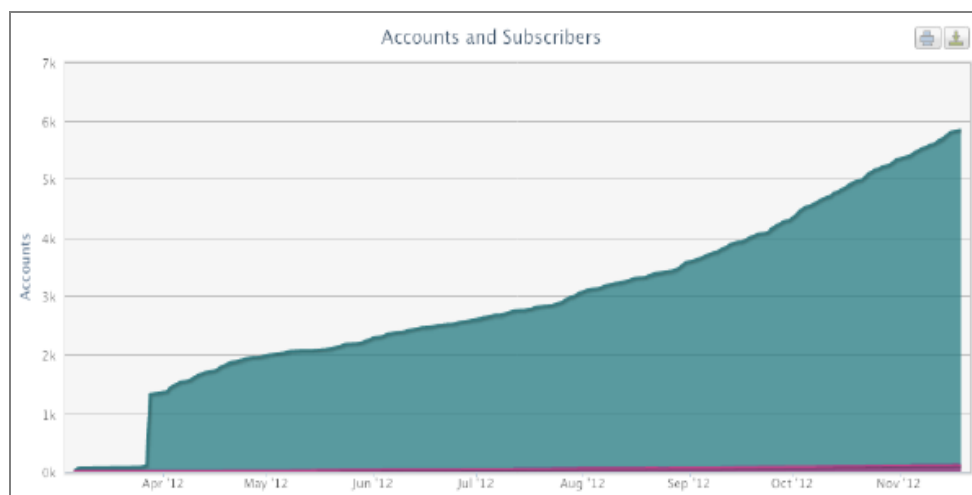

Figure 3 - Accounts and subscribers to CartoDB

COMBER[11]

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 to address that shortcoming. Recreational diving is most active over the summer season (July–August) and 5 diving clubs were involved in the COMBER activities over this period. They offered attractive discounts (20-25%) to divers who'd like to follow COMBER's underwater activities and also a discount for a ticket to the Cretaquarium and to an exhibition on Climate Change. Visitors to the Cretaquarium, on the other hand, had the option of a free

---

9   http://geocat.kew.org
10   http://cartodb.com/
11   http://www.comber.hcmr.gr

tour and training on the fish species living in the tanks of the facility and illustrated in the FishCard [12] (http://www.bio-watch.com/index.php?option=com_virtuemart&Itemid=95). The discount rate offered by the Cretaquarium was 50% of the ticket for both the aquarium and the Climate Change exhibition. More than 1,700 new records have been added to the COMBER web site of the pilot project that will soon approach 4,000 records over two seasons. Professional divers from HCMR are also interested and involved in COMBER's activities.

It is of interest to note that there were rather fewer recreational divers recruited in Year 2, reflecting a change in the tourism pattern. The 2012 season attracted far more holidaymakers from eastern Europe, where diving is a far less common hobby compared with northwestern Europe. Very few of those recruited in Year 1 returned to dive again in Year 2, so it is likely that such citizen science is a holiday activity and likely to be highly episodic. This does not diminish the value of the data generated, of course, and may lead to a more general engagement with observational science from the holidaymaker in the longer term.

Pensoft[13]

Pensoft have launched a radical new type of publication in the Biodiversity Data Journal [14]. This innovative journal is the first to introduce a full-cycle workflow for publication, from manuscript planning and creation, through submission, publication and archiving in relevant repositories. Authors prepare their manuscripts in templates through Scratchpads or Pensoft's own Pensoft Writing Tool, both of which are social media designed for collaborative working. The benefit this delivers is a very low-cost open-access journal that will accept a broad range of manuscripts, including traditional scientific papers, simple species descriptions or plain data sets. All will be peer reviewed and published through Pensoft's parallel output channels, as PDF, HTML and XML. Furthermore, key information types are automatically transmitted to appropriate central registers, including ZooBank, IPNI, Encyclopedia of Life, GBIF and others.

The launch of the BDJ created considerable interest within the biodiversity community and following a call for editors of the new journal, 194 people had applied in only 4 weeks.

MediaWiki

For a host of technical reasons, the use of Knowledge Organising Systems (KOS) are desirable. In Year 2 GBIF and JKI have agreed that MediaWiki represents the most stable solution to introducing KOS services to the broader community. This remains a huge sociological challenge because, for an average biodiversity data contributor, it is not yet clear to an early adopter that they will gain any benefit from their invested effort. Nevertheless, an architecture has been developed supported by several major players. http://terms.gbif.org/ is an end-user oriented product that is used by GBIF and the GBIF contributor community. Whether its use can be extended to a wider audience remains to be elaborated.

---

12   http://www.bio-watch.com/index.php?option=com_virtuemart&Itemid=95
13   http://www.pensoft.net/
14   http://www.pensoft.net/journals/bdj

WP2 - Technical architecture

Lead: Simon Rycroft, Natural History Museum 　　　　　Activity type: Research

Partner	Name	Acronym	Effort (PMs Year 2)
1	The Natural History Museum, London	NHM	30.07
15	Freie Universität Berlin	FUB-BGBM	3.55

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 2

Mirroring the Scratchpad server in Berlin (Year 1) was intended to provide stability and reliability in the event of hardware failure in the primary London site. In fact the major problem with reliability has been the design of the server architecture to maintain acceptable response times while usage continued to increase. Scratchpads are served from a virtual host and during Year 2 we have experimented with various configuration of the available resources, specifically by creating additional hosts for discrete tasks, such as running the Apache web server. It is unlikely that an optimum configuration can exist for long periods of time. As patterns of use change, then the nature of the load in the servers change. Such changes are not visible to the user and typically result in a change in the number and pattern of calls made to the underlying database.

Distributing Scratchpad servers

The Scratchpad code base is available from our GIT repository[15]. The software has been installed, although significantly modified, by GBIF to deliver the Nodes Portal Toolkit[16], and by the Royal Botanical Gardens (Kew) to deliver a collaborative working environment tailored to botanical communities, specifically through the eMonocot project[17]. Issues relating to long-term support and integration of diverse sites will be addressed below in Task 4 'sustainability'.

15　http://scratchpads.eu/develop/repository
16　http://www.gbif.org/participation/participant-nodes/nodes-portal-toolkit/
17　http://e-monocot.org/

Integration with other projects

The Scratchpad code base has been mounted on an experimental server by the EU funded agINFRA [18] project to assess the integration with the modules written for agriDRUPAL [19], specifically the Indexing of resources using Agrovoc, supporting Agrovoc multilingual features and supporting a Linked Open Data structure.

Dr. Sacha Spector has employed Trellon[20], a multinational Drupal development company, to realise a Scratchpad version of the ScarabNet portal[21] that he works on. This will be demonstrated at a meeting at the Natural History Museum in February.

Dr. Eli Sarnat has employed developers to extend the functionality of his Antkey Scratchpad, which have been integrated into the Scratchpads code base for all other Scratchpads.

Further work

We will continue to monitor the usage pattern of server resources so that an effective level of end-user service can be maintained from an economically realistic platform.

Task 2: Multi-site integration (NHM, BGBM)

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 2

Scratchpads registry

As reported in D2.2[22], the Scratchpads registry was delivered on time and is fully integrated into Scratchpads 2.0. The decision was made to not include any registry code in Scratchpads 1.0, pending their upgrade to Scratchpads 2.0. GBIF[23] are not currently running the latest version of their registry software, against which the Scratchpads registry module was developed, so we are running a clone of the GBIF registry using the new software until GBIF are confident that the new software is sufficiently robust for public use.

Scratchpads statistics module

The Scratchpads statistics module builds on the work carried out in meeting deliverable D2.2. Each Scratchpad provides, with the help of the Statistics module, an additional statistics endpoint, which is registered with the GBIF registry. Using these Scratchpads endpoints, which are registered with the GBIF registry, it is possible to query each Scratchpad to harvest its usage information. This usage information can, in turn, be aggregated to provide a general overview of the usage of all of the Scratchpads. The Scratchpads statistics module is currently only enabled on the Scratchpads website, and is currently capable of displaying the distribution of content across each Scratchpad. As part of ViBRANT's year 3 activities we plan to develop this further.

---

18 http://aginfra.eu/
19 http://www.agridrupal.org/
20 http://www.trellon.com/
21 http://scarabnet.amnh.org/
22 http://vbrant.eu/content/d22-scratchpad-registry
23 http://gbrds.gbif.org/

Publication of usage statistics

The re-developed Scratchpad home page[24] (see task 3 below), contains a dynamic reporting mechanism for usage statistics across all sites[25], while statistics for individual sites are available on the list of all sites[26].

The Darwin Core data transfer standard

The Darwin Core standard[27] (DwC) has become widely adopted within the biodiversity informatics community and has been extended by the Darwin Core Archive (DwC-A) to include additional materials not explicitly covered by the DwC itself. It is essentially a set of text (csv) files with a simple descriptor (meta.xml) bundled into a single zip file. Since its introduction in 2009, it has quickly become the format of choice to facilitate interaction between biodiversity resources. A module to produce DwC-A files from Scratchpads was introduced this year and mechanisms to use DwC-A files to bring content into a Scratchpad was similarly added. This module supports harvesting of Scratchpad data by the CDM and via externally funded activities like the Encyclopedia of Life and eMonocot project.

Data search portal

The data search portal will provide a means for a user to search all Scratchpads indexed in the EDIT Common Data Model (CDM). A test set of taxonomic data from different Scratchpads have been manually imported into the CDM using DwC-A and can be searched through a standard interface[28].

Future work

It is intended that the NHM, as a ViBRANT partner, will become a permanent mirror of the GBIF registry.

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 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 2

Release of Scratchpads 2.0

Scratchpads 2.0 was released to all users signing up for a new Scratchpad 1st March 2012. The initial release had a number of teething problems which were quickly ironed out with the help of our users. Since this initial release, we have quickly built up a large number of Scratchpads 2.0 sites (153 on 10/12/2012), as well as a large amount of data across all of the sites (>80,000 data objects as of 10/12/2012). See the statement on 'Impact' in section 3.2.2 above.

Scratchpads 2.0 publication module

The Scratchpads 2.0 publication module has built on the preliminary prototype developed in Scratchpads 1.0. The Scratchpads 2.0 module is now capable of automatic submission to Pensoft's Writing Tool (PWT) being developed by ViBRANT WP6. We have taken the opportunity to rewrite the module, changing its look and feel completely, creating a tool which better resembles a WYSIWYG word processor to make the tool feel more natural to end users.

---

24   http://scratchpads.eu/
25   http://scratchpads.eu/explore
26   http://scratchpads.eu/explore/sites-list
27   http://rs.tdwg.org/dwc/
28   http://dev.e-taxonomy.eu/vibrant_index/search/

Integration of Scratchpads 2.0 with RefBank

RefBank, the ViBRANT WP7 product, has been integrated into the Scratchpads in two ways. First, it is harvesting published bibliographies from all Scratchpads (see the WP7 report below). Second, the Scratchpads allow a user to search RefBank from their Scratchpad, and to incorporate any references that they find of interest. Currently the search from within a Scratchpad is done on author name(s) and year of publication, although we hope to extend this in the future to include taxonomic name.

Scratchpads.eu site redesign

The Scratchpads website[29] has gone through a considerable redesign, and re-branding. All of the content on the site has been rewritten and restructured to make the content as clear as possible and to help maintain a consistent brand message. This will greatly help us with promotion of the Scratchpads, presenting a clear message summarising what the Scratchpads do, and at whom they are aimed.

Issues Queue

We have installed and integrated into the Scratchpads platform an improved issue tracker called Redmine [30]. Redmine is much more flexible and feature-rich than our previous issue tracker. We can now assign issues to groups of people (e.g. Developers, Support team) and can have issues dependent on other issues. We are also able to create Gantt charts and similar reports, as issues now have deadlines and estimated times, to aid project management.

We also intend to use the Scratchpads installation of Redmine as a management tool in projects that are outside of ViBRANT but closely related to the Scratchpads. This would not have been possible with our previous issue tracking tool.

Continued module development

We have continued to develop the Scratchpads through the issues queue. A substantial number of refinements and new features, detailed in M2.17[31], have been added to support our users in their use of the Scratchpads. A few of these changes are listed here:

- Taxon pages: show data from synonyms under valid name

- Country maps for genera aggregating species data

- Display videos

- Taxon pages: Create EOL panel

- Add a license field to the image matrix

- Bulk upload of files/images

- Select TDWG distribution regions from a list

- EOL export

- Character project: Improve workflow

- Adding a Twitter feed

- Add scale bar to maps

Upgrade of Scratchpads 1.0 sites to Scratchpads 2.0

The process of upgrading the Scratchpads from 1.0 to 2.0 has taken a significantly longer than expected essentially because Scratchpads 1.0 were less structurally constrained than Scratchpads 2.0, which, in turn, has resulted in a much more heterogeneous mixture of Scratchpads. Conversion of over 300 Scratchpad

---

29 http://scratchpads.eu/
30 http://www.redmine.org/
31 http://vbrant.eu/content/m217-ongoing-module-development-supporting-prioritised-features-requested-users-through-wp3

1.0 sites demands automation, but the necessary scripts have to be tuned for sites individually, or at best in small groups. The bulk of the work has been channelled through one developer, to enable us to continue both Scratchpads 2.0 development and Scratchpads 1.0 and 2.0 support.

Unit testing

Unit testing[32] is a mechanism to automate the testing of software components for compatibility. All software funded by ViBRANT is, as specified in our Consortium Agreement, open source. Sustaining open source projects, however, is not free of costs. The Scratchpads core will continue to have some developer support from the NHM, as discussed above, but several new and independently funded projects, such as David Maddison's NSF funded Bembidiina of North America, and Andrew Hipp's NSF-funded *Carex* monograph[33], will be developing additional modules. Incorporation of new modules within the Scratchpad framework currently requires extensive compatibility testing on multiple sites with many browsers, which is inevitably time-consuming. During Year 2 we have implemented a standard tool that allows us to include tests into our own modules that can establish compatibility with the system core and with combinations of other modules. Unfortunately this does not allow us to test client-side components. Taken together, the server- and client-side unit testing will reduce the workload of sustaining future Scratchpad development.

Future work

To continue the development of new functionality and to incorporate, if possible, a mechanism for client-side unit testing. This is being managed through the Scratchpad issues queue developed as part of WP2 activities[34].

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 2

The NHM has appointed a new Director for Science and a new Head of ICT who have begun restructuring the way the Museum works. As part of that reorganisation, several new posts for developers have been created and work has started to integrate information-rich areas and to ensure their interoperability. The success of the Scratchpads project has led to their incorporation into the Museum's core strategy and the extension of Drupal as a management system into other areas of the Museum's work. This resulted in the WP2 leader being made a permanent member of NHM staff and essentially guarantees that the core services offered through Scratchpads will continue to be supported for the foreseeable future.

The work on data publication is led by Pensoft, an SME, and will be sustained by the Pensoft commercial model.

Beyond these two areas, work on D2.3 (Financial sustainability) has been focussed on defining actual costs of services, identification of service beneficiaries, and potential owners of those services. We expect the report to be ready as forecast.

---

32   http://vbrant.eu/sites/vbrant.eu/files/m2.16.pdf
33   http://systematics.mortonarb.org/carex/index.php/BioSynC_2011
34   http://support.scratchpads.eu/

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 2)
1	The Natural History Museum, London	NHM	1.10
3	Royal Belgian Institute of Natural Sciences	RBINS	14.80
5	Vrije Universiteit Amsterdam	VU	12.00
17	Università di Trieste	UNITS	0.60

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. 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.

5. 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.

6. 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 2

Training is provided as part of the support and outreach programme. The major objective of organising training courses is to enhance the use of ViBRANT tools and support users working with them. Hence strengthening existing virtual research communities and facilitating the formation of new ones.

As part of the larger and integrated support scheme, in addition to on-site training courses, other strategies are used to deliver training to a larger audience: an online help system, including detailed documentation (cf. task 4), and the Ambassadors' network (cf. task 2) have been developed.

In March 2012 a new version of Scratchpads (v.2.0) was released, so all existing training material including courses, manual and help pages have been updated to reflect the new version. In November 2012 the new Scratchpads.eu portal was also launched, holding more information for the ViBRANT project and the Scratchpads communities.

Three types of training courses were offered during ViBRANT year 2: (i) updater training for maintainers to inform them about the new Scratchpads version, (ii) basic training for those interested to start a site or for less experienced users and (iii) advanced courses for experienced users.

The one-day Scratchpads courses aim to guide current and prospective Scratchpads users through major Scratchpads features and illustrate Scratchpads' ability to support online biodiversity research communities. Participants learn best practices and gain a better understanding of what the Scratchpads can do for them and their research community.

As the on-site training includes hands-on practice, places are limited and applicants need to register on the Scratchpads site[35]. Due to the limited number of places (ca. 10-15 per course, max. 20) applicants are screened based on a list of criteria. Where the demand for places exceeds capacity we will either recommend alternative courses (other venues) or organise additional training events. There is no course registration fee and participants receive a printed training manual. In order to optimise training, some of the courses focus on needs of larger research communities and thus are organised in association with high profile national or international conferences.

Figure 4 - Scratchpads training course, Hellenic Centre for Marine Research, Heraklion, Crete, 17 October 2012

Figure 5 - Interviews are carried out by the VU on the expectations and use experiences of Scratchpads users. Scratchpads training course, Stockholm, 7 June 2012

The following table presents the training organised during ViBRANT year 2. Note that in February 2012 a Scratchpads course was held as part of the NHM/Imperial College MSc course in Taxonomy and, in cooperation with WP8, a training course was organised in the framework of the COMBER project (Fig. 4).

---

35   http://scratchpads.eu/support

ViBRANT Year 2 - Organisation Scratchpads training courses				
	Course level	Location	Date	Participants
1	Basic SP	Natural History Museum London/Imperial college	13.02.12	29
2	SP2 update maintainers	Royal Botanic Gardens Kew	20.03.12	8
3	SP2 update maintainers	Natural History Museum London	21.03.12	6
4	SP2 Basic	Natural History Museum London	23.05.12	12
5	SP2 Basic	Swedish Museum of Natural History Stockholm°	07.06.12	9
6	SP2 Advanced	Swedish Museum of Natural History Stockholm°	08.06.12	5
7	SP2 Advanced	Natural History Museum London	12.06.12	2
8	SP2 Basic	American University Blagoevgrad°	26.06.12	15
9	SP2 Basic	Natural History Museum London	16.10.12	5
10	SP2 Basic	Hellenic Centre for Marine Research Heraklion	17.10.12	20
11	SP2 basic	Kirstenbosch National Botanical Garden, Cape Town*°	19.10.12	12
12	SP2 Basic	People's Palace, Beijing*°	29.10.12	12
13	SP2 Basic	Joinville, Brazil*°	11.11.12	13
				148

* e-monocot funded
° organised in association with international conference

Whereas in the first year priority was put on the development of the Ambassadors network (cf. task 2) in the second year a lot of effort was put on the delivery of training as this is related to the launch of the new Scratchpads version. As reflected in the following table, during the second year training courses doubled and the number of different venues significantly increased.

Scratchpads training courses for Year 1 & Year 2			
	Participants	Training courses	Venues/host institutions
Year 1	70	6	3
Year 2	148	13	8

Planning for forthcoming training courses during Year 3 is ongoing and is advertised on the web [36].

In addition to the delivery of on-site training courses, a range of electronic resources is accessible to users at their home institution via the website:

- an illustrated training manual, and online documentation of all Scratchpads features;
- video tutorials;

---

36   http://scratchpads.eu/support/training/events

- an inbuilt help system in each Scratchpad;
- the Scratchpads Help wiki[37];
- training sites[38] can be requested for extensive testing;
- sandbox site[39] allowing people to try out all the available functions;
- dedicated issue tracker, through which users can register technical or support related requests (cf. task 4).

Future work

As the migration of Scratchpad sites gathers momentum we anticipate greater demand for training courses. Furthermore with Scratchpad technologies being used in other projects and with a growing range of external services through Scratchpads themselves we shall be evaluating the most effective place to direct our training resources.

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 2

In order to promote Scratchpads use and to foster long-term sustainability of the Scratchpads communities, the Scratchpads Ambassadors programme was launched in May 2011. Ambassadors are enthusiastic and experienced Scratchpads users that spread the word about Scratchpads, promote their use and either arrange or give training in their local community. Ambassadors are our points of contact for Scratchpads users in their local research community and in that way they help the Scratchpads team better understand the needs of users, guiding the development steps of Scratchpads.

To recruit the first group of Ambassadors, a passive approach was used. A call to become an Ambassador was announced on the Scratchpads site and ViBRANT site in May 2011. In addition an active approach was used by analysing the Scratchpads users database. A selection of potential Ambassadors was made based on their maintainers' role, their online activity, content and number of users of their sites. Also geographical location was taken into account as we were aiming for a global spread of the Ambassadors. Maintainers who met the criteria were contacted by email to become Ambassadors. After the Ambassadors initiative was announced on the Scratchpads website in September 2011, a global network of 9 Ambassadors was in place.

Through the website and word of mouth spread of the programme, potential ambassadors have since then contacted us directly. All the Ambassador applications are screened to assure that their level of Scratchpads knowledge is adequate. Novice applicants are encouraged to re-apply after they obtained more experience.

---

37   http://help.scratchpads.eu/w/Main_Page
38   http://scratchpads.eu/support/training/apply-for-site
39   http://scratchpads.eu/support/sandbox

Since the programme was launched, 16 Ambassadors from 15 countries were recruited. More information about the programme can be found on the Scratchpads website[40]. Ambassadors' activities such as beta testing of the new Scratchpads version, giving demos, organising interviews and training are described in M3.14[41].

A ViBRANT promotional strategy has been developed and tools to achieve the outreach activities were defined in M3.10 Delivery of a promotional strategy for ViBRANT services[42]. An evaluation of the promotional strategy, based on the usage of ViBRANT tools and services, is given in M3.14 Assessment of user support services and promotional activities[43].

In cooperation with WP1, the public-facing website has been updated and organised around tasks rather than workpackages per se.

Future work

The upgrades to both the ViBRANT and Scratchpad web sites serve to advertise the growing range of services available. We anticipate working with the ambassadors in Year 3 to maximise the benefits they can derive from the ViBRANT infrastructure, particularly to sustaining the human networks that have been established.

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 organisations - 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 2

Changing attitudes to data sharing

"Most academics would kill to be interviewed about their work, even as they cling tenaciously to copyright on their unread articles. Many diaries are written to be read (ultimately). As

---

40   http://scratchpads.eu/support/ambassadors
41   http://vbrant.eu/sites/vbrant.eu/files/M3.14_Assessment_usersupportservices_and_promotionalactivies-281112.pdf
42   http://vbrant.eu/sites/vbrant.eu/files/M.3.10_Delivery of a promotional strategy for ViBRANT services_0.pdf
43   http://vbrant.eu/sites/vbrant.eu/files/M3.14_Assessment_usersupportservices_and_promotionalactivies-281112.pdf

Wilde put it, 'there is only one thing in the world worse than being talked about, and that is *not* being talked about.'"[44]

The Open Science movement has burgeoned during ViBRANT's second year, with influential reports being published, e.g. the Finch[45] and the Royal Society[46] reports, and the Research Councils UK (RCUK) and the Wellcome Trust, a major UK biomedical funder, both creating policies that require funded research to be published in a free-to-access form. ViBRANT has been active in this arena, demonstrating the benefits of sharing data and the ease with which it can be done, e.g. as an invited delegate at the International Council for Science (ICSU) workshop "Revaluing Science in the Digital Age" and through the work of WP6 in creating systems to simplify open access publication at minimal cost.

Central to the success of open publication and sharing is a change in attitude within the community as to normal behaviour. Work by WP3 has detected a growing realisation that increased exposure brings benefits and recognition, "making invisible work visible"[47]. Hence, infrastructures like ViBRANT make pieces of work and the organisation of work visible to a large audience, where before these were only visible to colleagues sharing the same office or otherwise working in close proximity to each other. These work activities and work settings include among others: (informal) communication and brainstorming; sharing work in progress (annotated texts) and visibility of network relations. Tools like forums, bug/issue trackers, groupware, access to use-metrics (e.g. number of visits, downloads), and social network applications, make that today in science not only the final research article or product is shared, but also how the work settings are organised, what steps are taken to get to the final output, as well as how often the output is used. If facilitated well, these tools help researchers in advancing and communicating their work and interactions, and will give a more detailed picture of the impact and use of biodiversity research in general.

During ViBRANT's first year we found evidence that the mechanisms which help researchers to make invisible parts in their work visible, are also likely to give them more control over their work and so increase their autonomy[48]. The COMBER project shows that respect, given by professionals to volunteers, is another important element to make users engage.

Concepts and tools that extend user's autonomy, like access to data and people (not limited by organisational, or national boundaries) are also thought to be central ideas of ViBRANT, offering advantages over the traditional offline organisation of science. Highlighting these advantages will help increase user uptake. Below we summarise the lessons learnt from four activities and in doing so we give a list of mechanisms that we think have the potential to further increase user uptake in ViBRANT.

1. The user studies of the VU aim to provide insights on the social context of ViBRANT tools and services. In doing so the VU wants to contribute to the development of mechanisms to increase user uptake.

    - The 'measure the scope of the audience' method which is currently under development, is based on the principle that meaningful impact reports on online-work provide a tool for researchers to make their invisible work visible and so enhance their autonomy. In doing so an attractive condition for user participation is created. Furthermore, this study allows us to understand ViBRANT users better, which gives us valuable pieces of information. For instance, if we need to target our outreach and training activities, either to reinforce areas where we are already strong, tailoring the materials to that specific interest group, or rethink why we are missing some potential users altogether. This will inform our choice of materials to develop, at which conferences to offer workshops, and how to best use the available funds.

---

44 O'Hara, K. & Shadbolt, N. (2008). The spy in the coffee machine. Oxford: Oneworld. p20.

45 http://www.researchinfonet.org/publish/finch/

46 http://royalsociety.org/policy/projects/science-public-enterprise/report/

47 http://vbrant.eu/sites/vbrant.eu/files/Milestone_311_VU_130112.pdf

48 Autonomy is generally considered an important element of professionalism (cf. Engel G.A. 1970 Professional autonomy and bureaucratic organization. Administrative Science Quarterly **15**(1): 12-21. URL: http://www.jstor.org/stable/2391182).

- The method and theory of multiplex networks suggest that Scratchpad membership has an added value in network terms to traditional academic network relations, such as coauthor relations. Stressing this point in promotional activities should help increase user uptake. Hence, a more extended empirical study is needed to test the hypothesis.

