CReATIVE-B

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

Comparison of Technical Basis of Biodiversity e-Infrastructures

Document identifier:	D3.1 Comparison of technical basis of biodiversity e-infrastructures
Date:	**31 October 2012**
Activity:	**Work Package 3, Task 3.1**
Lead Partner:	**CU**
Document Status:	**Issue2.0**
Dissemination Level:	**PUBLIC**
Document Link:	http://creative-b.eu/documents/10826/9f2f2ed9-f6b6-443a-a06d-fd3c46875a84,

ABSTRACT

Deliverable D3.1 "*Technical Interoperability Specifications*", prepared on the basis of available information at the time of writing, is the output of CReATIVE-B task T3.1, which aims to "*Compare the technical basis of e-infrastructures (for biodiversity research)*". It provides a synopsis comparison of the technical approaches of the e-infrastructures analysed within the scope of the project and elaborates the interoperability analysis by defining it and making a quantitative comparison of the technical facts gathered thus far. It reflects as accurately as possible the technical findings, structured along dimensions of interoperability that match the functional areas and layers of the research infrastructures being analysed. Deliverable D3.1 aims at shedding light on existing similarities and differences between participating research infrastructures thus forming a solid information and knowledge basis for future interoperability guidelines developments in D3.2 and D3.3.

SEVENTH FRAMEWORK PROGRAMME

1. COPYRIGHT NOTICE

2. DELIVERY SLIP

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

3. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1.0	18-OCT-2012	ISSUE 1.0	A. Hardisty/CU, D. Manset/MAAT
2.0	31-OCT-2012	ISSUE 2.0	D. Manset/MAAT, A. Hardisty/CU

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, toward 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.

In order to do so, CReATIVE-B collaborates with the international 'sister' RIs to have a second edition of the e-Biosphere conference[7], to be organized in 2014, and to contribute to this conference. In achieving the goals of this coordination and support action, CReATIVE-B will further support the European Commission flagship, vision 2020[8].

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 D3.1 entitled "*Technical Interoperability Specifications*", prepared on the basis of available information at the time of writing, is the output of task T3.1, which aims to "*Compare the technical basis of infrastructures (for biodiversity research)*". It provides a synopsis comparison of the technical approaches of the infrastructures analysed. More specifically, D3.1 aims at shedding light on existing similarities and differences between participating research infrastructures thus forming a solid information and knowledge base onto which developing future interoperability guidelines in deliverables D3.2 and D3.3.

The following picture illustrates the conceptual framework into which it inscribes together with subsequent deliverables, which will be produced. This document constitutes the basis of this framework, where technical information will be gathered and analysed incrementally, to then get translated into interoperability guidelines and recommendations.

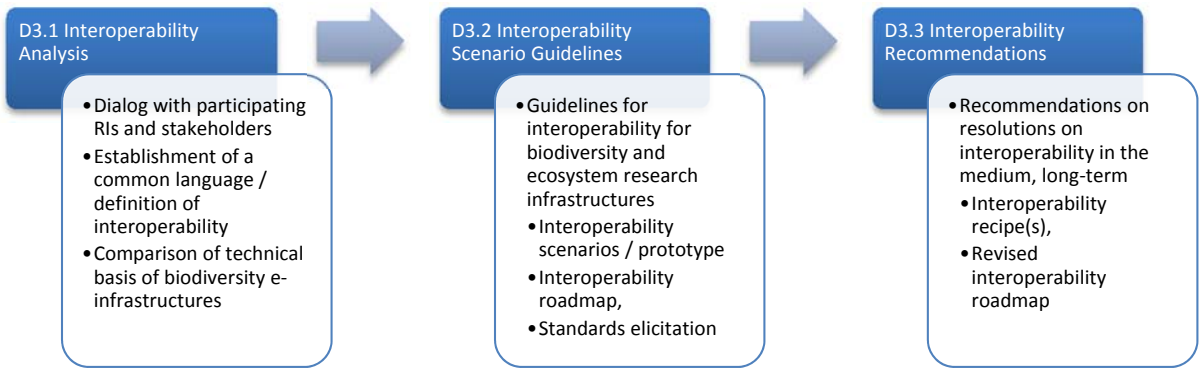


Figure 1. Deliverables conceptual framework

Deliverable D3.1 thus elaborates the interoperability analysis by defining research infrastructures and interoperability, and making a quantitative comparison of the technical approaches undertaken in each of the participating infrastructures. Deliverable D3.2 "*Guidelines for Interoperability*" will then build on this analysis outcome to define a set of interoperability scenarios and to introduce candidate standards where and when relevant. From these guidelines, a resolution containing a roadmap by which stakeholders can evolve towards future interoperability will be produced 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.

This document therefore attempts to reflect as accurately as possible the technical facts gathered thus far, structured along "dimensions of interoperability", that match the functional areas and layers of the infrastructures being analysed. The following sections introduce and decorticate technical information, while proposing a framework to structure and compare technical findings across participating RIs.

SEVENTH FRAMEWORK PROGRAMME

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Background

In Europe, the ESFRI LifeWatch research infrastructure is being constructed to support the biodiversity science community in carrying out research required to address the huge gaps being faced in understanding life on Earth. Pressing problems, such as measuring the impact of environmental change on biodiversity, and predicting / preventing further biodiversity loss arising from the activities of mankind 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 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[9]) in the understanding and the rational management of our ecosystems. It will form the European contribution to GEO BON[10], the biodiversity observation component of GEOSS.

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:

ALA	In Australia, the Atlas of Living Australia (ALA)[11] 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.
cria	In Brazil, the Reference Centre on Environmental Information (CRIA[12]) 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 contributing to this analysis, the SiBBr[13] and speciesLINK[14] projects.
	In China, the Germplasm Bank of Wild Species (GBoWS)[15] 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.

SEVENTH FRAMEWORK PROGRAMME

DataONE	In the USA, the Data Observation Network for Earth (DataONE)[16] 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)[17] 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)[18] 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. The present document draws upon a definition of interoperability and its dimensions when applied to e-infrastructures that has been pioneered in similar FP7 coordination and support action project, "outGRID" (EC FP7 no. 246690)[19]. This approach has been highly successful in allowing stakeholders (from neuGRID[20], LONI[21], CBRAIN[22] neuroscience e-infrastructures) to understand each other and to objectively compare their respective approaches and functions.

1.2 Purpose

This document reports and structures the knowledge gathered thus far from the biodiversity e-infrastructures within the scope of the CReATIVE-B project i.e. focussing on LifeWatch, DataONE, CRIA / SiBBr / speciesLINK, GBIF, GBoWS / CAS, ALA, SANBI and GEOSS / GEO BON). This work has been initially carried out by WP3 "*Technology and Interoperability*", and more precisely within the task entitled "*T3.1 Compare the technical basis of infrastructures*", with the following objective (extract from the project description of work): "*Tabulate the technical basis of cooperating infrastructure with available information on the basis of template. Highlight the similarities and differences of approach between the research infrastructures initiative. The deliverable of this task provides input for the workshop in task 3.2*".

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.

Note: It must be noted that the present document will attempt to avoid repeating information that can be found from respective infrastructures such as schematic system architectures, detailed technical descriptions of concepts and technologies. Instead it will list, classify and quantitatively compare them, while referencing external information sources as appropriate.

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.

Document Title	Description	Background
The CReATIVE-B project website and background materials[23]	*CReATIVE-B, Toward a Global Virtual Environment for Biodiversity Research. Project website and background materials*	*Research Infrastructures*
The LifeWatch Reference Architecture[24]	*Reference architecture model design specifications, based on the Open Data Processing Reference Model*	*Research Infrastructures*
The Principles of the GEO BON Information Infrastructure[25]	*GEO BON concept document covering data integration and interoperability in general terms*	*Research Infrastructures*
The Convention on Biological Diversity[26], AICHI Biodiversity Targets[27]	*Convention on Biological Diversity's AICHI Biodiversity Targets towards strengthening protected area implementation for the conservation of biodiversity*	*Biodiversity*
The Global Biodiversity Informatics Outlook[28]	*A framework for biodiversity intelligence*	*Biodiversity*
The outGRID Interoperability Cookbook[29]	*outGRID interoperability cookbook methodology developed as an adaptable tool to support other e-infrastructures convergence*	*Interoperability Methodology*
The outGRID Interoperability Final Specifications[30]	*outGRID interoperability analysis conclusions for neurosciences e-infrastructures, including interoperability recipes and recommendations*	*Interoperability Methodology*

Table 1. Reference documents

1.4 Logbook of actions carried out

Table 2 reports an exhaustive list of important meetings and actions carried out over the last 12 months of project activity, which have been key in the process of gathering and analysing the participating research infrastructures' information. Additionally, several email exchanges and shared documents editing took place online.

