



Project Acronym: **pro-iBiosphere**

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

Grant Agreement: **312848**

Project Duration: **24 months (Sep. 2012 - Aug. 2014)**

## D2.1.2 Towards a draft strategy for increased cooperation

Deliverable Status: **Final**

File Name: **pro-iBiosphere\_WP2\_Plazi\_VFF\_30092013.pdf**

Due Date: **30 September 2013 (M13)**

Submission Date: **30 September 2013 (M13)**

Dissemination Level: **Public**

Task Leader: **Donat Agosti (Plazi)**

Authors: **D. Agosti, D. Patterson, G. Hagedorn; T. Catapano, Q. Groom, A. Güntsch, D. Kirkup, E. Kralt, D. Mietchen, R. Morris, A. Paton, L. Penev, S. Sierra**

Copyright

© Copyright 2012-2014, the pro-iBiosphere Consortium. Distributed under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/).

Consisting of:

<b>Naturalis</b>	Naturalis Biodiversity Center	Netherlands
<b>NBGB</b>	Nationale Plantentuin van België	Belgium
<b>FUB-BGBM</b>	Freie Universität Berlin	Germany
<b>Pensoft</b>	Pensoft Publishers Ltd	Bulgaria
<b>Sigma</b>	Sigma Orionis	France
<b>RBGK</b>	The Royal Botanic Gardens Kew	United Kingdom
<b>Plazi</b>	Plazi	Switzerland
<b>Museum für Naturkunde Berlin</b>	Museum für Naturkunde Berlin	Germany

#### Disclaimer

*All intellectual property rights are owned by the pro-iBiosphere consortium members and are protected by the applicable laws. Except where otherwise specified, all document contents are: “© pro-iBiosphere project”.*

*All pro-iBiosphere consortium members have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the owner of that information.*

*All pro-iBiosphere consortium members are also committed to publish accurate and up-to-date information and take the greatest care to do so. However, the pro-iBiosphere consortium members cannot accept liability for any inaccuracies or omissions nor do they accept liability for any direct, indirect, special, consequential or other losses or damages of any kind arising out of the use of this information.*

Revision Control

Version	Author	Date	Status
1.0	Gregor Hagedorn (MfN), Donat Agosti (Plazi)	02 June 2013	First Draft
2.0	Quentin Groom (NBGB), David Patterson (Plazi), Eva Kralt (Naturalis), Daniel Mietchen (MfN), Soraya Sierra (Naturalis)	1-13 September 2013	Draft
3.0	David Patterson (Plazi)	13 September 2013	First elements for an Memorandum of Understanding (MoU)
4.0	Donat Agosti (Plazi)	16 September 2013	Draft reviewed
5.0	Donat Agosti , David Patterson (Plazi), Quentin Groom (NBGB), Anton Güntsch (BGBM)	27 September 2013	Draft
6.0	Lyubomir Penev (Pensoft), Daniel Mietchen, Gregor Hagedorn (MfN), Soraya Sierra, Eva Kralt (Naturalis)	29 September 2013	Draft
7.0	Donat Agosti, David Patterson (Plazi), Peter Hovenkamp (Naturalis), Don Kirkup, Alan Paton (RBGK), Eva Kralt, Soraya Sierra (Naturalis), Lyubomir Penev (Pensoft)	30 September 2013	Draft
8.0	Donat Agosti, David Patterson (Plazi)	30 September 2013	Final Draft reviewed
FF	Donat Agosti (Plazi)	30 September 2013	Final Draft converted to Portable Document Format (PDF)

## Table of Contents

<b>Executive summary .....</b>	<b>6</b>
<b>Introduction.....</b>	<b>9</b>
<b>Vision statement .....</b>	<b>9</b>
<b>Background .....</b>	<b>9</b>
<b>The process .....</b>	<b>10</b>
<b>Steps forward .....</b>	<b>12</b>
<b>Deliverables .....</b>	<b>12</b>
<b>Expert sources of information and informatics initiatives .....</b>	<b>12</b>
<b>Technical issues to improve sharing and interoperability .....</b>	<b>13</b>
<i>Identifiers .....</i>	<i>14</i>
<i>Biodiversity Knowledge Directory .....</i>	<i>14</i>
<i>Linked Data .....</i>	<i>15</i>
<i>Machine readability.....</i>	<i>15</i>
<i>Data standards .....</i>	<i>15</i>
<i>Applications Programming Interfaces (APIs) .....</i>	<i>15</i>
<b>Open Access .....</b>	<b>15</b>
<b>Attribution .....</b>	<b>16</b>
<b>Centralization and Oversight.....</b>	<b>16</b>
<b>Digitisation.....</b>	<b>17</b>
<b>Advanced Open Access publishing of scientific results .....</b>	<b>17</b>
<b>Access to legacy literature .....</b>	<b>18</b>

Tracking and crediting use .....	18
Global relevance.....	19
Benefits and costs of implementation schedule.....	19
<i>Benefits</i> .....	20
<i>Costs</i> .....	20
Incentives for successful implementation.....	21
<i>Trust</i> .....	21
<i>Efficiency</i> .....	22
<i>Awareness</i> .....	22
<i>Reward</i> .....	22
<i>Funding requirements</i> .....	22
Implementation.....	23
<b>Open Biodiversity Knowledge Declaration .....</b>	<b>24</b>
<b>Conclusions and Recommendations.....</b>	<b>26</b>
<b>Acknowledgements.....</b>	<b>29</b>
<b>References.....</b>	<b>30</b>
<b>Annexes.....</b>	<b>34</b>
<b>1.Abbreviations .....</b>	<b>34</b>
<b>2.Description of Work.....</b>	<b>35</b>
<b>3.Participants (23rd May 2013) .....</b>	<b>37</b>

## Executive summary

This report addresses general high-level policy and strategy coordination that will be needed for Open Biodiversity Knowledge Management (OBKM) to become widely adopted by the many current biodiversity initiatives. A centralized and explicit model is unlikely to gain traction, and more progress will be made with a flexible grass roots system that can adapt and evolve as new technologies appear, and as new needs and new opportunities arise. As a co-ordinating mechanism, we propose a declaration whose signatories voluntarily register their support of Open Biodiversity Knowledge Management. This will allow individuals, projects, and institutions to collaborate freely as they establish the policies, strategies, and technical requirements as they implement an Open Biodiversity Knowledge Management System (OBKMS). As a precedent, the members of the pro-iBiosphere consortium have demonstrated their commitment by developing and implementing several of the key features of the goals of the declaration. We expect such a system applied to biodiversity (taxonomic) descriptions to

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

In this report, we make an array of recommendations that, if addressed, will promote Open Biodiversity Knowledge Management.