Furthermore, the pilot study carried out by the VU[49] suggests there is a second mechanism at work that might help increase user uptake, which is social navigation. Collecting the data on multiplex, academic network relations of Scratchpad/ViBRANT users (co-authors; citations; bibliographies; editorial boards networks) and making them visible to others, will help users more easily to link to people and ideas and so contribute to the creation of new knowledge. If social navigation tools will actually contribute to the increase of user uptake of ViBRANT tools is something to be studied, although previous work on social navigation systems show their benefits to the overall usability of collaborative spaces and to the facilitation of learning[50].

2. Organising co-learning between developers and users by way of online tools such as the Scratchpad issue tracker is a clear example of a mechanism that encourages and facilitates the involvement of users in the design of Scratchpads. Being included in the improvement of the system, by identifying bugs, and by making feature requests gives users a sense of responsibility and autonomy. If something is not working as expected, users can actually do something about it.

3. The Scratchpad training courses lower the barrier for new users to join Scratchpads and for experienced users to help them bring their site use to a higher level. The courses give users information to help them organise their research community globally if they wish so, not limited by institutional or national boundaries nor by expensive travel costs. In addition, the ambassadors programme make power-users visible to other users and so offers a structure for help support between users, independent from the training and support given by the Scratchpad support team. Again an example of giving autonomy to users, which should be an asset for user uptake and long term sustainability of Scratchpads.

4. The COMBER project[51] [WP8] shows that citizens are keen to be involved in science, especially if their contributions are made visible and are linked to a larger data aggregator and if they receive training and the right training material. Key here is that they want to make a difference but on the condition that they and their expertise are taken seriously and treated with respect, e.g. providing them with professional training and material.

Service delivery and evaluation

Our aim is to build an IT infrastructure which is close to our user's needs and support users in achieving their goals. Vision and strategies to integrate the perspective that we applied are explained in a report [52], together with evidence that demonstrate where we were successful and where there is more work to do. From the evaluation of our services the following conclusions could be drawn:

49 Duin D, Van den Besselaar P (2011) Studying the effects of virtual biodiversity research infrastructures. In: Smith V, Penev L (Eds) e-Infrastructures for data publishing in biodiversity science. ZooKeys 150: 193–210. doi: 10.3897/zookeys.150.2164

50 Farzan R, Brusilovsky P (2006) Social navigation support in a course recommendation system. In: Wade V, Ashman H, Smyth B (Eds) Adaptive Hypermedia and Adaptive Web-Based Systems. Lecture Notes in Computer Science, 4018. Springer Berlin / Heidelberg, 91- 100. doi: 10.1007/11768012_11
Wu M & Bowles CT (2010) Principles for applying social navigation to collaborative systems. Proceedings of the 4th Symposium on Computer Human Interaction for the Management of Information Technology CHIMIT'10, November 12–13, 2010, San Jose, CA, U.S.A, 2-2.

51 Arvanitidis C, Faulwetter S, Chatzigeorgiou G, Penev L, Bánki O, Dailianis T, Pafilis E, Kouratoras M, Chatzinikolaou E, Fanini L, Vasileiadou A, Pavloudi C, Vavilis P, Koulouri P, Dounas C (2011) Engaging the broader community in biodiversity research: the concept of the COMBER pilot project for divers in ViBRANT. In: Smith V, Penev L (Eds) e-Infrastructures for data publishing in biodiversity science. ZooKeys 150: 211–229. doi: 10.3897/zookeys.150.2149

52 http://vbrant.eu/sites/vbrant.eu/files/D3.2_291112.pdf

Users, use of Scratchpads and its support services:

The evaluations show that we were successful in increasing our user base, number of sites, and increase in content.

We are happy to see that Scratchpads attract different audiences. We were in particular satisfied to see that audiences from the research and educational organisations find their way to our platform as this is the heart of where Scratchpads wants to make a difference.

Through the different feedback channels we have at place, we received positive evaluations from our users on what we try to achieve with Scratchpads. We are confident that user satisfaction will further increase next year when all users are experiencing the many advantages of Scratchpads 2.0.

During ViBRANT we invested in improving and expanding our support services. In the user feedback we see that most services are highly valued and actively used. The combination of distance support and local face-to-face support makes support easily accessible for all users (all levels of IT experience, in different languages). Local support will be further extended in the future with help of the ambassador programme. Users actively report to us by the issue tracker and email. We are confident that our support services are ready for further growth and new technology.

We have different media at place to communicate with users and from the evaluation of the different channels we learned that we reach many users and that many users are able to reach us. Yet some told us they find it hard to stay updated about the system improvements and that we need to reduce the IT jargon in our communication. Hence we aim to further promote the different communication channels among our users and rephrase our messages. Our overall communication and dissemination strategy is discussed in the report "Delivery of a promotional strategy for ViBRANT services"(M3.10).

We are aware that growth of services and products and migration to a new version of the platform also challenges the physical infrastructure like our server capacity. We recently installed a new statistics module to monitor in more detail, and in a more user friendly way, the number of maintainers, users and visitors to Scratchpad, as well as the volume of data storage and traffic. This information will help us to improve server capacity management and planning in the future.

Users, use of other ViBRANT services

The COMBER project links ViBRANT to members of the public and demonstrates how technology and (social) networks between the academic community and the public and can speed up data collection. It is an example of how with existing tools and expertise things can be done differently and smarter. The 2012 observation numbers lagged slightly behind those of 2011.

Geocat and CartoDB are flagships services offering tools for academic and non academic communities to process and visualise biodiversity information. The user uptake of the only recently launched CartoDB is remarkable and makes us keen to follow how the numbers will develop over next year. We are in particular satisfied with both services as they demonstrate the potential of open data services for research, policy and empowering citizens, an objective that is high on the VIBRANT agenda.

OBOE services offer Scratchpads users a cloud computing service according to the latest standards for efficient data processing. We are in particular proud to facilitate the development of such a service as it is a flexible, reliable and green way of data computing, three core values with which ViBRANT identifies itself. Next year we will focus on how to further increase the user numbers.

Future Work

The unfortunate departure of Daphne Duin from the project means that we need to re-assess the work for Year 3, while adhering to the goal of working towards a better understanding of the diffuse (potential) user community for biodiversity informatics. This will be done through different activities.

- First, we will map biodiversity research, as well as related fields that may be interested in biodiversity research. This is expected to reveal the extent of the potential scholarly user community.

- Secondly, we will use the developed tool (with WP7) to investigate the diversity of (also non-scholarly) user groups and the frequency of their use. This is based on Google Analytics data, covering all users and not only the registered ones. We will also analyse the use of the Scratchpads using the new registry information implemented by WP2.

- Thirdly, we will look at the changes that occur in behaviour as people start to use Scratchpads and to share their work through these new technologies. This will be done by

  i. a series of interviews with the users, and

  ii. network analysis of collaboration patterns.

  The latter may show whether scratchpads (and possible other ViBRANT services) reinforce existing collaboration networks, or whether they enhance and extend collaboration networks. If the latter, are these new networks also different in a cognitive sense, e.g. do these networks lead to new multidisciplinary collaboration.

- Finally, (on the wish list in case resources allow) we will analyse the use of other ViBRANT services (where data are available) and do a brief survey of biodiversity knowledge-use by non-scholarly audiences, such as industry, and governmental and non-governmental organisations.

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 2

The Scratchpads platform provides at present the following support systems for users:

- the help desk dealing with the issues, emails, calls and meetings related to user support;

- dedicated issue tracker, through which users can register technical or support related requests;

- an inbuilt help system in each Scratchpad;

- the Scratchpads 2.0 Help wiki;

- the sandbox and home training sites for self-training;

- training courses and online training manual;

- a blog to inform the community of new features, Scratchpads update, training courses etc.

With the launch of Scratchpads version 2.0 in March 2012, Scratchpads support services such as help pages, training manual and video tutorials have been adapted. A new help system has been developed, the Scratchpads 2.0 Help wiki[53]. The Help wiki contains an illustrated training manual, a glossary of Scratchpads terminology, video tutorials and documentation on advanced topics with examples. The wiki is a dynamic resource, being continuously expanded and updated. Taking into account the users' feedback, improvements

---

53   http://help.scratchpads.eu/w/Main_Page

have been made, e.g. incorporation of screenshots in the training manual. A glossary with Scratchpads terminology is available online.

The Scratchpads Help wiki can be used to create printable PDFs of pages should a training manual similar to those used on training courses be required. Currently the wiki is only editable by the Scratchpads Team but it is planned to open this up to ambassadors, some selected users or to all Scratchpads site maintainers in the future.

Getting to know the specific needs of the Scratchpads users and of potential users is important (Fig. 5) for further development of Scratchpads 2.0 and its support services. Within the framework of the sociological studies of the VU (cf. Task 3) a survey was distributed to all Scratchpads 1.0 maintainers in the period September-November 2011. The results of the feedback are discussed in detail in the report 'Scratchpad Survey Maintainers 2011' and recommendations for development, project management and future studies have been formulated. Both reports and the questionnaire are available on the ViBRANT website [54].

Future Work

In the coming year we will continue to update the help wiki and its support services: manual; help pages; video tutorials; etc..

---

54   http://vbrant.eu/content/scratchpad-maintainer-survey-2011

WP4 - Standardisation

Lead: Yde de Jong & Wouter Los, University of Amsterdam   Activity type: Co-ordination

Partner	Name	Acronym	Effort (PMs Year 2)
6	Julius Kühn-Institute	JKI	6.30
8	University of Amsterdam	UvA	1.00
12	PENSOFT Publisher	PENSOFT	1.00
13	Université Pierre et Marie Curie-Paris 6	UPMC	6.81
14	Global Biodiversity Information Facility	GBIFS	6.00
15	Freie Universität Berlin	FUB-BGBM	12.01

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 extent 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.

ViBRANT
Virtual Biodiversity

Progress in Year 2

Knowledge Organisation System (KOS)

The GBIF Vocabularies Service (Fig. 6) has been migrated from the Scratchpads server hosted by the Natural History Museum in London to new servers hosted by the GBIF Secretariat in Copenhagen for improved performance (M4.15)[55].
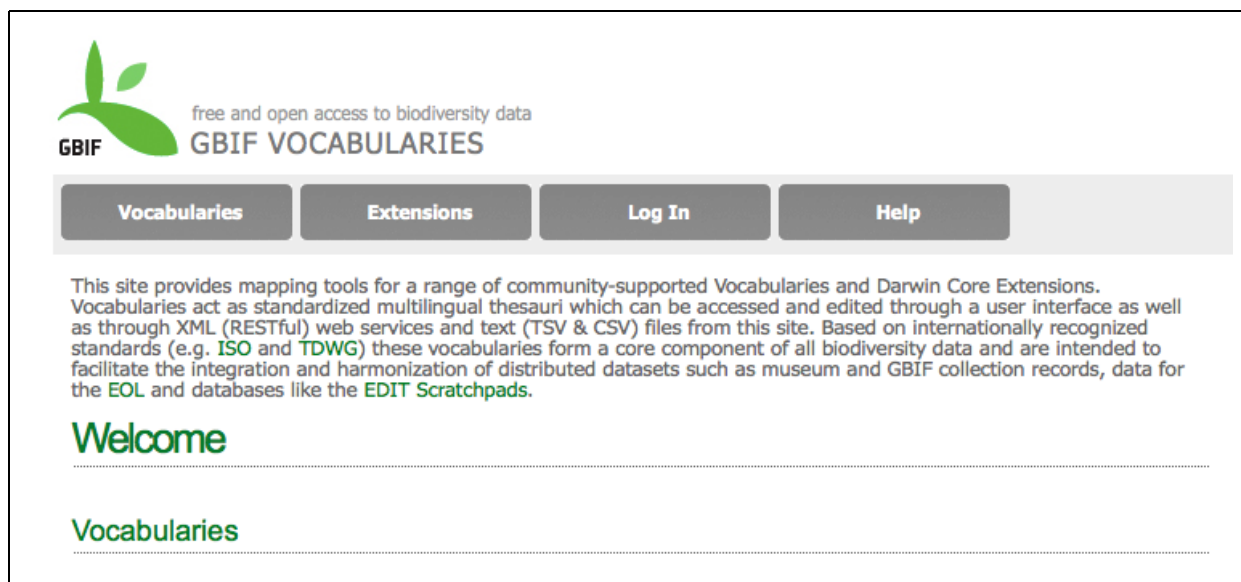
Figure 6 – GBIF Vocabularies Service

In ViBRANT year 2 the KOS Architecture[56] has been further developed (M4.16). GBIF has prepared a draft charter proposing to form a new TDWG task group on Vocabulary Management (VoMaG)[57]. The VoMaG proposal focuses on tools for managing a new glossary of basic terms. The present tools to create Darwin Core extensions and vocabularies for the GBIF infrastructure does not include such a basic overview list of the terms that are defined and reused by the extensions and vocabularies.

The vocabulary management plan has been to explore three alternative software tools for managing basic terms:

1. isoCAT[58]: the platform seems very well designed from the perspective of functionality, work-flows and structure for maintaining a list of basic terms in a collaborative manner, but the isoCAT user interface is still not mature and in the present form not sufficiently user-friendly;

2. Semantic Media Wiki: Initially SpeciesID wiki[59] on the biowikifarm and KOS wiki[60] at GBIF were used for testing; based on these results http://terms.gbif.org has been created on the ViBRANT biowikifarm;

3. GBIF Vocabularies Service will be explored for maintaining basic terms, although the present feeling is that this tool will need too many modifications to be extended to cover this purpose. The Vocabulary Server is well suited for building Darwin Core extensions and vocabularies for use by the GBIF IPT, but less suitable for the functionality we have in mind for the glossary of terms.

---

55 GBIF Vocabularies Service: http://vocabularies.gbif.org
56 KOS Architecture: http://kos.gbif.org/
57 Vocabulary Management Task Group (VoMaG): http://community.gbif.org/pg/groups/21382/
58 IsoCAT software tool: http://kos.gbif.org/isocat/interface/
59 SpeciesID Wiki: http://species-id.net/wiki/
60 KOS wiki: http://kos.gbif.org/wiki

Progress has been made on developing a new information system for exploring terms and concepts declared by the relevant Knowledge Organisation System (KOS). Prototype solutions for providing a human-centric information and a discussion forum for biodiversity vocabulary terms can be found at http://terms.gbif.org/. Year 2 has focused on identifying the most appropriate web hosting implementation to ensure the long-term persistence of such a service. After a detailed review the biowikifarm [61] Web environment seems to be the most appropriate current solution. The instalment of a technical backbone implementation for the recommended and published machine-readable vocabulary resources (RDF vocabularies) is established on the existing GBIF Resources Repository [62].

The first prototype implementation for RDF vocabularies of terms has been demonstrated using the sandbox at the GBIF Resources Repository. The GBIF Resources Browser [63] includes a Glossary of Terms for the resources included at the GBIF Repository of terms and machine readable services more aimed at technology developers than the more human-centric wiki forum. A summary of the new Ontology Platform prototype is given below (Fig. 7).
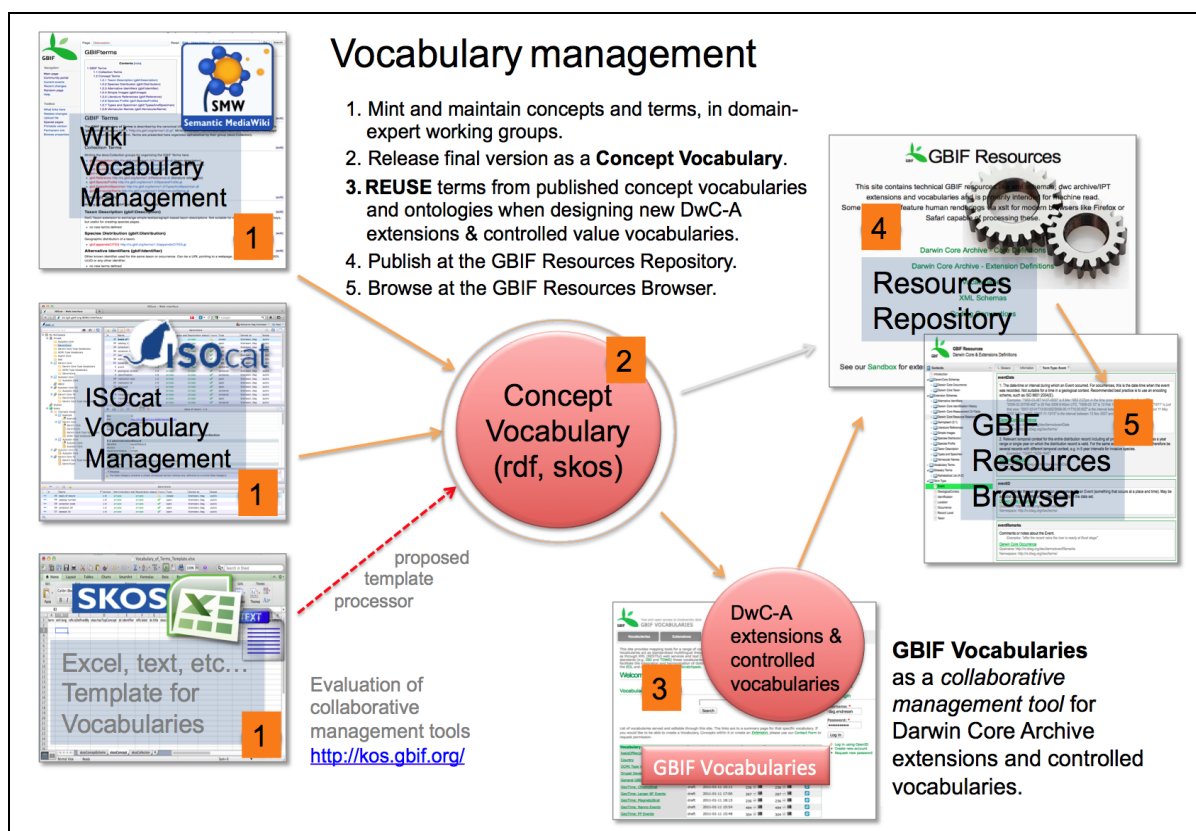
Figure 7 – Further scoping the KOS architecture

For biodiversity ontology management and evaluation two new tools have been prototyped, including a biodiversity "slice" at the NCBO BioPortal ontology repository [64] (M4.21) (Fig. 8) and the GBIF Term Browser [65] (M4.39), allowing users to browse for terms defined in widely used concept vocabularies such as Darwin Core, Dublin Core, and so on.

---

61   Biowikifarm: http://biowikifarm.net
62   GBIF Resources Repository: http://rs.gbif.org/terms
63   GBIF Resources Browser: http://tools.gbif.org/resource-browser/
64   NCBO BioPortal: http://bioportal.bioontology.org/ontologies/3058
65   GBIF Term Browser: http://kos.gbif.org/termbrowser
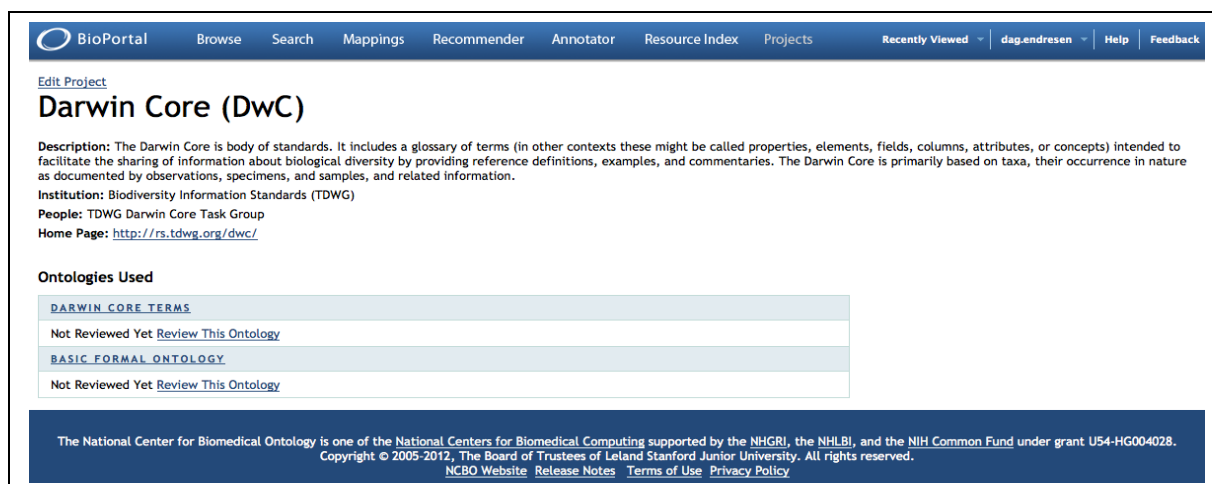
Figure 8 – NCBO BioPortal ontology repository

MediaWiki

To develop standard data interfaces and APIs, optimising a collaborative community interface to facilitate the (bottom-up) definition and management of ontologies, an instance of MediaWiki has been created [66]. In this context glossary-oriented potential contributors have been canvassed (M4.19) and as an example vocabulary the Audubon Core standard has been implemented on Mediawiki (Fig. 9), as part of the GBIF Terminology Platform (M4.27). The implementation of BiolFlor into the MediaWiki is awaiting the latest version of the data (M4.22).
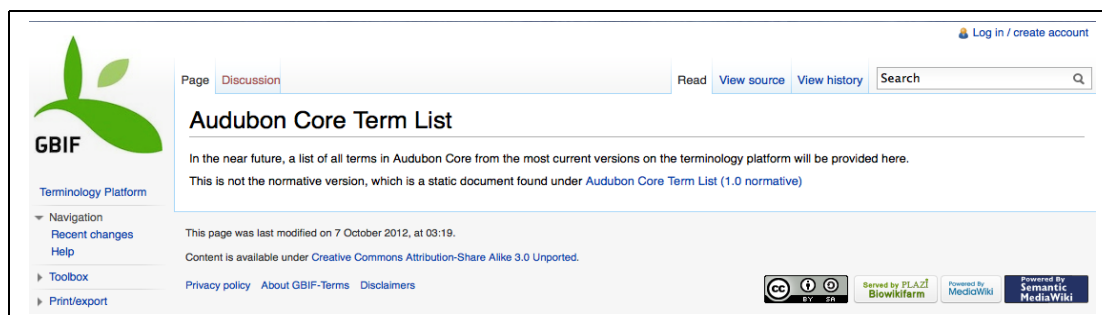
Figure 9 – Audubon Core as MediaWiki

The implementation of additional example vocabularies are scheduled (as part of canvassing) as well the integrating of the biowikifarm into ViBRANT authentication system.

Integration with other workpackages

The WP4 Ontology Platform activities are working together with WP2 on (taking over) hosting facilities by GBIF and on proper routines for Scratchpads consuming GBIF vocabularies, with WP5 (UPMC) on key development and with WP7 applying vocabularies to semantic entity extraction and automatic mark-up of documents; an effort which overlaps with the EU project agINFRA[67] (Grant No. 283770) and BioVel[68] (Grant No. 283359) with whom a joint meeting was held in Oxford at 24 February 2012. The function of the Media Wiki to provide a publication tool will be coordinated with WP6 (Pensoft).

---

66   http://kos.gbif.org/wiki/Main_Page
67   agINFRA: http://aginfra.eu
68   BioVel: http://www.biovel.eu

Future Work

The GBIF Resources Repository will be further developed, focussing on DwC-A tools and extensions. Tools to apply vocabularies will be developed as a mechanism to engage the broader user community, initially by extending some of the exemplar vocabularies that have already been created.

Task 2 Scratchpad common access point (UvA, BGBM)

ViBRANT will add internationalisation and localisation to the various platforms hosting ontologies and develop routines allowing a common access to the different platforms. This includes the Scratchpads platform, EDIT CDM platform, Xper2 platform, Pensoft publishing tools and Ontology Platform.

Progress in Year 2

This task was re-defined during the year, originally entitled "Development of taxonomic reference lists", the task is now rather broader. Fig. 10 shows the vision for interaction between various structural components.
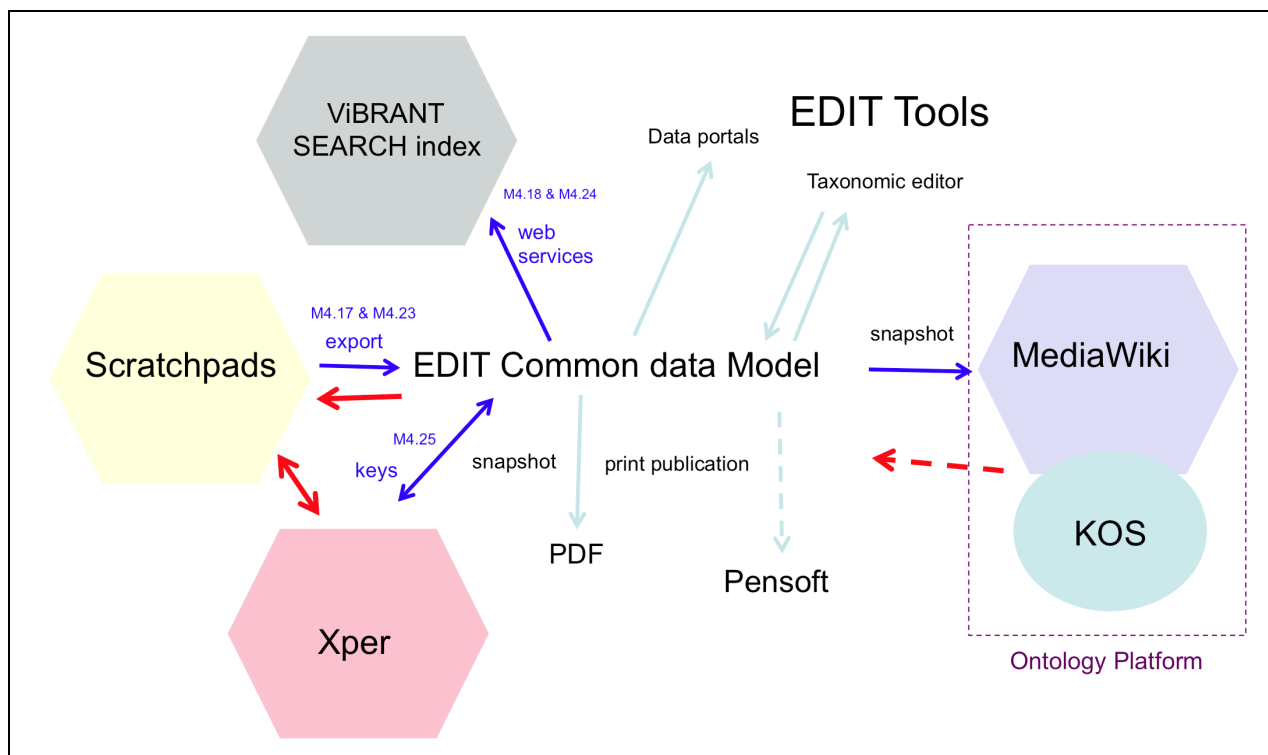
Figure 10 – Overview of Common Access Point of established in ViBRANT WP4 Y2 (dark blue) and anticipated

Scratchpads Export Module

The available DwC-A Scratchpads module to export data from the Scratchpads 2.0 database has been extended. A subset of the fields from DwC-A for each of the 5 extensions (reference, image, types and specimen, distribution and description) has been mapped to fields in Scratchpads 2.0 (Fig. 11). For this purpose views were created in Scratchpads 2.0 to specify the data required in each extension using the DwC-A module available from EDIT's subversion repository[69]. Once enabled the module provides an interface for the user to associate a view with a file name for export of data from the Scratchpads database into DwC-A format (M4.17). See also WP2.

