





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)

D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management

Final
pro-iBiosphere_WP6_RBGK_VFF_31082014.pdf
31 August 2014 (M24)
1 September 2014 (M24)
Public
Don Kirkup (RBGK)
C. Johns, D. Kirkup, A. Paton

pro-iBiosphere FP7 Project ■ Grant Agreement #312848





Copyright

© Copyright 2012-2014, the pro-iBiosphere Consortium. Distributed under the terms of the <u>Creative Commons Attribution 3.0</u> License.

Consisting of:

Naturalis	Naturalis Biodiversity Center
NBGB	Nationale Plantentuin van België
FUB-BGBM	Freie Universität Berlin
Pensoft	Pensoft Publishers Ltd
Sigma	Sigma Orionis
RBGK	The Royal Botanic Gardens Kew
Plazi	Plazi
MfN	Museum für Naturkunde Berlin

Netherlands Belgium Germany Bulgaria France United Kingdom Switzerland 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	Charlotte Johns	07/07/2014	First Draft
2.0	Charlotte Johns, Don Kirkup, Alan Paton	29/08/2014	Draft
3.0	Don Kirkup, Alan Paton, Soraya Sierra	31/08/2014	Draft
4.0	Peter Hovenkamp, Soraya Sierra, Don Kirkup	31/08/2014	Final Draft reviewed
FF	Don Kirkup	01/09/2014	Final Draft converted to Portable Document Format (PDF)

Page 3 of 47





Table of Contents

Executive summary
Introduction7
Business requirements for OBKMS9
Currently met OBKMS Business Requirements9
Additional OBKMS Business Requirements not yet met10
Model evaluation
Discussion14
Conclusions17
Acknowledgements
Appendix 1. Business requirements currently met19
Appendix 2. Unmet business requirements for OBKMS and actions needed22
Appendix 3. Business process models25
Data Mobilisation26
Digitisation and conversion of hardcopy to machine readable text26
Large-scale text conversion
Small-scale text conversion
Markup of legacy texts
Small-scale markup of legacy texts
Large-scale markup of legacy texts
Prospective markup of texts
Online XML authoring tools for prospective markup of text32





Data Dissemination & Discoverability	.34
Advanced Publishing	.34
Hosting systems	.38
Outreach	.39
Marketing and demonstration projects	.39
Fundraising	.41
Provider help desk function	.43
Appendix 4 Pre- and Post-evaluation business process diagrams	.45





Executive summary

In order to make fundamental biodiversity data digital, open and re-usable, the pro-iBiosphere project has the vision of implementing an Open Knowledge Biodiversity Management System (OBKMS). To achieve this, the project must address the challenge of sustainability.

This report recapitulates the pro-iBiosphere vision, describes the Business Requirements for the core functions of an OBKMS; identifies the services that do not yet exist but would be essential for a fully functioning OBKMS; and provides an evaluation of the models for the business processes underpinning these requirements, drawing criteria from WP6 reports and outputs of the workshop <u>MS24 "Model Evaluation</u>".

The main findings are:

- Existing business requirements will continue to be met by partners' current funding models.
- Funding for infrastructure and services will be a mixed model (public and private money) and there will be opportunities for specialists, for example in the provision of text conversion services. In this mixed model the following sources of income may play a role:
 - Grants
 - Joint project funding
 - Subscription
 - Charging for interpretative value added services aimed at commercial consumers
 - Additional funding to cover set up costs of new OBKMS product and service infrastructure
- Support must be found for the development of production level tools, not just for the prototypes.
- Practical demonstrations should be developed to demonstrate the value of an OBKMS to
 - the institutions themselves
 - users and
 - society at large.
- Governance is likely to evolve. To begin, the joint OBKMS governance could be quite small, focusing on the development of demonstration projects to showcase the value of OBKMS and on coordinating outreach activities to ensure the benefits of OBKMS are delivered to users and the broader scientific community. Additional funding would be required to establish an independent OBKMS entity with a "core" outreach role.

The key recommendations are:

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

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 6 of 47

D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August 2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme Coordination and support action FP7-INFRASTRUCTURES-2012-1 Subprogram area INFRA-2012-3.3





Introduction

At the outset of the pro-iBiosphere project, an Open Biodiversity Knowledge Management System was seen as a system that would facilitate the acquisition, curation, synthesis and sustainable provision of high quality biodiversity data to partners and users. It would include e-science infrastructure projects as well as global initiatives on biodiversity informatics. The system would create an authoritative framework for the naming of specimens, the development of identification tools and the generation of descriptions. It would also facilitate the open access of taxonomic data, creating synergies with other initiatives / projects and, hence, allow linkage of taxonomic data in a wider context. Users such as European natural history institutions, herbaria and botanic gardens would benefit from the OBKMS as it would help to fulfil their responsibility to ensure that taxonomic data within their collections, names servers, libraries and publications are generated, curated and effectively disseminated, and that semantic markup and technical interoperability are in place adequately facilitating these functions. The findings below derive from the workshop on model evaluation MS24 and from previous WP6 deliverables, Particularly D6.1.2 Report on cost delivery, efficiency and cost reduction through effective practices (2), D6.2.1 Report on benefits to users/clients, D6.2.2 Report on benefits to providers/suppliers, D6.3.3 Report on diversity and strengths of existing business plans and discussion of sustainability (3), D6.4.1 Risk analysis, D6.4.2 Sustainability Draft Report, and discussion of these at the workshop and final event.

The pro-iBiosphere workshop on <u>'Model Evaluations'</u> that took place in Brussels, June 2014 addressed the question "Once the system has been created, how will it remain sustainable?". The following conditions for sustainability were highlighted:

- Be clear about what we want to do
- Identify a niche
- Ensure that the products and services meet our understanding of user needs
- Continue to advance technologically and to demonstrate benefits to stakeholders through pilot projects

Over the course of the project, the vision of the OBKMS shared by consortium partners has evolved into the following vision: An OBKMS provides people and computers with free and open access to all digitally managed knowledge and information (as well as underlying research data), in a system that allows continuous collaborative review and improvement.

The goal of OBKMS is to allow scientists as well as the entire society to respond more effectively to present and future challenges related to biodiversity. It will support professional evidence-based policy advice as well as informed citizen participation in political decisions. OBKMS addresses the acquisition, mobilisation, curation, interoperability, synthesis and dissemination of biodiversity information. The pro-iBiosphere project deals with the technical, legal, and sustainability aspects of OBKMS.

In the case of pro-iBiosphere, sustainability would refer to ongoing and functioning data-flows and services, where benefits of services would outweigh costs. Work package 6 of the pro-iBiosphere project focuses on sustainability. Kew, Sigma Orionis and

pro-iBiosphere FP7 Project Grant Agreement #312848

D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August 2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme ■ Coordination and support action FP7-INFRASTRUCTURES-2012-1 ■ Subprogram area INFRA-2012-3.3

Page 7 of 47





Naturalis have been coordinating this work package, with input from other consortium partners and from the broader biodiversity community interested in the sustainability of an OBKMS.

To achieve the goal of delivering biodiversity data in open and re-usable forms, OBKMS-governance must address the sustainability of both the data and the means to access them. Sustainability requires that providers can curate, supply and exploit data effectively to ensure that the benefits of participation outweigh the costs. This can be facilitated through technical innovations, but also requires a full understanding of user requirements and how these can best be met.