### Recommendation 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.

**Recommendation 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.

**Recommendation 3:**

The assignment, use and process of managing identifiers should be given very high priority, as this will promote widespread use of persistent, dereferenceable identifiers for physical and digital data objects such as specimens, images and taxonomic treatments as well as their metadata representations.

**Recommendation 4:**

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

**Recommendation 5:**

Design and implement a system for tracking the use of any and all elements of information to ensure that sources and suppliers of data are assigned credit for their contribution to the creation and supply of data.

**Recommendation 6:**

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

**Recommendation 7:**

Establish multi-institutional OBKM working parties to pursue issues relating to collaboration, technical requirements, implementation schedules and sustainability for OBKM. The pro-iBiosphere team will be well suited to coordinate the dialogue that will refine the concept, priorities and technical requirements of OBKM.

**Recommendation 8:**

The pro-iBiosphere project to work with OBKM working parties and to make use of existing mechanisms such as Biodiversity Information Standards TDWG (Taxonomic Database Working Group), to establish technical requirements for standards, vocabularies and protocols for OBKM, to improve access to, and linking and use of, open data (Task 3.3); and to identify implementation processes and priorities; facilitate automation of these processes; and use existing standards as far as possible (Tasks 4.1 and 4.2).

**Recommendation 9:**

Establish an open mechanism for the election of an advisory and management board for the OBKMS to complement the current members of pro-iBiosphere and its board.

**Recommendation 10:**

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



## Introduction

### Vision statement

The Open Biodiversity Knowledge Management System (OBKMS) is a vision of a system of Open Biodiversity Knowledge Management (OBKM) in which European and international partners promote a shared infrastructure that will index, organise, and make accessible all available data that describes biodiversity. iBiosphere will improve the capacity of anyone interested in biodiversity research to take full advantage of all available information to deal with the grand challenges that face our community, open up new methods of biological discovery, and create new knowledge.

### Background

Within the European Union, there are many digital biodiversity initiatives involving individuals, institutions and commercial organisations (see pro-iBiosphere D2.1.1 [Report on ongoing biodiversity related projects, current e-infrastructures and standards](#)). Each project has its own mission. All can be enhanced and made more efficient by drawing on the content and achievements of each other. The pro-iBiosphere project addresses two main goals dealing with (i) technical and semantic interoperability and (ii) sustainability. The vision of pro-iBiosphere is to implement in the near future (through a second project) an integrative system for intelligent management of biodiversity knowledge, with emphasis on taxonomic data. The OBKMS will:

1. Include improved coordination among EU biodiversity data projects and platforms;
2. Reduce duplication of efforts and associated costs; and
3. Lead to integration and interoperability among EU and USA-based global initiatives, such as GBIF, EOL, Global Names Architecture (GNA), DataOne, LifeWatch, and PESI.

Integration will lead to the free exchange of digital biodiversity information among partners. This can be achieved by semantic information management (Berners Lee et al., 2006).

An OBKMS will require a commitment to Open Access, discoverability of existing systems through the use of registries, use of common data standards (Catapano et al., 2011), use of persistent digital

identifiers for data elements ([pro-iBiosphere Best practices for stable URIs](#)), and changed working practices (Thessen and Patterson, 2010; D3.1 report [Towards a Best Practices on Editorial Policies for the curation and publication of fundamental biodiversity data and information in an e-environment](#); D2.3 report [The Use of e-Tools among Producers of Taxonomic Knowledge](#)). It will also require political effort to bring about an appropriate change in the legislative framework for research, as it is being pursued by pro-iBiosphere and the KNEU project [http://biodiversityknowledge.eu/images/PDF/White-paper\\_BiodiversityKnowledge\\_01.pdf](http://biodiversityknowledge.eu/images/PDF/White-paper_BiodiversityKnowledge_01.pdf)).

This document elaborates some of the most pressing issues that will need to be addressed, and recommends that the process begin with a voluntary registration of support for OBKM.

## The process

pro-iBiosphere Task 2.1 "Coordination and routes for cooperation across organisations, projects and e-infrastructures" sets an agenda for high-level policy and strategy coordination. The task consists of two reports/deliverables. [Deliverable 2.1.1](#) addresses the specific expertise of the project partners and identifies potential collaborators, users, and gaps in the infrastructure. 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. Complementing that, the present report D2.1.2 "Towards a draft strategy for increased cooperation" addresses general high-level policy and strategy coordination that will be needed for Open Biodiversity Knowledge Management to become widely adopted by the many current biodiversity initiatives.

The process to identify the community that will support an Open Biodiversity Knowledge Management begins with the endorsement of the "Open Biodiversity Knowledge Declaration" by participants of the pro-iBiosphere workshop held in Berlin in May, 2013. We envisage that this community will grow, taking increasing responsibility for Work Packages 3 (Scientific content and workflow coordination) and 4 (Technical and infrastructure coordination), and will articulate the composition and design of the infrastructure.

In order to implement the recommendation made in report D2.1.1 to develop a 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 (e.g. individuals, projects and institutions). 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.

The registration process will ensure that active parties will declare their willingness to work towards OBKMS, and allow new potential collaborators to identify themselves. Together, we can list common goals and synergies and develop policies and actions that will lead to shared Open Biodiversity Knowledge Management.

In order to ensure continued dialogue as required for this project, we recommend the election of an Advisory Board for OBKMS, and the establishment of working parties. Collectively, these components will ensure cooperation, promote the guiding principles for OBKM, and will take responsibility for the design and implementation of selected elements of Open Biodiversity Knowledge Management. The group will continue to monitor and analyse existing digital infrastructures based on the initial report of ongoing biodiversity related projects ([D2.1.1](#)), address content such as legacy publications and curation systems, including regional and global monographs, as well as entire taxonomic treatments and how they can be semantically enhanced and linked. Two pro-iBiosphere workshops will take place in October 2013, in Berlin: [WS 1, October 8](#) and [WS2, October 8](#). These workshops will analyse how the different groups can cooperate and exchange data; what the needs of user communities are and the necessary software interfaces to serve them; and how to strengthen cooperation between European and non-European biodiversity projects and platforms.

With this in place, stakeholders may better identify potential collaborators with whom they may convene to identify common goals, synergies, gaps, policies and actions towards open biodiversity knowledge management and curation. The process can lead to the changes that will improve coordination, integration and interoperability in the field of biodiversity informatics

## Steps forward

There are already many biodiversity related projects ([D2.1.1](#)). Various levels of interoperability have been achieved, but unfortunately in a still incomplete and inconsistent manner. Often, biodiversity projects end up as silos of locked data and information. Many projects involve methods and tools to exchange content among two or a few players. This is in contrast to the vision of Open Biodiversity Knowledge Management that would be available to any player to engage with and benefit from. Our role is to change the scale of interactions that will enable OBKM.