---

69   EDIT Developer Tools: http://dev.e-taxonomy.eu/svn/trunk/drupal/7.x/modules/dwca_export

The mapping relied on the user specifying an appropriate view for the extension, with the correct field types in the correct order. Moreover some data consumers wishing to export data from Scratchpads are interested in different sub-sets of the data; simply providing a default set of extensions and fields to export is not sufficient. Therefore a flexible mapping solution was required and is available from the Scratchpads git repository[70] (M4.23).
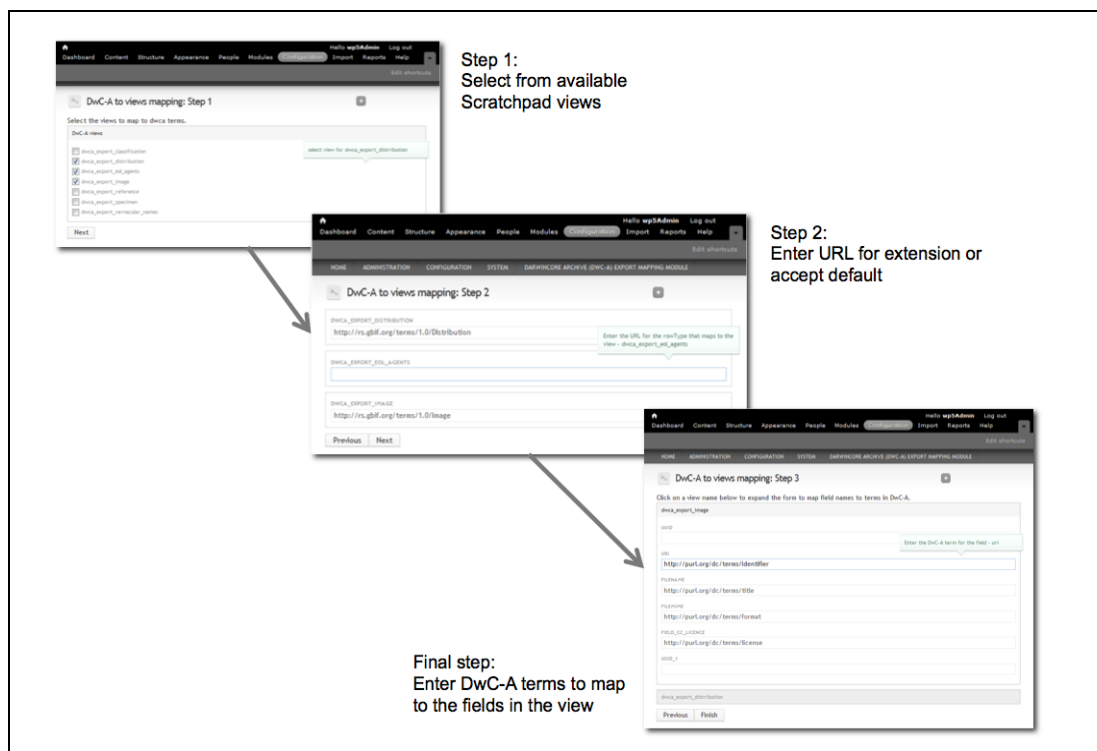
Figure 11 – Scratchpad to DwC-A Mapping and Export Module

ViBRANT Search Index
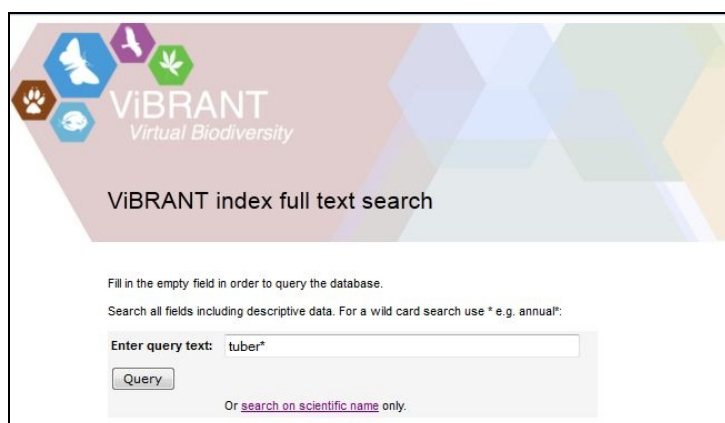
Figure 12 – Example querying the ViBRANT index for a full text search
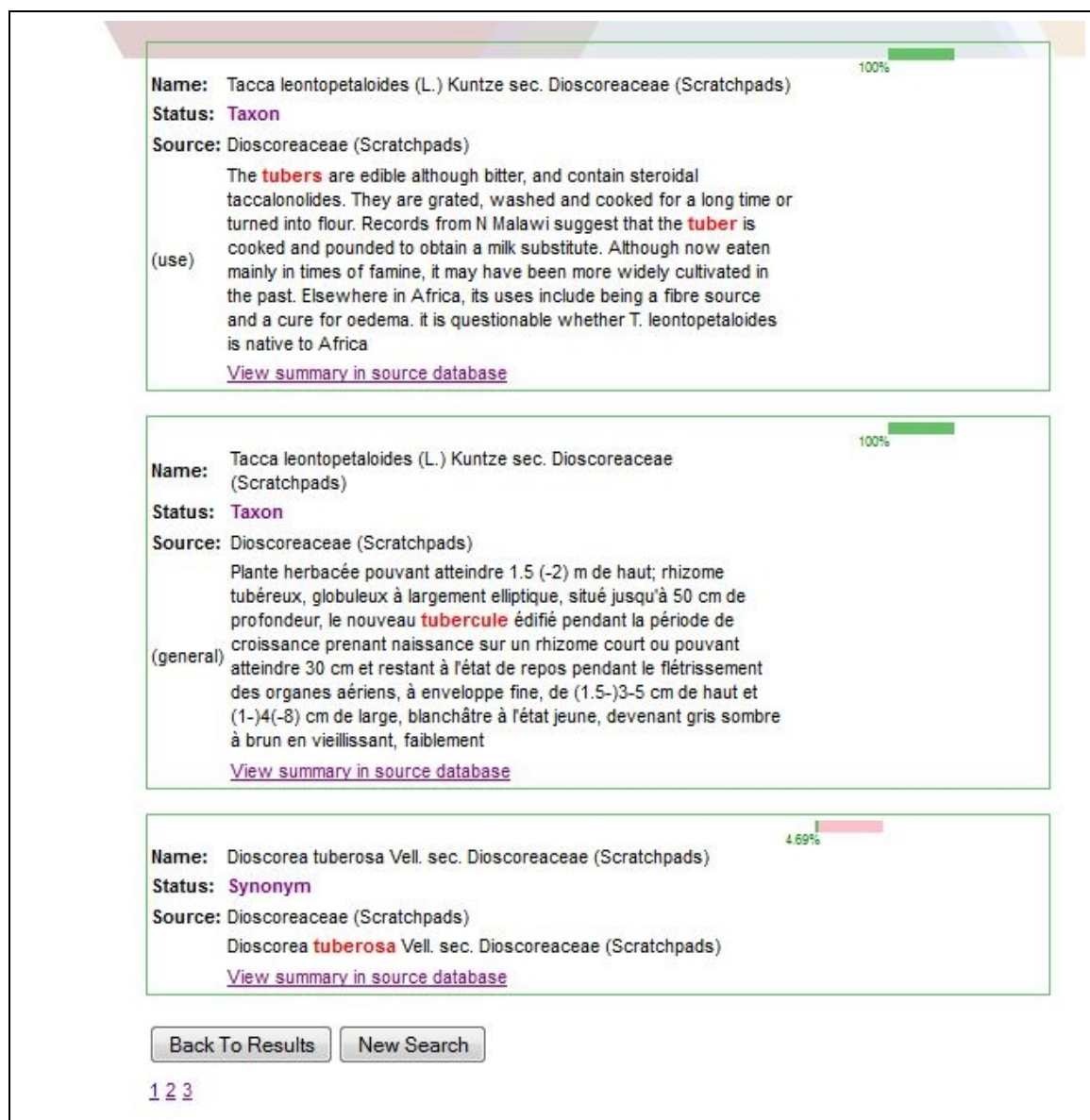
A CDM data store was implemented as a ViBRANT index to facilitate alternative views on ViBRANT data, especially for taxon names. For this purpose a simple web-based user interface has been created [71] to enable

---

70   DwC-A mapping module: https://git.scratchpads.eu/v/scratchpads-2.0.git
71   http://dev.e-taxonomy.eu/vibrant_index/search/

the querying of names of taxonomic data imported into the Common Data Model (CDM) from a variety of sources, providing a Human Interface for ViBRANT data in the CDM (M4.18) and a full text searching function (M4.24).

Web service methods were implemented in the CDM to enable full text searching using searches for the query string in the Taxon name and in the associated descriptive text. The XML generated from the web service query was transformed to generate web pages displaying the search results (Figs 12 & 13).

Figure 13 – Example showing the results returned from the search shown in Fig. 12

Taxonomic data from all available Scratchpads 2.0 sources will be imported into the CDM ViBRANT index database (using the Export Module). Further work is needed to create the Scratchpads URLs that link the Taxon back to the Scratchpads 2.0 source, requiring a term identifier for taxon centric and specimen centric data. Once this is in place we will be able to implement strategies for shared objects. Also the user interface will be extended enabling the user to filter the results e.g. on description type (Scratchpads description type data include for example: general description, distribution, conservation).

ViBRANT Single Access Key Service

Xper2 is a platform that utilises structured descriptive data to annotate taxonomic data, for the purpose of taxonomic identification developed by UPMC[72]. A web service is implemented for producing single access keys from the exchange format for structured descriptive data (SDD).

Figure 14 – Cichorieae test portal results showing the generated key generated in the CDM portal

To organise the integrated access of descriptive features across the different platforms, in the ViBRANT second year, a pilot implementation[73] of the web service into the EDIT platform was developed as part of a

---

72   Xper2 documentation and binaries: http://www.identificationkey.fr/index.php/aboutTheWebservice

73   Source code of pilot implementation: http://dev.e-taxonomy.eu/svn/branches/cdmlib/ikey-plus-import/

workshop and hackathon. For testing a TDWG-SDD file containing descriptive character of 300 Cichorieae taxa was exported from the CDM to Xper2 to generate a key and,to get the SDD data right into the CDM, transformed into a relevant CDM iKey-Plus single access key. The results have been visualised in the CDM data portal (see Fig. 14) (M4.25).

Standardised cross-platform integration of taxonomic data

The progress on implementing a cross-platform (CDM<->Scratchpad and CDM<->MediaWiki) pipeline will establish a functional taxonomic Clearing House, allowing Scratchpads, CDM and other users an integrated access to taxonomic resources and facilitating the production of regional checklists and taxonomic catalogues in co-ordination with the PESI project. Although not emphasised in ViBRANT year 2, this remains a key-objective, adding important internationalisation and localisation to the ViBRANT efforts.

Integration with other workpackages

This part of activities in ViBRANT WP4 is connected to WP2 on developing the Drupal export module for Scratchpads, to WP5 on linking Xper2 to the CDM platform and collaborates with WP8 on publishing Fauna Europaea as an e-Publication.

Future Work

The CDM SDD output for Xper2 will be optimised (M4.29) and the field mapping for the DwC-A module will be refined. The role of the ViBRANT Index will be clarified and marketed to increase its application within the community, particularly by working with other projects, such as BioVel and LifeWatch.

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 2

This task has been subsumed within Task 1, above.

WP5 - Interaction and services

Lead: Neil Caithness, Oxford e-Research Centre   Activity type: Service

Partner	Name	Acronym	Effort (PMs Year 2)
4	Oxford e-Research Centre	UOXF.E9	14.00
6	Julius Kühn-Institute	JKI	6.30
11	Vizzuality	VIZZ	4.05
13	Université Pierre et Marie Curie-Paris 6	UPMC	16.21
16	Universite de la Reunion	UdlR	1.57

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 2
OBOE is, as far as we are aware, a unique service in that it offers a drop-and-compute facility through a single API. The single API is the unique factor and is likely to be instrumental in opening access to a much wider range of potential users. This is in sharp contrast to GRID suppliers who are scrabbling for users to show their potential beyond particle physics, but who do not offer an interface that is easy to use, or in most cases, even access.

The two figures below show the growth in OBOE registered users from when the platform went live in March 2011 until November 2012, with a projection to the end of the project.
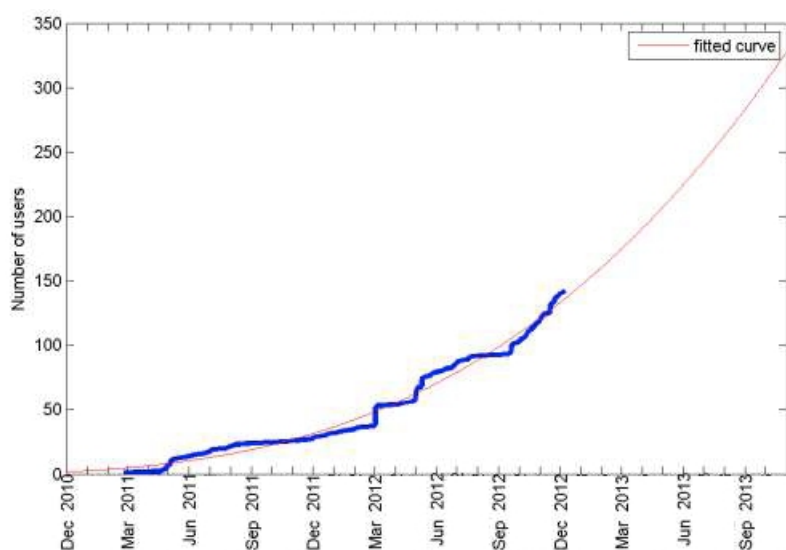
Figure 15 - Growth of OBOE userbase

There are currently 142 registered users and this is expected to more than double by December 2013. The number of job submissions has grown to 709 in November 2012, with an expectation that this will also double over the next year.

We see the steady growth in the OBOE user base as encouraging, especially as this phase falls prior to the main service delivery and dissemination activities of year three. We anticipate growth in users and job submissions significantly above the projections shown.
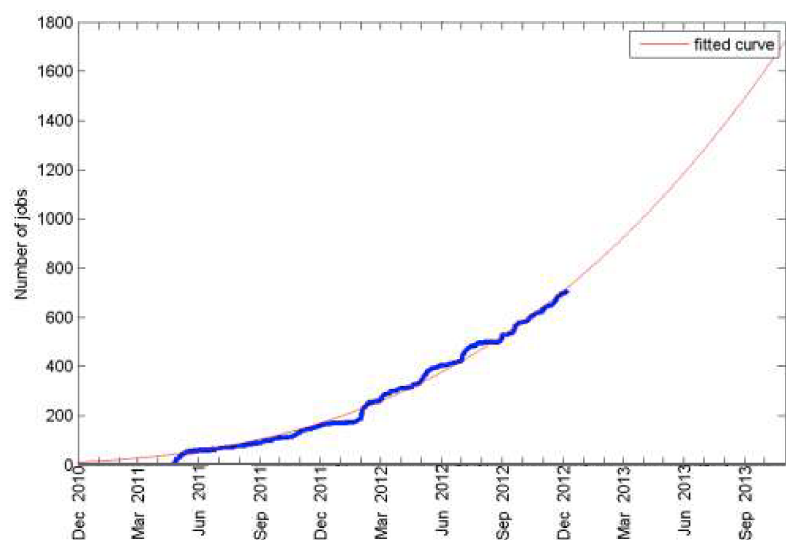
Figure 16 - Growth of OBOE job submissions

OBOE now supports the following tools:

- LEFT[74]: a paid-for service, the Local Ecological Footprinting Tool produces an ecological valuation report for a specified area. It is intended as a pre-planning tool to use before undertaking further field-based investigations and quickly highlights areas of high ecological value to avoid in the location of facilities with an environmentally damaging footprint

---

74   http://www.biodiversity.ox.ac.uk/LEFT

- LCC: Land Cover Change (see Task 10 below)

- BICT: Biological Index Calculation Tool (See WP8, page 70)

- Mr Bayes: a phylogenetic analysis tool (see Task 4 below)

- BEAST: a phylogenetic analysis tool (see Task 4 below)

- MUSCLE[75]: sequence alignment tool

- iKey+: creation of identification keys from SDD files (see Task 7 below)

- Durden[76]: Durden, created in WP2, will take a large image and create the tiles necessary for the image to be viewed using various client software such as the Bigimage Google map image viewer.

- GoldenGATE[77]: Golden Gate offers various web services to parse bibliographic references, dates, coordinates and quantities.

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 2
OBOE now supports services running on diverse platforms via simple, loosely coupled, middleware. Platforms supported include Windows and Linux operating systems, as well as the Oxford Supercomputing clusters.

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 2
The metadata repository is in place as an array of replicated MongoDB databases. This is an integrated part of the OBOE platform and provides complete metadata capture for job history and retrieval. This task is now complete.

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.

---

75   http://www.drive5.com/muscle/
76   https://git.scratchpads.eu/v/durden.git
77   http://idaho.ipd.uni-karlsruhe.de/GoldenGATE/

Progress in Year 2

Wrappers for two phylogenetics applications (BEAST[78] and Mr. Bayes[79]) have been implemented and are available via OBOE[80].

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 2

A review of target applications was completed and the following priorities were identified:

1. MaxEnt software

2. WorldClim environmental layers

3. GBIF locality data; or user supplied data.

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 2

The molecular alignment software Muscle[81] is now available via OBOE.

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 2

The Xper2[82] key generation service is available via OBOE.

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.

Progress in Year 2

Workpackage 5 has created, in a collaboration of Julius Kühn-Institute (JKI) and Université Pierre et Marie Curie-Paris 6 (UPMC), a user interface tools which combines the creation of matrix keys in Xper2 with the

---

78 http://beast.bio.ed.ac.uk/Main_Page
79 http://mrbayes.sourceforge.net/
80 https://oboe.oerc.ox.ac.uk/
81 http://www.drive5.com/muscle/
82 http://www.identificationkey.fr/

the ViBRANT webservice to create single-access keys (dichotomous or polytomous) keys and the wiki- and Android-based user interface of the ViBRANT biowikifarm.

The data can be curated in the java-based Xper interface, the single-access ViBRANT iKey+ webservice will automatically export tot he ViBRANT biowikifarm. The following examples show the appearance on the web in a web browser:
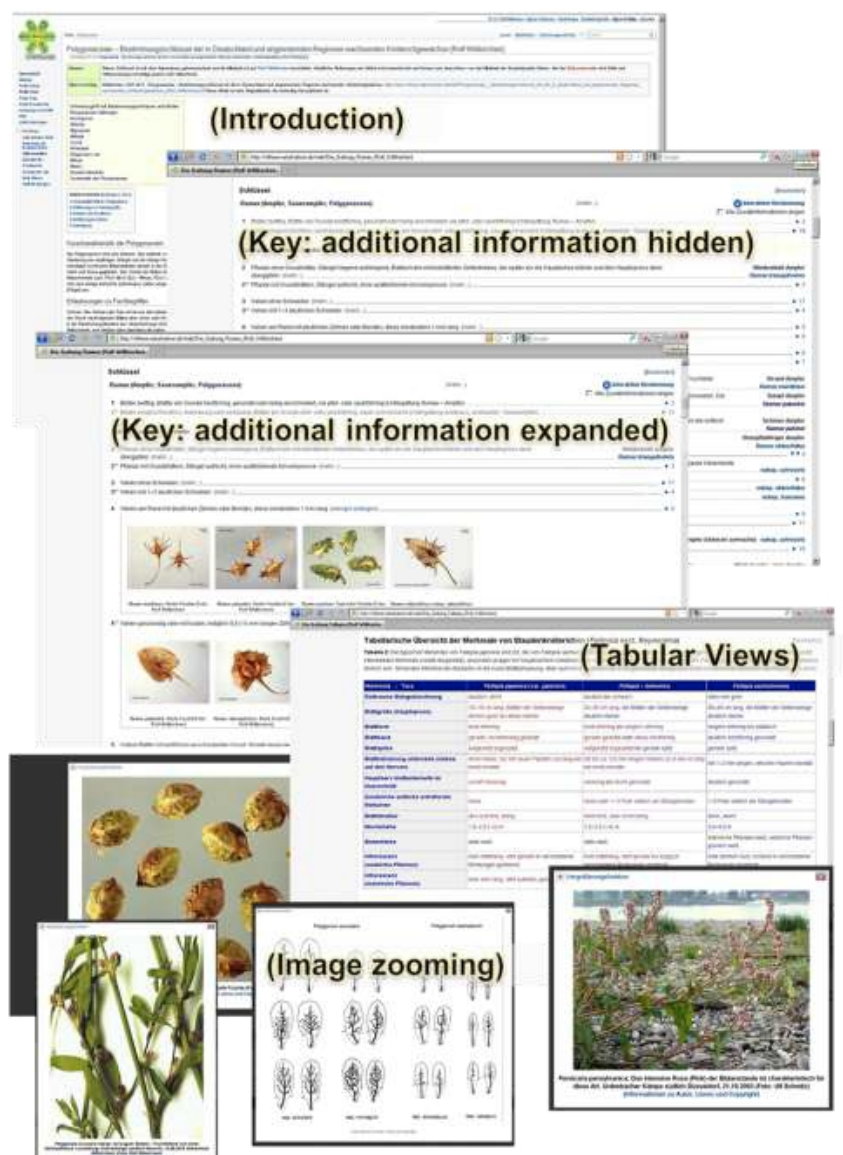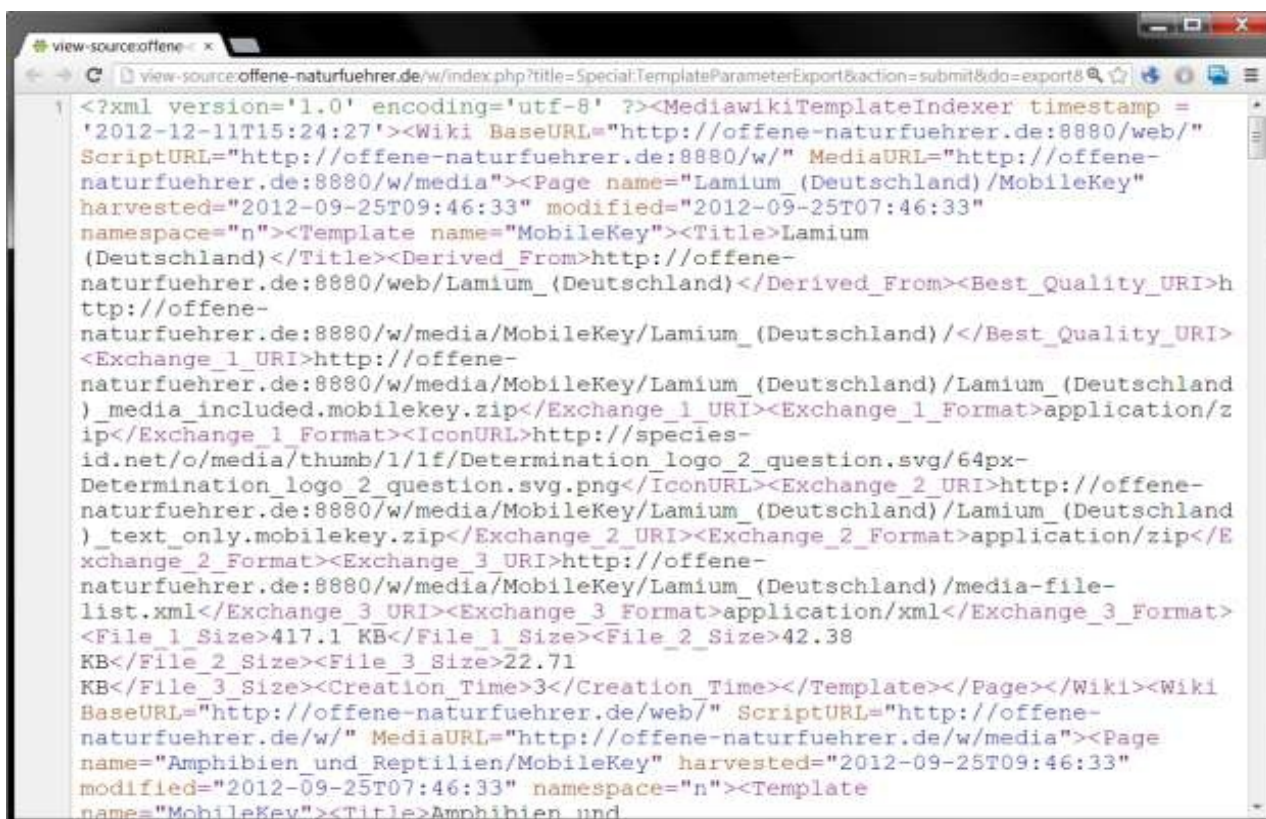

Figure 17 - information structure used in key building

The same information can be repurposed by the click of a button into a Smartphone-enabled format. A newly generated Android App accesses a webservice on the ViBRANT biowikifarm:

Figure 18 - the information can be rendered into XML for re-purposing
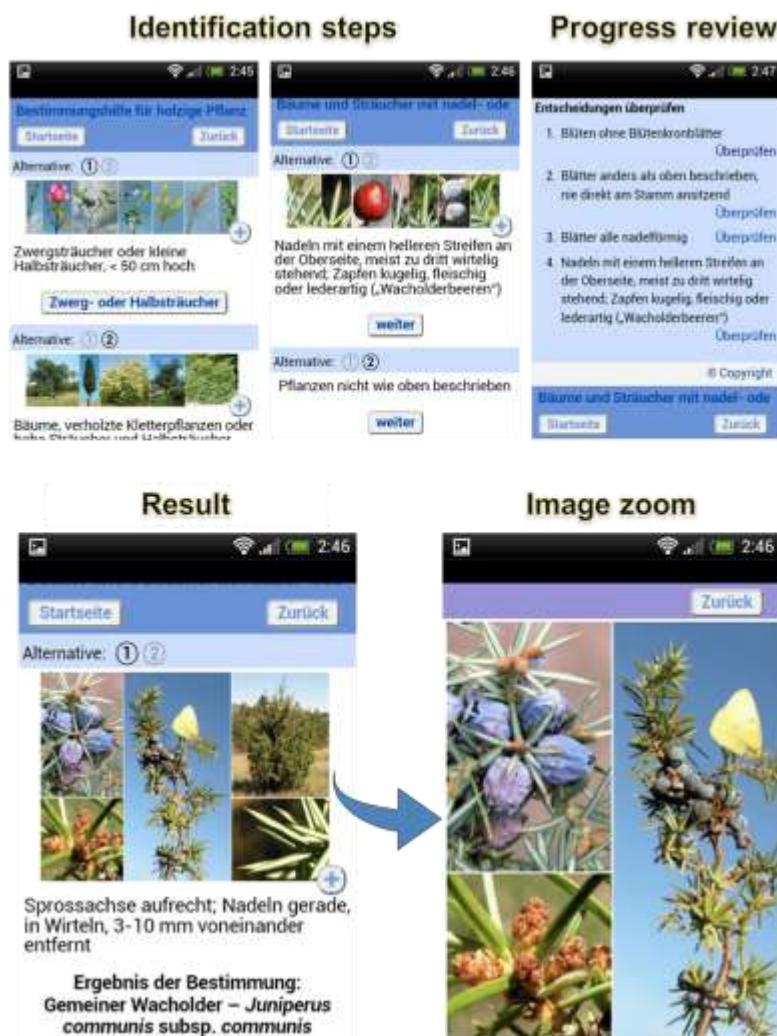
Figure 19 - the same data as Figs 17 & 18 presented in mobile-phone format

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 2

A suite of computational services for geo-spatial analysis has been developed for Scratchpads via the OBOE infrastructure[83].

