CReATIVE-B

Coordination of Research e-Infrastructures Activities Toward an International Virtual Environment for Biodiversity

Guidelines for Interoperability for Biodiversity and Ecosystem Research Infrastructures

Document identifier:	D3.2 Guidelines for Interoperability for Biodiversity and Ecosystem Research Infrastructures
Date:	**26 September 2014**
Activity:	**Work Package 3, Tasks 3.3 & 3.4**
Lead Partner:	**CU**
Document Status:	**Issue 1.0**
Dissemination Level:	**PUBLIC**
Document Link:	*<to be added at time of publication>*

ABSTRACT

Deliverable D3.2 "*Guidelines for Interoperability*", prepared on the basis of available information at the time of writing, is the output of CReATIVE-B tasks T3.3 and T3.4, which aims to "*Prepare guidelines for interoperability (for biodiversity and ecosystem research infrastructures)*". It provides a resume of the conclusions about the status and achievability of interoperability existing between the surveyed e-Infrastructures. It highlights the currently known obstacles and makes suggestions for overcoming these. Using a typical use-case drawn from contemporary work on Essential Biodiversity Variables (EBV) the document defines some typical scenarios of interoperability that can be supported by several or all e-Infrastructures. The use-case is mapped to the most likely scenario but further work is required at each of 3 levels of interoperability (applications, service logic, resources) to illustrate how each research infrastructure may support the use case. Building on Deliverable D3.1, on the existing similarities and differences between participating research infrastructures, Deliverable D3.2 aims at presenting a set of guidelines for overcoming obstacles to interoperability. It advises on a roadmap for medium-term (5-7 years) convergence towards worldwide technical interoperability of biodiversity and ecosystem research infrastructures. Thus, it forms a solid knowledge basis for recommendations on resolution of interoperability in the medium to long-term (deliverable D3.3).

1. COPYRIGHT NOTICE

2. DELIVERY SLIP

	Name	Partner/Activity	Date
From	Alex Hardisty	CU	26 Sep 2014
Reviewed by	**Moderator:** **Reviewers:**		
Approved by			

3. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1.0	26 Sep 2014	Issue 1.0	A.Hardisty / CU D.Manset / Gnubila

4. APPLICATION AREA

This document is a formal project deliverable for the European Commission, applicable to all members of the CReATIVE-B project, beneficiaries and Joint Research Unit members, as well as its collaborating initiatives and projects.

5. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors.

6. TERMINOLOGY

This document is based on the CReATIVE-B official glossary[1].

7. PROJECT SUMMARY

Research infrastructures supporting environmental sciences are increasingly crucial for advanced research and for contributing to the solution of global environmental problems. This holds specifically for our living natural environment, biodiversity and ecosystems. Understanding these systems however requires access to global data sets and consequentially the integration of several heterogeneous data sources. Fast running interoperable capabilities to analyse these data and to test new computationally demanding models of processes of change are thus essentials. In this context, the emerging LifeWatch[2] infrastructure for biodiversity and ecosystem research, funded by the European Strategy Forum on Research Infrastructures (ESFRI[3]) has to cooperate with international partners to: enhance global data facilities; to address bottlenecks for achieving interoperability; and to identify and implement common solutions.

Accompanying this long-term objective, several key initiatives worldwide have already expressed their interest and are committed to cooperate with the strategic EU FP7[4] CReATIVE-B[5] project, towards the development and governance of an international virtual environment for biodiversity. The immediate objective is to define a roadmap for interoperability on the technological level, on the governance level and on the interrelation with the scientific communities using the research infrastructures (RIs). The project will therefore be a catalyst for worldwide collaboration in this field by supporting and initiating coordination activities of these RIs. The ultimate goal of this collaboration is to support the Global Earth Observation System of Systems (GEOSS)[6] initiative. At the same time, the international outreach of LifeWatch can lead to further international collaboration(s) on interoperability of these infrastructures to even better serve research communities worldwide.

8. EXECUTIVE SUMMARY

In the biodiversity science domain, research communities work internationally with data originating from all over the world, stored and made available through a wide range of tools, services and mechanisms. The researchers are scattered in many regions of the world so there is a need for fast and reliable aggregation of the data and tools into large and interoperable research infrastructures supporting easy and accurate use of the capabilities.

This is what we refer to as an international virtual environment for biodiversity. Capabilities of such an environment include: sensors and sensor networks deployment, digitizing of biological specimens, ground level and remote observations, fast DNA sequencing facilities, interoperability and data sharing, data discovery and knowledge development, computation for modelling and simulation, virtual laboratories and e-services.

In CReATIVE-B, the challenge is thus to bring various regional, national and international community initiatives together and to add scientific value for users. This is not a merger of infrastructures as these are inherently distributed. Interoperability of data, tools and services and the infrastructure's management have to be achieved to promote advanced data mining and knowledge development in a coordinated and international setting.

CReATIVE-B thus is implementing a challenging work plan thereby organizing and chairing these activities. In the context of its work package 3 "*Technological Interoperability*", the present deliverable D3.2 entitled “*Guidelines for Interoperability*”, prepared on the basis of available information at the time of writing, is the output of tasks T3.3 and T3.4, which aims to “*Prepare guidelines for interoperability (for biodiversity and ecosystem research infrastructures)*”. It highlights the currently known obstacles. Using a typical use-case drawn from contemporary work on Essential Biodiversity Variables (EBV) the document defines some typical scenarios of interoperability that can be supported by several or all e-Infrastructures. The use-case is mapped to the scenarios at each of 3 levels of interoperability (applications, service logic, resources) to illustrate how each research infrastructure may support the use case.

The following Figure 1 illustrates the conceptual framework into which it inscribes together with previous and subsequent deliverables (to be produced). This document constitutes the core of this framework, where scenarios for interoperability are described and illustrated for the typical use case, and general guidelines on interoperability are derived. Guidance is given on application of relevant Standards.

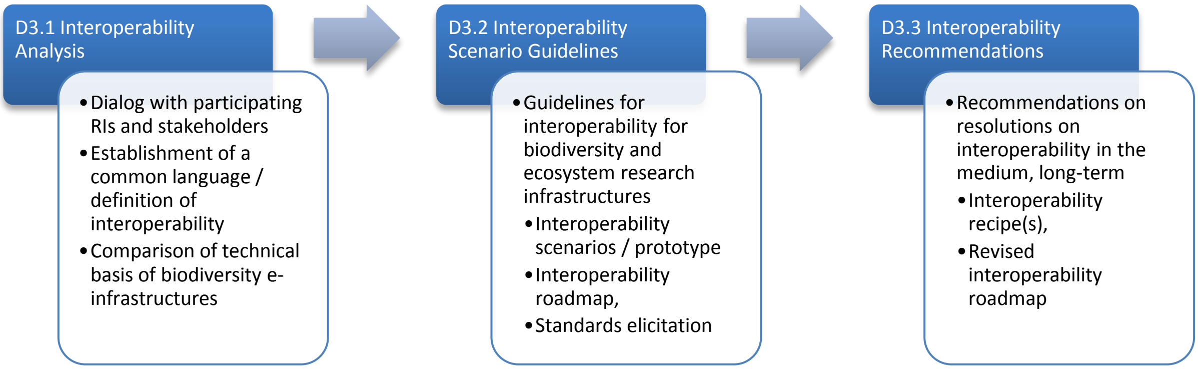

Figure 1. Deliverables conceptual framework

Deliverable D3.2 builds on the interoperability analysis contained in Deliverable D3.1, describing the existing similarities and differences between participating research infrastructures. Deliverable D3.2 aims at presenting a set of guidelines for overcoming obstacles to interoperability and advises on a roadmap for medium-term (5-7 years) convergence towards worldwide technical interoperability of biodiversity and ecosystem research infrastructures. From these guidelines, recommendations on resolution of interoperability in the medium to long-term can be determined i.e., deliverable D3.3 “*Interoperability Recommendations*”. D3.3 will be shared with collaborators and external stakeholders to promote awareness and to influence funding agencies, so laying the foundations of other initiatives developing the targeted full interoperability.

The following sections reprise the conclusions of deliverable D3.1; examine some possible ways to overcome existing barriers to interoperability; introduce contemporary work on Essential Biodiversity Variables (EBV) as an interoperability use-case to further deepen the analysis; and elaborate typical scenarios of interoperability that show how the use-case can be supported across multiple heterogeneous but complementary research infrastructures.

This page intentionally blank.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Background

In Europe, the ESFRI LifeWatch research infrastructure will support the biodiversity and ecosystem science community to carry out the research required to address the huge gaps we face in understanding life on Earth. Pressing problems such as measuring the impact of environmental change on biodiversity, predicting/preventing further biodiversity loss arising from human activities, and sustaining ecosystem services are key foci in such science.

LifeWatch will construct and bring into operation the facilities - hardware, software and governance structures - for all aspects of biodiversity and ecosystems research. It will consist of: facilities for data generation and processing; a network of observatories; facilities for data integration and interoperability; virtual laboratories offering a range of analytical and modelling tools; and a Service Centre providing specific services for scientific and policy users, including training and research opportunities for young scientists. The infrastructure has the support of all major European biodiversity research networks and stakeholders. Its innovative design supports scientists to enter new research areas with large-scale data resources, advanced analytical and modelling capabilities with computational power. LifeWatch will not only serve the scientific community, but will also be an essential tool for local and global policy makers (such as IPBES[7]) in the understanding and the rational management of our ecosystems. It will form the European contribution to GEO BON[8], the biodiversity observation component of GEOSS. Aspects of the LifeWatch vision are being piloted in the EC FP7 funded Biodiversity Virtual Laboratory (BioVeL) project[9].