A framework of collaboration should be established as a process that allows engagement by individuals, teams, institutions and projects. The nature of association should be flexible, but require support for fundamental principles of OBKM - those of open data, use of persistent dereferenceable identifiers, semanticization of content, credit tracking, and using sustainable business models ([D6.3.2](#)). Some issues deserving attention are listed below.

### Deliverables

Initial shared deliverables should be agreed upon. It is likely that initial deliverables will be driven by active members of the community. They should be tied to use cases in every positive opportunity that can stand as exemplars of prototype projects for the community at large.

### Expert sources of information and informatics initiatives

We need to become aware of the biodiversity science and informatics projects that provide content and services and that may benefit from OBKM.

The total number of internet accessible sites with content relating to biodiversity is not known, but is estimated to be in the region of tens of thousands. They include web sites of the estimated 6,000 (Hopkins and Freckleton, 2002) -40,000 (Costello et al., 2012) individual taxonomists, taxon specific web sites ([AlgaeBase](#), [Species Fungorum](#), [Global Lepidoptera Names Index](#), etc.), project web sites, web sites of museums, herbaria and other institutions with interests in biodiversity, state and national governments, international initiatives, etc.

There is as yet no comprehensive listing of biodiversity related projects. The [D2.1.1 report](#) and the [D6.3.1 report](#) provide an insight into 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. Of special interest are opportunities for cooperation among European and non-European biodiversity projects and platforms.

The situation will be dynamic. As a list of existing players is a prerequisite for progress, we give highest priority to voluntary registration through endorsement of the 'Open Biodiversity Knowledge Declaration'. For the declaration to be attractive, registrants need to benefit from signing it. We recommend that the registration is to be regarded as a declaration of interest in OBKM, with an assumed willingness to participate. This will ensure improved access to content, increased use of their own content and therefore its recognition, and capacity to influence the priorities of OBKM. In addition, we recommend an oversight group to promote coordination, and working parties to address particular tasks relating to that. Working Parties might, for example, address stakeholder needs, technical requirements and specifications, initial priorities and deliverables. Some of these issues are discussed below.

Projects such as pro-iBiosphere, as well as other EU-funded FP7 projects like [ViBRANT](#), [BioVel](#), [EU-BON](#), etc. will have a role in recruiting participants to OBKMS.

## **Technical issues to improve sharing and interoperability**

OBKM requires a technical infrastructure. High priority must be given to assembling the requirements of a system that complies with the principles of OBKM, and at the same time supports the broadest possible communities. Issues to be addressed in the first place are basic concepts that have the biggest impact, such as Open Access to scientific results and data, persistent identifiers to the different data elements, and standards for data exchange. Where possible, OBKM should adopt widely accepted standards to facilitate persistence and re-use of the data. Ultimate pre-selection of a standard should be avoided, but rather criteria for choosing the right standard in the right case should be listed.

## ***Identifiers***

Identifiers are simple strings of alphabetical and numerical characters that can be read by machine. Ideally they are stable, persistent, and unique. Good identifiers not only point to online digital resources (i.e., they are resolvable), but they also lead directly to the content it refers to - i.e., they are dereferenceable. Persistence and stability are required for all classes of data objects ([Best practices for stable URI](#)). They will need to be adopted by all OBKM partners and applied to core data such as specimen data, literature, treatments, or digital imaging and other digital objects. In the pre-digital world, objects acquire multiple identifiers over time. While it is desirable to have as few identifiers as possible in the digital world, there is no single solution that meets all needs. Progress is already being made to implement Linked Open Data systems - for example involving the Royal Botanical Garden Edinburgh, the University of Manchester, the Royal Botanic Gardens, Kew, Museum für Naturkunde, the Botanical Gardens in Berlin, iMarine, Plazi, Zoobank, and AntWeb. It is under discussion by the Consortium of European Taxonomic Facilities (CETAF) Information Science and Technology Commission. We will need to accept multiple digital identifiers and integrate identity relations between these identifiers (“a123 same as b123”) into OBKM. Management decisions are required to define stable web addresses (URIs) for specimens, images, treatments and similar digital assets. To be compatible with the Semantic Web and Linked Open Data, the recommended type of stable identifiers are web addresses (“http://...”-URIs). Both stable http-URIs based on the institutional domain name (preferred in the semantic web) and DOI technology in the form of <http://dx.doi.org/..doi> (preferred in the publishing industry) are possible implementations. The process of managing identifiers should be overseen and coordinated by an expert panel established during pro-iBiosphere [Workshop 1](#), October 8, 2013 and in coordination with [CETAF](#).

## ***Biodiversity Knowledge Directory***

This will serve as a registry for content, services and tools used to develop, maintain and manage biodiversity knowledge. The registry makes registered content discoverable. The Directory can set standards for documentation. A good candidate for such a directory would be the Biodiversity Catalogue (<https://www.biodiversitycatalogue.org/>) developed by [BioVeL](#).

### ***Linked Data***

OBKM should be based on the principle of linked data (see e.g. [Best practices](#)). Data linking relies on the adoption of appropriate identifiers. Linking data ensures that data can be enriched by context, that data can be interconnected, that sources of data can be identified and credited, and as a system to monitor use and for quality control.

### ***Machine readability***

Content of the OBKMS has to be machine readable. This requires appropriate data models and APIs.

### ***Data standards***

Standards for data and metadata are a basic requirement for interoperability and for data linking. OBKM should use the currently accepted data standards, collaborating as required in their extension or in developing new standards. This will reduce duplication of effort and ensure best practices.

### ***Applications Programming Interfaces (APIs)***

APIs are the devices that software systems use to interact, as is required for exchange of information or to activate computational tasks. All components of an OBKMS must use consistent mechanisms, and adopt best documentation practices. Service interfaces should be registered in a publicly accessible service registry such as the Biodiversity Catalogue.