The current focus of many (or even most) Scratchpads is on taxonomic groups – this was the original intended use. A new set of geo-spatial tools will facilitate an alternative focus on geographic regions – e.g. protected areas, game reserves and national parks. Equally, non-protected areas, agricultural lands, urban/rural interfaces, etc., may all be spatially demarcated areas of interest for the Scratchpad community of users.

---

83   https://oboe.oerc.ox.ac.uk/

These tools will provide a rich set of maps and other analytical graphics, as well as download links for the associated regional data subsets in standard formats for further analysis by the user.

A generic software framework and template for drop-and-compute services has been developed. This framework will provide for rapid and efficient development of the planned ecological informatics services detailed in milestones M5.34 M5.39

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 2

The GlobCover Land Cover project[84] has released two global land cover classifications for the periods December 2004 - June 2006, and January - December 2009. These are the highest resolution (300 metres) global land cover products ever produced. The Land Cover service will be available through OBOE and will deliver cover maps for the defined region and an analysis of the land cover change occurring in the region between 2005 and 2009 (Figure 20). An extended analysis will also provide an array of patch statistics using the FRAGSTATS analysis package[85].
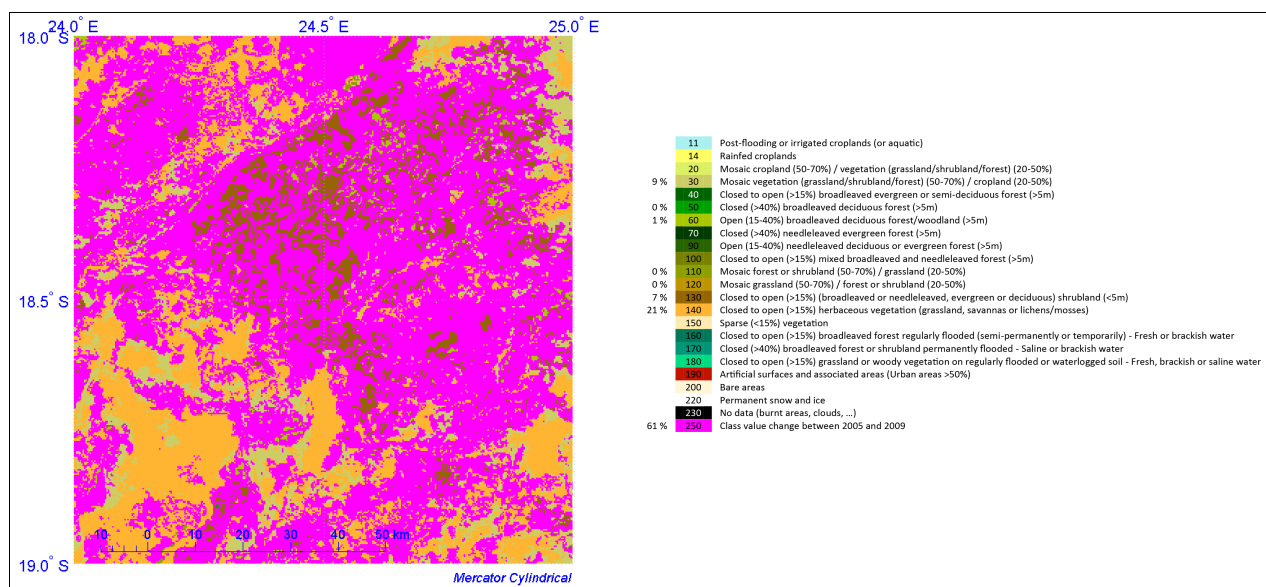
Figure 20 - Screen-shot from the LCC tool in OBOE, showing change/no change of a region of Chobe National Park, Botswana

---

84   http://due.esrin.esa.int/globcover/
85   http://www.umass.edu/landeco/research/fragstats/fragstats.html

WP6 - Scholarly publishing

Lead: Lyubomir Penev, Pensoft Publisher

Activity type: Service

Partner	Name	Acronym	Effort (PMs Year 2)
1	The Natural History Museum, London	NHM	1.32
12	PENSOFT Publisher	PENSOFT	12.00

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 2

Refinement of the XML schemas for submission of manuscript from Scratchpads to ZooKeys

The export of structured content in XML format from Scratchpads to ZooKeys has been exemplified by Faulwetter et al. in a ZooKeys paper published in 2011. The paper dealt with a newly described species of marine bristle worm found in the Mediterranean sea. The XML schema used in the manuscript was TaxPub, an extension to the Document Type Definitions (DTD) of the US National Library of Medicine Journal Archiving and Interchange Tag Suite (NLM). Since July 2009, TaxPub has been routinely implemented in the everyday publishing practice of Pensoft, to provide:

1. Semantically enhanced, domain-specific XML versions of articles for archiving in PubMedCentral (PMC);

2. Visualisation of taxon treatments on PMC;

3. Export of taxon treatments to various aggregators, such as Encyclopedia of Life, Plazi Treatment Repository, and the Wiki Species-ID.net.

The experience in implementation of TaxPub in Pensot's journals has been presented at the Journal Article Tag Suite Conference[86].

With the development of the Biodiversity Data Journal (BDJ) and Pensoft's Writing Tool (PWT), however, and after consultations with a number of leading bioinformaticians, a number of flaws in the use of TaxPub schema become apparent, especially in regard to the export of manuscripts from Scratchpads to ZooKeys. Primarily TaxPub was originally designed for taxonomic treatments and does not encompass the overall complex structure of the other types of research articles in biodiversity sciences. Besides simple taxonomic treatments several other types of manuscripts, such as data papers, software descriptions and checklists, are envisaged for acceptance and publication in BDJ, so the overall XML schema for export from Scratchpads to PWT through the PWT API were significantly expanded to reflect the specifics of each of the different publication templates (Fig X). In two subsequent workshops held in Sofia (Vibrant meeting, 16-18 April 2012) and Beijing (annual TDWG conference, 22-26 October 2012) a number of important decisions were taken in this connection and a new workflow and XML schema developed. The pilot test is expected to happen by the end of December 2012 when a new species of millipede will be described through the PWT and published in Biodiversity data Journal.

---

86   Penev, L., Catapano, T., Agosti, D., Georgiev, T., Sautter, G., Stoev, P. (2012) Implementation of TaxPub, an NLM DTD extension for domain-specific markup in taxonomy, from the experience of a biodiversity publisher. In: Journal Article Tag Suite Conference (JATS-Con) Proceedings 2012 [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2012. Available from: http://www.ncbi.nlm.nih.gov/books/NBK100351/
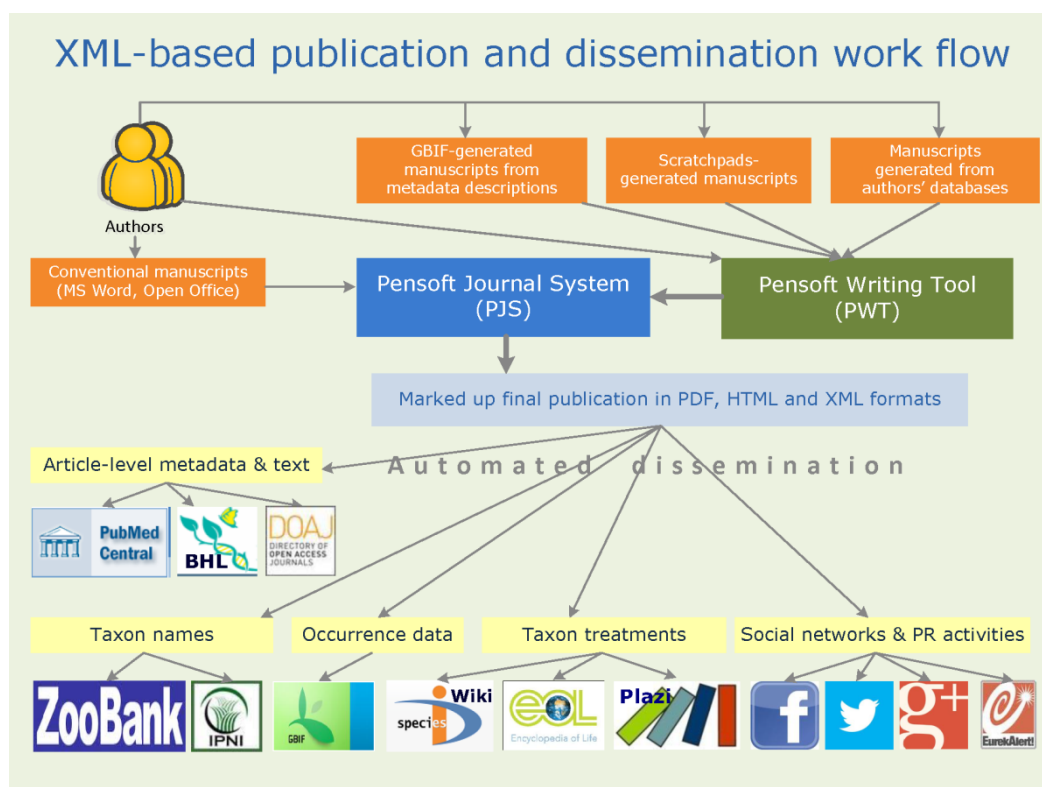
Figure 21 - A chart showing the XML work flow from Scratchpads and GBIF to PWT and integration of the published content into major data aggregators.

Converting a word-processed floristic checklist into Darwin Core Archive format

Data mining and converting texts to structured data, especially of historical biodiversity literature, is a major challenge in biodiversity informatics. Methods and tools developed to provide such conversions ideally should allow extraction of data from texts of both legacy literature and manuscripts submitted for publication, facilitating the incorporation data elements into globally accessible databases, thus enhancing discoverability. Remsen et al.[87] described a pilot project to convert a conventional floristic checklist, written in a standard word processing program, into structured data in the Darwin Core Archive format. A conventionally written Checklist of vascular plants of the Department of Ñeembucú, Paraguay [88], consisting of more than 4,100 taxon names, was submitted to PhytoKeys as a Microsoft Word file. After peer-review and editorial acceptance, the final revised version was converted into Darwin Core Archive format from the original manuscript and published both as a conventional paper in PhytoKeys and as DwC-A structured data through the Global Biodiversity Information Facility (GBIF) Integrated Publishing Toolkit (IPT). In addition and for convenience of the readers and data users, the same data are also published as a supplementary Excel file in an Appendix to the checklist[89]. The data are available through the GBIF infrastructure and can be re-used on their own or collated with other data.

---

87   Remsen D, Knapp S, Georgiev T, Stoev P, Penev L (2012) From text to structured data: Converting a wordprocessed floristic checklist into Darwin Core Archive format. PhytoKeys 9: 1–13. doi: 10.3897/phytokeys.9.2770

88   Egea, J. D., Peña-Chocarro, M., Espada, C. & Knapp, S. (2012). Checklist of vascular plants of the Department of Ñeembucú, Paraguay. PhytoKeys 15–179. doi: 10.3897/phytokeys.9.2279

89   doi: 10.3897/phytokeys.9.2279.app1

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 2

The automated registration of nomenclatural acts and the quality control of the bibliographic metadata in registries (IPNI, ZooBank, Index Fungorum, MycoBank) should, ideally, be primarily the responsibility of publishers and registry curators, and to a lesser extent of authors. Currently no existing biodiversity publisher has tried to implement such a workflow, and this was one of the main tasks undertaken by Pensoft within WP6, namely to develop a workflow and technical infrastructure that allows automatic registration of nomenclature acts in global taxon indices. In November 2012 Pensoft successfully tested the pre-publication export of metadata associated with the description of a new plant species, subsequently published in PhytoKeys[90] though the IPNI's API. This was made possible after negotiation with the IPNI IP team to clarify the structure of the input XML and the overall workflow. The model follows three-step process:

Step 1.   Upon acceptance of the manuscript, a publisher-based tool sends an XML query to the registry containing the type of act, the taxon names, and preliminary bibliographic metadata of the article (title, authors, and journal).

Step. 2.   The registry server sends back an XML report containing the unique identifier (e.g., LSID, PURL) of the act; the LSIDs are included and appear in the final publication.

Step 3.   On the day of publication, the journal sends an automated XML report to the registry that contains the exact bibliographic details of the published article (e.g., authors, title, journal, issue no, date of publication, pagination, etc.) to be completed for the specific act.

Next step will be to make this workflow a routine practice in all Pensoft's journals, and implement it also in other global organismic indexers, namely ZooBank and MycoBank. The pilot will be presented at Symposium 17 'Nomenclature for the Future' at the Global systematic conference[91] February 2013 in Vienna.

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.

---

90   Maas-van de Kamer, H., Maas, P. J. M. & Specht, C. (2012). Costus loangensis, an exciting new species from Gabon, Africa. PhytoKeys 18, 11–18. doi: 10.3897/phytokeys.18.3713
91   http://biosysteu.univie.ac.at/symposia-workshops/

4. 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 WP4 Task 2). For this purpose ViBRANT will extent 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 2

Three online user surveys were carried out in 2012 by Pensoft in the frame of the project ViBRANT aiming at receiving feedback from:

1. ZooKeys' authors and reviewers on the editorial workflow and the existing journal services;

2. bioinformaticians on the interface and functionalities of the Pensoft Writing Tool (PWT);

3. scientific (mainly taxonomic) community of their interest in participating in the Editorial process of Biodiversity Data Journal.

The results of the three surveys have been analysed in ViBRANT report D6.2 [92]. Based on the analysis and suggestions received, the authoring and editorial processes of Pensoft's new editorial platform, and the scope and general publication workflow of the BDJ have been significantly refined. In addition to several purely technical suggestions for improvement of e.g. the PWT Taxon treatments and Comments interface, some brand new features, such as Community and Public review processes were introduced. The scope of the BDJ and the general workflow, from generation of manuscripts in the PWT, submission and publication in the Pensoft Journal System (PJS), and the dissemination of the published BDJ content have been elaborated. Guidelines for authors, reviewers and editors have been prepared, explaining in detail the steps of manuscript creation, processing and the criteria for evaluation.

On 22 October Pensoft announced through several channels (Taxacom, ZooKeys, PhytoKeys and MycoKeys mailing lists) the preparation for launching the BDJ [93], emphasising it place as a next generation platform for publishing biodiversity science and data. BDJ is currently recruiting editorial specialists who can be involved in the innovative reviewing process and have expertise in dealing with taxon treatments, checklists, genomic-, ecological- and environmental-datasets, analytical methods and software. Potential editors should be interested in new ideas, methods and approaches to publishing, sharing and using biodiversity information. An online registration form [94] has been developed to facilitate of the process.

The announcement immediately attracted the interest of the world's biological community. In only 4 weeks 194 scientists from many countries have expressed their interest in becoming editors of BDJ. They show a wide range of expertise, from taxonomy (156), to phylogeny, evolution, genetics, biogeography. Insect specialists have shown highest interest in becoming Subject Editors, 83 (43%) of respondents have an entomological background. The second largest group of experts are botanists, mycologists and algologists, which amount for about 15% of all respondents. Relatively high interest was shown also by herpetologists (15) and Crustacea experts (14) (Fig. X). Among entomologists, beetle specialists dominate with 25 respondents, followed by Hymenoptera (7), Trichoptera (6) and Lepidoptera (5) experts.

92   http://vbrant.eu/sites/vbrant.eu/files/D6-2-Review & refine services.docx
93   http://biodiversitydatajournal.com/
94   http://www.pensoft.net/editor_form.html
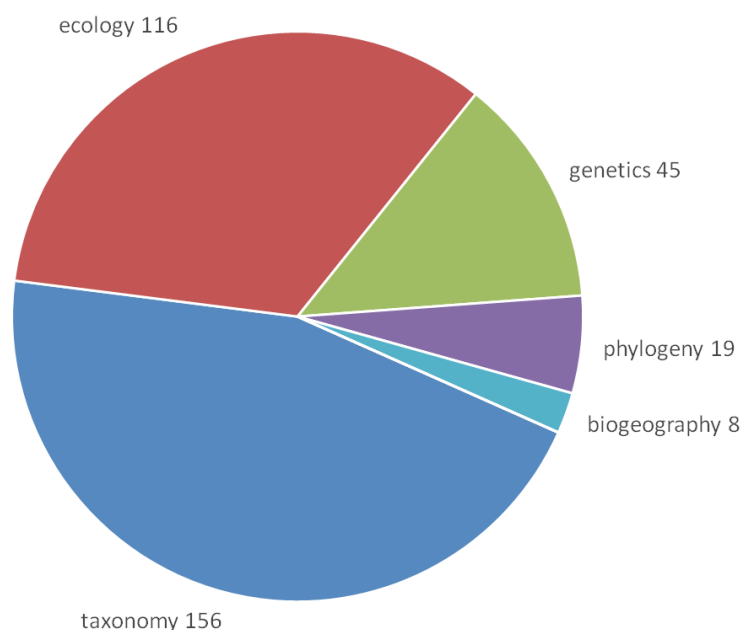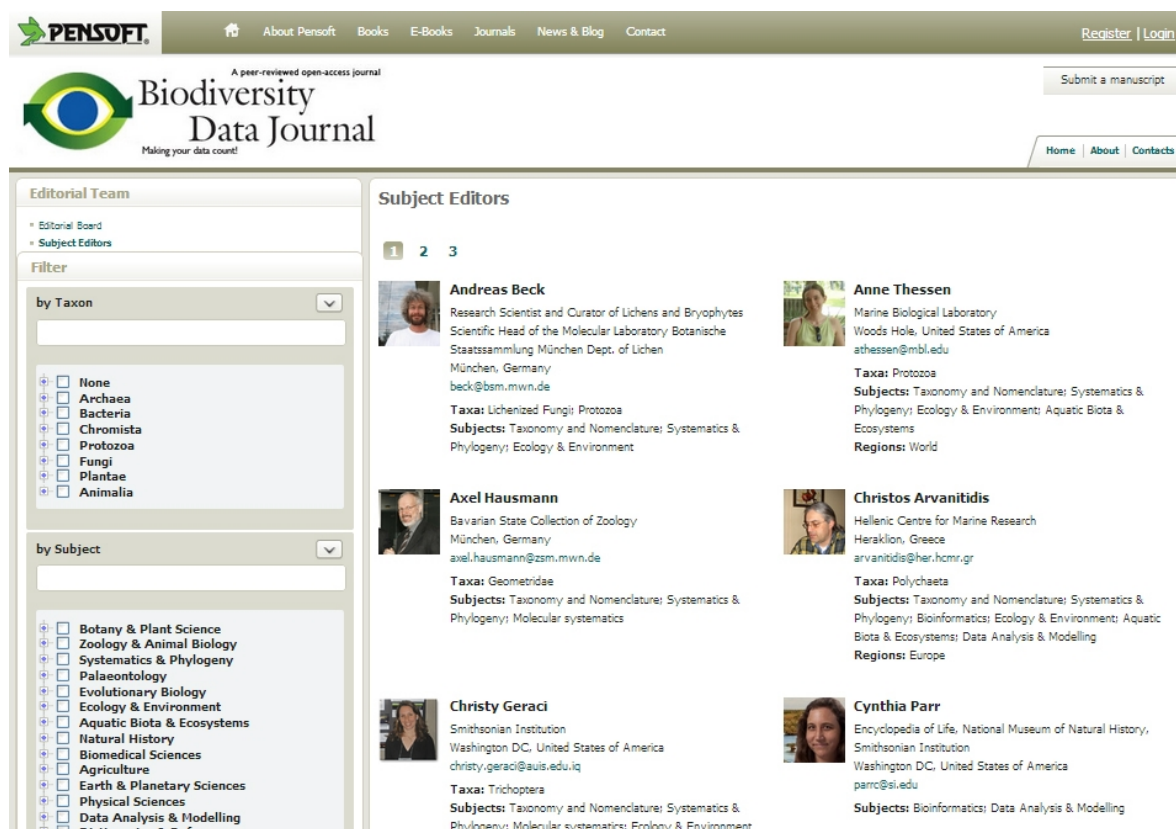
Figure 22 - A pie chart showing the background of respondents

Figure 23 - A screenshot of Subject Editor's page of the Biodiversity Data Journal website

Progress towards development of Pensoft Writing Tool (PWT) and Biodiversity Data Journal (BDJ)
A main effort of Pensoft's IT team in 2012 was directed towards development of the PWT and the BDJ. The PWT is a collaborative tool for online article authoring; it provides templates for different kinds of biodiversity articles, with upfront markup, links to external resources, and various options for data publishing. The tool is designed to solve one of the main difficulties with the implementation of TaxPub, namely to mark texts that have already been structured by the authors in text editors (usually MS Word or OpenOffice). Authors using the PWT have at their disposal a set of pre-defined, yet flexible, templates that they will fill in through sophisticated editing software. Taxon treatments are a core element in the PWT. Different types of taxon treatments, such as (re-)descriptions, nomenclatural acts (new synonymies, re-validation of names, designations of type specimens, etc.) are modelled in accordance with the slightly different requirements of the Biological Codes, e.g., for animals, plants, and fungi (Figs X, X).
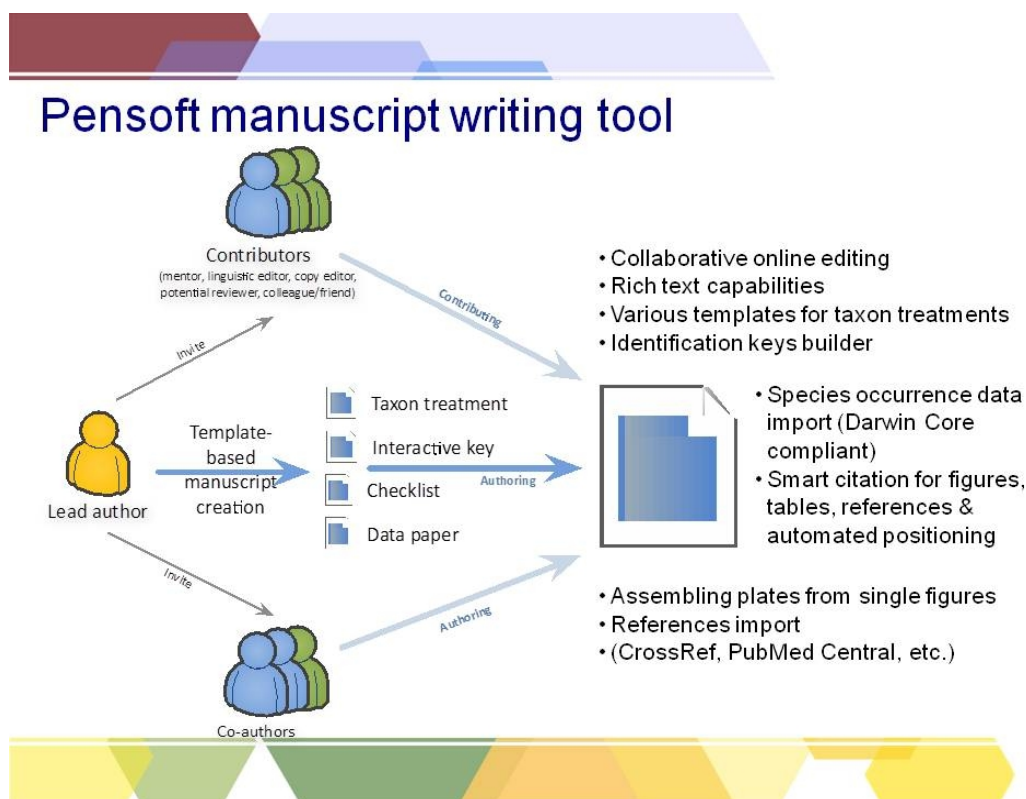
Figure 24 - A generalized workflow of the collaborative article authoring and editing software (Pensoft Writing Tool, PWT)

Figure 25 - Screenshots of the collaborative article authoring and editing software (Pensoft Writing Tool, PWT)

In addition, PWT will provide:

- a collaborative environment for authors to create and work on an online document (manuscript);

- authors may invite additional contributors (e.g., mentors, potential reviewers, linguistic and copy editors) to watch, comment, and edit the text during the writing process;

- e-mail and chat communication tools within the group of co-authors and contributors associated with a manuscript;

- automated import of data-structured manuscripts generated in various platforms (e.g., Scratchpads, authors' databases);

- revision history, version control, and version comparison;

- various modes of data publishing (supplementary files, multimedia, import of data tables, linking to external data repositories, etc.), in accordance with internationally accepted standards (e.g., species occurrence data in Darwin Core);

- semantic markup of text and data during the writing process, with no additional effort for the authors;

- rich-text editing, smart management of citation and placement for references/figures/tables;

- import of references from external bibliographic databases (CrossRef, PubMed and others);

- pre-submission validation of the manuscript.

