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

FRONT PAGE

PROJECT PERIODIC 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: (1st) 2nd 3rd 4th

Period covered: from Sep. 2012 to Aug. 2013

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

E-mail: soraya.sierra@naturalis.nl

Project website address: www.pro-ibiosphere.eu

Declaration by the scientific representative of the project coordinator

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

- The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;

- The project (tick as appropriate) [1]:

    X   has fully achieved its objectives and technical goals for the period;

    ☐ has achieved most of its objectives and technical goals for the period with relatively minor deviations.

    ☐ has failed to achieve critical objectives and/or is not at all on schedule.

- The public website, if applicable

    X is up to date

    ☐ is not up to date

- To my best knowledge, the financial statements which are being submitted as part of this report are in line with the actual work carried out and are consistent with the report on the resources used for the project and if applicable with the certificate on financial statement.

- All beneficiaries, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SMEs, have declared to have verified their legal status. Any changes have been reported under the section 3.2.3 (Project Management) in accordance with Article II.3.f of the Grant Agreement.

Name of scientific representative of the Coordinator: Soraya Sierra

Date: 31/ 10/2013

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

---

[1] If either of these boxes below is ticked, the report should reflect these and any remedial actions taken.

pro-iBiosphere FP7 Project ∎ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ∎7th Framework Programme
∎ Coordination and support action FP7-INFRASTRUCTURES-2012-1∎ Subprogram area INFRA-2012-3.3

Page 2 of 85

Table of Contents

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 3 of 85

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 4 of 85

3.1 Publishable Summary

Biodiversity data constitute an important source of knowledge for many disciplines. It is fundamental to informed and balanced policy making in areas such as conservation, development, and disaster response, and for understanding the potential impacts of climate change. To satisfy the increasing need for accessible and reliable biodiversity data, biodiversity science needs to modernise the production and dissemination of primary data, and knowledge and information products. Today's rapid growth in data volume is not matched by a commensurate increase in resources for curation and quality control (the available number of specialists: taxonomists). The pro-iBiosphere project was launched to investigate ways to increase the accessibility of biodiversity data, improve the efficiency of its curation and increase the userbase of biodiversity data consumers and applications. This two year project will specifically address (i) technical and semantic interoperability between different forms in which data are published and (ii) sustainability issues related to the maintenance and curation of biodiversity data and derived information and knowledge.

To meet these challenges and opportunities, pro-iBiosphere is preparing the ground for an Open Biodiversity Knowledge Management System (OBKMS), a network and e-infrastructure that will require implementation through a second project. An important aspect of the implementation of the envisaged OBKMS is the capability to make data from potentially hundreds of millions of pages of legacy literature accessible and to integrate this with contemporary publication processes. It also involves encouraging the biodiversity community to publish biodiversity data in a way that satisfies the need for accessibility from the start.

More specifically, the objectives of pro-iBiosphere are to:
- Coordinate towards and prepare the foundations for a viable long-term platform for knowledge management, and 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 forthcoming semantic mark-up of taxonomic literature, and link elements of biodiversity literature to the original data;
- Promote and monitor the development and adoption of common mark-up 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;
- Promote and increase cooperation between the major biodiversity projects, initiatives and platforms at EU and global levels.

During the first year of activities, the project organised a total of three meetings and six workshops with various stakeholders from around the globe (attendees from six continents were present) and a wide

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 5 of 85

spectrum of interested groups, e.g. traditional taxonomists, librarians, publishers, web designers, bioinformaticians. The meetings have been summarised below.

- Meeting 1 (27-28 September 2012):
    - Kick-off meeting

- Meeting 2 (11-15 February 2013)
    - Workshop on development and use of e-tools
    - Workshop on data acquisition and curation using e-tools for taxonomy
    - Workshop on semantic mark-up generation, data quality and user-participation infrastructure

- Meeting 3 (21-24 May 2013):
    - Workshop on stakeholder requirements
    - Workshop on measuring and constraining the costs of delivering services
    - Workshop on coordination & routes for cooperation across organisations, projects & e-infrastructures

In addition to the meetings, the project has:
- Conducted a series of pilots addressing technical and semantic interoperability challenges (including marking-up digitised text in preparation for databasing).
- Organised a training on the use of specific software for mark-up of digitised text (GoldenGATE, January 2013).
- Produced and disseminated various promotional materials consisting of: various posters, a fact sheet, four press releases, a social media postcard, and three eNewsletters.
- Sponsored & co-organised the BIH2013 conference that took place in September 2013 in Rome, Italy.
- Submitted a proposal, together with other infrastructure projects, to participate in the EC ICT2013 networking session and to staff a booth at the event. The proposal was accepted, and pro-iBiosphere is making preparations to participate in the event, to take place in November 2013, Vilnius, Lithuania.
- Published two articles.

The pro-iBiosphere project has contributed to a unified agreement by the biodiversity community on the use of identifiers for collections (this subject had been discussed for more than 10 years, without any agreement until now). Based on a process started by the Information Science and Technology Commission of the Consortium of European Taxonomic Facilities (CETAF) in 2012, the coordination workshops facilitated by pro-iBiosphere brought the discussions to a fruitful culmination. The technical foundation for integration was one the focus topics in three pro-iBiosphere workshops organised by WP3 in February 2013, WP2 in May 2013, and WP4 in October 2013. As a result of this joint pro-iBiosphere and CETAF effort, the adoption of semantic web and linked open data technology is increasing. At present (10/2013), the technology is being adopted by 7 institutions (i.e.: pro-iBiosphere institutions: Plazi, FUB-BGBM, MfN, RBGK, NBGB and two non pro-iBiosphere institutions: Royal Botanic Gardens Edinburgh - RBGE and Museum National d' Histoire Naturelle, Paris - MNHN). These institutions set a recognizable example for others to follow. In

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 6 of 85

order to make data interoperable, URIs are being assigned to specimen records, to key data items in treatments and to entire documents. A guide to best practices for stable URIs is being updated on the project wiki. When links based on these URIs are in place, a broad range of biodiversity information sources will be connected, including literature, organism names, classifications, occurrences, traits, sequences, and phylogenies.

In order to document possible business models for supporting an Open Biodiversity Knowledge Management System in a financially sustainable way, two reports on diversity and strengths of existing business models and discussion of sustainability have been published. These, along with the future deliverables documenting relevant costs and benefits, will allow an evaluation of alternative business models and recommendations for options to strengthen Open Science e-infrastructure. A number of issues have been addressed in these deliverables to date, namely:

- A methodology for gathering information from partners in a consistent way to allow analysis of exploitation plans and business models.
- A background document (combining input by partners and desktop research), providing the broader context, within which the Open Biodiversity Knowledge Management System is expected to exist.
- The Partners' vision of the project and their expectations.

In order to ascertain the costs of supplying services, a workshop in May 2013 was organised. The workshop brought together an international audience of both system providers and user communities to discuss the measurement of the cost of providing taxonomic content and the constraints placed on the delivery of this content by these of costs.

The project has produced a series of recommendations to make biodiversity data and information (now only accessible in unstructured form in the literature) digital, open, and linked. Recommendations are available in the following reports:

- "Ongoing biodiversity related projects, current e-infrastructures and standards" (D2.1.1). This report identifies the gaps, challenges and recommendations and summarises these in a white paper.
- "Draft strategy for increased cooperation" (D.2.1.2). This report includes an array of recommendations, which if adopted, will promote an Open Biodiversity Knowledge Management System.
- "Proof of concept reports on the use of e-tools" (D2.3). This report provides recommendations to improve the use of digital infrastructure among researchers in taxonomy and related areas.
- "Report on User Feedback" (D.2.2). This summarizes the major constraints preventing the optimal use of biodiversity information contained in classical Floras and Faunas ("Biotas").
- "Draft policy on Open Access for data and information (1)" (2.4.1). This report includes recommendations for promoting free and open access to and free re-use of biodiversity data and information at the level of institutions, the EU, and its member states.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 7 of 85

- "Towards a Best Practices e-Guide on Editorial Policies" (D3.1). This report includes an overview of (i) available systems and their respective properties and strengths, and (ii) standardized editorial policies that are needed for the curation and publication of biodiversity data in an e-environment.
- "The state and quality of biosystematics documents and survey reports" (D3.2.2). This report addresses how to improve semantic interoperability of biodiversity data in general, and especially their integration into the emerging Semantic Web. It also provides a set of practical recommendations and guidelines for achieving improved semantic interoperability of marked-up documents.

As a result of the interoperability pilots that are being conducted, the following results have been achieved:
- Specifications of minimal requirements defining the interface between different systems for storage and curation of biodiversity data have been produced.
- Transformation software allowing imports of mark-up documents into the EDIT-CDM platform for storage and curation has been developed.
- Key interoperability topics have been identified (and will be addressed in a workshop on interoperability (due in October 2013)).
- Active collaborations with Consortium of European Taxonomic (CETAF) and Biodiversity Virtual e-Laboratory (BioVeL) on these key topics have been established.
- An improved GoldenGATE version that allows marking-up over 1000 treatments has been produced.
- A best practice for stable URIs has been created.

Expected final results of the project are:
- A whitepaper describing an optimal dataflow, and identifying the elements where current data flows fall short of this.
- A work plan and roadmap for the semantic integration of biodiversity literature.
- A report on the state and quality of biosystematic documents and survey reports.
- Strategies for improved cooperation and interoperability between infrastructures.
- A strategy for the improvement and increased interoperability of XML schemas.
- A set of business requirements and scenarios for a sustainable Open Biodiversity Knowledge Management System as an alternative to current business models, and recommendations for achieving sustainable delivery of core biodiversity data and information.
- A draft policy on Open Access for data and information.
- A draft strategy for increased cooperation.

Potential impact of the project consists of, increased:
- Cooperation among projects, organisations and individuals;
- Trust among players;
- Exchange of content within and beyond the biodiversity domain;
- Possibilities for sharing costs involved in maintaining accessible biodiversity information.

Address of project website: http://www.pro-ibiosphere.eu/

Address of project wiki: http://wiki.pro-ibiosphere.eu/wiki/Main_Page

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 8 of 85

3.2 Core of the report for the period

Project Partners

Partner number	Partner short name	Partner name and country
1	**Naturalis**	Naturalis Biodiversity Center, Netherlands
2	**NBGB**	Nationale Plantentuin van België, Belgium
3	**FUB-BGBM**	Freie Universität Berlin, Germany
4	**Pensoft**	Pensoft Publishers Ltd, Bulgaria
5	**Sigma**	Sigma Orionis, France
6	**RBGK**	The Royal Botanic Gardens, Kew, United Kingdom
7	**Plazi**	Plazi, Switzerland
8	**MFN**	Museum für Naturkunde Berlin, Germany

3.2.1 Project objectives for the period

Workpackage 1 Management and coordination

Overview of project objectives

This Workpackage provides the financial, consortium and project management. The proposal is submitted and co-ordinated by Naturalis. This Workpackage will carry out:

- Result-oriented, efficient management of the project;
- Effective internal communications;
- Quality control of results and deliverables;
- Transparent financial management and control; and,
- Timely communication with the European Commission.

Workpackage 2 European and international policy coordination

Overview of project objectives

WP2 will lay the foundation for the overall concept of the Open Knowledge Management System that this coordination aims at. It will liaise between all consortium members and collaborating external partners.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 9 of 85

These include the key biodiversity infrastructure developers, knowledge creators, science publishers and information users.

The objectives of WP2 are, to:

- Consolidate and update the knowledge on the status of biodiversity informatics resources for taxonomic literature and to report on their current user base;
- Identify potential users of a new e-taxonomy infrastructure;
- Identify potential barriers for users presently expressing reluctance to use an open and shared e-taxonomy;
- Collect and analyze the requirements of taxonomic data providers and users;
- Learn about impediments of the existing infrastructure and workflow elements by providing exemplar training with focus groups;
- Identify common goals and reciprocal synergies that may induce a wide spectrum of organisations to collaborate and collaboratively develop policies that enable such collaboration.
- Communicate user feedback to technical coordination Workpackages and existing infrastructures;
- Mobilize international expertise;
- Identify gaps in the current e-taxonomy infrastructure, reporting on future requirements; and,
- Review the legal aspects of access to and sharing of taxonomy information towards an Open Knowledge Management System.

The planned activities will enhance the understanding about (i) the available and necessary elements of a digital infrastructure to curate and synthesize biodiversity information; and (ii) how the stakeholder cooperate, generate and exchange data.

Workpackage 3 Scientific content and workflow coordination

Overview of project objectives

The purpose of WP3 is to contribute to the establishment of new and standardized editorial policies for the curation and publication of biodiversity data in an e-environment, including editorial policies, Intellectual Property Rights (IPR) management, and involvement of citizen scientists in data enhancement, use and re-use. In addition, it will coordinate the efforts to provide access to the legacy literature by semantic tagging and linking the atomized content; for instance, to external sources like name servers (IPNI for plants, ZooBank for animals, Index Fungorum and MycoBank for fungi), observation data, ontologies, images of herbarium and other materials). Mark-up will be used for direct interpretation by machines on the web, or to facilitate import into databases.

The objectives of WP3 are, to:

- Review and analyze the existing methods for data acquisition, curation and semantic integration.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 10 of 85

- Review current methods for semantic mark-up and their use.
- Facilitate crowdsourcing/auth-sourcing studies with data curators.
- Align ongoing and forthcoming efforts to semantic mark-up of biodiversity literature and provide technical and social solutions for their use.
- Attract potential new users to biodiversity informatics infrastructures by emphasizing potential improvements in current workflows and collect user feedback during these events.

Workpackage 4 Technical and infrastructure coordination

Overview of project objectives

The purpose of WP4 is to facilitate joint studies and actions that will promote and enhance cooperation for a future technical and semantic interoperability between biodiversity platforms, including the promotion, monitoring the development and adoption of common mark-up standards. WP4 will exchange information and reach agreements that will facilitate the exchange of biodiversity-specific information (localities, treatments, keys, names, and references).

The objectives of WP4 are, to:

- Identify and develop strategies for increased cooperation among current and previously funded EU biodiversity related projects, and with non-EU projects;
- Identify and develop strategies to improve current taxonomic treatment servers and increase cooperation and coordination between the current existing electronic registers for species information;
- Identify current needs and promote future interoperability at technical and semantic level, through improvement and interoperability of the existing mark-up XML schemas and standards.

Workpackage 5 Dissemination, communication and public awareness

Overview of project objectives

The purpose of WP5 is to actively disseminate the results through various outreach activities, create stakeholder engagement and intense communication towards/with them; which will lead to increased public awareness.

Each Pro-iBiosphere Workpackage will target specific audiences, such as: coordinators of projects and e-infrastructures, and directors of natural history institutions (WP2 - European and international policy coordination); staff in charge of gathering data, such as scientists and/or citizen scientists, and technical staff responsible for the XML schemas (WP3 - Scientific content and workflow coordination); initiatives, projects and platforms at the infrastructure level (EU and non EU), IT experts, data providers (WP4 - Technical and infrastructure coordination); and institutions providing services within the taxonomic

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 11 of 85

community (WP6 - Sustainability planning). Through the various activities that are planned in WP5 (Dissemination, communication and public awareness), pro-iBiosphere strive to reach all those stakeholders that are interested in the project but that were not able to join the meetings and workshops.

The activities to be developed through this Workpackage are:
- Development of the project image, documentation and web platform
- Dissemination of the project results through outreach activities
- Stakeholder engagement and communication

Workpackage 6 Sustainability planning

Overview of project objectives

The purpose of WP6 is to evaluate possible business models for supporting Open Science in view of achieving financial sustainability. A sustainability plan will take into account the needs of the pro-iBiosphere community and best practices for other domains. WP6 activities will secure that users are able to make use of a European Open Biodiversity Knowledge system, addressing acquisition, curation, synthesis, integration, interoperability and dissemination of biodiversity knowledge. This will allow the pro-iBiosphere community to create new information (Floras and Faunas) in the future and to guarantee that the partners and users have a well-documented and standardized access to the services that will be offered (e.g. increased access to legacy documents). The main outcomes of WP6 will be i) an evaluation of alternative business plans and recommendations for options to strengthen Open Science e-infrastructures (M23), and ii) recommendations to policy makers with regard to achieving sustainable delivery of biodiversity information to target audiences (M24).

The objectives of WP6 are, to:

- Ascertain the costs of supplying services;
- Ascertain the benefits of these services for a) users or clients and b) suppliers;
- Identify business models in use within and beyond our community; and,
- Evaluate these models and do a SWOT analysis, arriving at i) conclusions about feasibility and ii) recommendations as to model (or models) to follow.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 12 of 85

3.2.2 Work progress and achievements during the period

Workpackage 1 Management and coordination

Summary of progress towards objectives

Partner number	Partner short name	WP 1 Person-months (PMs) per partner[2]	Effort (PMs Year 1)
1	Naturalis	6.00	5.95
2	NBGB	0.50	1.05
3	FUB-BGBM	0.50	0.57
4	Pensoft	1.50	1.18
5	SIGMA	1.50	0.90
6	RBGK	0.50	0.40
7	Plazi	3.50	1.60
8	MFN	0.50	0.125

The tasks of Workpackage 1 have been implemented successfully. The summary of progress towards objectives is described below.

- **Objective 1: Result-oriented, efficient management of the project.** The project is on track and all the activities that were envisaged for the first year have been accomplished.

- **Objective 2: Effective project internal communications.** Internal communication has been maintained among partners through email correspondence, face-to-face meetings, the pro-iBiosphere website and wiki.

- **Objective 3: Quality control of results and deliverables.** In order to provide high quality reports, a peer review mechanism, involving all partners and external reviewers, was created at the beginning of the project. Furthermore, various online tools are being used by the consortium. These tools are: Google Drive (used for writing all reports), pro-iBiosphere wiki (used for sharing information on meetings, articles, best practices, a.o.), Perfect It (editing tool that helps to check for inconsistency and errors).

- **Objective 4: Transparent financial management and control.** Four quarterly Management Progress Reports were prepared and submitted to the EU. These reports describe the major achievements and difficulties per task, the work performed per each partner, and provide an indication of the resources spent and justifications. This activity has allowed transparency of financial management and provided the coordinating institution with an updated overview of the finances of the project. The management reports are being written with the aid various online tools: EU-xpert (for management reports, EU-Fin (for financial reports), and Google drive.

---

[2] Based on Annex 1

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 13 of 85

- **Objective 5: Timely communication with the European Commission.** Online communication with the EC was maintained on a regular basis. Whenever possible, face-to-face communication has also been used.

Details for each task

Task 1.1 Administrative and financial management (lead: Naturalis. Start: M1, End: M24)

Project Management and Quality Assessment Plan. The pro-iBiosphere "Project Management and Quality Assessment Plan" for pro-iBiosphere was prepared and distributed among the consortium at the start of the project (see Ref. 111) The report contains templates, guidelines and requirements for the partners to contribute to the project and complements the procedures explained in Annex I – "Description of Work" and the Consortium Agreement documents.

Task 1.2 Quality management, assessment and reporting (lead: Naturalis. Start: M1, End: M24)

Reporting tools. The EU-Xpert (management, see Ref. 33) and EU-Fin (financial, see Ref. 34) tools are being used by the consortium to elaborate the quarterly progress reports and annual reports.

Online progress meetings. In order to discuss progress of the pilot activities and workshops, various online meetings have been organised within the reporting period. The minutes of all meetings organised are available for consortium members on the internal library of the pro-iBiosphere website.

Internal Quality. In order to guarantee the high quality of contents of deliverables and reports, the consortium actively participates in the review of all reports, and external reviewers are involved too. In order to facilitate the peer review of deliverables, a table with institutions and deadlines is available on the pro-iBiosphere wiki (see Ref. 21).

Deliverables. All deliverables were produced according to the original planning. Deliverables can be downloaded from the pro-iBiosphere wiki (see Ref. 21).

Milestones. All planned milestones were achieved according to the original planning (see Ref. 35). The only exception is MS20 "Meetings with clients/users- benefits", which was postponed to Y2 (for explanation on this, see WP6 subheading on "Reasons for deviations").

Evaluation of events organised by pro-iBiosphere events. In order to evaluate the workshops organised by pro-iBiosphere and assess satisfaction by participants, an online questionnaire was designed in Google Drive (see Ref. 36). The results of the questionnaire have been taken into account for follow-up meetings organised by the project.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 14 of 85

Task 1.3 Internal Communication (lead: Plazi; participants: PENSOFT. Start: M1, End: M24)

Internal Communication. To ensure that communication and interaction among consortium members and Workpackages are reliable, efficient, and transparent, a protected web platform (see Ref. 30) and a wiki (see Ref. 31) were developed at the start of the project. These platforms facilitate the secure exchange of project information, results, and various documents such as deliverables, agendas, meeting minutes, etc.. All project members were registered in the Internal Communication Platform (ICP) of pro-iBiosphere. Each project member was given a unique account that grants him/her access to the internal documentation of the project.

The ICP is an integral part of the project's website and provides a medium for communication among project participants. It serves for exchange of various types of information: data-sets, results, coordination decisions, timetables, presentations, materials, and for reporting among partners. The ICP was developed by Pensoft's IT team, in close collaboration with Naturalis. The ICP has the following main features:

i) **Mailing module –** project members can send emails to one or more project participants after logging into the system. Users are assigned to one or more mailing groups depending on their role in the project. Collective emails can be sent to one or more mailing groups and individual users. All emails are properly archived and can be sorted by date, sender, etc., which helps the quick discovery of older correspondences.

ii) **Internal Document Library -** all documents deriving from the ongoing activities are stored in the Internal Document Library. All consortium members are given access and can upload files there. The IDL is the place to sort out, index and host all files related to the project activities, from the branding products and templates to presentations from meetings and reports. Documents are arranged in folders depending on their content and aim. Publications and other pro-iBiosphere information (deliverables) that are open for access/download to the external users are being made public at the "results page" of the website. The IDL serves as a main repository for pro-iBiosphere documents. At present, it contains ca. 118 documents.

iii) **Internal events** – this feature of the ICP provides an opportunity for monitoring of the project tasks, which facilitates the overall project management. By sending regular reminders of approaching tasks, this tool helps consortium members to follow the time frames and submit their reports on time.