Dates	Actions	Notes	RI
2011-11-07	Kickoff 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 synopsis *Online*	Online sharing and collaborative edition of RI synopsis table and RI descriptions	ALL
2012-09 NOW	D3.1 consolidated draft *Online*	Technical information analysis and specification formalisation in draft D3.1	CReATIVE-B
2012-10-22	EUDAT 1st Conference *International Conference*	Conference attendance	CReATIVE-B

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

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. The present document draws significantly upon that approach and is destined to serve as the basis for further WP3 activities, and for CReATIVE-B's concluding guidelines on interoperability for biodiversity e-infrastructures. The document satisfies several general design objectives and qualities:

(1) To develop a coherent definition of research infrastructure and interoperability that will structure subsequent requirements gathering and technical facts finding. It gives a preliminary structure to interoperability dimensions, and therefore allows:
- Understanding and categorizing infrastructures' major functions,
- Decomposing infrastructures' functions into systems and sub-systems to provide more information about related sets of services, components, standards, formats and related technologies;

(2) To establish a framework for systems and sub-systems' control and communications, useful for harmonizing and gluing the potentially heterogeneous components/technologies from targeted infrastructures,

(3) 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,

(4) 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. Ultimately, it will serve as a solid basis for the following project deliverables, in an incremental approach to consolidating interoperability recommendations:

- D3.2 – "*Guidelines for interoperability for biodiversity and ecosystem research infrastructures*",
- D3.3 – "*Recommendations on resolution on interoperability in the medium to long-term*".

2.2 Approach to interoperability

It is the authors' belief that interoperability across borders and across complex distributed computing platforms being developed by multi-disciplinary consortia require non-invasive information gathering processes, in particular in order not to enter any Intellectual Property Rights (IPR) breaches. Such processes must promote the least possible invasive convergence / integration guidelines, thus facilitating their adoption and further implementation. They must also conclude in quantitative analyses of technical interoperability findings more than qualitative evaluations, thereby respecting each other's choices, which are often the results of complex and internal decision making processes. Interoperability guidelines must encourage as much as possible open standards and be in line with the state-of-the-art, at the time they are being issued.

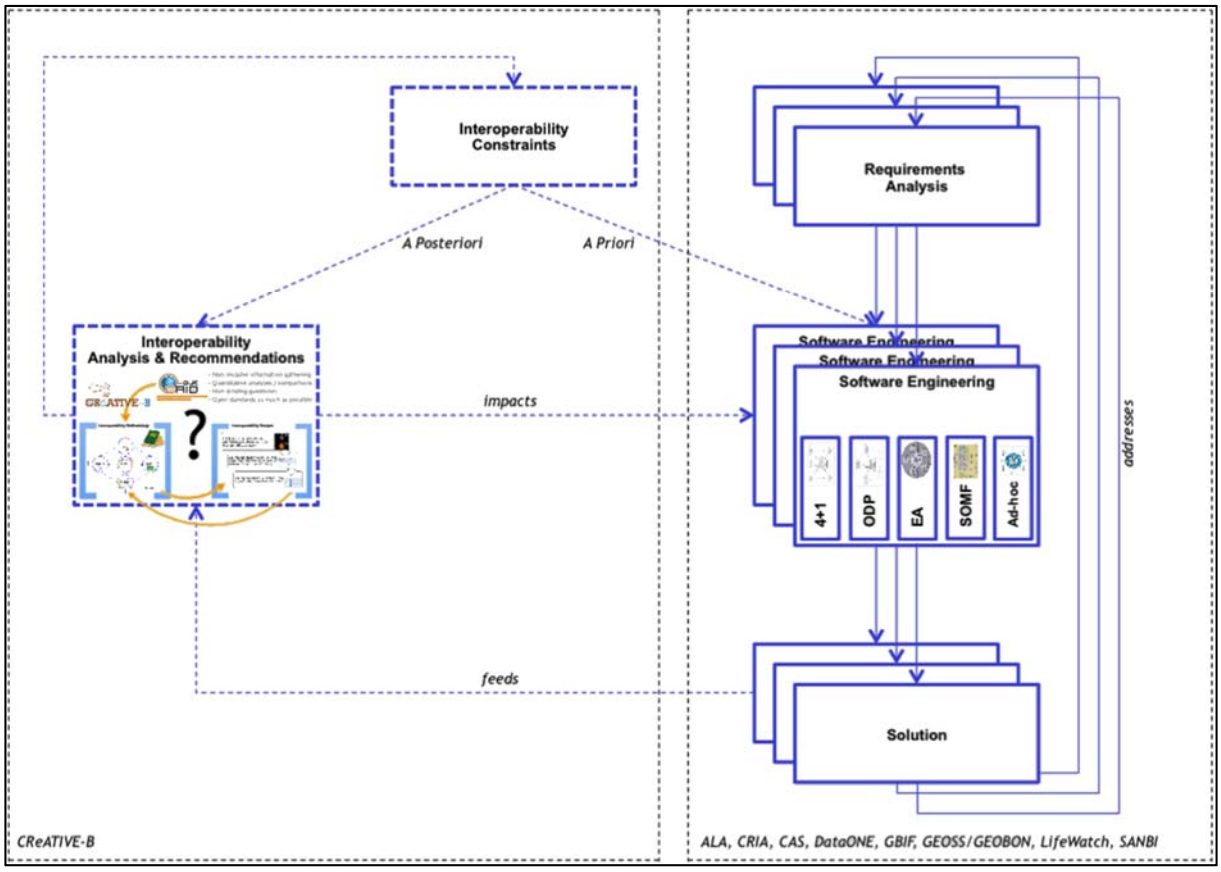


Figure 2. Approach to interoperability

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[31], to Service Oriented Modelling Framework (SOMF[32]) 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 analyse differences and similarities, towards identifying interoperability enablers and prioritizing 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 will be further explored, as the project will progress in the development of deliverables D3.1, D3.2 and D3.3:

(1) Carrying out an interoperability analysis

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 will be formalised in CReATIVE-B deliverable D3.1

(2) Identifying interoperability scenarios

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 will be formalised in CReATIVE-B deliverable D3.2

(3) Specifying an interoperability roadmap

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 will be formalised in CReATIVE-B deliverable D3.2

(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

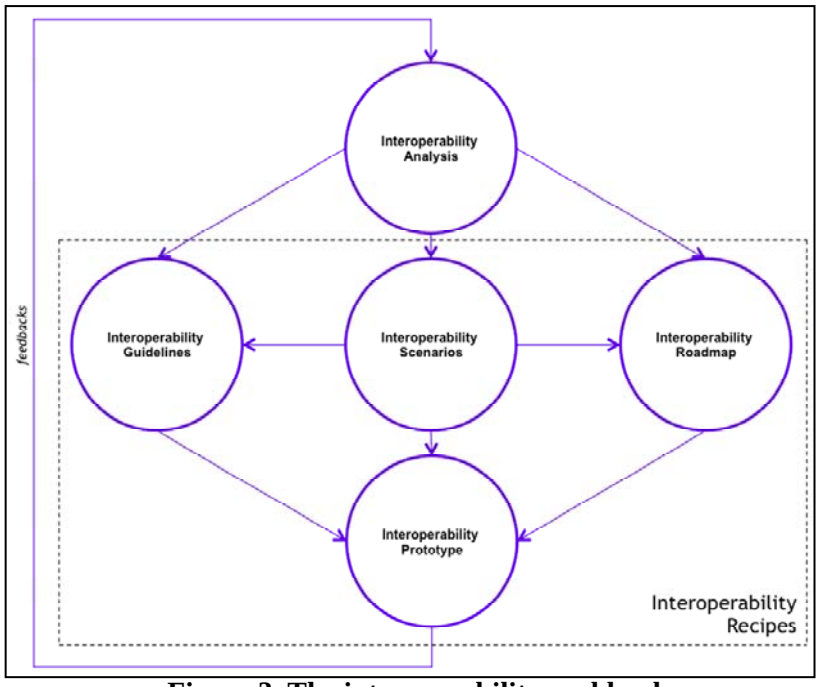


Figure 3. The interoperability cookbook

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

The following section elaborates on technical information gathered thus far from participating research infrastructures. A rather succinct description is first given of measured technical facts. The latter are then decorticated in details and compared along various dimensions to highlight differences and commonalities between infrastructures, in order to inform next steps of the process, in particular in prioritizing the interoperability roadmap.