Several other infrastructures with complementary aims are operational or under construction in other parts of the world. In addition to LifeWatch and GEOSS / GEO BON initiative, a significant subset of these contribute to the present analysis, as is listed herein below:

In Australia, the Atlas of Living Australia (ALA)[10] contains information on all known living species in Australia, aggregated from a wide range of data providers: museums, herbaria, community groups, government departments, individuals and universities.

In Brazil, the Reference Centre on Environmental Information (CRIA[11]) aggregates and disseminates biological information of environmental and industrial interest, as a means of organising the scientific and technological community of the country towards conservation and sustainable use of Brazil's biological resources. More specifically, the SiBBr[12] and speciesLINK[13] projects are subject to the present analysis.

In China, the Germplasm Bank of Wild Species (GBoWS)[14] is one of the 11 large research infrastructures managed by the Chinese Academy of Sciences (CAS). CAS is China's government organisation, founded in Beijing on 1 November 1949, as the nation's highest academic institution in natural sciences and its supreme scientific and technological advisory body, and national comprehensive research and development centre in natural sciences and high technologies.

DataONE	In the USA, the Data Observation Network for Earth (DataONE)[15] is developing the future foundations for environmental sciences with a distributed framework and sustainable e-infrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered earth observational data.
GBIF	Through a global network of countries and organizations, the Global Biodiversity Information Facility (GBIF)[16] encourages free and open access to biodiversity data, and promotes and facilitates the mobilization, access, discovery and use of information about the occurrence of organisms over time and across the planet.
SANBI Biodiversity for Life	The South African National Biodiversity Institute (SANBI)[17] leads and coordinates research, monitors and reports on the state of biodiversity in South Africa. Providing biodiversity information is central to SANBI's mandate and it does this by providing several databases and other resources developed by SANBI and its partners.

In light of the ongoing construction of such platforms, the European FP7 funded CReATIVE-B international coordination and support action project aims to promote the maximum level of interoperability between these infrastructures, with the ultimate objective of securing better cooperation of the global scientific communities dealing with the construction, operation and use of large infrastructures and facilities for biodiversity and ecosystems research. To do so, CReATIVE-B is arranging a series of workshops and enabling the collaboration of working groups in Europe, USA, China, South Africa, Brazil and Australia, among others, to encourage the exchange of key technical information to direct the developments of these infrastructures towards full interoperability, and to promote awareness on the timeliness of having international efforts converging.

Thus, the first task of CReATIVE-B's Work Package 3 (WP3) has been to make a comparison of the technical approaches adopted by the different infrastructures with the aim of developing better understanding of the opportunities and solutions for achieving greater interoperability between them. This has been documented previously in Deliverable D3.1 "*Comparison of Technical Basis of Biodiversity e-Infrastructures*".

Subsequent tasks have been concerned with identifying current obstacles to interoperability; determining a contemporary and typical use case to drive interoperability analysis; and elaborating typical scenarios of interoperability that show how the use-case can be supported across multiple heterogeneous research infrastructures.

The present document elaborates these aspects and shows how the use-case can be mapped onto the scenarios at each of the 3 levels of interoperability (i.e., applications, service logic, resources) to illustrate how each research infrastructure may support the use case. This provides a basis for deriving more general guidelines.

1.2 Purpose

This document reports and structures the analysis carried out thus far to identify current obstacles to interoperability between different research infrastructures for biodiversity and ecosystems science. It describes a contemporary and typical use case that has been used to drive interoperability analysis; and it elaborates typical scenarios of interoperability that show how the use-case can be supported across multiple heterogeneous research infrastructures.

It shows how the use-case can be mapped onto the scenarios at each of the 3 levels of interoperability (i.e., applications, service logic, resources) to illustrate how each research infrastructure may support the use-case. It derives more general guidelines for future interoperability.

This work has been carried out by WP3 "*Technology and Interoperability*", and more precisely within the tasks entitled "*T3.2 Technical workshop on obstacles to interoperability*" and "*T3.3 Prepare guidelines for interoperability*". These tasks had the following objectives (conflated and abridged extract from the project description of work): "*Organising a workshop to facilitate exchange of technical expertise at the global level. Based on workshop outcomes, prepare a set of guidelines for overcoming obstacles to interoperability and advise on a roadmap for medium-term (5-7 years) convergence towards worldwide technical interoperability of biodiversity and ecosystem research infrastructures*".

In the following sections, the reader will gain understanding of the approach that has driven the work. The approach establishes the major dimensions of interoperability, which the concerned development teams should address and as such is intended to become a useful reference throughout the future phases of biodiversity e-infrastructures development.

In the following sections the reader will be reminded of the conclusions of D3.1 before being introduced to contemporary work on Essential Biodiversity Variables (EBV). One of these is used as a use-case to drive the interoperability analysis reported here. Subsequent sections, elaborate on typical scenarios of interoperability to show how the use-case can be supported across multiple heterogeneous research infrastructures.

Note: It must be noted that the present document will attempt to avoid repeating information that can be found elsewhere.

1.3 Reference documents

Prior to reading this report, the reader should be familiar with additional documents, deliverables and information sources produced within the various initiatives, which have or are considered to potentially impact on the future developments of interoperability. Details of relevant websites and documents are given at the beginning of each section of the present document. Footnotes provide additional information.

Table 1. Reference documents

Document Title	Description	Background
The CReATIVE-B project website and background materials[18] **In particular, D3.1 "*Comparison of Technical Basis of Biodiversity e-Infrastructures*"**	*CReATIVE-B, Toward a Global Virtual Environment for Biodiversity Research. Project website and background materials*	*Research Infrastructures*
The LifeWatch Reference Architecture[19]	*Reference architecture model design specifications, based on the Open Data Processing Reference Model*	*Research Infrastructures*
The Principles of the GEO BON Information Infrastructure[20]	*GEO BON concept document covering data integration and interoperability in general terms*	*Research Infrastructures*
The Convention on Biological Diversity[21]**, AICHI Biodiversity Targets**[22]	*Convention on Biological Diversity's AICHI Biodiversity Targets towards strengthening protected area implementation for the conservation of biodiversity*	*Biodiversity*
The Global Biodiversity Informatics Outlook[23]	*A framework for biodiversity intelligence*	*Biodiversity*
The outGRID Interoperability Cookbook[24]	*outGRID interoperability cookbook methodology developed as an adaptable tool to support other e-infrastructures convergence*	*Interoperability Methodology*
The outGRID Interoperability Final Specifications[25]	*outGRID interoperability analysis conclusions for neurosciences e-infrastructures, including interoperability recipes and recommendations*	*Interoperability Methodology*
GEOBON Table of EBVs[26]	*GEOBON initial table of Essential Biodiversity Variables*	*Biodiversity*

1.4 Logbook of actions carried out

Table 2 reports an exhaustive list of important meetings and actions carried out over the duration of project activity, which have been key in the process of analysing interoperability. Additionally, several email exchanges and shared documents editing took place online.

Table 2. Logbook of important actions carried out in this analysis

Dates	Actions	Notes	RI
2011-11-07	Kick-off meeting *Amsterdam, Netherlands*	WP activities organisation & contact persons identification	ALL
2011-11 / 2012-02	Information sharing *Online*	Initial intra-consortium information sharing on key documents, reference models etc	CReATIVE-B
2012-02-15	Workshop preparation *Teleconference meeting*	Workshop organisation, methodology and materials preparation	CReATIVE-B
2012-02-27-29	1st Interoperability Workshop *Rio de Janeiro, Brazil*	Methodology introduction and technical information gathering	ALL
2012-03/04	Deliverable D3.1 first draft *Online*	Technical information analysis and specification formalisation in draft D3.1	CReATIVE-B
2012-05-03	Workshop preparation *Teleconference Meeting*	Workshop organisation and intermediary analysis brainstorming for presentation	CReATIVE-B
2012-06-07	European Data Forum (EDF12) *International Conference*	Conference attendance	CReATIVE-B
2012-06	Information collection *1st round*	Online technical information gathering and D3.1 drafting	CReATIVE-B
2012-07-05	2nd Interoperability Workshop *Copenhagen, Denmark*	Preliminary analysis presentation and technical information gathering	ALL
2012-09	Collaborative work on	Online sharing and collaborative	ALL

	synopsis *Online*	edition of RI synopsis table and RI descriptions	
2012-09	D3.1 consolidated draft *Online*	Technical information analysis and specification formalisation in draft D3.1	CReATIVE-B
2013-04-08	3rd Interoperability Workshop - Kunming, China	Analysis conclusions presentation and technical brainstorming on interoperability use-case(s)	CReATIVE-B
2012-10-22	EUDAT 1st Conference *International Conference*	Conference attendance	CReATIVE-B
2013-09-03	Biodiversity Informatics Horizons 2013 conference, Rome, Italy	Structuring the biodiversity informatics community at the European level and beyond	CReATIVE-B
2013-09-18	EGI TF COOPEUS Workshop - Madrid, Spain	Workshop attendance and contribution to discussions on interoperability	COOPEUS
2013-10-28	4th Interoperability Workshop – Santa Fe, NM, USA	Further discussion on EBVs as interoperability use case	CReATIVE-B
2014-02-16	5th Interoperability Workshop – Rome, Italy	Editing the Creative-B Roadmap	All

2 METHODOLOGICAL APPROACH

2.1 Objectives and qualities of the document

The methodological approach of the present activity has been derived from that pioneered in the outGRID project, described extensively by outGRID documents listed in Table 1 above.

