PRO-iBIOSPHERE
WWW.PRO-iBIOSPHERE.EU

Coordination & policy development in preparation for a European Open Biodiversity Knowledge Management System, addressing Acquisition, Curation, Synthesis, Interoperability & Dissemination

PROJECT FINAL REPORT

Grant Agreement number: 312848

Project acronym: pro-iBiosphere

Project title: Coordination & policy development in preparation for a European Open Biodiversity Knowledge Management System, addressing Acquisition, Curation, Synthesis, Interoperability & Dissemination

Funding Scheme: Coordination and Support Action

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

Periodic report: 2nd

Period covered: from 1 September 2012 **to** 31 August 2014

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

Dr Soraya Sierra

Naturalis

Darwinweg 2

Postbus 9517

2300 RA Leiden

Tel: +31 (0)71 5273565

Email: soraya.sierra@naturalis.nl

Project website address: www.pro-ibiosphere.eu

SEVENTH FRAMEWORK PROGRAMME

European Commission

e-infrastructure

Table of Contents

4.1 Final publishable summary report

4.1.1 Executive summary

The EU FP7 pro-iBiosphere project was launched to investigate ways to improve access to biodiversity data, improve the efficiency of its curation and increase the user base of data and applications. pro-iBiosphere made significant impact in four areas directly related to: (i) Open Access to high-quality biodiversity information, (ii) technical and semantic interoperability, (iii) a more integrated user community, and (iv) a sustainable provision of high-quality base data. The pro-iBiosphere project defined the goals, fundamental principles, and framework of the **Open Biodiversity Knowledge Management System (OBKMS)**. This OBKMS will: (i) embrace existing digital infrastructures; (ii) accommodate past publications, including regional or global monographs; (iii) benefit from new enhanced publication processes that integrate data and narrative (text); and, (iv) manage the data elements that constitute taxonomic treatments, such as specimen data, images, DNA sequences, taxon names and their concepts, morphological characters, ecological and biological traits. It will thereby help to remove the rigid distinctions in the current system between the owners, curators, users and providers of biodiversity data.

The project contributed towards the OBKMS in a number of ways. It identified several legal and sociological challenges that hamper the exchange and re-use of biodiversity data. It addressed technical and semantic interoperability between different forms and formats in which data are published. It addressed sustainability issues related to the maintenance and curation of biodiversity data and of derived information and knowledge. The project achieved the following key outputs:

- **Removing legal barriers to the exchange of taxonomic information.** Recommendations for promoting free and open access to and re-use of biodiversity data and information were drawn up.
- **Launch of the** Bouchout Declaration for OBKM. The Declaration allows the biodiversity community to support the principle of open data; and to use this to identify and remove impediments to open data sharing. By 14.10.2014, the Declaration had been signed by 173 individuals and 85 organisations from 46 countries around the globe.
- **Improved cooperation and interoperability of e-infrastructures**
    - Recommendations for the use of persistent stable HTTP-URI identifiers were drawn up and are implemented already by 8 major institutions for their collection objects and treatments. The system has been proposed as an official product of the Consortium of European Taxonomic Facilities (CETAF).
    - An agreement was reached on adopting the BiodiversityCatalogue, developed by the EU FP7 project, as a global registry for biodiversity related services.
    - Recommendations were drawn up on tools and standards for biodiversity projects and initiatives.
    - Interoperability was achieved between different XML schemas (FlorML and TaxPub) and implemented by republishing a Flora Malesiana volume on Pensoft's Advanced Books platform.
    - A workflow was developed that locates, identifies, and enhances data included in treatments from both legacy and newly published taxonomic literature, facilitating discovery, analysis, re-use and export to relevant aggregators.
    - The Core Archive interchange format (developed for Scratchpads) was tested and used as a data sharing tool for data mined from legacy and prospective literature.

- **Semantic enrichment and integration of biodiversity data.** A set of use cases for semantic markup, recommendations and a roadmap were drafted. Use cases were addressed during the Biodiversity Data Enrichment Hackathon.
- **Automated registration of taxon names for publishers and registries.** An automated registration workflow was implemented to facilitate the process of taxon registration with the appropriate registries. The registration workflow is free to use. The XML query and response formats are publicly available. A version of the tool was developed to incorporate an accepted community standard.
- **Requirements of taxonomic** data providers and the major constraints.
- **Recommendations for the** sustainability of OBKMS.

4.1.2 Summary description of project context and objectives

4.1.2.1 Aim

The aim of pro-iBiosphere was to prepare, through a coordination action, the ground for an Open Biodiversity Knowledge Management System (OBKMS). The project investigated ways to improve access to biodiversity data, improve the efficiency of its curation and increase the user base of data and applications.

Biodiversity provides society with a wide range of ecosystem services which are essential to mankind's survival on Earth. A sustainable management and effective conservation of biodiversity requires international cooperation and open exchange of data through accepted community standards.

The core of biodiversity knowledge is composed of information on species and specimens. A wealth of such knowledge has been assembled in taxonomic literature (Floras and Faunas volumes, taxonomic revisions, monographs, inventories, research articles) over hundreds of years, by and in European natural history institutions, herbaria, botanic gardens, and increasingly through ongoing biodiversity monitoring programs. Taxonomic literature provides authoritative syntheses of the information that is available on a particular group of organisms (= taxon) in a particular region, compiled and evaluated by specialist taxonomists. It thus provides access to data of the highest quality available, including links to external resources upon which analyses are based. However, both external and internal factors call for an urgent modernisation of the production of and access to these data, information and knowledge.

External factors include the need for biodiversity data to support conservation decisions such as the conservation of a particular region or a particular taxon, and for mitigation of climate change impact and adaptation studies. Internal factors are a consequence of the new opportunities arising from the digital revolution, and of the need to reconcile the increasing amount of data that needs to be collected and curated with the decreasing numbers of taxonomists available for this type of work.

Only during the last three decades has core biodiversity data been produced in digital formats (i.e. on-line publications, DNA-sequences, images and metadata of specimens, GPS data of specimen records). Existing taxonomic literature is being digitised from printed copies by many institutions around the world. The Biodiversity Heritage Library Europe achieved substantial progress in coordinating and integrating these efforts in the EU. Semantic enhancements (markup) of this digitised body of literature is a new field of interest that will increase its accessibility. This requires coordination to align ongoing and future efforts to perform this semantic mark up, and to provide the community with technical solutions for the enhancement, use and re-use of these data.

Managing the quality of the data is already a major challenge. In many institutes, the curation workflow for producing core taxonomic information, such as Floras and Faunas, has not changed much over the years and is a time-consuming process usually performed by individual specialists (Figure 1). A strategy to adapt the methods of acquisition, curation, synthesising and dissemination of biodiversity data to the digital era is therefore of paramount importance (Figure 2).

4.1.2.2 Key challenges of the project

The project (i) identified legal and sociological challenges that hamper access to and re-use of biodiversity data to implement the OBKMS, (ii) addressed technical and semantic interoperability challenges between different forms in which data are published and originating in different Biota; and (iii) addressed sustainability issues related to the maintenance and curation of biodiversity data and derived information and knowledge.

An OBKMS will play a major role in facilitating the synthesis of core biodiversity data by creating a framework for the discovery of new species, the (re)naming of specimens and species, identification tools, descriptions, and various other basic types of information.

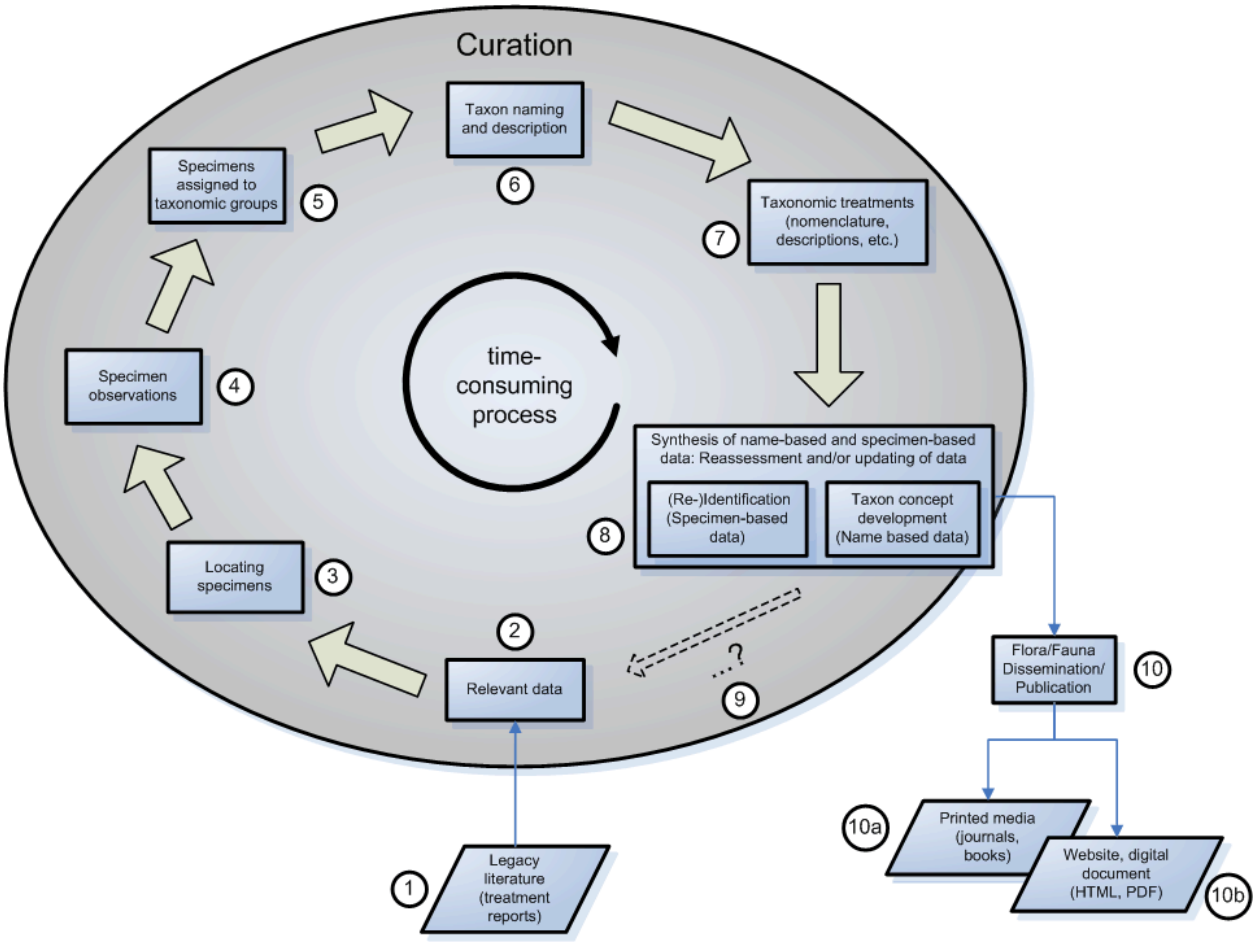

Figure 1. Traditional workflow for production of taxonomic information.
Legacy (i.e. existing) literature for a particular taxon is gathered (1) and relevant data are manually extracted (2). Legacy literature usually contains information that a scientist uses to locate the physical specimens that were used to describe a taxon (3). Specimens are studied and new observations are made (4), on basis of which, specimens are assigned to

taxonomic groups (5). Based on these observations taxa are distinguished (e.g. including where necessary new species), named and described (6). Taxonomic treatments describing the various taxa (families, genera, species, etc.) are written (7). Materials in collections are re-identified databases are updated as required (8). Due to the time-consuming nature of the process and the scarcity of experts, treatments are generally not regularly updated (9). Taxonomic treatments are published in journals or books specialised in taxonomy (10). Publications usually are of two types: print-only (10a), and/or online, usually consisting of HTML or PDF versions of the entire printed taxonomic treatments (10b), with only few enhancements. The print-only and online formats do not allow users to extract specific information in an easy or automated way. Typically, the whole process of producing major treatments with many taxa may take many months and up to 20 years or more in a large taxonomic group.

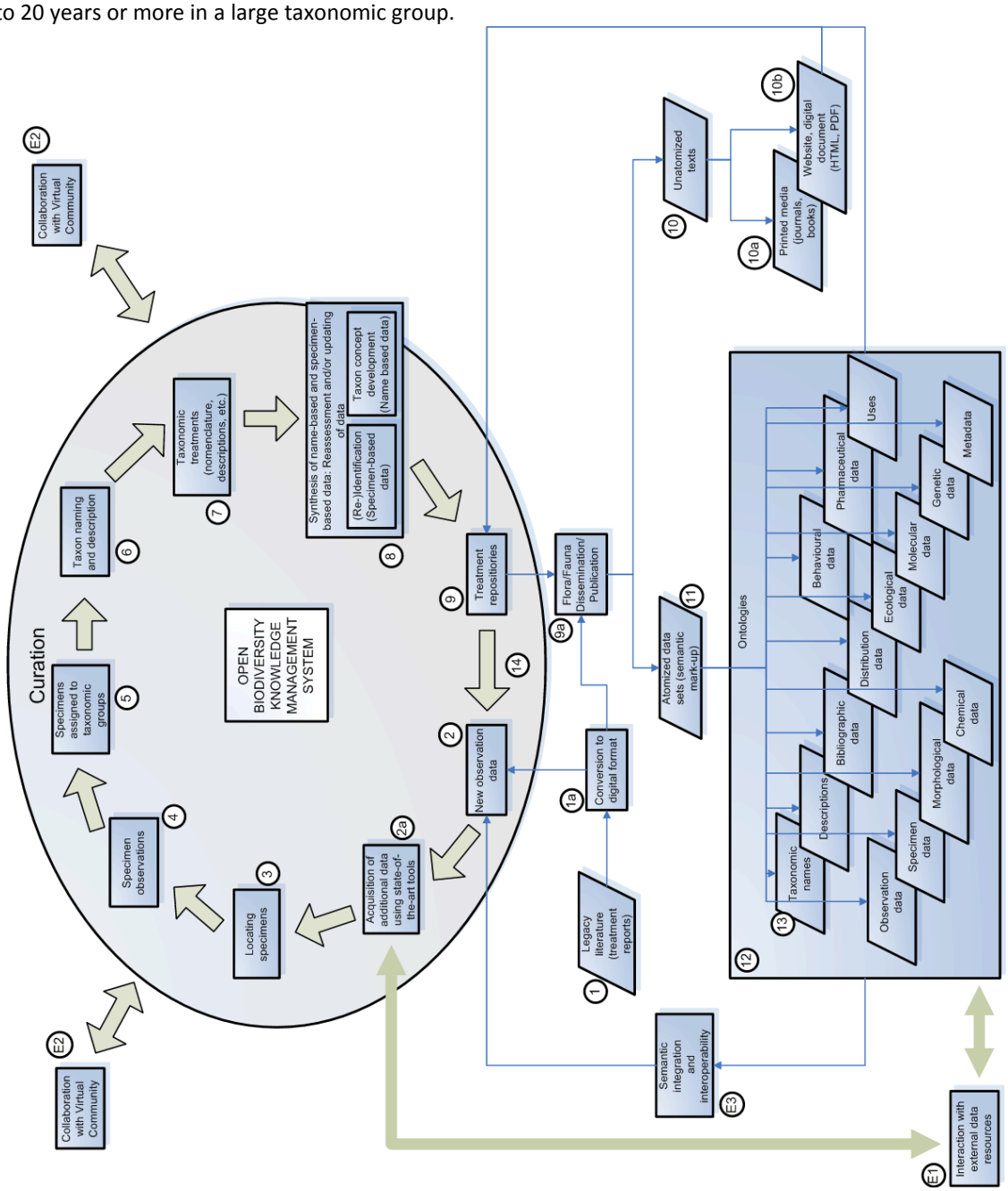

Figure 2. OBKMS workflow for production of core taxonomic Information.

Legacy literature (1) is digitised (e.g. by OCR) which makes further processing by computers possible. Steps (2)-(8) are essentially the same as in the traditional workflow, but can be assisted by tools that allow interaction and exchange with other, external, data resources (E1) to find additional information not present in the legacy literature (2a), and can be performed in real-time collaboration with other scientists (E2). Treatments are stored in an open access treatment repository (9). Taxonomic information dissemination and/or publication (9a) may follow one of two pathways: i) The traditional way *(10)* using printed media *(10a)*, or static websites and digital documents *(10b)*, and ii) dissemination/publication of atomised data sets, texts that use semantic markup for making their contents available to processing by both humans and computers and also made available in XML *(11)*. Ontologies are used to describe the data and its internal and external relationships *(12)*. Data come in many different types, each of which will be made available separately *(13)*. To the human reader, the information is presented in an integrated form similar to the original text. To computers, data exchange and semantic integration and interoperability has been made possible, both within *(E3)* the Open Biodiversity Knowledge Management System, and with data resources outside it *(E1)*. Furthermore, converted legacy taxonomic literature *(1a)* can be fed directly into the dissemination and publication process *(9a)*, again allowing a choice between unatomised texts and atomised datasets. This avoids the need of a full taxonomic revision before re-publication and is especially useful for legacy data the quality of which is uncontested. The curation cycle thus never stops; treatments can be updated on the fly, either using new information from specialist scientists *(14)*, information generated from within the semantic integration and interoperability process *(E3)*, or interaction with other data resources *(E1)*.

4.1.2.3 Objectives

The objectives of pro-iBiosphere were to:

- Coordinate towards and prepare the foundations for a viable long-term platform for knowledge management, data aggregation and integration;
- Provide new methods to synthesize distributed knowledge;
- Develop a strategy to adapt to the digital era the methods of acquisition, curation, and dissemination of core biodiversity data and derived information;
- Help to align ongoing and future semantic markup of taxonomic literature, and link elements of biodiversity literature to the original data;
- Promote and monitor the development and adoption of common markup standards and specifications for making biodiversity knowledge more accessible and re-usable;
- Provide the community with technical solutions for the enhancement and use of these data;
- Analyse and evaluate business models for supporting Open Science, and provide recommendations to achieve sustainable delivery of biodiversity information to target audiences;
- Develop and agree a shared data and IPR (Intellectual Property Rights) policy; and
- Promote and increase cooperation between the major biodiversity projects, initiatives and platforms at EU and global levels.

4.1.3. Description of the main S&T results / foregrounds

4.1.3.1 European and international policy coordination (WP2)

Coordination and routes for cooperation across organisations, projects and e-infrastructures

The framework of an Open Biodiversity Knowledge Management System was defined in the report "Towards a draft strategy for increased cooperation" (D2.1.2) as an e-infrastructure that will:

- Embrace existing digital infrastructures;
- Accommodate past publications, including regional or global monographs;
- Benefit from new enhanced publication processes that integrate data and narrative (text); and
- Manage the data elements that constitute taxonomic treatments, such as specimen data, images, DNA sequences, taxon names and their concepts, morphological characters, ecological and biological traits. The solution will involve semantic information management and use of wide accepted and community agreed systems of identifiers.

The following recommendations were made to further promote Open Biodiversity Data Management:

- **Recommendation 1:** That pro-iBiosphere implements the recommendation made in the report on "Ongoing biodiversity related projects, current e-infrastructures and standards" (D2.1.1) to develop a Memorandum of Understanding (MoU) to express the commitment of the participants to OBKM. We recommend this takes the form of a declaration that individuals, teams, initiatives, and institutions can support.
- **Recommendation 2:** That pro-iBiosphere promotes the free and open use of content, services and other resources by adopting, where necessary, licenses that grant all users - including automated tools - a free, irrevocable, worldwide, right of access to copy, use, distribute, transmit and display the work and data publicly and to make and distribute derivative works, in any digital medium for any responsible purpose; using community conventions rather than copyright to achieve proper attribution. This recommendation also notes that the providing institutions may also offer commercial services based on the data where appropriate to cover costs of production, maintenance and future development.
- **Recommendation 3:** The assignment, use and process of managing identifiers should be given very high priority, as this will promote widespread use of persistent, dereferenceable identifiers for physical and digital data objects such as specimens, images and taxonomic treatments as well as their metadata representations.
- **Recommendation 4:** To register content and services, and to explore the option of adopting existing facilities for this purpose, such as the BioVel BiodiversityCatalogue (https://www.biodiversitycatalogue.org/).
- **Recommendation 5:** Design and implement a system for tracking the use of any and all elements of information to ensure that sources and suppliers of data are assigned credit for their contribution to the creation and supply of data.

- **Recommendation 6:** Establish agreements on specialisation in services (example: one institution specialises in geographical analysis, another in visualisation tools), to facilitate providing services to other institutions or projects.
- **Recommendation 7:** Establish multi-institutional OBKM working parties to pursue issues relating to collaboration, technical requirements, implementation schedules and sustainability for OBKM. The pro-iBiosphere team will be well suited to coordinate the dialogue that will refine the concept, priorities and technical requirements of OBKM.
- **Recommendation 8:** The pro-iBiosphere project to work with OBKM working parties and to make use of existing mechanisms such as Biodiversity Information Standards TDWG (Taxonomic Database Working Group), to establish technical requirements for standards, vocabularies and protocols for OBKM, to improve access to, and linking and use of, open data (Task 3.3); and to identify implementation processes and priorities; facilitate automation of these processes; and use existing standards as far as possible (Tasks 4.1 and 4.2).
- **Recommendation 9:** Establish an open mechanism for the election of an advisory and management board for the OBKMS to complement the current members of pro-iBiosphere and its board.
- **Recommendation 10:** OBKM working parties to work together to identify funding, using existing and new sources, to implement the OBKMS.

Following Recommendation 1, negotiations to solicit support on a declaration started early February 2014. The first draft of the Declaration was circulated in April 2014 among pro-iBiosphere partner institutions, and other non-consortium members, including the Consortium of European Taxonomic Facilities - CETAF, NSC-USA, and Univ. Harvard. This activity helped to develop consensus. The comments by interested parties were integrated in the wording of the Declaration. It was decided to have two categories of signatories (organisations and individuals), and to broaden the target group beyond the natural history museum community to everybody with interest in, use of, production of, or keepers and curators of biodiversity data. A dedicated Website (http://bouchoutdeclaration.org) was designed to be fully operational at the launch of the Declaration. The Site includes the sections: Declaration, Background, Frequently Asked Questions, Signatories, 'Sign'; access to copies in 9 additional languages, a Twitter feed, a news section, and a contact section to submit questions.