The PWT will serve as gateway for the Biodiversity Data Journal (http://biodiversitydatajournal.com/) and also for other journals in the future. BDJ is the first journal ever to complete the cycle from writing a manuscript, through submission, community peer-review and editing, to publication and dissemination within a single, fully XML-based, online collaborative platform, the Pensoft Journal System (PJS). The publication in the PJS environment is intended to be very low-cost, and this is largely achieved by properly structured submission and thus minimising handling by human editors.

WP7 - Biodiversity literature access and data mining

Lead: David Morse, The Open University   Activity type: Research

Partner	Name	Acronym	Effort (PMs Year 2)
1	The Natural History Museum, London	NHM	0.14
9	The Open University	OU	11.45
10	Karlsruher Institut für Technologie	KIT	18.85
12	PENSOFT Publisher	PENSOFT	3.00

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 2

Bibliography of Life: RefBank + REFinder Workflow
Substantial progress has been made this year extending the capabilities of RefBank to meet the needs of ViBRANT's users. REFinder is a stand-alone application, developed to help [biodiversity] scientists to search for, quickly find and download bibliographic references.

RefBank functionality
At the time of last year's report there some 4,000 references in RefBank. At its public launch, at this year's TDWG conference[95] in Beijing (22–26 October 2012), RefBank contained approximately 160,000 references. Guido Sautter attended TDWG and gave a presentation[96] and demonstrations of RefBank. In the fortnight following TDWG, users added a further 6,000 references.

To enhance the utility of RefBank to its users we have delivered many new features in Year Two. We have extended the RefBank import routines to support widely used bibliographic formats, eg BibTeX and RIS [97]. This makes it easier for users to load their existing references if they have used reference manager software. It also makes our work as developers easier because we can now harvest existing online resources, and if they offer an export facility in one of these formats, we can directly upload the data without having to re-parse it. Similarly, we have added a bulk upload facility both to enable easier contribution of references by individual users and to facilitate automatic harvesting and loading of references from other sources [98].

A significant enhancement has been to incorporate RefParse[99] fully into RefBank[100]. This feature allows users to load plain text bibliographic references into RefBank and for the references to be automatically parsed by the system. Guido Sautter presented the development of his work underlying this enhancement at TPDL 2012[101], with the paper Sautter, G., and K. Böhm. 'Improved Bibliographic Reference Parsing Based on Repeated Patterns'. *Theory and Practice of Digital Libraries* (2012): 370–382[102]. This enhancement provides a valuable service because, from anecdotal reviews with working taxonomists, we are aware that many users do not use bibliographic management software, preferring to maintain lists in Word files, or other authoring software of their choice.

Also this year we worked with ViBRANT partner Pensoft[103], to automatically import references to, and contained in, newly published Pensoft journals[104]. Hence, month-by-month, the number of references in RefBank continues to grow, and these are quality, verified references to currently relevant articles. In addition, we have made minor usability changes to the web-based user interface, and some other minor changes to the back-end code to increase robustness of the system.

---

95 http://www.tdwg.org/conference2012/
96 http://vbrant.eu/content/towards-universal-bibliography-%E2%80%93-refbank-approach
97 http://www.wiki.scratchpads.eu/w/M723report
98 http://www.wiki.scratchpads.eu/w/M724report
99 Gupta, D., Morris, B., Catapano, T. and Sautter, G., (2009), 'A new approach towards bibliographic reference identification, parsing and inline citation matching', *Contemporary Computing: Communications in Computer and Information Science*, **40**(2), 93-102
100 http://www.wiki.scratchpads.eu/w/M725report
101 http://www.tpdl2012.org/
102 http://vbrant.eu/sites/vbrant.eu/files/RefParseTPDL2.ppt
103 http://www.pensoft.net/
104 http://www.wiki.scratchpads.eu/w/M722report
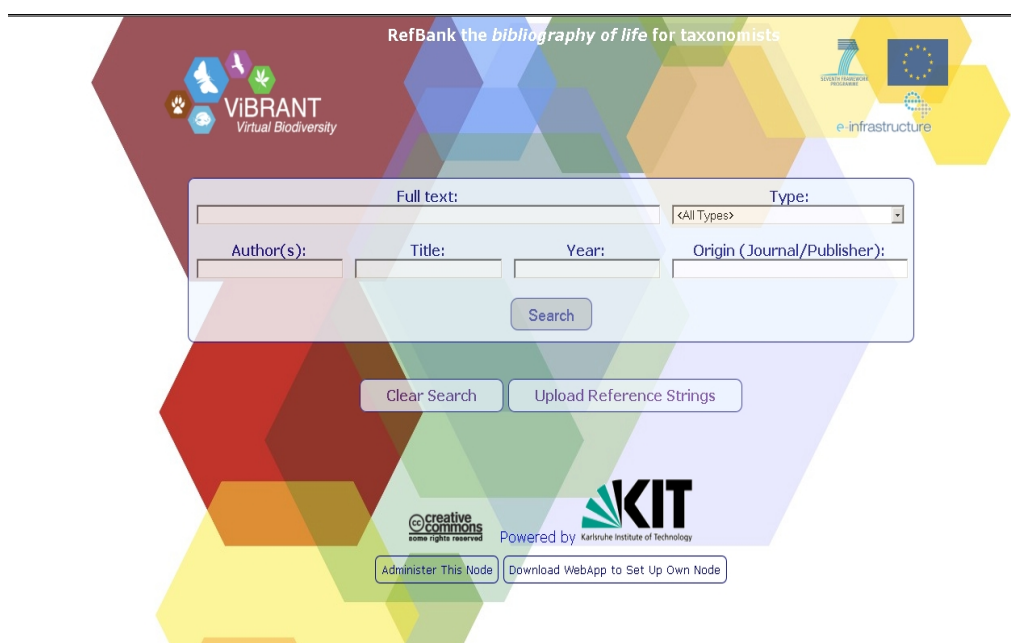
Hence we have gone from a tool that at the time of last year's report could only accept references as plain strings or formatted using RefBank's extension to the Metadata Object Description Schema (MODS) [105], and could not support download of references at all, to one that supports the full range of data formats that are widely used for storing and exchanging bibliographic information [106]. We can import and export these formatted references in bulk, which makes both the individual and bulk automatic populating of RefBank far easier, as well as data extraction should we need to re-purpose the collected data.

Support for plain text export, for those users who prefer not to work with a reference manager, has been delivered as part of *M7.23 Extend RefBank import routines to support other widely used bibliographic formats[107]*, eg BibTeX, RIS, etc..

For future use a record of the contributor was added to RefBank's data [108]. This enables us to track who contributed what, which will be invaluable when extending the tool to citizen scientists because we can then reward good contributors, acknowledging their contributions and granting their additions a positive status when they are used to reconcile conflicts, as well as to block contributors of poor quality or spurious data. A second related data field was added at the same time to track who parsed and re-parsed a reference. This serves a similar function for quality control allowing us to encourage good quality contributions from citizen scientists.

RefBank Nodes

A key feature of RefBank is replication. The ability to replicate content across RefBank instances builds in recoverability to the service and also exposes the service to more potential users whose data, wherever uploaded, will contribute to the bibliography. Last year we had two servers, three if you count the one that was internal to the OU. We now have four servers, having added one full server at the OU and one at Pensoft, and CiteBank have expressed an interest in hosting a RefBank instance.

Figure 26 - Screenshot of RefBank's front screen. The current state of RefBank content and functionality can be seen in the KIT hosted RefBank instance at http://vbrant.ipd.uka.de/RefBank/search.

---

105 MODS, http://www.loc.gov/standards/mods/;
   RefBank extensions, http://www.wiki.scratchpads.eu/w/WP7RefBankXML
106 http://www.wiki.scratchpads.eu/w/M710report
107 http://www.wiki.scratchpads.eu/w/M723report
108 M7.21 Add metadata to cover origin of bibliographies, http://www.wiki.scratchpads.eu/w/M721report

Aggregators: several aggregators have been developed, each tailored to the output of a specific website, emphasising the need for a tool such as the Bibliography of Life to provide a single point of access to bibliographic information. The early aggregators also had to rework the extracted references into RefBank's extended MODS format for loading. This year's enhancements to RefBank, especially M7.23 [109] and M7.24[110], render the reworking unnecessary because, having extracted references in a commonly used format such as BibTeX or EndNote, they can be loaded directly into RefBank, which saves considerably on development time.

De-duplication software: has been built based on graph theory to identify centroids of clusters. Each reference is rendered as a plain string, which is used to generate a MD5 checksum that forms a node in the graph. This MD5 checksum is also used as the unique persistent identifier for the record within RefBank. We use this because of the unresolved debate within the biodiversity community as to which identifier they should adopt. While we can contribute to the debate, in our opinion the Digital Object Identifier system (DOI)[111] is the preferred solution, we cannot impose a solution.

Linkages: being progressed in conjunction with the data mining task, for example to assist identification of features such as resolving pagination.

Scratchpads access: in progress, with integration with Scratchpads 2.0 being developed.

CiteBank: owing to the change in CiteBank's approach to references in non-Biodiversity Heritage Library hosted literature we have established contact with its new Director and are actively exploring methods of data interchange[112].

Test cases: were completed and reported last year, with the identification of example test cases being presented. This year, additional test data has been generated to directly test RefBank's reconciler software. This is not a general test suite, but specifically to test features of the software.

Training and standards: all external links to and from RefBank permit formats already in use by digital libraries and software-based reference managers to be used. Appropriate training materials will follow completion of the linkage and integration tasks.

Additional developments

An additional benefit of gathering bibliographic data into one source is that it enables further work on the references. This work is particularly useful for de-duplication and also allows us to add value in other ways. For example, it is possible to list all the authors (or journals, publishers, etc) recorded within RefBank by a simple RESTful query: http://vbrant.ipd.uka.de/RefBank/data?type=persons. These lists will facilitate the development of catalogues of synonyms.

There also is the potential, to be investigated in Year Three, to exploit these orthogonal views of RefBank's records with a view to enhance authority and name reference sources. In particular we will investigate the potential of ORCID[113] identifiers.

Alternatives to RefBank

We have kept abreast of developments elsewhere in bibliographic management. In particular we have closely followed developments led by the Open Knowledge Foundation [114] and their Bibserver[115] project and

---

109 M7.23 Extend RefBank import routines to support other widely used bibliographic formats, eg BibTeX, RIS, etc, http://www.wiki.scratchpads.eu/w/M723report

110 M7.24 Upload service for complete bibliographies, http://www.wiki.scratchpads.eu/w/M724report

111 http://www.doi.org/

112 We may even be able to profitably revisit the work on automatically populating CiteBank from Scratchpads that was begun and abandoned in Year One owing to CiteBank's then change of stance to restrict its range of references.

113 http://about.orcid.org/

114 http://okfn.org/

115 http://bibserver.org/

associated lightweight interchange format, BibJSON[116]. David King attended two hackathons[117] associated with this work, and was able to introduce the concepts behind RefBank to a wider audience.

We have also been in discussion with a linked open data project at the OU, Co-Reference Open Access Repositories (CORE)[118] and are setting up automated data interchange. There is already some overlap in that CORE harvests the Biodiversity Heritage Library. CORE offers several potentially interesting features including a similarity detection algorithm, which could be incorporated into our work to suggest similar references beyond those that explicitly match the search terms entered by a user when looking for references.

REFinder Workflow

In summary, a new service called REFinder is to be developed by WP6 that will act as both a search service and a front-end to RefBank. Work on this revised approach began following Management Committee discussions and was revised at a mini-workshop in August[119] led by Dauvit King of WP7.

Aims:

1. Assist authors in finding references in a few external trusted databases, known to be content-rich sources for biodiversity literature references, especially for the "historical" literature; the databases to be queried are currently RefBank, CrossRef, PubMed, Mendeley, CiteBank, Scratchpads Biblio Module, Pensoft's reference database (see below for a detailed list)

2. Create and display a dynamic webpage, called Reference Profile (RP); The RP will be displayed in a well-designed interface and will allow users to download references from there; depending on the level of markup at the original source of reference, the RP will provide options to download.

Functionality

1. To search for a literature reference through dedicated online resources (online bibliographies, indexers, reference collections) by specific content: DOI, PubMed ID, author(s), year of publication, title (all parts of), journal name and volume, and possibly also through a full text search in the reference;

2. The RP will NOT MAINTAIN its own database of records; rather it will provide the interface for searching through existing databases;

3. In the light of the above manual input of reference metadata seems to be superfluous.

Existing APIs that can be used are: CrossRef[120]; JSTOR[121]; PubMed[122]; & CiNii[123].

Depending on the external services available, the RP should be able to request and provide back information from those resources, such as:

1. Link(s) to online resource(s) containing information on the article, including (1) metadata only, through (2) metadata and abstract and, ideally, (3) full text;

---

116 http://bibserver.org/about/bibjson/

117 Bibliohack, 13-14 June 2012, Queen Mary, London, UK – notes: http://vbrant.eu/content/notes-bibliohack-2012-0. Future Citations, 27-28 September 2012, University of Aston, Birmingham, UK – notes: http://vbrant.eu/content/vibrant-future-citations-notes

118 http://core-project.kmi.open.ac.uk/

119 http://vbrant.eu/content/bibliography-life-workflow-mini-workshop

120 http://crossref.org, which offers a variety of services as explained on that page, though it is necessary first to register for a free account at http://www.crossref.org/requestaccount/

121 http://dfr.jstor.org/

122 http://www.ncbi.nlm.nih.gov/pubmed/, which offers several services such as single citation matching an online version of which is available at http://www.ncbi.nlm.nih.gov/pubmed/citmatch, though access to all services require an account such as a free NCBI account or a Google account

123 http://ci.nii.ac.jp/info/en/api/api_outline.html describes the API services available

2. Bibliographic metadata to be downloaded, possibly in a few leading bibliographic styles (e.g., Chicago, PLOS, Nature, etc.).
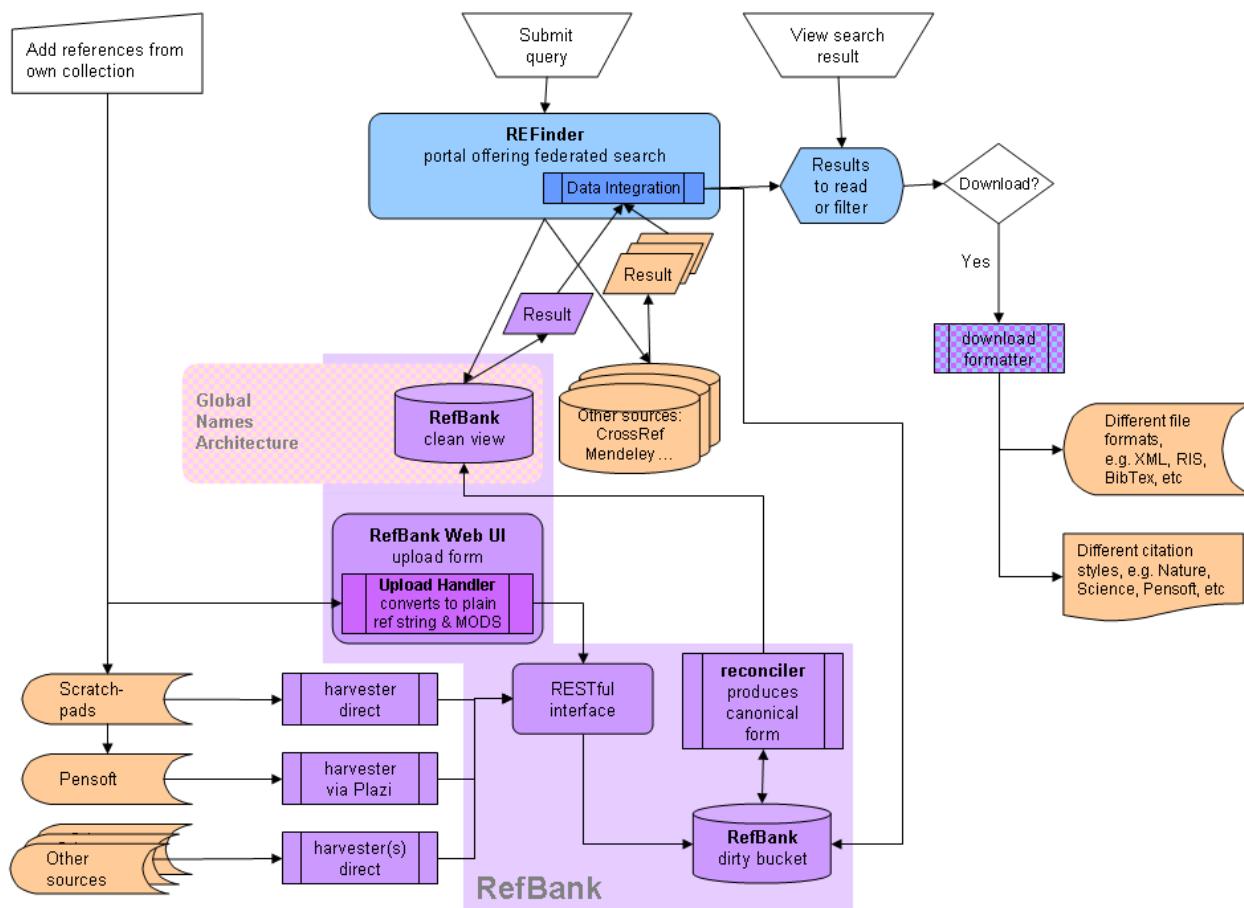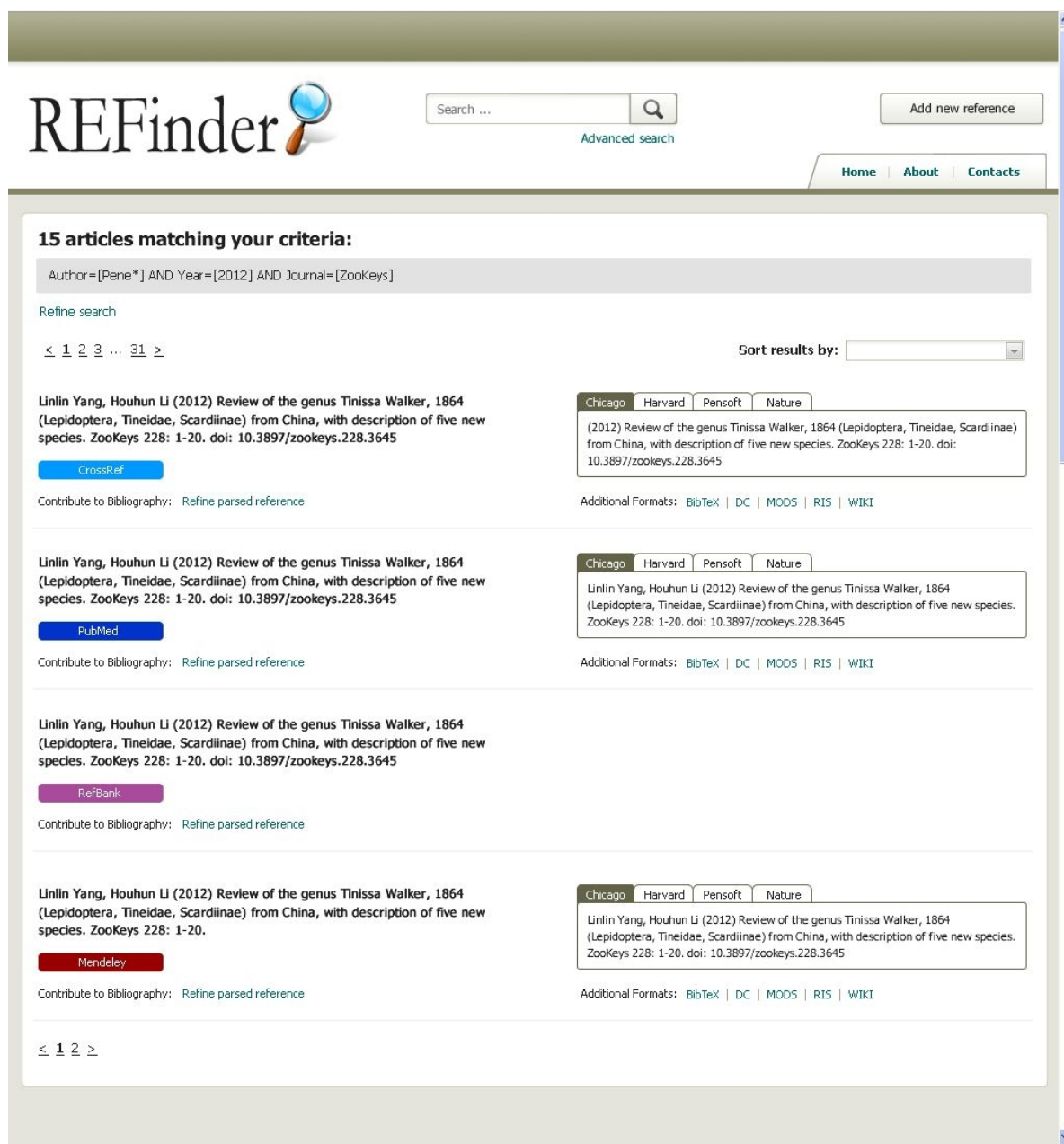
Figure 27 - A chart showing the general concept of Bibliography of Life

Figure 28 - Suggested interface of the Bibliography of Life portal

Future work

In the year ahead we expect our users to continue to contribute to RefBank. In addition, the automatic addition of references from Pensoft will continue, the full integration and automatic load of references from Scratchpads will be implemented, and we will extend our aggregating services to cover more biodiversity resources.

We will 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. Resolved in workshop 22 November 2012 [124] and in development now for delivery this calendar year.

---

124 http://vbrant.eu/content/notes-london-workshop-22-november-2012

A "parse / re-parse this reference" mechanism, will be displayed in the search result page. Demonstrated at this year's TDWG and implemented on the KIT RefBank instance (on searching select 'Refine Parsed Reference'). Awaiting deployment to other RefBank instances.

We plan to expose the content to Mendeley as part of a larger strategic decision for the ViBRANT project.

Our focus will return to data mining in Year 3 and, as part of this, the development of a reference extraction tool that will integrate the reference parser built in to RefBank with the reference discovery engine that is also being developed by WP7. Thus users can extract references from papers hosted by BHL and other online repositories.

For sustainability we plan to port the system to PHP. We also intend to implement a different database engine, possibly in 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. Outline planning and initial guidelines were prepared following Technical Workshop in Sofia (April 2012) and currently on hold pending larger review of the role of RefBank within ViBRANT and the development of REFinder.

We will investigate the use of Apache SOLR for indexing, should performance become an issue.

We will investigate exposing the content as RDF triples for data linking, as anticipated in agINFRA, in joint workshops being organised for Spring 2013. Representatives from outside the two projects with an interest in sharing biodiversity data will also be invited, eg from the BioVeL [125] project.

We will continue to refine and populate RefBank. However, we are all too aware that other initiatives are under way and will liaise with these to ensure that ViBRANT remains committed to delivering the most appropriate bibliographic solution to its users.

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 though 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 2

Having established that GoldenGATE could not integrate easily with the new infrastructure, the primary achievement of Year Two is the reworking of Golden GATE functionality into discrete web services [126]. These are available as standalone RESTful calls and through WP5's OBOE service. Through the application of the

---

125 http://www.biovel.eu/

126 M7.16 – Mark-up modules delivering outline mark-up, http://vbrant.eu/sites/vbrant.eu/files/M716.odt

individual GoldenGATE web services we will be able to deliver all elements of this task. Year Three will focus on enhancing the use of these services following successful completion of the pilot exercise [127].
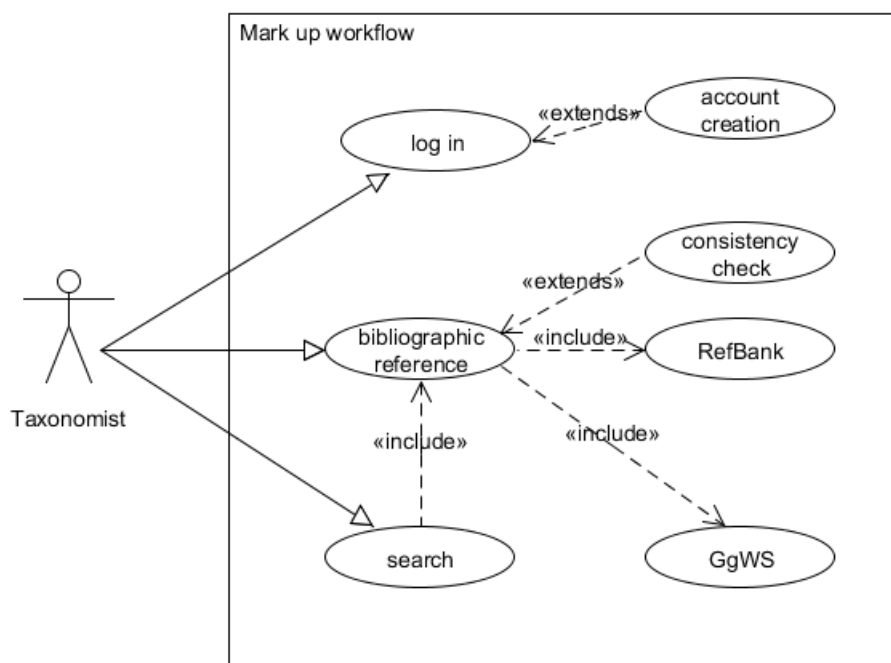
Figure 29 - Proposed mark-up workflow

The envisaged workflow can be described as follows: a working taxonomist wishes to apply mark-up to an existing text:

1. Taxonomist logs in to system, with account creation on first use.

2. In a web form:

    1. enter bibliographic reference data and a URL pointing to PDF or attach a PDF.

    2. enter some details about a document then search for it to complete the web form, then same process as above.

3. Consistency check for reference data – if inconsistent or incomplete taxonomist will be prompted for correction.

4. Reference, after passing consistency check, is automatically passed to RefBank. Hence, we grow RefBank at the same time as marking up text.