 The ICP allows the coordinator (and partners) to regularly monitor progress in data collation, analysis, and accomplished deliverables. The ICP is also used as an internal discussion forum for items emerging within Workpackages that need rapid decisions.

The ICP and IDL can be accessed by the consortium by logging into the website of pro-iBiosphere. The Internal Calendar feature of the website is also regularly used as a task manager, facilitating the overall project implementation, and it is also being by Naturalis and Sigma to follow up on deadlines and meetings and send automated reminders to the consortium.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 15 of 85

The **pro-iBiosphere Wiki** (see Ref. 31) was created by Plazi primarily for dissemination of the project outputs (e.g. information on meetings/workshops, including purpose of meetings, expected results, agendas and list of participants, pilot information, training activities). The Wiki has proven to be an excellent communications platform. It is being used by the consortium for exchanging ideas about the organisation of the workshops, preparing agendas, drafting reports and sharing the results achieved by the project with the stakeholders of the project.

Task 1.4 Consortium and review meetings (lead: SIGMA; participants: all partners. Start: M1, End: M24)

Project meetings. During the first reporting period, the following consortium meetings were organised:

- Kick-off meeting, on September 27-28 2012, in Leiden, the Netherlands (see Ref. 1).
- Second consortium meeting, on February 15 2013, in Leiden, the Netherlands (see Ref. 2).
- First advisory board and management committee meetings, on February 15 2013, in Leiden, the Netherlands (see Ref. 37).
- Third consortium meeting, on May 24 2013, in Berlin, Germany (see Ref. 38).

Workshop agendas, lists of participants, and information on the venue were shared with stakeholders on the wiki. Meeting minutes were prepared by the task leader in collaboration with the project coordinator and project partners, and made available on the project wiki.

During the meetings, partners reviewed the latest results and achievements of the different Workpackages, discussed the outcomes of the project workshops, highlighted topics of particular interest such as the progress of the pilots, the dissemination of project results and drafted an action plan to be performed in the next months prior to the next consortium meeting. A list of the planned meetings with the consortium bodies is available on the wiki (see Ref. 39).

Meeting template documents. Two specific templates for the preparation of meetings (i.e. agenda and minutes) were shared with partners in M1, and updated in M12. The templates are available on the pro-iBiosphere website (see Ref. 40).

Significant results

Significant results include:

- Two online communication platforms were set up to allow and encourage better collaboration among partners (and stakeholders).
- All deliverables were timely submitted
- All milestones were achieved.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 16 of 85

Reasons for deviations

Not applicable. There were no deviations from Annex I.

Reasons for failing to achieve critical objectives

Not applicable. All critical objectives were achieved.

Statement on the use of resources

The following budget was reallocated in Y1:

Plazi (person months). In September 2012, it was agreed to re-allocate four person months from Plazi to Naturalis. The transfer was made because only two person months were required to implement and maintain the wiki (T1.3. Internal Communication and T5.1. Development of the project image, documentation and external communication and web platform), whereas a total of six person months had been allocated. The four person months were used to hire a part-time project assistant by Naturalis for the project. The hiring took place in May 2013.

Plazi (money left from May meeting). Budget was allocated to Plazi for the organisation of the pro-iBiosphere May meeting (consisting of 3 workshops organised by Naturalis, RBGK and Plazi) that took place in May 2013, in Berlin Germany. A total of 4.771 euros were left from the pro-iBiosphere meeting. The amount left will be used for hiring a part time project quality assistant (desk editor) to help reviewing the quality of all pro-iBiosphere documents that will be produced in Y2 (i.e. 23 deliverables, articles, posters, factsheets, press releases, etc.).

Corrective actions taken

Not applicable.

pro-iBiosphere FP7 Project ∎ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ∎7th Framework Programme
∎ Coordination and support action FP7-INFRASTRUCTURES-2012-1∎ Subprogram area INFRA-2012-3.3

Page 17 of 85

Workpackage 2 European and international policy coordination

Summary of progress towards objectives

Partner number	Partner short name	WP 2 Person-months (PMs) per partner[3]	Effort (PMs Year 1)
1	Naturalis	2.25	1.00
2	NBGB	5.25	4.68
3	FUB-BGBM	0.75	0.29
4	Pensoft	2.75	2.20
6	RBGK	3.75	3.15
7	Plazi	4.50	3.00
8	MFN	0.75	0

The tasks of Workpackage 2 have been implemented successfully. The summary of progress towards objectives is described below.

- **Objective 1:** Consolidate and update the knowledge on the status of biodiversity informatics resources for taxonomic literature and report on their current user base. The report on "ongoing biodiversity related projects, current e-infrastructures and standards" (see D2.1.1 Ref. 27) documents the present activities, strategies, gaps, goals, use cases, interests and visions as well as cooperation and interrelations of the various European and international partners interested in participating in an Open Biodiversity Knowledge Management System (i.e., a taxon treatment-like knowledge management system). The report focuses on the (i) existing digital infrastructures; (ii) past publications and curation systems, including regional or global monographs; and (iii) data elements constituting taxonomic treatments, such as specimen data, images, sequences, taxon treatments, taxon names and their concepts, morphological characters, ecological and biological traits. The report is a result of the workshop "Routes towards cooperation" organised on 23ᵗʰ of May 2013 in Berlin that was aimed at increasing our reciprocal understanding and fostering progress towards multi-institutional action that will improve cooperation.

- **Objective 2:** Identify potential users of a new e-taxonomy infrastructure. The primary role of alpha taxonomists is the collection and dissemination of information about biodiversity, particularly the recognition of and the distinction between different taxa. They are potentially important users of these tools, and new e-taxonomy infrastructure(s), both as creators and consumers of digital information. The reports on "The Use of e-Tools among Producers of Taxonomic Knowledge (Proof of concept report on the use of e-tools)" (D2.3, see Ref. 11) and "User Feedback" (D2.2., see Ref. 41) list additional potential users of a new e-taxonomy infrastructure.

- **Objective 3:** Identify potential barriers for users presently expressing reluctance to use an open and shared e-taxonomy. The report on "The Use of e-Tools among Producers of Taxonomic Knowledge (Proof of concept report on the use of e-tools)" (D2.3, see Ref. 11) identifies barriers to the adoption of software tools by taxonomists. Main barriers found are a lack of training, a shortage of application support, the time required to learn a new system, the lack of recognition from Internet publications and a lack of user oriented focus by taxonomists. The report is based on a workshop

---

[3] Based on Annex 1

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 18 of 85

organised by the pro-iBiosphere project in February 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.

- **Objective 4:** Collect and analyze the requirements of taxonomic data providers and users. The workshop on the "Uses and Users of Fauna, Flora and Mycota Information" held in May 2013, in Berlin, Germany (see Ref. 4) provided information on how those working outside of the taxonomic disciplines currently use the information, and how they would like to be able to do so in future. The results of the workshop are available in the report on "User Feedback" (D2.2., see Ref. 41). According to the report, in addition to access to Biota information, there is a strong requirement from users for activities such as identification or validation of data.

- **Objective 5:** Learn about impediments of the existing infrastructure and workflow elements by providing exemplar training with focus groups. A training event on e-platforms and e-tools for taxonomy was organised on the 12th of February 2013, in Leiden, the Netherlands. The event helped to build interest in IT tools for taxonomy among participants and to learn about the impediments of the existing infrastructure and workflow elements. The report on "The Use of e-Tools among Producers of Taxonomic Knowledge (Proof of concept report on the use of e-tools)" (D2.3, see Ref. 11) summarises the major constraints to users of Biotas and the information they hold. According to the report, there is considerable variety of biodiversity informatics software, much of it created by the taxonomists themselves for individual projects. Support for international standards for data fields and data exchanges is 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 skepticism within the community regarding electronic publication and open access to the literature.

- **Objective 6:** Identify common goals and reciprocal synergies that may induce a wide spectrum of organisations to collaborate and collaboratively develop policies that enable such collaboration. The workshop on "Coordination and Routes for Cooperation" organised on May 24th 2013, in Berlin, Germany (see Ref. 6), helped us to (i) document the current position of data providers and discuss points for cooperation, and (ii) understand the needs of external communities and report on the necessary software interfaces for these users. Of special importance was to identify further potential routes for cooperation between European and non-European biodiversity projects and platforms. The report of the workshop (D2.1.2 due in September 2013) falls outside the scope of this report and hence progress on this objective will be documented in the Y2 report.

- **Objective 7:** Communicate user feedback to technical coordination Workpackages and existing infrastructures. The report on "The Use of e-Tools among Producers of Taxonomic Knowledge (Proof of concept report on the use of e-tools)" (D2.3, see Ref. 11) includes ten recommendations to improve the use of digital infrastructure among taxonomists. In order to provide feedback to other infrastructures, the report was broadly disseminated through a press release (see Ref. 46) and the pro-iBiosphere database of participants (which includes more than 200 people).

- **Objective 8:** Mobilize international expertise. Through the W2 workshops mentioned above and the various workshops organised in the other Workpackages, the project has been able to mobilise international expertise throughout year one. This objective will continue until the end of the project (2 meetings with stakeholders are envisaged in Y2). In particular, the workshop on "Coordination and Routes for Cooperation" organised in May 24th 2013, in Berlin, Germany (see Ref. 6), facilitated the continuation of discussions started at the workshops on legacy and prospective literature in Leiden on February 13th and 14th, with top international experts in the domains of taxonomic publishing and semantically enhanced literature. This led to a focus on a

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 19 of 85

small number of major issues to foster international cooperation, e.g. with respect to the implementation of identifiers and open access to data.

- **Objective 9:** Identify gaps in the current e-taxonomy infrastructure, reporting on future requirements. The report on "ongoing biodiversity related projects, current e-infrastructures and standards" (see D2.1.1 Ref. 27) documents the present activities, strategies, gaps, goals, use cases, interests and visions as well as cooperation and interrelations of the various European and international partners interested in participating in an Open Biodiversity Knowledge Management System (i.e., a taxon treatment-like knowledge management system). The most important gaps, challenges and recommendations are summarised in a white paper.

- **Objective 10:** Review the legal aspects of access to and sharing of taxonomy information towards an Open Knowledge Management System. Task 2.4 consists of a two-step deliverable. Deliverable D2.4.1 "Draft policy on Open Access for data and information (1) was submitted in September 2013. After targeted external review, D2.4.1 will then be developed to a final proposal of a draft policy that will be available by month 23. Both deliverables fall outside the scope of this report and hence progress on this objective will be documented in the Y2 report.

Details for each task

Task 2.1 Coordination and routes for cooperation across organisations, projects and e-infrastructures (lead: Plazi; participants: Naturalis, NBGB, FUB-BGBM, PENSOFT, RBGK. Start: M6, End: M13).

Task 2.1 serves as a baseline coordination platform for general high-level policy and strategy coordination. Activities of the task aimed at analysing the way how various groups (i.e. key biodiversity players) cooperate, generate and exchange data. In order to facilitate this analysis, a workshop on Coordination and Routes for Cooperation was organised in May 24ᵗʰ 2013, in Berlin, Germany (see Ref. 6). Specific objectives of the workshop were to:

(i) Document the current position of data providers and discuss points for cooperation in a memorandum of Understanding (MoU), which will facilitate the sharing of resources among biodiversity organisations, projects, and initiatives.

(ii) Understand the needs of external communities and report on the necessary software interfaces for these users. Of special importance was to identify further potential routes for cooperation between European and non-European biodiversity projects and platforms.

Before the workshop, a questionnaire was disseminated among all the participants that registered for the event, available as Ref. 42. The online survey targeted users and providers of biodiversity data. The survey was undertaken with the aim to assess the specifics of the available e-infrastructure and databases. The questionnaire consisted of 5 multiple choice questions and 12 open-ended questions. 60 completed survey questionnaires were received between March 5ᵗʰ 2013 and May 14ᵗʰ 2013. All presentations of the meeting are available online on the pro-iBiosphere wiki (see Ref. 6). A total of 45 participants were present during the meeting. Invited lectures were carefully selected to represent the potential collaborators, the topics and the interest of the users of the data, followed by discussions.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 20 of 85

Two reports were prepared. The first one, "Report on ongoing biodiversity related projects, current e-infrastructures and standards" (D2.1.1; Ref. 27), documents and updates the present activities, strategies, goals, use cases, interests and visions, and the cooperation and interrelations of the various European and international partners interested in participating in a taxon treatment-like knowledge management system (see Ref. 27). The second one, "Towards a draft strategy for increased cooperation" (D2.1.2; Ref. 29), updates and consolidates our knowledge of the project partners, identifying potential collaborators, users, and gaps in the infrastructure. The draft strategy (also known as the Open Biodiversity Knowledge Declaration) includes a set of recommendations on open access to content, identifiers to specimen and related data, as well as a registration of discovery mechanisms (see Ref. 29). The declaration was sent to one of the CETAF-ISTC members (in October 2013). The intention is that it will be signed by the consortium members and also others (i.e. CETAF members and other biodiversity initiatives) on the occasion of the final pro-iBiosphere event, and that some of the core elements will be implemented by the end of the project. As the task has ended in Month 13, a follow up of the recommended declaration will be addressed in the Year 2 activities of Task 4.1.

Advisory Board. At the start of the project, in conjunction with WP1, an Advisory Board (AB) was assembled (see Ref. 37). The present AB consists of 4 members representing data providers, IT infrastructure, publishers and users of taxonomic information. The minutes of the AB meeting held on the 15th of February 2013 in Leiden are available from Ref. 37 . The AB will help to develop recommendations for improvement of the data integration and interoperability in the three main directions:
(1) Improving coordination and management of biodiversity data and platforms through active discussion and identification of stakeholders' needs and development of strategies for reduction of duplicated efforts and associated costs;
(2) Improving the coordination between working groups that have participated or are currently participating in past and on-going EU projects;
(3) Analysing and developing of strategies for integration and interoperability in the field of bioinformatics between EU and USA-based global initiatives.

Task 2.2 Stakeholder requirements (lead: RBGK; participants: Naturalis, NBGB. Start: M2, End: M12).

Scope. In order to achieve sustainable services, it is vital that the producers of e-Floras and e-Faunas deliver the information that users want. Following discussion at the Kick-off meeting, it was agreed that T2.2 "Stakeholder Requirements" would therefore concentrate on aspects of information requirements, whereas stakeholder requirements for tools was realigned under T2.3.

Stakeholder Survey. The stakeholder survey is based on the results of a dedicated workshop, questionnaire and augmented with further desk based research. The list of attendees to the February 11th-14th 2013 pro-iBiosphere workshops was used as a basis for the workshop on "Uses and Users of Fauna, Flora and Mycota

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 21 of 85

Information" that was held in Berlin in May 2013 (see Ref. 4 and MS4 Draft Report available under Ref. 43). This original list was expanded by the consortium to include a much wider range of producers, providers and users of flora and fauna information (i.e. from a range of domains including conservation, ecology, taxonomy, facilitators, data aggregators, earth systems science). Invitations were sent out in the first week of March 2013, and about 250 potential participants were targeted by direct emailing. A very large positive response was received. This allowed representative use cases to be developed in the May workshop, with the possibility of 1:1 follow-up visits where necessary and where time would allow (MS20). The single largest stakeholder group was those involved in conservation assessments, particularly IUCN red list compilation, and covered a range of taxonomic domains. Although a large number of scientists working on Earth Systems, vegetation modelling and plant trait analysis were invited, relatively few responses were received, perhaps because the workshop subject seemed too peripheral to their main area of research. Nevertheless, there were 3 participants from this community represented at the workshop.

As several of these participants were UK based, it was proposed during the consortium meeting held in October 2013 that a separate UK workshop will be arranged at RBGK for these individuals by the end of December 2013. The 0.5 Person months that were left from Tasks 2.3 and 3.2 (which already ended in Year One) will be used by RBGK for organising the workshop.

In all, 53 people took part in the May workshop. Adopting a formal presentation followed by questions approach, the workshop comprised of a number of highly interactive, facilitated sessions, to try and maximise the information gathered from the participants - the agenda is available on Ref. 4. The aim of the workshop was to learn from users of Flora, Fauna and Mycota information (particularly those working outside of the taxonomic disciplines), how they currently work with the information, and how they would like to be able to do so in future. The two morning sessions were split into 3 groups in separate rooms, within which a total of 13 small groups were formed, each of which worked on their particular "use case" (listed below):

1. Making a IUCN Red list assessment (#1)
2. Making a IUCN Red list assessment (#2)
3. Plant trait database compilation
4. Linking ecophysiology to vegetation modelling
5. I want to describe a new species
6. How do I identify a plant
7. I want to prepare a quick and dirty flora account for a taxon
8. I want to publish and disseminate high quality taxonomy
9. I want to carry out a plant survey of a small national park for management's decision making
10. Producing a Digital Flora
11. Re-Publishing Floras, Fauna & Mycotas
12. Producing a Field Identification Tool
13. Ecological Niche Modelling (ENM) based on specimen records and observation from Floras

For each use case, the participants constructed an activity map. They then prioritised their activities according to those that they felt were the most time consuming, thus providing insights into what users

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 22 of 85

feel are the "pain factors" in their work. In the second part of the morning, the groups constructed information maps to show what information they used and where they obtained it from. The maps grouped sources by the information types; they next ranked these information types in order of importance (value) to their work. In the afternoon, the discussion was broadened out in plenary to take views from the audience and a workshop questionnaire was distributed to all participants. The results of the workshop were the documentation of the use-cases through transcripts of recordings of the presentations and photographs and diagrams of the workflows, as well as the completed flipchart papers.

Following the workshop, the information obtained from these exercises was consolidated, and where necessary the use cases further developed in collaboration with the participants, mainly by email.

Task 2.3 During the past decade, state-of-the-art tools to facilitate acquisition of core biodiversity data have been developed

Digital tools are increasingly being used for data acquisition, curation and distribution of taxonomic information. Some of these tools are mature products, while others are more experimental. Task 2.3 aims to assess the state-of-the-art of these tools and their adoption by the taxonomic community. It also covers identifying the advantages of using these tools for the taxonomic community.

A workshop on "e-platforms & e-tools for taxonomy" took place on February 11ᵗʰ and 12ᵗʰ 2013, in Leiden (for agenda, please see Ref. 3). The first day was devoted to presentations from the developers of taxonomic tools and to people actively using the tools in their work. The second day was dedicated to giving hands-on training of a sample of the available tools. Four parallel training rooms in three sessions were organised, and hence, allowed the trainees to experience as many tools as possible in only a short time. Verbal feedback from the participants was unanimously positive; several participants that adopted these tools as a result of the training wrote contributions to the second pro-iBiosphere newsletter. Participants included a wide range of taxonomists who work in European and non-Europeans institutions on plant, animal and fungal groups; and members of all the leading e-taxonomy initiatives and global projects in biodiversity.

The organisation of the second pro-iBiosphere meeting (including this and other two workshops coordinated by Plazi and Naturalis) took a significant proportion of the time related to this task. Activities included inviting speakers and managing the agenda; inviting representatives from Africa and Asia and organising their travel arrangements, documents for visa applications and accommodation; organising venue arrangements and the timetable for trainers; and answering questions of attendees. The main outcomes of this workshop are available in the "pro-iBiosphere meeting number 2" document. (see Ref. 2).

As part of Tasks 2.3 & 3.1 activities, the NBGB, RBGK and Naturalis developed a questionnaire on the use of digital taxonomic tools. This poll aimed to reveal the opinion on digital taxonomy of taxonomists and related research staff. It covered the core focus of Task 2.3 but also related subjects. For example, there

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 23 of 85

were basic questions on their knowledge of and use of digital tools, but also on their attitude towards intellectual property rights; data standards and their collaborations. We collected about 220 responses; about two thirds from pure taxonomists and editors of taxonomic publications, while the others are mainly ecologists, conservationists and IT developers. The respondents are largely European, but we do have contributions from 23 different countries, including several from Asia and Africa. The questionnaire was distributed among the consortium institution staff. The questionnaire can be seen on Ref. 44. The results of the questionnaire were used in the report on Task 2.3, and provided information useful for many other Workpackages.

A matter arising from the February workshops was how to progress the cause of semantic authorship and publication. Various ideas were suggested and discussed. Together with one of the participants, David Shotton (see Ref. 45 on Semantic Publishing), we created a short questionnaire to poll people on which are the simplest steps that could be taken to achieve this aim. The questionnaire was distributed among 2 mailing lists (i.e. Taxacom and pro-iBiosphere). This questionnaire received 45 responses. The low rate of participation is expected due to the highly technical nature of the questions.

In collaboration with WP4, we have been examining the availability and connectivity of digital data for taxonomy. The pilot on *Chenopodium* has helped highlight some of the advantages and limitations of current technology and workflows. The report 2.3, with supporting data gathered from a number of different sources, brings together a wide variety of technology specific reports to summarise the state-of-art in digital tools for taxonomists.

We also conducted another data gathering exercise on the international collaboration of taxonomists. Taxonomists frequently collaborate with each other. Digital tools for taxonomy potentially provide advantages for taxonomists in networking, data generation, writing and editing. We investigated the nature of international taxonomic collaboration using joint authorship information from a wide range of journals. International collaboration is thought to be one of the areas most helped by the adoption of digital technology, and an understanding of these networks of collaborating taxonomists helps us design systems that facilitate online collaboration.