Note: a concrete and end-to-end application of this interoperability cookbook can be found further detailed in the following case study[33]:

Field: Neurosciences	**Project**: outGRID www.outgrid.eu
Three e-infrastructures for computational neuroscience are presently active or under construction worldwide. In Europe, neuGRID (www.neugrid4you.eu) aims to provide large sets of brain images paired with grid-based computationally intensive algorithms for studies of neurodegenerative diseases. In Canada and the U.S.A, (http://www.loni.ucla.edu/) CBRAIN (https://cbrain.mcgill.ca) at McGill in Montreal and LONI (http://www.loni.ucla.edu) - Laboratory of Neuro Imaging at UCLA offer computational resources and algorithm pipelines Aim of outGRID is to ignite the process to lead the 3 e-infrastructures to converge into one unique worldwide facility. This case study analyzed native interoperability between the 3 e-infrastructures and produced a detailed set of scenarios and recommendations for implementation.	
References	• Interoperability analysis and conclusions http://www.outgrid.eu/public/outgrid/download/deliverables/D2.3.pdf • Interoperability prototype https://www.shiwa-workflow.eu/success-stories

3 KNOWLEDGE ACQUISITION & INTEROPERABILITY ANALYSIS

3.1 Background introduction

3.1.1 The Atlas of Living Australia (ALA) infrastructure

URL: http://www.ala.org.au/
URL: http://www.ala.org.au/about-the-atlas/

3.1.1.1 Overview and purpose

Figure 4. The ALA web portal

The intent of the Atlas of Living Australia (ALA) is to create a national infrastructure giving access to information about all of Australia's biota, accessed through a single, easy to use web site (see figure 4 illustration) and to provide web services to enable the construction of external portals. The ALA mobilizes and integrates biodiversity data and provides mechanisms for data discovery. It provides support for taxonomy by, for example offering services on top of national checklists, and collaboration environments such as TRIN wiki[34]. It supports natural history collections and herbaria - through data mobilization and accompanying tools[35] and provides support for ecological research[36], and specialized portals for community groups. The ALA provides an extensive suite of environmental and contextual layers, see [37], to complement biodiversity information. The ALA is built largely on web services, e.g. [38]. Information available through ALA is used to:

- Improve understanding of Australian biodiversity,
- Assist research scientists to build a more detailed picture of Australia's biodiversity,
- Assist environmental managers and policy makers develop more effective means of Managing and sustaining Australia's biodiversity and
- Provide an educational platform for ages 5 to university level.

Note: The ALA is the Australian GBIF node.

3.1.1.2 Architecture

Figure 5 below and its accompanying notes provide an overview of the technical architecture of the ALA e-infrastructure.

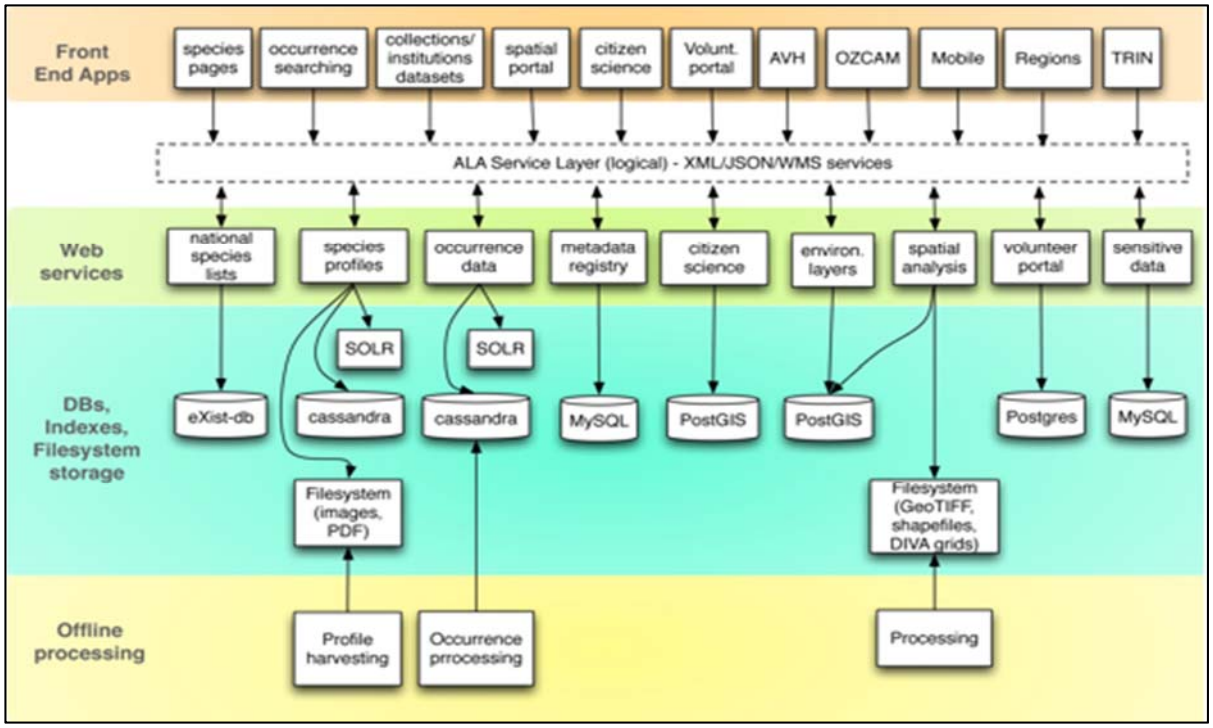


Figure 5. Architecture of the Atlas of Living Australia

Note 1: The following components are missing for brevity:

- *Geo Network* [39] *- geospatial metadata registry used to share metadata with other geospatial networks (e.g. Auscope)*
- *Email Alert Service*[40] *- providing email alerts on changes within ALA components (e.g. new data, layers)*
- *Sandbox Service*[41] *- an environment for uploading Darwin Core data and using ALA's occurrence data processing & visualisation*
- *Single Sign On Service*[42] *- provides single sign on access across the ALA*

Note 2: "Spatial analysis" consists of 2 software components:

- *Layers-service - this provides abilities to query a contextual (e.g., political boundaries) and environmental layers.*
- *Analysis toolkit - provides support for distribution modeling (e.g., MaxEnt and area classifications).*

Note 3: "Environmental layers" uses Geoserver[43].

Note 4: "Metadata registry" refers to a component known as the "Collectory" within the ALA.

Note 5: "Occurrence data" refers to a component known as the "Biocache" within the ALA.

Note 6: "Species profiles" refers to a component known as the Biodiversity Information Explorer (BIE) within ALA.

Note 7: For Volunteer portal, Citizen science, Collectory - the same deployed application provides the front end and the web services.

Note 8: The following links exist:

Spatial portal & Spatial Analysis - http://spatial.ala.org.au - http://spatial.ala.org.au/layers - http://spatial.ala.org.au/ws - http://spatial.ala.org.au/ws/examples -http://code.google.com/p/alageospatialportal/ Species pages - http://www.ala.org.au - http://bie.ala.org.au/ws - http://code.google.com/p/ala-bie/ Volunteer portal - http://volunteer.ala.org.au - http://code.google.com/p/ala-volunteer/ Citizen science - http://bdrs-uat.ala.org.au - http://cs.ala.org.au - http://code.google.com/p/ala-citizenscience/ Occurrence data - http://biocache.ala.org.au - http://biocache.ala.org.au/ws - http://code.google.com/p/ala-portal/ National Species Lists - http://biodiversity.org.au - http://code.google.com/p/ala-nsl/	OZCAM - http://ozcam.ala.org.au AVH - http://avh.ala.org.au Sensitive Data Service - http://sds.ala.org.au- http://code.google.com/p/alageospatialportal Sandbox - http://sandbox.ala.org.au/ - http://code.google.com/p/ala-portal/ Metadata registry - http://collections.ala.org.au - http://collection.ala.org.au/datasets - http://collections.ala.org.au/ws - http://code.google.com/p/ala-collectory/ Regions - http://regions.ala.org.au TRIN - http://wiki.trin.org.au/

Table 3. The ALA applications

3.1.1.3 Data

An overview of the data available through the ALA can be found at [44]. The ALA high-level goal is to incorporate and integrate all sources of biodiversity data, including:

· Occurrence data - specimen, observation records typically in Darwin Core format

- Environmental data and polygonal data such as climatic surfaces and bio-regionalization of Australia
- Australian and international gazetteer information (including gazetteer entries from contextual polygonal layers)
- Identification keys
- Literature (e.g., from [45])
- National species checklists - includes taxon concepts, taxon names, publications, references
- Ad-hoc lists of species properties e.g. Habitat, Conservation status
- Sensitivity data rules for species (the 'Sensitive Data Service) across national and State/Territory jurisdictions.
- Multimedia for taxa - video, sound, images
- Attribution information - relationships between institutions, datasets, collections, and metadata for all of these entities, and licensing information
- Species interactions data.

Data shared through ALA is subjected to checking against a pre-defined quality model[46].

3.1.1.4 Services and tools

ALA offers a range of services to its user community, namely:

- Data discovery and access,
- Data publication - datasets are submitted through the website and then exposed through the ALA tools,
- Distribution analysis, modeling through tools such as Maxent, GDM (accessed through the research portal,
- Sandboxing environment for data through[47],
- Web services for supporting external biodiversity portals,
- Citizen science portals for observation gathering and viewing,
- Crowd sourcing[48],
- Data licensing agreements to assist data publication,
- Taxonomy profile working areas.

3.1.1.5 Typical applications

Typical applications supported by ALA include:

- Taxonomy,
- Ecological research,
- Education,
- Community portals,
- Citizen science.

3.1.1.6 Standards

ALA uses a range of technical standards (e.g., for data formats, for metadata, protocols for data exchange, etc.), to name a few: Darwin Core, Darwin Core Archive, EML, TAPIR/DIGiR/BioCASe, OAI-PMH, OGC standards such as WMS, RIF-CIS, Support for EOL schemas for data publication, JSON including GeoJSON, KML, WKT and Shapefiles.