The Declaration was successfully launched in June 12 2014 during the Final Meeting of pro-iBiosphere preceded by a panel discussion on the impact of the Declaration. By the end of the launch day, 57 organisations and 58 individuals of 28 countries had signed the Declaration (Figure 3). During its few months of existence it has been broadly endorsed by the biodiversity community and beyond.

During initial discussions various institutions and individual scientists initially rejected the notion of Open Access. During the implementation phase it was very encouraging to see that many directors or presidents of organisations signed. Ongoing discussion between MfN, Pensoft, Plazi and FUB-BGBM are taking place on how to foster the adoption and implementation of the principles by the signatories and how to measure the impact of the Declaration. Plazi and the MfN have committed to continue to coordinate the Declaration in the future. A possible model to follow is the Berlin Declaration on Open Access to Knowledge in the Sciences

and Humanities (http://openaccess.mpg.de/Berlin-Declaration). Possible actions under discussion are, to: organise biannual follow up meetings; further promote the Declaration; and follow-up the implementation of the Declaration through indicators/metrics. Ultimately, the real impact of the Declaration will be the creation of new knowledge that is free, discoverable and integrated.

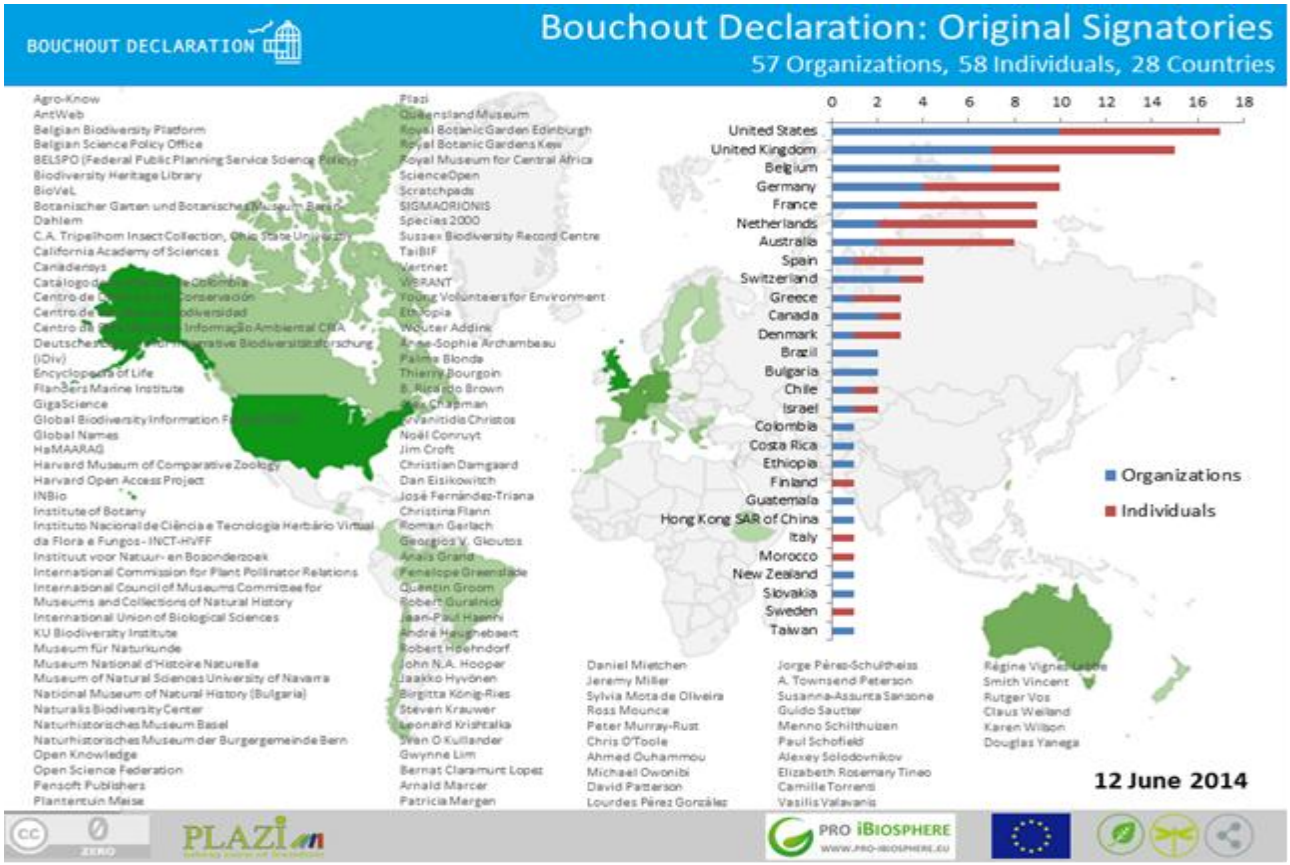

Figure 3. Visualisation of the signatories of the Declaration at the launch day.

By October 14 2014 the Declaration was signed by 173 individuals and 85 organisations from 46 countries around the globe (see here). The list of organisations extends well beyond the Natural History Museums community, it includes a wide array of individuals and organisations (e.g. those dealing with aggregation of data such as GBIF, DataOne; scientific institutions; associations with a wide array of interests such as the Open Knowledge Foundation, Creative Commons, International Union of Biological Sciences) (Figure 4).

The Declaration, with its very broad spectrum of signatories, is a strong indication that the biodiversity community is willing to collaborate by opening up their data, making them discoverable and providing standardised ways to access them. It is also shows the commitment by institutions to invest in a sustainable biodiversity e-infrastructure.

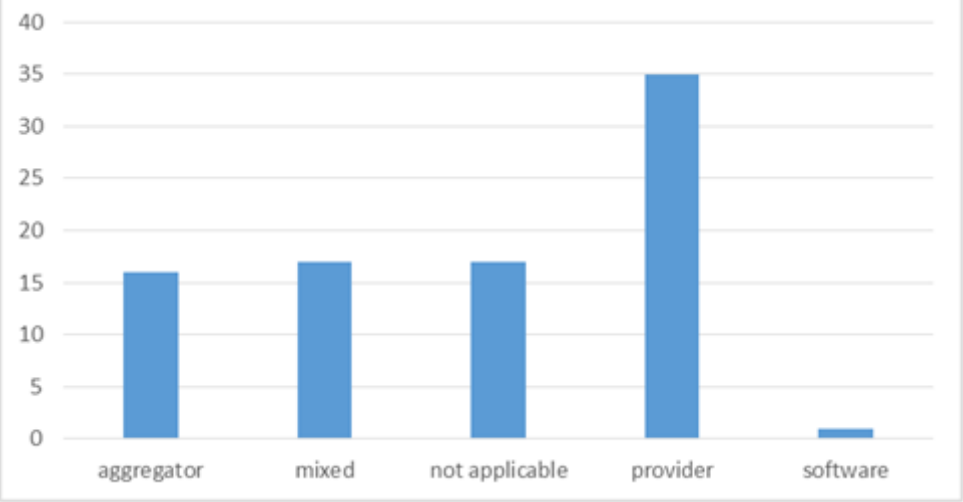

Figure 4. The nature of the organisations.

Stakeholder requirements

The project reviewed how information held in Biotas (i.e. monographs such as Faunas, Floras and Mycotas) is used by a variety of audiences. The methods used to gather evidence included interactive workshops, pre-workshop questionnaires, follow-up interviews and desk based research. The main audiences surveyed were taxonomists, informaticians, conservationists, ecologists, publishers and IT developers who routinely handle Biota information. Distribution, morphology, habitat and taxonomy are the most commonly used information types. The major constraints to users of Biotas and the information they hold were:

- The time needed to locate the relevant source work and to retrieve the information and present it in a more appropriate format for a particular reuse.
- The lack of easy to use technical solutions for mining and extracting data currently in Biotas for reuse in other products or activities.
- Extraction of atomised information by markup is time consuming, technically difficult and potentially very costly. However, such data underpins all further analysis.
- Interfaces to data marked-up and atomised from Biotas need to be easy to use and follow data standards where appropriate.
- Access to information is limited due to:
    - the limited access to the internet in some countries and segments of society
    - the costs of access to data providers who charge for their services
    - IPR restrictions imposed by data providers.
- Difficulties in interpreting the data. There is a strong requirement for human expertise to access or interpret Biota information for activities such as identification or validation of data.
- People (and their expertise) are a valuable resource but as there is no standard index of who is available, significant time is spent locating experts. User feedback channels are poor and this

hampers understanding of user requirements. The anonymity of users of some online systems contributes to this.

- Users of Biotas have requirements for information traditionally not published in Biotas but that is closely related and potentially available to Biota producers. Significant data gaps (such as species abundance) and monitoring could be resolved and undertaken by citizen scientists and local enthusiasts provided they have easy access to existing information and means to add that information to the online Biota.
- Identification of species is a major activity both in the field and lab. The kinds of tools needed to help with this varies from low-tech to high tech depending on the particular situation. Where appropriate, these tools should be designed to integrate with local expertise.

Detailed information is available in the report on "User Feedback".

Enhancing general knowledge on biodiversity tools

Taxonomists provide all the names and organise information about the life on Earth. The project documented the taxonomists' perspective on the requirements of and obstacles to a fully digital workflow for taxonomic publication. It also examined the strengths and weaknesses of existing software solutions for taxonomic publication; the obstacles to the adoption of these tools by taxonomists; and some of the benefits for taxonomists who adopt a digital approach.

There is considerable variety in biodiversity informatics software, much of which is created by the taxonomists themselves for individual projects. International standards for data fields and data exchanges are patchy, and there are few examples of a seamless flow of data from one system to another. Knowledge of biodiversity informatics systems among taxonomists is poor, and there is still considerable scepticism within the community regarding electronic publication and open access to literature. Nevertheless, there are examples where well-designed software and strong community groups have combined to create exemplary databases of taxonomic information.

Detailed information is available in the report on "The Use of e-Tools among Producers of Taxonomic Knowledge (Proof of concept report on the use of e-tools). The report is based on a workshop organised by the pro-iBiosphere project in May 2013 attended by about 100 people; a questionnaire distributed among taxonomists and related professionals, which received 220 responses; a literature review; and direct conversations with taxonomists.

The report makes the following recommendations:

- **Recommendation 1:** Focus on usability and interoperability of software, not just functionality.
- **Recommendation 2:** Promote information technology within the taxonomic community.
- **Recommendation 3:** Provide clear direct benefits in software for taxonomists, as well as downstream users.
- **Recommendation 4:** Set a realistic minimum level of IT-literacy that is necessary to function as a professional taxonomist, and to incorporate that level into curricula and professional training.
- **Recommendation 5:** Tackle the social obstacles to IT adoption, such as the disconnect between taxonomists and the users of taxonomy.

- **Recommendation 6:** Stop using journal impact factors to assess the value of taxonomic works, but use measures such as book sales and web hits, which reflect the value of the work to users, rather than to other taxonomists.
- **Recommendation 7:** Raise the profile of descriptive standards within taxonomy. For example, by creating translations for the Taxonomic Database Working Group (TDWG) world geographical scheme for recording plant distributions.
- **Recommendation 8:** Mandate institutional data archival policies.
- **Recommendation 9:** Research data has to be open access.
- **Recommendation 10:** Ensure firm, long term commitment of institutions to digital taxonomic infrastructure.

Legal issues of data acquisition, curation and dissemination

An analysis was made of legal barriers to scientific research in European copyright and database protection legislation; and recommendations were provided in the report "Draft policy on Open Access for data and information" (D2.4.2). The investigation that started in Y1 (including France, Germany and Sweden, see D2.4.1) was extended to three other EU member states (Denmark, Italy and UK) and to two non-member states (Norway and Switzerland).

The report includes the following recommendations:

A. At the level of fully or partially funded public institutions

- Publicly funded institutions should refrain from claiming intellectual property rights for biodiversity data and information collected and/or published by them. By default, all content referring to names and taxonomic information should be openly accessible.
- Publicly funded institutions should encourage re-use of biodiversity data and information for research purposes with a requirement for attribution of the source, but should impose no other requirements on re-use.
- As far as material owned by a publicly funded institutions is protected by copyright or by database rights, the institutions should dedicate these works or databases to the public domain by publishing them under a CC0 or similar License.

B. At the level of the European Union and the European Economic Area

- The EU should revise the Directive 2001/29/EC by declaring that the provision of a copyright exception for scientific research is compulsory for all member states. The new regulations should not refer to commercial or non-commercial scientific research as this distinction is neither useful nor applicable in practice. Nor should they refer to the place from where, nor the technical mode how, works are accessed as such restrictions hamper the research process.
- The EU should revise the Directive 96/9/EC by declaring that the re-use of protected databases for scientific research is authorised by a compulsory exception to database rights.

C. At the level of Member States of the European Union or the European Economic Area

- Member states of the EU or the EEA should introduce or, where it already exists, extend a copyright exception for the use of works for scientific research. This exception should not refer to commercial or non-commercial scientific research as this distinction is neither useful nor applicable in practice.

Nor should it refer to the place from where, nor the technical mode how, works are accessed as such restrictions hamper the research process.

- Member states of the EU or the EEA should introduce or, where it already exists, extend an exception of database protection for the re-use of databases for scientific research.

Findings were further reported in the following publications:

- Egloff W. et al. 2014. Open exchange of scientific knowledge and European copyright: The case of biodiversity information. ZooKeys 414, 109-135. DOI: 10.3897/zookeys.414.7717
- Patterson DJ. et al. 2014. Scientific names of organisms: attribution, rights, and licensing. BMC Research Notes 2014, 7:79. Doi:10.1186/1756-0500-7-79

4.1.3.2 Scientific content and workflow coordination (WP3)

Semantic integration of biodiversity literature

In recent years, a number of perspectives on the near future of biodiversity informatics have been put forward, including the Global Biodiversity Informatics Outlook (GBIO) report, the "Decadal view of biodiversity informatics" white paper, and the differences between the biodiversity literature and biological databases are steadily diminishing.

Now, the focus has to be on the ultimate goal of a new, capable, efficient OBKMS for biodiversity information, one that transparently leads from original data to reviewed conclusions, provides full version history as well as attribution, and is semantically interlinked between different knowledge domains. In order to achieve this, constraints on openness of biodiversity data need to be tackled.

Lessons learned from the Biodiversity Heritage Library (BHL) Europe initiative, other efforts towards mass digitisation of biodiversity literature, and the experiences with markup methodologies from Plazi, Pensoft, Naturalis, RBGKew, were used by pro-iBiosphere to (i) evaluate the state-of-the art in semantic integration of biodiversity literature; (ii) report on available tools, processes and web services and (iii) assesses the feasibility of semantically enhancing legacy literature and integrating these different kinds of information into coherent knowledge management workflows with current technology. Detailed information is available in the report on "State-of-the-art and research horizons of semantic integration of biodiversity literature" (D3.3.1).

Three viable paths for future improvement for semantic enrichment of legacy literature were considered:

- Fully automated natural language processing (NLP);
- Basic markup complemented by semi-automated processing and specialist correction; and
- Social crowd-sourcing models (citizen involvement).

Setting priorities towards a semantically-aware OBKMS. A set of use cases for semantic markup of biodiversity literature was used to test different approaches, make recommendations, and to draw up a priority list for future activities.

While significant and promising progress has been made along all the three viable paths, significant obstacles remain to be overcome before these techniques can be deployed at the scales required for the envisaged OBKMS. In parallel, institutional awareness for opportunities and challenges related to semantic enrichment has risen to a point where multiple parties, beyond the pro-iBiosphere partners, are actively working on different aspects of the generation and consumption of markup.

The pro-iBiosphere international workshop on "Markup of Biodiversity Literature" brought together key players from the biodiversity markup community. They discussed how semantic enrichment could be scaled up beyond individual articles or partners.

The following use cases were presented by participants of the workshop and discussed during the event:

- Data integration across multiple sources;
- Repurposing and redistribution;
- Metadata extraction and processing;

- Facilitating data mining; and
- Disambiguation and machine reasoning.

The following recommendations were made to achieve semantic integration of biodiversity literature:

- **Recommendation 1:** Semantic enrichment of selected revisionary works, capturing the most recent knowledge efficiently, rather than targeting the taxonomic literature as a whole.
- **Recommendation 2:** Differentiate between different use cases for semantic markup and annotations and choosing appropriate resource investments, with basic markup being the default and fine-grained markup reserved for special cases.
- **Recommendation 3:** Invest in the re-invention of publishing workflows, such that they automatically produce semantically integrated publication forms.
- **Recommendation 4:** Work towards an integrative and semantically enabled Open Knowledge System that integrates well with past knowledge, while allowing for new agile and collaborative approaches to publishing and curating, possibly building on the Wikidata model.

Roadmap and the first elements of a workplan. The project identified three paths that have to be pursued in an increasingly coordinated fashion:

- **Path 1: Publishing new knowledge in a semantically aware system.** This will be a collaborative and distributed system, and its implementation and maintenance will be a permanent activity requiring stable funding. This funding could be supplied by institutions with the commitment and capacity for long-term knowledge preservation commitment (museums), external funding bodies, or the private sector.
- **Path 2: Cost-efficient, largely automated basic enrichment of large - and especially revisionary - parts of the legacy literature with the primary goal of increased discoverability of taxon names, person names, geolocations, or citations.** This will be an ongoing activity requiring funding until the need for enrichment of legacy publications is saturated.
- **Path 3: Fine-grained markup of selected literature with the goal of seeding the semantic knowledge management system**. These activities will have to promote a transition from the currently very large-scale but rare publications of Faunas and Floras to a more agile system. The focus here has to be on efficient operation for a purpose rather than just-in-case operations. Capturing the newest taxon treatments in semantically enabled formats is more important than capturing the complete history of taxonomic publishing in the same way.

For most of the components of an OBKMS some basic functionality exists, but not at or near production level. Some of these gaps were investigated in the pro-iBiosphere pilots conducted in the framework of work package 4.

The project contributed to make progress towards two of these paths, mainly, “Publishing new knowledge in a semantically aware system (Path 1)” , and “Cost-efficient, largely automated basic enrichment of large - and especially revisionary - parts of the legacy literature with the primary goal of increased discoverability of taxon names, person names, geolocations, or citations” (Path 2).

A Best Practices for stable URIs document was created and is being updated by various stakeholders around the globe on a regular basis. A joint pro-iBiosphere & CETAF-ISTC initiative on this topic resulted in a pilot system for stable HTTP-URI based identifiers for collections. The system was further advertised by pro-iBiosphere and has been proposed as an official product of the Consortium of European Taxonomic Facilities (CETAF). The decision will be ratified during the general CETAF-meeting in Vienna (taking place in October 2014). The pilot system for stable HTTP-URI based identifiers for collections is also the basis for a EU proposal recently submitted addressing the implementation of a Linked Open Data infrastructure for Biodiversity Information including collection data, observations, literature, sequences, descriptive data, and taxa.

Plazi implemented a complementary HTTP URI identifier scheme for treatments allowing treatments to be linked and providing direct access to their content. Furthermore, the transformation processes between the document-centric Plazi repository and the EDIT Platform for Cybertaxonomy were equipped with several functions for automated parsing and semantic enrichment of free-texts as well as improved feedback of text strings requiring decisions from users conducting the markup.

During the pro-iBiosphere Biodiversity Data Enrichment Hackathon, biodiversity data were mobilised from their silos and enriched with meaningful links to related resources, such as links from taxon names to taxon concept URIs; links from described habitats to environment ontologies; links from character traits to trait ontologies; links from species treatments to relevant images, publications and specimens. The workflow tools and data publishing platforms that operate on such enriched data were enhanced to provide greater interoperability and data integration functionality.

Nine breakout groups addressed three main themes: i) mobilising heritage biodiversity knowledge; ii) formalising and linking concepts; and iii) addressing interoperability between service platforms (see Table 1). Prototype solutions for the nine use cases were developed, and areas of inadequacy were discussed and are being pursued further beyond the duration of the project. Technical solutions have been broadly documented in the scholarly publication here available: Doi: 10.3897/BDJ.2.e1125 and presented during the pro-iBiosphere final event through demonstrations.

Beyond deriving prototype solutions for each of the nine use case, areas of insufficiency were discussed and are being pursued further beyond the project duration. It was striking to see how many possible applications for biodiversity data there were and how quickly solutions could be put together when the normal constraints to collaboration were broken down for a week. Conversely, mobilising biodiversity knowledge from their silos in heritage literature and natural history collections will require formalisation of the concepts (and the links between them). The concepts define our research domain as well as increased interoperability between the software platforms that operate on these concepts.

The tangible outcomes of the Hackathon are sustainable homes in the appropriate code bases, registries and repositories, and proofs-of-concept applications for scientific publications and project proposals. The main intangible outcomes of the event are the fostering of a community of experts in biodiversity informatics and the strengthened human links between research projects and institutions. The event also demonstrated the ongoing need for data normalisation and integration, for example through the application of ontologies, as well as the opportunities for innovative research such integration will afford.

Table 1. pro-iBiosphere Biodiversity Data Enrichment Hackathon Themes and Aims

MOBILISING HERITAGE BIODIVERSITY KNOWLEDGE	
Biodiversity data analytics	Extract statistical data about specimens for visualisation in a dashboard.
OCR correction	Provide a simple interface for interactively editing of OCR'd text, as well as tools to track the edits, to provide feedback to improve the OCR.
Open access images	Liberate and showcase openly-licensed (e.g. CC-BY) images from journal article PDFs and republish on image sharing social media sites. Find images of phylogenetic trees for data re-extraction from the image.
FORMALISING AND LINKING CONCEPTS	
Trait ontology	Extract and ontologise plant trait data from digitised Floras.
SWeDe	Produce a standard for describing scientific web services.
Specimen links	- Link together name and specimen data, especially from Floras. - Link specimen citations in "Literature" to specimens from Kew, Brussels and Edinburgh. - Proof of concept and requirements gathering for Taxonomic MindMapper.
SERVICE PLATFORMS	
EDIT Platform Common Data Model API	Develop a web service to extract occurrences out of CDM instances (EDIT platform).
iPython notebook/Taverna	Access Taverna workflows from within iPython notebook.
BioVeL/NeXML services	Deliver RESTful services to merge and query phylogenetic data and metadata.