The report for deliverable of D2.3 "Proof of concept reports on the use of e-tools" was submitted in August 2013. This involved collating the information gathered since the workshop in Leiden that took place in February 2013. This information was further supplemented by (i) researching the systems and the uses made of information technology by taxonomists (botanists and zoologists); (ii) conducting interviews and discussions with taxonomists on their use of information technology; and (iii) on their attitude and awareness of software systems that have been built for them in the past years. The report was made available on the wiki on 1ˢᵗ September 2013 (see Ref. 11). To ensure successful dissemination of the report, a press release was published, and links to the report were posted on scientific social media sites (to view the press release on the website, see Ref. 46).

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 24 of 85

To give a broad view of IT within the taxonomic community, the report investigated a wide variety of aspects, including:

1. The attitude of taxonomists towards Open Access; international collaboration within taxonomy;
2. The advantages and disadvantages of digitisation to different groups within the taxonomic community; and
3. The barriers to adoption of digital technology.

Although there is a widespread international collaboration among taxonomists, their use and knowledge of IT tools is poor. In addition, they do not use international standards, except in the area of nomenclature. We also reported on two very successful examples of the use of IT in biodiversity study: i.e. FishBase and Mapmate. These systems have little in common, but the features that make them successful are that they:

1. provide direct benefits to the users
2. are simple to use and
3. have a large and well supported user base.

Task 2.4 Legal issues of data acquisition, curation and dissemination (lead: Plazi, participants: all partners. Start M9, End: M23).

Initial contacts with other EU-initiatives covering the topic of related legal issues were made during a number of meetings:

- Global Biodiversity Informatics Conference in Copenhagen (July, 2012)
- OpenAIRE conference in Göttingen (November 21-22, 2012)
- EU-BON meeting (February 11-14 2013, in Berlin)
- Names Attribution, Rights and Licenses" organised by the Global Names Project at the Arizona State University (April 15-17 2013).

Task 2.4 consists of a two-step deliverable. Deliverable D2.4.1 "Draft policy on Open Access for data and information (1). This provisional draft policy includes

- Legal aspects of access to and re-use of data.
- Issues related to building and maintaining a knowledge management system and their legal ramifications.
- Findings of the workshop by the Global Names Project (see Ref. 28) on copyright issues related to names of organisms and compilations of names.
- Aspects of copyright and database right in the EU.

The draft also includes examples of the transformation of this legal framework in several member states of the EU and the European Economic Area (EEA).

Part of the document is based on answers to a questionnaire that was distributed among legal experts in different countries. The questionnaire was introduced at the pro-iBiosphere workshop "Coordination and cooperation" in Berlin, May 23, including a request to support finding the proper legal experts in the EU countries.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme ■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 25 of 85

In order to be able to compare the various national and EU legislations, the draft was circulated among specialists in representative EU countries (see press release published on the website and disseminated on the project social media platforms (Ref. 47). After targeted external review, D2.4.1 will then be developed into a final proposal for a draft policy. The final report (Deliverable 2.4.2; see Ref. 21) will be available by month 23.

Significant results

Workshops. The workshops organised in February and May 2013 were attended by participants from various EU and non-EU countries, representing the world's leading bio-informatics platforms (see Ref. 75).

Ongoing biodiversity related projects, current e-infrastructures and standards - Gaps, challenges and recommendations
Deliverable D2.1.1 on Ref. 27 " Report on ongoing biodiversity related projects, current e-infrastructures and standards" includes gaps, challenges and recommendations that are summarised on the white paper. The white paper includes the following topics:

1. Transition to full and gold open access publishing models for biodiversity information;
2. Policies towards open data publishing, sharing and re-use;
3. Increased cooperation between EU and non-EU bioinformatics initiatives;
4. Support digitisation, markup, data mining and re-publishing of legacy literature in advanced open access, semantically enhanced mode;
5. Maximum interoperability and integration between traditionally produced printed works (so called "legacy literature") as in Biotas (e.g., Floras, Faunas, Mycotas, Protozoas) and newly produced data;
6. Principle of machine readability: pressure to adopt advanced open access publishing models that allow automated data harvesting and re-use;
7. Wide adoption of universal persistent and resolvable identifiers for biodiversity informatics elements;
8. Further development and implementation of standards for data and associated metadata and data management infrastructures;
9. Linked Data and associated tools and infrastructure (e.g., ontologies, vocabularies, registries) to become a predominant model of data management in the biodiversity domain;
10. Promoting the use of existing directories for registration of tools and services derived from EU-funded projects.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 26 of 85

Open Biodiversity Knowledge Declaration

Deliverable 2.1.2 "Towards a draft strategy for increased cooperation" on Ref. 29 was submitted in September 2013. The report includes an array of recommendations that, if implemented, will promote Open Biodiversity Knowledge Management.

The recommendations include:

1. That pro-iBiosphere implements the recommendation made in report D2.1.1 (Report on ongoing biodiversity related projects, current e-infrastructures and standards) 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. Once the declaration is made openly available on the web, it will offer a mechanism of voluntary participation of those interested in biodiversity data. The declaration will be signed by the members of the pro-iBiosphere consortium at the final meeting of the present project at the Bouchout castle at the National Botanic Garden of Belgium, 9-13th June 2014.

2. That pro-iBiosphere promotes the free and open use of content, services and other resources by adopting, where possible, licenses that grant all users - including automated tools - a free, irrevocable, world-wide, 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.

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.

4. To register content and services, and to explore the option of adopting existing facilities for this purpose, such as the BioVel Biodiversity Catalogue (https://www.biodiversitycatalogue.org/).

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.

6. Establish agreements on specialization in services (example: one institution specializes in geographical analysis, another in visualization tools), to facilitate providing services to other institutions or projects.

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.

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).

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 27 of 85

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.

10. OBKM working parties to work together to identify funding, using existing and new sources, to implement the OBKMS.

Recommendations on how to improve the use of digital infrastructure among taxonomists

Ten recommendations to improve the use of digital infrastructure among taxonomists have been compiled in Deliverable D2.3 on "Proof of concept reports on the use of e-tools" (see Ref. 11). These recommendations include:

1. Focus on usability and interoperability of software, not just functionality.
2. Promote information technology within the taxonomic community.
3. Provide clear direct benefits in software for taxonomists, as well as downstream users.
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.
5. Tackle the social obstacles to IT adoption, such as the disconnect between taxonomists and the users of taxonomy.
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.
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.
8. Mandate institutional data archival policies.
9. Research data has to be open access.
10. Ensure firm, long term commitment of institutions to digital taxonomic infrastructure.

Constraints identified to users of Biotas and the information they hold

The major constraints identified by the pro-iBiosphere project to users of Biotas and the information they hold were:

1. The time needed for information to be retrieved from a Biota and presented in a more appropriate format for a particular reuse, and locating the relevant source work in the first place;
2. The lack of easy-to-use technical solutions for mining or presenting data currently in Biotas for reuse in other products or activities.
3. Extraction of atomised information by markup is time consuming, technically difficult and potentially very costly. However, such data underpins analysis of information and may facilitate further synthesis, provided that data standardisation is adopted.
4. Interfaces to data marked-up and atomised from Biotas need to be easy to use and follow data standards where appropriate.
5. Limited access to information due to:
6. restrictive access conditions
7. difficulties in interpreting the data
8. inefficient discovery mechanisms
9. Access to information varies, depending on where you live and work, the resources at your disposal and on your available finance. IPR restrictions are only one of the factors limiting access.
10. There is a strong requirement from users for human expertise in addition to access to Biota information for activities such as identification or validation of data.
11. People (and their expertise) are a valuable resource but as there is no standard index to them, significant time is spent locating them. User feedback channels are poor, and this hampers

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■ 7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1 ■ Subprogram area INFRA-2012-3.3

Page 28 of 85

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

12. Users of Biotas have requirements for information traditionally not published in Biotas, but closely related and potentially available to Biota producers. Significant data gaps (such as species abundance) and monitoring could be resolved and undertaken by local people, provided they have easy access to existing information and means to add that information to the online Biota.

13. Identification of species is a major activity both in the field and in the lab. The kinds of tools needed to help with this vary from low-tech to high-tech, depending on the particular situation. Where appropriate, these tools should be designed to integrate with local expertise.

Recommendations for promoting free and open access to and a free re-use of biodiversity data and information at the level of institutions, the EU, and its member states

Deliverable D2.4.1 (see Ref. 28) entitled "Draft policy on Open Access for data and information (1)" includes the following recommendations for promoting free and open access to and a free re-use of biodiversity data and information at the level of institutions, the EU, and its member states. These recommendations include:

A. On the level of institutions

1. Publicly funded institutions should refrain from claiming intellectual property rights for biodiversity data and information collected and/or published by them. All content should be openly accessible without any form of further authorisation required.
2. Re-utilization of biodiversity data and information for research purposes should be allowed without any form of authorisation. The only legitimate claims at least from publicly funded institutions with regard to the reutilization of material collected and/or published by them refer to the attribution of the source.
3. As far as publicly funded institutions' own material - which is protected by copyright or by database rights - are concerned, they should dedicate these works or these databases to the public domain by publishing them under a CC0-Public Domain Dedication (see Ref. 112)

B. On the level of the European Union (EU) and the European Economic Area (EEA)

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

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

1. 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. It should neither refer to the place where, nor the technical mode how, works are accessed, as such technical restrictions hamper the research process.
2. Member states of the EU or the EEA should introduce or, where it already exists, extend an exception of database protection for the use of databases for scientific research.

pro-iBiosphere FP7 Project ∎ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ∎7th Framework Programme
∎ Coordination and support action FP7-INFRASTRUCTURES-2012-1∎ Subprogram area INFRA-2012-3.3

Page 29 of 85

Reasons for deviations

Not applicable. There are no deviations on this Workpackage.

Reasons for failing to achieve critical objectives

Not applicable. All critical objectives have been achieved.

Statement on the use of resources

The RBGK has a total of 0.5 Person months left from Tasks 2.3 and 3.2. Since the above mentioned tasks have already ended in Year One, the 0.5 person months will be used by RBGK for organising a UK workshop on `Stakeholders' Requirements´ (in the framework of Task 2.2 activities).

Several participants interested in joining the May 2013 stakeholders workshop were UK based. It was proposed during the consortium meeting held in May that a separate UK workshop will be arranged at RBGK for these individuals. During the consortium meeting held in October 2013, it was decided by RBGK that the workshop will take place by the end of December 2013.

Corrective actions taken

Not applicable.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 30 of 85

Workpackage 3 Scientific content and workflow coordination

Summary of progress towards objectives

Partner number	Partner short name	WP 3 Person-months (PMs)per partner[4]	Effort (PMs Year 1)
1	Naturalis	5.00	2.99
2	NBGB	0.50	0.50
3	FUB-BGBM	11.00	6.50
4	Pensoft	3.00	1.23
6	RBGK	1.00	0.80
7	Plazi	4.00	3.20
8	MFN	5.00	0

The tasks of Workpackage 3 have been implemented successfully. The summary of progress towards objectives is described below.

- **Objective 1:** Review and analyze the existing methods for data acquisition, curation and semantic integration.

  To achieve this objective, a workshop was organised as part of Task 3.1 on 'Prospective Literature – Toward Best Practices for data acquisition and curation using e-tools for taxonomy' (MS10) on the 14th February 2013 in Leiden (see Ref. 12). Expected results of the workshop were to find new and standardized editorial policies that are needed for the curation and publication of biodiversity data in an e-environment and the collection of data for writing the Best Practices Guide on editorial policies. A questionnaire distributed among consortium members also provided useful information for this guide. The report was submitted in month 9 (D3.1 "Towards a Best Practices Guide on editorial policies", see Ref. 13).

- **Objective 2:** Review current methods for semantic mark-up and their use. Deliverables incorporated in Task 3.3 will help to achieve a thorough review of methods for semantic mark-up in the second year of the project:
  D3.3.1 Report on state-of the art and research horizons of semantic integration of biodiversity literature (month16).
  D3.3.2 Report on progress during the coordination process of partners and non consortium partners (month20).

  A Wiki page for collecting relevant materials has been set up. A first draft of the report is already available (see Ref. 17).

- **Objective 3**: **Facilitate crowd-sourcing/auth-sourcing studies with data curators.** Task 3.2 addressed objective 3 in a workshop on "Legacy literature – Semantic mark-up generation, data quality and user-participation infrastructure", which took place on the 13th February 2013 in Leiden (MS11) (see Ref. 14). Present and alternative ways of publishing were discussed by

---

[4] Based on Annex 1

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 31 of 85

participants, and methods to increase expert and non-expert involvement. Two reports have been submitted in this task:

The **D3.2.1** Concept paper for involvement of individual experts, commercial vendors, and citizen scientists was submitted in month 9 (see Ref. 15). This paper tries to identify possible stakeholder collaborations.

The **D3.2.2** Report on the state and quality of biosystematics documents and survey reports was submitted in month 12 (see Ref. 16). A thorough assessment of biodiversity documents to find missing data and standards is addressed in this report.

- **Objective 4:** Align ongoing and forthcoming efforts to semantic mark-up of biodiversity literature and provide technical and social solutions for their use. The workshop held in February 2013 on legacy literature (MS11) and the D3.3.2 "Report on progress during the coordination process of partners and non consortium partners" (due in month 20) will help to achieve this objective. The integration of the results of the pilot implementation within WP4 (Tasks 4.1 and 4.2, see Ref. 7) and the realisation of a workshop on "mark-up of biodiversity literature" (MS12) in February 2014 in Berlin (see wiki) will also provide technical solutions to align efforts of this objective.

- **Objective 5:** Attract potential new users to biodiversity informatics infrastructures by emphasizing potential improvements in current workflows and collect user feedback during these events. The workshops and activities of WP3 as well as the use of different dissemination platforms of the project (Facebook, Twitter, Google+ and Linked-in, see Ref. 30) for outreach have been and will be used to attract potential new users to biodiversity informatics infrastructures. In addition, active cooperations with Important EU projects in the field of Biodiversity Informatics (e.g. BioVeL, Vibrant) helped to increase the awareness of e-Platforms promoted by pro-iBiosphere.

Details for each task

Task 3.1 Data acquisition and curation (lead: Naturalis; participants: NBGB, FUB-BGBM, RBGK. Start: M3, End: M9).

A workshop on 'Prospective Literature – Toward Best Practices for data acquisition and curation using e-tools for taxonomy' took place on the 14th of February 2013 at the Hilton Garden Inn, in Leiden, the Netherlands. The workshop was targeted at editors of Floras and Faunas and online data curators.
It included ca. 80 participants and 14 invited lectures (to cover the subjects of the meeting). The main objective of the workshop was to identify and promote good practices for entering new field data and collaborative writing of taxonomic treatments. Expected results were, to: (i) establish what are the new and standardized editorial policies that are needed for the curation and publication of biodiversity data in an e-environment; (ii) establish how to address IPR management; and (iii) collect data that will facilitate writing the Best Practices e-Guide on Editorial Policies. A detailed agenda (containing objectives, expected outcomes, target audience, a.o.) is available on the wiki (see Ref. 12).

The lectures covered topics such as: whether Floras and Faunas are the ways to communicate; advanced journal production workflows to technologies to link data, such as linking data to ontologies, references or names; and how to create semantically marked-up publications that allow direct and efficient harvesting and dissemination of the content or parts of it. Text mining was covered as an additional aspect. The demonstration of the production workflows by Pensoft and the European Journal of Taxonomy proved very

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 32 of 85

helpful as a real example to discuss not just an idea but work in progress. Explanation of the advanced technologies and concepts, along with workshop discussions, eased understanding for many of the participants and at the same time gave valuable feedback to the builders of the tools. Though some people voiced satisfaction by being able to retrieve a fixed PDF copy of a work, the overwhelming majority preferred open access advanced publications, with the content being linked to as many resources as possible and with the ability to re-use data in the future.

In order to gain more understanding on the present practices for data acquisition and curation, a questionnaire was designed by NBGB, RBGK and Naturalis. The questionnaire was distributed among the consortium institution staff. The results of the questionnaire and outcomes of the workshop are summarised in the reports on Tasks 2.3 and 3.1. For additional notes of this and the other two workshops that took place in February 2013, please see Ref. 3.

During the workshop, it became clear that some of the issues in legacy literature needed more discussion. This concerned especially the topic of "unique identifiers". In order to receive valuable input from institutions concerning the issue of DOIs in the publishing world, experts were invited to the workshop on " Coordination and routes for cooperation across organisations, projects and e-infrastructures", that took place on 23rd May 2013, in Berlin, Germany, in the framework of T2.1 activities (see Ref. 6).

This task finished in month 9. The deliverable D3.1 "Towards a Best Practices Guide on editorial policies" (see Ref. 13) was submitted in May 2013 (M9) and broadly distributed via the social media tools of the project and consortium contacts/networks.

Task 3.2 Semantic mark-up generation, data quality, and user-participation infrastructure (lead: Plazi; participants: FUB-BGBM, PENSOFT, Naturalis, RBGK. Start: M3, End: M12).

A workshop on "Legacy literature – Semantic mark-up generation, data quality and user-participation infrastructure" took place on the 14th of February 2013, at the Hilton Garden Inn in Leiden (see Ref. 14). The workshop included ca. 80 participants, 15 invited lectures (to cover the subjects of the meeting), one free unsolicited talk and time for discussion. The topics spun from an introductory talk to explain the subject (beginning with the scientific question to the format in which the publications might be published in the future), to various technical issues on how to extract information, how the community might be involved, and how to measure the quality.

During the discussion, the following topics were addressed:

- Why do we publish?
- Do we publish in the right format (given the changes of the underlying digital resources)?
- How to extract content from the legacy material?
- How can all the content be linked up?

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 33 of 85

The final discussion focused on two issues: (i) how to get the content out with limited resources, and thus, how the crowd could be used. Several models have been introduced in this regard. (ii) How the data can be linked and what unique identifiers might be used to achieve this goal.

The participation of experts on the field helped to make this meeting very successful and fruitful. These experts administrate large important databases and have been involved in their development, both from IT and content points of view. They also have a deep understanding of how to get people involved and of the costs involved to make projects sustainable. The combination of European, developing countries and US participants was useful to assure that everybody became aware of developments and to think about how resources could be shared: i.e. names do not know geographic boundaries and thus one global system is enough. A major lesson learned is that a triage system is needed to focus on content that is needed, i.e. content that is wanted by the users of the system, rather than content that is projected to be used.

Two series of hands-on sessions of mark-up at treatment (GoldenGATE tool) and character level (CharaParser tool) were taught twice to the participants of the workshop. For additional notes of this and the other two workshops that took place in February 2013, please see Ref. 3.

Deliverable D3.2.1 "Concept paper for involvement of individual experts, commercial vendors, and citizen scientists" was submitted in May 31ˢᵗ 2013 (see Ref. 15). Deliverable D3.2.2 "Report on the state and quality of biosystematics documents and survey reports" was submitted in August 31ˢᵗ 2013 (see Ref. 16). The state and quality of biosystematics documents were assessed and the recommendations for their future form and content were provided as well as a list of incentives to involve stakeholders in adapting more advanced publishing. This concerns their description in a semantic web environment, including necessary infrastructures like ontologies or reference databases. It includes the feedback from the pro-iBiosphere meetings that took place in Leiden (February 2013) and Berlin (May 2013).

Task 3.3 Semantic integration of biodiversity literature (lead: MFN; participants: FUB-BGBM, PENSOFT, Plazi. Start: M1, End: M24).

The deliverables and milestones for T3.3 are due in the second half of the project. In Y1, it was important to coordinate and align this task with the other tasks of WP3 (in particular T3.2), and the pilot development work in WP4. One of the goals of pro-iBiosphere is to achieve semantic and technical interoperability. Hence, the semantic integration activities of this task are very important for achieving these project goals. Preliminary activities that need to be undertaken by the MfN deal with identifying steps to achieve the semantic integration of biodiversity literature.

Most taxonomic mark-up schemas are presently defined as XML-schemas, which do not provide the means for making machine readable references to the semantic concepts defined by their element definitions. A system decoupling the concept definitions from schema specifications in a way that makes them addressable in a linked data environment would allow us to directly represent corresponding concepts

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 34 of 85

between the CDM (Common Data Model) and mark-up schemas, and to facilitate a greater automation of import processes and data integrity measures. This system will be specified on the pro-iBiosphere Wiki and summarized in a report in M24.

In December 2012, the MfN wrote the first draft for D3.3.1 - Report on state-of the art and research horizons of semantic integration of biodiversity literature. At present, the document describes the data enrichment processes used by BHL, following the three patches described in the Description of Work: i.e. (i) fully automated natural language processing, (ii) base mark-up complemented by automated processing and (iii) specialist correction, social crowd-sourcing models. The document was sent in December 2012 to the BHL-US directors for feedback and additional input. The document is being further developed in 2013 by the MfN and other institutions involved in the task (i.e. Pensoft, Plazi and the FUB-BGBM). In order to prepare the "Workshop on mark-up of biodiversity literature" (MS12), a meeting was held together with Pensoft and BGBM on Thursday, February 14ᵗʰ 2013, in Leiden. During this meeting, envisaged outcomes of the meeting to be held in February 2014 and potential participants were discussed. This information was used as a basis for updating the agenda of the workshop, which is now scheduled to take place on February 10ᵗʰ-11ᵗʰ, 2014 at MfN in Berlin, in conjunction with the workshop MS23 on alternative business models and the Fifth Consortium Meeting. A wiki page has been set up to facilitate the planning of the workshop on Ref. 105.

Wiki pages were also set up for collecting relevant information for and to facilitate the drafting of the upcoming deliverables:

- D3.3.1 Report on state-of-the-art and research horizons of semantic integration of biodiversity literature (due Dec. 2013; Ref. 17) and
- D3.3.2 Report on progress during the coordination process of partners and non-consortium partners (due April 2014; Ref. 106).