## **Open Access**

Open Access is the principle of providing unrestricted access to materials via the internet. It is enshrined in an array of initiatives, such as the [Budapest Open Access Initiative](#) (BOAI, 2002, 2012), the [Bethesda Statement on Open Access Publishing](#), and the [Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities](#) (Berlin Declaration, 2003), [Salvador Declaration, 2005](#), or UNESCO guidelines ([Babini, 2013](#)). Open Access is vital to achieve the goal of OBKMS. It is essential for data sharing. It promotes collaboration, developing new ways to manage the growing deluge of data, with new solutions to link, mine, extract and maintain the data and its derivative products. The Finch report ([Finch, 2012](#)) has pushed the UK Government in the direction of Open Access and prompted active discussion (see [Houghton & Swan, 2013](#), [Anonymous, 2013](#); [Poynder, 2013](#)). A similar trend towards OA is evident with other governments and funding agencies such as the [US Government](#), the [European](#)

[Commission, Austria](#), etc. Legal aspects of Open Access are now under active evaluation ([Dulong de Rosnay & Guadamuz, 2013](#)). Open Access should be accompanied by mechanisms to link content to sources, and to track usages both as a foundation of quality control and as a mechanism to give credit to data sources and data intermediaries. Therefore, we advocate that OBKM is based on the principle of Open Access, given that the original source of information is credited (attribution). The OBKM community will need institutional policy agreements on Open Access, applying the [Linked Open Data five star scoring](#) to data sets and data policies, and associating standardized machine readable Open Access licenses (e. g. the [Creative Commons](#) CC0 for data that may contain minor copyright portions, CC BY or CC BY-SA for text) with content. Not all stakeholders endorse Open Access, and we need to be vigilant and monitor the interpretation of copyright laws.

## Attribution

Attribution identifies the source of materials. The “Routes towards cooperation” workshop held in Berlin in May 23 2013 emphasized the importance of identifying the source of content (provenance) and directing the users of data back to the source. To this, we add the need to provide credit to sources and intermediaries for the subsequent re-use of data,. The source should obtain the greatest credit. Provenance systems are under development. An example is the PROV Data Model for provenance interchange on the web (<http://www.w3.org/TR/prov-primer/>). Another option is to call on annotation systems such as [FilteredPush](#) (Morris et al., 2009) and [Annosys](#), to track all transactions involving any data object. Such systems have the capacity to realize the agenda of providing usage statistics back to sources. By systematically providing links to any data used in a publication, the hypothesis testing and verification of results will be supported.

## Centralization and Oversight

The levels of centralization and decentralization have to be addressed. Centralization can achieve maximum functionality at most efficient running costs, but at risk of reducing innovation. We need to ensure that there is redundancy for critical components. In addition, we need to promote efficiency and an absence of bureaucracy to enable the quick emergence of specific services or their rapid modification as new needs and opportunities arise. To this end, it would be useful to find out what areas certain



institutions would wish to strengthen, and what areas they wish to abandon or delegate to others. The community should decide if it wishes to have an advisory group for OBKM, and working parties able to make recommendations on particular topics.

## Digitisation

Many projects, such as BHL and ADBC (Advancing the Digitisation of Biodiversity Collections), are investing in transforming large quantities of legacy information, such as the billions of biological specimens held in museums and half a million pages of legacy literature, into digital formats, so as to make it available to the "big data" pool. Increased targeting of effort is needed to cater for the changing nature of science, especially as genomics produces data on large numbers of so-called "dark taxa", that is taxa that are identified as such through sequencing but are either new to science (not named) or not matched to existing taxon names and data associated to them. Besides the needed infrastructure to find and link this legacy data (publications, specimens, etc.), a targeted effort would prioritise conversion of appropriate and relevant content, and to assemble reference libraries of genetic information using the resources that are contained within museums and herbaria. A mechanism must be put in place for ongoing digitisation efforts to be informed by the needs of end users, and funding might be targeted for creating content that is in high demand and directly helps to bridge the gap between genomic and traditional taxonomic studies. Further work is needed to understand the granularity of access needed in linking legacy data to ongoing research. What are the requirements for data containers? Does the user community need article level documents, or the treatments within documents, or individual observation record such as the specimen analysed in a DNA study?

## Advanced Open Access publishing of scientific results

Publications are a key element of the scientific process but their production and re-use have so far not been closely integrated with research workflows (Borne, 2010). A campaign is needed to make the scientists and publishers aware that traditional publishing practices add to the increasing amount of inaccessible literature that requires costly post-processing to convert them into semantic documents (e.g. [Agosti & Egloff, 2009](#)), and our goal should be enhanced content that is accessible in a semantic

world. With the emergence of OBKM, semantically enhanced linked publications are needed to allow the content to be openly accessed, for semantic mark-up and linking of content allowing machine reasoning and extraction, and for preservation. A state of the art example is the recently launched [Biodiversity Data Journal \(BDJ\)](#) ([Smith et al., 2013](#)), which is the first ever publishing platform to put together the authoring, peer-review and publication process within a single place. BDJ represents a completely novel workflow and infrastructure to mobilise, review, publish, store, disseminate, make interoperable, collate and re-use data through the act of scholarly publishing.

## Access to legacy literature

The legacy literature is defined to include all literature that does not comply with contemporary standards of linked open data ([Shotton, 2009](#)). Those standards will evolve, but it is foreseeable that traditional e-publishing will not change substantially within the next five years. Only a few years ago, legacy literature was seen as literature not published in a digital format. With the passage of time, legacy literature will refer to any literature that is not semantically marked up. There are an estimated 500 million pages of legacy literature, still expanding with descriptions of the estimated 17,000 newly discovered taxa each year ([Polaszek, 2005](#)) and with updates on known taxa. Taxonomic literature is irreplaceable both because of the quasi-legal nature of taxonomic descriptions and because it contains records of historical occurrences. Published data plays also an important role in the assessment by the Intergovernmental Platform of Biodiversity and Ecosystem Services ([IPBES](#)). The content must be made machine readable, semantically enhanced, and preserved in appropriate repositories.

## Tracking and crediting use

With OBKM, traditional metrics of scientific effort will need to be supplemented with systems that track and provide credit for data use and re-use. Lack of acknowledgement, and a fear of uncredited data use, deters sharing of data. A system that allows tracking all transactions involving any data elements is required, and where possible this should direct maximum credit to the sources of data. The use of universally unique identifiers for each data element, coupled with annotation and/or provenance tracking systems, offer a mechanism by which this may be implemented.

## Global relevance

The challenges that the European community is addressing are shared by biodiversity projects around the world. What we promote to serve the EU will serve the rest of the world, and what they promote is also going to serve the rest of us. We must promote global collaborations such as [GBIF](#), [Biodiversity Heritage Library](#), [EOL](#), [Atlas of Living Australia](#), [Global Names Architecture](#) (GNA), [DataOne](#), [PESI](#) and others.

OBKM requires links with international stakeholders (funding bodies, programs, projects, and individuals) and working together to achieve common goals, such as Open Access.

The Open Biodiversity Knowledge Declaration can be used to engage with international partners, and is a system that can expand, evolve and adapt as progress is made.