5. Thereafter is GoldenGATE Web Services in interactive mode. See footnote 116 for available services.

The work to progress task 2 is managed through four milestones – M7.11 and M7.13 delivered in Year One, and M7.16 and M7.17 mentioned earlier in this report – and brought together in Deliverable 7.2, which have been successfully completed to schedule. With the technology in place, such as RefBank and GoldenGATE web services, we can look forward to exposing it to the wider Scratchpads community in Year Three in collaboration with WP3 for training and outreach and WP8 for integration with other ViBRANT developed workflows.

---

127 M7.17 – Review of pilot mark up processes within the Scratchpad infrastructure, http://vbrant.eu/content/m717-review-pilot-mark-processes-within-scratchpad-infrastructure

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 2

In parallel with the use of GoldenGATE, we have investigated other annotation tools, and are currently making extensive use of the BioNLP rapid annotation tool (brat)[128]. brat is especially useful for work with the stand-off style of mark-up that is used in text mining (a sub-category of natural language processing, hence NLP) in preference to the XML favoured by the biodiversity community. brat's mark-up[129] originates with the BioNLP Shared Task[130] challenge. This file formats[131] is suitable to ViBRANT's needs.

This means we can go from XML like this:

```
<pb id="BCA-aves-v3p1-p325" n="[325]"/>In Guatemala it is common in
the forest-country north of Coban, where we obtained specimens in
1862, and whence others have been forwarded to us by native
collectors.
</p></div><div type="discussion"><p elementid="BCA-aves-v3p1-3599">A
Costa-Rican specimen of this species was described by Cassin as <hi
rend="italic">A. rufodorsalis
</hi>; but this seems to be nothing but a stained example of the
well-known bird. Salvin was of this opinion when he examined the type
in 1874<ref target="BCA-aves-v3p1-t445-taxft-17"><hi
rend="superscript">17
</hi></ref>. No similar specimens have since been obtained.
</p></div><div type="discussion"><p elementid="BCA-aves-v3p1-3600">As
already stated, <hi rend="italic">A. aurantiirostris
</hi> is the sole representative of the genus in Central America,
beyond the limits of which it is not found. It belongs to the same
group as <hi rend="italic">A. silens
</hi> of Guiana and Brazil, and has its nearest ally in <hi
rend="italic">A. spectabilis
</hi> of Colombia and Ecuador, from which it differs in having a
darker back and a broader black pectoral band.
</p></div></div></div><div type="taxon genus" id="BCA-aves-v3p1-t446"
n="BCA-aves-v3p1-t363" next="BCA-aves-v3p1-t452" prev="BCA-aves-v3p1-
t444" rend="implicit"><head elementid="BCA-aves-v3p1-3601">SALTATOR.
</head><div type="taxon synonymy"><p elementid="BCA-aves-v3p1-
3602"><hi rend="genus"><hi rend="italic">Saltator
</hi></hi>, <bibl rend="primary"><author>Vieillot
```

128 http://brat.nlplab.org/index.html
129 http://brat.nlplab.org/standoff.html
130 https://sites.google.com/site/bionlpst/
131 http://2011.bionlp-st.org/home/file-formats
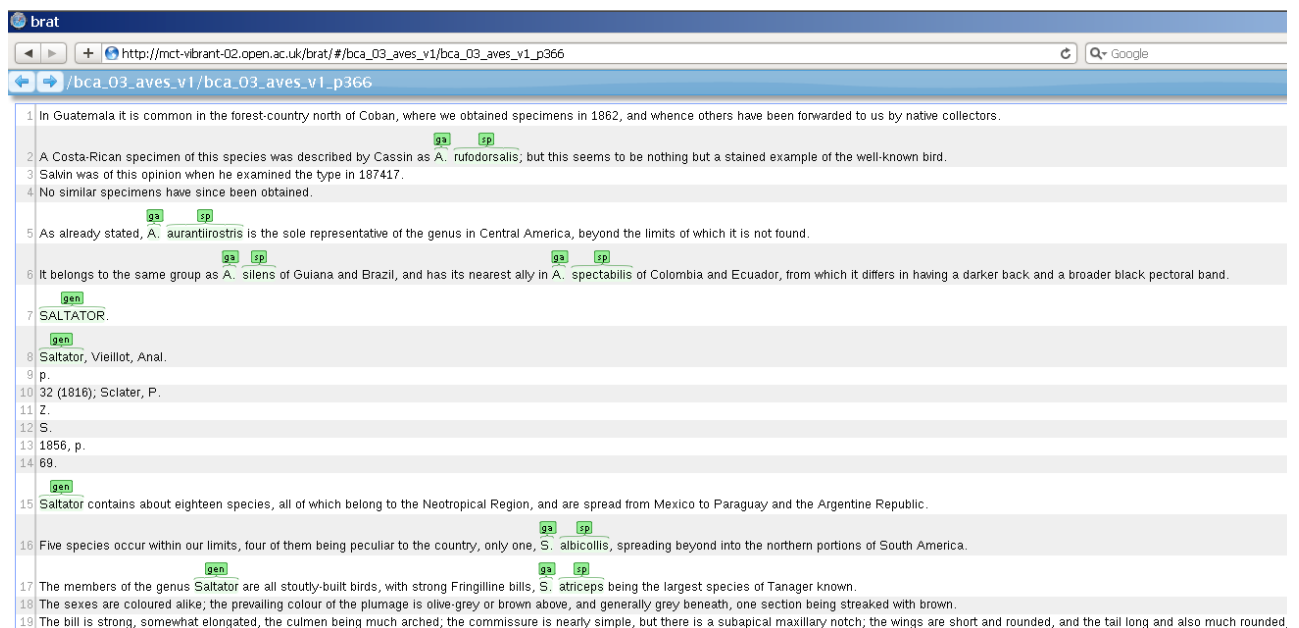
```
</author>, <title>Anal.
</title> <biblScope>p. 32
</biblScope> (<date>1816
</date>)
</bibl>; <bibl rend="secondary"><author>Sclater
</author>, <title>P. Z. S.
</title> <date>1856
</date>, <biblScope>p. 69
</biblScope></bibl>.
</p></div><div type="discussion"><p elementid="BCA-aves-v3p1-
3603"><hi rend="italic">Saltator
</hi> contains about eighteen species, all of which belong to the
Neotropical Region, and are spread from Mexico to Paraguay and the
Argentine Republic. Five species occur within our limits, four of
them being peculiar to the country, only one, <hi rend="italic">S.
albicollis
</hi>, spreading beyond into the northern portions of South America.
</p></div><div type="discussion"><p elementid="BCA-aves-v3p1-
3604">The members of the genus <hi rend="italic">Saltator
</hi> are all stoutly-built birds, with strong Fringilline bills, <hi
rend="italic">S. atriceps
</hi> being the largest species of Tanager known. The sexes are
coloured alike; the prevailing colour of the plumage is olive-grey or
brown above, and generally grey beneath, one section being streaked
with brown. The bill is strong, somewhat elongated, the culmen being
much arched; the commissure is nearly simple, but there is a
subapical maxillary notch; the wings are short and rounded, and the
tail long and also much rounded;
```

To a more human-friendly visualisation of the same content in brat like this:

Figure 30 - Visualisation of mark-up in brat

We have developed several mark-up conversion tools during Year Two, to facilitate interchange among the various XML formats used in biodiversity and the stand-off formats used by text mining practitioners. These will be fully written up and published during Year Three, with one repository already having been created in readiness to make the software freely available[132].

Future work

We will continue to evaluate and then to develop and to apply either GoldenGATE and/or other tools to meet the objective of refining automated mark-up.

Tasks 2 and 3 test data

A critical factor for the success of tasks 2 and 3 is adequate test data. This is especially important when using machine learning techniques, which require gold standard data to learn from as well as accurate test data. The absence of this data is major hindrance to all work in applying text mining techniques to biodiversity literature. While tools have been developed there is no matching corpus to evaluate them. GoldenGATE was used to produce the Madagascar corpus of some 2,500 page that is available on request via Plazi[133]. However, this is marked up in the TaxonX schema, which focuses on taxon treatments and is not suitable for a more general evaluation of new tools. The current most accurate automatic named entity recognition tool, NetiNeti[134] has one test document[135] of 172,656 words, which though large only covers American seashells and hence is limited in scope.

We attempted to use the data from the published output[136] from a PhD candidate researching the issue of taxon name recognition. However, we could not repeat the reported results using the publicly available data[137]. After some weeks' effort cleaning the data and in discussion with the data owners we decided not to continue with this work, but to develop our own corpus.

We have utilised two data sources.

- One is the mark-up of born-digital texts. For this we are using documents from Pensoft, converting them to text mining style stand-off mark-up as described earlier in this section. As such, because we are developing a process not a fixed body of texts, we can freely extend the corpus and ensure we cover the breadth of literature encompassed by biodiversity writings.

- Our other data source has been donated by the INOTAXA project[138] for which we are grateful. The complete 65 volumes of the Biologia Centrali-Americana (BCA)[139] have been re-keyed. This is an exceptionally valuable resource because it covers a wide variety of zoology data for which we have OCR textual content. Hence, we can develop tools to address the peculiarities of historic biodiversity literature, such as the interleaving of Latin and vernacular text, and how these peculiarities are further complicated by recognition errors introduced as part of the OCR[140] process, because we can compare the OCR text with the clean re-keyed text. Comparing clean and dirty versions of text to build training and testing data builds on previous work in the ABLE project[141], which involved several of ViBRANT's contributors from the NHM and the OU. We are enhancing tools developed in ABLE to progress this work, and will make the tools available through ViBRANT.

---

132 https://git.scratchpads.eu/v/taxpub2brat.git

133 http://plazi.org/

134 https://github.com/mbl-cli/NetiNeti

135 http://ubio.org/netinetifiles/

136 Wei, Qin, P. Bryan Heidorn, and Chris Freeland. 'Name Matters: Taxonomic Name Recognition (TNR) in Biodiversity Heritage Library (BHL)', in: *Proceedings of iConference 2010*, http://hdl.handle.net/2142/14919

137 http://bhlnameevaluation.wikispaces.com/

138 http://www.inotaxa.org/jsp/index.jsp

139 http://blog.biodiversitylibrary.org/2012/09/biologia-centrali-americana-hispanic.html

140 Optical Character Recognition – using a machine to convert text images into individual text characters

141 http://able.myspecies.info/

In the first phase of our work with the BCA we are creating a 10 volume, 4,442 page,1.9M word corpus, which will transform research in this area. If time permits we will extend the corpus.

In Year Three we will use the corpus in a workshop, with attendees from the EU's agINFRA [142] and BioVeL[143] projects and other interested parties to explore the ability to extract significant biodiversity entities and relationships from unstructured text.

Tasks 2 and 3 mobilising data

Having successfully extracted semantic metadata there remains the issue of exploiting this information. When ViBRANT was first proposed we envisaged using existing XML schemas and associated workflows to share the extracted semantic metadata. In particular, we planned to use TaxonX in Plazi and taXMLit in INOTAXA. This topic was the subject of much research during Year One [144]. However, neither of the XML schemas nor their workflows have grown beyond their original audience to engage the wider biodiversity research community. Hence, we risked expending considerable effort on building solutions that would not be used. To address this issue, during the past year we began investigating the use of linked open data (LOD) to share information, building on initiatives within biodiversity such as the TDWG RDF group [145] founded in October 2011, the extensive use[146] of linked open data in the agINFRA project[147], (an EU funded project to build an infrastructure for agricultural sciences very similar to what ViBRANT is doing for biodiversity science) and in which the OU is a partner, as well as existing experience within the OU. In a workshop at the NHM on Friday 23 November 2012, we agreed that LOD should be the emphasis for data mobilisation during Year Three. The outcome of that review echoes similar changes in emphasis seen in other work packages, particularly WP4 Standards. WP7 will work closely with WP4 in Year Three to utilise the controlled vocabularies and ontologies developed by that work package to mobilise extracted semantic metadata.

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 2
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.

---

142 http://aginfra.eu/
143 http://www.biovel.eu/
144 http://vbrant.eu/content/m710-agreement-standard-format-community-contributed-bibliographies-conjunction-wp4 and King, David, David Morse, Alistair Willis, and Anton Dil. "Towards the bibliography of life." ZooKeys **150** (2011): 151-166. doi:10.3897/zookeys.150.2167
145 http://code.google.com/p/tdwg-rdf/
146 Agricultural science has many years experience with LOD and has tools such as AgroVoc, http://aims.fao.org/standards/agrovoc/about, which we will investigate re-purposing for use in ViBRANT
147 http://aginfra.eu/

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	48.00
7	Museum für Naturkunde, Berlin	MfN	2.00
11	Vizzuality	VIZZ	5.55
12	PENSOFT Publisher	PENSOFT	0.40
14	Global Biodiversity Information Facility	GBIFS	6.00

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 EuBon 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 2

HCMR improved the web site[148] of the COMBER citizen scientist pilot project to make it easier to use and more efficient for data storage and processing. On the front page of the web site there's now a short video that catches user's attention and prompts them to play with the services available.

More diving clubs (5) were involved in the COMBER activities over this period. They offered attractive discounts (20-25%) to divers who'd like to follow COMBER's underwater activities and also a discount for a ticket to the Cretaquarium and to an exhibition on Climate Change.

---

148 http://www.comber.hcmr.gr/

Visitors to the Cretaquarium, on the other hand, had the option of a free tour and training on the fish species living in the tanks of the facility and illustrated in the FishCard[149]. The discount rate offered by the Cretaquarium was 50% of the ticket for both the aquarium and the Climate Change exhibition.

More than 1,700 new records have been added to the COMBER web site of the pilot project, which will soon approach 4,000 records over two seasons. Professional divers from HCMR are also interested and involved in COMBER's activities.

Task 2 Data quality improvement module (GBIF[150])

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 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 2

A substantial portion of this task has been rendered effectively unnecessary by work GBIF did in year 1 to auto-correct the data they serve. The immediate consequence of this is to re-define Deliverable 8.3 that was originally:

> D8.3 - Scratchpads module supporting correction and annotation of GBIF observational data

> Description: Correction and annotation of GBIF observational data: Scratchpads module supporting correction and annotation of GBIF observational data [month 36]

The new text reads:

> D8.3 - A start-up portal for discovery of GBIF mediated data

> Description: The Nodes Portal Toolkit (NPT) Startup enables countries to establish a web presence with relevant biodiversity information from GBIF and EOL. The NPT Startup will be built on the ScratchPads platform, and will extend developments of ViBRANT to a broader context of the use of biodiversity data portals. The first version of the Nodes Portal Toolkit (NPT) Startup will be available as a Drupal profile accompanied with relevant documentation for use and further development. The NPT will be deployed in several countries within and outside Europe. At the completion of this deliverable, there will also be some commitments from the GBIF network to further develop the NPT Startup with more advanced functionalities.

The new Deliverable clearly has the same overall intention as Task 2, but is expected to have far greater impact. Note that ViBRANT is an infrastructure project, so the Description of Work (DoW) is a dynamic document maintained on the project web site and is not routinely archived as paper versions.

The Nodes Portal Toolkit is based on the Scratchpad infrastructure and is designed to allow local GBIF nodes to operate a community web site. The first full version of NPT Startup will be released for review at the beginning of 2013, but the current version is already used by INBio in Costa Rica, and by GBIF Benin. INBio, as the GBIF Participant Node for Costa Rica, is, with the help of the GBIF Secretariat, is developing additional functionalities to the current version of NPT Startup. This will enable the development of a national biodiversity portal for Benin. More countries have indicated their interest in using the first version of

---

149 http://www.bio-watch.com/index.php?option=com_virtuemart&Itemid=95

150 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.

the NPT Startup. As the current NPT Startup version will be sent out for review, we expect an expansion of the user base in 2013.

The Biodiversity Data Journal (BDJ) and the Paper Writing Tool (PWT) are in beta-testing phase, but annotated data sets will be assembled and published as a mechanism to give credit to those building the datasets.

Future work

After reviewing the release of the NPT and receiving feedback from the Nodes community, a plan will be drafted on how further development will be organised to reach the Deliverable D 8.3 - A start-up portal for discovery of GBIF mediated data.

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 on 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 be able to be linked to other ecological quality index calculation web tools (e.g. M-AMBI).

Progress in Year 2

A report[151] on existing processing services and analytical services has been prepared. The field is vast and fast evolving with few widely accepted standards. The study included, therefore, only those tools and services available by major players in the field, that is projects, initiatives, fora and organisations. Two factors are assumed to have an impact on the technical and operational features of the services and tools available: (a) the provider (web site); and (b) its function (the user needs it addresses). Many groups of tools and services with comparable if not similar or even identical functions are made available by different providers, but a substantial number of tools and services seem to have a degree of complementarity and could, therefore, used in combination in different kind of analyses to achieve better results and interpretation potential. The latter should be seriously taken into account by the large ESFRI infrastructures which are at the construction and operational stage, such as LifeWatch.

Routines for another two ecological quality indices were developed during Year 2. These indices are: (a) the benthic index (BENTIX[152]) and, (b) the Benthic Quality Index (BQI[153]). The former largely resembles the AZTI Marine Biotic Index (AMBI[154]), widely implemented throughout Europe, in that it makes an *a-priori* classification of the species in terms of their tolerance level. These values are subsequently taken into account in the calculation of the index value from the data (replicate unit/station/assemblage). The latter index, BQI, is based on the species composition and abundance, the quality levels being determined from these components alone, but this index performs poorly for samples containing few species and so the routines also included a variant of the index in which the calculation is made at the level of family and not on the species. Samples with few species are common in severely impacted habitats. The issue of the

151 http://www.comber.hcmr.gr/Milestone815.doc

152 Simboura N. & Zenetos A. (2002) Benthic indicators to use in Ecological Quality classification of Mediterranean soft bottom marine ecosystems, including a new Biotic Index. *Medit. Mar. Sci.*, **3(**2): 77-111. http://www.michele.scardi.name/corsi/bentix.pdf

153 Rosenberg, R, Blomqvist M, Nilsson H, Cederwall H, Dimming A (2004) Marine quality assessment by use of benthic species-abundance distribution: a proposed new protocol within the European Union Water Framework Directive. *Marine Pollution Bulletin* **49**: 728–739. ftp://ftp.ifremer.fr/ifremer/delao/gt_benthos_dce/Swedish_BF.pdf

154 Borja A, Franco J, Perez V (2000) A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. *Marine Pollution Bulletin* **40**: 1100–1114. http://www.ecasa.org.uk/Documents/AMBI-MarineBioticIndex.pdf

implementation of these routines to the Infrastructure has been solved in collaboration with WP5 through OBOE. This mechanism greatly simplified access to the calculations. Scripts written in C++ for the calculation of the indices, as well as templates for input and output formats have been delivered to WP5.

Biological Index Calculation Tool (BICT) is an online tool available via OBOE (Oxford Batch Operation Engine, https://oboe.oerc.ox.ac.uk/). BICT calculates up to nine different biodiversity indices on data submitted by the user and returns calculated data as well as a graphical report.

The following indices can be calculated:

1. Species Richness: The total number of different species in the submitted file per station / sample.

2. Number of Individuals: The total number of individuals in the submitted file per station / sample.

3. Shannon Index: Shannon's diversity index:

$$H' = -\Sigma_i \, p_i \, \ln(p_i)$$

    where p_i is the proportion of the total count arising from the i-th species.

4. Pielou's Index: A Species evenness index, indicating how close in numbers the species in an environment are. Calculated as

$$J' = H'(observed) / H'max$$

    where H'max is the maximum possible diversity which would be achieved if all species were equally abundant.

5. Rarefaction: ES(50), the "expected number" of species from 50 individuals, as developed by Hurlbert

6. Average Taxonomic Distinctness: The Average Taxonomic Distinctness (Δ+) describes the average distance between species, taking into account their higher classification / phylogeny) Calculation is performed as described in Clarke & Warwick (1998)[155] and is based on presence/absence data.

7. Variation in Taxonomic Distinctness: The Variation in Taxonomic Distinctness (Λ+) describes the "evenness" or "variation" in taxonomic distances between each pair of species, across their taxonomic tree. Calculation is performed as described in Clarke & Warwick (1998) and is based on presence/absence data.

8. BENTIX index: An Index of biological quality, developed for the assessment of the ecological and trophic status of water bodies according to EU Water Framework Directive (Simboura & Zenetos, 2002[152]).

9. Biological Quality Index (BQI): An Index of biological quality, developed for the EU Water Framework Directive by Rosenberg et al. (2004)[153]. Here, it is used in its modified form as described by Leonardsson et al. (2009)[156]. The species' sensitivity values are taken from the publication by Dimitriou et al. (2012)[157].

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

---

155 Clarke, K. R. & Warwick, R. M. (1998). A taxonomic distinctness index and its statistical properties. Journal of Applied Ecology 35, 523–531. http://www.jstor.org/stable/2405167

156 Leonardsson, K., Blomqvist, M. & Rosenberg, R. (2009). Theoretical and practical aspects on benthic quality assessment according to the EU-Water Framework Directive – Examples from Swedish waters. *Marine Pollution Bulletin* **58**, 1286–1296. ftp://ftp.azti.es/neagig/Papers/Leonardssonetal2009.pdf

157 Dimitriou, P. D., Apostolaki, E. T., Papageorgiou, N., Reizopoulou, S., Simboura, N., Arvanitidis, C. & Karakassis, I. (2012). Meta-analysis of a large data set with Water Framework Directive indicators and calibration of a Benthic Quality Index at the family level. *Ecological Indicators* **20**, 101–107.

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 thread 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 2
Vizzuality has developed two tools[158], one for Endangered Species Red List Assessments called **GeoCAT**[159] (in collaboration with WP5) and a more generic tool to do analysis and visualisation called **CartoDB**[160].

GeoCAT was launched at the end of Year 1. To date more than 1,000 assessments had been made and at least 5 assessments have officially been published by IUCN that made use of GeoCAT. IUCN is working on a method to implement a batch upload from GeoCAT to their system. GeoCAT has been cited 5 times already and has been presented in more than 10 workshops and conferences.

From January 2012 to November 2012, we recorded 7,082 unique visitors to the application online. Around 40% of the users coming from UK, 20% Brazil, 10% US and the remainder countries around the world. The total amount of users, based on the visits and the participation in the forum is around 40 users worldwide with numbers increasing as more training/workshops are performed using the tool.

CartoDB is an Open Source visualisation platform built to make it easier for people to tell their stories by providing them with flexible and intuitive ways to create maps and design geospatial applications. It is now used by more than 7,000 users around the world (Fig 3). Many users are through Biodiversity Informatics projects, like Map of Life, Vernet or Canadensys, but the majority are currently not attached to the Biodiversity field. The developed platform is remarkably flexible with usages going from Carbon Calculators (UNEP-WCMC) up to visualising the results of US elections (Wall Street Journal). More than 100 applications have already been developed using CartoDB, and more than 20 Million visits had been made to visualisations using CartoDB. The current growth rate is of 1,500 new users per month.

Integration of visualisation applications with Scratchpads has been delayed because of the migration from Drupal 6 to 7. A workshop on data visualisation using CartoDB/GeocAT was organised in November in which developers from WP2 and WP8 have begun to plan the integration of GeoCAT and Scratchpads.

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.

Progress in Year 2
The development of a generic Scratchpads module for data recording (Citizen Science) was postponed to the first quarter of the third project year because it needed more time to collaborate with WP2. Migration from Drupal 6 to Scratchpads2 requires additional effort both from WP8 and WP2 to deliver the Alpha version of the data recording module.

---

158 http://vizzuality.s3.amazonaws.com/vibrant_deliverables/D8.2_Vizzuality.pdf
159 http://geocat.kew.org/
160 http://cartodb.com/

3.2.3. Project Management

Consortium management tasks and achievements

The Consortium Management Committee has met 4 times in the second 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 priorities. The board has not met since, because no substantive issues of conflict or deviations from the initial strategy have been identified. Individual board members have been frequently consulted on an ad-hoc basis about issues relevant to their area of expertise..

Communication between beneficiaries

Interaction between project partners has been the outstanding feature of the project's second year, both within and between workpackages.

WP3 (VU) is working together with WP2 for the collection of web data for user studies and to build an integrated Scratchpads user database (access log files) and employing the OU's data mining expertise in WP7 to develop an alternative metric for e-Science.

WP3 (RBINS) is working together with WP1 to updating the public pages of the ViBRANT website, with information related to the various tasks and activities going on in the work packages, and to produce information leaflets used for project publicity.

WP3 (RBINS) is working together with WP2 developers for help desk support, training, beta testing and website content delivery, and with WP8 for training delivery in the framework of the COMBER project.

WP4 is developing a Scratchpad common access point by working with WP2 to develop a Drupal export module for Scratchpads, and with WP5 to linking Xper2 to the CDM platform.

WP4 is working with WP8 and WP6 to publishing Fauna Europaea as an e-Publication.

The WP4 Ontology Platform activities are working together with WP2 on (taking over) hosting facilities by GBIF and on proper routines for Scratchpads consuming GBIF vocabularies, with WP5 (UPMC) on key development and with WP7 applying vocabularies to semantic entity extraction and automatic mark-up of documents; an effort which overlaps with the EU project agINFRA[161] (Grant No. 283770) and BioVel[162] (Grant No. 283359) with whom a joint meeting was held in Oxford at 24 February 12. The function of the Media Wiki to provide a publication tool will be coordinated with WP6 (Pensoft).

WP7 continued collaboration with WP5 and to deliver GoldenGATE web services through OBOE.

WP6 & WP7 have aligned development decisions that relate to the overlap of literature mark up and processing. This will culminate in the development of a simple literature search tool through REFinder to be progressed by WP6 and the development of more sophisticated search options by WP7.

The current version of the NPT Startup makes use of much functionality already present in Scratchpads or functionality that has been developed in the framework of ViBRANT. There are clear connections with WP2, WP4, WP5 and WP8 in this product.

Co-operation with other projects

ViBRANT has been co-operating closely with the Biodiversity Heritage Library (BHL[163]) particularly through WP7 in the construction and population of CiteBank (see WP7 report above). Collaboration with BHL is on-going.

---

161 agINFRA: http://aginfra.eu
162 BioVel: http://www.biovel.eu
163 http://www.biodiversitylibrary.org/

Plazi[164] is an open repository of structured, biodiversity-related publications. Donat Agosti, the project leader 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.