All partners involved in this task were invited to provide their input on the D3.3.1 draft, and the structure and content of the report are being fine-tuned further following the October 2013 meetings at FUB-BGBM in Berlin.

Significant results

Making data interoperable

The pro-iBiosphere workshop on "Prospective Literature – Toward Best Practices for data acquisition and curation using e-tools for taxonomy" (February 2013, see Ref. 12) brought together representatives from across the globe (attendees from six continents were present) with a wide spectrum of interested groups, e.g. traditional taxonomists, librarians, publishers, web designers, bioinformaticians. The plurality of ideas and feedback received during the pro-iBiosphere workshop facilitated writing the report on "Towards a Best Practices e-Guide on Editorial Policies". The report includes an overview of (i) available systems and their respective properties and strengths, and (ii) standardized editorial policies that are needed for the

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 35 of 85

curation and publication of biodiversity data in an e-environment. A guide to best practices for stable URIs is being updated on the project wiki (see Ref. 74). URIs are important because they facilitate making data interoperable. With links in place, a broad range of biodiversity information sources will be connected, including literature, organism names, classifications, occurrences, traits, sequences, and phylogenies.

The biodiversity community has discussed about this subject for more than 10 years, and until now, no unified agreement on the use of identifiers for collections had been reached. Based on a process started by the Information Science and Technology Commission of the Consortium of European Taxonomic Facilities (CETAF) in 2012, the coordination workshops available in pro-iBiosphere brought the discussions to a fruitful culmination. The topic of the technical foundations for integration was one of the focus topics in 3 pro-iBiosphere workshops organised by WP3 in February 2013, WP2 in May 2013, and WP4 in October 2013. As a result of this joint pro-iBiosphere and CETAF effort, the adoption of semantic web and linked open data technology is increasing. At present (10/2013), the technology is being adopted by 7 institutions (i.e.: pro-iBiosphere institutions: Plazi, FUB-BGBM, MfN, RBGK, NBGB and two non pro-iBiosphere institutions: Royal Botanic Gardens Edinburgh - RBGE and Museum National d' Histoire Naturelle, Paris - MNHN). These institutions set a recognizable example for others to follow.

URIs are being assigned to key data items and entire documents. For instance, in the case of Plazi, URIs will be assigned to every new treatment imported into the Plazi repository. In the case of the natural history museums and gardens, URIs will be assigned to each specimen record.

Mobilising data from legacy literature
An important aspect of the implementation of a future Open Biodiversity Knowledge Management System (OBKMS) is the capability to mobilise data from potentially hundreds of millions of pages of legacy literature and their integration into contemporary publication processes. It also involves encouraging the biodiversity community to publish biodiversity data and information in non-traditional formats (in order to avoid the accumulation of PDF's, and hence, subsequent mark-up activities). In the workshop "Legacy literature – Semantic mark-up generation, data quality and user-participation infrastructure" (February 2013, see Ref. 14), international stakeholders interested in semantic mark-up activities discussed improved strategies for integration and interoperability of services and data as well as potential incentives for collaborating on mark-up activities. The results have been summarized in a concept paper which provides a valuable basis for future optimised data acquisitions processes (D3.2.1, see Ref. 15). The report on "The state and quality of biosystematics documents and survey reports" (D3.2.2, see Ref. 16) addresses how to improve semantic interoperability of biodiversity data in general, and especially their integration into the emerging Semantic Web. It also provides a set of practical recommendations and guidelines for achieving improved semantic interoperability of marked-up documents.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 36 of 85

Reasons for deviations

Not applicable. There are no deviations on this Workpackage.

Reasons for failing to achieve critical objectives

Not applicable. All critical objectives have been achieved.

Statement on the use of resources

Not applicable.

Corrective actions taken

Not applicable.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 37 of 85

Workpackage 4 Technical and infrastructure coordination

Summary of progress towards objectives

Partner number	Partner short name	WP 4 Person-months (PMs) per partner[5]	Effort (PMs Year 1)
1	Naturalis	3.00	0.26
3	FUB-BGBM	8.00	6.17
4	Pensoft	7.00	3.18
6	RBGK	2.00	0.95
7	Plazi	3.00	4.20

Objective 1. Identify and develop strategies for increased cooperation among current and previously funded EU biodiversity related projects, and with non-EU projects. The work towards this objective is based on the results of, the (i) various workshops organised by the project during its first year of activities; (ii) Deliverables D2.1.1 "Report on ongoing biodiversity related projects, current e-infrastructures and standards" (see Ref. 27) and D2.1.2 "Towards a draft strategy for increased cooperation" (see Ref. 29); and (iiI) online surveys and questionnaires (published in Deliverable D2.1.1). In particular, D2.1.2, includes several recommendations towards an increased cooperation.
The declaration was sent to the CETAF-ISTC members (14th October 2013). A follow up of the recommended declaration will be addressed in the Year 2 activities of this task.

In the framework of Task 4.1 activities, a workshop on "How to improve technical cooperation and interoperability at the infrastructure level" was held in October 2013 (Y2 of the project). The outcomes of the project will be documented in the report "Strategies for improved cooperation and interoperability between infrastructures" (due in December 2013). This draft will also include the results and experience gained through the implementation of the pro-iBiosphere Pilot 3 (see Ref. 53) and partly Pilot 4 (see Ref. 54).

Objective 2. Identify and develop strategies to improve current taxonomic treatment servers and increase cooperation and coordination between the current existing electronic registers for species information. The ongoing work towards this objective is based mostly on an analysis of interoperability issues between sources, repositories and consumers of taxonomic treatments - for example, Plazi, Encyclopedia of Life (EoL) and Species-ID as main repositories for treatments, pilot mark-up activities and as suppliers of treatments from legacy literature; and Pensoft journals ZooKeys, PhytoKeys, MycoKeys and Biodiversity Data Journal as suppliers of prospectively published treatments. Conversion and harvesting mechanisms have been built to extract treatments published in the TaxPub NLM JATS schema and convert these into TaxonX schema. The conversion of treatments from both legacy and prospectively published literature into Darwin Core Archive format has been achieved and tested in beta version in cooperation

---

[5] Based on Annex 1

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 38 of 85

with the EoL, during the pro-iBiosphere workshop in Berlin in October 2013. EOL is the main user of this kind of format to import and interchange treatment and traits data.

A common registration model for new taxon names and other nomenclatural acts with the registries in three organismic domains: plants (International Plant Name Index, IPNI), fungi (Index Fungorum, MycoBank) and animals (ZooBank) has been modeled. The results are provided in an extensive description on the project's wiki (see Ref. 55). The model has been presented at the Biosystematics 2013 Conference meeting in Vienna (February 2013: see Ref. 56). Currently, the model has already been tested in a real-time beta version with IPNI. Index Fungorum will use the IPNI model based on the same International Code of Nomenclature for algae, fungi and plants (ICNafp). The ZooBank registration workflow has also been discussed and modeled. Its implementation was elaborated during a workshop to be held on October 22nd - 26th 2013, in Sofia, Bulgaria (organised by Pensoft). The workshop was attended by the ZooBank manager Dr. Richard Pyle, and developer Mr. Robert Wharton, from the Bishop Museum, Honolulu. The final discussion on registration will take place in October 2013, during the TDWG meeting in Florence. The outcomes of the discussion will be reported in the Deliverable D4.1 "Report on strategies for improved and interoperability between infrastructures" and 4.2. "Report on a strategy for improvement & interoperability of the XML schemas". Both reports are due by the end of December 2013.

Objective 3. Identify current needs and promote future interoperability at technical and semantic level, through improvement and interoperability of the existing mark-up XML schemas and standards. The work towards Objective 3 will be reported in the Deliverable D4.1 "Report on strategies for improved and interoperability between infrastructures" and 4.2 "Report on a strategy for improvement & interoperability of the XML schemas". Both reports are due by the end of December 2013.

In order to make biodiversity data and information (now entrapped in literature) digital, open and linked, various mark-up pilots, using Floras, Faunas and Mycotas are being conducted on digitised texts. These approaches will be tested by Plazi (and participant institutions of Task 4.2) by the end of December 2013.

The pilot activities are also helping to identify the main challenges that prevent efficient interoperability of biodiversity data, and to outline strategies for increasing data exchange and interoperability. At the level of treatments, primary biodiversity data and references, TaxonX and TaxPub (i.e. XML schemas for markup of biodiversity documents) allow an efficient interchangeable level through harvesting mechanisms and XSLT conversions. Based on this, two routes of data integration and interoperability are being tested. The first one, on technical interoperability, will be demonstrated at the level of the Plazi Treatment Repository. Treatments from different sources (e.g., legacy literature that is obtained through OCR and mark-up, and new publications, harvested directly from the XML of recently published articles in Pensoft journals) and originating in different organismic kingdoms (fungi, plants, animals) are placed into a single store in a common format, so that questions of compatibility and data reuse can be tested. A second route of technical interoperability explored is the interchange of treatments between infrastructures, such as Plazi, EDIT Common Data Model (CDM), EOL and Pensoft. The semantic interoperability will be piloted through

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 39 of 85

one of the most prominent semantic Web technologies for data integration: the Resource Description Framework (RDF). Treatments are modeled and expressed in RDF statements and exported into a triple store in the Linked Open data (LOD) cloud. This allows data from different sources and different levels of mark-up granularity to be placed together and be explored to answer specific questions.

Details for each task

Task 4.1 Improve technical cooperation and interoperability at the e-infrastructure level (lead: FUB-BGBM; participants: Naturalis, PENSOFT, RBGK. Start: M4, End M24).

A total of four pilots are being conducted in Tasks 4.1 and 4.2 using the following key groups: Ants, Bryophytes, *Chenopodium*, Chilopoda (*Eupolybothrus*), Ferns (*Nephrolepidaceae*), Fungi, Mistletoes, and Spiders). An overview of the workflow is mentioned below:

1. Mark-up activities, using different approaches, are being performed by domain specialists (in the case of Naturalis: 4 people working on a voluntary basis; in the case of the NBGB, Pensoft, RBGK and Plazi consisting of one person each, affiliated to the project). Prior to the start of mark-up activities, a two day training on the use of the mark-up tool "GoldenGate - GG" (developed by Plazi) was organised (see Task 4.2 for further information on this). With the exception of Loranthaceae, all mark-up pilots used GoldenGATE to carry out semantic mark-up, and the resulting treatments were uploaded to the Plazi repository, which will output treatment in TaxonX output format as one of four possible formats (html, GoldenGATE native XML, RDF). The Flora of China account was marked up by combining GoldenGATE with the other methods.

   Each XML document contains a HTTP request and a response header that includes all the relevant bibliographic data. At present, over 1000 treatments have been marked up and uploaded to Plazi. An export to TaxonX will be studied to facilitate the integration of Loranthaceae marked-up data into the EDIT Platform.

2. Plazi is a digital taxonomic literature repository that allows the delivery of marked up documents (taxonomic treatments) based on the original document name or on a taxonomic name described in the original document. The treatments are saved in a dedicated database with the resolution according to the level of markup, and as entire versioned XML documents. Due to copyright issues, only written treatments are open access, since they are considered data that does not qualify as work in the legal sense. This stands in contrast to contained images and illustrations, which are not saved ([Ref. 113]).

3. Imports to the EDIT Common Data Model (CDM): The CDM is the domain model for the core EDIT components. Wherever possible, the CDM has been made compatible with existing community data Standards. The CDM library provides import and export packages for taxonomic classifications, descriptive data, specimens and observations, and media in many standardised or quasi standardised data formats.

The results will be presented in the Deliverable D4.1 "Report on strategies for improved and interoperability between infrastructures" due in December 2013.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 40 of 85

Pilot 3. Task 4.1 will produce strategies to improve interoperability and coordination between initiatives, projects and platforms at the infrastructure level. A pilot on "Interoperability model between PLAZI and the EDIT Platform for Cybertaxonomy based on transformations between XML-repositories and Common Data Model (CDM)-stores" is being conducted (pilot 3). An important outcome of the pilot will be a workflow to facilitate linking legacy and prospective biodiversity literature and data through mark-up standards and tools. The number of pilot marked-up treatments that have been uploaded in September 2013 is:

- Ants: 20 treatments from 3 publications
- Bryophytes: 25 treatments from 2 publications
- *Chenopodiu*m: 236 treatments covering 17 publications
- *Chilopoda*: 30 treatments from 7 publications
- Fungi: 5 treatments from 1 publication
- Spiders: 17 treatments from 3 publications

Additionally, in the case of *Chenopodium* (see Ref. 59) and Ants, a CDM online portal has been created (see Figure 1).

During the mark-up and upload to CDM, the following issues and challenges were encountered:

- For the workflow from markup to upload (GoldenGATE) to a different system (CDM), it is necessary to understand the data, its structure and granularity on both the provider and reviewer sides, respectively. It also has to be negotiated on which side the data structure has to be changed and what format will be needed to get from a manual import to close to blind interoperability.

- The quality and granularity of the mark-up varies greatly between the different taxonomic groups and has to be improved to a predefined level.

- GoldenGATE (GG) is explicitly loose (i.e., not based on a schema or DTD) and allows to mark-up any elements in a text. Users can therefore not be enforced to follow a specific mark-up syntax and structure. However, workflows (pipelines) implemented for specific projects would essentially prescribe the markup to be used to create the documents required. Software systems processing GoldenGATE output documents have to rely on example structures and informal specifications rather than defined schemas. As a consequence, documents often have to be fed back to the persons conducting the mark-up process in order to correct the data. More data quality checks and validation mechanisms at the GG level would help to streamline this process and should therefore be considered.

- The GG visualisation of mark-up data is complex and can become an obstacle for less experienced users. An HTML preview function transforming the mark-up into a more readable web-page would greatly help to speed up the mark-up process and improve its quality.

- Due to the very heterogeneous structure of legacy literature, GG does not capture some information in old literature, and this requires that text must be respectively tagged manually. Such cases are, for instance, the tags [elevation] and [quantity], when 'm' (meters) and 'mm' (millimeters) precede the figures, and tag [species count], when very often the symbols for male and female are used.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 41 of 85

- In particular ▪ when dealing with long documents, users may lose track of mark-up steps already performed and steps that still need to be performed. A function for monitoring the mark-up process and informing users about what still needs to be done to meet the minimal quality requirements would be helpful.

To address the above mentioned challenges, the FUB-BGBM compiled a list of minimal quality requirements which should be fulfilled by mark-up documents to be successfully imported (see Ref. 58). At present, the agreed mark-up standards refer to:

nomenclature: scientific name, author and nomenclatural reference – the latter at least roughly marked as a bibref or a citation.

material_citations (i.e. specimens, observations, types): locality and observation/collecting event details.

description / supplemental taxon data: diagnosis (for technical reasons, here and not under descriptive data), biology_ecology, scientific name.

distribution: a summary of the distribution based also on the data given in detail in the material_citations as string.

references: bibliographic references as string.

descriptive data: at present only with very low granularity. The data cannot be marked-up in TaxonX, but can be inserted as SDD, or referred to as auxiliary material. We have started communicating with the developer of the CharaParser software.

polytomous keys are also important elements that cannot be precisely marked-up with TaxonX yet, but are accommodated in the CDM.

treatment: a block of text that is explicitly linked to a particular taxon. This taxon is defined in the nomenclature section of the treatment. A treatment will be given a unique identifier.

citation: a reference to a treatment.

Developing a mark-up quality checker function is being discussed.

The above-mentioned standards document is being updated on a regular basis. Problems were regularly reported to pilot-leaders through online folders created on GoogleDrive. Regular Skype meetings between Plazi and FUB-BGBM have helped to resolve technical issues of the interoperability workflow. In order to increase progress of the pilots in Y2, the FUB-BGBM and Plazi plan to increase communication between IT staff and pilot leaders, so that problems can be addressed in real time.

Because of the above-mentioned challenges and issues (i) the alignment and improvement of mark-up outputs into the preferred schema, TaxonX, is taking more time than originally expected; and (ii) the goal to increase time efficiency using a mainly automatic mark-up process has not been reached in a way that would allow to benefit from its full potential. Different approaches to streamlining the mark-up workflows supported by different mark-up editor systems need to be further discussed between mark-up system developers, users, and import software developers.

pro-iBiosphere FP7 Project ▪ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ▪7th Framework Programme
▪ Coordination and support action FP7-INFRASTRUCTURES-2012-1▪ Subprogram area INFRA-2012-3.3

Page 42 of 85

The use of inexperienced staff members to perform the mark-up showed the importance to produce services that pre-process documents to a level that can be done by professional vendors. This might end up with the training of a vendor for as many as possible mark-up projects. Mark-up that needs domain experts can then be done by domain specialists.

More data are needed to decide optimal strategies in terms of mark-up procedures (GoldenGATE, perl-scripts, MSWord macros). Alternative ways to mark-up have been discussed at the pro-iBiopshere meeting in October 8, in Berlin, Germany (see Ref. 72).

Ongoing activities:

- Specification of PLAZI to CDM pipeline and interfaces. In the past months, PLAZI has implemented a new web-service for delivering the list of documents that have been marked up and stored in their repository. An important attribute, the original page number, has been added to the document metadata.
- Assessment and discussion of Plazi/GoldenGATE export formats. The native format of GoldenGATE is not XML-valid. The mark-up level of TaxonX has been investigated and will be reviewed in the "Report on a strategy for improvement & interoperability of the XML schemas" (D. 4.2, due in December 2013).
- Imports into the EDIT Platform for Cybertaxonomy. The new version of the ABCD import has been useful for the specimen data, which is linked to treatments. The TaxonX import implementation is continuously being improved.

Linking of cited treatments within a treatment has been tackled by re-designing the mark-up policy by using specific elements for treatment citations. In this way, the import of TaxPub-based documents from Pensoft to Plazi will retain the respective treatment citation element and in return will allow imports of the mark-up to CDM.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 43 of 85

Figure 1. Chenopodium data portal containing species information on nomenclature, description, habitat, common names, images and distribution maps.

Preparations of workshop on "how to improve technical cooperation and interoperability at the e-infrastructure level"(MS13, in October 2013). In year 1, the aim has been to build a specialized user group, including external experts and users, to identify problems in the interoperability at the e-infrastructure level, and ways on how to improve technical cooperation. The feedback received from the participants of the workshops in May 2013 (see Ref. 66) and the report on "Ongoing biodiversity related projects, current e-infrastructures and standards" (i.e. D2.1.1., submitted by Plazi in June 2013; see Ref. 27) prepared the ground and delivered the information needed for the planning of the workshop on "How to improve technical cooperation and interoperability at the e-infrastructure level" (held on the 8th of October 2013).

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 44 of 85

The aim of the workshop is to create a productive work atmosphere with key experts (spread across smaller expert groups), who can give valuable input. Instead of trying to tackle the interoperability issue in general, the project decided to focus on two key interoperability aspects where considerable progress can be made during the two year lifetime of the pro-iBiosphere project. The two work-groups reflect the discussion focus of the past months: identifier work-group (group 1) and biodiversity registries work group (group 2). For the participants list, see Ref. 65). The identifier and registry topics will also be the key elements of the interoperability report D4.1.

In the previous months, there have been ongoing discussions and intensive consultations on various platforms, to (i) analyse the accumulated experience in the use of identifiers (e.g., DOIs, LSIDs) and (ii) propose a model that will make these identifiers universally usable within the biodiversity domain. These discussions will be followed up with the identifier work-group (see Ref. 62).

pro-iBiosphere meeting 4. The FUB-BGBM was allocated a budget to organise this pro-iBiosphere meeting 4 in October (consisting of four workshops, including the technical cooperation workshop), and as such, BGBM were responsible for organising the logistics of the event. In August 2013, official invitations were sent to envisaged participants. Prior to the workshops in October 2013, regular online meetings with the organisers of the workshops have been held in order to discuss the agendas, envisaged participants, and specific requirements of each workshop. General information on the event, detailed agendas, lists of participants and travel information are available on the pro-iBiosphere Wiki (see Ref. 60).

Deliverable D4.1 The "Report on strategies for improved cooperation and interoperability between infrastructures" (D4.1) is due in December 2013 (see Ref. 21). The main topic of this report is the service interoperability between taxon-level biodiversity informatics infrastructures, such as taxonomic working platforms (CDM, Scratchpads), taxon treatment repositories (EOL, Plazi, Species-ID) and biodiversity publishers (Pensoft). The report will focus on two key aspects of a distributed biodiversity informatics infrastructure which are identifier systems and registries for biodiversity-related web-services. The report will also assess the pilot implementation for pipelining information from XML-based PLAZI repositories to CDM-stores. The FUB-BGBM made available a collection of materials for a first draft in August 2013. Results from the (i) pro-iBiosphere May workshop on "Coordination and routes for cooperation" (see Ref. 6), (ii) pro-iBiosphere October workshop on "How to improve technical cooperation and interoperability at the e-infrastructure level" (Ref. 62), and (iii) the pilot implementations on an "Interoperability model between PLAZI and the EDIT Platform for Cybertaxonomy" (to view the poster on the model, see Ref. 64) will be used as the basis of this report.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 45 of 85

Task 4.2 Promote and monitor the development and adoption of common mark-up standards and interoperability between schemas (lead: Plazi; participants: Naturalis, FUB-BGBM, PENSOFT, RBGK. Start: M4, End M24).

Pilots. Within this task, the following three pilots are being conducted:

- **Pilot 1.** "Interoperability model between taxon treatments from both legacy and prospective literature from three organismic domains (fungi, plants and animals)", led by Plazi.
- **Pilot 2.** "Common query/response model for automated registration of higher plants (International Plant Names Index, IPNI), fungi (Index Fungorum, MycoBank) and animals (ZooBank)", led by Pensoft.
- **Pilot 4.** "Revision of a tool (CharaParser) that generates identification keys by reusing morphological characters from published species descriptions", led by Plazi.