Benefits to the users include increased availability of data though a central portal as well as increases in the volume, consistency, reliability and currency of information available through searches. The benefits to data-providers include greater use of their data, more accurate crediting, improved awareness of impact, avoidance of duplicated effort, increased opportunities for innovative collaboration and tools for curation, analysis and research.

The particular niche for (pro)-iBiosphere (identified in D6.3.1 and D6.3.2) is a focus on the mobilisation of knowledge contained within the legacy literature allowing the synthesis of research data. This does not involve massive spatial data or climate change evaluation, but allows a concentration on taxonomic treatments (primary biodiversity information, in particular the information contained in existing literature. Issues related to the curation of this data once mobilised will become an increasingly important element as an OBKMS is being developed.

To fulfill our vision we identify and implement a core set of services that allows us to go forward and make a viable proposition. The main challenge is to continue the development of these services to meet the growing need for them. That the need will be growing can be inferred from the interest in the <u>Bouchout Declaration</u> in which the goal is clearly formulated that data should be available and can be viewed by a maximum number of people.

In the following section we outline the business requirements for the OBKMS core services.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848 D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August 2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme ■ Coordination and support action FP7-INFRASTRUCTURES-2012-1 ■ Subprogram area INFRA-2012-3.3 Page 8 of 47





Business requirements for OBKMS

A list of the business requirements that must be delivered to provide value (as opposed to products, systems, software, and processes which identify how the business requirements are to be met) is presented below The list is derived from the core products and services identified in D6.4.2 draft sustainability report.

The list includes Business Requirements that the community currently meets (from Appendix 1), together with those which are yet to be implemented (Appendix 2).

Currently met OBKMS Business Requirements

- Acquisition and mobilisation: Acquisition is the gathering of data and records, these can be in multiple formats or media. Biological information found within legacy literature is often hard to gather as it can be in hard copy form and held in old repositories. Mobilisation refers to the digitisation of documents into a machine/computer readable format, and depends upon digitisation technology. Mobilisation allows relevant data to be processed easily via computer so that it can be easily found and extracted. This service also enables data to be linked to external data networks. Tools promoting acquisition and mobilisation of data exist in prototype form, though more resources are needed to convert these tools and procedures into production level workflows. Funding to date for mobilisation and acquisition has been provided by institutions and grant funding. Some institutions specialise on some areas of the workflow and the possibility of outsourcing to specialists has been investigated.
- **Curation**: It is important that data remains current and up to date. Data curation is the active and on-going management of data through its life cycle (creation, management, and long-term care of data). It allows links between the legacy biodiversity literature and currently published biodiversity data, facilitates reviews and updates data. Curation depends on services for online editing (such as that provided by the EDIT Platform for Cybertaxonomy) and may include broader community involvement. As OBKMS grows, the need for curation will be an increasingly important issue.
- Dissemination & Discoverability: It is important that users are able to find and extract information easily. The
 discoverability of this information depends on services for publication, indexing, adoption of global unique identifiers and
 name reconciliation. Advanced publication services, such as the republication of Biotas (for example the online
 republication of volume 14 of Flora Malesiana) provide exemplars which illustrate the potential of semantically enriched
 publications to generate value for both users and publishers, built on top of open access information.
- Interoperability: It is important that the partner repositories, networks and aggregators are linked and compatible, so data can move freely between systems. The biological community uses many different data collection protocols, data formats and tools. The OBKMS therefore needs services which can pull different information sources together and also to prioritise common data handling methods to enable the system to be compatible with the optimum number of partners

pro-iBiosphere FP7 Project Grant Agreement #312848

Page 9 of 47

D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August 2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme Coordination and support action FP7-INFRASTRUCTURES-2012-1 Subprogram area INFRA-2012-3.3



A way in which these can be bridged is by using the emerging DarwinCore-Archive format as a common data and metadata harvesting format. This area of OBKMS compliance relies heavily on the adoption of data standards.

- Aggregation & Synthesis: This area contains services which pull together, analyse and answer research questions from the aggregated data.
- **Outreach**: A service that attracts and maintains positive user relations. This service educates and raises awareness among the user and provider communities. Specific areas include provider help desk functions, fundraising, marketing, the promotion of standards and demonstration projects.

The scenarios for the provision of these Business requirements are elaborated further in the Discussion section below.

Additional OBKMS Business Requirements not yet met

Three core areas were highlighted during the 'Model Evaluation Workshop' MS24. These are products and services that we currently do not provide, but that we envisage as necessary for the OBKMS. The focus upon these core areas sets the OBKMS apart from other systems. These new areas of the OBKMS will require major new funding, so it is important that it is understood what needs to be done to implement them.

To identify aspects of the products and services we do not have and need to further develop, each product and service has been evaluated to isolate what we do have and what we do not, according to the sustainability headings in Appendix 2.

- Taxonomic Name Reconciliation services (TNRS) necessary to reconcile and disambiguate organism names. This is important as information is stored in disparate systems, using different names and taxonomies. These services are required to aggregate data about particular organisms. Prototype services have been developed in projects such as Global Names, IPlant and BioVel.
- Annotation services. Openly accessible data can be curated by a broad audience if that audience have the ability to annotate the data, and such annotations can be managed. Annotation systems thus facilitate a broader network of data curation and improve the prospects of sustainable data curation. Examples include FilteredPush and Annosys
- **Cross referencing services and infrastructure (identifiers stored and linked)** Persistent identifiers and the means to discover and resolve them resolve them allows greater and more reliable access to data, facilitating services.
- Linked open data (technical accessibility of linked open data machine and human) Linked open data (LOD) is the online connection and open licence publication of related data, information and knowledge that was not previously linked on the Semantic Web. Methods of supporting this include the use of Uniform Resource Identifiers (URIs), the Hypertext Transfer Protocol (HTTP) and the Resource Description Framework (RDF).

In order to implement the correct technology for LOD a number of areas surrounding LOD have been explored. These consist of the best practice for Persistent URIs as noted above; a strategy for extended use of RDF in the biodiversity domain (D4.2); use of ontologies that can be used as controlled vocabularies; and treatments in RDF. However, the implementation of LOD service needs to be further explored through use cases and pilots. Areas that need to be further addressed are greater accessibility, infrastructure and interfaces to allow easy use, and a broader scope interoperability of ontologies.

pro-iBiosphere FP7 Project
Grant Agreement #312848

Page 10 of 47





Policy support is a very important factor when considering LOD. <u>The Bouchout Declaration http://www.bouchoutdeclaration.org/</u> for Open Biodiversity Knowledge Management has been created and released, its purpose being to raise awareness and help to make digital data about our biodiversity openly available. It offers members of the biodiversity community a way to demonstrate their commitment to open science. Further work can be done to push for additional support for the declaration and its implementation.

The outreach services will be needed to stimulate the use of data by other LOD communities such as Genomics. Interlinking would also occur where commonalities in data support citizen science and non-institutional systems by providing hooks for long-tail small data.