3.1.2 The Brazilian SiBBr and CRIA SpeciesLINK infrastructures

URL: http://splink.cria.org.br/index?&setlang=en
URL: http://www.cria.org.br/

3.1.2.1 SiBBr overview and purpose

Figure 6. The SiBBr web portal

The goal of the recently announced[49] Brazilian Biodiversity Information System (SiBBr) is to consolidate the entire 'library of life' in Brazil into a unique scientific information system. With Brazil, the most biodiverse in the world being home to about 13 percent of all known plant and animal species, SiBBr will help policymakers to identify the best options for conservation and sustainable use of Brazilian biodiversity. SiBBr will be built by:

(i) consolidating the infrastructure, instruments, tools, and technology required to qualify, gather and make the biodiversity information contained in the resources of the country's biological collections freely available online;

(ii) strengthening institutional and taxonomic capacities to ensure continuous uploading and updating of information; and

(iii) development of products and services that will allow key decision-makers to establish policies that integrate biodiversity conservation and sustainable use objectives into the operations of the productive sectors.

SiBBr builds upon and integrates an existing base of infrastructure that has been emerging over the past years, in particular the speciesLINK.

3.1.2.1.1 Architecture

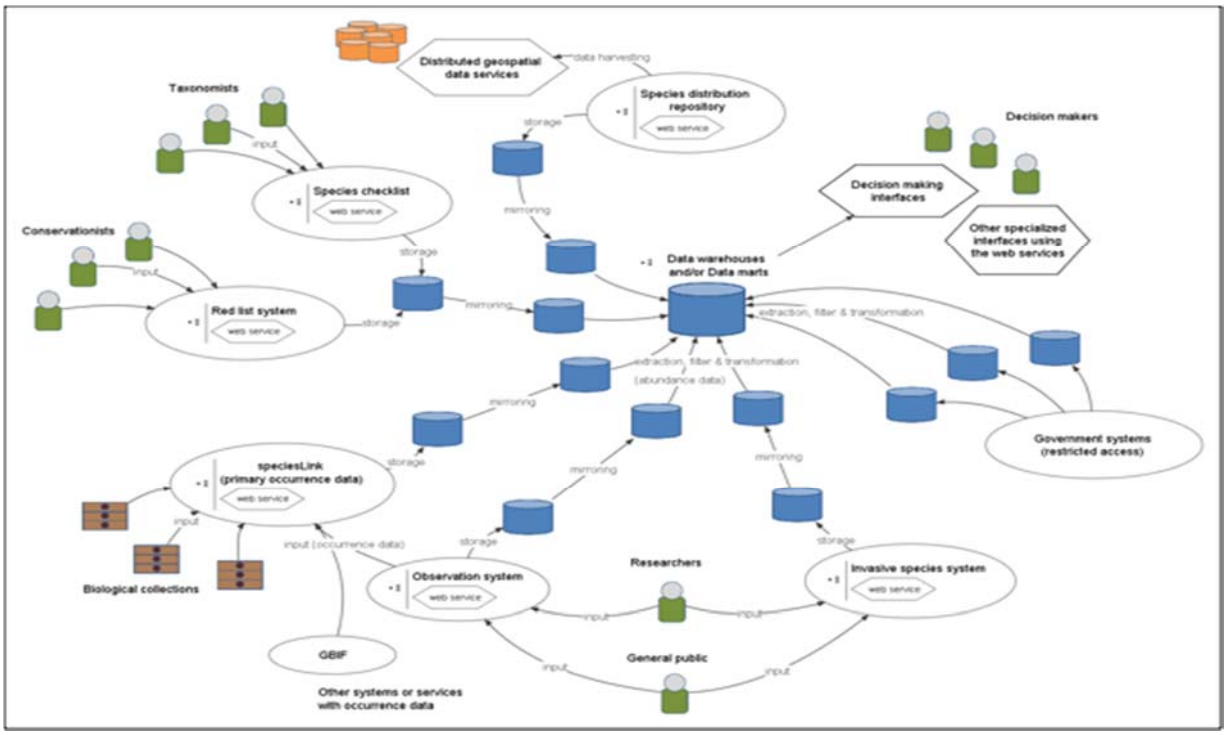


Figure 7. The SiBBr architecture

3.1.2.1.2 Data

Not available at the time of writing.

3.1.2.1.3 Services and tools

Not available at the time of writing.

3.1.2.1.4 Typical applications

Not available at the time of writing.

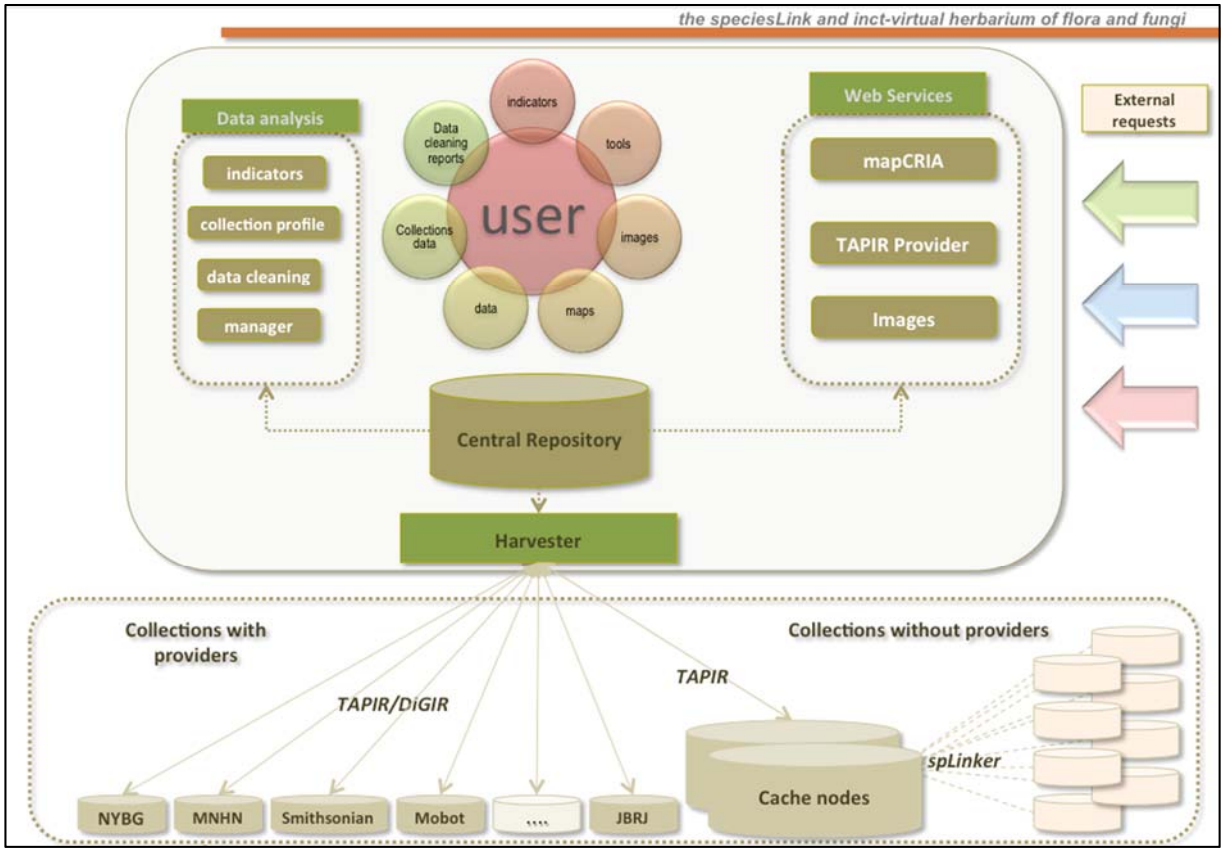
3.1.2.1.5 Standards

Not available at the time of writing.