Detailed information on the pilots is available on the pro-iBiosphere wiki (see Ref. 7).

Pilot 1. "Interoperability model between taxon treatments from both legacy and prospective literature from three organismic domains (fungi, plants and animals".

The pilot aims at obtaining content out of the published record in a form that can be re-used in the cybertaxonomy environment. Naturalis, Pensoft and Plazi drafted in the starting phase of the project, and in close cooperation with all partners, a proposal for a workflow focusing on the realisation of pilots and the basic criteria for selection of taxa. Taxa and criteria were discussed during face to face and online meetings, which led to a successful selection of (i) pilot taxa, dealing with zoological, botanical and fungi data, and (ii) granularity of mark-up. The description of the pilots and the associated workplan are available on the project wiki (see Ref. 7). Pilot projects include the following taxa: Ants, Bryophytes, *Chenopodium*, Chilopoda (*Eupolybothrus*), Ferns (*Nephrolepidaceae*), Fungi, Mistletoes, and Spiders.

PIlot activities were discussed with the respective specialists, and the processing of files started in March 2013. A Google Drive folder was set up for the pilots, whereby the data can be uploaded and made accessible to all participants. For discussions on mark-up with the persons leading the pilots, online and face-to-face meetings were organised. Throughout the pilot, Plazi provided help to improve the mark-up of publications (generated with the tool GoldenGATE) that were causing problems during the process. The mark-up of all pilot groups produced marked-up treatments. These treatments were uploaded to the Plazi repository (see Figure 1 below) and imported into the Common Data Model (CDM; see Ref. 67).

The pilot mark-up process on different organism groups has allowed us, to:
1. Monitor and improve the workflow from legacy literature to digital information available through Plazi and the EDIT Platform.
2. Set common standards.
3. Provide estimates of the time frame necessary to create fine-grained mark-up.
4. Compare the different technical approaches used for mark-up. A pro-forma spreadsheet was set up online for this purpose at the beginning of the pilots). Where possible, time budgets for different stages of the mark-up for the different methods have been recorded, so it should be possible to get

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 46 of 85

comparative ballpark estimates of the time required to mark-up elements such as treatment boundaries.
5. Identify bottle necks and find approaches for solving them.

Acquired insights will be used in Y2 to:
1. Assess the strengths and weaknesses of the workflows used to build the pilots. Including text capture, semantic enhancements, integration, and underlying services needed for automatic data integration.
2. Define a workflow and define specifications deemed necessary for an OBKMS.

The "Report on a strategy for improvement & interoperability of the XML schemas" (D4.2, due in December 2013) will document these and other findings.

Training on GoldenGATE. An introductory course to GoldenGATE-based mark-up was prepared and run by Plazi, and locally organised by Naturalis on January 14 and 15, 2013. The purpose of the workshop was to:

- teach the participants of the pilot studies how to mark-up their legacy documents using GoldenGATE,
- give the participants the opportunity to test the entire workflow from document to structured data,
- enable an efficient development of the pilots.

A total of 18 participants attended the training, including two tutors from Plazi. A customized manual and new version of GoldenGATE allowing direct input of PDF files was prepared and afterwards improved using the feedback received from participants. During the hands-on sessions, participants had the opportunity to upload files into the GoldenGATE Editor and work through all subsequent mark-up steps with the assistance of the organisers. After the training course, the manual for GoldenGATE has been revised to accommodate elements that have been ambiguous or missing and will in the second period be moved onto a Wiki for community editing. Though the PDF import function was deemed necessary to shortcut an external OCR step, the results have been ambiguous, not least because of the very variable input quality of the PDF documents, both image and text based. This step clearly has to be addressed and resolved to operate efficiently within OBKMS and thus will be discussed in the forthcoming pro-iBiosphere meeting taking place in February 2014.

A short course has been repeated as side event at the pro-Biosphere meeting in Leiden, February 13, 2013 (see Ref. 68). Google docs have been set up for the task and for each of the pilots projects on Ref. 69. Skype and email have been used for further support. A final technical meeting took place on October 7, 2013, prior to the pro-iBiosphere Meeting in Berlin, October 8 (see Ref. 60).

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 47 of 85

PLAZI
taking care of freedom

GoldenGATE Search & Retrieval Server by Guido Sautter, IPD Böhm, Universität Karlsruhe (TH), 2009
Search through 26554 treatments (2136 documents) in the archive. view statistics

Back to Search Form

GoldenGATE SRS Search Result: 116 Treatments [more] GoogleMaps					
Scientific Name	Status	Publication	Pages	ModsID	GoogleMaps
Chenopodium album		Charles Joseph Pitard and Louis Proust, 1908, Chenopodieae, Les îles Canaries. Flore de l'archipel. Paul Klincksieck, Paris, pp. 326-327: 326-327, (download)	326-327	3784-0748-2261	
Chenopodium glaucum		Charles Joseph Pitard and Louis Proust, 1908, Chenopodieae, Les îles Canaries. Flore de l'archipel. Paul Klincksieck, Paris, pp. 326-327: 326-327, (download)	326-327	3784-0748-2261	
Chenopodium urbicum		Charles Joseph Pitard and Louis Proust, 1908, Chenopodieae, Les îles Canaries. Flore de l'archipel. Paul Klincksieck, Paris, pp. 326-327: 326-327, (download)	326-327	3784-0748-2261	

Figure 1. Marked-up treatments uploaded onto the Plazi repository.

The progress of these pilots is briefly described below:

- The centipede genus *Eupolybothrus* (Lithobiidae, Chilopoda) comprises 24 valid species and 15 doubtful (sub-)species assigned to 7 subgenera ranging from mostly Southern Europe and North Africa to the Near and Middle East, including the largest Mediterranean islands of Corsica, Sardinia, Sicily, Crete and Cyprus (Zapparoli and Edgecombe 2006, 2011). The genus shows the highest species diversity in the Italian and Balkan peninsulas (Zapparoli 2003), where a number of cave-dwelling species present restricted distribution ranges. Further 66 species-level taxa proposed in *Eupolybothrus* are currently considered junior synonyms, although this taxonomic status might change in the light of future taxonomic and molecular study. Several new species have been revealed by the recent molecular analysis based on Cytochrome C Oxidase Subunit I gene (COI) barcoding. There are ca. 200 major treatments related to 104 taxon names, published in 50 different articles between 1847 and 2011. Within the pilot, a total of 130 taxon treatments have hitherto been marked up by Pensoft, of them 90 uploaded to the Plazi platform. By mid-October 2013, we envisage to have all original description treatments for the species of *Eupolybothrus* and all treatments for the *Eupolybothrus caesar* group marked-up and uploaded on Plazi. A publication based on the pilot will be prepared by Pensoft' team and is scheduled to be published in the Biodiversity Data Journal. The publication will demonstrate for the first time an enhanced Cybertaxonomy checklist encompassing virtually all legacy literature available on the known species. The checklist will be accompanying the description of a new species of the genus, which is also a part of the pilot demonstrating how legacy and recently published treatments could be put together within one treatment repository (Plazi) and linked to in a recent publication. The description of a new species of *Eupolybothrus - E. cavernicolus* - has already been published in the journal, with rich data accompanying the morphological description.

- Spiders. The purpose of this pilot is to demonstrate a two-pronged approach to dealing with the massive challenge of populating a taxonomic scaffolding (represented by the World Spider Catalog) with treatments. The first part is to markup taxonomic articles in which spider treatments are concentrated. To represent this, we have marked up all the open access spider articles in Zootaxa. The second part is to have specialists mark-up the remaining articles relevant to a particular taxonomic domain. One of the open access articles in Zootaxa deals with the spider family Penestomidae. A handful of other papers have contributed treatments on Penestomid species, and these have now been added.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 48 of 85

- In the *Loranthaceae* pilot, we have started to transform existing mark-up into TaxonX. For this pilot, we have been concentrating on the feasibility of fine grained atomisation of geographical and morphological information, both of which rank high in the feedback from the user requirements workshop (May 2013), and the mark-up of identification keys. A schema for mark-up of the morphological descriptions and the detailed geographical distribution has been developed and applied to existing marked-up text for *Helixanthera* (from three publications - Flora of Tropical East Africa, Mistletoes of Africa, Flora Malesiana) and de novo to account for *Helixanthera* from the Flora of China. The existing mark-up was produced using the "Kew method" or derivations of it. The Flora of China account was done using a combination of GoldenGATE with the other methods. It is apparent that different methods have their strengths and weaknesses for different stages of the mark-up process.

- The *Chenopodium* pilot concentrated on two aspects of the mobilisation of legacy literature, i.e.: assessing (i) the value of the legacy literature as a source of distributional records, and (ii) the ease of workflow from paper to the EDIT Common Data Model database. Seventeen different texts containing 127 treatments of *Chenopodium* species were marked up using the GoldenGATE editor. These texts are a broad cross-section of the taxonomic literature spanning the 19th, 20th and 21st centuries. These texts were mostly written in English, but several other languages were included that are also written in Latin script. Treatments were converted to digital text in various ways, including through commercial OCR systems, crowd-sourced transcription (Wikisource) and manual transcription. During the process, details of difficulties were documented and some improvements to the GoldenGATE editor were made. Completed texts were deposited in the Plazi repository, from where they were uploaded into the CDM store, but not before mark-up errors and inconsistencies had been resolved. Resolving errors involved a certain amount of iteration between editing in GoldenGATE and the CDM upload scripts. As a parallel activity, we have gathered localised records of one species of *Chenopodium* from as many different sources as possible, such as databases, legacy literature and herbaria. These data will be used to assess the value of botanical literature as a source for biogeographical information.

- *Nephrolepidaceae:* A published account of the Nephrolepidaceae (Hovenkamp & Miyamoto, 2005) was marked up using the GoldenGATE editor and has been deposited in the Plazi repository. During the mark-up process, there has been intensive contact with the developer of GG, and as a result, various updates of GG were installed during the process. This is probably the reason that the level of detail of mark-up is at present very uneven throughout the document, and additional mark-up is necessary. As the automated GoldenGATE pipelines do not allow this, this is much more time-consuming, and will be a task for the coming months.

- Fungi. Mycorrhizal fungi play a central role in the functioning of terrestrial arctic ecosystems, providing minerals and water to arctic plants in nutrient-poor environments and harsh climates. Ectomycorrhiza (ECM) are the predominant mycorrhiza type in arctic and alpine environments. ECM fungi are crucial for the survival of arctice shrubs (e.g*., Betula, Dryas, Salix*). The goal of the fungi pilot is to provide tools that will facilitate taxonomic work by making comparisons of morphological and ecological data more efficient. For this purpose, we have focused on the mark-up of the taxonomic literature of arctic ECM *Lactarius* species. This genus works well as a study case, because of the manageable number of arctic species (ca. 20) and available monographs. Moreover, we have recently detected 2-3 putatively new species lineages in the Arctic using molecular phylogenetics, and these can serve as a test case for the above comparisons. We are finishing up the scanning and the mark-up of the selected literature using the GoldenGATE editor. Validation and quality check of data integrity are still required before depositing the data in the Plazi repository. The next step is to implement character parsing tools to utilize the deposited dataset to facilitate morphological comparison of the taxa of focus.

- Bryophytes. The genus *Campylopus* comprises a total of 13 species occurring in the Guianas region (Guyana, Surinam, French Guiana). The genus was recently treated in the Flora of the Guianas

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 49 of 85

(Series C, fascicle 2, year 2011) but some species descriptions can only be found in the Flora of Suriname (vol. VL, part 1, Musci I and II, year 1964). The pilot focuses on marking up descriptions that are either available only as hard copy or no longer available. Various text files were scanned and marked up using the GoldenGATE editor. Upon completion of mark-up activities in June 2013, the treatments were uploaded onto Plazi. A second upload was required due to initial omission of the associated metadata. At present, Plazi and the FUB-BGBM are solving a few technical issues to allow the pipeline between Plazi and CDM.

- Ants. For the ant genus *Anochetus* from Madagascar, including 5 known and one new species, all the publications from which treatments have been cited have been marked up and those published in English have been further parsed with CharaParser to extract characters and character states. The parsing of the characters using CharaParser is in process and will be finished by the end of November. The entire workflow and logistics has been discussed and agreed in August. This includes as next steps parsing of the characters, transfer of the characters into the matrix system at the Hymenoptera Name Server at Ohio State University (Norm Johnson; Ref. 114), editing the matrix with new data, and supplying all the auxiliary data (links to images, Barcodes, specimen data) by Brian Fisher (Ref. 115), creating a taxpub based manuscript semiautomatically to be submitted to Zookeys in the second period of the project. Additionally, an effort has been made to develop and implement stable identifiers to be able to embed links from referenced treatments to the treatment proper (see below).

Pilot 2. "Common query/response model for automated registration of higher plants (International Plant Names Index, IPNI), fungi (Index Fungorum, MycoBank) and animals (ZooBank)".

Two presentations on the "Common query/response model for automated registration of higher plants (International Plant Names Index, IPNI), fungi (Index Fungorum, MycoBank) and animals (ZooBank)" were given by Pensoft at the 2ⁿᵈ pro-iBiosphere meeting in Leiden and at the BioSyst 2013 conference in Vienna. A poster on the latest results from the pilot has been presented at the Bioinformatics Horizon 2020 meeting in Rome in September 2013 and the pro-iBisophere meeting in Berlin in October 2013 (see Ref. 73). This resulted in an ongoing testing of the preliminary XML schemas underpinning the common XML-based automated publisher-registry pipeline. The model has already been tested in a real time beta version with IPNI. Index Fungorum will use the IPNI model as being based on the same International Code of Nomenclature for algae, fungi and plants (ICNafp). The ZooBank registration workflow has also been discussed and modeled. Its implementation will be elaborated at a workshop in Sofia organised by Pensoft on October 22-26ᵗʰ 2013 and attended by the ZooBank manager Dr Richard Pyle from the Bishop Museum, Honolulu, and the ZooBank developer Robert Wharton from the same institution. The final discussion on registration will happen at the TDWG meeting in Florence in the end of October 2013 and reported in the Deliverable D4.2. The full description of the pilot and sample XMLs of the first successful tests are available from Ref. 55.

The registration of taxonomic and nomenclatural acts could be done by three main groups: (1) authors, (2) registry curators, and (3) publishers, hence the workflow should be based on a multiple-choice principle (Figure 2). The registration process proposed in the current workflow is "journal-centric".

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 50 of 85
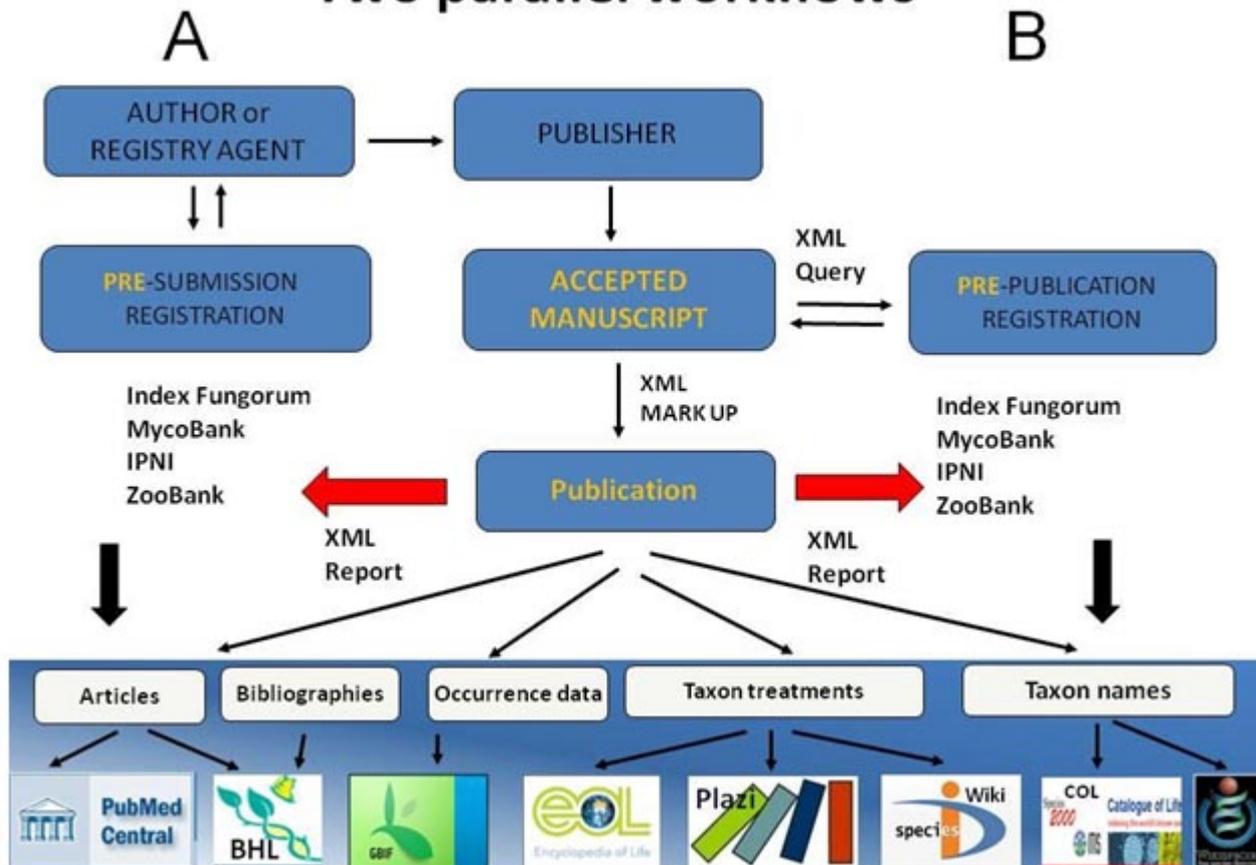
Figure 2. Two possible workflows for registration of taxonomic and nomenclatural acts, provided by an author or registry curator (A) or by a publisher (B). In both cases, the published information will automatically be amended in the registry through implementation of XML- or JSON-based server-to-server services.

In case a publisher takes the responsibility for the registration of taxonomic and nomenclatural acts in the above mentioned electronic registries, the workflow can be performed following a common model (Fig. 3):

Step 1. XML query from the publisher to the the registry on the day of acceptance of the manuscript (containing the type of act, taxon names, and preliminary bibliographic metadata).

Step. 2a. Response XML report containing the unique identifier (e.g., LSID, PURL, or other resolvable URLs) of the act and potential error messages.

Step. 2b. Error correction and de-duplication performed manually: human intervention, at either registry's or publisher's side (or at both).

Step. 3. Inclusion of identifiers in the published treatments (protologues, nomenclatural acts).

Step 4. Final XML report sent by publisher on the day of publication (exact bibliographic details of the published article: authors, title, journal, issue no, date of publication, pagination).

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 51 of 85

Automated registration
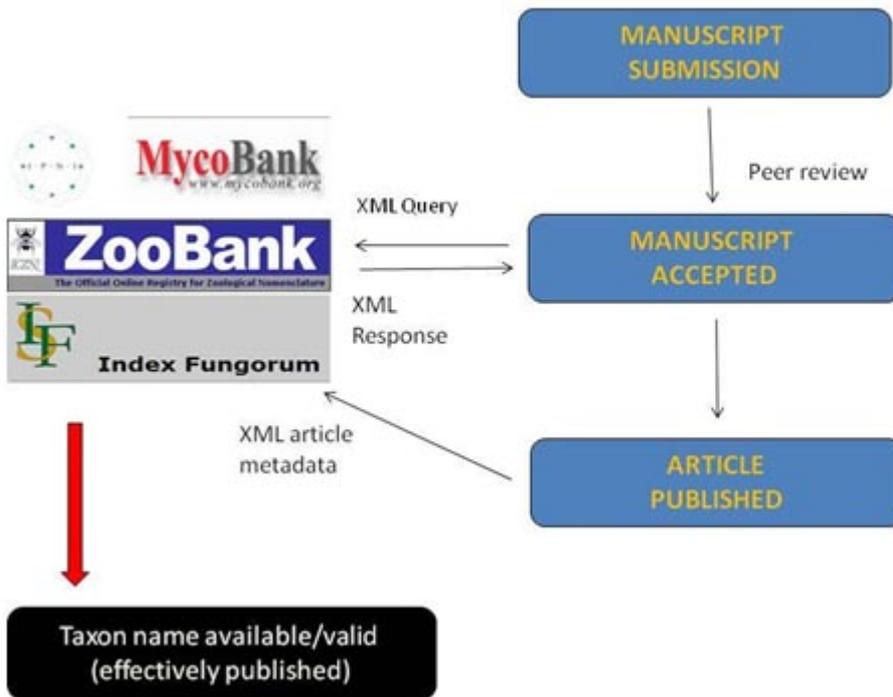
Figure 3. Automated registration process and validation of finally published data and metadata between publisher and registry.

There are several reasons to maximize 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
- Decreased risk of errors caused by human intervention (e.g. re-typing)
- Disambiguation of the dates of acceptance and publication of a manuscript
- Efficient and accurate validation of final published data and metadata through automated export from the publisher to the registry on the day of publication

Pilot 4 "Revision of a tool (CharaParser) that generates identification keys by reusing morphological characters from published species descriptions".

The pilot aims to revise a tool (CharaParser) that generates identification keys by reusing morphological characters, locality and bibliographic citations.

In order to discuss the pilot and the training on CharaParser, Plazi attended an initial three day technical workshop by Hong Cui in Tucson, on December 5-7th, 2012. Subsequently, a training workshop on the use of CharaParser was held during the pro-iBiosphere meeting that took place in February 2013. The training had a total of 11 participants.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 52 of 85

The current workflow includes:

- Mark-up 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.
- Production of a linked open access Taxpub based publication.