Short term funding is required to help stimulate development of these needed services. However this funding should also support the development of production level tools in addition to developmental prototypes.

pro-iBiosphere FP7 Project Grant Agreement #312848 D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August 2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme Coordination and support action FP7-INFRASTRUCTURES-2012-1 Subprogram area INFRA-2012-3.3 Page 11 of 47





Model evaluation

In deliverable <u>D6.4.2</u> 'Draft Sustainability Report' the proposed key products and services (i.e. the business requirements) of an OBKMS were identified. In **Appendix 3**, the business process models that are required to fulfill these business requirements are evaluated by mapping their costs, benefits, strengths, weaknesses, opportunities and threats of each approach (ex D6.1.2, D6.2.1, D6.2.2, D6.4.1, D6.4.2). These were further examined during the pro-iBiosphere workshop <u>MS24</u> 'Model Evaluations'. The key points emerging from the workshop discussions (see also the presentations by <u>Anton Güntsch, Openup Sustainability, and Wouter</u> Addink, Catalogue of life) are :

- It is recommended that the large scale digitisation of legacy texts should involve outsourcing to specialist providers. To make this service sustainable it is important that:
 - Relationships can be maintained through effective outreach and communication (D6.4.1)
 - The chosen third party OCR providers are willing to take long-term responsibility for this infrastructure component
- Only a large scale approach to markup for legacy literature would appear to have the potential to be viable (assuming that technological improvements are possible). XML authoring tools for advanced publishing (e.g. the Pensoft system) illustrate one route for new data with benefits to authors, publishers and users. To make these service sustainable it is important that:
 - The chosen markup markup providers are willing to take long-term responsibility for this infrastructure component.
 - Production level tools are developed but development (prototypes) of entirely new systems should be discouraged.
 - o Funding is available kick-start technological innovations
- Data and information will be made discoverable through the Edit platform and through advanced publishing such as integrated text and data publishing, scholarly publishing and re-publishing. To ensure sustainability continues, the OBKMS must:
 - Try to find organisations willing to take long-term responsibility for infrastructure components (ideally before the start of a project/proposal)
 - o Convince institutions that subscribing to services (data, computational) can be a worthwhile investment.
 - o Avoid new developments if possible
 - o Think of sustainability and feed the biological network that already exists.
 - Improve the IPR basis (Creative Commons) and strengthen data security to ensure that basic sources (providers) benefit from these developments.
- Three important outreach products and services have been identified; Provider help desk function, Marketing and demonstration projects to demonstrate value of the system and Fundraising. To ensure that these areas of outreach remain sustainable, the OBKMS must:
 - Think of the OBKMS as a business. Identify Products and Services, make the OBKMS a business dependency and sell it
 - o Plan for the future, think of succession and provide incentives for tomorrow's taxonomists.
 - Engage in international collaboration:. It must promote existing initiatives (GNA, CoL, GBIF) It must work towards an ecosystem of biodiversity info services (Horizon 2020) It must promote a shared approach to hosting, e-infrastructures, data resources, business models.
 - o Be explicit about the costs involved

pro-iBiosphere FP7 Project ■ Grant Agreement #312848





 \circ \quad Need to decide upon a governance structure and hosting agreement

The pre- and post-evaluation business process diagrams are presented in Appendix 4 (figures 1&2 respectively)





Discussion

Sustainability of biodiversity information services is a concern. Many services are based on subsidised projects or grants, rather than anchored in core-funding or as part of the academic mission and output of the taxonomic institutions and international organisations.. This can lead to services not being permanently adopted and integrated into core business, and so long term support is lacking.

The funding of a research project is only a step in the development of a more ambitious project by the consortium partners. In some cases the research results can lead to a pre-commercial or even a commercial phase during which no EU funding will be necessary any longer. Suited business models ensure full exploitation of project results, hence, the continuity of the partners' efforts, and the sustainability of the overall project they envisioned.

It would appear that many of the existing business requirements can continue to be met by partners' current business processes and funding models with the following provisos:

- the current climate for core funding for many institutions is declining and policy support is needed (D6.4.4).
- current best practice needs to be adopted for the technological processes underpinning data mobilisation (e.g. markup) and support for further research is required.
- the rate of mobilisation of the legacy literature is slow and likely to remain so, resulting in a delay in the realisation of the benefits.

In this scenario, funding for infrastructure and services will continue to be a mixed model (public and private money):

- Institutions that currently provide resources for products and services, will continue to do it more efficiently:
 - Explore Mixed funding:
 - Grants
 - Joint project funding
 - Subscription
 - Charging for interpretative value added services aimed at commercial consumers
 - Additional funding to cover set up costs of new OBKMS product and service infrastructure
- The development of production level tools needs to be funded, not just the prototypes.

In the MS22 workshop "<u>Meeting to evaluate business models currently in use by partners and relevant non-partners</u>" workshop we considered what unique roles an OBKMS entity could perform. We outlined two extremes for an OBKMS model and two intermediate scenarios (see diagram below): At one end of the spectrum, there is an inclusive entity which takes on all the major roles such as data mobilisation, hosting, publishing etc. At the other we have a completely dispersed consortium without any central OBKMS body.

pro-iBiosphere FP7 Project
Grant Agreement #312848





As a group we were inclined to the latter (no coordinating role), but having gone through the evaluation, a case could be made for an intermediate model where an outreach role could only be performed efficiently by a separate coordinating OBKMS entity, at least to begin with. This would kick-start production level mobilisation of literature, after which an existing body (such as GBIF) might take over.

The advantages of having this small separate core would be that;

- it would build on the existing expertise of the group and the momentum that has been created and
- the governance structure would be much simpler than if subsumed into a larger entity.
- It could coordinate beyond what would be possible by individual partner institutions.

Additional funding perhaps via subscription, may be required for a OBKMS legal entity which would;

- Coordinate with existing e-infrastructures and users to continue to promote standards and develop user requirements.
- Promote practical demonstrations of value to;
 - the institutions themselves (providers)
 - o users
 - o traditional publishers seeking to develop new sustainable services in the open access environment
 - o society at large
- Initially the governance structure could be informal and lightweight focusing on securing institutional commitment to key components and coordinating demonstration projects.







4 Scenarios for an OBKMS entity with core functions (as defined by the yellow area)

A – OBKMS with no unique core function (all functions lie outside in the red area) B - OBKMS with comprehensive core functions, C - OBKMS with selected specialised core functions, D - OBKMS with a small core function (coordination & outreach), all other functions are carried out by consortium, utilise and feed existing e-infrastructures or yet to be created services.

pro-iBiosphere FP7 Project Grant Agreement #312848 D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August 2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme Coordination and support action FP7-INFRASTRUCTURES-2012-1 Subprogram area INFRA-2012-3.3 Page 16 of 47





Conclusions

OBKMS offers many benefits to both providers and to users – advanced publishing services based on open access data are better for the authors of taxonomy as their data is more widely and effectively disseminated. Publishers can also benefit by developing premium services to target specific user groups. The missing components of OBKMS will require increased funding- through grants and added value services; but the benefits are increased and more effective use of information.

- There are inefficiencies in the current system
 - There is no system to prevent duplication of specialist tasks
 - There is a reliance on work done by workers using highly specialized tools without allowing time for the necessary training and skills development
 - There is no system to ensure that prototypes will, when successful, be developed into functional production level workflows
- Although many elements of the OBKMS are available as prototypes, we need clear examples to demonstrate the benefits of OBKMS to users and providers.
- Certain critical functions of the OBKMS still need to be developed: linked open data, taxonomic resolution, cross referencing and annotation.