4.1.3.3 Technical and infrastructure coordination (WP4)

pro-iBiosphere addressed technical cooperation and interoperability at the e-infrastructure level, and promoted the development and adoption of common markup standards and interoperability between schemas. This resulted in the creation of a first technical strategy on the interoperability of biodiversity data in the future framework of an Open Biodiversity Knowledge Management System (OBKMS), Figure 5.

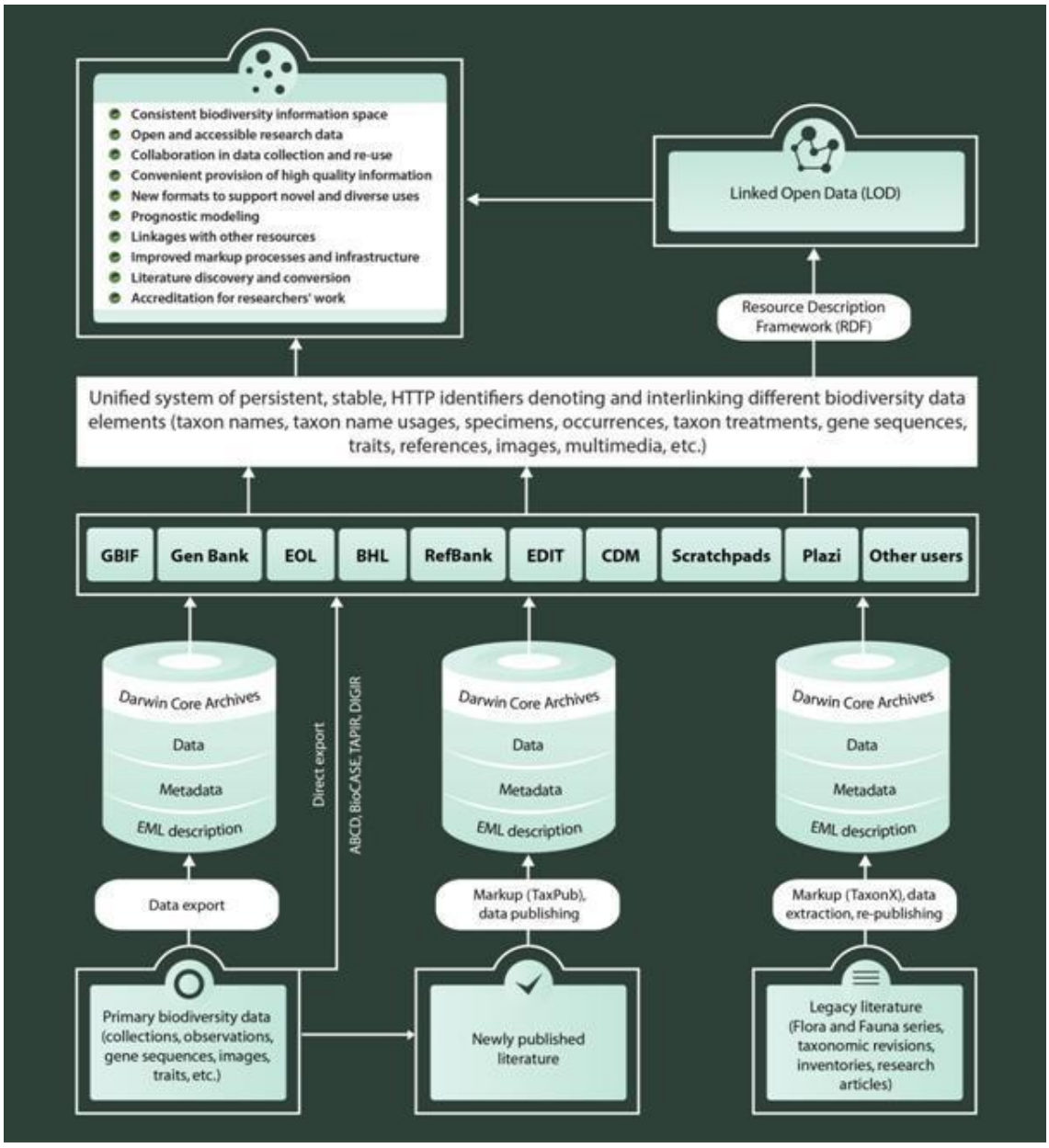

Figure 5. Interoperability between e-infrastructures and data standards in the future Open Biodiversity Knowledge Management System (OBKMS).

Improving the technical cooperation and interoperability at the e-infrastructure level

pro-iBiosphere identified four major issues hindering true interoperability at the e-infrastructure level:

- Competing standards;
- Ambiguous, poor or absent documentation of standards;
- Lack of stable identifier systems; and
- Absence of semantic interoperability.

In order to avoid duplication of effort with other ongoing initiatives (e.g. Taxonomic Databases Working Group (TDWG) Biodiversity Information Standards, LifeWatch), pro-iBiosphere focused on two topics having the potential to make significant progress during its two year lifetime, in line with the recommendations of the Bouchout Declaration. These topics were:

- Implementation review and roadmap of collection identifiers;
- Service registries and the BiodiversityCatalogue.

The pro-iBiosphere workshop on "How to improve technical cooperation and interoperability at the e-infrastructure level" reviewed the topics thoroughly and planned a roadmap for further implementations (for additional information see here). Findings have been documented in the report "Strategies for improved cooperation and interoperability between infrastructures" (D4.1).

Roadmap of collection identifiers: In cooperation with the Information Science and Technology Committee of the Consortium of European Taxonomic Facilities (CETAF-ISTC), pro-iBiosphere defined and documented a Linked Open Data compliant system for using HTTP-URIs as stable identifiers for collection objects and their associated metadata (see Best Practice Guidelines).

The project piloted a working identifier system for adoption by the wider community. The system has been successfully implemented by 9 institutions in Europe and the USA:

- Botanischer Garten und Botanisches Museum Berlin-Dahlem, Germany;
- Harvard University Herbaria, US;
- Harvard Museum of Comparative Zoology, US;
- Muséum National d'Histoire Naturelle, Paris, France;
- Museum für Naturkunde, Berlin, Germany;
- National Botanic Garden Belgium, Belgium;
- Naturalis Biodiversity Center, the Netherlands;
- Royal Botanic Gardens Kew, UK, and;
- Royal Botanic Gardens Edinburgh, UK

The successful and rapid implementation of HTTP-URIs by several European and US collection institutions provides a convincing model for other institutions to follow. By adopting Linked Open Data compliant technical mechanisms, the solution complies with the vision of Open Biodiversity Knowledge Management, demonstrating the benefits of discoverability and interoperability of data.

Service registries and the BiodiversityCatalogue: The 7th Framework project BioVeL implemented a global registry of biodiversity-related services to help potential users discover and understand service functions, interfaces, and behaviour; and to foster the growth of service-based applications and workflows. The

catalogue provides unified and rich metadata records for each service registered, a structured service classification system, search and browsing mechanisms, as well as basic service monitoring capabilities. In addition, the catalogue is a service by itself, so that machines can search for services and their access points. It is part of the MyExperiment platform that supports developers and users of Taverna-based scientific workflows. However, the core functions of the catalogue are not bound to their deployment in workflow environments.

The participants of the workshop concluded that:

- The Biodiversity Catalogue had the necessary functions needed for a global registry of biodiversity-related services;
- It was ready for use and service providers should register their services as soon as possible;
- Metadata standards and exchange protocols will need to be added for synchronisation with other initiatives; and
- With a growing number of services being registered, the Biodiversity Catalogue will need curation and quality control.

pro-iBiosphere made recommendations for enhancements that will secure the registry of biodiversity-related web services at the heart of a future Open Biodiversity Knowledge Management System (OBKMS).

In the course of 2014 web-services produced by the project to increase the technical and semantic interoperability were registered in the Catalogue, for example, occurrence services providing access to point location data captured in pro-iBiosphere taxonomic pilots.

Interoperability between biodiversity-related e-infrastructures: workflow between Plazi – CDM-based EDIT Platform of Cybertaxonomy

The project explored interoperability between biodiversity-related e-infrastructures with the development of a pilot workflow between the Plazi treatment repository and the CDM-based EDIT Platform of Cybertaxonomy. The workflow uses markup standards and tools that enable the linking of data from 3 different sources: (1) legacy literature, (2) prospectively published literature, and (3) unpublished data, onto a common data management platform, the CDM EDIT data store (see Figure 6). The pilot required transformations between XML-repositories and CDM-stores. Seven sub-pilots (markup of taxonomic treatments of mistletoes, fungi, ants, bryophytes, centipedes (*Chilopoda*), spiders, and for the plant genus *Chenopodium* were conducted to validate this pipeline.

The exercise showed that the mappings of semi-structured XML based text data (here TaxonX) into a fully structured data model (CDM) must be kept flexible in terms of:

- Completeness of markup;
- correctness of markup;
- level of atomization; and
- incompatibilities.

The exercise also demonstrated the feasibility of the concept and improved technical and "human" interoperability. The pilot involved close cooperation between taxonomists, biologists, bioinformaticians and computer scientists, and resulted in an evolving workflow that can be used for the markup of legacy data

from historical literature. Thus, legacy data joins other data within a common data model platform, and can be used for the generation and mobilisation of new knowledge.

To inform experts carrying out taxonomic markup about different granularity levels and their consequences on portal and service functionalities, a comprehensive table based user guide has been created.

Depending on different project requirements, it is now possible to import data with very basic markup and still get reasonable results in CDM portals and services. For instance, marking up coordinates for a specimen will allow showing these specimens on a map while the markup can be omitted if this is not a requirement. Similarly, marking up all parts of a name, including original citations and name status, will allow sophisticated search and structured and prettily formatted synonymies for online and printed publication, while for projects with a different focus a basic markup with no further atomization in the nomenclatural section will lead to a fully acceptable "free text" synonymy with only basic search and formatting functionality.

The source code for all implementations is freely available from the FUB-BGBM subversion repository at http://cybertaxonomy.eu/cdmlib/source-repository.html.

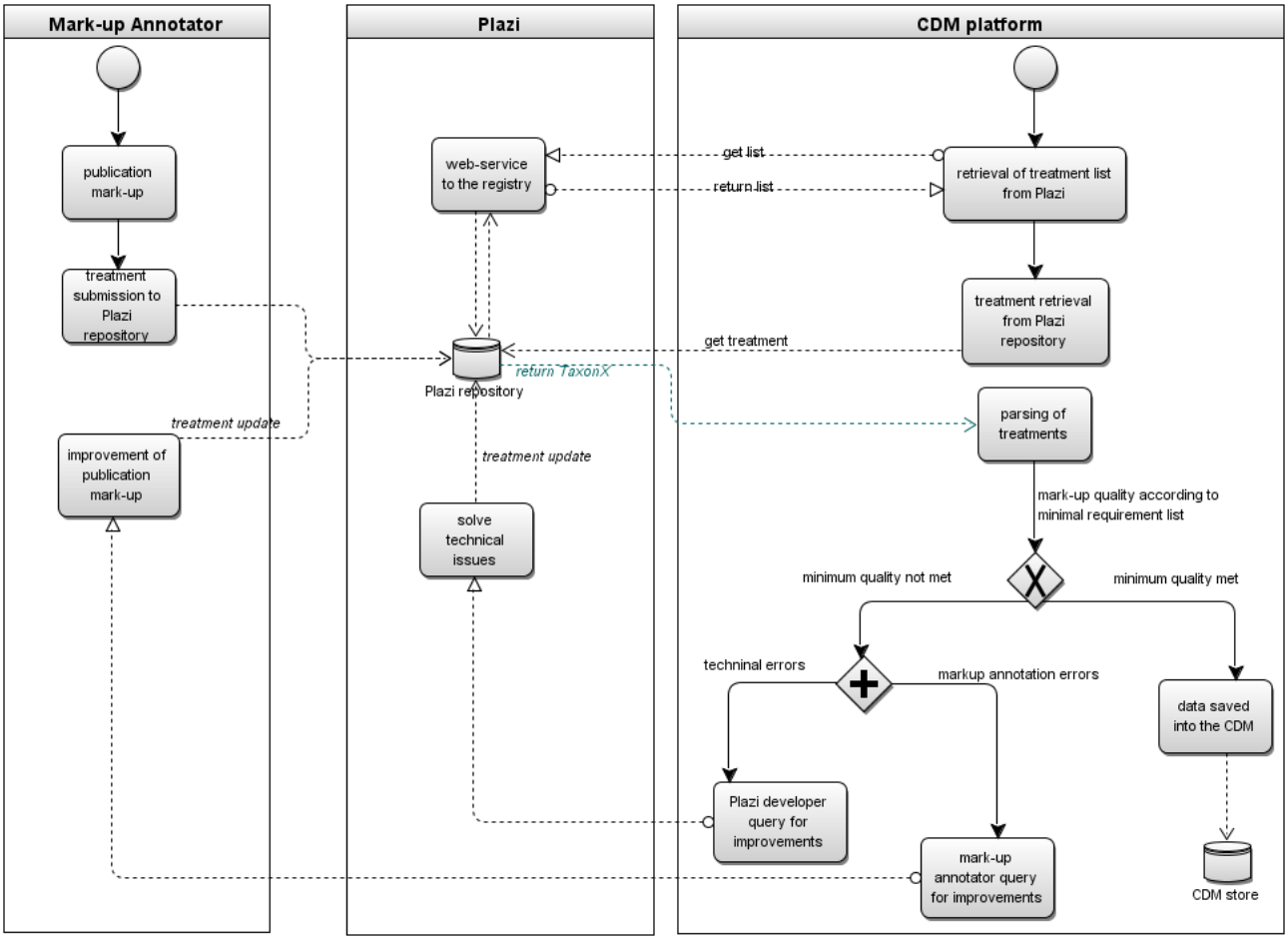

Figure 6. Workflow for taxonomic literature markup and import into Plazi and CDM information systems.

Promoting and monitoring the development and adoption of common markup standards

In order to promote and monitor the development and adoption of common markup standards, the project conducted three pilots (drawing on a broad spectrum of organisms: i.e. animals, higher plants, fungi and bryophytes) and used the outcomes to analyse the currently applied XML schemas for markup of taxonomic information. A workshop on "How to promote and foster the development & adoption of common markup standards & interoperability between schemas" (held at the FUB-BGBM in October 2014) that brought together 20 participants involved in pilots in WP4 to assess their results of the pilots and discuss a strategy for development and adoption of common markup standards. The results of the discussions and conclusions are reported in the report "Strategy for improvement & interoperability of the XML schemas (D4.2). The meeting notes can be found here.

The project made the following additions and improvements to the e-infrastructure that is available within Europe:

- Adoption of the TaxPub Journal Article Tag Suite (JATS) to serve as a standard system for the semantic markup of publications in the area of biological systematics. JATS is now being used in the Biodiversity Data Journal and all other Pensoft journals.
- Development of an Extensible Stylesheet Language Transformations (XSLT) mechanism was developed to inter-convert content from TaxPub and TaxonX.
- Development of a common XML query model to automatically register names in nomenclatural registries such as the International Plant Name Index (IPNI) for plants and ZooBank for animals.
- Adaption of the Darwin Core Archive to transfer data within literature among Pensoft, Plazi, GBIF and EOL.
- Development of a proof-of-concept workflow to link publishers (in this case Pensoft), repositories (in this case Plazi), and dedicated taxonomic platforms (in this case CDM).

The project also:

- Explored and established feasibility of the use of Resource Description Framework (RDF) for modelling and transfer of information in taxonomic treatments.
- Launched a workflow for re-publication of Floras and Faunas in an advanced, semantically enhanced, open access platform, using Flora Malesiana as an example. The workflow achieved interoperability between the FlorML and TaxPub schemas and demonstrated multiple uses of the investment of effort in markup and data extraction.

The results show that markup requires substantial effort and human resources - the most important limiting factors in achieving a sufficient degree of interoperability. The project concluded that a strategy to improve machine-based dialogue of semantic enhanced biodiversity information relating to taxonomic treatments should contain at least the following elements:

- ***Recommendation 1:*** *Adoption of TaxPub Journal Article tagging Suite (JATS) to serve as a standard* system for the semantic markup of newly published journal articles in the area of biological systematics.
- **Recommendation 2:** adoption of TaxonX for markup of legacy literature, and investment in the further development that will overcome current limitations.

- **Recommendation 3:** development of instruments (e.g. XSLT transformations) for lossless conversion among environments such as TaxPub and TaxonX that will ensure integration of data from historical and recent literature.
- **Recommendation 4:** development of industrial strength digitization workflows that will format the corpus of legacy literature into useful machine-readable content.
- **Recommendation 5:** endorsement of Darwin Core Archive as an interchange format for occurrence data and taxonomic treatments.
- **Recommendation 6:** utilisation of RDF as a format for distribution of data extracted from the content of marked-up treatments.
- **Recommendation 7:** substantial investment in markup tools with improved user interfaces and efficiency so as to achieve high accuracy and reduced labour costs.

The outcomes of three pilots are summarised below. For additional information see "Strategy for improvement & interoperability of the XML schemas" (D4.2) and "Report on ongoing biodiversity related projects, current e-infrastructures and standards" (D4.1).

Pilot on Common query/response model for automated registration of higher plants (International Plant Names Index, IPNI), fungi (Index Fungorum, MycoBank) and animals (ZooBank)

To ease the process of taxon registration, Pensoft and the registries Zoobank and IPNI successfully developed, tested and implemented an automated workflow for registration with IPNI and ZooBank. The registration process proposed in the workflow is "journal-centric" (Figure 7).

In order to make the registration tool more generally available, a version of the tool was developed by Royal Botanic Gardens Kew to incorporate an accepted community standard: the Taxonomic Concept Schema. A pilot on a general purpose registration tool was released at the end of August 2014 and is currently being tested by members of the Special Committee on Registration of Names of the International Association of Plant Taxonomists prior to release for broader community testing.

In addition to this, ZooBank and IPNI created interfaces for registration of new taxa by any user.

There are several reasons to maximise automation of registration, the most significant being:

- Increasing cases of bulk descriptions of new taxa within a single paper, sometimes counted in hundreds, which creates significant overhead on the authoring and editorial process;
- Decreasing the risk of errors caused by human intervention (e.g., re-typing);
- Disambiguation of the dates of acceptance and publication of a manuscript; and
- Efficient and accurate validation of final published data and metadata through automated export, from the publisher to the registry on the day of publication.

The registration workflow, formats and schemas will be published in 2014 in a special issue of ZooKeys devoted to nomenclatural and Codes-relevant issues. It is expected that the envisaged publication and accumulated experience will allow easy adoption and use by all interested journals and publishers.

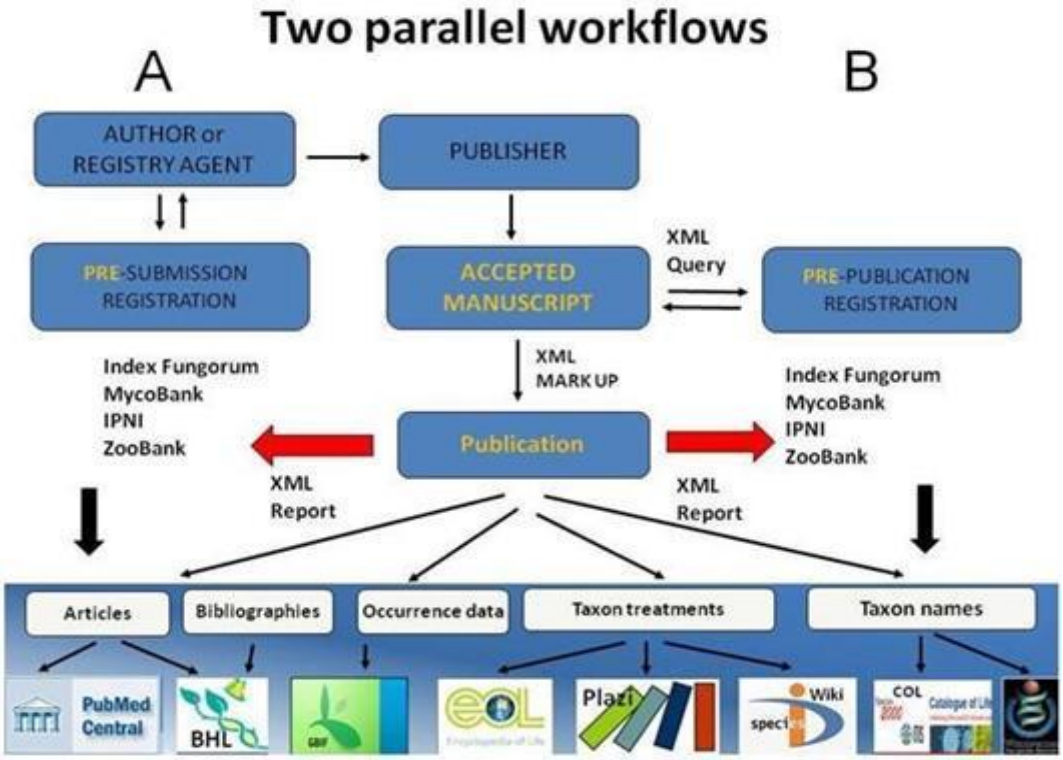

Figure 7. Two possible workflows for registration of taxonomic and nomenclatural acts, provided by an author or registry curator (A) or by a publisher (B). In both cases, the published information will automatically be amended in the registry through implementation of XML- or JSON-based server-to-server services.

Pilot on Interoperability model between taxon treatments from both legacy and prospective literature from three organismic domains (fungi, plants and animals)

The goal to demonstrate interoperability between taxon treatments extracted from both legacy and prospective literature, as well as between treatments coming from three organismic domains (fungi, plants and animals) was achieved by combining all the treatments in one single system (Figure 8). All treatments reside in Plazi, and any query can be submitted irrespective of the origin, biota, or markup technology applied, and be successfully answered.