Encyclopedia of Life (EoL[165]) closely modelled their LifeDesks on Scratchpads as a means for users to contribute data to the Encyclopedia. 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[166] 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 co-leading 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 12 between BioVeL, agINFRA and ViBRANT. We will maintain our links with BioVel to share and benefit from mutually relevant development of end-user workflows.

agINFRA (Grant No. 283770) is 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. In the year ahead we expect to collaborate with agINFRA in areas of common interest such as Linked Open Data and bibliographies.

4D4Life & i4Life: the pipelines built 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 2011 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 continues to be the retention of skilled staff in crucial positions. WP4 has benefitted immeasurably from Yde de Jong's involvement as WP4 manager, but his position at UvA has been terminated in the first month of Y3. Currently, the possibility of a temporary position at RBINS for Dr de Jong is being investigated. If positive, the budget will be transferred from UvA to RBINS to cover his committed time. The management committee is aware of the risk to the project should we lose Dr de Jong's involvement and are monitoring the situation closely.

The NHM has undergone a major internal re-organisation that significantly increases its commitment to biodiversity informatics generally. This is an extremely positive development in terms of the sustainability of the e-infrastructure that ViBRANT is constructing, but the timing means that the project has lost a full-time developer, Ben Scott, to an internal NHM role. The project's lead developer, Simon Rycroft, has similarly been appointed to a permanent position in the NHM, but is now only able to devote 80% of his time to ViBRANT. In recognition of the impact that this will have, the NHM has made two short-term positions available and we have been fortunate to recruit two talented developers in Alice Heaton and Katherine

---

164 http://plazi.org/
165 http://eol.org/
166 http://www.lifewatch.eu/

Bouton. The advertisement to replace Ben Scott published in November failed to attract any suitable applicants. We are re-advertising now, but given the central role played by Scratchpads in the project, and the general need to integrate components into a straightforward user environment, this absence represents a significant risk to the project.

Various project elements and partners have felt somewhat disconnected from the project as a whole in the absence of general project meetings (cut as part of the budget reduction in negotiation phase). To compensate the management committee issued a call for partners to organise focused workshops based on particular issues rather than within individual workpackages. The first of these, on mapping, was held in November and widely considered a success. The workshops are open to all partners. Future workshops are planned for Citizen Science (January); .workshop on RefBank/RefFinder; Workshop on PWT/BDJ; Workshop on the nature and structure of the biodiversity research field; Workshop on literature mining and markup using OBOE; joint workshop with text mining and markup with agINFRA and BioVeL; and a workshop with JEMU on phylogenetics in Scratchpads.

Changes in the consortium

There have been no changes to the consortium in Year 2.

List of project meetings, dates and venues

Meeting	from	to	location	url
Management Committee 5	18/01/12		virtual	http://vbrant.eu/content/man-comm-5
WP3 meeting	24/02/12	24/02/12	virtual	http://vbrant.eu/content/notes-skype-meeting-vurbins
Management Committee 6	17/04/12	18/04/12	Sofia	http://vbrant.eu/content/man-comm-6
WP6/7 workshop	17/04/12	18/04/12	Sofia	
Management Committee 7	19/07/12		virtual	http://vbrant.eu/content/man-comm-7
ViBRANT WP4 Technical meeting	08/10/12	09/10/12	Berlin	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_WP4_meeting_8-9_November_2012.jpg
Management Committee 8	18/10/12		Crete	http://vbrant.eu/content/man-comm-8
CartoBD workshop	16/11/12		Madrid	

Publications

Ariño, Arturo H., Vishwas Chavan and Nick King (2011) The Biodiversity Informatics Potential Index. *BMC Bioinformatics* **12**(Suppl 15): S4. doi:10.1186/1471-2105-12-S15-S4

Burguiere, Thomas, Florian Causse, Visotheary Ung, and Régine Vignes-Lebbe. (2012) Ikey+: a new single-access key generation web service. *Systematic Biology*, **62**(1):157-161 doi:10.1093/sysbio/sys078

Chatzigeorgiou, G., S. Faulwetter, C. Arvanitidis, E. López and R. Sardá. (2012) Can coastal biodiversity measured in four Mediterranean sites be representative of the region? A test for the robustness of the NaGISA protocol based on syllid (Annelida, Polychaeta) taxocommunities. *Hydrobiologia* **691:**147-156. http://www.springerlink.com/content/x8238224rj286106/

Chavan, V. and L. Penev. (2011) Data Paper: Mechanism to incentivise discovery of biodiversity data resources." *BMC Bioinformatics* **12**(Suppl 15):S2. doi: 10.1186/1471-2105-12-S15-S2

Christodoulou, M., A. Antoniou, A. Magoulas, and A. Koukouras. (2012) Revision of the freshwater genus Atyaephyra (Crustacea, Decapoda, Atyidae) based on morphological and molecular data." *Zookeys* **229**:53-110. doi: 10.3897/zookeys.229.3919

Conruyt, Noël, David Grosser, and Régine Vignes-Lebbe. (2012) Knowledge Discovery for Biodiversity: from Data Mining to Sign Management. In *International Environmental Modelling and Software Society (iEMSs)*, R. Seppelt, A. A. Voinov, S. Lange and D. Bankamp (Eds). International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet, Sixth Biennial Meeting. Leipzig, Germany, 12. http://vbrant.eu/sites/vbrant.eu/files/iEMSs2012_Conruyt_et_Al.pdf

Conruyt, Noël, Véronique Sébastien, Olivier Sébastien, David Grosser and Didier Sébastien. (2012) From Knowledge transmission to Sign sharing: Semiotic Web as a new paradigm for Teaching and Learning In *The Future Internet In 20th European Conference on Artificial Intelligence.* Eunika Mercier-Laurent, Nada Matta, Mieczyslow L. Owoc and Inès Saad. (Eds.) ECAI 12 1st International Workshop on Artificial Intelligence for Knowledge Management, AI4KM 12. Montpellier, France, 12. http://www2.lirmm.fr/ecai2012/images/stories/ecai_doc/pdf/workshop/W22_AI4KM_proceedings.pdf

Duin, Daphne, David King and Peter van den Besselaar (2012) Identifying Audiences of E-Infrastructures - Tools for Measuring Impact. *PLoS ONE* **7**: e50943. doi:10.1371/journal.pone.0050943

Förstner, Konrad U., Gregor Hagedorn, Claudia Koltzenburg, Fabiana M. Kubke and Daniel Mietchen (2011) Collaborative platforms for streamlining workflows in Open Science. In *Proceedings of the 6th Open Knowledge Conference,* Sebastian Hellmann, Philipp Frischmuth, Sören Auer & Daniel Dietrich (Eds.) http://ceur-ws.org/Vol-739/paper_8.pdf

Goddard, Anthony, Nathan Wilson, Phil Cryer and Grant Yamashita (2011) Data hosting infrastructure for primary biodiversity data. *BMC Bioinformatics* 12(Suppl 15): S5 doi:[10.1186/1471-2105-12-S15-S5](10.1186/1471-2105-12-S15-S5)

Ingwersen, Peter and Vishwas Chavan (2011) Indicators for the Data Usage Index (DUI): an incentive for publishing primary biodiversity data through global information infrastructure. *BMC Bioinformatics* **12**(Suppl 15): S3 doi:[10.1186/1471-2105-12-S15-S3](10.1186/1471-2105-12-S15-S3)

Mietchen, Daniel, Gregor Hagedorn, Konrad U. Förstner, Fabiana M. Kubke, Claudia Koltzenburg, Mark J. Hahnel and Lyubomir Penev. (2011) Wikis in scholarly publishing. *Information Services and Use,* **31**:53-59. doi:[10.3233/ISU-2011-0621](10.3233/ISU-2011-0621)

Moritz, Tom, S. Krishnan, Dave Roberts, Peter Ingwersen, Donat Agosti, Lyubomir Penev, Matthew Cockerill and Vishwas Chavan. (2011) Towards mainstreaming of biodiversity data publishing: recommendations of the GBIF Data Publishing Framework Task Group. *BMC Bioinformatics* **12**(Suppl 15): S1. doi:[10.1186/1471-2105-12-S15-S1](10.1186/1471-2105-12-S15-S1)

Penev, Lyubomir, Pierfilippo Cerretti, Hans-Peter Tschorsnig, Massimo Lopresti, Filippo Di Giovanni, Teodor Georgiev, Pavel Stoev and Terry L. Erwin (2012) Publishing online identification keys in the form of scholarly papers. *ZooKeys* **205**: 1-3. doi:10.3897/zookeys.205.3581

Roberts, Dave and Tom Moritz (2011) A framework for publishing primary biodiversity data. *BMC Bioinformatics* **12**(Suppl 15): I1. doi:[10.1186/1471-2105-12-S15-I1](10.1186/1471-2105-12-S15-I1)

Tzomos, Th, A. Antoniou, A. Magoulas, C. Arvanitidis and A. Koukouras (2012) Systematics and phylogeography of the genus *Palaemonetes* (Decapoda: Palaemonidae) in the cirum-Mediterranean region. In *The Crustacean Society Summer Meeting & Tenth Colloquium Crustacea Decapoda Mediterranea*. Royal Olympic Hotel, Athens, 12.

Dissemination activities

Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
C	KIT	Taxonomic Databases Working Group Annual Meeting	Towards A Universal Bibliography – The RefBank Approach	22/10/12 - 26/10/12	Beijing, China	Scientific community	90	International	Yes	http://www.tdwg.org/fileadmin/2012conference/slides/RefBank-TDWG-12.pdf
C	BGBM	Taxonomic Databases Working Group Annual Meeting	Interoperability between Scratchpads and the EDIT Platform for Cybertaxonomy	22/10/12 - 26/10/12	Beijing, China	Scientific community	90	International	Yes	http://www.tdwg.org/fileadmin/2012conference/slides/Morris_ViBRANTa.pdf
W	RBINS		Scratchpads 2 - Basic training course	17/10/12	Hellenic Centre for Marine Research, Heraklion	Scientific community	20	Greece	Yes	
W	GBIF	NCBO Webinar Series Talks	GBIF vocabulary management for biodiversity informatics	17/10/2012	Webinar	Scientific community	?	All	Yes	http://www.bioontology.org/GBIF-vocabulary-management-for-biodiversity-informatics
W	RBINS		Scratchpads 2 - Basic training course	16/10/12	NHM, London	Scientific community	5	EU	Yes	
C	NHM	Biodiversity Technologies Biodiversity Institute Symposium	Scratchpad virtual research environments: sharing, linking and publishing biodiversity data the ViBRANT way	27/09/12 - 28/09/12	Oxford, UK	Scientific community	200	International	Yes	http://vbrant.eu/sites/vbrant.eu/files/12-Scratchpads-VSmith-BioSympOx_opt.pdf
C	KIT	TPDL: Conference on Theory and Practice of Digital Libraries	Improved Bibliographic Reference Parsing Based on Repeated Patterns	23/09/12 - 27/09/12	Pafos, Cyprus	Scientific community	50	European	Yes	

Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
W	VU	Shadows and lights on global biodiversity: taxonomy's revival EASA 12 (European Association of Social Anthropologists)	Studying effects of e-infrastructures for taxonomy	11/07/12	Paris	Scientific community	20	France, Switserland, Netherlands, Slovenia	Yes	
Pr	VU	Annual Network Institute Event 12	The social design of a science (e)-infrastructure	26/06/12	Amsterdam	Scientific community	130	local	No	
W	RBINS	Sixth European Hemiptera Congress	Scratchpads 2 - Basic training course	26/06/12	American University, Blagoevgrad (EHC6)	Scientific community	15	international	Yes	
W	RBINS		Scratchpads 2 - Advanced training course	12/06/12	NHM, London	Scientific community	2	International	Yes	
W	RBINS	in conjunction with the "International Symposium on Flatworm Biology" held at the SMNH	Scratchpads 2 - Advanced training course	08/06/12	Swedish Museum of Natural History, Stockholm	Scientific community	5	EU	Yes	
W	RBINS	in conjunction with the "International Symposium on Flatworm Biology" held at the SMNH	Scratchpads 2 - Basic training course	07/06/12	Swedish Museum of Natural History, Stockholm	Scientific community	9	EU	Yes	
W	RBINS		Scratchpads 2 - Basic training course	23/05/12	NHM, London	Scientific community	12	EU	Yes	

Type of Activity	Lead partner	Conference	Title	Date	Place	Type of audience	Size of audience	Countries	Presentation	Slides
P	NHM	Joint meeting of LifeWatch with related projects	ViBRANT	13/04/12	Rome	Scientific community	25	EU	Yes	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_Reading290312.pdf
P	NHM	Taxonomy and the Catalogue of Life: Ecosystem of Services	ViBRANT	28/03/12 - 29/03/12	Reading	Scientific community	25	EU	Yes	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_Reading290312.pdf
W	RBINS		Scratchpad 2 update training session for Scratchpad maintainers	21/03/12	Natural History Museum London	Scientific community	6	EU	Yes	
Pr	NHM	International Conference on Research Infrastructures	Integrating services across biodiversity science	21/03/12 - 23/03/12	Copenhagen	Scientific communityPolicy makers	750	Global	No	http://vbrant.eu/sites/vbrant.eu/files/ICRI_poster.pdf
W	RBINS		Scratchpad 2 update training session for Scratchpad maintainers	20/03/12	Royal Botanic Gardens Kew	Scientific community	8	UK	Yes	
W	VU	Digital Social Research: A Forum for Policy and Practice	Social science for e-science in biodiversity research A position paper on behalf of ViBRANT	13/03/12	Oxford	Scientific communityPolicy makers	30	EU, Australia, US	No	
W	OU	Social Science and Digital Research: Interdisciplinary Insights	'Wish you were here before!'#Who gains from collaboration between computer science and social research?	12/03/12	Oxford	Scientific community	50	EU, Australia, US	Yes	http://vbrant.eu/sites/vbrant.eu/files/Presentation_DuinKingBesselaar.ppt

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 3 are set towards the end of year 2. See Table 2 for details. Note that the definitive list of milestones and deliverables, as befits an e-infrastructure project, is maintained on the project web site, not in Annex 1. This was a project design intention from project conception.

Impact of possible deviations from the planned milestones and deliverables

Deliverable D8.3 has been changed from:

> D8.3 - Scratchpads module supporting correction and annotation of GBIF observational data

> Description: Correction and annotation of GBIF observational data: Scratchpads module supporting correction and annotation of GBIF observational data [month 36]

To:

> D8.3 - A start-up portal for discovery of GBIF mediated data

> Description: The Nodes Portal Toolkit (NPT) Startup enables countries to establish a web presence with relevant biodiversity information from GBIF and EOL. The NPT Startup will be built on the ScratchPads platform, and will extend developments of ViBRANT to a broader context of the use of biodiversity data portals. The first version of the Nodes Portal Toolkit (NPT) Startup will be available as a Drupal profile accompanied with relevant documentation for use and further development. The NPT will be deployed in several countries within and outside Europe. At the completion of this deliverable, there will also be some commitments from the GBIF network to further develop the NPT Startup with more advanced functionalities. [month 36]

The reason for the change was that the original deliverable was rendered effectively unnecessary by some work GBIF did to auto-correct the data they serve. The new deliverable extends the goal of increasing the usability of GBIF's data but offers a mechanism to get significantly greater engagement.

Project milestones have developed in a way that reflects the dynamic environment in which we are working. Any changes to milestones are recorded on the project web site, that acts as a dynamic Annex 1 to the project.

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 in Year 2.

Development of the Project website

The project web site (http://vbrant.eu) has been refurbished during the year in an effort to increase the accessibility of the project and its products. The site 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 act as a dynamic Annex 1 to the project, i.e. 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	http://scratchpads.eu/	The social media component of ViBRANT. Born in the FP6 EDIT project (http://www.e-taxonomy.eu).
biowikifarm	http://biowikifarm.net/ with many hosted sites, especially http://species-id.net/ and http://terms.gbif.org/	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	https://oboe.oerc.ox.ac.uk	A front-end for non-scratchpad users to the database and scheduling services being developed as part of WP5.
Identification key webservice	http://www.identificationkey.fr	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	http://www.comber.hcmr.gr/	A citizen science portal for recreational divers and snorkelers to record observations of fish species, developed by WP8
GeoCAT	http://geocat.kew.org/	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	http://www.aginfra-project.eu/	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	http://www.cs.cf.ac.uk/biovel/	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	http://www.cost.esf.org/domains_actions/essem/Actions/ES1003	Development and implementation of a pan-European Marine Biodiversity Observatory System (EMBOS), is an EU project funded by the COST instrument.
eMonocot	http://e-monocot.org/	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	http://www.eol.org/	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'	http://jemu.myspecies.info	Uses Scratrchpads 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	http://www.lifewatch.eu/	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!	http://open-up.eu/	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	http://www.marine-vectors.eu/	Vectors of Change in Oceans and Seas Marine Life, Impact on Economic Sectors (VECTORS) is a European Commission Seventh Framework Programme (FP7) project.

3.3. Deliverables and milestones tables

Table 1: Deliverables

Del.no.	Deliverable name	Version	WP no.	Lead beneficiary	Nature	Dissemination level	Delivery date from Annex I (proj month)	Actual / Forecast delivery date Dd/mm/yyyy	Status No submitted/ Submitted	Contractual Yes/No	Comments
D1.1	Overall Management		1	NHM		PU	36	30/11/13	No submitted		
D2.1	Distributing servers	Final	2	NHM	Report	PU	12	30/11/2011	Submitted		http://vbrant.eu/sites/vbrant.eu/files/d2.1.pdf The Scratchpad Mirror server is: http://e090.bgbm.fu-berlin.de/ An example mirrored site is: http://mastigoteuthidae.myspecies.info/ => http://mastigoteuthidae.m1.myspecies.info/
D2.2	Scratchpad Registry	Final	2	NHM	Report	PU	18	31/05/12	Submitted		http://vbrant.eu/sites/vbrant.eu/files/d2.2.pdf
D2.3	Financial sustainability		2	NHM		PU	30	31/05/13	No submitted		
D2.4	Unit testing		2	NHM		PU	36	30/11/13	No submitted		
D3.1	Training strategy	Final	3	RBINS	Report	PU	12	29/11/2011	Submitted		http://vbrant.eu/sites/vbrant.eu/files/D3.1-Training strategy.pdf
D3.2	Service delivery and evaluation	Final	3	VU	Report	PU	24	30/11/12	Submitted		http://vbrant.eu/sites/vbrant.eu/files/D3.2_291112.pdf
D3.3	Community delivery and evaluation		3	RBINS		PU	36	30/11/13	No submitted		
D4.1	Scratchpad common access point	Final	4	UvA	Report	PU	12	30/11/2011	Submitted		http://vbrant.eu/sites/vbrant.eu/files/ViBRANT-D4.1.doc http://vbrant.eu/sites/vbrant.eu/files/Descriptive_data_workshop, June 6-7, 2011.doc
D4.2	Ontology tools		4	UvA		PU	25	31/12/12	No submitted		http://vbrant.eu/sites/vbrant.eu/files/D4.2_ProgressReport301111.pdf
D4.3	Design of robust services		4	UvA		PU	36	30/11/13	No submitted		https://oboe.oerc.ox.ac.uk/docs
D5.1	Prototype workflows and API	Final	5	UOXF.E9	Report	PU	12	30/11/2011	Submitted		http://vbrant.eu/content/d51-prototype-workflows-and-api

D5.2	Functional tools	Final	5	UOXF.E9	Report	PU	24	30/11/12	Submitted	http://vbrant.eu/sites/vbrant.eu/files/D5_2.pdf
D5.3	Sustainable software services		5	UOXF.E9		PU	36	30/11/13	No submitted	
D6.1	XML mark up tool & service	Final	6	PENSOFT	Report	PU	12	30/11/2011	Submitted	http://vbrant.eu/sites/vbrant.eu/files/D6-1-PENSOFT-XML mark up tool.pdf
D6.2	Review, refine & evaluate services	Final	6	PENSOFT	Report	PU	24	30/11/12	Submitted	http://vbrant.eu/sites/vbrant.eu/files/D6-2-Review & refine services.docx
D6.3	Data publication workflow		6	PENSOFT		PU	36	30/11/13	No submitted	
D7.1	Community contributed bibliography	Final	7	OU	Report	PU	12	30/11/2011	Submitted	Prototype Service: http://plazi2.cs.umb.edu:8080/RefBank/search Description: http://vbrant.eu/sites/vbrant.eu/files/D7-1_report_final.pdf
D7.2	Mark-up modules	Final	7	OU	Report	PU	24	29/11/12	Submitted	http://vbrant.eu/sites/vbrant.eu/files/D7-2_report_final.pdf
D7.3	Literature search		7	OU		PU	35	31/10/13	No submitted	
D8.1	Scratchpad modules engaging citizen scientists	Final	8	HCMR	Report	PU	12	30/11/2011	Submitted	http://www.comber.hcmr.gr/ http://vbrant.eu/sites/vbrant.eu/files/Deliverable8.pdf
D8.2	Analytical and reporting tools	Final	8	Vizz	Report	PU	24	30/11/12	Submitted	http://vizzuality.s3.amazonaws.com/vibrant_deliverables/D8.2_Vizzuality.pdf
D8.3	A start-up portal for discovery of GBIF mediated data		8	GBIF		PU	36	30/11/13	No submitted	

Table 2: Milestones

Milestone no.	Milestone name	Part of Deliverable	WP	Lead beneficiary	Due date	Achieved Yes/No	Date delivered	Comment
M1.10	Confirmation of year 1 meeting dates	D1.1 - Overall Management	1	NHM	31/12/10	Yes	31/12/10	http://vbrant.eu/content/management-committee
M1.11	Circulation of next meeting agenda	D1.1 - Overall Management	1	NHM	31/12/10	Yes	31/12/10	http://vbrant.eu/content/management-committee-meeting
M1.12	Organise kick-off meeting	D1.1 - Overall Management	1	NHM	31/12/10	Yes	31/12/10	http://vbrant.eu/scripting-life
M1.13	Minutes of management committee meetings	D1.1 - Overall Management	1	NHM	28/02/11	Yes	28/02/11	http://vbrant.eu/content/management-committee-meeting
M1.14	Minutes of strategic board meetings	D1.1 - Overall Management	1	NHM	28/02/11	Yes	28/02/11	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT _StrategicBoardMinutes_Jan11.pdf
M1.15	Circulation of next meeting agenda	D1.1 - Overall Management	1	NHM	30/04/11	Yes	30/04/11	http://vbrant.eu/content/man-comm-2
M1.16	Minutes of management committee meetings	D1.1 - Overall Management	1	NHM	30/06/11	Yes	30/06/11	http://vbrant.eu/content/man-comm-2
M1.17	Minutes of strategic board meetings	D1.1 - Overall Management	1	NHM	30/06/11	Yes	30/06/11	http://vbrant.eu/content/strategic-board-2
M1.18	Circulation of next meeting agenda	D1.1 - Overall Management	1	NHM	30/09/11	Yes	30/09/11	http://vbrant.eu/content/man-comm-4
M1.19	Preparation of first JPA templates	D1.1 - Overall Management	1	NHM	31/08/11	Yes	31/08/11	http://vbrant.eu/content/m119-preparation-first-jpa-templates
M1.20	Minutes of management committee meetings	D1.1 - Overall Management	1	NHM	18/11/11	Yes	18/11/11	http://vbrant.eu/content/man-comm-4
M1.21	Confirmation of year 2 and 3 meeting dates	D1.1 - Overall Management	1	NHM	30/11/11	Yes	30/11/11	http://vbrant.eu/content/man-comm-3
M1.22	Circulation of next meeting agenda	D1.1 - Overall Management	1	NHM	31/12/11	Yes	31/12/11	http://vbrant.eu/content/man-comm-5
M1.23	First year financial reports	D1.1 - Overall Management	1	NHM	31/12/11	Yes	31/12/11	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT _AnnReptYear1_public.pdf

M1.24	First year annual progress report	D1.1 - Overall Management	1	NHM	31/01/12	Yes	31/01/12	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_AnnReptYear1_public.pdf
M1.25	Minutes of management committee meetings	D1.1 - Overall Management	1	NHM	29/02/12	Yes	29/02/12	http://vbrant.eu/content/man-comm-5
M1.26	Minutes of strategic board meetings	D1.1 - Overall Management	1	NHM	29/02/12	Yes	29/02/12	http://vbrant.eu/content/strategic-board-2
M1.27	Organise mid-project meeting	D1.1 - Overall Management	1	NHM	31/03/12	Yes	31/03/12	http://vbrant.eu/content/m127-organise-mid-project-meeting
M1.28	Circulation of next meeting agenda	D1.1 - Overall Management	1	NHM	30/04/12	Yes	30/04/12	http://vbrant.eu/content/man-comm-7
M1.29	Minutes of management committee meetings	D1.1 - Overall Management	1	NHM	30/06/12	Yes	30/06/12	http://vbrant.eu/content/man-comm-7
M1.30	Minutes of strategic board meetings	D1.1 - Overall Management	1	NHM	30/06/12	Yes	30/06/12	http://vbrant.eu/content/strategic-board-2
M1.31	Circulation of next meeting agenda	D1.1 - Overall Management	1	NHM	31/08/12	Yes	31/08/12	http://vbrant.eu/content/man-comm-8
M1.32	Preparation of second JPA templates	D1.1 - Overall Management	1	NHM	31/08/12	Yes	31/08/12	http://vbrant.eu/content/m132-preparation-second-jpa-templates
M1.33	Minutes of management committee meetings	D1.1 - Overall Management	1	NHM	31/10/12	Yes		http://vbrant.eu/content/man-comm-8
M1.34	Circulation of next meeting agenda	D1.1 - Overall Management	1	NHM	31/12/12	No		
M1.35	Second year financial reports	D1.1 - Overall Management	1	NHM	31/12/12	No		
M1.36	Second year annual progress report	D1.1 - Overall Management	1	NHM	31/01/13	No		
M1.37	Minutes of management committee meetings	D1.1 - Overall Management	1	NHM	28/02/13	No		
M1.38	Minutes of strategic board meetings	D1.1 - Overall Management	1	NHM	28/02/13	No		
M1.39	Organise final meeting	D1.1 - Overall Management	1	NHM	31/03/13	No		