3.1.2.2 CRIA speciesLINK overview and purpose

Figure 8. The CRIA speciesLINK web portal

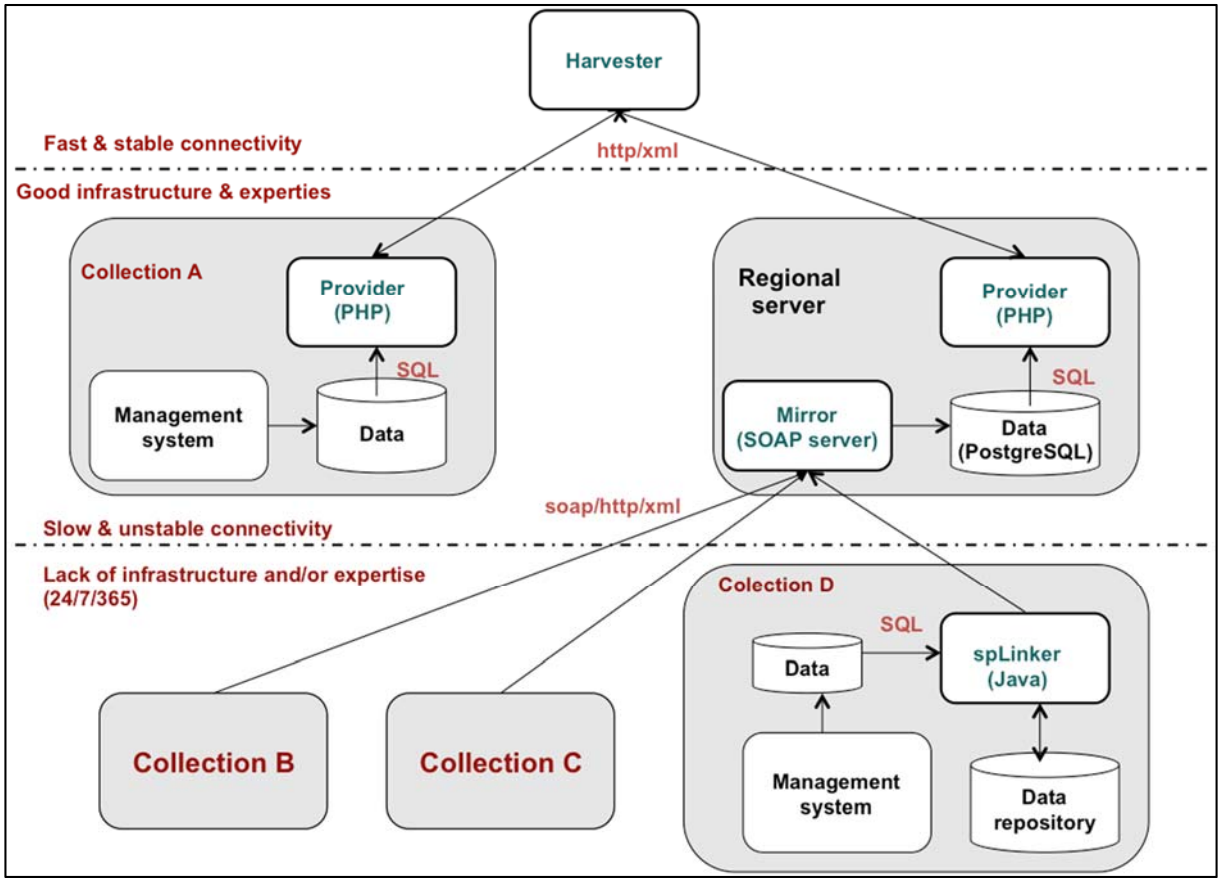
3.1.2.2.1 Architecture



3.1.2.2.2 Data

Not available at the time of writing.

3.1.2.2.3 Services and tools



3.1.2.2.4 Typical applications

Not available at the time of writing.

3.1.2.2.5 Standards

Not available at the time of writing.

3.1.3 The GBoWS / Chinese Academy of Sciences infrastructure

URL: http://www.genobank.org/

3.1.3.1 Overview and purpose

Figure 9. The GBoWS genobank web portal

Not available at the time of writing.

3.1.3.2 Architecture

Not available at the time of writing.

3.1.3.3 Data

Not available at the time of writing.

3.1.3.4 Services and tools

Not available at the time of writing.

3.1.3.5 Typical applications

Typical applications provided by the CAS Institutes include:

1) Institute of Botany[50]
   - Chinese Virtual Herbarium = digitization of specimens. 3.31m specimens. i) specimen info ii) associated literature from Flora China iii) images iv) names (from CoL china node). Generally much more occurrence data than GBIF holds.

- Specific CoL CD for China, produced in China. 65k species.
- Chinese Field Herbarium – color photos upload. Citizen science? 2.5m images of which 600k have GPS coordinates[51].

2) Institute of Zoology[52]

Group on Biodiversity Informatics -

- Demonstration of Species 2000 Multi-Hub System[53]
- Taxonomic Tree Tool[54]

3) Kunming Institute of Botany[55]
   - China Germplasm Bank of Wild Species
4) Kunming Institute of Zoology
   - Mitotool: a database and web server for the retrieval and analysis of human mitochronidal DNA sequence variations[56]
   - China Animal Scientific Database: a database of animal resources in China[57]
   - Database of animal resources in South-west of China[58]
   - Antimicrobial Peptide Database[59]

3.1.3.6 Standards

Not available at the time of writing.

3.1.4 The DataONE infrastructure

URL: https://www.dataone.org/
URL: http://mule1.dataone.org/

3.1.4.1 Overview and purpose

Figure 10. The dataONE web portal

Data Observation Network for Earth (DataONE) is a distributed e-infrastructure for environmental sciences that enables new science and knowledge creation by providing open, persistent, robust, and secure access to well-described and easily discovered data about life on earth and the environment that sustains it.

DataONE Member Nodes are the distributed data centres, science networks or organizations that expose their data within the DataONE network. In addition to scientific data, Member Nodes can provide computing resources, or services such as data replication, to the DataONE community. Member Nodes may eventually number in the 1000's as many smaller data providers put their data on-line. A small number of mirrored DataONE Coordinating Nodes provide coordination and indexing services, making it easy for scientists to discover data held on Member Nodes, and also enables data repositories to make their data and services more broadly available to the international community. Facilitated by community outreach, engagement, training and education activities, the DataONE initiative aims for its infrastructure to be commonly used by researchers, educators, and the public to better understand and conserve life on earth and the environment that sustains it.

3.1.4.2 Architecture

As illustrated by Figure 11, the DataONE e-infrastructure consists of: Member Nodes which represent data repositories; Coordinating Nodes which serve data management and discovery services; and the Investigator Toolkit which contains a variety of end user tools for interacting with the infrastructure.

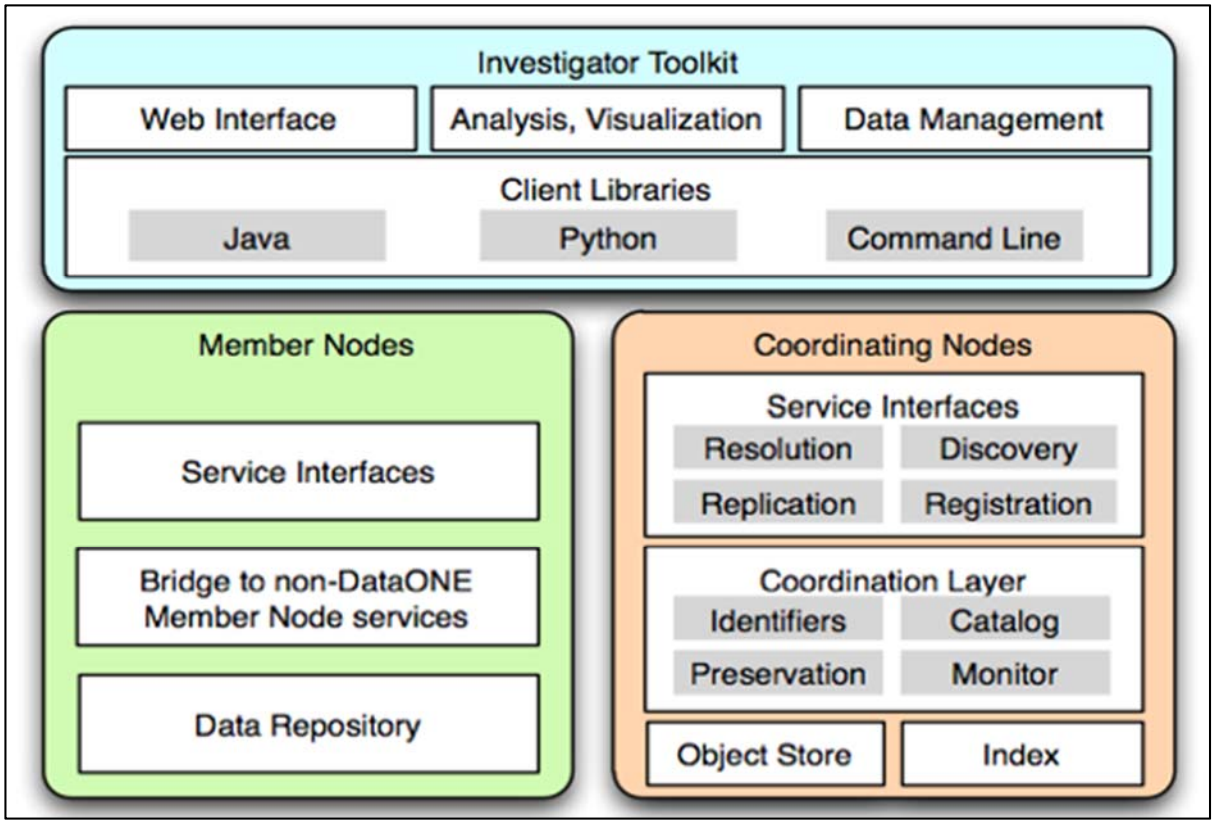


Figure 11. Major components of the DataONE e-infrastructure[60]

Member Nodes are primarily existing data repositories (e.g. Dryad, the Knowledge Network for Biodiversity, ORNL DAAC) that already fill an important role in their respective communities supporting data management, curation, discovery and access functions. Existing or new repositories can participate in the DataONE infrastructure by implementing a simple set of APIs (Application Programming Interfaces), which represent a convergence of functionality expressed in a variety of existing systems. These APIs include basic operations such as listing and retrieving objects, support for creation of content, and the ability to generate low level system metadata describing the various objects (data, metadata) exposed by the service. Member Nodes may implement a subset of the full suite of Member Node APIs, and in this way participate in the network with minimal effort (e.g. as a “read only” data source). Member Nodes that implement the full suite of APIs will be able to accept data from other Member Nodes which in turn assists with data preservation by ensuring multiple copies of all content are available, thus reducing the risk that content will be lost or inaccessible if a Member Node should go offline.