Our key recommendations are:

- Governance system should be lightweight initially enabling Individual institutions to commit to sustaining core components of OBKMS and identification and relationship building with outsourcing partners.
- Funding for production level tools to be encouraged at national and EU level.
- Practical demonstration projects to illustrate the benefits of OBKMS (recommendation from 6.4.1).
- Funding to encourage the missing components: linked open data, taxonomic resolution, cross referencing and annotation.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 17 of 47





Acknowledgements

On behalf of the pro-iBiosphere consortium, we would like thank the following workshop participants for their professional and

enthusiastic participation in the MS24 workshop

Soraya Sierra - Naturalis Lyubomir Penev - Pensoft/ Representing EU BON Donat Agosti - Plazi/ Representing EU BON Anton Güntsch - FUB-BGBM Sabrina Eckert - FUB-BGBM Quentin Groom - BGM Daniel Mietchen - MfN Jeremy Miller - Naturalis Biodiversity Center Chuck Miller - Missouri Botanical garden David Patterson - Global Names project Alan Paton - RBGK Aaike De Wever - Royal Belgian Institute of Natural Sciences Alex Hardisty - Cardiff University Wouter Los - LifeWatch Wouter Addink - Catalogue of life Patricia Mergen - Royal Museum for Central Africa Marc Reynders - Botanic Garden Meise Henry Engledow - Botanic Garden Meise Michel Chauvet - UMR AMAP Vince Smith - Natural History Museum London, Vibrant project Natacha Beau Marion Le Claire Don Kirkup - RBGK Alan Paton - RBGK Charlotte Johns - RBGK

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 18 of 47





Appendix 1. Business requirements currently met

Theme	Requirement	User / Use
Mark up	Provide coarse-grained markup	Biodiversity informatics for content discovery. Markup major text sections and essential nomenclature. Treatments in such form can be aggregated on sites such as the Encyclopedia of Life and Wikispecies.
Mark up	Provide fine-grained markup	Biodiversity informatics for analysis. Detailed markup of species names and their relationships, characters and traits, and materials citations can contribute to a wide range of analyses, indices and applications, including biodiversity catalogs, ecological models, identification tools, space-time visualization of species occurrence records, and monitoring of and feedback of biodiversity collections.
Publication	Scholarly publication	Other publishers, institutional/society journals, researchers, libraries, electronic archives, students, organizations, policy makers
Publication	Data publication	Other publishers, institutional/society journals, researchers, libraries, electronic archives, students, organizations, policy makers, data repositories or centres, data aggregators, text mining institutions
Publication	re-publication of Biotas	Other publishers, institutional/society journals, researchers, libraries, electronic archives, students, organizations, policy makers, data repositories or centres, data aggregators, text mining institutions
Core technical	Development of requirements for digitisation and semantic enhancement of literature.	Consortium, third party digitisation service providers
Core technical	Hosting systems such as treatment repositories, CDM, Scratchpads	All
Core technical	IT infrastructure: development	All such as cross referencing services, reconciliation tools; API architecture, data sharing tools; repositories, annotation infrastructure.
Core technical	IT infrastructure: performance and scalability, archiving	All.
Core technical	Software maintenance	All
Core technical	Semantic enrichment tools	Publishers, researchers
Core technical	Develop markup tools:	Biodiversity informatics centres, taxonomists, researchers.
	Editorial mark up tools; XML-based authoring tools;	Scientific institutions, scientific authors

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August 2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme Coordination and support action FP7-INFRASTRUCTURES-2012-1 Subprogram area INFRA-2012-3.3

Page 19 of 47





	XML-based publishing platform	
Core technical	Software tools to curate content e.g. Taxonomic Editor, SRS, Linnaeus	Providers of data and community annotators
Core technical	Creation of schemas and vocabularies	Institutions, informatics projects, publishers, developers, biodiversity informatics centres
Outreach	Provider help desk function	All users
Outreach	Marketing and demonstration projects to demonstrate value of the OBKMS education, engagement with potential users/ stakeholders	Targeted user community eg. Taxonomists, functional ecologists, conservationists, policy decision makers
Outreach	Fundraising	All
Outreach	Education on Legal issues understanding of legal issues	Suppliers and publishers of data
core technical	Legal issues dealt with- machine readable	Suppliers and publishers of data
core technical	Interoperability between treatments and databases	 Registries Publishers Catalogue of life The Plant List World Flora Online Biodiversity institutions informatics projects (channels=Portals) Taverna based workflows (Indirect = researchers) publishers
core technical	Technical accessibility of linked open data (machine and human)	
core technical	User annotation	
core technical	Issue unique identifiers for e.g. treatments	
core technical	Machine to machine services to data (making available for others to manipulate - interface and documentation) metadata, QC tools, workflows	 Registries Publishers Catalogue of life Plant list World Flora Online
core technical	Research and development of technical core functionality	- Consortium - Anyone publishing plant names

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August

2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme ■ Coordination and support action FP7-INFRASTRUCTURES-2012-1 ■ Subprogram area INFRA-2012-3.3

Page 20 of 47





core technical	Data publication through portals or repositories	- Tool makers - System builders
core technical	Machine to human services to data (making available for others to manipulate - interface and documentation), metadata, QC tools, workflows	 Biodiversity institutions informatics projects (channels=Portals) (Indirect = researchers)
core technical	Consensus taxon classification service	 Consortium Biodiversity institutions Informatics projects (channels=Portals) addressing scientific workflows (e.g. based on Taverna/Kepler), Indirect = researchers, publishers

Page 21 of 47





Appendix 2. Unmet business requirements for OBKMS and actions needed

The following table lists business requirements for OBKMS currently not or only partially met. This is an outcome of the MS24 workshop on Model Evaluation, held in Brussels, June 2014. The business requirements (column 1) are listed against sustainability criteria (columns 2-7). The green cells in the table indicate criteria we currently meet, the pink cells give the criteria yet to be fulfilled.

Unmet or not fully met business requirement	Appropriate Technology	Economic and financial capacity	Policy Support	Institutional and management capacity	Socio-cultural aspects	Environmental protection
1. cross referencing services and infrastructure (identifiers stored and linked	 Plazi prototype for treatments and names Zoobank & GNUB literature identifiers linked to taxonomic acts BioNames and BioStore. Pensoft Taxon Profile These above are working models which work at a small scale for particular bits of data. BioSciCol +iDigBio in the USA BOLD system for voucher specimen, sequences, occurrences and images Global Genome Biodiversity Network References within Plazi treatments can be linked from a text reference to an older treatment (e.g., a synonym or original description) to the stable URI of that treatment on Plazi. This creates links between treatments 	Not sustainable in current paradigm- projects currently short lived.				