The present document draws significantly upon that approach and is destined to serve as the basis for CReATIVE-B's concluding guidelines on interoperability for biodiversity and ecosystem research infrastructures. The document satisfies several general objectives and qualities:

(1) To develop a coherent set of interoperability scenarios that will structure subsequent brainstorming and technical developments.
(2) To identify inadequacies in objectives, requirements, standards, formats and technologies which could prevent concerned infrastructures from converging, thus supporting partners to make appropriate decisions over time,
(3) To provide a reference tool readable by developers, testers, maintainers as well as concerned researchers. Beyond formalizing what interoperability is about, the document also serves as a conceptual map that can be consulted at any time to better understand/locate/solve technical issues.

As such, the document also attempts to exhibit several qualities:

(a) Complete: everything that is essential must be described,
   - Rigorous: expressed in a well-defined notation. Diagrams are formalized, as much as possible, following standard notations when possible,
   - Uniform: the entire document is at the same level of detail and remains an abstract description of the targeted interoperability;
(b) Desensitized to change: it voluntarily hides implementation details to remain a high-level specification,
(c) Modifiable: this document may change over time. As expressed earlier, the presented specifications may be revised as technical information changes,
(d) Confirmable, verifiable and testable: the resulting recommendations should elaborate on the established technical facts while supporting future interoperability developments.

The present document will, as much as possible, enforce the formerly listed objectives and qualities for the duration of the project. It will serve as a solid basis for the subsequent project deliverable:

- D3.3 – "*Recommendations on resolution on interoperability in the medium to long-term*".

2.2 Approach to interoperability

Interoperability across national and organisational borders and across complex distributed computing platforms need non-invasive processes of interoperability. This is true in particular in order to avoid breaches of Intellectual Property Rights (IPR). Such processes must promote the least possible invasive convergence / integration guidelines, bearing in mind the different choices made by the different research infrastructures' managements; which are often the results of complex and internal decision making processes. Non-invasive processes are amenable to more rapid adoption and implementation. Interoperability guidelines must encourage as much as possible the use of open standards and be in line with the state-of-the-art, at the time they are being issued.

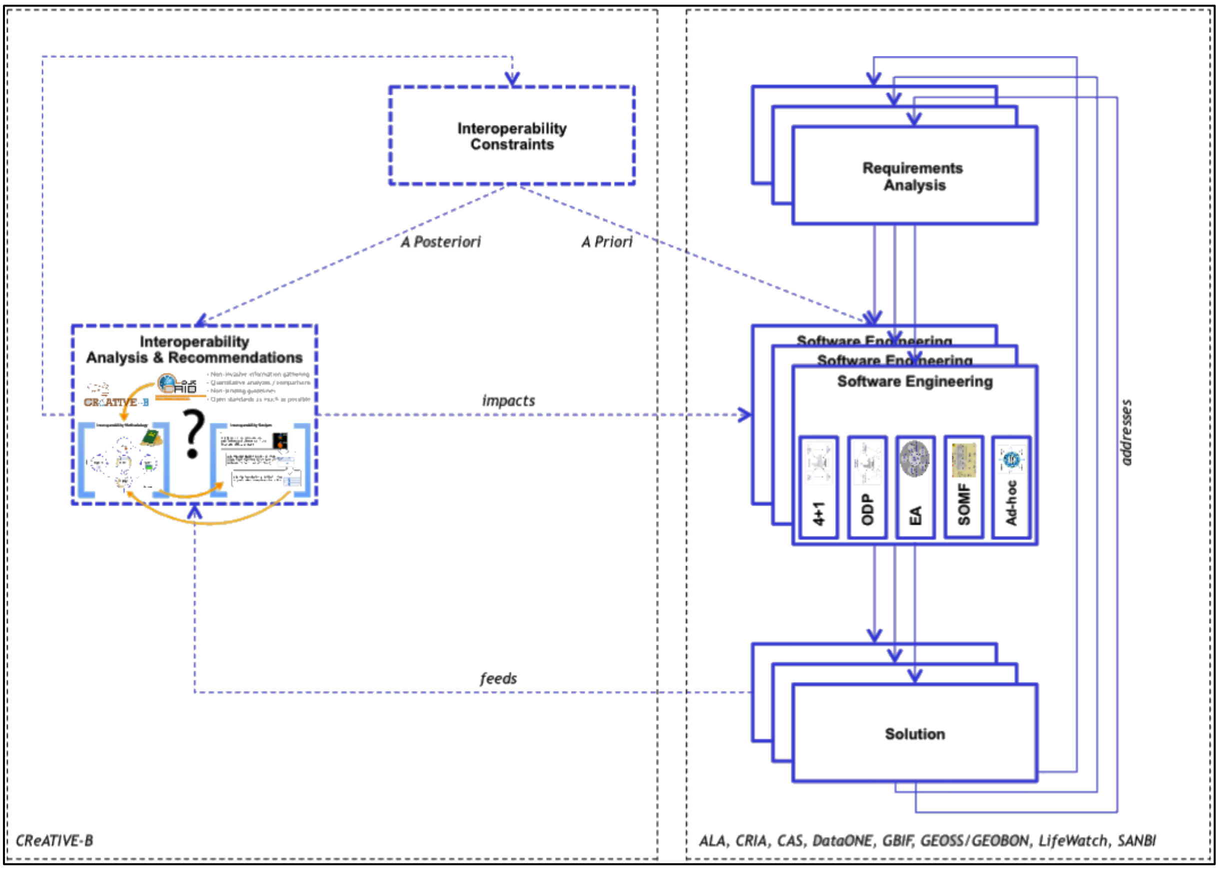

Figure 2. Approach to interoperability

Figure 2 illustrates where CReATIVE-B thus positions itself. Participating e-infrastructures (on the right of Figure 2) have been developing solutions for several years already. All of these potentially used different software engineering techniques and requirements analysis approaches, from 4+1 architectural view model[27], to Service Oriented Modelling Framework (SOMF[28]) or even ad-hoc processes.

CReATIVE-B (on the left) therefore aims to carry out an interoperability analysis "à posteriori", with the ultimate goal of impacting on targeted future developments of participating infrastructures. To do so, CReATIVE-B focuses on the solutions produced by the various infrastructures to identify obstacles to interoperability, to propose different scenarios of interoperability and to prioritize next actions to be carried out.

The present document and methodology were produced following these gold principles and aim at accompanying concerned e-infrastructures by letting them consensually agree upon, define and implement simple yet efficient guidelines, corresponding to their specific realities and coinciding when possible with their respective development roadmaps.

Implementing the proposed methodology, this document thus refers to 5 major steps, which compose the whole process of analysing and implementing interoperability. These steps are explained hereinafter and are further explored, as the project progresses in the development of deliverables D3.1, D3.2 and D3.3:

(1) Carrying out an <u>interoperability analysis</u>

This first step implies refining the proposed interoperability definition, in particular its inner dimensions, to be used as classification criteria to compare quantitatively measured technical facts within targeted e-infrastructures. The outcome of this analysis should allow precisely understanding similarities and differences between e-infrastructures, offer a synthetic overview of respective facilities and accompanying portfolios, and last but not least inform the process of identifying realistic interoperability scenarios together with preliminary actions to be carried out with accompanying guidelines. Step (1) should conclude with defining the objective and thus sketching major directions to be further investigated in consecutive stages of the process.

Step (1) should deliver:

- A detailed interoperability analysis,
- A synthetic overview of concerned systems, enabling their quantitative comparison, with identified similarities and differences,
- Early insights on possible interoperability scenarios and next actions to be carried out, towards step (2).

➔ This information has been formalised in CReATIVE-B deliverable D3.1

(2) Identifying <u>interoperability scenarios</u>

Based on the outcome of (1), realistic interconnection scenarios must be defined, which highlight accurately the potential added value for stakeholders. These scenarios must support the decision process towards an appropriate solution, and as such offer a panel of possibilities (when/if possible), towards progressive convergence. At the same time, these scenarios will give a taste on the technical complexity associated with all identified possibilities.

Step (2) should deliver:

- A list of realistic interoperability scenarios, together with pros and cons for decision making,
- A preliminary cut of high-priority actions to be carried out to achieve the scenario(s).

➔ This information is formalised in CReATIVE-B deliverable D3.2, the present document.

(3) Specifying an <u>interoperability roadmap</u>

Technical actions to be carried out by participating e-infrastructures should be listed, prioritized, and structured following the interoperability definition and subsequent dimensions, as defined in (1). This coordinated roadmap must be refined to gear towards the interoperability scenario(s), selected in (2). The roadmap is key to the process, as it is a working and operational document, which must be used as a central decision support tool for technical management boards and development teams to refer to. The roadmap must therefore be revised periodically and as much as possible take into consideration the technical and time constraints of participating e-infrastructures so to maximize impact onto respective developments.

Step (3) should deliver:
- An interoperability roadmap, identifying actions, priorities and achievements per participant, corresponding to formerly introduced interoperability dimensions.

➔ This information is formalised in CReATIVE-B deliverable D3.2, the present document.

(4) Defining and refining interoperability guidelines

Accompanying the roadmap defined in (3), interoperability guidelines must be expressed which advocate (when/if possible) standards to be used for achieving the targeted scenario(s). The guidelines are here to facilitate convergence by either promoting “de facto” or open standards, thus confining complexity down to well-located interconnections. The guidelines must be consensually agreed among development teams for further integration. Hereinafter, a so-called “standards matrix” is proposed, which allows listing standards according to the combination of different criteria such as interoperability dimensions and scenarios.

Step (4) should deliver:
- A standards matrix, defining de facto and/or open standards for interconnecting systems and their subsystems, per interoperability dimensions and scenarios.