Stronger international involvement in an infrastructure will improve data flow. A considerable amount of scientific data has been produced locally but is of global importance. Local scattered data acquire scientific value with the integration into a species- or similar concept and ultimately into the published record and/or database. In this way, local data may eventually acquire global relevance, and local investments should ultimately be incorporated into a global cyberinfrastructure.

Our goal should be an infrastructure supported by major funding institutions irrespective of political boundaries and to strive for agreements in which particular agencies take global responsibility for parts of the infrastructure in exchange for reciprocal actions. The participating institutions in OBKM should act similarly.

## Benefits and costs of implementation schedule

The costs and benefits are being explored under [Workpackage 6](#) as Task 6.1: 'Measuring and Constraining the costs of delivering services'. Two reports will be submitted to the EC on this subject, [D6.1.1](#) in November 2013 and [D6.1.2](#) in May 2014.

A [workshop](#) in May 2013 ([MS19](#)) was organised to discuss how cost is measured among providers of taxonomic works such as faunas and floras. The aim of the workshop sought to understand how

providers currently deliver the information to their customers, who their users are and what are the costs in the provision of their services. By understanding these, the project aims to develop strategies towards a sustainable model for the delivery of biodiversity information. A list of recommendations may be developed on how to sustain useful workflows and pipelines that are needed for a future implementation of Open Biodiversity Knowledge Management under D6.1.2.

### ***Benefits***

The fundamental benefits of OBKM are improved access and delivery of data to users; and more efficient creation, maintenance and dissemination of the data by the producers. This is to be achieved by the elimination of barriers and through greater cooperation. We do not believe that complex and explicit infrastructures will have the flexibility to suit all users. Instead, we believe that the adoption of a number of general principles and methodologies will allow new infrastructures and methodologies to emerge that increase the flow of biodiversity information. These include semantic enhancements of taxonomic treatments, licensing that focusses on attribution, use of persistent unique identifiers as advocated by the W3C http technology that avoid proprietary systems such as DOIs that require central resolving services (for an illustration see this blog: [Hagedorn, 2013](#)). Such changes are relatively simple and can help underpin the free exchange and re-use of content. Immediate benefits include:

- Access to more information, and lowered costs of data acquisition
- Cost sharing (economic efficiencies) in respect of community organised activities or duplicated content - OKMS provides a mechanism for collaboration and cost sharing in the development of new features
- Open up mechanisms for continuous quality control
- Increase use and relevance for authors of content and intermediaries, increasing their relevance and improving their arguments for continued funding

### ***Costs***

OBKM will incur new costs to design and implement an open system. pro-iBiosphere will report estimated costs under Workpackage 6 of Task 6.1, reports [6.1.1](#) and [6.1.2](#). In addition to the implementation, costs will also be incurred for long-term maintenance of a shared system. This continuing support is needed to ensure persistence of high quality services.

Where possible, players use existing systems as far as possible to reduce costs, to enhance interoperability and to enhance their use. The adoption of automated systems where possible will also reduce costs.

Work package 6 is looking at alternative business models and further analysing the costs of production and dissemination, and the benefits to users and providers. Forthcoming work will seek to identify opportunities for efficiencies and explore how the benefits of OBKM can be maximized. Results of WP6 will feed into the final recommendations of the pro-iBiosphere project.

The provision, maintenance and future development of biodiversity data has a cost. The partnership will seek to make the best use of existing funding through increased collaboration, and to explore new avenues of funding to support this.

## **Incentives for successful implementation**

There is in some cases a reticence to share information, because the rewards are not immediately visible and there are threats (inappropriate use of data, or unscrupulous consumers may publish before the original author of the data has the opportunity to do so) (Thessen and Patterson, 2010). Motivation is needed to stimulate the free flow of data. Re-use requires trust and visibility of content, and needs to demonstrate that it is efficient and rewarding.

### ***Trust***

The emergence of OBKM adds a new feature to the organisation and operations of participants. All, whether individuals, initiatives or institutions, must trust in the system to ensure that vital parts are delivered continuously and in a persistent way. Trust is achieved in part through the adoption of agreed, well documented principles, standards and vocabularies.

- Co-ownership of OBKM including a commitments to seek and provide funds to invest in the resulting OBKMS,
- A governing board to provide oversight, with an agreed and documented procedure for election of members

- A process that will ensure the evolution and adaptation of the OBKMS, but with clearly documented goals and deliverables for each stage of development of the system,
- A minimal level of redundancy to ensure stability of services,
- Development of agreed policies on identifiers, open access, names,
- Development of mechanisms to assess growth, performance, and client satisfaction.

### ***Efficiency***

The sharing of content and services, involving appropriate identifiers and APIs will reduce the need for different agents to duplicate the efforts of others and will reduce costs. Shared responsibility for new services will also reduce costs. The adoption of existing standards, vocabularies, and automated solutions to problems reduce costs, too.

### ***Awareness***

To make content and services discoverable, biodiversity web services should be registered. A potential registry is the BioVel Biodiversity Catalogue, hosted at the University of Manchester.

### ***Reward***

Many sources of information are rewarded when the re-use of content is acknowledged. This is equivalent to the traditional system of citing earlier papers in publications. A widespread system that will allow the tracking of data and recording any transactions such as re-use or involvement in services, has yet to be developed. Such systems are now technically possible by modifying annotation systems such as FilteredPush or Annosys. These provide a virtual layer that allows annotations to be attached to any uniquely identified digital object, and incurs little or no cost to a data source that uses persistent, stable and unique identifiers for their digital objects.

### ***Funding requirements***

Though open access, structured data and tools to access data and embed it easily on users' applications might seem a convincing incentive, building, maintaining and populating this infrastructure is not part of a traditional scientific workflow. Only about 25% of scientists use Open Access self-repositories if there is no pressure to do so. Therefore, contributing to these systems has to be mandated (for example, as with the change from request to required in the National Institutes of Health public access policy in

---

2007). The costs of ensuring adherence to the principle of making data available to others must be added to the costs of building or participating in the Biosphere.

## Implementation

The roadmap and timeline to implement the Open Biodiversity Knowledge Declaration will be resolved at the forthcoming fourth management meeting on October 11 in Berlin. There will be a formal signing ceremony of the final Declaration at Chateau Bouchout in the Belgian National Botanical Garden, June 9-13, 2014. A roadmap that includes but is not necessarily limited to the definition of responsibilities for the implementation process, establishment of an Advisory Board that would include members of the pro-iBiosphere consortium, participants from the initial May meeting (Appendix 3) among others, and priorities will be agreed at that time. Dissemination will include various channels, such as a circulation of the initiative within CETAF, the creation of a dedicated Website and the dissemination strategy developed in pro-iBiosphere. We will ask participants of the May workshop to help publicize the Declaration.