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 22 of 47





	and their associated names and taxon status.					
1. cross referencing services and infrastructure (identifiers stored and linked	Need Need to upscale individual components and integrate these Data in repository needs to be open access. Multiple taxon concepts exist. Alternative concepts need to be mapped Reference citations in treatments (prototyped by Plazi) extract relationships among taxonomic concepts from the reference group section of treatments. The result is much like a catalog, that can be used to resolve taxonomic names Bio Sci Coll - what lessons can be learned	Funds needed to provide incentives collaboration Long-term funding for repositories needed	A system is needed for giving credit for creating infrastructure and feeding content flexibility in international projects to involve most appropriate partners inside and outside of EU. mixed model funding: institutional subscription; external funding, grants and projects developing existing infrastructures and initiatives;	Institutional commitment is needed to support and use infrastructure such as integrating OBKMS into institutional IT support and maintenance Staff time and other resources must be allocated	A community needs to be formed, supported and engaged User requirements gathered. Feedback on how these are being met by OBKMS informs future development. Good documentatio n and user interfaces allow broad participation Potential for education and increasing public awareness	Linking to GEOSS allow better monitoring of biodiversity. Potential for environmental education
2.Linked open data	Available: Best practice for Persistent URIs; strategy for extended use of RDF D4.2 in biodiversity domain; ontologies that can be used as controlled vocabularies; treatments in RDF		Bouchout declaration			
2.Linked open data	Needed: use cases; pilots; greater accessibility . and broader scope and interoperability of ontologies;	short term funding to enable transition Business case studies	support for openness and LOD policy ensuring net	Support for LOD provision and use Additional support for	Use outreach to stimulate use of data by other LOD communities e.g Genomics;	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 23 of 47





	infrastructure and interfaces to allow easy use.	demonstrating value of LOD to providers and users	neutrality Additional support for Bouchout Declaration and its implementation	Bouchout Declaration and its implementatio n	interlinkage where commonalitie s in data. Support citizen science and non- institutional systems by providing hooks for long-tail small data	
3. Reconciliation and annotation Service	Available: prototype names reconciliation and resolution via Global Names, iPlant and BioVel FliteredPush and Anosys annotation structures; Bibliography Of Life and RefBank for literature references; Global Registry of Biological repositories GRBIO.org DOIs- literature					
3. Reconciliation and annotation Service	Needed: Production names reconciliation and resolution services via Global names resolvers for treatments; occurrences; comprehensive bibliographic references; specimens Reconciliation of molecular OTUs informally diagnosed and those having a formal name following formal nomenclature.	Funding to develop production level services	funding required for production level services in addition to research prototypes	Support for open data provision and use Additional support for Bouchout Declaration and its implementatio n	be aware of and collaborate with existing communities and reconciliation practices	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 24 of 47





Appendix 3. Business process models

In report <u>D6.3.2.</u>, the current business models (BM) for pro-iBiosphere partners were described using the 'Business Model Canvas' template (<u>http://en.wikipedia.org/wiki/Business_Model_Canvas</u>). These business models allowed us to identify the products and services offered by individual partners and the tools and actions needed to provide them.

These models often comprised of one or more business processes (BP), such as re-publication contained OCR, markup and publication. Some BP's were also duplicated in different business models (e.g. several BMs were identified under which markup is carried out).

In order to evaluate these chosen products and services, common BPs have been extracted and mapped to a single diagram each, onto which the following elements have been added :

- Providers (blue boxes)
- Activities (orange boxes)
- Costs (red boxes)
- Products and services (purple boxes)
- Channels (green boxes)
- Users (yellow boxes)

Evaluation criteria that may be encountered during the development or use of products and services have also been added to each of the diagrams. These criteria are represented by triangles, and have been taken from the risk analysis report (D6.4.1), Report on user feedback (D2.2), Report on benefits to users (D6.2.1), Report on benefits to providers (D6.2.2), Report on cost delivery, efficiency and cost reduction through effective practices (D6.1.1) and Measuring and constraining the costs of delivering services (D6.1.3).



A benefit to users and/or providers



A constraint, indicates uncertainty, but can be mitigated and probably in itself not fatal to a BM

A critical factor, without obvious mitigation, any one of which might be fatal to a BM

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August 2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme Coordination and support action FP7-INFRASTRUCTURES-2012-1 Subprogram area INFRA-2012-3.3

Page 25 of 47





Data Mobilisation

The Mobilisation of data, in the form of OCR and coarse and fine grained markup is a key compliance of the OBKMS and is the starting activity for many other products and services.

Digitisation and conversion of hardcopy to machine readable text



Figure 4. Sub-business model diagram showing two OCR business processes: Large-scale text conversion and Small-scale text conversion.

Large-scale text conversion

Table 1. Evaluation criteria associated with large-scale text conversion

Critical factor	Explanation
A ₅	The OCR third party provider will receive economic benefits of having regular work and income and social benefits such as gaining useful connections.
Benefits to provider	
 6	Users for the digitised text, such as markup providers, will benefit from OCR/ double entry typing to HTML as it will give them access to a reliable, cheap and quick OCR service.
Benefits to user	
<u>_</u> 2	A close and long-term relationship with the third party providers is required to develop specifications (manual – accuracy specifications) and domain knowledge (enables the
Provider Relationship	third party company to do more of the taxonomic the markup). (D6.4.1)
neutronship	Relationships can be maintained through effective outreach and communication. It is important that the chosen OCR providers are willing to take long-term responsibility for this infrastructure component (ideally before the start of the project/proposal)

pro-iBiosphere FP7 Project
Grant Agreement #312848

Page 26 of 47





Costs to Provider	OCR will be outsourced to a third party provider. Costs to the OCR service providers include manpower, scanning of documents, checking output of OCR, correcting errors and retyping. The OCR service is estimated to cost €287 per 600 taxa (D6.1.2).
	However, these costs can be constrained by the use of up-to-date and specifically built equipment. The start-up cost is also counteracted by the large volume of ongoing work received. If the right third party provider is chosen, the staff cost will not be a factor.

Small-scale text conversion

Table 2. Evaluation criteria associated with small-scale text conversion

Critical factor	Explanation
Costs to Provider	Costs are largely the time taken for the taxonomist to perform OCR. Costs are further increased due to working with less advanced technology than available to specialist firms and the taxonomists' (providers) lack of expertise. This process of OCR produces a low volume output and little profit. The provider may be unable to find the funds to mobilise legacy literature this way and therefore lacks benefits for performing the service.
	There are few benefits to the provider performing OCR this way. This process is usually done out of goodwill by the provider, making the document and data usable to many. There is little that can be done to reduce costs and increase benefits. Outreach demonstration projects may be a way of increasing the providers understanding of the OCR to PDF process.
Quality of data	There is a risk that small scale user needs will not be met because the digitised text is inaccurate due to processing errors. We know from our experience that the quality of OCR and PDF can vary and some approaches to markup are prone to propagate these errors. If this small scale process of OCR to PDF is used, users may be discouraged to use the system due to the unknown quality of the data. This high risk factor is therefore considered as a serious critical factor (D6.4.1).
	Mitigation steps for this risk include the establishment of appropriate standards for text conversion, the provision of annotation tools to help improve the quality of the data, performing quality control of critical index terms carried out and the promotion of a collaborative platform where authenticated users can correct OCR documents rendered on web pages (Further mitigation steps can be found in D6.4.1).