Coordinating Nodes[61] provide network-wide services to enhance interoperability of the Member Nodes and to support indexing and replication services. Coordinating Nodes implement critical services through the APIs to enable identifier resolution, data preservation, data discovery, and to supplement the federated identity system. Coordinating Nodes replicate all content between themselves, and in doing so create a small set (3-6 nodes) of geographically and institutionally isolated systems that ensure ongoing operation of the infrastructure should any particular node be inaccessible. Coordinating Nodes maintain complete copies of all science metadata (detailed descriptions of science data objects and collections) and system metadata (low level metadata describing the type, size, ownership, and locations of data and) and index this information to enable data discovery services.
The **Investigator Toolkit** (ITK, see section 3.1.4.4 below) enables access to customized tools that are familiar to scientists and that can support them in all aspects of the data life cycle.

These components and their relations are illustrated in more detail in Figure 12 and further described in detail in the DataONE Architecture Documents[62].

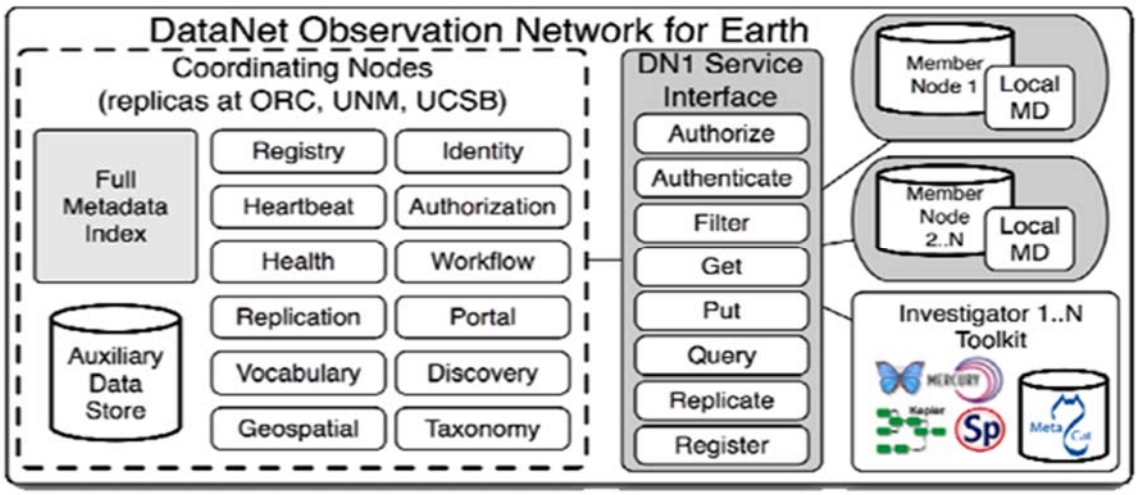


Figure 12. Detailed architecture of DataONE

3.1.4.3 Data

Data, in the context of DataONE[63], is a discrete unit of digital content that is expected to represent information obtained from some experiment or scientific study. The data is accompanied by science metadata, which is a separate unit of digital content that describes properties of the data. Each unit of science data or science metadata is accompanied by a system metadata document that contains attributes that describe the digital object it accompanies (e.g., time stamps, ownership, relationships). This is for internal use.
In the initial version of DataONE, science data are treated as opaque sets of bytes stored on Member Nodes (MN). A copy of the science metadata is held by Coordinating Nodes (CN) and is parsed to extract attributes to assist the discovery process (i.e. users searching for content). The opaqueness of data in DataONE is likely to change in the future to enable processing of the data with operations such as translation, extraction, and merging.

3.1.4.4 Services and tools

Investigator Toolkit is a suite of software libraries, tools, and applications that support interaction with the DataONE infrastructure through the REST service APIs exposed by the Coordinating and Member Nodes. Low level libraries are initially available in Python and Java which assist application developers to take advantage of the core services exposed by DataONE participants. For example, an R plugin has been developed using the Java library. Enabling this plugin within a R script enables discovery, retrieval, and storage of content directly in the DataONE infrastructure. Similar extensions are being developed for workflow tools such as Kepler, VisTrails and Science Pipes to enable interaction with the core DataONE services. A catalogue of software tools[64] capable of making use of data accessible through DataONE is provided. The listed tools are wide-ranging in scope.

3.1.4.5 Typical applications

Because DataONE intends to be a general purpose data infrastructure there are no typical applications referred to as such. The DataONE Best Practices database[65] provides individuals with recommendations on how to effectively work with their data through all stages of the data lifecycle (Figure 13). Users can access best practices within the database based upon a stage of the lifecycle.

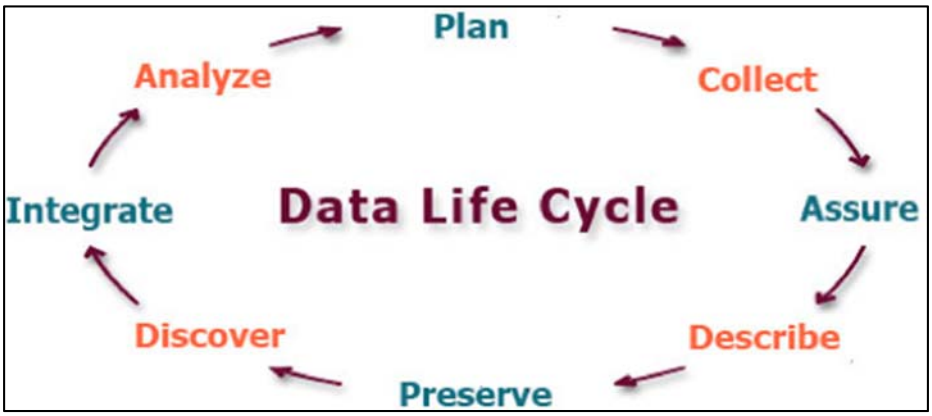


Figure 13. The data life cycle (source: DataONE)

Users new to data management can use a Best Practices Primer as an introduction to the DataONE Best Practices database and data management in general.

3.1.4.6 Standards

DataONE utilizes existing standards where possible. For science metadata documents, this includes FGDC, ISO19115 (and derivations), EML, and similar standards (more can be supported as necessary). Content packaging is represented using OAI-ORE. DataONE is identifier agnostic, and will work with any existing identifier format including, but not limited to: UUIDs, URI, URL, PURL, DOI, ARK, etc. Authentication is performed using client side RFC5280 X.509 certificates generated by CILogon, which in turn utilizes a number of standards. All messaging in DataONE is XML encoded in structures defined with XML-Schema.

3.1.5 The GBIF infrastructure

URL: http://www.gbif.org/
URL: http://data.gbif.org/welcome.htm

3.1.5.1 Overview and purpose

Figure 14. The GBIF web portal

The Global Biodiversity Information Facility (GBIF) was established by governments in 2001 to encourage free and open access to biodiversity data via the Internet. Through a global network of countries and organizations, GBIF promotes and facilitates the mobilization, access, discovery and use of information about the occurrence of organisms over time and across the planet. At the time of writing, 422 data providers and 10,067 resources have been registered with and published through the GBIF Global Biodiversity Resources Discovery System (GBRDS)[66]. GBRDS is an annotated index of publishers, institutions, networks, collections, datasets, schemas and services that allows GBIF to harvest data into a dynamic, regularly refreshed index which is fronted by a web portal and web services providing unified search and retrieval across the whole network[67].

As a mega-science initiative, GBIF aims to provide an essential global informatics infrastructure for biodiversity research and applications worldwide. The GBIF indexing covers primary biodiversity occurrence records, dataset descriptive content (metadata), and also checklists (taxonomic, nomenclatural or other).

3.1.5.2 Architecture

Figure 15 illustrates how the major informatics components are integrated within the GBIF infrastructure. Note that not all functions described below are shown in Figure 15.