A description of the use of CharaParser is planned using ants and fungi (data and information).

The ant pilot focuses on the extraction of treatments from the legacy publications (4 publications cited in Fisher and Smith, 2008) covering the ant genus *Anochetus* in Madagascar, including publications that have been cited in the last revision by Fisher & Smith (2009). Character and character state data will be extracted using the CharaParser tool. The characters will be uploaded to the Hymenoptera Name Server. The character matrix will then be updated with data from a new *Anochetus* taxon. A new publication will be created to describe a new taxon and redescribe the existing ones using the tools at the Hymenoptera Name Server to export semi-automatically a manuscript in Taxpub JATS, including links to the names, bibliographic references, materials observed and treatments.

Due to the heterogeneous nature of the ant input documents (description paragraphs are usually mixed with diagnosis, distribution, or discussion paragraphs), the use of CharaParser appeared more time consuming than expected.

The pilot is now being conducted with fungi (description paragraphs do not have the issues mentioned above). A detailed comparison between the CharaParser's output, Matrix Generation output, and correct output (sent by the domain specialist to the IT CharaParser team) will be conducted in Y2..
Like GoldenGATE or a cascade of scripts to mark-up floras, CharaParser is a research tool in a stage where the human interface has not been properly developed and thus makes it difficult for non specialists to use. It is thus hardly possible to create a customized version, which has been explored for GoldenGATE. The interest of the participants and users of biodiversity data for morphological data will make the development of industrial strength software for mark-up adamant.

Significant results

Significant results include:

- Specifications of minimal requirements defining the interface between the mark-up information flows and the CDM imports have been produced.

pro-iBiosphere FP7 Project ∎ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ∎7th Framework Programme
∎ Coordination and support action FP7-INFRASTRUCTURES-2012-1∎ Subprogram area INFRA-2012-3.3

Page 53 of 85

- Transformation software allowing imports of mark-up documents into the EDIT-CDM platform have been developed.
- Key interoperability topics tackled in both interoperability workshop and report have been identified and agreed upon.
- Active collaborations with CETAF and BioVeL on these key topics have been established.
- The technical ground for the establishment of i) a consistent space of stable identifiers for collection objects across European taxonomic institutions and ii) a central registry for biodiversity-related services has been prepared.
- An improved GoldenGATE version that allows marking up over 1000 treatments has been produced.
- A best practice for stable URIs has been created.
- The ground work for the common query/response model between Pensoft and Zoobank.

Tasks 4.1 and 4.2, have outlined strategies for increased interoperability of standards and at technical level. Interoperability will be achieved by wider application of stable HTTP identifiers, Darwin Core Archive interchange formats and RDF (activities due in Y2).

Reasons for deviations

Not applicable. There are no deviations on this Workpackage.

Reasons for failing to achieve critical objectives

Not applicable. All critical objectives have been achieved.

Statement on the use of resources

Not applicable.

Corrective actions taken

Not applicable.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3
Page 54 of 85

Workpackage 5 Dissemination, communication and public awareness

Summary of progress towards objectives

Partner number	Partner short name	WP 5 Person-months (PMs) per partner[6]	Effort (PMs Year 1)
1	Naturalis	0.95	1.22
2	NBGB	0.75	0.52
3	FUB-BGBM	0.75	0.28
4	Pensoft	5.75	5.59
5	SIGMA	8.00	3.20
6	RBGK	0.75	0.05
7	Plazi	3.80	0.40
8	MFN	0.75	0

The tasks of Workpackage 5 have been implemented successfully. The summary of progress towards objectives is described below.

Objective 1. Development of the project image, documentation and web platform. The project graphical identity (logo and documentation) were developed in September 2012. Two online platforms (wiki and website) were designed, developed, tested and publicly released at the beginning of the project (October 2012). The pro-iBiosphere website domain (see Ref. 30) was obtained for a period of 10 years. Various project materials were produced for outreach activities, including three versions of the project poster and three pilot posters, one leaflet and one post card (see Ref. 8).

Objective 2. Dissemination of the project results through outreach activities. Three newsletters were released (see Ref. 21) and four press releases published (see Ref. 22). A dissemination and communication implementation plan (dissemination strategy) was produced and is being implemented by the consortium. The project has given presentations during thirteen international events (see Ref. 23) and published two articles on Ref. 109 and Ref. 110. The list of contributions to events and articles is available on the wiki (see Ref. 10).

Objective 3. Stakeholder engagement and communication. The efforts undertaken in dissemination activities ensured stakeholders' engagement and public awareness of the project. Stakeholder engagement can be assessed through their level of participation and interest in the project events, the number of social media members, and their contribution to the project newsletter.

In order to facilitate engagement and communication with stakeholders, a dissemination database was

---

[6] Based on Annex 1

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 55 of 85

created (at present, it includes more than 400 pro-iBiosphere contacts), and representations of the project on four social media platforms (Facebook, LinkedIn, Twitter, Google+) were developed.

Details for each task

Task 5.1 Development of the project image, documentation and external communication web platform (lead: PENSOFT; participants: SIGMA, Plazi. Start: M1, End M24).

Dissemination and Communication Implementation Plan (DCIP). In order to prepare the dissemination strategy and gather inputs from partners, a first draft on the Dissemination and Communication Implementation Plan was shared with the consortium at the start of the project. The DCIP provides information on, the: dissemination strategy of the project, aim of the dissemination actions, communication and dissemination tools to be used, and awareness of activities and mechanisms for information exchange with various stakeholders. Hence, it is meant to be consulted and used by all partners as the dissemination activities guidelines for the project. The DCIP was released in M4 and placed on the project website (on Ref. 40).

In order to foster synergies and engage in cross promotional activities with various initiatives, two lists were created and shared with partners on the wiki. The lists contain information on (i) other EU projects and biodiversity initiatives; and (ii) social networks to follow-up online dissemination external channels; contribution to events, publications and other dissemination activities. The lists are being updated by all partners on a regular basis.

Logo. Several alternative versions of the project logo were designed before the start of the project to be selected by partners though a vote. Following the logo selection (Ref. 78), the WP5 leader designed the overall graphical identity at M1, setting up the project font and colours.

Documents templates. In M1, three project templates were created: letter, deliverable, and PowerPoint templates. All these templates were shared among partners for improvement and approval on the occasion of the Kick-off meeting. These templates are available on the internal library of the website (Ref. 40).

In M11, to correspond to the policy of pro-iBiosphere that encourages open access, the disclaimer of various templates was updated by including the terms of the Creative Commons Attribution License 3.0 (CC-BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 56 of 85

Leaflet. In collaboration with the consortium partners, the task leader drafted and designed a concise presentation of the project goals and outcomes in a promotional leaflet (see Ref. 79). At present, 1000 copies have been printed for initial on-site distribution.

Website. The web domain www.pro-ibiosphere.eu was registered for a period of 10 years. The website was designed, developed and tested by the task leader at M2. All partners were registered as users and assigned a login and password to access the internal communication platform (ICP). The Guidelines for use of the website and internal communication platform (ICP) were distributed to all partners in M3.

The platform consists of several informational and operational subpages and tools, such as:

- Registration module,
- External and internal document libraries that allow upload and archiving of different types of documents alongside with their descriptions (metadata),
- ICP to facilitate easy communication between partners by Workpackages and task groups,
- Electronic newsletter with email subscription module,
- Media centre and social network profiles,
- RSS feeds to disseminate news and events,
- Feedback/comment options at the level of documents and for the project website as a whole.

All along the project period, the website has been used as a major dissemination channel of the project while displaying news, articles, presentations, internal and external documents. During the reporting period the pro-iBiosphere website and the wiki platform have continued to serve as main sources of information about the project results, 58 news stories have been posted in the News section. Both the website and wiki have been updated in real time with information on the project activities and its partnership. For the M1 – M12 period, the website had 61 432 page loads and it had been visited 12 651 times, of which 5 346 are unique visitors and 7 305 returning visitors. Most visited pages are: Homepage (16 037), News (4 443), ICP Login (3 464), Events (2 194), and Library (2 158). The website has been visit from 117 countries, main traffic comes from Bulgaria (2 679), Germany (2 313), the United Kingdom (1 059), the Netherlands (1 016) and France (870). The website was visited via social media 264 times, of which 96 via Facebook, 66 via LinkedIn, 58 via Twitter and 10 via Google+.

Wiki. The wiki is being used by the consortium to (i) facilitate sharing of documents that can be edited and modified by partners; (ii) prepare project events; (iii) share lists of partners' contribution to events and publications, other initiatives, pilots; (iv) promote on-going activities and outcomes of the project, a.o.

The wiki is being regularly updated by all partners to share project events information (agenda, participants list, useful information) and to report on their contribution to the different Workpackages.

eNewsletter. The pro-iBiosphere eNewsletter was produced and disseminated by the task leader through the various contact networks and the pro-iBiosphere dissemination database. Three newsletters were produced during the 12 month reporting period. The project newsletters are disseminated broadly via the pro-iBiosphere dissemination database and and to the project consortium. Partners were encouraged to

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 57 of 85

further share the eNewsletter to potentially interested parties across their own and institutional networks.

All newsletters are available for download in PDF format in the news section of the website (Ref. 9).

Documentation. The various dissemination materials have been produced by the consortium are available on the media center section of the website (Ref. 8). These materials include the:

- Project posters (Ref. 80)
- eNewsletter (Ref. 9)
- Leaflet presenting the project goals and outcomes (Ref. 79)

Press release. A total of four press releases were produced by the consortium and disseminated through EurekAlert and various bioinformatics and biodiversity related mailing lists (e.g. TDWG Structure of Descriptive Data, TDWG Content, Open Science, Flora Malesiana).

In M2, a press release communicating the launch of the project was issued (Ref. 81). In addition to English, the document was translated and distributed to important biodiversity institutions in Bulgaria.

Three additional press releases were produced in M12, after achievement of key deliverables (i.e. D2.3, D3.2.2 and D4), and posted via EurekAlert.

Registration and questionnaires for meetings. In order to: (i) facilitate registration by participants to the three workshops organised in M9, an online registration form was uploaded on the website; (ii) gather information previous to the workshops, three questionnaires were designed and uploaded on the website; (iii) provide feedback from participants after the workshop, an evaluation questionnaire was uploaded and distributed to all participants.

The registration template and online questionnaire are available on the internal library of the project's website (Ref. 40; a login account is required).

Presentations. All presentations given during the pro-iBiosphere workshops that have been organised during the reporting period are available for download on the project wiki (Ref. 39).

Deliverables and milestones. During the reporting period, the following milestones were achieved, and deliverables submitted:

- MS15: First set of promotional materials - M4
- D5.1.1: Web-based platform and wiki - M4 (Ref. 83)
- D5.1.2: Electronic newsletters - M4, M8, M12 (Ref. 21)

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 58 of 85

Task 5.2 Dissemination of the project results through outreach activities (lead: SIGMA; participants: all partners. Start: M3, End M24).

Contribution to events. In order to plan and report on partners' contribution to events, a table was prepared in M1, listing relevant external events (biodiversity events as well as European Commission's events). The table also includes a section on articles to enable partners to report on submitted articles ([Ref. 23](#)).

In order to facilitate dissemination of outreach activities on the project website, a specific form was designed ([Ref. 84](#)); to access the link, a login account is required). This form enables partners to promote on their contribution and to add their presentation or article in order to publish a news item on the project website.

In M9, the pro-iBiosphere consortium, in collaboration with other e-infrastructure projects, drafted a proposal for organising a networking session and holding an exhibition booth at ICT2013 (an event organised by the European Commission's DG CONNECT, to be held from November 6 to 8, 2013 in Vilnius, Lithuania). As a result, the European Commission has approved both the networking session and the exhibition booth, enabling the project to actively participate in this major event on ICT research and innovation.

In M11, the project participated in the "Open Access" event organised by the European Commission in Brussels. Following this event, an Open Access statement from the pro-iBiosphere consortium was submitted to the European Commission and is available on the project website ([Ref. 85](#)).

The project actively supported the Biodiversity Informatics Horizons Conference ([Ref. 87](#)) as a sponsor and participant.

In total, project partners participated in 13 events over the first project period (Sep. 2012 to Aug. 2013). 3 of these events took place overseas (China, Taiwan, the United States) and 3 events were organised by the European Commission. In the case of non-European Commission related events, project partners disseminated information about the project by delivering a presentation (event agendas and presentations are available in the internal library of the project website). Detailed information on partners' contributions is reported in D5.2.1.

Contribution to articles. In M1, the task leader listed the target scientific journals in which project articles and other materials can be published (see above). This list has been updated on a regular basis during the reporting period by all project partners. As for the events , a specific form has been drafted and placed on the website to enable partners to publish a news item on a released article (on [Ref. 88](#); in order to access the link, a login account is required).

In in the first project period, the project published two articles:

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 59 of 85

- "Implementation of TaxPub, an NLM DTD extension for domain-specific mark-up in taxonomy, from the experience of a biodiversity publisher " in the National Center for Biotechnology Information (NCBI) website on Ref. 89 (full text available in Annex 23).
- "An appraisal of megascience platforms for biodiversity information " in MycoKeys Journal on Ref. 90 (full text available in Annex 24).

Complementary information on these articles and on the list of target scientific journals is available in D5.2.1.

All articles published by the project are available on the pro-iBiosphere website (Ref. 9) and wiki (Ref. 91).

Social Media. The social media accounts of the project (i.e. Twitter, LinkedIn, Google + and Facebook) have been created at M2 and are linked with the project website, where links to these different social media appear as well..

From M7 onwards, the WP5 leader has taken the lead of the different project social media accounts and prepared a "Social Media Action Plan" in order to analyse the status of social networks to-date and to decide upon the specific actions that need to be undertaken to improve pro-iBiosphere presence on these social media. This strategy document was shared with all partners at M8 and placed on the project website and wiki (link), it has also been integrated into D5.2.1.

Following the release of the Social Media Action Plan, further promotional efforts have been provided by project partners from M8:

- A promotional email was sent to all partners to promote the project's Social Media presences to their contacts
- A postcard has been designed by the WP5 leader to promote the different project groups on social media to be distributed on the occasion of project workshops and other events (Ref. 92)
- Specific Twitter hashtags have been systematically used during project workshops to enable participants to tweet and share their comments
- A Twitter stream functionality has been added to the project website (Ref. 77)

Helpdesk services to the project community have been ensured through the use of LinkedIn and enabled the project, for instance to get in contact with new stakeholders.

For complementary information on the Social Media Action Plan and the different promotional efforts displayed in the period, see D5.1.2 and the wiki page (Ref. 93).

Calendar. In M9, a Google Calendar was created to monitor and ensure the contribution of partners in dissemination activities. The different project deliverables, milestones and prospective events have been added to this calendar to have a clear vision of project activities and outcomes and to enable the project

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 60 of 85

coordinator and the WP5 leader to encourage partners to contribute to post news on the website and discussions on the social media.

In M10, a Skype meeting was organised between the WP5 leader and the project coordinator to discuss the Google Calendar functionalities and usage. Following this meeting, a calendar file was shared with all partners to be integrated into their own calendar.

Dissemination database. In M9 the WP5 leader, with the help of project partners, consolidated all project contacts:

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

Then, a mailing list was created containing all these 350 contacts: dissemination@pro-ibiosphere.eu. This mailing list has been used to disseminate the second eNewsletter in M8 and will be used as a major dissemination channel to communicate with stakeholders regarding project activities, progress, events, outcomes, a.o.

Other dissemination activities. The project has liaised with the Advisory Board members to develop synergies and cross-promotional activities, for instance, by encouraging them to contribute to project documents (e.g., list of events, list of other initiatives available on the project wiki).

Deliverables and milestones. During the reporting period, the following deliverables and milestones have been submitted/completed:

- MS16: First report on articles and contributions to events (part of deliverable D5.2.1) - M12
- D5.2.1: Report on dissemination, communication and public awareness (on Ref. 94) - M12

Task 5.3 Stakeholder engagement and communication (lead: SIGMA, participants: all partners. Start: M18, End M22).

The task aims at presenting the outputs of the project and its sustainability perspectives to a targeted audience during the project final event. T5.3 will start in M18.

Further developments will be reported in the next reporting period.

Significant results

Main results achieved so far are:

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■ 7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 61 of 85

Task 5.1
1. Development of project graphical identity (logo and templates)
2. Development of Dissemination and communication implementation plan (dissemination strategy)
3. Development of Website (including an internal communication platform)
4. Development of Wiki
5. Release of 3 eNewsletters
6. Publication of 4 press releases
7. Elaboration of registration and questionnaire templates for meetings

Task 5.2
1. Elaboration of Outreach materials (pilot posters) and tools (dissemination database)
2. Development of Social media Action Plan and corresponding actions
3. Links developed with stakeholders and helpdesk services
4. Contribution to 13 events and 2 articles

Reasons for deviations

Not applicable. There are no deviations on this Workpackage.

Reasons for failing to achieve critical objectives

Not applicable. All objectives have been successfully achieved.

Statement on the use of resources

No deviations to date.

Corrective actions taken

Not applicable

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 62 of 85

Workpackage 6 Sustainability planning

Summary of progress towards objectives

Partner number	Partner short name	WP 6 Person-months (PMs) per partner[7]	Effort (PMs Year 1)
1	Naturalis	9.00	3.68
2	NBGB	1.00	0.50
3	FUB-BGBM	3.00	0.04
4	Pensoft	3.00	0.17
5	SIGMA	3.50	2.80
6	RBGK	14.00	1.40
7	Plazi	3.00	0.20
8	MFN	1.00	0

The tasks of Workpackage 6 have been implemented successfully. The summary of progress towards objectives is described below.

Objective 1. Ascertain the costs of supplying services. At the T6.1 May workshop (MS19) the participants developed a range of use-cases based on their own experiences. From these we were able to identify and document activity based costing across a broad spectrum of different activities and reached agreement on which metrics to use. The workshop outcomes were checked in consultation with external expertise (eg. eFiscal project) after which we proceeded to distribute template spreadsheets for partners to enter cost information for their activities (this work is ongoing). The established method and the first returns of baseline costs are on track for publication in deliverable D6.1.1 at the end of November 2013.

Objective 2. Ascertain the benefits of these services a) users or clients and b) suppliers.The main work on this objective has been the preparation work for the T6.2 workshop (MS21) on "user engagement and benefits" in October 2013. This comprised identifying and ensuring the participation of what might be termed "real users" (particularly those who benefit from use of biodiversity information who are not taxonomists or who are otherwise involved in it's production) and the development of the agenda and workshop activities. Outcomes of the workshop will be reported in the Y2 report.

Objective 3. Identify business models in use within and without our community. T6.3 has produced 2 deliverables towards this objective. D6.3.1 summarises information gathered via a questionnaire which was circulated to partners. It presents partners exploitation plans, a first analysis of their current business models, and information on the market background.
D6.3.2 is the first six monthly update of this information. It includes two possible business models at the pro-iBiosphere enterprise level for discussion at the T6.3 workshop "to evaluate business models in

---

[7] Based on Annex 1

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 63 of 85

currently in use by partners and relevant non-partners" (MS22). Other preparation completed for the October 2013 workshop are the participant list and agenda.

Objective 4. Evaluate these models and do a SWOT analysis, arriving at 1) conclusions about feasibility and ii) recommendations as to model (or models) to follow. Evaluation will take place in Y2 based on the CANVAS (and STOF) BM frameworks.

Details for each task

Task 6.1 Measuring and Constraining the costs of delivering services (lead: Naturalis; participants: FUB-BGBM, PENSOFT, RBGK, Plazi. Start: M6, End M21).

Task 6.1 explores questions associated with how the services being provided within the taxonomic community can be sustained beyond the end of the project. It will consider, for example, who will host the EDIT Platform Portal and Plazi Repositories, how they are deployed, who will check that they are running, who will check that they are secure, and how they are accounted and charged for. Furthermore, it will address how to sustain useful workflows and pipelines that are needed for a future implementation of the Open Biodiversity Knowledge Management System (OBKMS). Examples of such workflows are: (1) Aggregation of marked up legacy and prospectively published literature into a common place, such as CDM and Plazi servers; (2) Automated XML query into electronic registers of biological names; (3) Automated dissemination of atomized content (taxon names, taxon treatments, localities, bibliographic references, etc.) to indexing services and end users; and; (4) Automated export of treatments from databases to peer-reviewed publishing platforms, such as the Biodiversity Data journal established within the ViBRANT FP7 project. pro-iBiosphere will analyze the already established workflows and pipelines from the viewpoint of their current maintenance and suitability model and will develop strategies towards increased automation, cost-sharing and increased interoperability, which shall lead to reducing maintenance costs in the future. RBGK participation will ensure a good fit between Task 6.1 and the requirements of business model analyses in Task 6.4.

Workshops. The Workshop on "measuring and constraining the costs of delivering services" organised in May 2013 was attended by various participants, including Biota producers and publishers, data resources managers and providers of technology infrastructure and services. The input of the workshop will be included in the follow up of two reports of Task 6.1 in the Year 2 project.

In order to (i) understand what costs are involved in providing Biota information and services, (ii) how to measure the baseline costs for comparison with new business models; and (iii) to document approaches to sustainability within the wider biodiversity information community, a workshop on "Measuring and constraining the costs of delivering services" was organised on Wednesday the 22nd of May 2013 in Berlin. A total of 50 persons participated in the workshop. The aim of the workshop was to develop strategies towards increased automation, cost-sharing and increased interoperability which shall lead to reducing

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 64 of 85

maintenance costs in the future (e.g. for writing software, for sustainable delivery of content). The objectives of the workshop were, to: (i) document what are important services that are or need to be provided within the taxonomic community, and (ii) gain a better understanding of how these services can contribute to the reduction of costs (e.g. for mark-up and delivery of legacy biodiversity information, the provision of identification tools, and the production of the "New" floras and faunas, a.o.).