➔ This information will be formalised in CReATIVE-B deliverable D3.3

(5) Developing an interoperability prototype and challenge

With the two-fold objective of exercising achieved interoperability and showcasing its benefits to stakeholders, step (5) advocates the definition of a prototype application addressing part of the targeted scenario requirements, and allowing the design of a so-called “interoperability challenge”. The latter shall demonstrate the technical feasibility and benefits to users.

Step (5) should deliver:
- A prototype application focusing on a subset of technical interoperability issues,
- An interoperability challenge, technically and if possible scientifically testing the prototype, while showcasing the benefits of interoperability

➔ This information will be formalised in CReATIVE-B deliverable D3.3

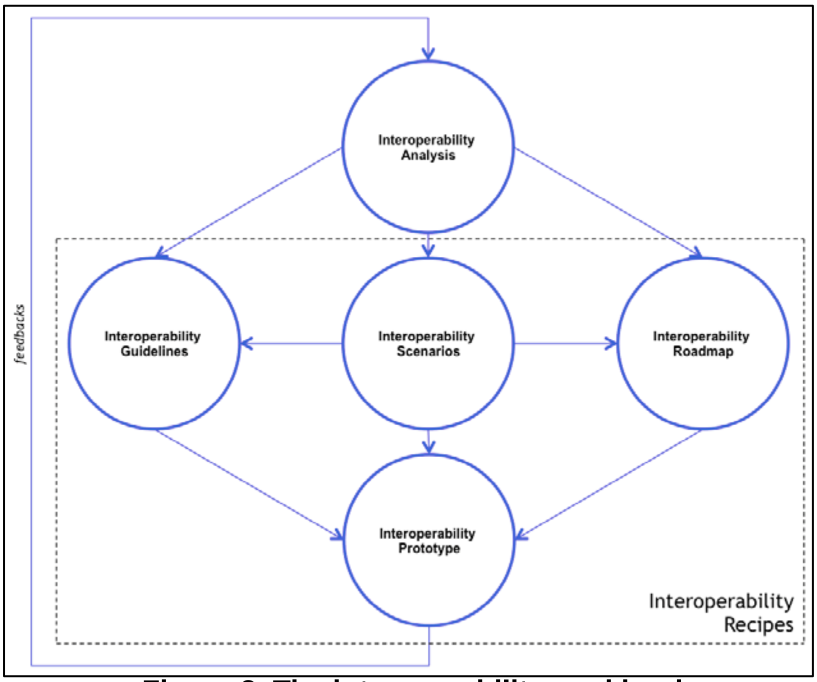

Figure 3. The interoperability cookbook

Figure 3, on the left, recalls these steps in the global process and introduces the notion of interoperability recipes, corresponding to identifiable scenarios and providing functional objectives together with interoperability recommendations to be followed.

The following section reprises the conclusions of the first step of carrying out the interoperability analysis (documented in D3.1).

3 CONCLUSIONS OF THE INTEROPERABILITY ANALYSIS

3.1 Background and research infrastructure definition

Information about 9 research infrastructures for biodiversity and ecosystems science has been gathered, collated and compared in the interoperability analysis step (described as step 1). This is reported in Deliverable D3.1 "*Comparison of Technical Basis of Biodiversity e-Infrastructures*". The 9 research infrastructures are:

- The Atlas of Living Australia (ALA)[10]
- Brazilian SiBBr and CRIA SpeciesLINK infrastructures[11,12,13]
- GBoWS / Chinese Academy of Sciences infrastructure[14]
- DataONE[15]
- GBIF[16]
- GEOSS / GEO BON6'8
- LifeWatch2
- The South Africa National Biodiversity Institute (SANBI) infrastructure[17]

Biodiversity communities work internationally with data originating from all over the world, stored and made available through a wide range of tools, services and mechanisms. Researchers are geographically scattered, thus creating a need for fast and reliable aggregation of data and tools into large and interoperable research infrastructures supporting easy and accurate use of the capabilities. This is what we refer to as an International Virtual Environment (IVE) for Biodiversity[29]. Capabilities of such an environment include: sensors and sensor networks deployment, digitizing of biological specimens, ground level and remote observations, fast DNA sequencing facilities, interoperability and data sharing, data discovery and knowledge development, computation for modeling and simulation, virtual laboratories and e-services.

In CReATIVE-B, the challenge is to bring several of these regional, national and international community initiatives together and to add scientific value for users, towards the long-term development of an IVE for Biodiversity. This is not a merger of research infrastructures, as these are inherently distributed and technologically heterogeneous. It is instead an interoperability enabler towards international governance. Indeed, interoperability of data, tools and services as well as the infrastructure's management have to be achieved to promote advanced data mining and new knowledge discovery in a coordinated and international setting.

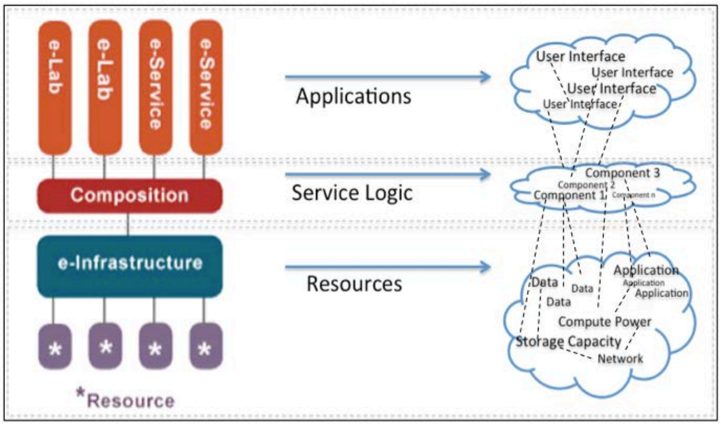

Figure 4. Research infrastructure first-class objects

As illustrated in Figure 4, a research infrastructure (RI) is characterized by three major layers. Resources useful to users, at the bottom, where one can find network bandwidth, storage capacity, computational power as well as scientific data and applications, which a given RI offers access to. Integrating these, the service logic in the middle, provides harmonized and standardized interfaces to mediate accesses from value added business logic to higher levels of the system, namely user applications. Finally, users seamlessly access (composed sets of) resources, thanks to dedicated interfaces such as virtual laboratories, web portals and other specific e-services.

These first-class concepts, i.e. resources, service logic and applications, support and structure a more elaborated definition of RIs (see section 3.2.2 in D3.1); in particular to elicit further elements on which to base a quantitative comparison of participating infrastructures useful to identify key differences and similarities.

3.2 Potential for achieving interoperability

Comparing the 3 major levels of RIs, as introduced in Figure 4, the radar maps in Figure 5 - **Figure 6** below, highlight most important aspects with red borderlines, in order to give a concrete feel on enabled/ing facts, overall feasibility and beyond anything else major obstacles to interoperability.

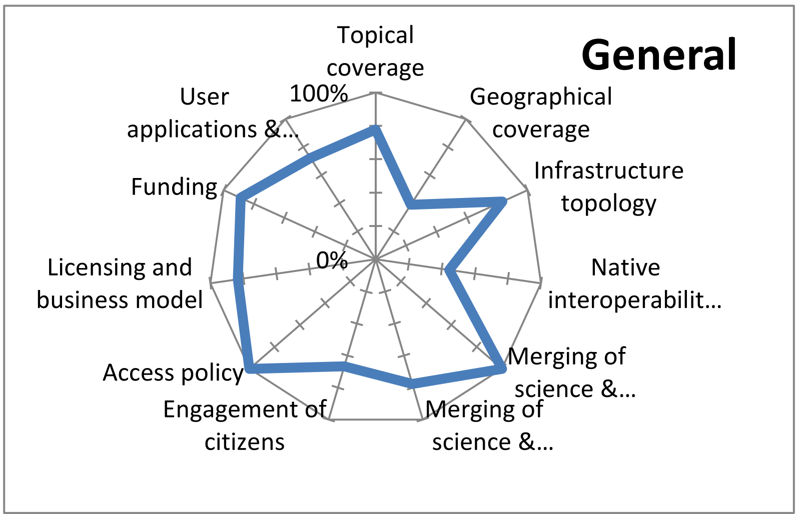

Figure 5: Potential for interoperability at the general level

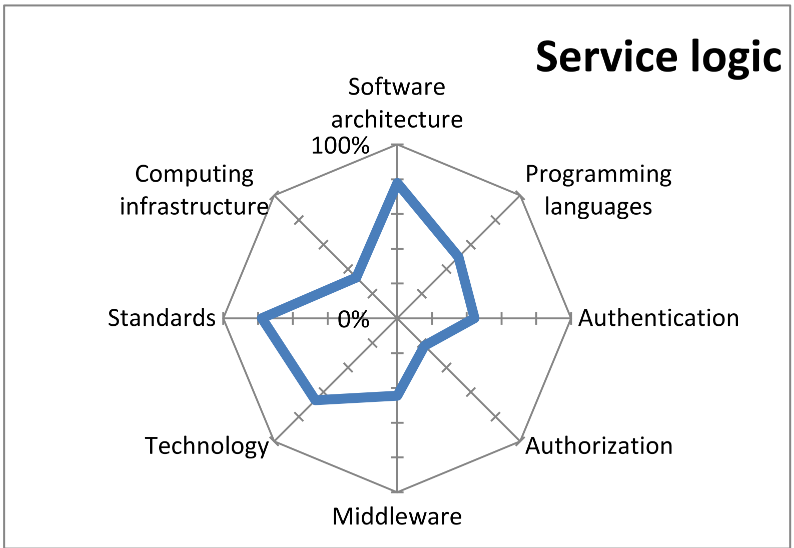

Figure 6: Potential for interoperability of service logic

Generally speaking, all the cooperating research infrastructures exhibit a satisfactory level of potential interoperability (Applications radar map); in particular, in the way they offer access to biodiversity data, available applications and related resources. Each RI pursues similar objectives in terms of business models, industry and policy involvement, and overall sustainability plans. This facilitates achievement of a future international virtual environment (IVE) for biodiversity and ecosystems research and the accompanying governance. Participating RIs have complementary geographical and topical coverage, while differing in their actual implementations. This begins with the physical topologies of their networks and resources. As is to be expected, differences in implementation become more obvious in the "service logic" (Service logic radar map). Despite similar approaches to software