M1.40	Circulation of next meeting agenda	D1.1 - Overall Management	1	NHM	30/04/13	No		
M1.41	Minutes of management committee meetings	D1.1 - Overall Management	1	NHM	30/06/13	No		
M1.42	Minutes of strategic board meetings	D1.1 - Overall Management	1	NHM	30/06/13	No		
M1.43	Circulation of next meeting agenda	D1.1 - Overall Management	1	NHM	31/08/13	No		
M1.44	Minutes of management committee meetings	D1.1 - Overall Management	1	NHM	31/10/13	No		
M1.45	Third year annual progress report	D1.1 - Overall Management	1	NHM	30/11/13	No		
M1.46	Third year financial reports	D1.1 - Overall Management	1	NHM	30/11/13	No		
M2.10	Test version of distributed Scratchpad server	D2.1 - Distributing servers	2	NHM	31/05/11	Yes	01/06/11	http://vbrant.eu/sites/vbrant.eu/files/m2.10.pdf
M2.11	Report on Scratchpad usage statistics options	D2.2 - Scratchpad Registry	2	NHM	31/05/11	Yes	21/11/11	http://vbrant.eu/sites/vbrant.eu/files/m2.11.pdf
M2.12	Ongoing module development supporting prioritised features requested by users through WP3	D2.1 - Distributing servers	2	NHM	30/11/11	Yes	21/11/11	http://vbrant.eu/sites/vbrant.eu/files/m2.12.pdf
M2.13	Define further milestones in the light of usage and feedback	D1.1 - Overall Management	2	NHM	30/11/11	Yes	21/11/11	http://vbrant.eu/sites/vbrant.eu/files/m2.13.pdf
M2.14	Report on the data search portal providing a single point of entry to all Scratchpad data completed	D2.1 - Distributing servers	2	BGBM	31/05/12	Yes	05/06/12	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT _M2_14.docx
M2.15	Biodiversity data citation metric	D6.3 - Data publication workflow	2	NHM	31/05/13	No		
M2.16	Localised unit testing mechanism	D2.4 - Unit testing	2	NHM	30/11/12	Yes	06/12/12	http://vbrant.eu/sites/vbrant.eu/files/m2.16.pdf

M2.17	Ongoing module development supporting prioritised features requested by users through WP3	D5.3 - Sustainable software services	2	NHM	30/11/12	Yes	05/12/12	http://vbrant.eu/sites/vbrant.eu/files/m2.17.pdf
M2.18	Data search portal providing a single point of entry to all Scratchpad data completed	D5.3 - Sustainable software services	2	BGBM	31/05/13	No		
M2.19	Report on the options for a biodiversity data citation metric for data published through ViBRANT	D6.3 - Data publication workflow	2	NHM	31/05/13	No		
M2.20	Ongoing module development supporting prioritised features requested by users through WP3	D5.3 - Sustainable software services	2	NHM	30/11/13	No		
M3.10	Delivery of a promotional strategy for ViBRANT services	D3.2 - Service delivery and evaluation	3	RBINS	30/09/11	Yes	27/09/11	http://vbrant.eu/sites/vbrant.eu/files/M.3.10_Delivery of a promotional strategy for ViBRANT services_0.pdf
M3.11	Potential Mechanisms to increase user uptake	D3.3 - Community delivery and evaluation	3	VU	13/01/12	Yes	13/01/12	http://vbrant.eu/sites/vbrant.eu/files/Milestone_311_VU_130112.pdf
M3.12	Assessment of user support services, with the refinement of subsequent milestones	D3.1 - Training strategy	3	RBINS	29/11/11	Yes	29/11/11	http://vbrant.eu/sites/vbrant.eu/files/M3.12-Assessment of user support services.pdf
M3.13	Recommendations to modify products	D3.2 - Service delivery and evaluation	3	VU	30/11/12	No		
M3.14	Assessment of user support services and promotional activities	D3.2 - Service delivery and evaluation	3	RBINS	28/11/12	Yes	28/11/12	http://vbrant.eu/sites/vbrant.eu/files/M3.14_Assessment_usersupportservices_and_promotionalactivies-281112.pdf
M3.15	Ambassadors' network is fully operational	D3.3 - Community delivery and evaluation	3	RBINS	30/04/13	No		
M3.16	Established user support services to keep the open source network community alive and vibrant	D3.3 - Community delivery and evaluation	3	RBINS	30/11/13	No		

M3.17	Suggestions for potential new user groups	D3.3 - Community delivery and evaluation	3	VU	31/10/13	No		
M4.10	Define methods to use the Scratchpad service layer deliverable within the CDM store	D4.1 - Scratchpad common access point	4	GBIF	31/05/11	Yes	25/07/11	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT M410 Specification CDM Scratchpads_DwC-A_module.doc
M4.11	Prototype of collaborative community interface	D4.1 - Scratchpad common access point	4	JKI	31/05/11	Yes	31/05/11	http://vbrant.eu/sites/vbrant.eu/files/M4.11%E2%80%94Prototype of collaborative community interface_1.pdf
M4.12	Liaison and networking with ontology experts and existing ontology providers	D4.2 - Ontology tools	4	GBIF	31/07/11	Yes	02/12/11	http://vbrant.eu/sites/vbrant.eu/files/Smith-Submitted-v2Revised_ed_final.doc http://vbrant.eu/sites/vbrant.eu/files/D4.2_ProgressReport301111.pdf
M4.13	Release an API on the catalogue of resources	D4.1 - Scratchpad common access point	4	JKI	31/08/11	Yes	05/10/11	http://vbrant.eu/sites/vbrant.eu/files/M4.13 - Release an API on the catalogue of resources - report.pdf
M4.14	Define further milestones in the light of usage and feedback	D1.1 - Overall Management	4	UvA	30/11/11	Yes	24/12/11	http://vbrant.eu/sites/vbrant.eu/files/WP4_milestones_2.xls
M4.15	Ontology Tools: migrate GBIF Vocabularies Service	D4.2 - Ontology tools	4	GBIF	31/12/11	Yes	31/12/11	http://vocabularies.gbif.org

M4.16	Ontology Tools: scoping document on a KOS architecture	D4.2 - Ontology tools	4	UvA	31/12/11	Yes	26/01/12	GBIF has prepared a draft charter proposing to form a new TDWG task group on a Vocabulary Management (VMTG). The VMTG proposal focuses on tools for managing a new glossary of basic terms. The present tools to create Darwin Core extensions and vocabularies for the GBIF infrastructure does not include such a basic overview list of the terms that are defined and reused by the extensions and vocabularies.The plan is to start by exploring three alternative software tools for managing basic terms. (1) A demo using the isoCAT software tool has recently been prepared. The isoCAT platform seems very well designed from the perspective of functionality, work-flows and structure for maintaining a list of basic terms in a collaborative manner. However the isoCAT user interface is still not mature and in the present form not sufficiently user-friendly. (2) We have explored the SpeciesID Wiki and have started to develop a new tool using Semantic Media Wiki for describing basic terms. We hope this could be a option and a solution to work well together with the SpeciesID and other ViBRANT tools. (3) We also wish to explore the possibilities for using the present http://vocabularies.gbif.org for maintaining basic terms. However the present feeling is that this tool will need too many modifications to be extended to cover this purpose. The Vocabulary Server is well suited for building Darwin Core extensions and vocabularies for use by the GBIF IPT, but less suitable for the functionality we have in mind for the glossary of terms.We would very much welcome feedback and suggestions to this first draft!
M4.17	Extensions for DwC-A export functionality	D4.1 - Scratchpad common access point	4	UvA	28/02/12	Yes	02/03/12	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4 17 %E2%80%94 Extensions for DwC-A export functionality.pdf
M4.18	Human Interface for CDM-ViBRANT Index	D4.1 - Scratchpad common access point	4	UvA	28/03/12	Yes	13/12/12	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4.18_%E2%80%94_Human_Interface_for_CDM-ViBRANT_index.pdf http://dev.e-taxonomy.eu/vibrant_index/search/
M4.19	Canvassing glossary-oriented potential contributors	D4.2 - Ontology tools	4	UvA	16/05/12	Yes	11/07/12	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4.19 %E2%80%94 Canvassing glossary-oriented potential contributors.pdf
M4.20	Ontology Tools: new prototype (GBIF) Glossary of Terms registry	D4.2 - Ontology tools	4	GBIF	30/04/12	Yes	26/04/12	http://vbrant.eu/sites/vbrant.eu/files/2012-04 ViBRANT M420 status report_0.pdf http://terms.gbif.org/ http://kos.gbif.org/wiki/ http://species-id.net/ http://rs.gbif.org/ http://rs.gbif.org/sandbox/terms/ http://tools.gbif.org/resource-browser/

M4.21	Ontology Tools: new prototype Biodiversity BioPortal	D4.2 - Ontology tools	4	GBIF	31/07/12	Yes		http://bioportal.bioontology.org/ontologies/3058
M4.22	BiolFlor Data as semantic web	D4.2 - Ontology tools	4	UvA	28/02/13	No		
M4.23	Scratchpad to DwC-A Mapping Module	D4.1 - Scratchpad common access point	4	UvA	30/09/12	Yes	01/10/12	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4.23%E2%80%94Scratchpad_to_DwC-A_Mapping_Module.pdf http://vbrant.eu/sites/vbrant.eu/files/Darwin_Core_Archive_Export_Module_EoL.docx
M4.24	Full text search CDM-ViBRANT Index	D4.1 - Scratchpad common access point	4	UvA	30/09/12	Yes	13/12/12	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4_24.pdf
M4.25	Pilot implementation of webservice at BGBM	D4.1 - Scratchpad common access point	4	UvA	30/09/12	Yes	01/10/12	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4.25%E2%80%94SingleAccessKeyWebservice_EDITPlatform.pdf
M4.26	Practical application milestone for mediawiki (ViBRANT X party)	D4.2 - Ontology tools	4	UvA	30/11/13	No		
M4.27	Audubon Core standard on Mediawiki	D4.2 - Ontology tools	4	UvA	30/09/12	Yes	30/09/12	http://terms.gbif.org/wiki/Audubon_Core_Term_List
M4.28	Statistical output provided for CDM-ViBRANT Index query Interface	D4.1 - Scratchpad common access point	4	UvA	30/11/12	Yes	03/12/12	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4_28.pdf
M4.29	CDM SDD output optimised for Xper2	D4.1 - Scratchpad common access point	4	UvA	30/11/12	Yes	30/11/12	http://jackson.snv.jussieu.fr/thumbnailsWS/resources/doc/ViBRANT_M4.29-Xper2AppletDigestSDD.pdf
M4.30	Integrating the biowikifarm into ViBRANT authentication	D4.2 - Ontology tools	4	UvA	30/11/12	Yes	08/01/13	http://vbrant.eu/sites/vbrant.eu/files/ViBRANT_M4.30 — Integrating the biowikifarm into ViBRANT authentication.pdf
M4.31	Webservice-Layer adapted for ViBRANT	D4.1 - Scratchpad common access point	4	UvA	31/01/13	No		
M4.32	Integrating a wiki-glossary into the scratchpad	D4.2 - Ontology tools	4	UvA	31/01/13	No		
M4.33	XML Tranformations CDM export	D4.1 - Scratchpad common access point	4	UvA	30/04/13	No		
M4.34	Prototype for taxonomic generalisation	D4.1 - Scratchpad common access point	4	UvA	31/05/13	No		

M4.35	Filter functionality CDM export	D4.1 - Scratchpad common access point	4	UvA	31/05/13	No		
M4.36	XPer2-CDM-relation for taxonomic hierarchy specification	D4.1 - Scratchpad common access point	4	UvA	30/06/13	No		
M4.37	Xper2 and EDIT Platform publishing capability on Mediawiki	D4.2 - Ontology tools	4	UvA	31/05/13	No		
M4.38	ZooKeys special issue on Fauna Europaea data papers	D4.1 - Scratchpad common access point	4	UvA	28/02/13	No		
M4.39	Term browser	D4.2 - Ontology tools	4	GBIF	01/08/12	Yes	27/08/12	http://kos.gbif.org/termbrowser/
M5.10	API design plan	D5.1 - Prototype workflows and API	5	UOXF.E9	31/12/10	Yes	31/12/10	http://vbrant.eu/content/ms-510-api-design-plan
M5.11	Review existing key construction software and workflow interaction with both Scratchpads and the CDM	D5.2 - Functional tools	5	UPMC	28/02/11	Yes	28/02/11	http://www.infosyslab.fr/vibrant/M5_11_deliverable_UPMC_february2011.pdf
M5.12	Review user requirements for the visualisation tool for Scratchpads	D5.2 - Functional tools	5	Vizz	28/02/11	Yes	25/04/11	http://vbrant.eu/sites/vbrant.eu/files/M5-12.pdf
M5.13	API baseline documentation	D5.1 - Prototype workflows and API	5	UOXF.E9	31/03/11	Yes	19/04/11	http://vbrant.eu/sites/vbrant.eu/files/M5.13.api_baseline.txt
M5.14	Review of target applications (phylogenetics)	D5.2 - Functional tools	5	UOXF.E9	31/03/11	Yes	31/03/11	http://vbrant.eu/content/m514-review-target-applications-phylogenetics
M5.15	API version control	D5.1 - Prototype workflows and API	5	UOXF.E9	30/04/11	Yes	30/04/11	https://oboe.oerc.ox.ac.uk/docs
M5.16	Review of target platforms and requirements	D5.3 - Sustainable software services	5	UOXF.E9	30/06/11	Yes	30/06/11	http://vbrant.eu/content/m516-review-target-platforms-and-requirements
M5.17	Define further milestones in the light of usage and feedback	D1.1 - Overall Management	5	UOXF.E9	31/05/12	Yes	31/05/12	http://vbrant.eu/content/m517-define-further-milestones-light-usage-and-feedback#comment-579

M5.18	Review user requirements for enhanced user interface	D5.1 - Prototype workflows and API	5	UOXF.E9	30/11/11	Yes	30/11/11	http://vbrant.eu/content/m518-review-user-requirements-enhanced-user-interface
M5.19	Deliver prototype key-generating service through Scratchpads	D5.2 - Functional tools	5	UPMC	30/11/11	Yes	22/12/11	http://vbrant.eu/sites/vbrant.eu/files/M5_19_report2.doc http://identificationkey.fr
M5.20	Metadata repository design plan	D5.2 - Functional tools	5	UOXF.E9	31/12/11	Yes	31/12/11	http://vbrant.eu/content/m520-metadata-repository-design-plan#comment-622
M5.21	Implement visualisation tool for Scratchpads	D5.2 - Functional tools	5	Vizz	29/02/12	Yes	29/02/12	http://vizzuality.s3.amazonaws.com/M5.21.pdf
M5.22	Review algorithms for biodiversity indices	D5.2 - Functional tools	5	Vizz	29/02/12	Yes	29/02/12	http://vizzuality.s3.amazonaws.com/M5.22.pdf
M5.23	Implementation of platforms specific middleware	D5.3 - Sustainable software services	5	UOXF.E9	31/03/12	Yes	31/03/12	http://vbrant.eu/content/m523-implementation-platforms-specific-middleware#comment-495
M5.24	Implement application specific wrappers (phylogenetics) that can be used by Scratchpads	D5.3 - Sustainable software services	5	UOXF.E9	31/03/12	Yes	31/03/12	http://vbrant.eu/content/m524-implement-application-specific-wrappers-phylogenetics-can-be-used-scratchpads#comment-494
M5.25	Implement custom wrapper for the identification service that can be used by Scratchpads	D5.2 - Functional tools	5	UOXF.E9	31/03/12	Yes	31/03/12	http://vbrant.eu/content/m525-implement-custom-wrapper-identification-service-can-be-used-scratchpads#comment-496
M5.26	Review of target applications (bioclimatic modelling)	D5.2 - Functional tools	5	UOXF.E9	30/06/12	Yes	30/06/12	http://vbrant.eu/content/m526-review-target-applications-bioclimatic-modelling#comment-578
M5.27	Metadata repository implementation	D5.3 - Sustainable software services	5	UOXF.E9	30/09/12	Yes	30/09/12	http://vbrant.eu/content/m527-metadata-repository-implementation#comment-623
M5.28	Implement enhanced user interface tools for managing keys in Scratchpads	D5.3 - Sustainable software services	5	JKI	30/11/12	Yes	11/12/12	http://vbrant.eu/sites/vbrant.eu/files/M5.28 - Implement enhanced user interface tools for managing keys in Scratchpads.pdf
M5.29	Implement application specific wrappers (bioclimatic modelling)	D5.3 - Sustainable software services	5	UOXF.E9	31/12/12	No		
M5.30	Implement biodiversity indices tool	D5.3 - Sustainable software services	5	Vizz	28/02/13	No		

M5.31	Review of target applications and services (molecular identification)	D5.3 - Sustainable software services	5	UOXF.E9	31/03/13	No		
M5.32	Tests and validation of the services in coordination with Workpackage 3	D5.3 - Sustainable software services	5	UOXF.E9	31/05/13	No		
M5.33	Generic computational framework for geo-spatial analysis via OBOE.	D5.2 - Functional tools	5	UOXF.E9	30/09/12	Yes	30/09/12	http://vbrant.eu/content/m533-%E2%80%93-generic-computational-framework-geo-spatial-analysis-oboe
M5.34	Land cover classification tool	D5.2 - Functional tools	5	UOXF.E9	30/11/12	Yes	30/11/12	http://vbrant.eu/sites/vbrant.eu/files/output.pdf
M5.35	Environmental variables tool	D5.2 - Functional tools	5	UOXF.E9	31/01/13	No		
M5.36	GBIF specimen records tool	D5.2 - Functional tools	5	UOXF.E9	31/03/13	No		
M5.37	Threatened species tool	D5.2 - Functional tools	5	UOXF.E9	31/05/13	No		
M5.38	Alpha diversity tool	D5.2 - Functional tools	5	UOXF.E9	31/07/13	No		
M5.39	Beta diversity tool	D5.2 - Functional tools	5	UOXF.E9	30/09/13	No		
M5.40	Sustainable services. Completion of the documentation of services for D5.3	D5.3 - Sustainable software services	5	UOXF.E9	30/11/13	No		
M6.10	Use cases of existing standards of XML mark up tagging and semantic enhancement collected and review	D6.1 - XML mark up tool & service	6	PENSOFT	28/02/11	Yes	04/04/11	http://vbrant.eu/sites/vbrant.eu/files/Milestone 6 10-Review of mark up and tagging tools.pdf
M6.11	Prototype and beta-version of XML submission from Scratchpads to publishers	D6.1 - XML mark up tool & service	6	NHM	31/07/11	Yes	11/11/11	http://vbrant.eu/sites/vbrant.eu/files/M6.11-Prototype and beta-version of XML submission from Scratchpads to publishersmark up tool.pdf
M6.12	Workshop on mark up tagging tools and implementations	D6.1 - XML mark up tool & service	6	PENSOFT	31/10/11	Yes	11/11/11	The milestone report consists of two meetings on XML mark up schemas and tagging tools held in Paris (during the kick-off) minutes and presentations and in London (24-25 Oct 2011) minutes and presentations.

M6.13	Define further milestones in the light of usage and feedback	D1.1 - Overall Management	6	PENSOFT	30/11/11	Yes	11/11/11	List of two additional milestones planned for the second period if ViBRANT.
M6.14	Review of existing standards of semantic enhancements to taxonomic papers	D6.2 - Review, refine & evaluate services	6	PENSOFT	31/12/11	Yes	31/12/11	http://www.pensoft.net/inc/journals/download.php?fileId=3401&fileTable=J_GALLEYS
M6.15	Sample papers testing the XML-based editorial workflow elaborated and published	D6.3 - Data publication workflow	6	PENSOFT	30/11/12	Yes	17/12/12	http://vbrant.eu/sites/vbrant.eu/files/M6.15-TestingXML.docx
M6.16	Refinements to XML-based workflow for peer-review based on user feedback	D6.3 - Data publication workflow	6	PENSOFT	30/11/12	Yes	13/12/12	http://vbrant.eu/sites/vbrant.eu/files/M616-Review & refine services.docx
M6.17	New open access data publishing journal	D6.3 - Data publication workflow	6	PENSOFT	31/03/13	No		
M6.18	Testing automated tools of dissemination of published results to aggregators and indexing services	D6.3 - Data publication workflow	6	PENSOFT	31/08/13	No		
M6.19	XML export format of metadata on identification keys to Key Central (ALA)	D6.3 - Data publication workflow	6	PENSOFT	31/01/12	Yes	09/04/12	http://vbrant.eu/sites/vbrant.eu/files/M6.19-XML export format of metadata on identification keys to Key Central_0.pdf
M6.20	Beta version of XML queries to IPNI and ZooBank	D6.3 - Data publication workflow	6	PENSOFT	31/03/12	Yes	09/04/12	http://vbrant.eu/sites/vbrant.eu/files/M6.19-XML export format of metadata on identification keys to Key Central_0.pdf
M7.10	Agreement of standard format for community contributed bibliographies in conjunction with WP4	D7.1 - Community contributed bibliography	7	OU	17/06/11	Yes	17/06/11	http://www.wiki.scratchpads.eu/w/M710report
M7.11	Review of options for interactive mark up tools within the Scratchpad infrastructure	D7.2 - Mark-up modules	7	OU	17/06/11	Yes	17/06/11	http://www.wiki.scratchpads.eu/w/M711report
M7.12	A suite of test cases that will be used to test the de-duplication software	D7.1 - Community contributed bibliography	7	OU	31/07/11	Yes	16/12/11	http://vbrant.eu/sites/vbrant.eu/files/M7-12_report.pdf

M7.13	Review of options to use typographical information and other contextual clues	D7.2 - Mark-up modules	7	OU	31/07/11	Yes	20/12/11	http://vbrant.eu/sites/vbrant.eu/files/M7-13_report.pdf http://vbrant.eu/sites/vbrant.eu/files/Report%20-%20GoldenGATE%20Modules.doc
M7.14	Collaboration with WP6 on mark-up tagging tools and implementation	D6.1 - XML mark up tool & service	7	OU	31/10/11	Yes	16/12/11	http://vbrant.eu/sites/vbrant.eu/files/M7-14_report.pdf
M7.15	Define further milestones in the light of usage and feedback	D1.1 - Overall Management	7	OU	29/02/12	Yes	29/02/12	http://www.wiki.scratchpads.eu/w/M715report
M7.16	Mark-up modules delivering outline mark-up	D7.2 - Mark-up modules	7	OU	28/09/12	Yes	19/11/12	http://vbrant.eu/sites/vbrant.eu/files/M716.odt
M7.17	Review of pilot mark up processes within the Scratchpad infrastructure	D7.2 - Mark-up modules	7	OU	23/11/12	Yes	23/11/12	http://www.wiki.scratchpads.eu/w/M717report
M7.18	First integration phase complete	D7.3 - Literature search	7	OU	31/03/13	No		
M7.19	Review of pilot of reference de-duplication software	D7.3 - Literature search	7	OU	31/07/13	No		
M7.20	Workpackage software documentation produced	D7.3 - Literature search	7	OU	31/10/13	No		
M7.21	Add metadata to cover origin of bibliographies	D7.3 - Literature search	7	OU	13/04/12	Yes	13/04/12	http://www.wiki.scratchpads.eu/w/M721report
M7.22	Import bibliographies from Pensoft to RefBank	D7.3 - Literature search	7	OU	04/05/12	Yes	04/05/12	http://www.wiki.scratchpads.eu/w/M722report
M7.23	Extend RefBank import routines to support other widely used bibliographic formats, eg BibTex, RIS, etc	D7.3 - Literature search	7	OU	01/06/12	Yes	01/06/12	http://www.wiki.scratchpads.eu/w/M723report
M7.24	Upload service for complete bibliographies	D7.3 - Literature search	7	OU	01/06/12	Yes	01/06/12	http://www.wiki.scratchpads.eu/w/M724report
M7.25	Enhance reference parser to parse references in bulk uploads	D7.3 - Literature search	7	OU	31/10/12	Yes	31/10/12	http://www.wiki.scratchpads.eu/w/M725report

M8.10	Prototype ecological and conservation applications	D8.1 - Scratchpad modules engaging citizen scientists	8	HCMR	31/05/11	Yes	03/06/11	http://vbrant.eu/sites/vbrant.eu/files/Milestone_8.10_Indices_C_code_v0.1.zip https://git.scratchpads.eu/viewgit?p=indices.git;a=summary
M8.11	Review of relevant field recording tools and applications	D8.1 - Scratchpad modules engaging citizen scientists	8	MfN	25/11/11	Yes	25/11/11	http://vbrant.eu/sites/vbrant.eu/files/a review mobile apps_20111125.doc
M8.12	Identify rules for an intelligent algorithm to identify suspicious data records in GBIF data	D8.3 - A start-up portal for discovery of GBIF mediated data	8	GBIF	30/11/11	Yes	30/11/11	http://www.slideshare.net/DavidRemsen/tdwg-1remsen
M8.13	Define further milestones in the light of usage and feedback	D1.1 - Overall Management	8	HCMR	30/11/11	Yes	30/11/11	http://vbrant.eu/sites/vbrant.eu/files/Milestones WP8Overview_FINAL_Year1_2.xls
M8.14	Prototype demonstration project with non-professional marine naturalists	D3.1 - Training strategy	8	HCMR	30/11/11	Yes	30/11/11	http://www.pensoft.net/journals/zookeys/article/2149/abstract/engaging-the-broader-community-in-biodiversity-research-the-concept-of-the-comber-pilot-project-for-divers-in-vibrant
M8.15	Report on existing processing services and analytical services	D5.2 - Functional tools	8	HCMR	30/09/12	Yes	18/10/12	http://www.comber.hcmr.gr/Milestone815.doc
M8.16	Prototype analytical tools and reporting tools implemented on the visualization tool	D8.2 - Analytical and reporting tools	8	Vizz	30/09/12	Yes	30/09/12	https://github.com/Vizzuality/cartodb
M8.17	APIs to integrate various data sources in support of analytical and reporting tools	D5.2 - Functional tools	8	HCMR	30/09/12	Yes	30/09/12	https://oboe.oerc.ox.ac.uk/
M8.18	Scratchpads module for data recording (Citizen Science) - Alpha version	D8.1 - Scratchpad modules engaging citizen scientists	8	HCMR	30/04/13	No		
M8.19	Data flow from GBIF annotation module to Biodiversity Data Journal	D6.3 - Data publication workflow	8	PENSOFT	30/09/12	Yes	30/09/12	BDJ: http://biodiversitydatajournal.com/ PWT: http://pwt.pensoft.net/
M8.20	Ecological and conservation applications implemented as a service	D5.2 - Functional tools	8	HCMR	30/04/13	No		

M8.21	Integration of visualization applications with Scratchpads	D8.2 - Analytical and reporting tools	8	Vizz	30/04/13	No		
M8.22	Prototype version of a start-up portal for discovery of GBIF mediated data	D8.3 - A start-up portal for discovery of GBIF mediated data	8	GBIF	01/03/13	No		