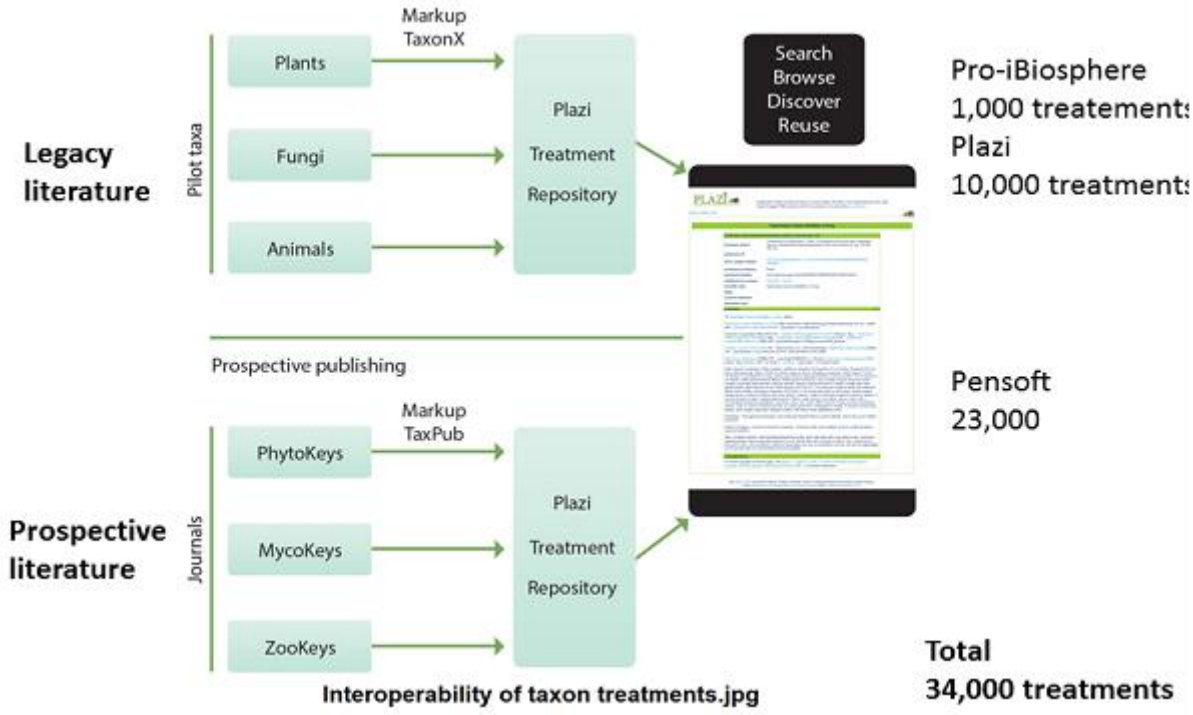

Figure 8. Overview of taxon treatment interoperability among biotas achieved during pro-iBiosphere.

Table 2 lists the documents that were processed and taxonomic treatments that were obtained within the pro-iBiosphere pilots. The documents complement treatments produced by Pensoft and submitted to Plazi as well as treatments resident in Plazi and are now accessible for download by other systems like CDM, Encyclopedia of Life (EOL) or Global Biodiversity Information Facility (GBIF).

Table 2. The documents processed and treatments obtained within the pro-iBiosphere pilots

Taxa	# Documents	# Treatments
Plants		
Loranthaceae	3	124
Chenopodium	15	174
Bryophyta	2	25
Nephrolepis	1	35
Fungi		

Basidiomycetes	5	5
Animals		
Centipedes	50	154
Ants	40	486
Spiders	30	219
Total	**146**	**1222**

Pilot on the "Evaluation of the CharaParser tool"

The legacy literature represents a huge pool for traits and, hence, there is a clear value for tools that can extract characters. The pilot aimed to evaluate the CharaParser tool. The tool generates identification keys by re-using morphological characters, locality and bibliographic citations.

The current workflow includes:

- Markup of treatments from legacy literature to morphological characters, locality and bibliographic citations;
- Review of other data types (e.g. phylogenies, ecological data) along with other ongoing initiatives;
- Key generation functions in the CDM Single-access keys;
- Generation of Multi-access keys;
- Production of Highly Structured descriptive data; and
- Production of a linked open access Taxpub based publication.

Ants and fungi were used for the evaluation.

The ant pilot included a set of publications in different languages, in different formats, and with mixture of proper descriptions and discussion of traits. The pilot demonstrated that CharaParser works best when used for a particular set of articles, like an entire Flora or a highly homogenous set of revisions in English. The fungi pilot focused on "Species identification and the detection of newly discovered taxa using character parsing and query-based searches based on digitised taxonomic literature". The goal was to provide tools that will facilitate taxonomic work by making comparisons of morphological and ecological data more efficient. In order to achieve this, the following method was followed: (i) Markup of scanned literature using GoldenGATE Editor, (ii) Parsing trait data, and (iii) Query-based search and filtering of quantitative and qualitative trait values using Shiny in R (see here for additional information). Ongoing work is still being conducted by the CharaParser team beyond the pro-iBiosphere project duration.

Based on the outcomes of these pilots, it was clear that future activities focusing on the integration in the markup workflow will require specific tools that are customised for traits, materials citations, bibliographic references, and linking to external resources.

Interoperability between schemas

The interoperability between FlorML and TaxPub XML schemas was tested and the model implemented in the re-publication of a Flora Malesiana volume on Pensoft's Advanced Books platform (Figure 9). The re-publication demonstrated:

- The multiple re-use of the effort spent for markup (usually a time consuming and expensive activity) and data extraction (Figure 10), and;
- Marked up information stored in Plazi and CDM can be used to produce new knowledge.

With the new platform, scientifically important historical monographs are enriched with additional information from up-to-date external sources about organismal names, species treatments, ecology, distribution and conservation value, morphological characters, etc. The re-publication in advanced open access shows the benefits of the digitization and markup efforts such as data extraction and collation, of the distribution and re-use of content, and of the archiving of different data elements in relevant repositories. This approach ensures that legacy can become freely usable for anyone at any place in the world.

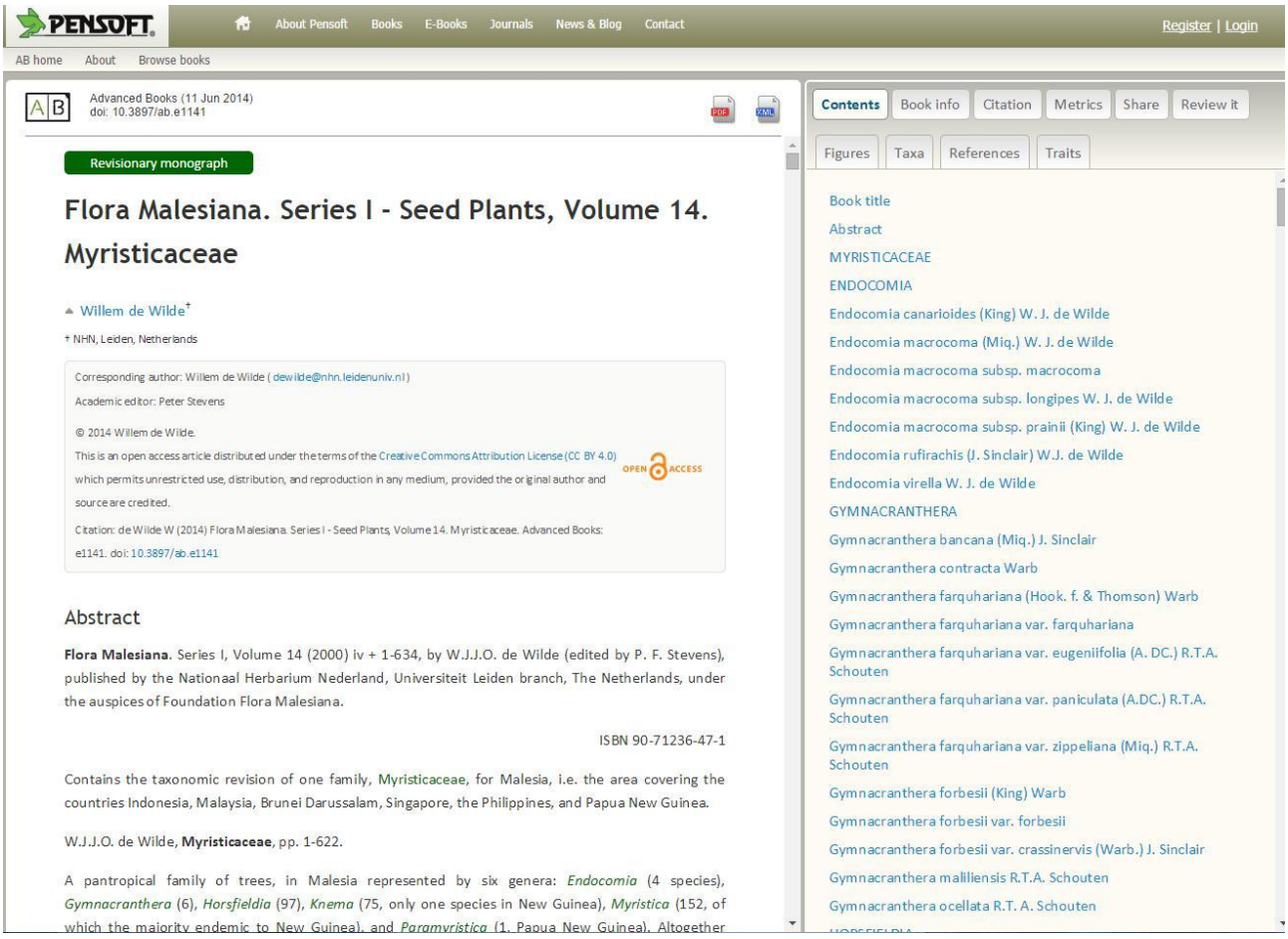

Figure 9. Flora Malesiana re-published in semantically enriched HTML format. (Publication: de Wilde W (2014), Flora Malesiana. Series I - Seed Plants, Volume 14. Myristicaceae. Advanced Books: e1141.doi: 10.3897/ab.e1141)

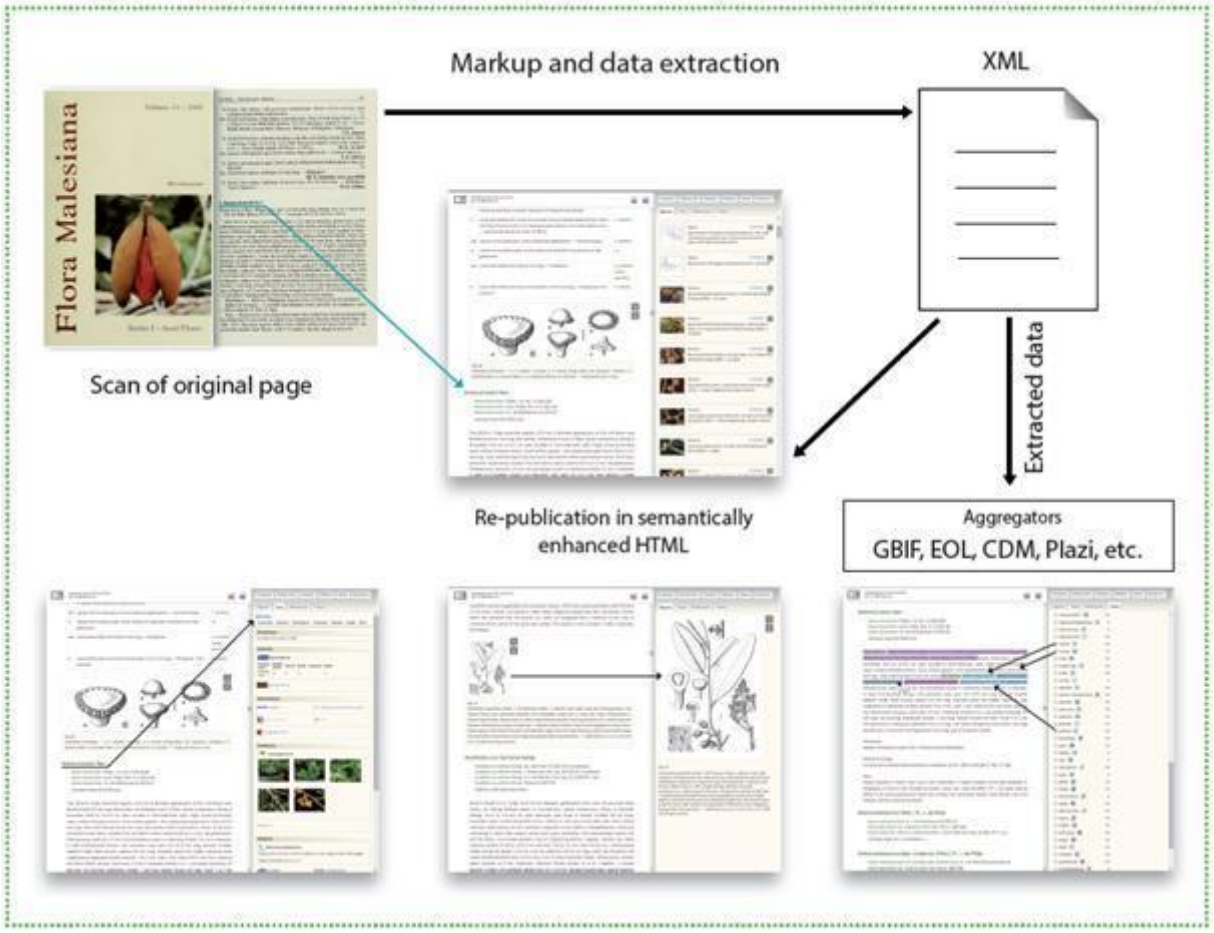

Figure 10. Multiple use of the markup effort: (1) digitisation and markup to produce a machine-readable XML file, (2) data extraction and submission to aggregators, and (3) advanced open access semantically enhanced HTML re-publication on Advanced Books.

4.1.3.4 Sustainability planning (WP6)

WP6 aimed to provide information on the costs and benefits of an OBKMS and to explore the factors necessary to support its development and sustainability.

To summarise, sustainability of OBKMS requires that:

- A base line of costs is established against which improvements can be made;
- User communities identified along with their needs are and the ways in which they benefit from the new services;
- Technical advances are embedded within biodiversity institutions (e.g. specimen identifiers);
- Good pilot projects are available to demonstrate best practice (e.g. the republication of Flora Malesiana in enhanced digital format);
- Institutes are motivated to continue to work together on new initiatives which build on the work of pro-iBiosphere; and
- Partners are actively soliciting new partners to join, e.g., by promoting the Bouchout Declaration on open access.

Measuring and Constraining the costs of delivering services

The project sought to identify the major costs associated with managing biodiversity information and making it widely available in digital form as part of an OBKMS.

The major costs elements identified are:

- High costs for the ongoing processes involved in mobilisation of legacy literature; including the costs of developing user friendly tools to make old literature digitally available;
- High costs owing to a lack of a scalable technological solution for production scale markup of the digitised text, particularly at the more detailed markup level;
- High startup costs required as investment in research and development of the new technology for enhanced publication services which are vital to demonstrate and deliver benefits to user and provider communities; and
- The large investment needed for outreach and support of the user and provider communities in order to demonstrate and deliver benefits of an OBKMS.

Recommendations were made as these costs may be reduced in the future were presented in the report "Cost delivery, efficiency and cost reduction through effective practices 2" (D6.1.2).

Identifying and measuring the benefits of delivering services

The project organised a workshop on "User engagement and benefits", attended by 26 users and providers of biological information (including educators, researchers, taxonomists and conservationists).

Main outcomes of the workshop consisted of:

- Identifying the user communities who could benefit from the implementation of the OBKMS. These include: conservationists, publishers, taxonomists, researchers, modellers, educators, regulators, citizen scientists, and commercial users (i.e. forestry, agriculture and aquaculture);
- Identifying the main benefits to potential users of an OBKMS. These include:
    - Benefits that can also be considered as enabling functions and that underpin all other benefits, these consist of 'increased availability & immediate access to data', 'availability of a central portal or repository to information' and 'consistent, accurate & current, up to date, information';
    - Direct benefits (those that are dependent upon data), consisting of 'comprehensive multilevel data, suitable for a variety of audiences', 'specialised data sets', 'local data' and 'non-biased, thorough data containing few gaps in biological knowledge'; and
    - Indirect benefits, consisting of 'increased accessibility of biological knowledge', 'creating income and raising money', 'food and natural materials security', 'evidence based decision making', 'increase biological knowledge and expertise', 'conservation of species, habitats and ecosystems' and 'improved public health'.

The benefits to users have been broadly summarised in the report "Benefits to users" (D6.2.1).

Technical, economic and sociological constraints arising from delivering these benefits were identified. It was established that the greatest constraints are likely to be sociological limitations that may take a long time to overcome. Challenges to cross disciplinary work, lack of political will or institutional or financial support, lack of taxonomic specialists and distrust of citizen science records were all identified as potential barriers.

The benefits to providers were also identified and have been summarised in the report "Benefits to providers" (D6.2.2). An understanding of these is critical if institutions are to change current practices and evolve towards more open knowledge management. The main benefits were:

- Increased access to data held by the provider;
- Collaboration and sharing of that data;
- Consistent, accurate and up-to-date information can be gathered from others;
- More open systems may be cheaper to run due to sharing of resources; and
- Provider gets recognition for data provision

Evaluating business models currently in use by partners

The project documented partners' current business models, essential baseline information which paved the way towards a more precise description of business models at the project's level which was further developed in Task 6.4 "Towards Sustainability for Services". The gathering of market background information informed the circumscription of the consortium's niche within OBKMS. Partners' exploitation plans were collated, updated and discussed, and allowed progress towards a consensual vision of what should and should not be done at the consortium level, taking advantage of the pro-iBiosphere developments.

These activities allowed a useful exchange of viewpoints, a good level of agreement on major concepts, a more detailed identification of what could be done together, representing a solid ground on which the project sustainability plan can be prepared.

The following key recommendations have been made:

- Consortium and non-consortium partners should identify a collaborative activity, time scale, and distribution of tasks in order to capitalise on the strong interest in collaboration with a view to identifying opportunities beyond the project time-frame.
- The project vision should be adjusted in light of any new priorities arising from the above.
- Consortium partners need to ensure that the Business Model, vision to implement the Open Biodiversity Management System, and Bouchout Declaration are each consistent with one another.

Towards Sustainability for Services

The "**Sustainability Draft Report**" (D6.4.2) was the first step in creating and fine-tuning a sustainability model for the OBKMS. Within this report, 22 products and services were highlighted, demonstrating the scope of OBKMS. Of these, four were chosen (as a result of the pro-iBiosphere workshop on 'alternative business models'), as high priority services and products:

- Linked open data publication;
- Providing markup tools;
- Conducting low granular markup of treatments; and
- Demonstration projects to demonstrate value of the OBKMS.

The list of core 22 iBiosphere products and services was grouped into 4 themes:

- Markup;
- Advanced publishing;
- Outreach; and
- Core technical functions.

An analysis of the four themes viewed from the partners perspective indicated that there are slightly more opportunities expected over threats to the individual partners business for two out of the four themes.

Higher threats to the markup theme are in the infrastructure (i.e. the technological solution and skills), and in the product/service (reflecting uncertainty over how atomised the data should be). With the outreach theme, there are higher than expected opportunities and also higher than expected threats connected with user relations. This emphasises that great opportunities are available when the user audience is appropriately addressed, but also that investment is needed to obtain a better understanding of the user requirements and to avoid the risks associated with misunderstanding the user relationship.

We also considered the situation where all partners would work together as a single enterprise. The preliminary analysis of the enterprise level benefits versus the constraints in realising the enterprise show

that in general, the expected number of benefits equals the expected number of constraints. Two exceptions to this are apparent. For the Advanced Publishing theme, the expected number of benefits is significantly larger than the expected number of constraints. This may reflect dissatisfaction with current traditional publishing and limited effective delivery to users. For the markup theme, the constraints to realising the infrastructure and the (right) product/service are significantly greater than the number of expected benefits. This reflects uncertainty as to the optimal level of markup, and perhaps concern over start-up costs and lack of current skills.

The report on "Risk Analysis" (D6.4.1) highlights the major risks to the long-term sustainability of the core products and services of an OBKMS, and outlines mitigation steps to minimise these risks. It distinguishes different categories of risks and identifies specific mitigation measures for each category.

Risks which would impact on the users include data not obtained in an usable format, not of the requested nature, or not of sufficient quality. Important mitigation steps include conducting outreach exercises with users; developing user-friendly tools including those promoting annotations to the original data; examining validation mechanisms; and building on existing networks between partners, and partners and users.

Risks which would be faced during the development and deployment of products and services fall into two categories: technical and economic. Technical risks include that efficient and user-friendly markup methods are not developed in time, and that new advanced publication mechanisms might not be taken up by a broad range of partners or other publishers. Economic risks include the cost of some innovations outweighing the benefits. There are also economic risks incurred by participating partners during the transition from a model where partners are working with separate systems, to a situation where greater integration and data sharing is the norm. Important mitigating actions include the use of standards; outreach to publishers to ensure the OBKMS can integrate with a variety of publication platforms; to develop easy to use tools; building on existing tools which allow more efficient work and linking of data; outsourcing markup to specialists, increasing efficiency and lowering costs; and to perform demonstration projects, which are important both for establishing efficient tools/processes and for illustrating the benefits to providers and to users.

The report "Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management" (D6.4.3) recapitulates the pro-iBiosphere vision, describes the Business Requirements for the core functions of an OBKMS; identifies the services that do not yet exist but would be essential for a fully functioning OBKMS; and provides an evaluation of the models for the business processes underpinning these requirements, drawing criteria from WP6 reports and outputs of the workshop on "Model Evaluation".