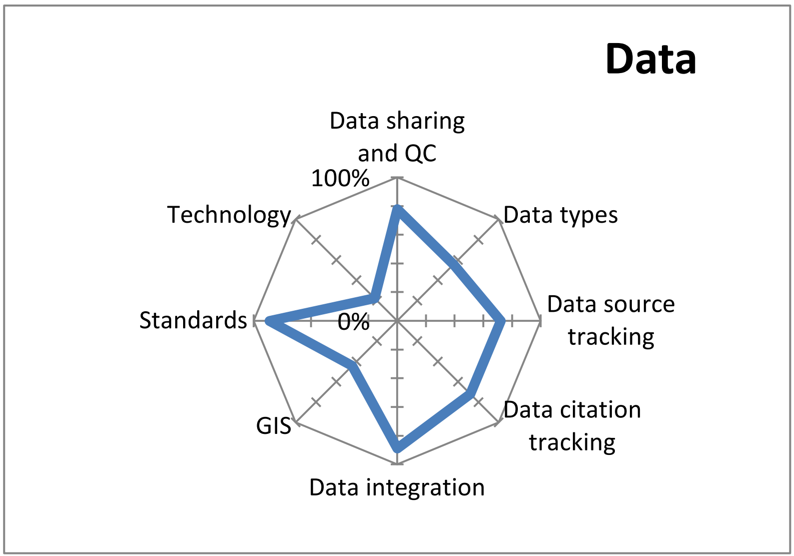

Figure 7: Potential for interoperability in context of data use

architectures and standards adopted, the service logic in the RIs is the place where the most differences can be found.

Proprietary middlewares have been deployed with (among many aspects) different security infrastructures, programming languages and technologies. It is in areas such as these where most work is needed to make systems syntactically and semantically compatible. However, a long-term goal of service orientation is not fundamentally compromised.

On the data side (Data radar map) some domain-specific standards (e.g., Darwin Core, TAPIR, Ecological Metadata Language (EML)) are emerging that begin to address the needs of data integration and organisation. Some similar sharing and quality control processes are in place for initiatives dealing with data collection, and traceability is a shared concern for scientific citations and raw data tracking.

3.3 Overcoming barriers to interoperability

From the various fora and exchange opportunities created by CReATIVE-B, feedback has been gathered from RI stakeholders and biodiversity researchers on actual needs felt unaddressed in the community and on the potential barriers to achieving interoperability. Some of these were already mentioned in D3.1 (i.e., in section 4.2) but are elaborated here again, in addition to several new ones.

Overcoming the barriers to global interoperability in RIs means: i) promoting understanding of the value of interoperable RIs; ii) using coordination to achieve interoperability; iii) emphasising and increasing the importance of standards; and iv) solving specific technical challenges.

The remainder of this section introduces these and suggests some recommendations supporting the above axes that could be included in the roadmap.

3.3.1 The value of interoperable RIs

Concrete benefits of interoperability have to be (made) visible to stakeholders. These have to be emphasized and promoted in order that the necessary attention is paid to encourage and achieve interoperability. Use-cases thus have an important role to play in illustrating these benefits. One significant use-case, presently the focus of wide discussions, is that of biodiversity indicators and more specifically the concept of “Essential Biodiversity Variables” (EBV), as introduced by GEOSS-GEO BON. Potentially, EBVs or similar indicators are the core future business for RIs and the converged IVE.

Suggested recommendation(s) for inclusion in roadmap:

Illustrate the benefits of interoperability with Essential Biodiversity Variables (EBV). Conceived by GEO BON collaborators (Pereira et.al. 2013[35]), EBVs are endorsed by the Convention on Biological Diversity (CBD) and in line with 2020 Aichi Targets. They provide a focus for GEO BON and related monitoring activities and have a role to play in biodiversity assessments (e.g., IPBES) and prediction of the future state of the biosphere. As a general principle, it should be possible to calculate the value of any chosen EBV:

- For any geographic area, small or large, fine-grained or coarse;
- At a temporal scale determined by need and/or the frequency of available observations;
- At a point in time in the past, present day or in the future;
- As appropriate, for any species, assemblage, ecosystem, biome, etc.
- Using data for that area / topic that may be held by any and across multiple RIs;
- Using a standardised, widely accepted protocol (workflow) capable of executing in any RI;
- By any (appropriate) person anywhere.

3.3.2 Coordination to achieve interoperability

As the Global Biodiversity Informatics Outlook GBIO[23] makes clear, there is considerable complexity and interconnections between the components that contribute to making a RI. There are multiple activities essential to the successful delivery of integrated e-infrastructures for managing and using biodiversity data in support of science and policy. Many of these are already underway but continuous support, increased technical capacity over time are all essential. It is indeed necessary to have global coordination and mutual understanding, to ensure that the benefits are realised at the lowest possible cost and within a reasonable timeframe. Alongside coordination, investment in training for skills development is also critical.

Suggested recommendation(s) for inclusion in roadmap:

Coordinating interoperability around a consensually agreed technical roadmap of joint, as well as, specific actions to be carried out by concerned RIs.

Capacity building. Structuring and supporting education and training that encourage interoperability. Such activities should be organized around/in a specialized biodiversity “market place”, e.g., http://www.biodiversityinformatics.org/culture/, together with access to RIs’ resources, thus facilitating adoption and harmonizing practices.

3.3.3 Emphasising the importance of standards

Regimes of unstable, constantly changing, or even using few standards, are fundamental barriers to interoperability. Stable standards for data formats and exchange protocols, with widespread adoption is the basis for good interoperability. Standards however need time to mature and stability accrues when players are actively involved in their simultaneous specification and implementation. Greater clarity is thus needed on standards that should apply in this domain. New standards may not be necessary (although that may be the case for data and semantics integration). Adoption of existing, well-used industry standards should therefore be encouraged and increased. Promoting and coordinating this (e.g., through specification in procurement) is essential. Technical enforcement of security, intellectual property protection and data licensing becomes easier when standards are widespread and industrially based.

Suggested recommendation(s) for inclusion in roadmap:

Learn lessons from other domains, such as the healthcare sector. Here, the modus operandi has been to proceed case-by-case, solving each use-case in turn. Interoperability "profiles" were thus introduced, specifying the standards needed at every level (e.g., of an architecture) and to be adopted by each provider within the sector. This approach could work for biodiversity science too.

Publish and promote standards best practices on a central and well-known Website, such as e.g., https://www.earthobservations.org/geobon.shtml, and ensure these are considered when roadmapping technical developments across RIs.

3.3.4 Solving specific technical challenges

Tackling interoperability implies addressing a set of technical challenges internally and externally to cooperating research infrastructures. Similarly to problems faced by enterprises in the past when making respective information systems to talk to each other, the biodiversity infrastructure community, but generally also other international research infrastructures, are in need of interconnecting their services, formatting messages, synchronizing actors and materializing their workflows. There are thus 5 key technical recommendations on the roadmap to achieving interoperability.

Suggested recommendation(s) for inclusion in roadmap:

1. Develop enabling, global and federated AAA infrastructures - 3 years

AAA. Overcoming barriers to Authentication, Authorization and Accounting (AAA) when composing complex applications across multiple RIs requires alignment and interworking of security infrastructures. User applications in one RI should be able to enact services and access data within another RI seamlessly. In practice, AAA interoperability at the global level is likely to be based on the Shibboleth model[30] and on identity federations established more broadly than just for the RIs of interest in this roadmap (e.g., eduGAIN[31]).

Trust. It will be necessary to establish mutual trust relationships between RIs as an essential prerequisite to supporting delegation of users' credentials throughout the flow of enacted services. This is a non-trivial and unsolved problem arising from multiple levels of trust relationships that exist between: a) users and the applications they use; b) the applications and the domain-specific specialised Service Providers offering services upon which the applications depend (such as GBIF); and c) Service Providers and the providers of foundational computing, storage and networking infrastructures (e.g., general-purpose cloud computing and cloud storage).

2. Encourage the use of consistent quality control, semantics - 5 years

Quality control. Data and metadata need to be better qualified in terms of quality, i.e. whether they were quality assured and following what protocol. Moreover, the granularity between data and metadata also requires a subtle and well-balanced thinking to be turned into meaningful information for users.

Semantics. The lack of use of consistent vocabularies and the absence of a comprehensive ontology for biodiversity and ecosystem science impedes the semantic integration of data. Alignment of concepts and agreed (meta)structures (copying, for example, the approach of UMLS - Unified Medical Language System) would contribute to better understanding, integration and interoperability.

EBVs. Work in the area of EBVs may help in overcoming ontologies alignment and complex new developments by introducing an intermediary semantic but simplified layer, closer to end-users expectations.

3. Promoting the development, sharing and use of workflows of services - 5 years

Services. Web services play a significant role nowadays in separating technological dependencies arising from specific engineering decisions of RIs. The use of Web services should be encouraged to expose RIs' functions and to allow their interoperation, without implying intrusive integration nor complex reengineering.