## Open Biodiversity Knowledge Declaration

The following draft declaration has been reviewed by the representatives of the pro-iBiosphere consortium, and will be circulated more widely, with the expectation of a final release at the final meeting of this project in June 2014 in Belgium at the Bouchout castle at the National Botanic Garden of Belgium to attract additional signatories.

### Open Biodiversity Knowledge Declaration

Our natural world is a source of wonder, food, water, resources, protection and pleasure that our society needs. The richness and complexity of nature, and the speed of new discoveries made possible by genomics and digital technologies, challenge us to find new ways to use information to benefit from and be better custodians of the natural world. Emerging information management systems can bring together the wealth of information now dispersed in a myriad of different locations. With such systems, we can harness the benefits of rapid discovery and open up our legacy of over 200 years of biological observations. Intelligent information management provides mechanisms to link our understanding of biodiversity to the biomedical research that seeks new solutions to health-care; to tracking change as it affects our agricultural activities and food security by supporting the modeling of life on earth, and to enable new discoveries. To take advantage of the opportunities, information must be made freely and openly available. This declaration is a commitment by the signatories to free and open access to information about our biodiversity and to an inclusive and shared knowledge management infrastructure that will allow us to respond more effectively to society's challenges. This infrastructure will

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

By promoting free and open access to content, the Open Biodiversity Knowledge Management System (OBKMS) will bring together the achievements of many independent biodiversity projects, yet will allow them to retain their identity and missions. The resulting virtual pool of information will allow new services to emerge for those who are addressing societal challenges. Awareness of, access to, and preservation of information will be enhanced by a shared and seamless infrastructure. Finally, by tracking data use, all who create, organise, or mobilise data will be fully recognized for their contributions.



The “**Open Biodiversity Knowledge Declaration**” promotes an overarching approach to Open Biodiversity Knowledge Management by committing its signatories to:

- The free and open use of content, services and other resources by adopting, where possible, licenses that grant all users and automated tools a free, irrevocable, world-wide, right of access to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution consistent with community practices;
- Promote policy development that will foster open biodiversity data;
- Tracking the use of information to ensure that sources and suppliers of data are assigned credit for their contribution to the creation and supply of data;
- Adoption of the agreed infrastructure, standards and protocols established by signatories to OBKM to improve access to and use of open data;
- Register content and services to allow access and use of open data;
- Adopt persistent, dereferenceable identifiers for data objects such as specimens, images and taxonomic treatments;
- Link data with agreed vocabularies that enable participation in the Linked Open Data Cloud;
- Participation in dialogue coordinated by iBiosphere that will refine the concept, priorities and technical requirements of OBKM.

The resulting Open Biodiversity Knowledge Management will improve availability to information; will increase the role and relevance of its participants; will increase their impact, and will reduce costs. As a society, we will understand our natural world better, manage it better, enable new types of discovery, return greater benefits to biomedical and agricultural endeavours, and increase food security. Universities, research institutions, funding agencies, foundations, publishers, libraries, museums, archives, learned societies, professional associations and individuals who share the vision expressed Open Biodiversity Knowledge Declaration are invited to join the signatories by registering at (location to be determined).

## Conclusions and Recommendations

Biodiversity knowledge is based on an extremely wide array of information and data. Similarly, the users and creators of data have a wide range of different goals and requirements. For that reason, a centralized and explicit model is unlikely to gain traction, and more progress will be made with a flexible grass roots system that can adapt and evolve as new technologies appear, and as new needs and new opportunities arise. As a co-ordinating mechanism, we propose a declaration (see above) whose signatories voluntarily register their support of Open Biodiversity Knowledge Management. This will allow individuals, projects, and institutions to collaborate freely as they establish the policies, strategies, and technical requirements as they implement an Open Biodiversity Knowledge Management System (OBKMS). As a precedent, the members of the pro-iBiosphere consortium have demonstrated their commitment by developing and implementing several of the key features of the goals of the declaration. We expect such a system applied to biodiversity (taxonomic) descriptions to

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

We believe that, if addressed, the following recommendation will promote Open Biodiversity Knowledge Management.

### **Recommendation 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 Open Biodiversity Knowledge Management (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.

**Recommendation 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 where appropriate to cover costs of production, maintenance and future development.

**Recommendation 3:**

The assignment, use and process of managing identifiers should be given very high priority, as this will promote widespread use of persistent, dereferenceable identifiers for physical and digital data objects such as specimens, images and taxonomic treatments as well as their metadata representations.

**Recommendation 4:**

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

**Recommendation 5:**

Design and implement a system for tracking the use of any and all elements of information to ensure that sources and suppliers of data are assigned credit for their contribution to the creation and supply of data.

**Recommendation 6:**

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

**Recommendation 7:**

Establish multi-institutional OBKM working parties to pursue issues relating to collaboration, technical requirements, implementation schedules and sustainability for OBKM. The pro-iBiosphere team will be well suited to coordinate the dialogue that will refine the concept, priorities and technical requirements of OBKM.

**Recommendation 8:**

The pro-iBiosphere project to work with OBKM working parties and to make use of existing mechanisms such as Biodiversity Information Standards TDWG, to establish technical requirements for standards, vocabularies and protocols for OBKM, to improve access to, and linking and use of, open data (Task 3.3); and to identify implementation processes and priorities; facilitate automation of these processes; and use existing standards as far as possible (Tasks 4.1 and 4.2).

**Recommendation 9:**

Establish an open mechanism for the election of an advisory and management board for the OBKMS to complement the current members of pro-iBiosphere and its board.

**Recommendation 10:**

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

---

## Acknowledgements

The pro-iBiosphere consortium acknowledges the participants in the workshop that was held in Berlin (May 2013) and the reviewers of this report. We thank Peter Suber for his input on drafting the declaration.

## References

Anonymous, 2013. Government mistaken in focusing on Gold as route to full open access, says Committee. <http://www.parliament.uk/business/committees/committees-a-z/commons-select/business-innovation-and-skills/news/on-publ-open-access/>

Babini, D 2013. UNESCO Guidelines provide a detailed review of Open Access. Scielo in Perspective (blog) <http://blog.scielo.org/en/2013/09/13/unesco-guidelines-provide-a-detailed-review-of-open-access/#.UjZmocbTuoM>

Berlin Declaration 2003. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. <http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>

Bethesda Statement on Open Access Publishing. 2003.  
<http://legacy.earlham.edu/~peters/fos/bethesda.htm>

BOAI, 2002. Budapest Open Access Initiative. <http://www.budapestopenaccessinitiative.org/read>

BOAI, 2012. BOAI 10, ten years on from the Budapest Open Access Initiative: setting the default to open. <http://www.budapestopenaccessinitiative.org/boai-10-recommendations>

Bourne, P. E. 2010. What Do I Want from the Publisher of the Future? PLoS Comput Biol 6: e1000787.  
doi:10.1371/journal.pcbi.1000787

Catapano, T., Hobern, D., Lapp, H., Morris, R.A., Morrison, N., Noy, N., Schildhauer, M & D. Thau. 2011. Recommendations for the use of knowledge organisation systems by GBIF. GBIF secretariat.