The main findings are:

- Existing business requirements will continue to be met by partners' current funding models, and;
- Funding for infrastructure and services will be a mixed model (public and private money) and there will be opportunities for specialists, for example in the provision of text conversion services.

In this mixed model the following sources of income may play a role:

- Grants;
- Joint project funding;
- Subscription;
- Charging for interpretative value added services aimed at commercial consumers;
- Additional funding to cover set up costs of new OBKMS product and service infrastructure;

- Support must be found for the development of production level tools, not just for the prototypes;
- Practical demonstrations should be developed to demonstrate the value of an OBKMS to:
    - The institutions themselves;
    - users; and
    - society at large

Governance is likely to evolve. To begin, the governing body could be quite small, focusing on the development of demonstration projects to showcase the value of OBKMS and on coordinating outreach activities to ensure the benefits of OBKMS are delivered to users and the broader scientific community. Additional funding would be required to establish an independent OBKMS entity with a "core" outreach role.

The key recommendations are:

- The governance system should be lightweight, initially directed at enabling individual institutions to commit themselves to sustaining core components of OBKMS, and at identification and relationship building with outsourcing partners;
- Demonstration projects should be initiated to illustrate the benefits of OBKMS;
- Attention should be given to the missing components of OBKMS: linked open data, taxonomic resolution, cross referencing and annotation; and
- Funding for production level tools should be encouraged at national and EU level.

The report **"**Recommendations to policy makers with regard to the sustainability of Open Knowledge Management" (D6.4.4) presents recommendations for sustainability of an Open Biodiversity Management System (OBKMS). The fundamental principles underlying OBKM are set out in the Bouchout Declaration. The recommendations presented here are principally aimed at policy makers at European, national and institutional levels. The recommendations draw on the results of the pro-iBiosphere work package on sustainability (WP6), but also on relevant recommendations from other work packages which addressed EU policy coordination (WP2); scientific content and workflows (WP3); technical infrastructure and coordination (WP4); dissemination, coordination and public awareness (WP5); and from discussions held at the pro-iBiosphere final event.

The recommendations on sustainability are grouped under the following categories: policy; financial; appropriate technology; socio-cultural aspects, including relationships within the biodiversity community; and institutional and management capacity.

Through the **workshop on "Alternative business models"** and pre-workshop questionnaire, a number of factors that underpin the systems economic viability were highlighted. It was discussed that for a system to be sustainable, not only does the income/revenue need to be equal or greater than the outgoings/costs, but also a number of social and environmental factors are needed, these include:

1. An appropriate supportive institutional framework;
2. An understanding and response to OBKMS's market environment;
    - Understand and meet current and future demand;
    - Be clear regarding OBKMS's niche and how its activities complement or compete against other initiatives;
    - Ensure OBKMS's target audiences are aware of its services;

- Ensure other audiences who may have an indirect interest in the project are aware of our services. Thus creating high visibility and may help fundraising;
- Ensure users and customers of OBKMS are supported appropriately in use of services and products;
- Recognition of funders and support; and
- Ensure that there is high visibility and popular press surrounding OBKMS;

3. An effective governance;
   - Trust among partners;
   - Appropriate and sustainable structure;
   - Clear roles and responsibilities;
   - Stakeholder role in governance; and
   - Planning for future changes.

In conclusion, WP6 has developed a clear baseline for measuring costs against which the saving of an evolving OBKMS can be monitored. It has determined communities of users and the benefits that they might expect from an OBKMS. Core products and services of an OBKMS have been determined and the potential opportunities and constraints identified. The key technical, financial and sociological factors needed to support the development of an OBKMS have been identified, the main risks considered, and recommendations have been made to help secure the development of an OBKMS.

4.1.4. Potential impact, main dissemination activities and exploitation of results

4.1.4.1 Potential impact

International and national efforts for conservation, sustainable use and public education demand accurate biodiversity information, as this information is crucial to monitoring programmes that seek to halt the rapid decline of our biodiversity. For information to be of use for such important purposes, it needs to be sufficiently detailed, complete, of high quality, and accessible, which is often not the case with biodiversity information. pro-iBiosphere made significant impact in four areas directly related to these issues, namely: (i) Open Access to high-quality biodiversity information, (ii) Technical and Semantic Interoperability, (iii) A more integrated user community, and (iv) A sustainable provision of high-quality base data.

i. **Open Access to high-quality biodiversity information.** pro-iBiosphere investigated the legal ramifications of copyright issues and recommended solutions that should satisfy both content providers and users. Reducing and/or eliminating the problems with sharing copyrighted biodiversity data will lead to increased scientific output and will put more biodiversity data in the spotlight of an ever increasing user base. The Bouchout Declaration for Open Biodiversity Knowledge Management, prepared and launched within the pro-iBiosphere project will help the community to make progress with data sharing and common tool development. The Declaration allows the biodiversity community to support the principle of open data; and to use this to identify and remove impediments to open data sharing. It also provides an important contribution to achieve the goals of the EU Digital Agenda, see here. **The Declaration was launched in June 2014 and during its few months of existence it has been broadly endorsed by the biodiversity community and beyond. By October 14 2014, the Declaration had been signed by 173 individuals and 85 organisations from 465 countries around the globe.** The Declaration, with its very broad spectrum of signatories, is a strong indication that the biodiversity community is willing to collaborate by opening up their data, make them discoverable and provide standardised ways to access them. Plazi and the Museum für Naturkunde have committed to continue to coordinate the Bouchout Declaration for Open Biodiversity Knowledge Management in the future.

ii. **Technical and Semantic Interoperability.** Based on current examples of integration provided by existing tools and platforms such as Scratchpads, Plazi's Treatment Repository and GoldenGATE Editor, Pensoft's Mark Up Tool (PMT), the Biodiversity Data Journal, BioWikiFarm, Species-ID, and others, the project facilitated the creation, comparison and exchange of protocols for biodiversity-specific information (occurrence data, localities, treatments, keys, names, and references), and, hence, increased cooperation and interoperability of methods and standards.

   a. pro-iBiosphere issued recommendations and guidelines for promoting future technical and semantic interoperability in the biodiversity data domain. These recommendations have already been taken into account by the most relevant international initiatives in the field:

      - **Consortium of European Taxonomic Facilities – Information Science and Technology Committee (CETAF - ISTC).** The recommendation addressed by the project to publish new knowledge in a semantically aware system using HTTP-URI based identifiers (for collection objects and beyond) has been addressed exemplary by the CETAF - ISTC. **The**

system will become a recommended strategy of CETAF (the decision will be taken by the time this report is submitted).

- **FP7 funded project EU BON** plans to include data from legacy publications as input to its biodiversity modelling. Tools and standards **recommended in the report "Strategy for improvement & interoperability of the XML schemas" are being used in the EU funded EU BON project and in the everyday activities of PubMed Central, GBIF, EOL and project partners such as Plazi and Pensoft.** In the framework of EU BON, the re-publication workflow and platform tested and modelled by pro-iBiosphere to **re-publish a Flora Malesiana volume on Pensoft's Advanced Books platform** (using the FlorML and TaxPub XML schemas) was used to: **(i) re-publish the Flora of Northumberland and Durham originally published in 1831; and (ii) geotag, extract and submit to GBIF historical localities of plants as a distinct dataset.**
- **EU FP7 funded project BioVel.** The BiodiversityCatalogue.org developed by the EU FP7 project BioVel allows to register web services provided by different players in the biodiversity domain. The Catalogue is already been used by a wide range of projects and initiatives across Europe (see here). BioVel and pro-iBiosphere agreed to register the web services developed in pro-iBiosphere and a future OBKMS in the Biodiversity Catalogue.
- **EU FP7 funded project European Distributed Institute of Taxonomyn (EDIT).** Promoting and monitoring the development and adoption of common markup standards has led to further improvements in interoperability and user experience. The technical pipelines for transforming literature markup into the CDM-based EDIT Platform for Cybertaxonomy implemented by pro-iBiosphere have opened up literature data in a standardised and semantically well-defined way. **This has facilitated new query mechanisms on literature published via EDIT Platform portals as well as provision of data via standardised and robust web services which can be integrated in workflows and e-platforms**.

b. The pro-iBiosphere Biodiversity Data Enrichment Hackathon helped achieve semantic interoperability of biodiversity data. The tools that were improved and developed during the Hackathon are facilitating re-use and enhancement of biodiversity knowledge by a broad range of stakeholders, such as taxonomists, systematists, ecologists, niche modellers, informaticians, and ontologists. **Technical solutions and their potential impact have been documented in the scholarly publication here available: Doi: 10.3897/BDJ.2.e1125**.

c. The Darwin Core Archive interchange format developed for Scratchpads (in the framework of the EUFP7 funded project ViBRANT) was tested and used by the pro-iBiosphere partners Plazi, FUB-BGBM and Pensoft as a data sharing tool for data mined from legacy and prospective literature. To facilitate the process of harvesting materials observation data from Plazi and exporting primary biodiversity data to GBIF for further indexing and re-use (as a result of the markup and data mining), a change from the Tapir service to the use of Darwin-Core Archive was discussed with GBIF during one of the pro-iBiosphere meetings**. The Darwin-Core Archive has been successfully implemented and the Tapir service has been discontinued. At present, the GBIF service uses the same technology for the transfer of treatments from Pensoft and Plazi to EOL.**

iii. **Sustainable provision of high-quality base data.** To cope with the urgent demand for information (from public, education, governments and NGOs), the project increased the possibilities for data retrieval and exchange between different data providers and other stakeholders. Taxonomic information has been published in the past in numerous scattered taxon-specific outlets and in different formats. As a result of this, the production of a taxonomic revision or a Flora or Fauna series required that the appropriate text was discovered and retyped manually.

   a. An accelerated workflow for interoperability of taxon treatments that takes advantage of the informatics developments of pro-iBiosphere was developed by Plazi. **The workflow allows users of the Plazi treatment repository to locate, identify, and enhance data included in treatments from both legacy and newly published taxonomic literature (e.g., marked up treatments supplied by Pensoft's journals), facilitating discovery, analysis, and re-use.**

   b. Markup activities were extended from training to real data extraction which led to more than 1000 marked up treatments from various taxa submitted to Plazi. These treatments are accessed to and linked back from newly published articles. **Most of the treatments are exported to relevant aggregators, such as EOL, AntWeb, GBIF and EDIT CDM.**

   c. To achieve its goal of delivering biodiversity data in open and re-usable forms, the pro-iBiosphere consortium addressed the sustainability of both the data and the means to access them. Sustainability requires that providers can curate, supply and exploit data effectively to ensure that the benefits of participation outweigh the costs. This can be facilitated through technical innovations, but also requires a full understanding of user requirements and how these can best be met. Benefits to the users include increased availability through a central portal as well as increases in the volume, consistency, reliability, and currency of information available through searches. **The benefits to data providers include greater use of their data, accreditation, improved awareness of impact, avoidance of duplicated effort, increased opportunities for innovative collaboration and tools for curation, analysis and research. To ensure that the benefits are realised, the implementation of OBKMS requires that stronger relationships are built and sustained between the provider and user communities.**

   d. Business requirements and scenarios for sustainable OBKM include, the (i) Identification of a mixed funding model which includes value added services; (ii) an initial model with a small core function; (iii) the importance of outreach and of good demonstrations to showcase the benefits; and (iv) the importance of the development of production level services rather than pilots. The deliverable "Recommendations to policy makers" summarises our significant results on sustainability under the following headings: policy; financial; appropriate technology; socio-cultural aspects, including relationships within the biodiversity community; and institutional and management capacity.

iv. **A more integrated user community**. The various workshops and events organised by the pro-iBiosphere have helped to further foster a community of experts in biodiversity informatics and to build human links between research projects and institutions, in response to recent calls to further such integration in this research domain. The various workshops and events organised by the project have **helped to inform users of the biodiversity data of the existing possibilities, and inform developers of technical and other barriers encountered**.

pro-iBiosphere coordinated and mobilised distributed international expertise and helped to increase community cooperation in- and outside Europe, leading to innovative use of biodiversity data.

a. Registries. The project worked closely with the registries of new taxon names and nomenclatural acts - International Plant Name Index (IPNI), Index Fungorum, MycoBank and ZooBank - to create a common automated registration model. Building on the work done in the project, a pilot system for registration of nomenclatural acts in IPNI has been further developed at RBGK. T**his system is now being tested by members of the Special Committee for the Registration of Names of the International Association of Plant Taxonomists prior to release for broader community use. In addition to this, the workflow established between Pensoft and ZooBank has been put in place as a separate interface page at ZooBank for free use for any publisher or journal**. This was facilitated by a strong collaboration with the ZooBank manager and developer, based in the Bishop Museum, Honolulu.

b. Publishers. In the framework of the pro-iBiosphere project, the TaxPub **Journal Article tagging Suite** (JATS) was adopted to serve as a standard system for the semantic markup of newly published journal articles in the area of biological systematics. After inconsistencies in publishers' use of the JATS delivered to PubMed Central were **discovered**, **a standardization group (JATS 4 Reuse) has been formed in June 2014 to address these issues.**

4.1.4.2 Main dissemination activities

During the project lifetime (September 2012 - August 2014), pro-iBiosphere conducted various dissemination activities aimed at promoting its research to the widest and varied audience possible. The Dissemination and Communication Implementation Plan (DCIP) was released at M2 and placed on the project wiki. The DCIP provided guidelines for dissemination activities by project partners by presenting information on the dissemination strategy of the project, the aim of the dissemination actions, the communication and dissemination tools to be used, and the activities and mechanisms for information exchange with various stakeholders.

Key dissemination tools and activities conducted by the project are described below.

- **Project website.** The website (www.pro-ibiosphere.eu) was made available in November 2012 (month 3 of the project). It has been used as an important dissemination channel, describing project activities and outcomes such as latest news, articles, presentations, internal and external documents. The domain name was registered for a period of 10 years and the website will be maintained at least 5 years after the project ends.

  The statistics as of October 2014 are:

  - News items posted: 108
  - External documents shared with stakeholders: 31
  - Sessions: 4.574
  - Average number of unique visits per month: 381,2
  - % of returning visitors: 53,7
  - Number of pages per visit: 2,49
  - New visitors: 46,3 %
  - Average visit duration: 0:02:26
  - Internal Communication Platform users: 47I
  - % of visits in Europe: 77.0

- % of visits in America: 9.1
- Number of news per month: 5,5
- Size of the dissemination database: 614

- **Project wiki.** The wiki (http://wiki.pro-ibiosphere.eu/wiki/Main_Page), available since August 2012 (month 1 of the project), has been used by the consortium and project stakeholders to facilitate collaborative work and promote on-going activities and outcomes of the project. It provides access to (i) all project events information (agenda, directions, list of participants, proceedings, presentations given); (ii) deliverables and milestones; (iii) partners' contribution to events and publications; (iv) tasks and pilots developments and achievements information. Per October 2014, the wiki had a total of 627 registered users, 862 content pages, and 394 uploaded files.

- **Bouchout Declaration website**. The Bouchout Declaration was launched in June 2014. The launch was followed online via Twitter (50 tweets on @bouchoutdec and estimated reach of 60,180 accounts). During its few months of existence it has been broadly endorsed by the biodiversity community and beyond. By October 14 2014, the Declaration had been signed by 173 individuals, 85 organisations from 46 countries around the globe.

- **Helpdesk services.** Helpdesk services were provided to the project community, they have been implemented through the use of:

  - The project LinkedIn group (http://www.linkedin.com/groups/pro-iBiosphere-4682845?trk=myg_ugrp_ovr)
  - The contact email (info@pro-ibiosphere.eu)
  - The online feedback form on the project website (http://www.pro-ibiosphere.eu/feedback/)
  - A contact email specific for the Final Event (final-event@pro-ibiosphere.eu).

  Around 30 requests have been received via the website, social media and the contact emails and dealt with by the project during the project period.

- **Logo, PPT and deliverable templates.** A logo (http://www.pro-ibiosphere.eu/media/center/3691) was created to establish an identity for the project. Templates for PowerPoint presentations and deliverables were designed to ensure congruent presentation to external audiences. Templates for deliverables include the terms of the Creative Commons Attribution License 4.0 (CC-BY-4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

- **Dissemination database.** Project contacts were consolidated into a mailing list (dissemination@pro-ibiosphere.eu) containing:

  - Participants of project events
  - Contacts from other initiatives
  - Experts identified during the preparation of the project
  - Partners' contacts

- **eNewsletter.** A total of six e-Newsletters (http://www.pro-ibiosphere.eu/news/) were disseminated during the project period, highlighting project activities, outputs, upcoming events and other biodiversity information of interest. e-Newsletters were sent on a quarterly basis to the project dissemination database.

- **Press releases.** A total of 7 press releases (http://www.pro-ibiosphere.eu/media/center/3693) were published during the project period to world news media, various bioinformatics and biodiversity related mailing lists and the project dissemination database.
- **Project documentation.** Dissemination material (http://www.pro-ibiosphere.eu/media/center/) produced by the project include:
    - Posters (13 in total)
    - Leaflet summarising the concept of the project
    - Social media postcard
    - Brochure presenting the project key outcomes
    - Flyer for Final Event
- **Publications.** The project produced a total of 12 peer reviewed publications, for additional information please see Template A1 of this report.
- **Social Media.** The social media accounts of the project (i.e. Twitter, LinkedIn, Google + and Facebook) were created in M2. They are accessible from the project website.
- **Calendar.** In M9, a Google Calendar was created to monitor the contribution of partners in dissemination activities and ensure their timely delivery. The different project deliverables, milestones and prospective events were added to this calendar to give a clearer vision of project activities and expected outputs. A calendar file was then shared with all partners to be integrated into their own calendar. They would then receive reminder emails to (i) manage their deliverables/milestones and to (ii) publish news on project tasks, pilot's achievements, etc.
- **Project trainings.** A total of three training events (including demonstrations) were organised, on: (i) The use of specific software (GoldenGATE) for markup of digitised text (14-15 January 2013 - Leiden - the Netherlands); (ii) e-platforms and e-tools for taxonomy (12 February 2013, Leiden - the Netherlands); WikiMedia (11 June 2014, Meise - Belgium).
- **Project workshops.** A total of 14 workshops were organised, including a wide spectrum of interested groups, including scientists from different disciplines (such as ecologists, niche modellers, historians, taxonomists), librarians, publishers, web designers, and bioinformaticians.
- **Participation in events.** In total, project partners participated in 32 events during the project period (Sep. 2012 to Aug. 2014). 25% of these events took place outside Europe (13% in Asia, 3% in Africa, 6% in the United States and 3% in South America) and 5 events were organised by the European Commission.

4.1.4.3 Exploitation of results

The pro-iBiosphere consortium will continue to advertise its outcomes beyond the project duration, promoting the Standardised HTTP URI-based identifiers for collection objects, automated registration workflows for taxon names, the possibilities for integrating data from literature in a handbook (part of the 2015 work programme of the FUB-BGBM Biodiversity Research Group), the Bouchout Declaration (including a review within approximately two years after the launch), among others.

pro-iBiosphere has issued recommendations and guidelines that will promote future technical and semantic interoperability in the biodiversity data domain. These recommendations have been adopted by ongoing projects (e.g. EU-BON) and the most relevant international initiatives in the field (e.g. GBIF, EOL).

The results and conclusions from the legal studies of the project are being used by Plazi to propose legal exemptions for scientific use of works both at EU as well as national level (Switzerland).

The activities organised by the project and the outcomes obtained have led to new ideas for research proposals and plans to better integrate existing projects. As a result of this, contacts with potential funding organisations is being carried out to provide longer-term funding opportunities. For instance, a EU proposal was submitted in August 2014 by various CETAF members addressing the implementation of a Linked Open Data infrastructure for Biodiversity Information including collection data, observations, literature, sequences, descriptive data, and taxa.

4.1.5. Address of the project public website and relevant contact details

pro-iBiosphere website: www.pro-ibiosphere.eu; wiki: http://wiki.pro-ibiosphere.eu/wiki/Main_Page

No	Beneficiary name	Short name	Principal investigator	Country
CO1	STICHTING NATURALIS BIODIVERSITY CENTER	Naturalis	Soraya Sierra	The Netherlands
P2	NATIONALE PLANTENTUIN VAN BELGIE	NBGB	Quentin Groom	Belgium
P3	FREIE UNIVERSITAET BERLIN	FUB-BGBM	Anton Güntsch	Germany
P4	PENSOFT PUBLISHERS LTD.	PENSOFT	Lyubomir Penev	Bulgaria
P5	SIGMA ORIONIS SA	SIGMA	Camille Torrenti	France
P6	ROYAL BOTANIC GARDENS KEW	RBGK	Don Kirkup	United Kingdom
P7	PLAZI VEREIN	Plazi	Donat Agosti	Switzerland
P8	MUSEUM FUR NATURKUNDE - LEIBNIZ-INSTITUT FUR EVOLUTIONS- UND BIODIVERSITATSFORSCHUNG AN DER HUMBOLDT-UNIVERSITAT ZU BERLIN	MfN	Daniel Mietchen	Germany

4.2 Use and dissemination of foreground

4.2.1. Section A

4.2.1.1 Publications

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES

NO.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers[1] (if available)	Is/Will open access[2] provided to this publication?
1	Scientific names of organisms: attribution, rights, and licensing	David Patterson	*BMC Research Notes*	Volume 7 - February 2014	BioMed Central	London, UK	2014	79	doi:10.1186/1756-0500-7-79	Yes
2	Open exchange of scientific knowledge and European copyright: The case of biodiversity	Willi Egloff	ZooKeys	Volume 414 - 6 June 2014	Pensoft Publishers	Sofia, Bulgaria	2014	109-135	10.3897/zookeys.414.7717	Yes

[1] A permanent identifier should be a persistent link to the published version full text if open access or abstract if article is pay per view) or to the final manuscript accepted for publication (link to article in repository).
[2] Open Access is defined as free of charge access for anyone via Internet. Please answer "yes" if the open access to the publication is already established and also if the embargo period for open access is not yet over but you intend to establish open access afterwards.