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 27 of 47





Markup of legacy texts



Figure 5. Sub-business model diagram showing two business processes for markup of legacy texts

Small-scale markup of legacy texts

Table 4. Evaluation	criteria a	ssociated with	desktop	authoring tools
----------------------------	------------	----------------	---------	-----------------

Critical factor	Explanation
12 Benefits to user	Users will receive benefits including open accessibly to quality data, a reduction in time to find and extract information and easier to edit data.
Funding & Economic viability	Funding for digitisation projects involving legacy literature can be difficult to secure, although there have been notable successes (D6.4.1). Steps can be followed to reduce this risk, one of which is to seek funding for particular bodies of literature, e.g. African Floras. Another factor that needs consideration is that small scale markup will not be as economically viable as large scale markup (D6.4.1).

pro-iBiosphere FP7 Project
Grant Agreement #312848





	Due to the accessibility of the markup tool, a large number of providers can perform markup at any one time, ensuring a high volume output. However, these providers will not be as efficient as markup specialists, nor can they easily markup using specialised templates.
	This risk can be reduced by piloting the system, adjusting tools according to customer feedback, considering cost savings (e.g. optimising outsourcing vs in house domain specific tasks), exploring income generation (e.g. grants, premium services, consultancy; provider subscription for specialist services), explaining nature of value to providers/ suppliers and capturing cost information accurately to illustrate savings if they exist.
Costs to users - Quality of data	There is a risk that user needs will not be met because the mobilized data is inaccurate due to processing errors. We know from our experience that the quality of OCR and PDF can vary and markup may propagate these errors. Users may be discouraged to use the system due to the unknown quality of the data the taxonomist markup providers will produce (D6.4.1).
	There is also a worry that information concerning new taxonomic circumscriptions, addition of new species and changes in names could quickly become outdated. Taxonomic changes are inevitable, and updates are usually not performed due to: (i) financial constraints (natural history museums/botanic gardens have no funding available for this activity). (ii) format issues (many works (Floras, Faunas, Mycotas) are only available as hard copies and updates will therefore involve re-publishing the whole work). (iii) copyright issues (the institution coordinating the Flora signed a contract with a publisher, and hence cannot re-publish online without their authorisation). (D6.4.1)
Costs to users - Accessibility of data	The accessibility of the data and information is critical. Markup tools work with legacy literature, which is mostly available in hard-copy format. This format does not allow easy and immediate access to the information by computers or people, so therefore will have to go through the process of OCR (D6.4.1).
Costs to users loss of focus	There is a worry that the tools will fail to meet user needs, because the requirements of end-users are subordinated as a result of focusing on the requirements of the third party funding (Government, charities, commercial parties). Loss of focus on user needs is a widely reported concern with 3rd party funding models which may favour cutting edge research over production development of tools (D6.4.1). Steps to mitigate this risk include to ensure that users' views are well represented in funding bids and in the design of tools, and to provide strong coordination of all stakeholders (users, intermediaries, partner providers) (D6.4.4).
Costs to Provider	There is a high cost associated with the development and maintenance of the tools as software developers' time and equipment needed is costly. This high development cost leads to problems with economic scalability. Software developers may be unable to find the funds to develop the tools.
	Mobilisation of legacy literature using is performed by taxonomists who receive no funding to provide the service. Inexperienced taxonomists first have to learn how to use the tools, then perform markup; this process is also often time consuming and costly (D6.4.1). The taxonomists also receive few benefits for providing this service, and it is usually done through goodwill.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 29 of 47





Large-scale markup of legacy texts

Table 3. Evaluation criteria associated with Large-scale Markup of legacy texts

Critical factor	Explanation
10	Economies of scale relating to proportionate reductions in set-up and staff training costs.
Benefits to	
provider	
1 1	Benefits received by users include, open accessibly to quality data, a reduction in time to find and extract information and easier to edit data.
Benefits to user	
Costs to provider	Biological institutions will incur many costs associated with performing markup. Startup activities include setting up the markup service and tools and management and training of the staff that will carry out the service. It was discovered that, not including software costs, the highest costing start up activity for markup would be 'management and training', which costs €4,108 and equates to 51% of the total cost for setting up markup of floristic text (D6.1.2, case study 'Markup of RBGK's African Floras'). However, start up activities for the markup of prospective text can be reduced by up to 75% if a template/ set format has already been developed for the markup of legacy literature of a similar type. (D6.1.2)
	There are high costs associated with ongoing markup activities. These consist of markup and three sub activities; transforming, proofing and markup. The main ongoing activities identified in Figure 5 and Figure 6 due to their high cost and high time consumption are 'markup', which costs €301 per 600 taxa and 'proofing', which costs €444 per 600 taxa (D6.1.2, case study 'Markup of RBGK's African Floras'). These costs are reduced by 75% when marking up prospective text, as standards and markup styles will automatically be applied when marking up the document. (D6.1.2)
2	Some partners may be unable to find and secure the funds to mobilise legacy literature.
Funding	Steps to overcome this critical factor include making use of existing structures such as BHL. and seeking funding for particular bodies of literature, e.g. African Floras. (D6.4.1)
Costs to users - accessibility of legacy data	The accessibility of the data and information is critical. This form of markup works with legacy literature, which is mostly available in hard-copy format. This format does not allow easy and immediate access to the information by computers or people, so therefore will have to go through the process of OCR (D6.4.1).
	In order to mitigate this risk, it is important that a useful format for the data and information is used (i.e. from hard-copy to digital), to avoid creating new data which cannot be accessed and to identify key user requirements to refine format. Partner institutes could share the effort to digitise legacy literature (hard copies from Institutional libraries) and waive copyright (when it applies). This could also increase the amount of legacy literature being digitised.(D6.4.1)
Costs to users - outdated information	Information concerning new taxonomic circumscriptions, addition of new species, changes in names could quickly become outdated. Taxonomic changes are inevitable. Updates are usually not performed due to: (i) financial constraints (natural history museums/botanic gardens have no funding available for this activity). (ii) format issues (many works (Floras, Faunas, Mycotas) are only available as hard copies, and updates will therefore involve re-publishing the whole work. (iii) copyright issues (the institution coordinating the Flora signed a contract with a publisher, and hence

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 30 of 47





	cannot re-publish online without their authorisation). (D6.4.1)
	To ensure that this critical factor does not prevent sustainable markup, systems must be in place for the curation of data, for example readily available taxonomic authority files and annotation systems; and taxonomists must have access to them. Simple and responsive feedback systems for users are needed. Natural history institutions/ botanic gardens should agree on what tools they want to support and host at their institutions (e.g. Scratchpads, EDITor), so that taxonomics are encouraged to use them and are assured that the system will be sustained. Develop taxonomic resolution services that "validate" names against authoritative nomenclatural and taxonomic sources, such as IPNI, Index Fungorum, ZooBank, and for such services to provide automatic updates. (D6.4.1)
Costs to users - Loss of focus	Failure to meet user needs because the requirements of end-users are subordinated as a result of focusing on the requirements of the third party funding (Government, charities, commercial). Loss of focus on user needs is a widely reported concern with 3rd party funding models which may favour cutting edge research over production development of widely accessible, easy to use tools. (D6.4.1)
	To mitigate this risk we must ensure that users' views are well represented in funding bids and in the design of OBKMS products and services and to provide strong coordination of all stakeholders (users, intermediaries, partner providers). Increased availability of high granular data will help understand how users wish to interact with that data.





Prospective markup of texts



Figure 6. Sub-business model diagram showing two business processes for prospective markup of text

Online XML authoring tools for prospective markup of text

Critical factor	Explanation
12	Researchers will receive benefits including open accessibly to quality data, a reduction in
15	time to find and extract information and easier to edit data.
Benefits to users	
1 4	Authors will benefit from rapid and efficient dissemination of their work
Benefits to users	
<u></u> 8	Small scale markup will not be as economically viable as large scale markup (D6.4.1)
Funding &	
Economic	

Table 5. Evaluation criteria associated wi	th integrated online XML a	authoring tools
--	----------------------------	-----------------

pro-iBiosphere FP7 Project
Grant Agreement #312848

D6.4.3 Alternative business requirements and scenarios for sustainable Open Biodiversity Knowledge Management, 31 August 2014; Task Leader: Don Kikup, RBGK. 7th Framework Programme ■ Coordination and support action FP7-INFRASTRUCTURES-2012-1 ■ Subprogram area INFRA-2012-3.3 Page 32 of 47





viability	
9 Costs to users - Quality of data	New taxonomic circumscriptions, addition of new species and changes in names could become quickly outdated. Taxonomic changes are inevitable. Updates are usually not performed due to: (i) financial constraints (natural history museums/botanic gardens have no funding available for this activity). (ii) format issues (many works (Floras, Faunas, Mycotas) are only available as hard copies, and updates will involve re-publishing the whole work, (iii) copyright issues (the institution coordinating the Flora signed a contract with a publisher, and hence cannot re-publish online without their authorisation). (D6.4.1)
Costs to tool provider	There is a high cost associated with the development and maintenance of XML authoring Tools, as software developers time is costly. A huge initial investment to create these specific tools is also needed.