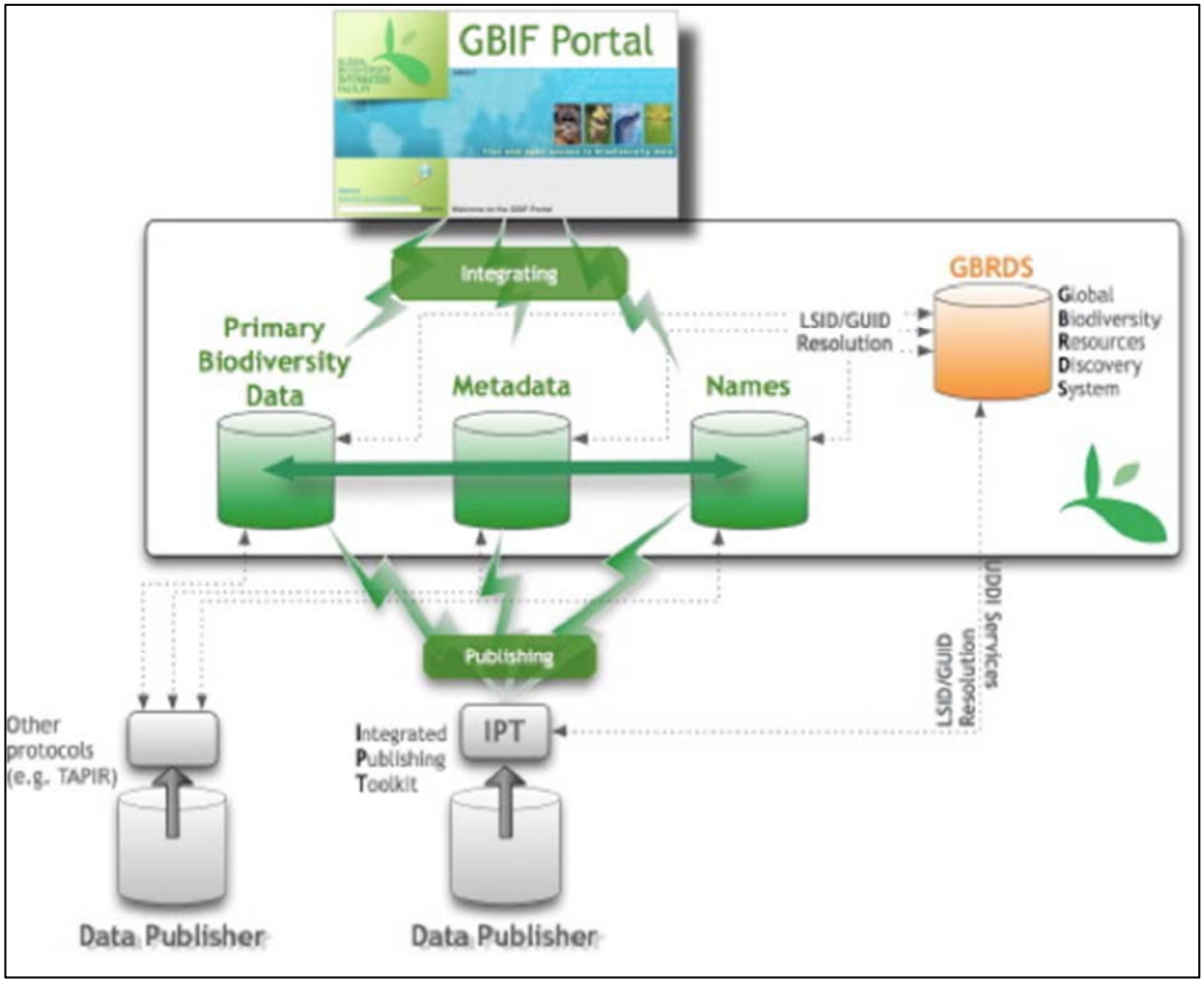


Figure 15. Architecture of GBIF

The infrastructure is divided into five major components:

1) **Publishing** is concerned with managing and publishing three types of data (taxon primary occurrence data, taxonomic checklists and resource metadata) imported from multiple data providers. Data can be published in standards-compliant DarwinCore Archives (DwC-A), BioCASe, TAPIR, DiGIR and Ecological Modeling Language (EML v2.1.1) and indexed through the GBRDS for discovery. The Integrated Publishing Toolkit (IPT) is one recommended approach for data publishing although other mechanisms (DiGIR, BioCASE or TapirLink) are also supported.

2) **Discovery** (both interactive via a portal and programmatic via Web services) is carried out via the data portal - a single annotated index of publishers, institutions and collections, a schema repository and associated services. Discovery of all kinds of information about digitised and not digitised biodiversity resources (i.e., the 'who, what, where, when and how') is possible. During 2012-13 GBIF will be rolling out a new set of web services and a web application enhancing discoverability.

3) **Indexing** (using the Harvesting Index Toolkit (HIT)) simplifies the otherwise complicated process of harvesting biodiversity data from a distributed network of data publishers. It also supports building an index of all harvested data, providing a means by which the data can be assimilated into a central point. This can be useful for creating regional or national data aggregations.

4) **Integrating** as single point of access through the GBIF Data Portal allows users to perform complex searches. These can be on any taxon, country, dataset or combinations of these parameters, with filters such as altitude, depth etc. They are performed in parallel across the worldwide network of biodiversity data providers that publish data through GBIF.

5) **Retrieving** data in various ways allows its use in a wide range of different ways. Occurrence records can be downloaded in a variety of formats, including those suitable to allow OGC-compliant clients to include data as layers in the generation of maps, and other geospatial analysis functions. A range of Web services allows direct access to XML-formatted GBIF data so that it can be used in different programmatic applications. Future APIs will be built around JSON based service layers.

3.1.5.3 Data

The GBIF infrastructure is designed to serve the following types of data:

- Metadata: data about data, including both the minimum set required for data discovery and retrieval, and more comprehensively to describe in detail the characteristics of the data retrieved.
- Primary biodiversity data: the digital text or multimedia data records that detail the instance of an organism – the 'what, where, when, how and by whom' of the organism's occurrence and recording. At the time of writing 388 million records of this data were available through the GBIF Data Portal
- Names data: All information about a species is tied to a scientific name. Many species, however, may be known by multiple scientific and vernacular names. At the same time, the same name may refer to different species. These and other issues affect accessibility and interoperability within the GBIF network and thus GBIF aims to provide access to multiple various sources of names information.
- Support for the aggregation and delivery of genomic data is also being planned.

SEVENTH FRAMEWORK PROGRAMME

3.1.5.4 Services and tools

GBIF provides a comprehensive set of REST based web services.
For more information, see[68].

3.1.5.5 Typical applications

GBIF does not provide applications as such. However, a number of typical applications making use of GBIF data have been explained / showcased by GBIF. These include:

- Taxonomic revisions,
- Ecological niche modeling,
- Compiling 'red lists' of threatened species,
- Biodiversity assessments.

Other projects and initiatives increasingly deliver applications that make use of GBIF data and there are increasing numbers of studies published that have made use of data sourced through GBIF. Since 2008, more than 500 peer-reviewed papers in academic journals have cited use of GBIF-mediated data. 204 of those papers were published in 2011.

3.1.5.6 Standards

GBIF's informatics infrastructure builds on existing and emerging standards and tools and GBIF plays an active part in their development. This is carried out in close collaboration with Biodiversity Information Standards (TDWG)[69] – the group focuses that on the development of standards for the exchange of biological/biodiversity data. Darwin Core, Darwin Core Archive, DiGIR, BioCASE and TAPIR are all important standards supported by GBIF. GBIF supports multiple metadata models natively including: Ecological Metadata language (EML), ISO 19115/19139, Natural Collections Descriptions (NCD), and FGDC Biological Profile.

3.1.6 The GEOSS / GEO BON infrastructure

URL: http://www.earthobservations.org/geobon.shtml
URL: http://www.earthobservations.org/index.shtml
URL: http://www.geoportal.org/web/guest/geo_home
Source documents:
- Principles of the GEO BON Information Architecture, Version 1.0 - 24 June 2010
- The GEO Biodiversity Observation Network Implementation Overview[70]
- GEO BON Detailed Implementation Plan, Version 1.0 – 22 May 2010[71]

3.1.6.1 Overview and purpose

Figure 16. The GEOSS / GEO BON web portal

GEO BON (Group on Earth Observations / Biodiversity Observation Network) is the main task of GEO in the biodiversity "Social Benefit Area". As part of GEOSS (Global Earth Observation System of Systems), GEO BON is concerned with data integration and interoperability[72]. Building on existing networks and initiatives, GEO BON proposes an implementation plan for "*an informatics network in support of the efficient and effective collection, management, sharing, and analysis of data on the status and trends of the world's biodiversity, covering variation in composition, structure and function at ecosystem, species and genetic levels and spanning terrestrial, freshwater, coastal, and open ocean marine domains.*"

The GEO BON system will be built largely from contributing systems that have their primary responsibility at regional, national or sub-national scales. At the European level, for example, it is envisaged that LifeWatch will form part of the European contribution to GEO BON[73]. EU BON is a new FP7 project that will be building the actual linkages and the required infrastructure components in 2012-2016. Among many aspects, GEO BON will need to harmonise observation standards, to promote use of multidisciplinary interoperability

standards, and to define and update interoperability arrangements – applying the System of Systems approach promoted and implemented by GEOSS. GEO BON will also help to promote data publication principles in support of full and open availability of data and information, recognizing relevant international instruments and national policies and legislation.