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
NO.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers[1] (if available)	Is/Will open access[2] provided to this publication?
	information									
3	A common registration–to–publication automated pipeline for nomenclatural acts for higher plants (International Plant Names Index, IPNI), fungi (Index Fungorum, MycoBank) and animals. 2014	Lyubomir Penev	ZooKeys	In press	Pensoft Publishers	Sofia, Bulgaria	2014	t.d.b.	Not known yet	Yes
4	Enriched Biodiversity Data as a resource and service	Rutger Vos	Biodiversity Data Journal	Volume 2 - June 2014	Pensoft Publishers	Sofia, Bulgaria	2014	e1125	10.3897/BDJ.2.e1125	Yes
5	Using legacy botanical literature as a source of phytogeographical data	Quentin Groom	Plant Ecology and Evolution	2014	Botanical Garden Meise	Meise, Belgium	In peer review	t.b.d.	Not known yet	Yes
6	A new species of Orthosiphon (Lamiaceae) from Angola	Alan Paton	Biodiversity Data Journal	Volume 2 - June 2014	Pensoft Publishers	Sofia, Bulgaria	2014	e1162	10.3897/BDJ.2.e1162	Yes

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
NO.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers[1] (if available)	Is/Will open access[2] provided to this publication?
7	*Eupolybothrus cavernicolus* Komerički & Stoev sp. n. (Chilopoda: Lithobiomorpha: Lithobiidae): the first eukaryotic species description combining transcriptomic, DNA barcoding and micro-CT imaging data	Pavel Stoev	Biodiversity Data Journal	Volume 1 - October 2013	Pensoft Publishers	Sofia, Bulgaria	2013	e1013	Doi: 10.3897/BDJ.1.e1013	Yes
8	Tracking two centuries of phytogeographic change using legacy botanical literature and collections	Quentin Groom	Neobiota	2014	Pensoft	Sofia, Bulgaria	In peer review	t.b.d.	Not known yet	Yes
9	Detailed markup of semi-monographic legacy taxonomic works using FlorML	Thomas Hamann	Taxon	Volume 63 (2) - April 2014	International Association of Plant Taxonomy	Bratislava, Slovakia	2014	377-393	http://dx.doi.org/10.12705/632.11	No

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
NO.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers[1] (if available)	Is/Will open access[2] provided to this publication?
10	Implementation of TaxPub, an NLM DTD extension for domain-specific markup in taxonomy, from the experience of a biodiversity publisher	Lyubomir Penev	Journal Article Tag Suite Conference (JATS-Con) Proceedings	Issue 2012 - October 2012	National Center for Biotechnology Information	Bethesda, USA.	2012		Bookshelf ID: NBK100351	Yes
11	Biodiversity research in the "big data" era: GigaScience and Pensoft work together to publish the most data-rich species description	Scot Edmunds	GigaScience	Volume 2 - October 2013	BioMed Central	London, UK	2013	14	doi:10.1186/2047-217X-2-14	Yes
12	An appraisal of megascience platforms for biodiversity information	Dagmar Triebel	MycoKeys	Volume 5 - December 2012	Pensoft Publishers	Sofia, Bulgaria	2012	45-63	doi: 10.3897/mycokeys.5.4302	Yes

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
NO.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers[1] (if available)	Is/Will open access[2] provided to this publication?
13	Some thoughts after Cusco	Donat Agosti	Hamuli	2014	International Society of Hymenopterists		2014	5(2): 4-5	10.5281/zenodo.11671	Yes

In addition to the publications listed, 7 publications are in progress to be finished and submitted after the duration of the project.

For more information please see here: http://wiki.pro-ibiosphere.eu/wiki/List_of_planned_publications

4.2.1.2 Dissemination activities

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
1	Presentation during the "Flora of the Guianas Biennial meeting and seminars"	Naturalis	Coordination and Policy Development in Preparation for a European Open Biodiversity Knowledge Management System	October 22-24 2012	Leiden, The Netherlands	Scientific community	c. 50	Global
2	Presentation during the "TDWG Annual Meeting"	Pensoft	Making small data big! The Biodiversity data Journal (BDJ)	October 22-26 2012	Beijing, China	Scientific community, policy makers	>150	Global
3	News item published on the EU FP7 BESAFE project website	Pensoft	Toward an European Open Biodiversity Knowledge Management System Making high quality biodiversity knowledge open, accessible and re-usable	October 29 2012	n.a.	Scientific community, policy makers	not specific, large community	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
4	Email sent to the mailing list Bioinfo	BGBM	Toward an European Open Biodiversity Knowledge Management System Making high quality biodiversity knowledge open, accessible and re-usable	November 2 2012	n.a.	Scientific community	not specific, large community	French +- Global
5	Email sent to the mailing lists: (i) TDWG Structure of Descriptive Data, (ii) TDWG Content, (iii) Open Science, (iv) Taxacom	Plazi	Toward an European Open Biodiversity Knowledge Management System Making high quality biodiversity knowledge open, accessible and re-usable	November 6 2012	n.a.	Scientific community, policy makers	not specific, large community	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
6	Email sent to the Taxacom mailing list	NBGB	Toward an European Open Biodiversity Knowledge Management System Making high quality biodiversity knowledge open, accessible and re-usable	November 7 2012	n.a.	Scientific community	not specific, large community	Global
7	Facebook posts	Plazi	Workshops and trainings to be organised by pro-iBiosphere in February 2013	November 15 2012	n.a.	Scientific community, policy makers, Civil society, Media	not specific, large community	Global
8	Presentation during a colloquium organised by Naturalis	Naturalis	Coordination and Policy Development in Preparation for a European Open Biodiversity Knowledge Management System	November 20 2012	Leiden, The Netherlands	Scientific community, Media	c. 300 persons	The Netherlands

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
9	Email sent to Flora Malesiana mailing list	Naturalis	pro-iBiosphere workshops to be held in February 2013 in Leiden, The Netherlands	November 26 2012	n.a.	Scientific community	not specific, large community	Global
10	Press release - EurekAlert	Naturalis	Communication on the release of the 1st newsletter of pro-iBiosphere	November 28 2012	n.a.	Scientific community, Media	not specific, large community	Global
11	Poster during the annual meeting on plant ecology and evolution	NBGB	Coordination and Policy Development in Preparation for a European Open Biodiversity Knowledge Management System	November 30 2012	Meise, Belgium	Scientific community	70	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
12	Email sent to the pro-iBiosphere dissemination database	Pensoft	pro-iBiosphere Newsletter (1 Sept-Dec 2012)	December 2012	n.a.	Scientific community, policy makers, civil society, industry	630	Global
13	Presentations during a series of visits in the U.S.: (i) Dec 3 2013, Meeting with C. Miller and W. Ulate (BHL project) at Missouri Botanical Garden (ii) Dec 5 2013, Lecture at the biodiversity group at Univ of Kansas, Lawrence (iii) Dec 6 2013, Participation in CharaParer workshop, Tucson, Arizona (iv) Dec 8 2013, Meeting at CAS	Plazi	Coordination and Policy Development in Preparation for a European Open Biodiversity Knowledge Management System: Goals, objectives, expected outcomes and on-going activities of the project	December 3-8 2012	St Louis, Missouri; Kansas, Texas; Tucson, Arizona; USA	Scientific community	>50	US

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
14	Email sent to the Naturalis mailing list	Naturalis	pro-iBiosphere Newsletter 1 (Sept-Dec 2012)	January 8 2013	Leiden, The Netherlands	Scientific community, Media	c. 300 persons	The Netherlands
15	Training	Plazi	The use of GoldenGATE for markup of digitised text	January 14-15 2013	Leiden, The Netherlands	Scientific community	20	Global
16	Training	NBGB	e-platforms and e-tools for taxonomy	February 12 2013	Leiden, The Netherlands	Scientific community	101	Global
17	Workshop	NBGB	pro-iBiosphere workshop on e-platforms and e-tools for taxonomy	February 12 2013	Leiden, The Netherlands	Scientific community	101	Global
18	Workshop	Plazi	pro-iBiosphere workshop on Legacy literature	February 13 2013	Leiden, The Netherlands	Scientific community	101	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES

NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
19	Workshop	Naturalis	pro-iBiosphere workshop on Prospective Literature	February 14 2013	Leiden, The Netherlands	Scientific community	101	Global
20	Workshop	RBGK	pro-iBiosphere workshop on Requirements of users of Flora, Fauna or Mycota publications or services	February 14 2013	Berlin, Germany	Scientific community	46	Global
21	Presentation during "BioSyst 2013"	Pensoft	Automated registration model for eukaryotic organisms: The opportunities for and responsibilities of publishers	February 18-22 2013	Vienna, Austria	Scientific community	>500	Global
22	Individual emails sent to >100 persons	Plazi	pro-iBiosphere May workshops	March 6 2013	n.a.	Scientific community, policy makers	> 100	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
23	Participation in the 10th Concertation Meeting on e-Infrastructures	Naturalis	pro-iBiosphere project	March 6-7 2013	Brussels, Belgium	Scientific community, policy makers, EU, Other	c.100	European
24	Email sent to the Flora Malesiana mailing list	Naturalis	pro-iBiosphere May workshops	March 12 2013	n.a.	Scientific community	not specific, large community	Global
25	News item for the GBIF (German node) newsletter	Naturalis	pro-iBiosphere May workshops	March 12 2013	n.a.	Scientific community, policy makers, civil society, industry	not specific	Germany

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
26	Participation in meeting "Towards a Roadmap for Research Infrastructures on Biodiversity and Ecosystem research in Europe"	Naturalis	pro-iBiosphere project	March 19-20 2013	Brussels, Belgium	Scientific community, policy makers, EU, Other	>50	Global
27	Email sent to the pro-iBiosphere dissemination database	Pensoft	pro-iBiosphere Newsletter 2 (Jan-April 2013)	April 2013	n.a.	Scientific community, policy makers, civil society, industry	630	Global
28	Presentation during the Symposium of the Biodiversity Research Group	Plazi	pro-iBiosphere project	May 12 2013	Lawrence, Kansas; USA	Research community	<50	US
29	Online meeting of the World Flora Technical group	RBGK	pro-iBiosphere project	May 20 2013	Online	Scientific community	c.20	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
30	Workshop organised by pro-iBiosphere	RBGK	Measuring and constraining the costs of delivering service	May 22 2013	Berlin, Germany	Scientific community	61	Global
31	Workshop organised by pro-iBiosphere	Plazi	Coordination & routes for cooperation across organisations, projects & e-infrastructures	May 23 2013	Berlin, Germany	Scientific community	66	Global
32	Participation in Hackathon	FUB-BGBM	Stable Identifiers Hackathon	June 4-5 2013	Edinburgh, UK	Computer scientists from Natural History Institutions	15	Global
33	Presentation during seminar series organised at the National Botanic Garden of Belgium	NBGB	Towards a European Open Biodiversity Knowledge Management System	June 10 2013	Brussels, Belgium	Scientific community	c.30	Belgium

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
34	Email sent to (i) German Systematic Association Email List/ Young Systematists and (ii) MfN's PR Department	MfN	Towards a European Open Biodiversity Knowledge Management System	June 17 2013	n.a.	Scientific community	>300	Germany
35	Presentation during the 19th International Congress of Arachnology (ICA 2013)	Naturalis	The structure of Taxonomic data	June 23-28 2013	Taiwan	Scientific community	>100	Global
36	Email sent to the pro-iBiosphere dissemination database	Pensoft	pro-iBiosphere Newsletter 3 (May-Aug 2013)	August 2013	n.a.	Scientific community, policy makers, civil society, industry	630	Global
37	Presentation during the 9th "International Flora Malesiana Symposium"	RBGK, FUB-BGBM	pro-iBiosphere project pilots	August 27-31 2013	Bogor, Indonesia	Scientific community	c.300	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
38	Press release - EurekAlert	Pensoft, Naturalis	Recommendations for removing copyright hurdles to scientific research	September 3 2013	n.a.	Scientific community, Media	not specific	Global
39	Press release - EurekAlert	Pensoft, NBGB, Naturalis	Recommendations on how to move the naming of organisms from paper and on to the Internet	September 3 2013	n.a.	Scientific community, Media	not specific	Global
40	Press release - EurekAlert	Pensoft	The future of biodiversity publishing	September 3 2013	n.a.	Scientific community, Media	not specific	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
41	Presentation and posters during the Biodiversity Informatics Horizon Conference 2013	Pensoft, FUB-BGBM, Plazi, NBGB, RBGK, Naturalis, MfN	pro-iBiosphere project pilots	September 3-6 2013	Rome, Italy	Scientific community, policy makers, EU, Other	180	Global
42	Presentation during the "Information Science & Technology Commission"	FUB-BGBM	Stable identifiers	September 10-11 2013	Edinburgh, UK	CETAF Directors	30	Global
43	Presentation during the workshop on "Strengthening the cooperation between the US and the EU in the field of Environmental Research Infrastructures and e-infrastructures meeting"	Naturalis	pro-iBiosphere towards Open Biodiversity Knowledge	September 18-20 2013	Madrid, Spain	Scientific community, policy makers, EU, Other	<100	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
44	Presentation during the "Identification and eTaxonomy: Morphometrics" workshop	Naturalis	Opening and facilitating re-use of Biodiversity knowledge	September 26 2013	Leiden, The Netherlands	Scientific community	c.30	Global
45	Presentation during the "Identification and eTaxonomy: Morphometrics" workshop	Naturalis	The Fungi Pilot: artic Russulaceae	September 26 2013	Leiden, The Netherlands	Scientific community	c.30	Global
46	News item for the FUB-BGBM newsletter	FUB-BGBM	pro-iBiosphere workshops taking place in October 2013	October 2013	n.a.	Scientific community	not specific	Global
47	Workshop organised by pro-iBiosphere	FUB-BGBM	How to improve technical cooperation and interoperability at the e-infrastructure level	October 8 2013	Berlin, Germany	Scientific community	21	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
48	Workshop organised by pro-iBiosphere	Plazi	How to promote and foster the development & adoption of common markup standards & interoperability between schemas	October 8 2013	Berlin, Germany	Scientific community	20	Global
49	Workshop organised by pro-iBiosphere	RBGK	User engagement and benefits	October 9 2013	Berlin, Germany	Scientific community	28	Global
50	Workshop organised by pro-iBiosphere	RBGK	Evaluation of business models	October 10 2013	Berlin, Germany	Scientific community	21	Global
51	Workshop organised by pro-iBiosphere	RBGK	Alternative business models	October 11-12 2013	Berlin, Germany	Scientific community	20	Global
52	Presentation during TDWG Annual Meeting 2013	MfN	What Taxonomic Information is Open Access? (Copyright Issues...)	October 28 – November 1 2013	Florence, Italy	Scientific community, policy makers	>150	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
53	Presentation during TDWG Annual Meeting 2013	Pensoft	Next generation sequencing requires next generation publishing	October 28 – November 1 2013	Florence, Italy	Scientific community, policy makers	>150	Global
54	Press release - Digital Agenda of Europe website	Soraya Sierra	Toward an European Open Biodiversity Knowledge Management System Making high quality biodiversity knowledge open, accessible and re-usable	October 29 2013	n.a.	Scientific community, policy makers, civil society, EU	not specific, large community	Global
55	News item for the BioVel newsletter	FUB-BGBM	Collaboration between BioVeL and pro-iBiosphere	November 2013	n.a.	Scientific community	not specific, large community	Global
56	Press release - Phys.org	Pensoft	Next-generation global e-infrastructure for taxon names registry	November 2013	n.a.	Scientific community, Media, policy makers	Large, non-specified	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
57	Presentation during the ICT 2013 event, Networking session on Big Data	Plazi	Coordination and Policy Development in Preparation for a European Open Biodiversity Knowledge Management System	November 6 2013	Vilnius, Lithuania	Scientific community, policy makers	<100	Global
58	Flyer dissemination during the ICT 2013 event	Plazi, Naturalis, Sigma	Coordination and Policy Development in Preparation for a European Open Biodiversity Knowledge Management System	November 6-8 2013	Vilnius, Lithuania	Scientific community, policy makers	c.5000	Global
59	News item for the GBits newsletter - November edition	Naturalis and FUB-BGBM	Contribution sent to Maren Gleisberg: GBIF-D Team Coordination	November 7 2013	n.a.	Scientific community, policy makers, civil society, industry	not specific, large community	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
60	Email sent to the pro-iBiosphere dissemination database	Pensoft	pro-iBiosphere Newsletter 4 (Sept.-Dec. 2013)	December 2013	n.a.	Scientific community, policy makers, civil society, industry	630	Global
61	Publication on Website	FUB-BGBM	Top tips on creating Web Services (http://www.software.ac.uk/blog/2013-12-04-top-tips-creating-web-services)	December 2013 -	n.a.	Web service developers	unknown	Global
62	Flyer dissemination during the 6th Africa-EU Cooperation Forum on ICT	Sigma	pro-iBiosphere project	December 2-3 2013	Addis Ababa, Ethiopia	Scientific community, policy makers	Not specific, large audience	Global
63	Workshop organised by pro-iBiosphere	MfN	Markup of biodiversity literature	February 10-11 2014	Berlin, Germany	Scientific community	21	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
64	Presentation during a seminar at the University of Bern	Plazi	Towards an (European) Open Biodiversity Knowledge Management System	March 17 2014	Bern, Switzerland	Scientific community	30	Switzerlandd
65	Hackathon organised by pro-iBiosphere	Naturalis	Biodiversity Data Enrichment Hackathon.	March 17-24 2014	Leiden, The Netherlands	Scientific community	37	Global
66	Presentation during a colloquium organised by Naturalis	Naturalis	Outcomes of the Biodiversity Enrichment Hackathon	March 24 2014	Leiden, The Netherlands	Scientific community, Media	c. 300	The Netherlands
67	Presentation during the 3rd Plenary of the Research Data Alliance, Legal Interoperability workgroup	Plazi	The Plazi wokflow and legal implications	March 27 2014	Dublin, Ireland	Scientific community, Policy makers	40	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
68	Press-release - Phys.org	Pensoft	New species discovery, description and data sharing in less than 30 days	March 27 2014	n.a.	Scientific community	not specific, large community	Global
69	Press-release - GBIF news	Naturalis	Providing instant access to data behind species discovery	March 27 2014	Copenhagen, Denmark	Scientific community, Media, policy makers	Large, non-specified	Global
70	Email sent to the pro-iBiosphere dissemination database	Pensoft	pro-iBiosphere Newsletter 5 (Jan-April 2014)	April 2014	n.a.	Scientific community, policy makers, civil society, industry	630	Global
71	Participation during the 2nd International Conference on Research Infrastructures	Naturalis	pro-iBiosphere project	April 2-4 2014	Athens, Greece	Scientific community, policy makers, EU, Other	763	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
72	Interview	Plazi	Big data, copyright and attribution in taxonomy	April 9 2014	n.a.	Scientific community	not specific, large community	Global
73	Email to EU BON dissemination database	Pensoft	Promotion of pro-iBiosphere Final Event	May 2014	n.a.	Scientific community, policy makers	not specific, large community	Global
74	Presentation during the iDiv workshop on "Plant Trait Semantics"	Biodiversity and Climate Research Centre	Presentation on FLOPO (use case addressed during the pro-iBiosphere Biodiversity Data Enrichment Hackathon)	May 6 2014	Leipzig, Germany	Scientific community	97	Global
75	Presentation during the Information Science & Technology Commission Report to CETAF35	FUB-BGBM	Update stable identifier initiative & discussion Bouchout Declaration	May 6-7 2014	Oslo, Norway	Scientific community, Other	<50	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
76	Presentation during the meeting of "Curators of Swiss Museums and Collections"	Plazi	Bouchout Declaration	May 23 2014	Frauenfeld, Switzerland	Scientific community, Other	80	Switzerland
77	Dissemination of Bouchout Declaration (in Bulgarian) to relevant Bulgarian NGOs and institutions	Pensoft	Bouchout Declaration	June 2014	Sofia, Bulgaria	Scientific community, policy makers	Large, non-specified	Bulgaria
78	Workshop organised by the pro-iBiosphere project	RBGK	Model Evaluation	June 9-10 2014	Mesie, Belgium	Scientific community	35	Global
79	Press release – Eurekalert	Pensoft	Classical monographs re-published in advanced open access	June 11 2014	n.a.	Scientific community, Media	not specific	Global
80	Press release - EU BON news	Pensoft	Classical monographs re-published in advanced open access	June 11 2014	n.a.	Scientific community, policy makers	not specific, large community	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
81	Dissemination of pro-iBiosphere Final Brochure – EU BON and pro-iBiosphere social media	Pensoft	OBKMS	June 11 2014	n.a.	Scientific community, policy makers	not specific, large community	Global
82	Press release – EurekAlert	Pensoft Publishers	Classical monographs re-published in advanced open access	June 11 2014	n.a.	Scientific community, Media	not specific, large community	Global
83	Demonstration organised by pro-iBiosphere	Naturalis	Outcomes of the pro-iBiosphere Data Enrichment Hackathon	June 11 2014	Meise, Belgium	Scientific community	49	Global
84	Workshop organised by pro-iBiosphere	FUB-BGBM	Biodiversity Catalogue	June 11 2014	Meise, Belgium	Scientific community	49	Global
85	Workshop organised by pro-iBiosphere	Sigma	pro-iBiosphere Final Conference	June 11 2014	Meise, Belgium	Scientific community	76	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
86	Press release – EurekAlert	Pensoft, Plazi, Naturalis	The Bouchout Declaration: A commitment to open science for better management of nature	June 12 2014	n.a.	Scientific community, Media	not specific	Global
87	News item on EU BON news	Pensoft	The Bouchout Declaration: A commitment to open science for better management of nature	June 12 2014	n.a.	Scientific community, policy makers	not specific, large community	Global
88	Press-release on EurekAlert	Pensoft Publishers	The Bouchout Declaration: A commitment to open science for better management of nature	June 12 2014	n.a.	Scientific community, Media	not specific, large community	Global
89	Press release - CORDIS Wire	Pensoft	Classical monographs re-published in advanced open access	June 13 2014	n.a.	Scientifc community, policy makers, other	not specific, large community	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
90	Meeting during the "Publisher XML standardization group"	MfN/ Pensoft	Harmonization of JATS formats delivered to PMC by OA publishers, based on recommendations from JATS Con 2014	June 18 2014	Cambridge, UK	Scientific community, Media, Other	10	Global
91	News on project website, tweet and LinkedIn post	Sigma	BioVeL and pro-iBiosphere collaboration	June 19 2014	n.a.	Scientific community	not specific, large community	European
92	Presentation and participation during the "JATS Data Citation group"	Pensoft, MfN	Sharing experiences and lessons learned from developing the TaxPub extension for JATS	June 19 2014	London, UK	Scientific community, Media, Other	20	Global
93	Presentation during the "Meeting of the Swiss scientific and academic libraries"	Plazi	Open Access and the Future of (Biodiversity-) Research	June 23 2014	Bern, Switzerland	Scientific community, Other	100	Switzerland