During the workshop, participants were divided into three break-out groups: (i) Biota producers and publishers; (ii) Data resources managers; (iii) Technology infrastructure and services. The workshop consisted of four phases:

- Phase one – Where are costs incurred? In order to document where costs arise in the workflows of Biota producers and service providers, each of the break-out groups was split into small groups, each to describe a particular workflow (e.g. taxonomic writing, desk editing, publication, software and service development, etc.).
- Phase two – How can the costs be measured? In order to decide how the costs identified in the previous exercise can be measured, each of the break-out groups brainstormed measurement methods for the costs they identified in the small groups and then they evaluated their feasibility. The main outcome of this exercise was a list of metrics that could be used against the costs identified above for each use case. A selection of the best (most practical) metrics was made with the help of all participants.
- Phase three – Constraining the costs of delivering information and services. During this exercise, participants brainstormed a wide range of possible cost-saving measures. Subsequently, their palatability with each stakeholder group was tested. The outcome of this phase was a list of possible cost-constraining measures, grouped by category and then sorted in order of preference within each category. Participants then gave short presentations to introduce potential ways of constraining costs.
- Phase four – General discussion. In order to document provider experiences, the business models they currently use and their approaches to sustainability, participants shared their experiences through short lightning talks. The open discussion helped to draw out views on the key points on costs, sustainability, sharing and layers of services.

Consortium members have contributed to these discussions, facilitating the targets of this task. Main outcomes of the workshop are to provide recommendations on (i) how the services being provided within the taxonomic community can be sustained on the long term (e.g. by sharing of resources for computing, storage, etc.) and (ii) how to sustain useful workflows and pipelines that are needed for a future implementation of the Open Biodiversity Knowledge System.

In order to obtain a better understanding of the activities of the taxonomic community and the costs on these activities, correspondences to follow up on the costs analysis among participants and the consortium are ongoing. Desktop research will also be performed during the months after the workshop. The information obtained from these exercises will be consolidated and the costs will be gathered in collaboration with the consortium and participants. The information will provide a solid basis on which to build the report D6.1.1 (due November 2013).

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 65 of 85

Task 6.2 Identifying and measuring the benefits of delivering services (lead: RBGK; participants: Naturalis, FUB-BGBM, PENSOFT, NBGB, Plazi. Start: M7, End: M18).

Task 6.2 explores the benefits to both users and clients. It can be viewed as a complementary exercise to Task 6.1 which explores the costs, and the method to be employed for estimating them. Both tasks feed into the work on sustainability covered in Task T6.4.

Task 6.2 activities started in M7 with background research at RBG Kew into the various business modelling approaches that are available. These should serve as a descriptive frameworks for the benefits information that will be collected in this task, together with the cost information from T6.1. Ideally these should further provide analytical frameworks for model evaluation (T6.3) and sustainability analysis (T6.4). Two leading approaches were investigated, CANVAS (A. Osterwalder & Y. Pigneur (2010) "The Business Model Generation" John Wiley and Sons (see [Ref. 114]) and STOF (H. Bouwman et al 2008 "Conceptualising the STOF model". Springer), together with the possibility of mapping between the two. CANVAS would look to be applicable in the early stages (especially as a brainstorming tool) and gives easily accessible displays of results of SWOT type analyses. STOF on the other hand explicitly includes organisational level detail and might offer benefits for risk analysis or stress testing. Mapping between the two methods is possible so both can be employed.

Concurrent with the above, efforts were underway to identify stakeholder groups for participation in the "user engagements and benefits" workshop (meeting 4, on October 9th 2013, Berlin). Following a 1 day face to face planning meeting with Sigma on July 4th 2013 in Nice, France, a list of 17 non-project candidates for participation in the two WP6 workshops was drawn up (WS3, [Ref. 96] and WS4, [Ref. 97], respectively), together with an agenda ([Ref. 95]). The Task delivered the workshop (M14) and D.6.2.1 "report on benefits to users" (M15) and D6.2.2 "report on benefits to suppliers" (M18), both of which feed in to Task 6.3 and Task 6.4. The recruitment process for a project assistant was initiated in July 2013 and the post was widely advertised; internally at Kew, on national recruitment websites and in the national press. The adverts ran throughout the month of August., nearly 70 applications were received by the closing date at the end of the month. This project assistant post will provide support at RBG Kew as the planned WP6 activities increase from the beginning of the second year, and this support will be essential to help avoid deviations from project objectives.

Task 6.3 Evaluating business models currently in use by partners (lead: SIGMA; participants: all partners. Start: M2, End M21).

In M1, a suggested methodology for elaborating exploitation plans was shared by the task leader. This methodology has been revised following partners' comments and in collaboration with WP6 leader to correspond to project objectives and partner profiles ([Ref. 98]).

In M2, following a Skype meeting held between WP6 leader, the task leader and the project coordinator, a questionnaire has been designed to guide partners towards describing their envisioned exploitation plans.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 66 of 85

In M4, an office meeting has been held between the task leader and the WP6 leader in London to establish the foundations for WP6 activities and to "clarify relationships between tasks 6.3 and 6.4". As a result of this meeting, the final draft of the questionnaire on "existing business plans currently in use by partners" was designed. A meeting report (MS18) detailing the main conclusions of this meeting along with the methodology and milestones is available on the project wiki (Ref. 99).

This questionnaire has been sent to partners at M4 to gather their inputs on three aspects: (i) exploitation plans at the level of partners' organisation; (ii) business models currently in use in partners' organisation; and (iii) towards the sustainability of the project (Ref. 100).

To complement the findings of the questionnaire, a desktop research with information on the different biodiversity platforms has been initiated by the task leader and shared with project partners to gather their inputs. Partners' inputs from the questionnaire and on the desktop research have been inserted into the first deliverable D6.3.1. submitted at M6.

In M10, so as to prepare the update of the second deliverable (D6.3.2), the task leader and the WP6 leader jointly decided that an office meeting should be held at the Sigma Orionis headquarters in Sophia Antipolis, France. An online meeting was then organised between them to discuss the preparations for this office meeting; the minutes of this meeting are available on the pro-iBiosphere wiki (Ref. 101).

In M11, both project partners met in Sophia Antipolis to work on (i) Task 6.3 activities, (ii) the release of the deliverable D6.3.2, and (iii) the organisation of the "Meeting to evaluate business models currently in use by partners and relevant non-partners" (MS22) held in October 10, 2013 in Berlin. Following this meeting, a revised work plan for Task 6.3 has been prepared (Ref. 102). The workshop agenda and list of participants were made available on the wiki (Ref. 103), and external participants have been invited.

To gather additional inputs from partners for the release of D6.3.2., new tables have been created and shared on the wiki at M11 (Ref. 104), the desktop research has been updated at M11, and D6.3.2 has been submitted at M12.

Further developments will be reported in the next reporting period with the organisation of MS22 (e.g., final agenda, participants, event outcomes).

Deliverables and milestones
During the reporting period, the following deliverables and milestones have been submitted/completed:

- MS18: Meeting to clarify relationships between tasks 6.3 and 6.4 (Ref. 99 to report) - M4
- D6.3.1: "Report on diversity and strengths of existing business plans and discussion of sustainability (1)" (Ref. 24) - M6
- D6.3.2: "Report on diversity and strengths of existing business plans and discussion of sustainability (2)" (Ref. 25) - M12

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 67 of 85

Task 6.4 Towards Sustainability for Services (lead: RBGK, participants: Naturalis. Start: M15, End M24).

Task 6.4 is scheduled to commence in M15. Tasks 6.3 and 6.4 will provide an evaluation of business models incorporating the costs assessment from task 6.1 together with the benefits to a wider audience that will be identified in 6.2. The task will then identify how key business model activities and resources (ie. the workflows and systems) can be made sustainable in the future. The information gathered will be used to recommend a model(s) to follow. An inventory of the requirements and services (e.g.services that provide information to our community, services that deliver information to the broader public) will be assessed. This will allow us to address questions such as how we might reduce costs and maximise benefits both to the producers and the users of information held in Floras, Faunas and related products.

The need for the close coordination of the tasks within WP6 was discussed at the project kick-off meeting. Following these discussions, a meeting was arranged between Sigma and RBGK in December 2012 in London (RBG Kew's premises) dedicated to further clarify the relation between tasks 6.3 & 6.4 (Ref. 99). The RBG Kew project assistant will be in post from November 2013 onward to support this task.

Upcoming deliverables from Task 6.3 will form the basis of an analysis of partners' exploitation plans. At the same time, an early draft of alternative business models will be circulated among partners (produced from a summary of on-going work in Tasks 6.1, 6.2 and 6.3).

Significant results

Main results achieved include:

1. A method (including a descriptive and analytical framework) and a work scheme for exploitation plans and business models
2. Agreement on the methods to be used for measuring baseline costs
3. Tools to gather partners' inputs
4. A market background document (partners' inputs consolidation and desktop research)

Reasons for deviations

MS20 (Task 6.2), consisting of a series of small scale meetings with the providers and the users of Flora, Fauna, and Mycota, to be take place at the end of June 2013, was re-scheduled to the second year of activities of the project.

The re-scheduling does not interfere with the WP6 deliverables that need to be submitted in the next months. There is in fact a benefit in having these meetings closer to the timing of the remainder of the task activities, in that the need and the practicalities can be much better assessed after the workshop on "user engagement and benefits" has taken place (in October 2013) and the results of the workshop analysed.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 68 of 85

Reasons for failing to achieve critical objectives

Not applicable. There are no deviations on this Workpackage.

Statement on the use of resources

Not applicable.

Corrective actions taken

Not applicable.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 69 of 85

3.2.3 Project management during the period

Consortium management tasks and achievements

The Consortium Management Committee has met 3 times in the first year, as anticipated. Minutes are available on the project wiki.

The Advisory Board has met once, at the beginning of the project, to validate the overall direction of the project and establish priorities. The board has not met since, because no substantive issues of conflict or deviations from the initial strategy have been identified. Individual board members have been frequently consulted on an ad-hoc basis about issues relevant to their area of expertise. A second meeting with the Board will take place in Y2.

Communication between beneficiaries. The wiki has provided an excellent platform for communicating, exchanging ideas and fostering discussions. Besides the wiki, communication also takes place on a regular basis via the email and online meetings.

In order to organise joint meetings and workshops, interaction among the various partners is a pre-requisite.

For instance, the topic on "best practices for stable URIs" was addressed by workpackages 2, 3 and 4 during the pro-iBiosphere workshops that took place in February 2013, WP2 in May 2013, and WP4 in October 2013. These first meetings helped to bring the community together to identify activities that needed to be tackled in order to achieve interoperability of biodiversity data. Subsequent meetings helped to propose suggestions in order to address those needs. As a result of this joint pro-iBiosphere and CETAF effort, the adoption of semantic web and linked open data technology is increasing. At present (10/2013), the technology is being adopted by 7 institutions (i.e.: pro-iBiosphere institutions: Plazi, FUB-BGBM, MfN, RBGK, NBGB and two non pro-iBiosphere institutions: Royal Botanic Gardens Edinburgh - RBGE and Museum National d' Histoire Naturelle, Paris - MNHN).

WP2 and WP6 are working together to enhance understanding on how stakeholders cooperate, generate and exchange data, and to make sure that stakeholder needs are taken into account prior to the design of the envisaged research infrastructure (the Open Biodiversity Knowledge Management System - OBKMS). As well as measuring baseline costs of partners current activities, WP6 will also need to understand how changes to current activities suggested by outcomes from WP3 and WP4 may impact costs, either as a route to constraining them, or else indicating where additional costs might be incurred.

WP5 is working together with all workpackages to produce the e-newsletter, for instance by contacting participants of meetings that could post relevant contributions.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 70 of 85

Understanding the technical and social requirements for IT infrastructure is critical to ensure that the beneficiaries, users, funders and developers are satisfied with the OBKMS. On the technical side, work packages 3 and 4, are examining standards and methods for connecting and mobilizing data, whereas, work package 2 and 5 address the social aspect of the venture. Work package 2 has examined the requirements that such an IT infrastructure will need to fulfill, whereas work package 5 has promoted the idea of an open infrastructure for biodiversity infrastructure to the scientific community. There has been considerable overlap between work packages 2 and 5, for example the participants in our workshops have provided much of the material we have been able to use for our newsletters.

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

There were three problems faces in Y1:

- Delay in the recruitment of support staff (i.e. project assistants) to provide support to the various institutions that are responsible for a large number of deliverables and milestones in the project, such as Naturalis, RBGK and Plazi. As a result of this, the various task leaders involved had a very high workload. In the case of Naturalis, a project assistant was hired 6 months ago. RBGK has found a suitable candidate (Charlotte Johans) who will start working on the project in November 2013. Dr. David Patterson has joined Plazi to provide support in writing scientific deliverables. However, a part-time assistant who can help with the review and timely submission of the non-scientific information (like the quarterly and year reports) still needs to be appointed.

- A problem faced by the consortium at the start of the project was the retention of skilled and expert staff in crucial positions, as was the case in Task 3.3, with the departure of Dr. Henning Scholz from the MfN in December 2012. Following his departure, Dr. Jana Hoffmann was assigned by the MfN to take over responsibility for the MfN involvement in the project activities. In June 2013, Dr. Gregor Hagedorn joined the MfN as head of the Digital World at the MfN. His association with Plazi remains; he does not charge Person months on behalf of the MfN towards pro-iBiosphere. Responsibilities for the project after the start of maternity leave of Jana Hoffmann starting in August had to be shifted; it was decided that Dr. Gregor Hagedorn will take over general responsibilities for pro-iBiosphere at MfN, with the specific scientific work being carried out by newly hired staff. In June 2013, the MfN started the hiring process for a position (6 PM, 100%). Interviews for the open position were held on the 3rd of July 2013. As a result of this, Dr. Daniel Mietchen has been hired for the project for the next 6 months. He is familiar with the project and actively contributed to some other tasks in pro-iBiosphere. A hand-over meeting took place on the 15ᵗʰ of August 2013. Dr. Mietchen is now focusing on preparing the Task 3.3 workshop (MS12) in February 2014 and the two related deliverables.

- Two work packages functioned throughout year one at the task level. All leaders involved in these work packages are very skilled, efficient and motivated, and hence, milestones and deliverables were achieved and timely submitted, respectively. However when work package reports needed to be made or decisions to be taken (at the WP level), the task leader approach did not work well. During the last management meeting that took place in October 2013, it was decided Quentin Groom (NBGB) and Anton Güntsch (FUB-BGBM) will be the new WP2 and WP3 leaders.

pro-iBiosphere FP7 Project ∎ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ∎7th Framework Programme
∎ Coordination and support action FP7-INFRASTRUCTURES-2012-1∎ Subprogram area INFRA-2012-3.3

Page 71 of 85

Changes in the consortium

No changes to report.

List of project meetings, dates and venues

A complete and detailed list of all meetings that have been organised of the project is available on the project wiki, on the following pages:

- http://wiki.pro-ibiosphere.eu/wiki/Meetings
- http://wiki.pro-ibiosphere.eu/wiki/Pro-iBiosphere_Milestones

In addition to these meetings, the following office meetings to clarify relationships between tasks 6.3 and 6.4 were conducted:

- Office meeting to clarify relationship between tasks 6.3 and 6.4 on December 7, 2012 in London (to view the report, see Ref. 99)
- Office meeting to prepare the release of D6.3.2 and to discuss the Workshop to evaluate business models in use by partners and relevant non-partners on July 4, 2013 in Sophia Antipolis, France (to view the work plan and report, see Ref. 102)

Project planning and status

The project is functioning as planned and is currently on schedule and on budget. See Section 3.3 for complete list of deliverables and milestones, including delivery date from Annex I, and actual/forecast delivery date.

Impact of possible deviations from the planned milestones and deliverables

The deviations of MS20 has no further impact on the envisaged activities and deliverables that will take place in Y2 of the project.

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

No changes to report.

Development of the Project website

The pro-iBiosphere website was created to meet two major needs of the project: internal communication within the consortium and external communication and dissemination of the projects objectives and results. It was developed by Pensoft's IT team and Plazi in close collaboration with Naturalis, using a Content Management System and a Wiki.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 72 of 85

- An Internal Communication Platform (ICP) was integrated with the project's public website to provide a medium for communication among project participants. The ICP allows the exchange of various types of information: datasets, results, coordination decisions, timetables, presentations, materials, and reporting among partners. It allows each partner, the Workpackages leaders, and the coordinator to regularly monitor progress in data collation, analysis, and accomplished deliverables. The ICP is also used as an internal discussion forum for items that may emerge within Workpackages between the main project meetings and which need live discussions for rapid decisions.

- The public website is designed to act as an information hub about the objectives, activities and results of pro-iBiosphere and serves as a prime public dissemination tool making available the project deliverables and the published materials. The news and events deriving from and organised by the consortium are also announced through the website. The website is being updated on a regular basis to keep the audience informed and ensure continued interest of already attracted visitors.

The Wiki is especially important because it is both public and editable. Many collaborations and discussions in the project require the involvement of external partners; for which the wiki is the prime communication platform.

Short comments and information on co-ordination activities

Possible cooperation with other programmes.

LifeWatch. In the first semester of 2013, various online meetings with LifeWatch took place with the purpose of establishing synergies with pro-iBiosphere. A joint document was written in order to establish areas of mutual interest, and discuss the next steps for a future collaboration. After these fruitful conversations, LifeWatch entered in a busy phase, organising two meetings with the main governing bodies of the infrastructure. In July 2013 LifeWatch informed pro-iBiosphere that the outcomes of these governing bodies meetings might provide new perspectives and a better path to continue thinking about a future cooperation.

EU BON. With the purpose of fostering cooperation with EU BON, representatives of the project (Hannu Saarenmaa, Finland, Florian Wetzel, Germany) participated in the pro-iBiosphere coordination workshop that took place in February 2013 (organised by Plazi, Task 2.1 - Coordination and routes for cooperation across organisations, projects and e-infrastructures)..

The Plazi system used in WP4 for the pilots will similarly be used in the EU-BON project with the goal to provide observation data needed for the biodiversity modelling activities by semantic enhancing of publications covering selected. Coordination is especially assured through Plazi and Pensoft, both partners in the EU-BON consortium (Plazi Task 3.4 lead).

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 73 of 85

ViBRANT. Two of the pro-iBiosphere partners are involved also in the FP7 project ViBRANT (http://vbrant.eu), i.e. Pensoft and FUB-BGBM. ViBRANT's coordinator and project manager Vincent Smith and David Roberts took part in the pro-iBiosphere workshop held on 22 - 25th of May 2013, in Leiden, the Netherlands. The future collaboration on building up the OBKMS has been discussed. The Biodiversity Data Journal established within ViBRANT will be used as a publication venue for some of the pilots of pro-iBiosphere, and especially for the *Eupolybothrus* cybertaxonomic checklist. The Bibliography of Life (http://biblife.org) and its two main modules RefBank and ReFinder, will be used as a reference handling tool in OBKMS. ViBRANT and pro-iBiosphere agreed to cooperate in HORIZON 2020 bioinformatics intitiatives. Darwin Core Archive interchange format developed for Scratchpads will be tested and used by the pro-iBiosphere partners Plazi, FUB-BGBM and Pensoft as a data sharing tool for data mined from legacy and modern literature.

BioVel. The EU project BioVel deals with workflows and webservices provided by different players in the biodiversity domain. BioVel and pro-iBiosphere agreed to (i) register the workflows and associated web servies developed in pro-iBiosphere (i.e. mark-up workflows and taxon name registration workflow) in the Biodiversity Catalogue and (ii) to link these to the appropriate existing BioVel services.

GBIF. All primary biodiversity data gathered as a result of the markup and data mining are exported to GBIF for further indexing and re-use. To facilitate the process, a TAPIR service has been created allowing GBIF to harvest materials observation data from Plazi (for data provided see Ref. 113). In the second period of the project, on the occasion of the pro-iBiosphere meeting that will take place in February 2013 in Berlin, a change from the Tapir service to the use of Darwin-Core Archive will be discussed. This will use the same technology for the transfer of treatments from Pensoft and Plazi to EOL.

Encyclopedia of Life (EOL). The Darwin Core Archive will be used by Plazi and Pensoft to model and export taxon treatment data into Encyclopedia of Life. The Darwin Core Archive is an important interchange format, which makes primary data largely interoperable. The technical implementation and constraints were discussed with the EOL representative Dr Katja Schultz at the pro-iBiosphere meeting in Berlin in October 2013.

Biological nomenclature registries. pro-iBiosphere 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 (pro-iBiosphere pilot 2).