Workflows. As progress is made in exposing analytical tools, data and other resources as standard Web services, it becomes more important to adopt robust workflow management systems, (e.g., Taverna and Kepler). Workflows make it possible to combine coarse-grained functions into complex applications (such as calculating EBVs) requiring access to resources located in various RIs. Workflows, especially peer-reviewed ones offer standardisation of approach, becoming a standard way of doing something or being an approved procedure in a regulatory environment. They are repeatable; allowing the same or similar task to be done again, and again – possibly with different data and/or control parameters. Workflows fit the "ISA" management model of "Investigations, Studies, Assays" that is finding favour nowadays in the wider life sciences. Workflows allow reproducibility and act as a provenance mechanism for capturing the way work was done – provenance of the data, the tools used and the precise steps followed. Workflows offer a faster, cheaper, and integrative way of linking and utilising resources across multiple RIs.

4. Creating a scientific market place for biodiversity services - 5 years

Market. Allowing workflows of services to be composed and executed cross-enterprise, cross-infrastructure require globally accessible catalogues of data, services and associated semantics. Catalogues will be used to publish, discover, share and manage global portfolios of data and services. DataONE and GBIF, for instance have already made much progress in these areas. Multiple catalogues for data lead to the need for federated search and discovery that can be addressed with openSearch technology[32]. Catalogues of services, such as the Biodiversity Catalogue (www.biodiversitycatalogue.org) should be promoted as well-known and well-founded directories of Web services for biodiversity and ecosystems analysis applications. In both cases, enhanced capabilities permitting semantic searching in and across catalogues are needed for the future.

Enterprise Service Bus (ESB). Comparable to current ESBs, a de-centralised and voluntary "Service Network" approach that accounts for independence and autonomy of individual Service Providers is most likely to find favour among cooperating RIs. Data and service brokering components, such as those investigated by EuroGEOSS[33] or Earthcube[34] take away from Service Providers much of the responsibility for interworking heterogeneous resources – even if they are encouraged to comply with relevant sector standards in order to maximise usage of their service(s).

5. Managing the provenance of resources in RIs - 10 years

Digital Objects Identifiers. All resources of involved RIs have to be assigned with a unique and global identifier, in the same way that scientific publications (DOIs) and data are. Thus, it would be possible to identify, manage these resources and ultimately to store provenance information when creating, modifying and utilizing them individually or collectively in workflows. A common mechanism across RIs is needed.

Provenance. Details of all actions carried out in RIs, the users involved, as well as all modifications of the state of resources should be monitored, traced and preserved in order to make it possible to define precisely the provenance of every single digital object, to assure IP ownership, define responsibilities, identify the root causes of problems, improve quality processes and support repeatability of processes. Open models for structuring provenance information, such as the Open Provenance Model (OPM) should be considered.

4 ESSENTIAL BIODIVERSITY VARIABLES AS A USE-CASE

4.1 Introduction to Essential Biodiversity Variables

Essential Biodiversity Variables (EBV) facilitates data integration by providing an intermediate abstraction layer between primary data observations and indicators / measures of biodiversity. EBVs aim to help observation communities to harmonise monitoring, by specifying how variables should be sampled, measured and computed. They standardise ontology for biodiversity and harmonise measurements, observations, and protocols. For the first time, they provide a tractable and holistic approach to monitoring (and eventually predicting) the state of the biosphere.

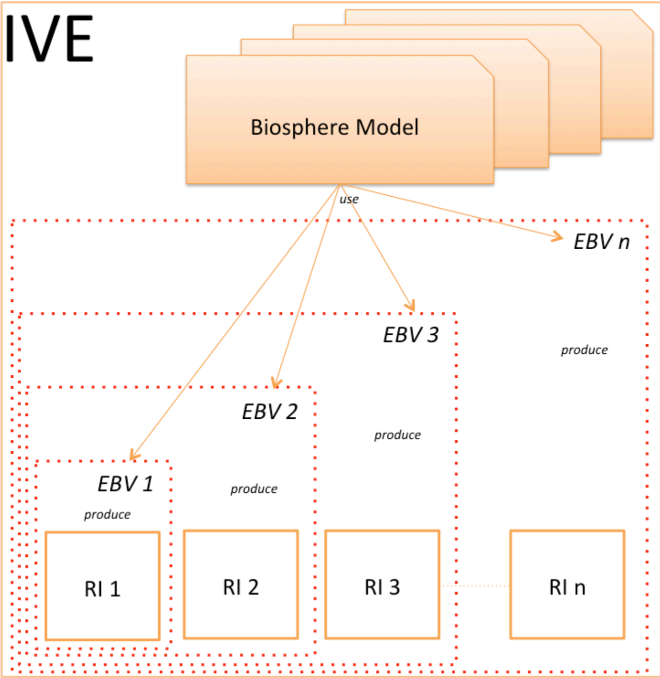

Figure 8. IVE & EBV, toward a Biosphere Model

Conceived by GEO BON collaborators (Pereira et.al. 2013[35]), EBVs are endorsed by the Convention on Biological Diversity (CBD) and in line with the 2020 Aichi Targets. They provide a focus for GEO BON and hence for the interoperability thrust within GEO BON activities. GEO BON has set the goal to have an operational system by 2015, meaning that some EBV production / computation is operational by then.

Essential Biodiversity Variables (EBV) are thus core future business for the RIs and the converged IVE, as Figure 8 above illustrates. RIs interoperating together can ultimately provide ubiquitous access to the data and tools needed for EBV computation and management. For the purposes of the present analysis, EBVs are thus a good use-case for driving thinking about interoperability, its challenges and solutions.

4.2 Calculating EBVs

As a general principle, it should be possible to calculate the value of any chosen EBV:

- For any geographic area, small or large, fine-grained or coarse;
- At a temporal scale determined by need and/or the frequency of available observations;
- At a point in time in the past, present day or in the future;
- As appropriate, for any species, assemblage, ecosystem, biome, etc.
- Using data for that area / topic that may be held by any and across multiple RIs;
- Using a standardised and widely accepted workflow capable of executing in any RI;
- By any person anywhere.

As illustrated in Figure 9 below (extracted from EU BON works) a workflow approach such as that envisioned by LifeWatch and piloted by the FP7 funded BioVeL project[9] for example, is appropriate.

The workflow approach offers to support the automatic data processing needed for computing the Essential Biodiversity Variables. Workflows can represent a rigorous and standardised method that is repeatable and that can be used anywhere.

It is intuitive from the foregoing that this ability to compute EBVs demands interoperability across RIs on multiple levels. For a selected typical EBV therefore, we consider from the interoperability perspective at each level (application, service logic, infrastructure) what are the implications for each RI to serve up the EBV i.e., to utilise data it holds or needs and to be able to calculate the EBV on request.

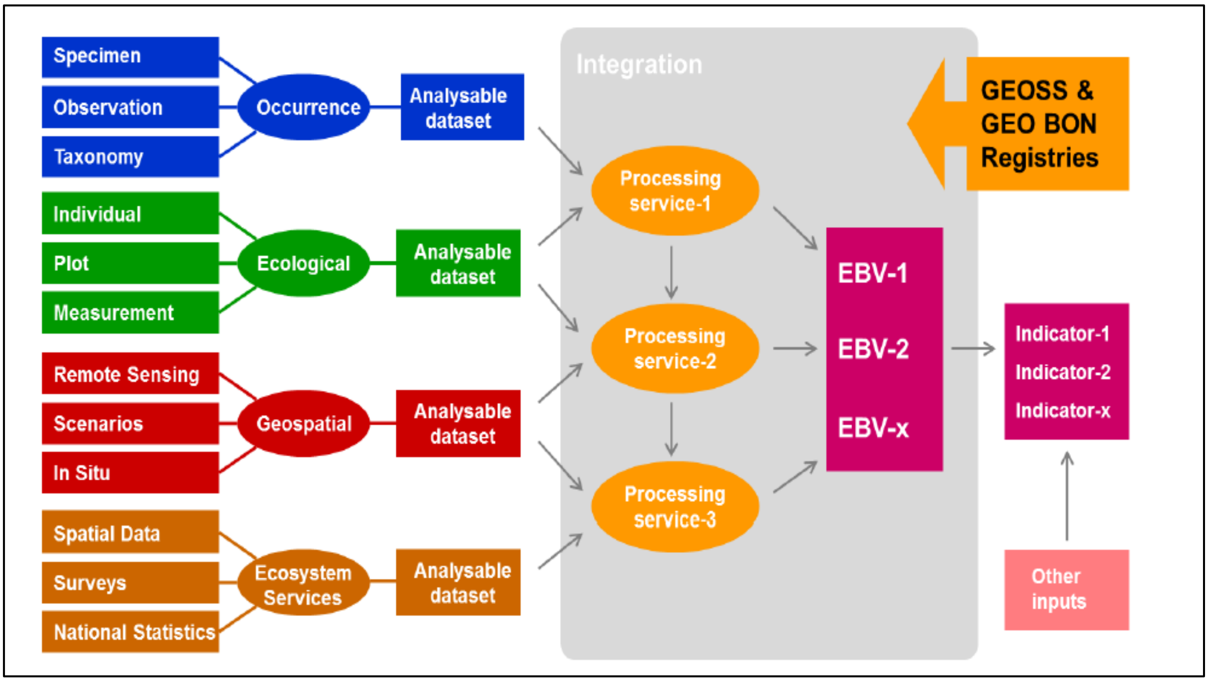

Figure 9. GEO BON vision of automated streamlined dataflow (Saarenmaa et al. 2013, redrawn by Hoffman et al. (submitted)

As a European project contributing towards GEO BON, the FP7 funded EU BON project[36] has adopted EBVs as a use-case to focus on[35,34]. In CReATIVE-B, we pick the Species Distribution EBV as one to try and to make further progress in that direction. This is because substantial data already exists and much investigative work in species distribution mapping and modelling has already been successfully undertaken. Many projects have been or are active in this area so there is much experience to draw upon.