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
94	Presentation during the "8th International Congress of Hymenopterists"	Plazi	Cusco, Peru - A Step Towards (From) Read to Write Access to Taxonomic Publications	June 23 2014	Cusco, Peru	Scientific community	140	Global
95	Presentation and flyer dissemination during the "Forum Herbulot 2014"	Pensoft	How to accelerate the inventory of Biodiversity	July 1-4 2014	Schlettau, Germany	Scientific community	>100	Global
96	Presentation during the "Open Data Week"	Plazi	Bouchout Declaration	July 12 2014	Montpellier, France	Scientific community	50	France / global
97	Presentation during workshop	MfN	Data mining	July 18 2014	Berlin, Germany	Scientific community	30	
98	Presentation during the "16th International Congress of Myriapodology"	Pensoft	Blowing the dust off the old publications: an enhanced cybertaxonomic checklist of Eupolybothrus Verhoeff, 1907 (Chilopoda: Lithobiomorpha: Lithobiidae)	July 20-25 2014	Olomouc, Czech Republic	Scientific community	c. 100	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
99	Presentation and distribution of brochures during the Hymenopterists Meeting	Pensoft	pro-iBiosphere results	July 20-25 2014	Cusco, Peru	Scientific community	> 100	Global
100	Email sent to the pro-iBiosphere dissemination database	Pensoft	pro-iBiosphere Newsletter 6 (May-August 2014)	August 2014	n.a.	Scientific community, policy makers, civil society industry	630	Global
101	Presentation and dissemination of brochures during the "10th International Mycological Congress"	Pensoft	pro-iBiosphere project	August 3-8 2014	Bangkok, Thailand	Scientific community	>400	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
102	Presentation during the 8th International Congress of Dipterology	Plazi	Data conversion and pro-iBiosphere	August 10-15 2014	Potsdam, Germany	Scientific community	<300	Global
103	Presentation and dissemination of brochures during the 8th International Congress of Dipterology	Pensoft	pro-iBiosphere project	August 10-15 2014	Potsdam, Germany	Scientific community	>500	Global
104	Presentation during the Research Data Alliance 4th Plenary - Legal Interoperability work group	Plazi	The Plazi workflow and legal implications	September 22 2014	Amsterdam, The Netherlands	Scientific community, Policy makers	30	Global
105	Presentation and dissemination of flyers during the 8th Botanists of the Twenty-first Century	RBGK	Roles, Challenges and Opportunities	September 22-25 2014	Paris, France	Scientific community, policy makers	c. 3700	Global

TEMPLAT A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
106	Presentation during the SE-FINS Cluster Event: Bridging the Gap: Working together to tackle invasive non-native species in Europe	NBGB	Data: Opportunities, problems and solutions	September 24 2014	Norwich, UK	Scientific community, policy makers	>150	Belgium, France, UK, the Netherlands
107	Presentation during the Research Data Alliance, 4th plenary	Plazi	Persistent Identifiers and the Bouchout Declaration	September 24 2014	Amsterdam, The Netherlands	Scientific community, Media, policy makers	100	Global
108	Presentation during Neobiota 2014	NBGB	Using historic botanical literature to chart the geography, habitats and pathways of invasive species	November 3-8 2014	Antalya, Turkey	Scientific community, policy makers	c.200	Global

4.2.2. Section B

4.2.2.1 Part B1

Not applicable, the project did not produce any patents, trademarks or registered designs.

4.2.2.1 Part B2

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	Software interface between the PLAZI Repository and the EDIT Platform for Cybertaxonomy	No	No	The source code is available from the EDIT Platform Code Repository at http://cybertaxonomy.eu/cdmlib/source-repository.html.	M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service activities	The interface has been integrated into the core EDIT Java library and API. It is therefore automatically available in all present and future EDIT Platform instances.	The software is published under the open Mozilla Public Licence scheme.	The software is owned by the Freie Universität Berlin, BGBM.

[3] 19 A drop down list allows choosing the type of foreground: General advancement of knowledge, Commercial exploitation of R&D results, Exploitation of R&D results via standards, exploitation of results through EU policies, exploitation of results through (social) innovation.

[4] A drop down list allows choosing the type sector (NACE nomenclature) : http://ec.europa.eu/competition/mergers/cases/index/nace_all.html

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	Automated registration workflow for taxon names	No	No	IPNI (http://ipni.org) ZooBank (http://zoobank.org), Pensoft Journal System (PJS) (http://pensoft.net/ journals)	J58 - Publishing activities. M72 - Scientific research and development (Biodi versity Informatics). J62.0.9 - Other information technology and computer service activities	The automated registration wokflows are already implemented in the routine publishing practices of all Pensoft journals. IPNI and ZooBank created their registration user interfaces based on the XML formats and workflows developed by the automated registration pilot.	public domain	The workflow developed by IPNI, ZooBank and Pensoft is free to use for any journal and publisher
General advancement of knowledge	Bouchout Declaration for Open Biodiversity Management	No	No	The declaration can be found at: http://www.boucho utdeclaration.org/	J58 - Publishing activities. M72 - Scientific research and development (Biodi versity Informatics)	The Declaration was launched in June 2014	n.a.	Owner: Signatories of the Declaration. Other Beneficiaries involved: Plazi, MfN, will continue to coordinate activities and implementation of the Declaration

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	Registration system for IPNI and plant names. In order to make the registration tool generally more available, a version of the tool has been developed to incorporate an accepted community standard: the Taxonomic Concept Schema.	No	No.	Repository available at: www.ipni.org	J58 - Publishing activities. M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service activities	A pilot on a general purpose registration tool was released at the end of August 2014 and is currently being tested by members of the Special Committee on Registration of Names of the International Association of Plant Taxonomists prior to release for broader community testing. Recommendations arising from the review process (end of Sept. 2014) will be considered.	The workflow has been designed to be open source, but as its still being tested, it is not yet. We would hope to make it so once testing is complete. Once tested, the botanical community will benefit from a mechanism to register new nomenclatural acts such as new species or new combinations. Such acts will be more easily found, referenced and cited using their unique identifiers.	Owner: Partnership Royal Botanic Gardens Kew, Harvard University, Australian National Herbarium. Other beneficiaries involved: International Association of Plant Taxonomists

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	Markup workflow to convert traditionally published biodiversity literature into semantically enhanced data	No	No	Converted legacy publications into semantically enhanced data and linked open data	J58 - Publishing activities. M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service activities	June 2015	All the programs are open source	Owner: Plazi

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	Marked up Flora nd Fauna monographs are re-published in semantically enhanced, advanced open access. This allows unlimited and free use by anyone, as well as facilitates data and text mining, mobilization and re-use.	No	No	Re-publication platform, Advanved Books: is available on: http://advancedbooks.org The re-published volume 14 of Flora Malesiana is available at: http://ab.pensoft.net/articles.php?id=1141 Guidelines for submission of manuscripts for publication in the Advanced books platform available here: http://ab.pensoft.net/about#WhyAdvancedBooks	J58 - Publishing activities. M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service activities	The platform was launched in June 2014 with the re-publication of volume 14 of Flora Malesiana. Currently two more books are published on it, as a result of the EUBON and SCALES FP7 projects.	No. All re-published books an d underlying data are free to use	The underlying publishing platform on which advanced books are displayed is owned by Pensoft.

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	Expertise in copyright relating to biodiversity works and data.	No	NA	Deliverable 2.4.1 & 2.4.2	J58 - Publishing activities. M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service activities	Available and ongoing development process	n.a.	Plazi EU-BON is building their copyright and data sharing policies on Plazi expertise
General advancement of knowledge	Biodiversity data analytics. Aim: Extract statistical data about specimens for visualisation in a dashboard.	No	n.a.	*Code repository* https://github.com/Dauvit/Data_enrichment *Demo* http://plazi.cs.umb.edu/GgServer/srsStats	J58 - Publishing activities. M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service activities	The tangible outcomes of this effort, are being deployed on the Plazi server.	Open source licensing	Plazi, Open University, Naturalis

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	OCR correction. Aim: Provide a simple interface for interactively editing of OCR'd text, as well as tools to track the edits, to provide feedback to improve the OCR.	No	.No	*NCode repository* https://github.com/rdmpage/ocr-correction *Demo* http://bionames.org/~rpage/ocr-correction/index.php	J58 - Publishing activities .M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service activities	BHL will assess how to incorporate the code into their web presence. There are also plans to incorporate the code into a future release of BioStor (http://biostor.org, Page 2011).	Open source licensing	University of Glasgow, Landcare Research, New Zealand, Université de Montréal / Canadensys, University of Eastern Finland
General advancement of knowledge	Open access images. Aim: (i) Liberate and showcase openly-licensed (e.g. CC-BY) images from journal article PDFs and republish on image sharing social media sites. (ii) Find images of phylogenetic trees for data re-extraction from the image.	No	n.a.	*Code repository* https://github.com/rossmounce/LeidenPDFhack *Demo* http://www.flickr.com/photos/79472036@N07/sets/72157642597074643/	J58 - Publishing activities. M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service activities	Potential scopes for follow-up include teaming up with open access journals to find a stable home for this novel method of exposing images, and harvesting and processing images of phylogenetic trees (e.g. using TreeRipper, Hughes 2011).	Open source licensing	University of Bath, Naturalis

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	Trait ontology. Aim: Extract and ontologise plant trait data from digitised Floras.	No	n.a.	*Code repository* https://github.com/leechuck/plantphenotypes *Demo* http://bioportal.bioontology.org/ontologies/FLOPO	M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service	Ongoing work aims to connect the extracted taxon names in the floras using identifiers from the International Plant Names Index (IPNI, http://ipni.org) and represent them together with their environment and FLOPO annotations as Linked Data in an RDF store. Work will also continue to support multiple languages: the current examples are in English and French, but there is a strong case for supporting other languages.	Open source licensing	Royal Botanic Gardens Kew, National Botanic Garden Belgium, Naturalis, Aberystwyth University, Biodiversity and Climate Research Centre - Senckenberg Nature Research Society
General advancement of knowledge	SWeDe. Aim: Produce a standard for describing scientific web services.	No	n.a.	*Code repository* https://github.com/njall/XS-SWeDe, https://github.com/njall/SWeDe-Farmer *Demo* http://swede-farmer.herokuapp.com/	M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service	A rudimentary application (code named "SWeDe farmer") to simplify generation of SWeDe documents was developed. BiodiversityCatalogue, the registry of biodiversity web services (http://biodiversitycatalogue.org), will be adopting this standard by implementing a SWeDe parser that can periodically update catalogue entries based on the state of a service provider's SWeDe document.	Open source licensing	Institute of Biomembranes and Bioenergetics - Italian National Research Center, Freie Universität Berlin, University of Manchester

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	Specimen links. Aim: (i) Link together name and specimen data, especially from Floras. (ii) Link specimen citations in "Literature" to specimens from Kew, Brussels and Edinburgh. (iii) Proof of concept and requirements gathering for Taxonomic MindMapper.	No	n.a.	*Code repository* https://github.com/RBGKew/leiden-hackathon *Demo* http://gist.neo4j.org/?9684109	M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service	The application of graph database technology (Neo4J, http://neo4j.org) towards the Taxonomic Mind Mapper demonstrated an attractively low entrance barrier to linking, exploring and visualising disparate collection data. As follow-up, a publication discussing this novel way of representing taxon concepts enabled by this approach is in preparation.	Open source licensing	Pensoft, Royal Botanic Gardens Kew, National Botanic Garden Belgium, Landcare Research, New Zealand, University of Eastern Finland

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General advancement of knowledge	Occurrence Catalogue REST API for the EDIT Platform for Cybertaxonomy	No	n.a.	*Code repository* EDIT Platform Code Repository: http://cybertaxonomy.eu/cdmlib/source-repository.html. *Demo:* http://dev.e-taxonomy.eu/cdmserver/chenopodiumPilot/occurrence_catalogue.json?query=95402484-06cc-4284-a84c-193be51489ea&pageNumber=0&pageSize=10. The service itself is accessible at: http://cybertaxonomy.eu/cdmlib/rest-api-occurrence-catalogue.html	M72 - Scientific research and development (Biodiversity Informatics) J62.0.9 - Other information technology and computer service	The product is available for all instances of the EDIT Platform (presently 25). Each new instances is automatically equipped with the service without any restrictions. The web service has been developed in the CDM library and has been committed to the CDM Subversion repository. The JSON output can be, for example, re-injected into the BioVel refinement workflow. In a follow-up, the facility could be enhanced by using a Lucene-based index to speed up the export function.	The software is published under the open Mozilla Public Licence scheme	Freie Universität Berlin-BGBM, National Botanic Garden Belgium

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
Exploitation of R&D results via standards	iPython notebook/Taverna. Aim: Access Taverna workflows from within iPython notebook.	No	n.a.	*Code repository* - https://github.com/myGrid/DataHackLeiden. *Demo* - http://nbviewer.ipython.org/urls/raw.githubusercontent.com/myGrid/DataHackLeiden/alan/Player_example.ipynb?create=1	M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service	One of the deliverables of the BioVeL project (Vicario et al. 2011) is a workflow "player" that can be embedded in HTML, exposing the functionality of a given Taverna workflow within a web page. This breakout group sought to combine these two tools, such that iPython notebook users are able to submit data assembled during their analysis to a Taverna workflow embedded in a player and retrieving the results. The participants succeeded in this, resulting in a more integrated, powerful research environment that exposes remote HPC resources (such as BioVeL services) to iPython notebook users. The results are available for easy installation as the Pypi package *tavernaPlayerClient*.	Open source licensing	Naturalis; University of Bath; Software Sustainability Institute, myGrid; University of Manchesterer

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
Exploitation of R&D results via standards	BioVeL/NeXML services. Aim: Deliver RESTful services to merge and query phylogenetic data and metadata.	No	n.a.	*Code repository* - https://github.com/naturalis/biovel-nbc. *Demo* - http://biovel.naturalis.nl/biovel?service=NeXMLExtractor&nexml=http://bit.ly/1mR11Yz&object=Trees	M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service	To make the NeXML merger and extractor tools easily accessible for the biodiversity research community and to enable their integration into existing workflows, they are implemented as RESTful web services hosted by Naturalis Biodiversity Center and made available in the BiodiversityCatalogue (https://www.biodiversitycatalogue.org/services/70). Preliminary tests of the NeXML merger and extractor have been conducted using data inputs and outputs used by the phylogenetic service set of BioVeL (Vicario et al. 2012) (https://www.biodiversitycatalogue.org/services/31/service_endpoint); NeXML extractor output has also been visualised as a phylogenetic tree with its taxon-associated metadata by implementing an ITOL (http://itol.embl.de/)tool wrapper within a Taverna workflow.	Open source licensing	Institute of Biomembranes and Bioenergetics - Italian National Research Center, University of Manchester

Type of Exploitable Foreground[3]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[4]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
Exploitation of R&D results via standards	Standardised HTTP URI-based identifiers for collection objects	No	n.a.	Best practices for HTTP-based stable identifiers: http://wiki.pro-ibiosphere.eu/wiki/Best_practices_for_stable_URIs. Source code and example documents: http://sourceforge.net/projects/stablecollectionidentifiers/	M72 - Scientific research and development (Biodiversity Informatics). J62.0.9 - Other information technology and computer service activities	Implementation in Natural History Institutions started during the pro-iBiosphere project. The proposed mechanisms for identifying collection objects are becoming an official CETAF product in 2014. Partner institutions have adopted identifiers building on the work done in pro- iBiosphere including: RBGK,	The standard itself follows Linked open data principles and is therefore free to use. Software components helping to implement stable identifiers in Natural History Institutions are freely available from a SourceForge repository.	n.a.

Software Interface between PLAZI and the EDIT Platform for Cybertaxonomy. The software has been implemented as part of the EDIT Platform for Cybertaxonomy Java API and provides a mechanism for importing semi-structured literature documents held in the PLAZI repository into the highly-granular CDM-based data store coming with the EDIT Platform. The interface is capable of parsing free-text strings and provides feed-back mechanisms for ambiguous content. The target user group are biodiversity research groups or individual scientists who would like to incorporate data contained in literature in web portals and data services. The software comes with every installation of an EDIT Platform instance (presently 25). We foresee the number of instances to double over the next two years with the increasing interest of research groups using the platform. We will actively promote the possibilities for integrating data from literature in a handbook (part of the2015 work programme of the BGBM Biodiversity Research Group), in workshops, and conferences.

Automated registration workflow for taxon names. The pilot developed in the project has been used by the International Plant Name Index to further develop its registration of plant names. The new registration system is currently being tested by members of the Special Committee for Registration of the International Association of Plant Taxonomists. ZooBank, which is the official registry of the International Commission of Zoological Nomenclature, used the pilot to establish a workflow and corresponding interface that allows free automated registration of new taxon names by any journal or publisher which want to use that. Implementation of automated workflow and invention of XML-based tools will facilitate the process of publication and dissemination of biodiversity information and greatly benefit publishers of biodiversity information, global taxonomic indexers and the taxonomic community as a whole. The automated registration wokflows are already being incorporated in Pensoft's open access journals.

Further development of a automated registration system for taxon names Registration system for IPNI and plant names. The registering of nomenclatural acts will make referencing, citing and discovering the plant names easier, facilitating greater reuse and efficiencies in the scientific process. Once tested and released the system will provide the ability to find and unambiguously use plant names which will benefit the whole biodiversity community. Without such a system, digitally published name information can easily be missed, leading to duplication and confusion. A registration system , supported by several registries, learning from the model being currently tested, will greatly improve the discovery and synthesis of biodiversity knowledge. Registration acts will be open access. Testing is currently underway. Recommendations arising will be pursued. Expected impact: Over 6000 nomenclatural acts at species level alone are published every year. Registration will enable better management of this information.

Bouchout Declaration. The Declaration will be used to further promote its principles established in the Pro-iBiosphere project. This will be done through continued lecturing), and a review within approximately two years after the launch. It is also discussed to provide specific answers for question in regards of the implementation of the Declaration.

Markup workflow to convert traditionally published biodiversity literature into semantically enhanced data. Plazi is working on the implementation of a workflow including several sub-workflows to expedite the markup of legacy literature (see Fig 16 Year two report) that is based on discussion during the pro-iBiosphere project.

Copyright and biodiversity. Plazi will use the results and conclusions from the legal studies of the Pro-iBiosphere project to propose legal exemptions for scientific use of works both at EU as well as national level (Switzerland).

Biodiversity data analytics. The advent of biodiversity data aggregators such as Plazi (Agosti and Egloff 2009) have made it possible to perform analytics on biodiversity data use and re-use. The initial use case pitch for this group proposed a "dashboard" with various graphs showing, for example, specimen citations, or the output of researchers or institutions, in order to assist with data analytics by being able to visualise and thereby better understand it, and assisting with its quality control. At the Hackathon, the members of the breakout group addressed this challenge from two ends: on the server side, GoldenGATE was enhanced with a search facility that outputs selected search predicates as structured data in various views; on the web browser (i.e. client) side, this structured data was consumed by the jQuery JavaScript framework (http://jquery.com/) in conjunction with the plugins jqPlot (http://www.jqplot.com/) and jVectorMap (http://jvectormap.com/) for data visualisation. The tangible outcomes of this effort, once stable, will be deployed on the Plazi server.

Standardised HTTP URI-based identifiers for collection objects. The system provides a standardised mechanism to expose collection objects and associated metadata as Linked Open Data (LOD) on the web. It consists of a wikipage providing best practices for identifier syntax and redirection mechanisms as well as a SourceForge repository for (example) software scripts helping intuitions to adopt the technology. The target audience is Natural History Institutions as well as data centres who need a persistent identifier system for publishing data on the web. The primary target audience is the Consortium of European Taxonomic Facilities with presently 32 member organisations. In autumn 2014 the proposed identifier system will become an official CETAF-product and will be further promoted in the CETAF Information Science and technology Commission (CETAF-ISTC). The pro-iBiosphere consortium will continues to advertise the system as a fundamental component for the emerging global biodiversity informatics infrastructures in working groups and conferences.

OCR correction. The Biodiversity Heritage Library (BHL) is a collaborative effort of natural history and botanical libraries to digitise their physical medium literature and make it available as open access publications, forming a "biodiversity commons". The effort has resulted in large volumes of digital documents obtained by optical character recognition (OCR) of scanned, physical documents. Although the OCR quality is generally very high, errors that require human intervention to correct them are unavoidable. This breakout group addressed this challenge by developing a collaborative platform where authenticated users can correct OCR documents rendered on webpages (via DjVu XML format, http://djvu.org), with the provenance of the edits recorded in the margin of the document. The facility is "intelligent" in that it attempts to detect patterns in OCR errors (e.g. ü is recognised as ii) in order to suggest subsequent corrections to the same class of errors. In addition, the facility cross-references taxonomic names against web services provided by GlobalNames (http://globalnames.org) to normalise these and correct them consistently throughout a document. The participants of the breakout group have communicated with BHL to assess how this code might be incorporated into their web presence. There are also plans to incorporate the code into a future release of BioStor (http://biostor.org, Page 2011).