Data Dissemination & Discoverability

The OBKMS will make its biological data discoverable through two core service areas, 'Advanced publishing' and 'Hosting systems and repositories'.

Advanced Publishing



Figure 7. Sub-business model diagram showing business processes for three OBKMS publishing services, 'republication', 'scholarly publication' and 'integrated text and data publication'.

<u>Re-publication</u>

Table 6. Evaluation criteria associated with Re-publication

Critical factor	Explanation
13 Benefits to	Benefits to the provider include regular work and funding for the publisher, increased accessibility to high quality legacy literature.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 34 of 47





provider	
14 Benefits to user	Increased usability of data (D6.2.2).
Costs to provider	The high cost associated with re-publishing poses as a critical factor. The provider incurs high start up activity costs which include exporting atomised content (€2844, per start up) and getting identifiers (€7128, per start up). Other costs such as the development of fine granularity markup tools would also be included (D6.1.2). High ongoing activity costs consist of enriching the text semantically (€2053 per 600 published taxa), database extracted data (€1800 per 600 taxa) and to publish as a semantically enhanced HTML edition (€1800 per 600 taxa). OCR and markup service would have to be used prior the re-publishing process. Further re-publishing activities and costs can be found in D6.1.2.
Market background	Re-publication is a newly created service, and adoption by partners may be slow. Electronic publication/digital media are well established in some partner organisations, but others still maintain traditional publishing activities. Many formerly loss-making publications (e.g. Floras) are now cost neutral, offering services such as print on demand for users without routine access to computers or internet. (D6.4.1)

Integrated text and data publication

Table 7. Evaluation criteria associated with integrated text and data publication

Critical factor	Explanation
15	Will receive a number of benefits for providing this service which include steady reliable work and income, provider acknowledgement and increased usability of data (D6.2.2).
Benefits to the provider	
16 Funding	Value added service
Costs to provider	There are high costs associated with data publication. Main start up activities include the creation and maintenance of article templates, export on online taxon profiles and the export of atomized content. The most costly activities include exporting data to semantically enhanced HTML (€7128 per startup) and the adoption and maintenance of standards (€5364 per startup). Another main cost is the development and maintenance of authoring tools, these costs are discussed under prospective markup, Section 3.2.3. (D6.1.2) Ongoing activities involved in data publishing are adoption of standards, creation and maintenance of data imports and the export of Darwin core archives, online taxon profiles and atomized content. The most costly ongoing activity is copyediting, €1027 per 600 taxa (D6.1.2). Benefits to the provider include regular work and funding for the publisher, increased accessibility to high quality legacy literature.

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 35 of 47





Costs to user – Loss of focus	Failure to meet user needs because the requirements of end-users are subordinated as a result of focusing on the requirements of the third party funding (Government, charities, commercial). Loss of focus on user needs is a widely reported concern with 3rd party funding models, which may favour cutting edge research over integrated text and data publication of legacy and prospective literature (D6.4.1).
	Steps to mitigate this risk include to ensure that users' views are well represented in funding bids and in the design of OBKMS products and services, and to provide strong coordination of all stakeholders (users, intermediaries, partner providers) (D6.4.4).
7 Market background	Adoption by partners of integrated text and data publication with in house electronic publication services may be slow. Electronic publication/digital media are well established in some partner organisations, but others still maintain traditional publishing activities. Many formerly loss-making publications (e.g. Floras) are now cost neutral, offering services such as print on demand for users without routine access to computers or internet. (D6.4.1)
	This is also a risk that new entrant electronic publication services compete with OBKMS- comliant offers. New entrant e-journals include Journal of European Taxonomy, Frontiers In. (D6.4.1)
Cost of the review process	Cost of review process outstrips funding. Publishers find suitable reviewers difficult to recruit and this is time consuming and hence costly; however an open OBKMS model offers possibilities for more open review by having wider accessibility as implemented in the Biodiversity Data Journal. (D6.4.1)

Scholarly publication

Table 8. Evaluation criteria associated with Scholarly publication

Critical factor	Explanation
17 Benefits to provider	Will receive a number of benefits for providing this service, these include steady reliable work and income, provider acknowledgement and increased usability of data (D6.2.2).
18 Benefits to user	University students benefit from the data being indexed and made discoverable, browsable and searchable through biodiversity infrastructures (e.g., the Global Biodiversity Information Facility (GBIF), Dryad and others). Researchers/taxonomists from biological institutions can disseminate information more effectively through interlinked datasets, metadata and respective data papers. Users such as policy makers may prefer this more common format when researching for biological information (D6.4.2).
19 Funding	Value added service

pro-iBiosphere FP7 Project
Grant Agreement #312848

Page 36 of 47





Costs to provider	High costs are associated with scholarly publication. High cost start up activities include the implementation of a markup tool (€72,000, 74% of total start up cost), the setup of journal's website and associated services (€9000, 9% of total start up cost) and exporting online taxon profiles (€4464, 5% of total start up cost). (D6.1.2). Many ongoing activities provided by Pensoft are associated with scholarly publication, these include the export of the publication into XML, HTML and PDF format, linking to external resources, PR activities etc. The most costly ongoing scholarly publication activities include markup and the development of markup tools (€446, 22% of ongoing costs) and the export of online taxon profiles (€357, 18% of ongoing costs). Ongoing activities 'copyediting', 'layout' and 'proofreading and inserting corrections' also hold big costs (D6.1.2). Benefits received by Pensoft for providing this publishing service include recognition, credibility and increased reputation. (D6.1.2) Benefits to the provider include regular work and funding for the publisher, increased accessibility to high quality legacy literature.
10 Costs to user – Loss of focus	Failure to meet user needs because the requirements of end-users are subordinated as a result of focusing on the requirements of the third party funding (Government, charities, commercial). Loss of focus on user needs is a widely reported concern with 3rd party funding models, which may favour cutting edge research over scholarly publication of legacy and prospective literature (D6.4.1).
	Steps to mitigate this risk include to ensure that users' views are well represented in funding bids and in the design of OBKMS products and services, and to provide strong coordination of all stakeholders (users, intermediaries, partner providers) (D6.4.4).
Market background	Adoption of scholarly publication by partners with in house electronic publication services may be slow. Electronic publication/digital media are well established in some partner organisations, but others still maintain traditional publishing activities. Many formerly loss-making publications (e.g. Floras) are now cost neutral, offering services such as print on demand for users without routine access to computers or internet. (D6.4.1)
	This is also a risk that new entrant electronic publication services compete with OBKMS- compliant offers. New entrants e-journals include Journal of European Taxonomy, Frontiers In. (D6.4.1)
Cost of the review process	Cost of review process outstrips funding. Publishers find suitable reviewers difficult to recruit and this is time consuming and hence costly; however an open OBKMS model offers possibilities for more open review by having wider accessibility, e.g., as implemented in the Biodiversity Data Journal. (D6.4.1) As well as staff man power costs incurred by the publisher, scholarly publication also requires a payment by the author, this covers subscription to the journal, submission of their data to a repository, printing and an open access fee. Some institutions and many funders have now budget lines for author paid submission. However, this requires budget reallocation in some institutions. (D6.4.1)