4.3 The Species distribution EBV example

4.3.1 Characteristics

Table 3 below is adapted from the GEO BON table of candidate EBVs[26].

Table 3. Characteristics of species distribution candidate EBV

EBV class:	*Species populations*
EBV:	*Species distribution – nominally, the presence of a taxon in a place at a time, measured repeatedly over time with consistent methodology*
Measurement and scalability:	*Presence surveys for groups of species easy to monitor, over an extensive network of sites with geographic representativeness. Potential role for incidental data from any spatial location.*
Temporal sensitivity:	*1 year to >10 years*
Feasibility:	*Presence surveys are available for a larger number of species than population counts and can make use of existing distribution atlas. Some efforts for data compilation and integration exist (GBIF, IUCN, Map of Life). There is an increasing trend for data contributed by citizen scientists (Observado, iNaturalist).*
Relevance:	*Abundance & distribution of populations/taxon per se is an intuitive biodiversity metric with public resonance. Abundance & distribution contributes to extinction risk indicators and indicators of supply of ecosystem services associated with particular species. Range shifts are expected under climate change.*
Related CBD targets:	*4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15.*

4.3.2 Method of calculation

A possible method of calculation may involve the elements mentioned in Table 4 below.

Table 4. Elements involved in calculation of species distribution EBV

Involved RIs	• Any holding species observation/occurrence data, e.g., GBIF
Primary data	• Occurrence data points for hundreds of species (e.g., plants, butterflies, …) • Each individual species may have thousands or tens of thousands of data points
Applications, models and / or services	• Use OpenModeller based Ecological Niche Modelling[37] for all the data points o Choose predictive layers from WorldClim and GEOSS sources o Run BioClim workflow for selecting environmentally unique data points o Run Ecological Niche Modelling (ENM) workflow for large number of species
Scientific questions	• How many species are increasing? • How many are decreasing? • Does the flora/fauna move to any direction? • Is distribution fragmenting? • Is distribution shrinking? • Are species becoming marginalised?

5 SCENARIOS OF INTEROPERABILITY

5.1 Introduction to interoperability scenarios

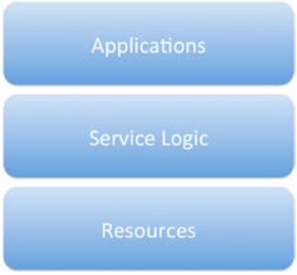

Figure 10. The 3 layers of RIs

Using an n-Tier architectural pattern at the micro-scale, e-infrastructures' reveal 3 major conceptual layers at the macro-scale, as illustrated by the simplified representation in Figure 10, on the right. The lowest layer (resources) represents the data, software and computational resources. The middle layer (Service Logic) integrates these, while the top layer (Applications) allows for composing preferred workflows. *Note: see deliverable D3.1 for further details on RI and layers definitions.*

Applications, through e-laboratories or e-services connected to the Service Logic, at the top, seamlessly benefit from underlying assets thanks to offered abstraction and virtualization.

This simple conceptual representation can be further used to facilitate understanding of interoperability and to sketch key interoperability scenarios - from the most basic interconnection of computing resources at the network level to the more complex enterprise-level integration of added-value business intelligence in targeted platforms and applications.

Table 5. Summary of interoperability scenarios

Scenario ID	**Short Description**
	**Scenario 1 – Interoperable Infrastructural Resources** This scenario is concerned with the interconnection and interoperability of computational, storage and other low level resources exploited by the RIs; and whether they are exposed, used and managed utilizing the same or similar standards, protocols and software implementations.
	**Scenario 2 – Interoperable Platform Middleware** This scenario is concerned with the middleware compatibility across concerned research infrastructures, from encapsulation, to business logic and communications services.
	**Scenario 3 – Interoperable User Applications** This last scenario represents the ultimate interoperable architecture, where any applications can interact with other RIs' applications, services or computational resources.

In the remainder of this section, 3 interoperability scenarios as summarised in Table 5 above are further introduced. This includes explanations about the benefits of having these layers made interoperable across disparate e-infrastructures.

5.2 Scenario 1 – Interoperable Infrastructural Resources

In this first scenario, infrastructural resources are interoperable. The participating initiatives interact via common physical networks. Communication and security protocols are shared or interworked and, optionally, computing resources are harmonized (for example, thanks to virtualization). This allows interconnected systems to communicate and to potentially share low-level infrastructural resources in an integrated and secure environment. This also facilitates cross-infrastructure maintenance of computing resources (e.g., processing power, storage capacity, etc.) and allows their mobility from one site to another (potentially from data centre to data centre). This interoperability scenario implies that the necessary network cabling and routing has been operated so as to have participating research infrastructures communicating over interconnected communication channels. Secondly, it implies that infrastructures do interact in a sound security framework in terms of communication protocols, encryption standards etc. Thirdly, underlying computing resources may be virtualized with the same or compatible hypervisor technology(ies) and / or standard(s), thus making it possible to move virtual machines from one node to another, expanding services on demand and facilitating overall maintenance without bothering about geographical locations and resources fabrics.

Users	Benefits
System Administrator	Facilitated maintenance of low-level computing resources
	Improved mobility and scalability of computing resources
	Facilitated network routing and improved communication performances
	Facilitated communication protocols and encryption security configurations
Computer Scientist	Improved portability of developed services
	Transparent access to computing resources from participating infrastructures
Biodiversity Scientist	Transparent access to computing resources from participating infrastructures
	Improved portability of developed service logics

In the opinion of the authors this is the least likely scenario of interoperability because it requires significant alignment of underlying technology choices across RIs.

5.3 Scenario 2 – Interoperable Service Logic

In this second scenario, the Service logics of participating initiatives are interoperable. It makes it possible for participating initiatives to interoperate in a synchronized manner, for a common purpose, such as (for instance) supporting the execution of a scientific workflow that utilises services / resources across multiple research infrastructures. This interoperability scenario implies that Service Logics are compatible and with similar semantics; not only in terms of data and metadata structures but also workflow and provenance specifications and engines. It is one of the most expensive scenarios in terms of technical developments as it means technological bridging (e.g., by mapping gateways or brokering) both in terms of syntactic communications and semantic operations. It therefore requires a deep understanding of services logics between (inter-) and within (intra-) infrastructures, data structures and higher-level communication protocols.

Users	Benefits
System Administrator	Facilitated maintenance of all computing resources from low-level virtual machines and networks to higher-level software services
	Improved mobility and scalability of computing resources
	Facilitated network routing and improved communication performances
	Facilitated communication protocols and encryption security configurations
Computer Scientist	Improved portability of developed services and full provisioning possibilities
	Transparent access to computing resources from participating infrastructures
	Increased functional possibilities towards addressing new user requirements
Biodiversity Scientist	Transparent access to computing resources from participating infrastructures
	Improved portability of developed service logics
	Access to cumulated scientific portfolio in terms of data sets, data sources and service logics
	Increased functional possibilities towards creating new scientific experiments

This is a desirable scenario and is, in the opinion of the authors the most likely scenario to be realised. It is a scenario based upon abstraction of underlying physical technology choices and presentation of services / APIs based on widely used industry standards.

5.4 Scenario 3 – Interoperable User Applications

In this third scenario, user applications exposed by participating initiatives are interoperable. Data and service logics can be shared and manipulated by all users and across user applications, which interact with respective middleware. Resources management from applications scheduling, execution, to data storage and distribution can be controlled / operated from within all participating infrastructures' middleware. It is the most expensive scenario in terms of technical developments, as it implies technologically bridging applications and middleware services both in terms of syntactic communications but also in terms of semantics. This allows users to access a virtually cumulated portfolio of scientific resources through a set of compatible interfaces, to setup new scientific experiments spanning several infrastructures in parallel. This technically allows all former interoperability scenarios to be performed, while opening the pathway to richer use-cases of the resulting unique IVE facility. It therefore requires a deep understanding of inner applications and services business logics, data structures and higher-level communication protocols of all involved RIs.

Users	Benefits
System Administrator	Facilitated maintenance of all computing resources from low-level virtual machines and networks to higher-level software services
	Improved mobility and scalability of computing resources
	Facilitated network routing and improved communication performances
	Facilitated communication protocols and encryption security configurations
Computer Scientist	Improved portability of developed services and full provisioning possibilities
	Transparent access to computing resources from participating infrastructures
	Increased functional possibilities towards addressing new user requirements

Biodiversity Scientist	Transparent access to computing resources from participating infrastructures
	Improved portability of developed algorithms/pipelines
	Access to cumulated scientific portfolio in terms of data sets, data sources, algorithms and pipelines
	Increased functional possibilities towards creating new scientific experiments
	Harmonized and user-friendly interfaces

This scenario is not very likely to be pursued in the opinion of the authors because it requires standardisation of and gateways between applications to be developed. This is considered unlikely in view of the sufficiently differing missions of the various RIs (cf. table 4 in Deliverable D3.1).

5.5 Disregarded Interoperability Scenarios

It is to be noted that the authors judged two intermediary interoperability scenarios not to be relevant to the present purpose, i.e. that of involving interoperability at the middleware level only (i.e. hereinafter referred to as scenario 1.5) and that of implying interoperability at the user application level only (i.e. hereinafter referred to as scenario 2.5).