Declarations in support of Open Access.  
[http://oad.simmons.edu/oadwiki/Declarations\\_in\\_support\\_of\\_OA](http://oad.simmons.edu/oadwiki/Declarations_in_support_of_OA)

---

Dulong De Rosnay, M. and Guadamuz, A. 2013. Open access to biodiversity data: a comparative survey. 17th ICABR Conference “Innovation and policy for the bioeconomy”. Ravello (Italy) (pers. comm.)  
<http://hal.archives-ouvertes.fr/docs/00/83/34/80/PDF/DulongGuadamuz.PDF>

Finch, J. 2012. Accessibility, sustainability, excellence: how to expand access to research publications  
<http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf>

Hagedorn, G. 2013. Part 3 of a series about LSID vs Semantic Web. Blog  
<https://plus.google.com/u/0/117201190352607228695/posts>

Hess Ch., Ostrom E. 2006. Understanding knowledge as a commons. From theory to praxis. MIT Press  
ISBN: 9780262083577 <http://mitpress.mit.edu/books/understanding-knowledge-commons>

Houghton, J. and Swan, A. 2013 Planting the green seeds for a golden harvest: Comments and clarification on “Going Gold”. D-Lib Magazine 19 /1/2) [doi:10.1045/january2013-houghton](https://doi.org/10.1045/january2013-houghton)

KNEU. 2013. A recommended design for “Biodiversity Knowledge”, a network of knowledge to support decision making on biodiversity and ecosystem services in Europe, White paper.  
[http://biodiversityknowledge.eu/images/PDF/White-paper\\_BiodiversityKnowledge\\_01.pdf](http://biodiversityknowledge.eu/images/PDF/White-paper_BiodiversityKnowledge_01.pdf)

Berners Lee, T., Hall, W. Hendler, J. A., O'Hara, K., Shadbolt, N. and Weitzner, D. J. 2006. [A Framework for Web Science](#), *Foundations and Trends in Web Science*, ISBN: 1-933019-33-6

Costello, M. J., Wilson, M. and Houlding, B. 2012. Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Syst. Bio.*, 61: 871-883.

Hopkins G. W. and Freckleton R. P. 2002. Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Animal Conservation* 5: 245-249.

Morris, P.J., Kelly, M., Lowery, D.B., Macklin, J.A., Morris, R., Tremonte, D. and Wang, Z. 2009. Filtered Push: annotating distributed data for quality control and fitness for use analysis. Eos Transactions of the American Geophysical Union (AGU) 90(52) Fall Meeting Supplement, Abstract available at <http://adsabs.harvard.edu/abs/2009AGUFMIN34B..08M>.

Polaszek, A., Agosti, D., 24 coauthors, 2005. A universal register for animal names. Nature, 437, 477 [doi:10.1038/437477a](https://doi.org/10.1038/437477a)

Poynder, R. 2013 UK House of Commons Select Committee publishes report criticising RCUK's Open Access Policy. Blog <http://poynder.blogspot.de/2013/09/uk-house-of-commons-select-committee.html>

pro-iBiosphere. Best practices for stable URIs. [http://wiki.pro-ibiosphere.eu/wiki/Best\\_practices\\_for\\_stable\\_URIs](http://wiki.pro-ibiosphere.eu/wiki/Best_practices_for_stable_URIs)

pro-iBiosphere. Report [on ongoing biodiversity related projects, current e-infrastructures and standards](http://wiki.pro-ibiosphere.eu/w/media/c/c9/Pro-iBiosphere_WP2_PLAZI_D2.1.1_VFF_30062013.pdf). [http://wiki.pro-ibiosphere.eu/w/media/c/c9/Pro-iBiosphere\\_WP2\\_PLAZI\\_D2.1.1\\_VFF\\_30062013.pdf](http://wiki.pro-ibiosphere.eu/w/media/c/c9/Pro-iBiosphere_WP2_PLAZI_D2.1.1_VFF_30062013.pdf).

Salvador Declaration 2005. Declaration of Salvador - Commitment to Equity. <http://www.icml9.org/channel.php?lang=en&channel=91&content=438>

Smith V, Georgiev, T., Stoev, P., Biserkov, J., Miller, J., Livermore, L., Baker, E., Mietchen, D., Couvreur, T., Mueller, G., Dikow, T., Helgen, K.M., Frank, J., Agosti, D., Roberts, D., Penev, L. 2013. Beyond dead trees: integrating the scientific process in the Biodiversity Data Journal. Biodiversity Data Journal 1: e995. DOI: [10.3897/BDJ.1.e995](https://doi.org/10.3897/BDJ.1.e995)

Swan A. 2012. Policy Guidelines for the development and promotion of Open Access. United Nations Educational, Scientific and Cultural Organization, Paris <http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>



Thessen, A. E., and Patterson, D. J. 2011. Data issues in the life sciences. *ZooKeys* 150: 15–51. doi:  
10.3897/zookeys.150.1766

Vision T. 2010. Open data and the social contract of scientific publishing. *Bioscience* 60(5): 330-331. doi:  
10.1525/bio.2010.60.5.2

## Annexes

### 1. Abbreviations

<b>ADBC</b>	NSF program, Advancing the Digitisation of Biodiversity Collections
<b>API</b>	Application Programming Interface; a mechanism that allows computers to exchange information and instructions
<b>BHL</b>	Biodiversity Heritage Library, digitising biodiversity literature
<b>DOI</b>	Digital Object Identifier, a system that provides a simple means of identifying digital objects, most widely used by publishers to identify publications.
<b>GBIF</b>	Global Biodiversity Information Facility, a major international initiative to share biodiversity data
<b>GNA</b>	Global Names Architecture, a collaborative project to build a names-based infrastructure to help manage biodiversity data.
<b>IPR</b>	Intellectual Property Rights
<b>MfN</b>	Museum für Naturkunde, Berlin
<b>MoU</b>	Memorandum of Understanding, a semi-formal mechanism to clarify the nature and terms of agreement between two or more parties who seek to cooperate
<b>NBGB</b>	National Botanic Garden Belgium
<b>NSF</b>	National Science Foundation, Major Funding agency with the United states of America.
<b>OBKM</b>	Open Biodiversity Knowledge Management
<b>OBKMS</b>	Open Biodiversity Knowledge Management System
<b>PROV</b>	W3C Provenance model
<b>RDF</b>	Resource description framework: an agreed mechanism of describing information on the internet, so that it can be understood by computers; a component of the Semantic Web.
<b>TDWG</b>	Taxonomic Database Working Group
<b>URI</b>	Universal Resource Identifier, a string of alphanumeric characters that identifies a digital object that is accessible through the Internet.