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 37 of 47





Hosting systems



EDIT platform for cybertaxonomy

Table 9 Evaluation criteria associated with the EDIT Platform for cybertaxonomy

Figure 8. Sub-business model diagram showing activity routes for the Edit platform business process

Critical factor	Explanation
A ₂	A platform to store and edit information.
Benefits to the	
data provider	
1 3	Easy access data.
Benefits to the	
users	
1	High start up costs associated with the Edit Platform. Startup activities include CDM-
Costs to provider	Server and data portal setup, configuration of roles and rights and floristic/ faunistic data
	imports. The most costly activity is the CDM setup which costs €33900. (D6.1.2)
	Ongoing activities include 'server maintenance' and 'CDM model changes in development
	and implementation'. The high cost ongoing activities are 'Fulfillment of special
	requirements' and 'Server hardware and data storage costs'. This service also requires the
	use of a technical helpdesk, these costs are discussed in Section 5.3. (D6.1.2)

pro-iBiosphere FP7 Project
Grant Agreement #312848

Page 38 of 47





Outreach

Marketing and demonstration projects



Figure 9. Sub-business model diagram showing the Marketing and demonstration project outreach business process. (The sub-business model diagram has have been divided into key service providers (Blue), activities (orange), costs (red), markup service (purple) and user (green). Red and orange critical factor triangles have also been added)

Table 10. Evaluation criteria associated with marketing and demonstration projects

Critical factor	Explanation
7	Marketing will help to increase trust within the OBKMS (D6.2.1)
Benefits to user - Increased trust	

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 39 of 47





8 Benefits to user – Increase capability	Demonstration projects will help to increase the skills of the users, and their capability to understanding and interpret results (D6.2.1) For example Taxonomists will benefit when they learn about the possibilities of the system, how the system targets the user group and when they receive face to face help with any tools they may want to use.
Benefits to user – understanding needs	Face to face discussions with users to understand their needs will increase the providers' understanding of the needs of others across disciplines
10 Benefits to user – multiple user groups targeted	Marketing will increase the number and diversity of users, as user communities such as policy makers and citizen scientists will understand how the services within the system suit their needs. Consortium members, suppliers of data, and publishers will also use this service ('Model Evaluation workshop' MS24 and deliverable <i>D6.4.2, case study 'Outreach products and services'</i>).
L Costs to provider	There are large costs to the provider associated with start up and ongoing outreach products and services. The highest cost start up activities are 'Niche exploration/stakeholder mapping' €36998, 'Customer business requirement and defining benefits' €36998 and 'User group setup' €37445. The highest costing ongoing activities are 'Customer business requirement and defining benefits' €30,441, 'Continued additional user engagement' €16,111 and 'Technical requirements including mock ups and testing' €16,111. (D6.1.2) In return, the provider will receive a large amount of funding and increased connections.
Costs to user – Loss of focus	A high risk to users of outreach is that the demonstration projects will fail to meet user's needs because the requirements of end-users are subordinated as a result of focusing on the requirements of the third party funding (Government regulators and commercial companies). Ways to mitigate this risk include to ensure that users' views are well represented in the design of OBKMS products and services and to provide strong coordination of all stakeholders (users, intermediaries, partner providers).(D6.4.1)
Costs to user – Loss of user relationship	A high risk to both user and providers of the outreach services is that user relationships will deteriorate because new electronic networks (such as social media services) are presented as a substitute for existing user specific (face to face/ phone) provider-user relations, but these prove inadequate. To mitigate this risk, providers should maintain and further strengthen their traditional networks with users and to specifically design systems to low cost and low bandwidth. (D6.4.1)

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 40 of 47





Fundraising



Figure 10. Sub-business model diagram showing the fundraising outreach business process.

Critical factor	Explanation
Increased funding	Increased funding to support the OBKMS
12 Increased connections	Increased connections
Costs to provider	Activities: Activities and costs for fundraising are unknown
A Meeting users needs – Loss of	A high risk to both user and providers of the outreach services is that user relationships will deteriorate because new electronic networks (online social media) are presented as a substitute for existing (face to face/ phone) provider-user relations , but these prove

pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 41 of 47





user relationship	inadequate. To mitigate this risk providers should maintain and further strengthen their
	traditional networks with users and to specifically design systems to low cost and low
	bandwidth. (D6.4.1)





Provider help desk function



Figure 11. Sub-business model diagram showing the provider help desk and legal education outreach business process.

Critical factor	Explanation
13	Benefits include a wider access to biological data and increased collaboration between data providers and users.
Costs to providers	The small individual activities of the helpdesk are unknown. The cost to the provider for providing a helpdesk per service is €4777. This large cost of the helpdesk, greatly depends on how the helpdesk operates, whether users access it via phone-call, email, social media or whether it should be an automated service. Automated service would save money, however may not provide potential users with satisfactory answers.
<u> </u>	A high risk to users of outreach is that the helpdesk will fail to meet user needs because

pro-iBiosphere FP7 Project ■ Grant Agreement #312848





Costs to user – Loss of focus	the requirements of end-users are subordinated as a result of focusing on the requirements of the third party funding (Government, charities, commercial). Ways to mitigate this risk include to ensure that users' views are well represented in funding bids and in the design of OBKMS products and services, and to provide strong coordination of all stakeholders (users, intermediaries, partner providers)(D6.4.1)
6 Costs to user – Loss of user relationship	A high risk to both user and providers of the outreach services is that user relationships will deteriorate because new electronic networks (remotely accessed, social media) are presented as a substitute for existing provider-user relations (face to face/ via phone), but these prove inadequate. To mitigate this risk, providers should maintain and further strengthen their traditional networks with users and specifically design systems to require low investment and low bandwidth from the users. (D6.4.1)

Page 44 of 47





Appendix 4 Pre- and Post-evaluation business process diagrams Figure 1. Partners' current business processes for OBKMS business requirements



pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 45 of 47





Figure 2. Post evaluation, preferred business processes for OBKMS business requirements



pro-iBiosphere FP7 Project ■ Grant Agreement #312848

Page 46 of 47





Two types of data will be submitted to the OBKMS, legacy and prospective. The first step to making data and information accessible is though data mobilisation. Legacy literature first goes through the digitization process (OCR), that will convert the document into a HTML file, making further processing by computers possible (1). The newly digitised legacy literature, along with prospective literature, will then go through the process of large scale markup (2). Dissemination is the second area on which the OBKMS will focus. Data will be made discoverable through databases (3) and portals such as the Edit platform (4). Advanced publication is another way of disseminating data and making mobilized data discoverable. The OBKMS will support three types of publication, republication (5), data publication (6) and scholarly publication (7). Following mobilisation, discovery and dissemination of the data, curation is an ongoing issue that needs to be addressed. Lastly the OBKMS products and services within the outreach area will contribute towards all other products and services, providing funding, marketing and advice (8).

Page 47 of 47