Scenario 1.5, on the one hand, is disregarded as it would imply situations in which different middleware platforms are interoperable but not communicating with / on the same physical network, nor using compatible standards and protocols. While the scenario is likely to materialize when involving different types of networks at the same time (e.g., a combination of satellite and broadband), it is not considered to be relevant to the present research e-infrastructures.

Scenario 2.5, on the other hand, is disregarded as it would imply interconnecting user applications with middleware services logics and thus moving interoperability up to every single application. Although this scenario is sometimes used to quickly prototype a demonstrator illustrating the benefits of interoperability, it does not scale nor is it financially acceptable as it implies re-engineering each time a new application or middleware is to be integrated.

6 MAPPING THE USE-CASE TO SCENARIOS

6.1 Mapping the selected EBV to a possible scenario

From the formerly introduced characteristics of the chosen EBV (i.e., species distribution) it appears that concerned data and models are wrapped up and made available through different RIs, service logics and infrastructural resources.

The nature of the proposed applications and service logics is such that integrating them into a greater set aiming at calculating the species distribution EBV would well fit the workflow concept, in computing terms.

Authors thus hereby make the assumption that the most adapted interoperability scenario is scenario 2 (see 0 above), i.e. that of interoperable service logic.

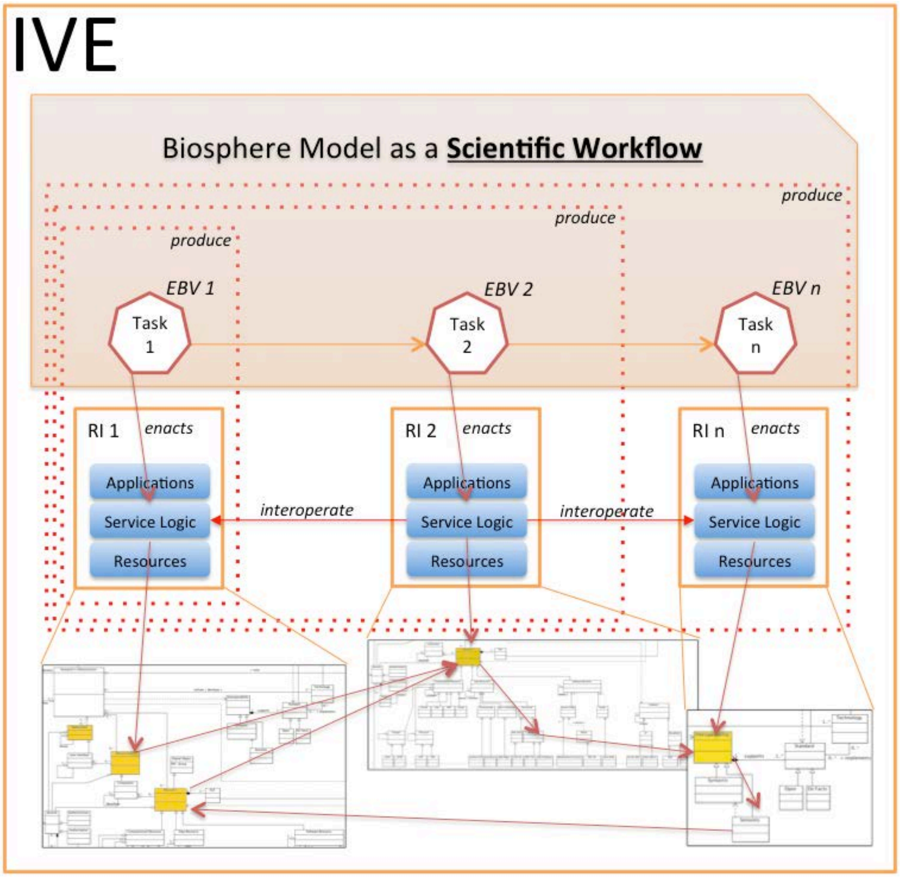

Figure 11. Model and EBVs as scientific workflows

This scenario is concerned with the middleware compatibility across involved research infrastructures, from encapsulation, to business logic and communications inter-services. This is what Figure 11 illustrates, with the biosphere model represented at the top as a scientific workflow made of tasks enacting underlying research infrastructures' service logics, towards the common objective of calculating required EBVs. The resulting "exo-workflow" is able to interact with RIs to enact respective services, synchronize them, pre/post process the results thus abstracting from underlying architectural and technological complexities, as illustrated at the bottom of Figure 11.

The detailed implications at each of the 3 levels – resources, service logic and applications require further study.

7 CONCLUSIONS & ADVICE ON ROADMAP

7.1 General conclusions

Analysis of interoperability is complex, as has been demonstrated by the work so far in the CReATIVE-B project but the storyline on how to solve it has, in the end to be simple.

Simply put, the way forward to achieving interoperability is based on the coordinated implementation of selected Standards in support of specific use cases i.e. a case-by-case approach, using precise definitions ("Profiles") of how specific Standards can be adopted and implemented to solve specific biodiversity and ecosystem science needs.

In the roadmap deliverable D3.3 we will propose recommendations (based on section 3 above) to support this conclusion.

7.2 Future works

In the context of the above conclusion, international cooperation between the globally distributed biodiversity data infrastructures with a focus on targeted services to support research on "predicting the biosphere" is the essential future direction for the research infrastructures globally. This grand challenge was the primary outcome of the international Biodiversity Informatics Horizon Conference in September 2013[38], from among 160 delegate experts.

An initiative is needed to liaise among these research infrastructures with a strong user community and top scientists addressing the challenge of "predicting the biosphere", including the societal grand challenges with respect to biodiversity (and related environmental) change. This initiative should focus on research infrastructure services supporting the requirements as identified by top scientists and other user communities. More specifically, this focus should deal with the concept of Essential Biodiversity Variables (EBV), introduced by the Group on Earth Observations Biodiversity Observation Network (GEO BON) arm of the Global Earth Observation System of Systems (GEOSS). The EBVs are considered to provide a tractable and holistic approach to monitoring (and ultimately, predicting) the state of the environment, thus directly supporting the "understanding, monitoring and conserving biodiversity" societal benefit area of GEO. As such, the cooperating research infrastructures are expecting that support for measuring and calculating EBVs will become a core business in the future.

SEVENTH FRAMEWORK PROGRAMME

8 APPENDIXES

8.1 List of figures

8.2 List of tables

8.3 Bibliography

[1] The CReATIVE-B glossary of terms http://www.creative-b.eu/glossary
[2] The LifeWatch project http://www.lifewatch.eu/
[3] The EC ESFRI http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
[4] The European Commission Framework Programme 7 http://cordis.europa.eu/fp7/home_en.html
[5] The CReATIVE-B project http://creative-b.eu/
[6] The Group on Earth Observations (GEOSS) http://www.earthobservations.org/
[7] The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services http://www.ipbes.net/
[8] The GEO Biodiversity Observation Network (BON) http://www.earthobservations.org/geobon.shtml
[9] The Biodiversity Virtual e-Laboratory (BioVeL) project http://www.biovel.eu/
[10] The Atlas of Living Australia (ALA) http://www.ala.org.au/
[11] The Centro de Referência em Informação Ambiental (CRIA) http://www.cria.org.br/
[12] The CRIA SiBBr Project www.sibbr.gov.br
[13] The CRIA speciesLINK Tool http://www.splink.cria.org.br/
[14] The China Germplasm Bank of Wild Species (GBoWS) http://www.genobank.org/
[15] The DataONE initiative http://www.dataone.org/
[16] The Global Biodiversity Information Facility (GBIF) http://www.gbif.org/
[17] The South African National Biodiversity Institute (SANBI) http://www.sanbi.org/
[18] The CReATIVE-B Project Website and Background Materials http://creative-b.eu/background
[19] The LifeWatch Reference Model http://subs.emis.de/LNI/Proceedings/Proceedings154/gi-proc-154-15.pdf
[20] The GEO BON Information Architecture http://www.earthobservations.org/documents/cop/bi_geobon/geobon_information_architecture_principles.pdf
[21] The Convention on Biological Diversity (CBD) http://www.cbd.int/
[22] The CBD AICHI Biodiversity Targets http://www.cbd.int/sp/targets/
[23] The Global Biodiversity Informatics Outlook (GBIO) http://www.biodiversityinformatics.org/
[24] The outGRID Interoperability Cookbook http://www.outgrid.eu/public/outgrid/download/deliverables/D2.4.pdf
[25] The outGRID Interoperability Analysis http://www.outgrid.eu/public/outgrid/download/deliverables/D2.3.pdf
[26] GEOBON EBV table http://earthobservations.org/documents/cop/bi_geobon/ebvs/201303_ebv_table.pdf
[27] The 4+1 Architectural View Model http://en.wikipedia.org/wiki/4%2B1_architectural_view_model
[28] The Service Oriented Modeling Framework (SOMF) http://www.modelingconcepts.com/pdf/SOMF_2.1_Conceptualization_Model_Language_Specifications.pdf
[29] "A decadal view of biodiversity informatics: challenges and priorities." http://dx.doi.org/10.1186/1472-6785-13-16
[30] https://shibboleth.net/
[31] http://www.geant.net/service/eduGAIN/Pages/home.aspx
[32] http://www.opensearch.org/
[33] http://www.eurogeoss.eu/broker/Pages/AbouttheEuroGEOSSBroker.aspx

[34] http://workspace.earthcube.org/building-blocks, http://workspace.earthcube.org/conceptual-designs
[35] Pereira et.al. (2013) "Essential Biodiversity Variables", Science, Vol. 339.6117 p277-278.
[36] http://www.eubon.eu/
[37] a.k.a. species distribution modelling
[38] http://conference.lifewatch.unisalento.it/index.php/EBIC/BIH2013