## 2. Description of Work

Biodiversity core data and information constitutes an important source of knowledge for many disciplines. In order to facilitate access to this knowledge, technical and semantic interoperability barriers need to be addressed. The objectives of pro-iBiosphere are, to: a) Coordinate towards and prepare the foundations for a long-term viable, evolving knowledge management, aggregation and integration platform needed to replace and improve the present system of taxonomic literature, especially as presented in Floras and Faunas; b) Provide new methods to synthesize distributed knowledge and a strategy to adapt methods of acquisition, curation, and dissemination of biodiversity data to the digital era; c) Help to align ongoing and forthcoming semantic mark-up of taxonomic literature, and to link elements of biodiversity literature (ie. taxonomic treatments) to the original data, such as the individual observation record (being the essential foundation of any biodiversity information); d) Promote and monitor the development and adoption of common mark-up standards and specifications for making biodiversity knowledge more accessible and re-usable; e) Provide the community with technical solutions for the enhancement and use of these data; f) Analyse and evaluate business models for supporting Open Science and provide recommendations to achieve sustainable delivery of biodiversity information to target audiences; g) Develop and agree on a shared data and Intellectual Property Rights (IPR) policy; h) Promote and increase cooperation between the major biodiversity projects, initiatives and platforms at EU and global levels. These activities will prepare the ground for an integrative system for intelligent management of biodiversity knowledge. A system that facilitates open access to taxonomic data is essential because it will allow a sustainable provision of high quality data to partners and users, including e-science infrastructure projects as well as global initiatives on biodiversity informatics.

Pro-iBiosphere task 2.1 addresses coordination and routes for cooperation across organisations, projects and e-infrastructures. This task serves as a baseline coordination platform for general high-level policy and strategy coordination. As a prerequisite, a report will be collaboratively prepared, documenting and updating the present activities, strategies, goals, use cases, interests and visions as well as cooperation and interrelations of the various European and international partners interested in participating in a taxon treatment-like knowledge management system. This report will update and consolidate the present knowledge of the project partners, identify potential collaborators, users, and gaps in the infrastructure. The major stakeholders, with special emphasis on organisations cooperation in existing biodiversity e-Infrastructures, will then convene to identify common goals and reciprocal synergies and coordinate their policies and actions towards an open and shared knowledge curation system. The coordination will analyse existing digital infrastructures but also past publications and curation systems, including regional or global monographs as well as entire taxonomic treatments. This task will analyse the way the various groups cooperate, generate and exchange data. Two workshops will be organised for this purpose. A specific goal will be to understand the needs of external communities and report on the necessary software interfaces for these users. Of special importance are further potential routes for cooperation between European and non-European biodiversity projects and platforms. An Advisory Board of representatives from major global biodiversity projects will be established 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 ongoing EU projects; (3) Analysing and developing of strategies for integration and interoperability in the

---

field of bioinformatics between EU and USA-based global initiatives, such as GBIF, EOL, Global Names Architecture (GNA), DataOne, PESI and others.

---

### 3. Participants (23rd May 2013)

Wouter Addink – Naturalis

Donat Agosti – Plazi

Nicolas Bailly – WorldFish Center – FIN

Henk Beentje – RBGK

Laurence Bénéichou – MNHN

Walter Berendsohn – FUB-BGBM

Jordan Biserkov – Pensoft

Jan Brase – Data Cite

Terry Catapano – Plazi

Christopher Chapano – National Herbarium and Botanic Garden (SRGH), Zimbabwe

Joe Cora – Ohio State University

Alastair Culham – University of Reading

Eduardo Dalcin – Rio de Janeiro Botanical Garden

Pablo Demaio – IUCN Temperate South American Plants Specialist Group

Daphne Duin – Naturalis

Sabrina Eckert – FUB-BGBM

Willi Egloff – Plazi

Henry Ford – Ecoflora of the British Isles, Bath University

Teodor Georgiev – Pensoft

Anton Güntsch – FUB-BGBM

Quentin Groom – NBGB

Gregor Hagedorn – JKI

Jana Hoffmann – MfN

Peter Hovenkamp – Naturalis

Norman Johnson – Ohio State University

Patricia Kelbert – FUB-BGBM

Yong-Shik Kim – Yeungnam University

Paul Kirk – Index Fungorum

Don Kirkup – RBGK

---

Eva Kralt – Natuarlis  
Alexander Kroupa – MfN  
Wouter Los – Life Watch  
Karol Marhold – SAVBA  
Patricia Mergen – Royal Museum for Central Africa  
Alexa Michel – FUB-BGBM  
Daniel Mietchen – Wikipedia, Wikimedia Commons, Open Knowledge Foundation  
Chuck Miller – TDWG standards  
Jeremy Miller – Naturalis  
Norman Morrison - BioVeL, The University of Manchester  
Luciana Musetti – Ohio State University  
Richard Old – XID Services, Inc.  
Sylvia Mota de Oliveira – Naturalis  
Alan Paton – RBGK  
David Patterson – GNA  
Deborah Paul – Florida State University  
Lyubomir Penev – Pensoft  
Richard Pyle – Bishop Museum  
Eckhard von Raab-Straube – FUB-BGBM  
Greg Riccardi – Florida State University  
Vincent Robert – CBS-KNAW Centraalbureau voor Schimmelcultures  
Dave Roberts – ViBRANT  
Marianne Le Roux – SANBI  
Hannu Saarenmaa– EU BON, University of Eastern Finland  
Guido Sautter – Plazi  
Ben Scott – Natural History Museum  
Soraya Sierra – Naturalis  
Yashica Singh – KZN Herbarium, Durban, SA  
Vince Smith – NHM  
Charlotte Thinois – MNHN  
Jonathan Timberlake – RGBK

Jan van Tol – Naturalis

Dagmar Triebel – SNSB

William Ulate – BHL

Florian Wetzel – EU BON

Zerihun Woldu – Addis Ababa University

Shuangxi Zhou – Macquarie Univ., AU