Biodiversity Heritage Library. pro-iBiosphere will use the PDF scans provided by BHL to elaborate and test technologies used for markup of legacy data. Discussion on feeding back documents that have been retrieved from BHL and have been semantically enhanced is ongoing.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 74 of 85

3.3 Deliverables and Milestones tables

Deliverables

Deliverable number	Deliverable title	Version	WP number	Lead beneficiary number	Nature	Dissemination level	Delivery Month	Actual / Forecast delivery date Dd/mm/yy	Status Not submitted/Submitted
D1.1	Project Quality Assessment Plan	Final	1	1	R	PU	1	30/09/2012	Submitted
D1.2.1	Management report 1	Final	1	1	R	PU	3	30/11/2012	Submitted
D1.2.2	Management report 2	Final	1	1	R	PU	6	28/02/2013	Submitted
D1.2.3	Management report 3	Final	1	1	R	PU	9	31/05/2013	Submitted
D1.2.4	Management report 4	Final	1	1	R	PU	12	30/08/2013	Submitted
D1.2.5	Management report 5		1	1	R	PU	15	11/2013	
D1.2.6	Management report 6		1	1	R	PU	18	02/2014	
D1.2.7	Management report 7		1	1	R	PU	21	05/2014	
D1.2.8	Management report 8		1	1	R	PU	24	08/2014	
D2.1.1	Report on ongoing biodiversity related projects, current e-infrastructures and standards	Final	2	7	R	PU	10	30/06/2013	Submitted
D2.1.2	Draft strategy for increased cooperation	Final	2	7	R	PU	13	30/09/2013	Submitted
D2.2	Report on user feedback	Final	2	6	R	PU	12	31/08/2013	Submitted

Deliverable number	Deliverable title	Version	WP number	Lead beneficiary number	Nature	Dissemination level	Delivery Month	Actual / Forecast delivery date Dd/mm/yy	Status Not submitted/Submitted
D2.3	Proof of concept reports on the use of e-tools	Final	2	2	R	PU	12	30/08/2013	Submitted
D2.4.1	Drafts policy on Open Access for data and information (1)	Final	2	7	R	PU	12	30/08/2013	Submitted
D2.4.2	Drafts policy on Open Access for data and information (2)		2	7	R	PU	23	07/2014	
D3.1	Best Practices Guide on editorial policies	Final	3	1	R	PU	9	31/05/2013	Submitted
D3.2.1	Concept paper for involvement of individual experts, commercial vendors, and citizen scientists	Final	3	7	R	PU	9	31/05/2013	Submitted
D3.2.2	Report on the state and quality of biosystematics documents and survey reports	Final	3	7	R	PU	12	31/08/2013	Submitted
D3.3.1	Report on state-of the art and research horizons of semantic integration of biodiversity literature		3	8	R	PU	16	12/2013	
D3.3.2	Report on progress during the coordination process of partners and non consortium partners		3	8	R	PU	20	04/2014	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 76 of 85

Deliverable number	Deliverable title	Version	WP number	Lead beneficiary number	Nature	Dissemination level	Delivery Month	Actual / Forecast delivery date Dd/mm/yy	Status Not submitted/Submitted
D4.1	Report on strategies for improved cooperation and interoperability between infrastructures		4	3	R	PU	16	12/2013	
D4.2	Report on a strategy for improvement & interoperability of the XML schemas		4	7	R	PU	16	12/2013	
D5.1.1	Web-based platform and wiki	Final	5	4	O	PU	4	15/03/2013	Submitted
D5.1.2	Electronic newsletters 4	Final	5	5.25	PU	4		22/12/2012; 24/04/2013; 30/08/2013	Submitted
D5.2.1	Report on dissemination, communication and public awareness (1)	Final	5	5	R	PU	12	31/08/2013	Submitted
D5.2.2	Report on dissemination, communication and public awareness (2)		5	5	R	PU	24	08/2014	
D6.1.1	Report on cost delivery, efficiency and cost reduction through effective practices (1)		6	1	R	PU	15	11/2013	
D6.1.2	Report on cost delivery, efficiency and cost reduction through effective practices (2)		6	1	R	PU	21	05/2014	
D6.2.1	Report on benefits to users/clients		6	6	R	PU	15	11/2013	
D6.2.2	Report on benefits to suppliers		6	6	R	PU	18	02/2014	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 77 of 85

Deliverable number	Deliverable title	Version	WP number	Lead beneficiary number	Nature	Dissemination level	Delivery Month	Actual / Forecast delivery date Dd/mm/yy	Status Not submitted/Submitted
D6.3.1	Report on diversity and strengths of existing business plans & discussion of sustainability	Final	6	5	R	PU	6	28/02/2013	submitted
D6.3.2	Report on diversity and strengths of existing business plans and discussion of sustainability (2)	Final	6	5	R	PU	12	31/08/2013	submitted
D6.3.3	Report on diversity and strengths of existing business plans and discussion of sustainability (3)		6	5	R	PU	18	02/2014	
D6.3.4	Report on diversity and strengths of existing business plans and discussion of sustainability (4)		6	5	R	PU	21	05/2014	
D6.4.1	Risk analysis of the different models or a mixed model		6	6	R	PU	16	12/2013	
D6.4.2	Draft summary of sustainability models for consultation		6	6	R	PU	17	01/2014	
D6.4.3	Alternative business requirements and scenarios for a sustainable Open Biodiversity Knowledge System		6	6	R	PU	23	07/2014	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 78 of 85

Deliverable number	Deliverable title	Version	WP number	Lead beneficiary number	Nature	Dissemination level	Delivery Month	Actual / Forecast delivery date Dd/mm/yy	Status Not submitted/Submitted
D6.4.4	Recommendations to policy makers with regard to achieving sustainable delivery of biodiversity		6	6	R	PU	24	08/2014	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 79 of 85

Milestones

Milestone number	Milestone name	Workpackage no	Lead benificary number	Delivery date form annex I	Achieved yes/no	Actual / Forecast achievement date dd/mm/yyyy	Comments
MS1	Kick-Off meeting with Pro-iBiosphere consortium and stakeholders	WP1	1	1	yes	27-28/09/2012	
MS2	Mid-term review	WP1	1	13		19/11/2013	
MS3	Project closure	WP1	1	24		08/2014	
MS4	Initial contact meetings and requirements gathering	WP2	6	5	yes	25/01/2013	
MS5	Workshop on development and use of e-tools	WP2	2	6	yes	11-12/02/2013	
MS6	Workshop on coordination & routes for cooperation across organizations, projects & e-infrastructures	WP2	7	9	yes	23/05/2013	
MS7	Workshop on stakeholder requirements	WP2	6	9	yes	21/05/2013	
MS8	Follow-up (small) meetings on selected topics with the relevant stakeholders	WP2	6	13	no	09/2013	MS 8 was not necessary in 09/2013; Kew will decide in the course of Year 2 if MS8 is necessary
MS9	Review process of Open Access Policy closed	WP2	7	20		04/2014	
MS10	Workshop on data curation and acquisition of Floras and Faunas	WP3	1	6	yes	14/02/2013	
MS11	Workshop on semantic mark-up generation, data	WP3	7	6	yes	13/02/2013	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 80 of 85

Milestone number	Milestone name	Workpackage no	Lead benificary number	Delivery date form annex I	Achieved yes/no	Actual / Forecast achievement date dd/mm/yyyy	Comments
	quality and user-participation infrastructure						
MS12	Workshop on mark-up of biodiversity literature	WP3	8	18		10-11/02/2014	
MS13	Workshop on how to improve technical cooperation and interoperability at the e-infrastructure level	WP4	3	14	yes	8/10/2013	
MS14	Workshop on how to promote the development & adoption of common mark-up standards & interoperability	WP4	7	14	yes	8/10/2013	
MS15	First set of promotional materials	WP5	4	4	yes	1/09/2012; 31/10/2012;18/12/2012	
MS16	First report on articles and contributions to events	WP5	5	12	yes	07/2013	
MS17	Final project event	WP5	5	22		12/06/2014	
MS18	Office meetings to clarify relationships between tasks 6.3 and 6.4	WP6	5	5	yes	01/2013	
MS19	Workshop on measuring and constraining the costs of delivering services	WP6	1	9	yes	22/05/ 2013	
MS20	Meetings with clients/users - benefits	WP6	6	10	yes	08/2013	
MS21	Workshop on user engagement and benefits	WP6	6	14	yes	9/10/2013	
MS22	Meeting to evaluate business models currently in use by partners and relevant non-partners	WP6	6	14		10/10/2013	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 81 of 85

Milestone number	Milestone name	Workpackage no	Lead benificary number	Delivery date form annex I	Achieved yes/no	Actual / Forecast achievement date dd/mm/yyyy	Comments
MS23	Workshop on alternative business models	WP6	6	18		11-12/02/ 2014	
MS24	Workshop on model evaluation	WP6	6	22		10-11/06/2014	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 82 of 85

3.4 References

1. http://wiki.pro-ibiosphere.eu/wiki/Pro-iBiosphere_meeting_notes_27_%26_28_September_2012
2. http://wiki.pro-ibiosphere.eu/wiki/Meeting_Notes_Workshops_Leiden_February_2013
3. http://wiki.pro-ibiosphere.eu/wiki/Workshops_Leiden_February_2013
4. http://wiki.pro-ibiosphere.eu/wiki/Workshop_Berlin_1:_Requirements_of_users_of_Flora,_Fauna_or_Mycota_publications_or_services
5. http://wiki.pro-ibiosphere.eu/wiki/Workshop_Berlin_2:_Measuring_and_constraining_the_costs_of_delivering_services
6. http://wiki.pro-ibiosphere.eu/wiki/Workshop_Berlin_3:_Coordination_and_routes_for_cooperation
7. http://wiki.pro-ibiosphere.eu/wiki/Pilots
8. http://www.pro-ibiosphere.eu/media/center/
9. http://www.pro-ibiosphere.eu/news/
10. https://docs.google.com/spreadsheet/ccc?key=0AgLFH-K5Y6_cdF9oalRPWE15VVhVYlctQUp5d3hBUUE#gid=0
11. http://wiki.pro-ibiosphere.eu/w/media/2/24/Pro-iBiosphere_WP2_NBGB_D2.3_VFFa_30082013.pdf
12. http://wiki.pro-ibiosphere.eu/wiki/Workshops_Leiden_February_2013#Workshop_.233:_Prospective_Literature_.E2.80.93_Toward_Best_Practices_for_data_acquisition_and_curation_using_e-tools_for_taxonomy
13. http://wiki.pro-ibiosphere.eu/w/media/4/42/Pro-iBiosphere_WP3_Naturalis_D3.1_VFF_31052013.pdf
14. http://wiki.pro-ibiosphere.eu/wiki/Workshops_Leiden_February_2013#Workshop_.232:_Legacy_literature_.E2.80.93_Semantic_mark-up_generation.2C_data_quality_and_user-participation_infrastructure
15. http://wiki.pro-ibiosphere.eu/w/media/c/cb/Pro-iBiosphere_WP3_Plazi_D3.2.1_VFF_31052013.pdf
16. http://wiki.pro-ibiosphere.eu/w/media/c/cc/Pro-iBiosphere_WP3_Plazi_D3.2.2_VFFa_31082013.pdf
17. http://wiki.pro-ibiosphere.eu/wiki/D3.3.1_Semantic_integration_biodiversity_literature
18. http://www.pensoft.net/journals/phytokeys/article/5556/lettowia-a-new-genus-of-vernonieae-from-east-africa-asteraceae-
19. http://wiki.pro-ibiosphere.eu/wiki/Mark-up_Roadmap
20. http://taxonx.org/documentation/taxonx1_xsd.html
21. http://wiki.pro-ibiosphere.eu/wiki/Pro-iBiosphere_Deliverables
22. http://wiki.pro-ibiosphere.eu/wiki/Pro-iBiosphere_Press_releases
23. http://wiki.pro-ibiosphere.eu/wiki/Partners%27_contribution_to_events
24. http://wiki.pro-ibiosphere.eu/w/media/6/6f/Pro-iBiosphere_WP6_Sigma_D6.3.1_VFF_28022013.pdf
25. http://wiki.pro-ibiosphere.eu/w/media/9/9f/Pro-iBiosphere_WP6_Sigma_D6.3.2_VFFa_31082013.pdf
26. http://wiki.pro-ibiosphere.eu/wiki/Report_on_cost_delivery,_efficiency_and_cost_reduction_through_effective_practices_(1)
27. http://wiki.pro-ibiosphere.eu/w/media/c/c9/Pro-iBiosphere_WP2_PLAZI_D2.1.1_VFF_30062013.pdf
28. http://wiki.pro-ibiosphere.eu/w/media/7/7f/Pro-iBiosphere_WP2_Plazi_D2.4.1_VFF_30082013.pdf
29. http://wiki.pro-ibiosphere.eu/w/media/a/a0/Pro-iBiosphere_WP2_Plazi_D2.1.2_VFF_30092013.pdf
30. www.pro-ibiosphere.eu
31. http://wiki.pro-ibiosphere.eu/wiki/Main_Page
32. http://www.intelligentediting.com/standardversion.aspx
33. https://pro-ibiosphere.eu-xpert.eu/
34. http://www.eufin-fp7.com
35. http://wiki.pro-ibiosphere.eu/wiki/Pro-iBiosphere_Milestones
36. https://docs.google.com/spreadsheet/viewform?formkey=dGRYQ2pMNlpLYkUyc19nRTdrbXpKdEE6MA
37. http://wiki.pro-ibiosphere.eu/wiki/Meeting_Notes_Advisory_Board_Meeting_Leiden_February_15,_2013
38. http://wiki.pro-ibiosphere.eu/wiki/Meeting_Notes_3rd_Consortium_Meeting_Berlin_May_24,_2013
39. http://wiki.pro-ibiosphere.eu/wiki/Meetings
40. http://www.pro-ibiosphere.eu/library
41. http://wiki.pro-ibiosphere.eu/w/media/6/6c/Pro-iBiosphere_WP2_RBGK_D2.2_VFFa_31082013.pdf
42. https://docs.google.com/forms/d/1vhD6xhq6C97OUqI1U2du6fcQPBmJoDhaPosdEPSqCB4/viewform
43. http://wiki.pro-ibiosphere.eu/w/media/2/24/MS4_report.docx
44. https://www.surveymonkey.com/s/eTools
45. http://semanticpublishing.wordpress.com/2013/02/26/ten-next-steps/
46. http://www.pro-ibiosphere.eu/news/0_9_2013#4395
47. http://www.pro-ibiosphere.eu/news/0_9_2013#4393
48. http://wiki.pro-ibiosphere.eu/wiki/February_Workshops_Participant_list

49. http://wiki.pro-ibiosphere.eu/wiki/21st_May_Workshop_Participant_list
50. http://wiki.pro-ibiosphere.eu/wiki/22nd_May_Workshop_Participant_list
51. http://wiki.pro-ibiosphere.eu/wiki/22nd_May_Workshop_Participant_list
52. http://wiki.pro-ibiosphere.eu/wiki/Report_on_strategies_for_improved_cooperation_and_interoperability_between_infrastructures
53. http://wiki.pro-ibiosphere.eu/wiki/Pilots#Pilot_3
54. http://wiki.pro-ibiosphere.eu/wiki/Pilots#Pilot_4
55. http://wiki.pro-ibiosphere.eu/wiki/Common_query/response_model
56. http://wiki.pro-ibiosphere.eu/w/media/e/ea/BioSyst_2013_Pensoft.pdf
57. http://wiki.pro-ibiosphere.eu/wiki/Report_on_progress_during_the_coordination_process_of_partners_and_non_consortium_partners
58. https://docs.google.com/document/d/1XPwsqJETEWISJPAaR4EMDzyHRWXdtsj4VM9qw5LTw08/edit#heading=h.hx5h2853quxq
59. http://dev.e-taxonomy.eu/dataportal/chenopodiumPilot/
60. http://wiki.pro-ibiosphere.eu/wiki/Workshops_Berlin,_October_2013
61. http://www.mapress.com/phytotaxa/content/2012/f/pt00073p030.pdf
62. http://wiki.pro-ibiosphere.eu/wiki/Workshop_Berlin_1:_How_to_improve_technical_cooperation_and_interoperability_at_the_e-infrastructure_level
63. http://www.pro-ibiosphere.eu/news/#4296
64. http://wiki.pro-ibiosphere.eu/w/media/9/95/Interoperability_model_between_PLAZI_and_the_EDIT_CDM_Platform.pdf
65. http://wiki.pro-ibiosphere.eu/wiki/Participants
66. http://wiki.pro-ibiosphere.eu/wiki/List_of_Workshops_in_Berlin_May_2013
67. http://wp5.e-taxonomy.eu/blog/cdm
68. http://wiki.pro-ibiosphere.eu/wiki/Workshops_Leiden_February_2013#Day_3:_13th_of_February_2013
69. https://drive.google.com/a/plazi.org/#folders/0B_yrQwn4yBySN1FhOHl6dTJqN3M
70. http://plazi.org:8080/GgSRS/search
71. http://conference.lifewatch.unisalento.it/index.php/EBIC/BIH2013
72. http://wiki.pro-ibiosphere.eu/wiki/Workshop_Berlin_2:_Promote_and_foster_the_development_%26_adoption_of_common_mark-up_standards_and_interoperability_between_schemas
73. http://wiki.pro-ibiosphere.eu/w/media/5/58/A_common%2Cautomated%2Cpre-publication_registration_model_for_higher_plants_%28International_Plant_Names_Index%2C_IPNI%29%2C_fungi_%28Index_Fungorum%2C_MycoBank%29_and_animals_%28ZooBank%29.pdf
74. http://wiki.pro-ibiosphere.eu/wiki/Best_practices_for_stable_URIs
75. http://wiki.pro-ibiosphere.eu/wiki/Workshops_Berlin,_May_2013
76. http://wiki.pro-ibiosphere.eu/wiki/Pro-iBiosphere_Posters
77. http://www.pro-ibiosphere.eu/
78. http://www.pro-ibiosphere.eu/media/center/3691
79. http://www.pro-ibiosphere.eu/media/center/3704
80. http://www.pro-ibiosphere.eu/media/center/4366
81. http://www.pro-ibiosphere.eu/getatt.php?filename=pro-iBiosphere_press_release_3694.docx
82. http://www.pro-ibiosphere.eu/media/center/3693
83. http://wiki.pro-ibiosphere.eu/w/media/c/cc/Pro-iBiosphere_D5_1_1_PENSOFT_VFF_13032013.pdf
84. http://www.pro-ibiosphere.eu/library/simpledoc/
85. http://www.pro-ibiosphere.eu/news/0_8_2013#4310
86. http://www.pro-ibiosphere.eu/getatt.php?filename=pro-iBiosphere_Posters_BIH2013_4343.pdf
87. http://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&ved=0CDUQFjAA&url=http%3A%2F%2Fconference.lifewatch.unisalento.it%2Findex.php%2FEBIC%2F&ei=sqlFUryoCMSX0AWc0oD4BA&usg=AFQjCNH5NU4Sllk7TlZjhXAZdc0FjOmVbQ&sig2=BHFMk9aGUDGJ-yZsk3OThw&bvm=bv.53217764,d.d2k
88. http://www.pro-ibiosphere.eu/library/adddocument/
89. http://www.ncbi.nlm.nih.gov
90. www.pensoft.net/journals/mycokeys\
91. http://wiki.pro-ibiosphere.eu/wiki/Partners%27_contribution_to_articles
92. http://www.pro-ibiosphere.eu/media/center/4210
93. http://wiki.pro-ibiosphere.eu/wiki/Partners%27_contributions_to_WP5
94. http://wiki.pro-ibiosphere.eu/w/media/f/f0/Pro-iBiosphere_WP5_Sigma_D5.2.1_annex_VFFa_31082013.pdf
95. http://wiki.pro-ibiosphere.eu/wiki/Workshop_Berlin_3:_User_Engagement_and_Benefits
96. http://wiki.pro-ibiosphere.eu/wiki/ParticipantListWS3

---

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 84 of 85

97. http://wiki.pro-ibiosphere.eu/wiki/Oct10_Workshop_Participants_list
98. http://wiki.pro-ibiosphere.eu/w/media/d/d7/Pro-iBiosphere_WP6_Task63_Sigma_V01_09112012-2.pdf
99. http://wiki.pro-ibiosphere.eu/w/media/2/24/Pro-iBiosphere_WP6_SIG_R_MS18_V1.2_20122012.pdf
100. http://wiki.pro-ibiosphere.eu/w/media/8/88/Pro-iBiosphere_WP6_SIG_ESPQ_V3.0_20122012.pdf
101. http://wiki.pro-ibiosphere.eu/w/media/2/24/Skype_Meeting_Minutes_June10_2013.pdf
102. http://wiki.pro-ibiosphere.eu/w/media/3/3b/Pro-iBiosphere_WP6_SIG_Work_Plan_D6.3.pdf
103. http://wiki.pro-ibiosphere.eu/wiki/Workshop_Berlin_4:_Evaluation_of_business_models_currently_in_use_by_partners_and_relevant_non-partners
104. http://wiki.pro-ibiosphere.eu/wiki/Exploitation_Plans_and_Business_Models_-_Task_6.3
105. http://wiki.pro-ibiosphere.eu/wiki/MS12_-_Workshop_on_mark-up_of_biodiversity_literature
106. http://wiki.pro-ibiosphere.eu/wiki/D3.3.2_-_Report_on_progress_during_the_coordination_process_of_partners_and_non_consortium_partners
107. http://wiki.pro-ibiosphere.eu/wiki/Meeting_Notes_Workshops_Leiden_February_2013
108. http://wiki.pro-ibiosphere.eu/wiki/Pilots#Fungi_-_Russulaceae_in_the_Arctic
109. http://www.ncbi.nlm.nih.gov/books/NBK100351/
110. http://www.pensoft.net/journals/mycokeys/article/4302/an-appraisal-of-megascience-platforms-for
111. http://wiki.pro-ibiosphere.eu/w/media/2/24/Pro-iBiosphere_D1.1_Naturalis_VFF_29092012.pdf
112. http://creativecommons.org/choose/zero/
113. http://www.gbif.org/dataset/5cc70440-9e93-11dc-8738-b8a03c50a862
114. http://www.businessmodelgeneration.com

pro-iBiosphere FP7 Project ■ Grant Agreement #312848
Project Periodic Report 1, 31 October 2013; Authors: pro-iBiosphere Consortium ■7th Framework Programme
■ Coordination and support action FP7-INFRASTRUCTURES-2012-1■ Subprogram area INFRA-2012-3.3

Page 85 of 85