Open access images. As public access biodiversity literature is growing in volume, so too are the graphs and pictures embedded in these publications. Unfortunately, these images are "buried" in

that up till now they could only be located within their publication. To mobilise and expose these images, this breakout group developed a pipeline that extracts embedded images from open access publications (as a proof of concept this was done by harvesting PDF documents from the journal *Phytotaxa*), pre-processes them (e.g. discard spurious, small figures such as journal logos; invert image negatives) and uploads them to the social web, i.e. to the photo sharing site Flickr, with periodic notifications to the Twitter account @PhytoFigs. To date over 2000 figures have been uploaded this way. Of these, some are the very first *discoverable* images (not embedded in PDF) for those taxa to be made available on the internet under an OKD-compliant open license (http://opendefinition.org) - allowing easy re-use even for commercial purposes without the hassle of asking for permission(s). Potential scopes for follow-up include teaming up with open access journals to find a stable home for this novel method of exposing images, and harvesting and processing images of phylogenetic trees (e.g. using TreeRipper, Hughes 2011).

Trait ontology. Floristic treatments of geographic regions have a long history and have resulted in a wealth of literature. To increase the accessibility of the knowledge contained in these floras, past efforts have focused on structured markup, resulting in websites such as the Digitised Flora of Central Africa and formats such as FlorML (Hamann et al. 2014). For the hackathon, marked-up floras were available that cover a large proportion of tropical plants. These were the *Flora Malesiana, Flora Zambesiaca, Flore d'Afrique Centrale* and *Flore du Gabon*. At the hackathon, a use case was presented calling for the creation of an RDF knowledge base of plant phenotypes by extracting trait data from such digitised floras. The ensuing breakout group addressed this challenge by text mining of documents to mark up relevant concepts with terms from the Plant Ontology (PO, to normalise the inhering plant anatomical parts) and the Phenotypic Quality Ontology (PATO, to normalise both "phenotypic qualities", such as red, serrated and small, and "traits", such as color, shape, and size). Under this approach, phenotypes are extracted from flora descriptions as Entity/Quality (EQ) statements (Gkoutos et al. 2005). From these EQ statements, an ontology was constructed that takes into account the anatomical relationships between plant parts (as represented in PO) as well as information about traits and values (as represented in PATO, Hoehndorf et al. 2010). The resulting ontology, the Flora Phenotype Ontology (FLOPO), was deposited at NCBO Bioportal. FLOPO consists of more than 25,000 classes describing plant traits and phenotypes, and every class in FLOPO has at least one taxon annotation in one of the processed floras. Work subsequent to the hackathon has additionally yielded annotations of the floras with terms obtained from the Environment Ontology (ENVO), providing a proof of concept of an ontology-mediated knowledge base that can be brought to bear on a variety of research questions on phenotype/environment interactions, functional diversity, and community ecology. Furthermore, ongoing work aims to connect the extracted taxon names in the floras using identifiers from the International Plant Names Index (IPNI, http://ipni.org) and represent them together with their environment and FLOPO annotations as Linked Data in an RDF store. Work will also continue to support multiple languages: the current examples are in English and French, but there is a strong case for supporting other languages.

SWeDe. The wide adoption of the REST (Representational State Transfer) design pattern for simple, stateless web services has resulted in a proliferation of different ways in which clients need to interact with biodiversity data and computational web services. Although the adoption of the design pattern has considerably lowered the barrier for developers to implement and deploy simple services, it poses challenges to clients and end users, as the way in which they are expected to interact with these heterogeneous service interfaces is not always documented clearly. Examples of

this include the various "RESTful" as well as SOAP/WSDL-based web services delivered to the BioVeL project (Vicario et al. 2012). This breakout group sought to address this issue by defining a standard for documenting such web services using XML. The resulting XML schema, SWeDe (Scientific Web-service Description), provides developers with a standard way to define metadata of their service (e.g. authorship, license, suggested citation), the inputs and outputs with the available parameters and their ranges, and usage examples. The SWeDe schema re-uses several components from the Access to Biological Collections Data (ABCD) Schema (Berendsohn 2005). In addition, the members of this breakout group have developed a rudimentary application (code named "SWeDe farmer") to simplify generation of SWeDe documents. BiodiversityCatalogue, the registry of biodiversity web services (http://biodiversitycatalogue.org), will be adopting this standard by implementing a SWeDe parser that can periodically update catalogue entries based on the state of a service provider's SWeDe document.

Specimen links. Several related use cases were addressed with the aim of better linking biodiversity data to form a navigable "knowledge graph". As a specimen-oriented example: specimens obtained from single collection events are often distributed among herbaria, where they tend to take on an isolated life of their own, with their provenance history and annotations uncoupled from other specimens collected for the same species, even during the same collection event. This fragmentation poses challenges when integrating collection data and floristic knowledge. Software previously developed at RBG Kew to build links between data sets based on configurable text-based rules was used to link up datasets, particularly to detect duplicates between digitised herbaria - Kew (K), Meise (BR), Edinburgh (E) and Naturalis - and to create specimen level citation links from scientific papers published in *PhytoKeys*. Numerous duplicate specimens among the collections were encountered, drawing attention to a greater need for collaboration and data integration among natural history collections. The software toolkit is implemented in Java, Spring and Lucene, with a JSON format web interface conforming to the Open Refine (formerly Google Refine) reconciliation service API, and the group will follow up by open-sourcing this toolkit and the rule sets used to configure matches. Frictionless integration of collection and specimen information would permit novel ways of specifying taxon concepts, relying on the emergent properties of graphs linking specimens by conspecificity (including provenance and evidence for the assertions). This was suggested by a use case pitching a "Taxonomic Mind Mapper" to enable exploration of this method of modeling and representing taxon concepts. The application of graph database technology (Neo4J, http://neo4j.org) towards the Taxonomic Mind Mapper demonstrated an attractively low entrance barrier to linking, exploring and visualising disparate collection data. As follow-up, a publication discussing this novel way of representing taxon concepts enabled by this approach is in preparation.

Occurrence Catalogue REST API for the EDIT Platform for Cybertaxonomy. The Common Data Model (CDM, Berendsohn et al. 2011) consists of a database schema and application programming interface that provide a persistence and publishing mechanism for nomenclature, taxonomy, descriptive data, media, geographic information, literature, specimens, and persons. As such, instances of CDM databases provide a wealth of structured data that can be re-used in a variety of research contexts, such as in species distribution modeling. To enable such re-use, CDM databases should be enhanced with a machine-readable data harvesting service. To address this challenge, this breakout group developed a service to look up and export occurrence data (specimen, observation) from CDM instances. The lookup is based on a single Taxon UUID that can be obtained from the Name-Catalogue web service (provided by the CDM). To prevent memory overload, both on the server and

on the client side, a paging system is used. The web service has been developed in the CDM library and has been committed to the CDM Subversion repository. The JSON output can be, for example, re-injected into the BioVel refinement workflow. In a follow-up, the facility could be enhanced by using a Lucene-based index to speed up the export function.

The new API for opens up occurrence data held in EDIT Platform instances for re-use in other service-oriented systems and in workflows such as the Taverna-based workflows developed by BioVeL. The target user groups are scientists who are using the EDIT Platform already and want to conduct analyses of species distributions using advanced workflow environments. The software comes with every installation of an EDIT Platform instance (presently 25). We foresee the number of instances to double over the next two years with the increasing interest of research groups using the platform. The possibility of combining EDIT Platform instances with BioVeL workflows will be actively promoted in EDIT workshops and tutorials.

iPython notebook/Taverna. iPython notebook (Perez and Granger 2007) provides an interactive computational environment within a web browser in which users can write and execute code written in the Python programming language. This code may be combined with text, mathematical and statistical calculations, production of plots and HTML display to produce shareable and re-usable notebooks. The notebooks can be shared on the iPython Notebook Viewer. Taverna (Wolstencroft et al. 2013) provides a suite of tools for workflow design, editing and execution. This includes the Taverna Workbench, the main creation tool for workflows. The workflows allow the coordination of services, including RESTful or SOAP web services, R scripts and command line tools. Taverna Server enables you to set up a dedicated server for executing workflows remotely, and it can be accessed by a WSDL or a REST API. One of the deliverables of the BioVeL project (Vicario et al. 2011) is a workflow "player" that can be embedded in HTML, exposing the functionality of a given Taverna workflow within a web page. This breakout group sought to combine these two tools, such that iPython notebook users are able to submit data assembled during their analysis to a Taverna workflow embedded in a player and retrieving the results. The participants succeeded in this, resulting in a more integrated, powerful research environment that exposes remote HPC resources (such as BioVeL services) to iPython notebook users. The results are available for easy installation as the Pypi package *tavernaPlayerClient*.

BioVeL/NeXML services. Biodiversity-oriented phylogenetics workflows usually involve various software tools connected in series that consume and produce different types of data. NeXML is an XML standard that supports the representation of (among others) taxa, character-state matrices, phylogenetic trees and semantic annotations within one single document and is therefore specifically tailored to ease the interplay of different tools in phylogenetic analysis (Vos et al. 2012). Since XML documents are generally intended to be handled by software rather than by users directly, a software tool to easily manipulate NeXML files appears desirable. To this end, the scope of this task group was to develop software that can i) construct NeXML documents from data encoded in commonly-used phylogenetic file formats or add metadata to an existing NeXML document (*NeXML merger*), and ii) extract information defined by the user from a given NeXML file (*NeXML extractor*). To make the NeXML merger and extractor tools easily accessible for the biodiversity research community and to enable their integration into existing workflows, they are implemented as RESTful web services hosted by Naturalis Biodiversity Center and made available in the BiodiversityCatalogue (https://www.biodiversitycatalogue.org/services/70). Preliminary tests of the NeXML merger and extractor have been conducted using data inputs and outputs used by the phylogenetic service set of

BioVeL (Vicario et al. 2012) (https://www.biodiversitycatalogue.org/services/31/service_endpoint); NeXML extractor output has also been visualised as a phylogenetic tree with its taxon-associated metadata by implementing an ITOL (http://itol.embl.de/)tool wrapper within a Taverna workflow.

Re-publication of legacy data. The re-publication workflow based on the FlorML and TaxPub XML schemas and executed through Pensoft's new Advanced Books platform was tested through volume 14 of Flora Malesiana. The new workflow demonstrates a re-publication of a volume of Flora Malesiana in a semantically enriched HTML edition and opens new horizons for semantic book publishing. The re-publication workflow and platform tested and modelled by pro-iBiosphere was later used to re-publish the Flora of Northumberland and Durham originally published in 1831, and to geotag, extract and submit to GBIF historical localities of plants as a distinct dataset. The target audience are scientists, but this product will also be of benefit for publishers and librarians. With the new platform, such scientifically important historical monographs, enriched with additional information from up-to-date external sources related to organisms' names, species treatments, information on their ecology, distribution and conservation value, morphological characters, etc., become freely usable for anyone at any place in the world.

4.3 Report on societal implications

A General Information *(completed automatically when **Grant Agreement number** is entered.*

Grant Agreement Number:	312848
Title of Project:	Coordination and policy development in preparation for a European Open Biodiversity Knowledge Management System, addressing Acquisition, Curation, Synthesis, Interoperability and Dissemination
Name and Title of Coordinator:	Dr. Soraya Sierra

B Ethics

1. Did your project undergo an Ethics Review (and/or Screening)? • If Yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project reports? Special Reminder: the progress of compliance with the Ethics Review/Screening Requirements should be described in the Period/Final Project Reports under the Section 3.2.2 *'Work Progress and Achievements'*	***No***
2. Please indicate whether your project involved any of the following issues (tick box) :	***No***
RESEARCH ON HUMANS	
• Did the project involve children?	No
• Did the project involve patients?	No
• Did the project involve persons not able to give consent?	No
• Did the project involve adult healthy volunteers?	No
• Did the project involve Human genetic material?	No
• Did the project involve Human biological samples?	No
• Did the project involve Human data collection?	No
RESEARCH ON HUMAN EMBRYO/FOETUS	
• Did the project involve Human Embryos?	No
• Did the project involve Human Foetal Tissue / Cells?	No
• Did the project involve Human Embryonic Stem Cells (hESCs)?	No
• Did the project on human Embryonic Stem Cells involve cells in culture?	No
• Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?	No
PRIVACY	
• Did the project involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	No

• Did the project involve tracking the location or observation of people?	No
RESEARCH ON ANIMALS	
• Did the project involve research on animals?	No
• Were those animals transgenic small laboratory animals?	No
• Were those animals transgenic farm animals?	No
• Were those animals cloned farm animals?	No
• Were those animals non-human primates?	No
RESEARCH INVOLVING DEVELOPING COUNTRIES	
• Did the project involve the use of local resources (genetic, animal, plant etc)?	No
• Was the project of benefit to local community (capacity building, access to healthcare, education etc)?	No
DUAL USE	
• Research having direct military use	No
• Research having the potential for terrorist abuse	No

C Workforce Statistics

3. Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).

Type of Position	Number of Women	Number of Men
Scientific Coordinator	1	0
Work package leaders	1	5
Experienced researchers (i.e. PhD holders)	2	16
PhD Students	0	0
Other	7	8

4. How many additional researchers (in companies and universities) were recruited specifically for this project?	3
Of which, indicate the number of men:	1

D Gender Aspects

5. Did you carry out specific Gender Equality Actions under the project?

○	Yes
X	No

6. Which of the following actions did you carry out and how effective were they?

		Not at all effective				Very effective
❑	Design and implement an equal opportunity policy	○	○	○	○	○
❑	Set targets to achieve a gender balance in the workforce	○	○	○	○	○
❑	Organise conferences and workshops on gender	○	○	○	○	○
❑	Actions to improve work-life balance	○	○	○	○	○
X	Other:	An equal opportunity policy has been part of the standard recruiting strategy of the partners.				

7. Was there a gender dimension associated with the research content – i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?

○ Yes- please specify

X No

E Synergies with Science Education

8. Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?

○ Yes- please specify

X No

9. Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?

○ Yes- please specify

X No

F Interdisciplinarity

10. Which disciplines (see list below) are involved in your project?

X Main discipline[5]: 1.5 Biological Sciences

X	Associated discipline[5]: 1.1 Mathematics and computer sciences	X	Associated discipline[5]: 1.4 Earth and related environmental sciences

G Engaging with Civil society and policy makers

[5] Insert number from list below (Frascati Manual).

11a Did your project engage with societal actors beyond the research community? *(if 'No', go to Question 14)*	X ○	Yes No

11b If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)?

- ○ No
- X Yes- in determining what research should be performed
- ○ Yes - in implementing the research
- X Yes, in communicating /disseminating / using the results of the project

11c In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?	X ○	Yes No

12. Did you engage with government / public bodies or policy makers (including international organisations)

- ○ No
- X Yes- in framing the research agenda
- ○ Yes - in implementing the research agenda
- X Yes, in communicating /disseminating / using the results of the project

13a Will the project generate outputs (expertise or scientific advice) which could be used by policy makers?

- X Yes – as a **primary** objective (please indicate areas below- multiple answers possible)
- ○ Yes – as a **secondary** objective (please indicate areas below - multiple answer possible)
- ○ No

13b If Yes, in which fields?

Agriculture		Energy		Human rights	
Audiovisual and Media		Enlargement		Information Society	
Budget		Enterprise		Institutional affairs	
Competition		**Environment**	**X**	Internal Market	
Consumers		External Relations		Justice, freedom and security	
Culture		External Trade		Public Health	
Customs		Fisheries and Maritime Affairs		Regional Policy	
Development Economic and Monetary Affairs		Food Safety		**Research and Innovation**	**X**
		Foreign and Security Policy		Space	
Education, Training, Youth		Fraud		Taxation	
Employment and Social Affairs		Humanitarian aid		Transport	

13c If Yes, at which level?

- ○ Local / regional levels
- ○ National level
- X European level
- X International level

H Use and dissemination

14. How many Articles were published/accepted for publication in peer-reviewed journals?	**13**
To how many of these is open access[6] provided?	**12**
How many of these are published in open access journals?	**12**
How many of these are published in open repositories?	**0**
To how many of these is open access not provided?	**1**
Please check all applicable reasons for not providing open access:	
❑ publisher's licensing agreement would not permit publishing in a repository ❑ no suitable repository available ❑ no suitable open access journal available ❑ no funds available to publish in an open access journal ❑ lack of time and resources ❑ lack of information on open access X other[7]: ……	Taxon is a well-established journal in the taxonomic community with subscriptions available for almost all European Natural History Facilities. It is important for pro-iBiosphere to use this publication channel in addition to open access journal.

15. How many new patent applications ('priority filings') have been made? *("Technologically unique": multiple applications for the same invention in different jurisdictions should be counted as just one application of grant).*		**0**
16. Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).	Trademark	**0**
	Registered design	**0**
	Other	**0**
17. How many spin-off companies were created / are planned as a direct result of the project?		**0**

[6] Open Access is defined as free of charge access for anyone via Internet.
[7] For instance: classification for security project.

Indicate the approximate number of additional jobs in these companies:	

18. Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:

❑	Increase in employment, or	❑	In small & medium-sized enterprises
❑	Safeguard employment, or	❑	In large companies
❑	Decrease in employment,	X	None of the above / not relevant to the project
❑	Difficult to estimate / not possible to quantify		

19. For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (*FTE = one person working fulltime for a year*) jobs:	*Indicate figure:*
Difficult to estimate / not possible to quantify	X

I Media and Communication to the general public

20. As part of the project, were any of the beneficiaries professionals in communication or media relations?

X Yes ○ No

21. As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?

X Yes ○ No

22 Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?

X	Press Release	X	Coverage in specialist press
X	Media briefing	X	Coverage in general (non-specialist) press
❑	TV coverage / report	❑	Coverage in national press
X	Radio coverage / report	❑	Coverage in international press
X	Brochures /posters / flyers	X	Website for the general public / internet
❑	DVD /Film /Multimedia	X	Event targeting general public (festival, conference, exhibition, science café)

23 In which languages are the information products for the general public produced?

❑	Language of the coordinator	X	English
X	Other language(s) The Bouchout Declaration (http://www.bouchoutdeclaration.org) has been translated into various languages		

***Question F-10*:** Classification of Scientific Disciplines according to the Frascati Manual 2002 (Proposed Standard Practice for Surveys on Research and Experimental Development, OECD 2002):

FIELDS OF SCIENCE AND TECHNOLOGY

1. NATURAL SCIENCES
1.1 Mathematics and computer sciences [mathematics and other allied fields: computer sciences and other allied subjects (software development only; hardware development should be classified in the engineering fields)]
1.2 Physical sciences (astronomy and space sciences, physics and other allied subjects)
1.3 Chemical sciences (chemistry, other allied subjects)
1.4 Earth and related environmental sciences (geology, geophysics, mineralogy, physical geography and other geosciences, meteorology and other atmospheric sciences including climatic research, oceanography, vulcanology, palaeoecology, other allied sciences)
1.5 Biological sciences (biology, botany, bacteriology, microbiology, zoology, entomology, genetics, biochemistry, biophysics, other allied sciences, excluding clinical and veterinary sciences)

2 ENGINEERING AND TECHNOLOGY
2.1 Civil engineering (architecture engineering, building science and engineering, construction engineering, municipal and structural engineering and other allied subjects)
2.2 Electrical engineering, electronics [electrical engineering, electronics, communication engineering and systems, computer engineering (hardware only) and other allied subjects]
2.3. Other engineering sciences (such as chemical, aeronautical and space, mechanical, metallurgical and materials engineering, and their specialised subdivisions; forest products; applied sciences such as geodesy, industrial chemistry, etc.; the science and technology of food production; specialised technologies of interdisciplinary fields, e.g. systems analysis, metallurgy, mining, textile technology and other applied subjects)

3. MEDICAL SCIENCES
3.1 Basic medicine (anatomy, cytology, physiology, genetics, pharmacy, pharmacology, toxicology, immunology and immunohaematology, clinical chemistry, clinical microbiology, pathology)
3.2 Clinical medicine (anaesthesiology, paediatrics, obstetrics and gynaecology, internal medicine, surgery, dentistry, neurology, psychiatry, radiology, therapeutics, otorhinolaryngology, ophthalmology)
3.3 Health sciences (public health services, social medicine, hygiene, nursing, epidemiology)

4. AGRICULTURAL SCIENCES
4.1 Agriculture, forestry, fisheries and allied sciences (agronomy, animal husbandry, fisheries, forestry, horticulture, other allied subjects)
4.2 Veterinary medicine

5. SOCIAL SCIENCES
5.1 Psychology
5.2 Economics
5.3 Educational sciences (education and training and other allied subjects)
5.4 Other social sciences [anthropology (social and cultural) and ethnology, demography, geography (human, economic and social), town and country planning, management, law, linguistics, political sciences, sociology, organisation and methods, miscellaneous social sciences and interdisciplinary ,

methodological and historical S1T activities relating to subjects in this group. Physical anthropology, physical geography and psychophysiology should normally be classified with the natural sciences].

6. HUMANITIES

6.1 History (history, prehistory and history, together with auxiliary historical disciplines such as archaeology, numismatics, palaeography, genealogy, etc.)

6.2 Languages and literature (ancient and modern)

6.3 Other humanities [philosophy (including the history of science and technology) arts, history of art, art criticism, painting, sculpture, musicology, dramatic art excluding artistic "research" of any kind, religion, theology, other fields and subjects pertaining to the humanities, methodological, historical and other S1T activities relating to the subjects in this group]

FINAL REPORT ON THE DISTRIBUTION OF THE European Union FINANCIAL CONTRIBUTION

The "Final report on the distribution of the European Union Financial contribution" will be submitted to the Commission within 30 days after receipt of the final payment of the European Union financial contribution.

Name of beneficiary	Final amount of EU contribution per beneficiary in Euros
1. Naturalis Biodiversity Center, the Netherlands (Naturalis)	
2. Nationale Plantentuin van België, Belgium (NBGB)	
3. Freie Universität Berlin, Germany (FUB-BGBM)	
4. Pensoft Publishers Ltd, Bulgaria (Pensoft)	
5. Sigma Orionis, France (Sigma)	
6. The Royal Botanic Gardens, Kew, United Kingdom (RBGK)	
7. Plazi, Switzerland (Plazi)	
8. Museum für Naturkunde Berlin, Germany (MfN)	
Total	