3.1.6.2 Architecture

In keeping with the GEOSS conceptual approach, the informatics infrastructure for GEO BON will be based on a decentralized and distributed Service Oriented Architecture.

GEO BON fits into the broader conceptual space of GEOSS and exploits the functionality provided by the GEOSS Common Infrastructure (GCI). Components, services, standards and special interoperability arrangements contributing to GEO BON will be discoverable through the GEOSS Clearinghouse as a result of being entered in the appropriate registry (Components Registry, Services Registry, Standards and Special Arrangement Registry[74]). One of the main tasks for GEO BON is thus to identify the main components contributing to the network, list the services they provide and the standards or special interoperability arrangements used by those services.

3.1.6.3 Data and services

Because GEO BON builds on existing networks and infrastructures and strives to create interoperability between them, it does not specifically offer access to data and services. Instead, it offers a broker model in which data and service interoperability are delivered to the data / service consumer via mapping and mediation components. Domain-specific metadata catalogues and registries (such as those from GBIF, ILTER, etc.) are expected to play a role in GEO BON. These catalogues would connect to the GEOSS Clearinghouse by implementing the required interface based on the OGC CSW (Catalog Service for the Web) specification.

3.1.6.4 Technical standards

GEO BON infrastructures must implement the SOA international standards and Earth system science multidisciplinary best practices, e.g., GEOSS Standards and Interoperability Forum (SIF) interoperability arrangements, including relevant standards of OGC, W3C, etc. Table 3 in the document: Principles of the GEO BON Information Architecture summarizes the important standards for the biodiversity domain.

3.1.7 The LifeWatch research infrastructure

URL: http://www.lifewatch.eu/, http://portal.lifewatch.eu/
Source document(s): LifeWatch Reference Model v1.0[75]

3.1.7.1 Overview and purpose

Figure 17. The LifeWatch web portal

The LifeWatch infrastructure for biodiversity and ecosystem research will allow scientists to tackle the big basic questions in biodiversity research, as well to address the urgent societal and fundamental scientific challenges concerning our living planet. Advanced systems oriented research on the complex biodiversity system is supported through dedicated virtual environments, enabling integrated access to data, analytical and modeling workflows and computational capacity. It is a new approach for large-scale cooperation in simulation and scenario development experiments. Managing our planet's biosphere is one of our greatest challenges – but meeting it requires deeper knowledge. In LifeWatch, policy makers can work on these problems directly or with assistance from local researchers in platforms actively investigating a policy-science interface. Environmental policy can move from uncertainty to increased confidence and safety.

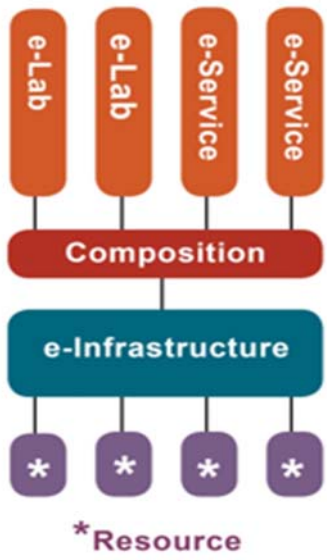
LifeWatch will construct and bring into operation the facilities, hardware, software and governance structures in an e-Infrastructure for research on the protection, management and sustainable use of biodiversity. The infrastructure includes enabling facilities for data

generation and processing; a network of observatories and sensors; facilities for data integration and interoperability; capabilities to create work flows of analytical and modeling tools; and a Service Centre providing special services for scientific and policy users, including training and research opportunities for young scientists. Its Grid-enabled Service-oriented design supports access to and the integration of external resources such as data from associated infrastructures and distributed computational capacity from high performance clusters. It can be characterized as a Collaborative Spatial Data Infrastructure sitting on top of available distributed computing infrastructures.

User groups may create their own e-laboratories or e-services within the common architecture of the infrastructure. Any new data will be channeled to the appropriate external resources (such as GBIF). LifeWatch enables distributed large scale and collaborative research on complex and multidisciplinary problems.

3.1.7.2 Architecture

Figure 18 and 19 below illustrate the architecture of the LifeWatch e-infrastructure.



The lowest layer (purple) shows the (external) data, software and computational resources. The e-infrastructure layer (turquoise) integrates these, while the next layer (red/orange) allows for composing preferred workflows. Users can access these services through specialized e-laboratories, or e-services, and can build their own ones. This illustration is expanded in Figure 8, which gives a service-oriented view. Further details can be found in the LifeWatch Reference Model document.

Figure 18. Functional domains of the LifeWatch architecture

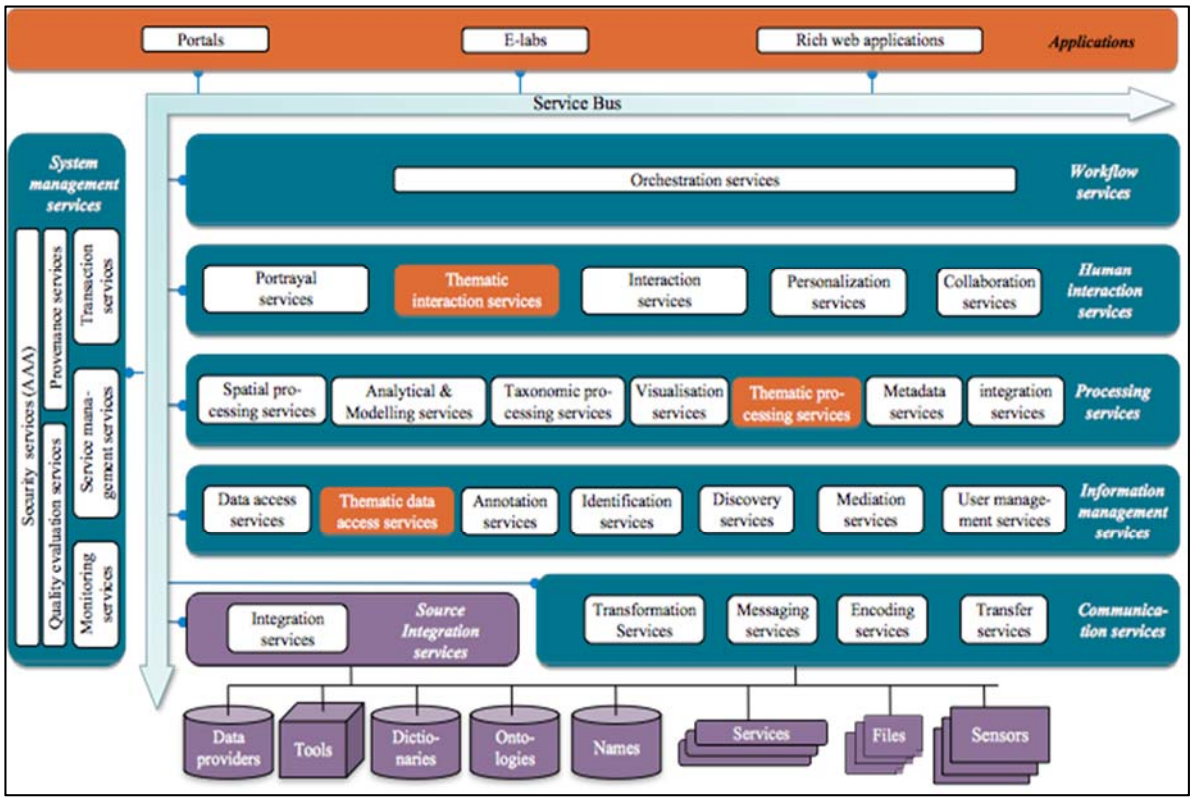


Figure 19. Service-oriented architecture of LifeWatch

3.1.7.3 Data

LifeWatch is not a research infrastructure generating data, but an environment to discover, process, model and collaborate with data (including software tools). The primary data available in LifeWatch originates from various sources, such as: EBONE, GEOBON, GBIF, BioCASE, PESI, LTER-Europe, MARS, EBI-ELIXIR, ESA, EVOLTREE, Species2000, TWReferenceNET, Lagunet, ALTERnet, EMBRC, EUMON, DAISY, ALARM, BALLOON, ELNET, BIOFRESH, EUROCEANS.

Secondary data (e.g., model results, provenance data, etc.) can be generated as a result of processing and analysing primary data in the infrastructure. Principal data products include: Species names lists (and various concepts; life history and ecological attributes), specimen data, field observations (all these with associated spatial coordinates, date, time, size, environmental data as salinity, collection or observer name), monitor data (temp, height, precipitation, humidity, plant physiological responses, CO2, genetic data (sequences, genes), habitat and landscape data (data aggregates), earth observation data.

All processed and published data automatically have metadata assigned, which may include information about data collection, how it has been processed, etc. Currently a diverse set of ontologies describes various aspects of the data. Further work to link such ontologies (e.g., through a thesaurus and associated semantic network) is presently being considered.

LifeWatch does not store data per se. This is the responsibility of the data providers. LifeWatch provides mechanisms for accessing such data through linkages with external data resources. As well as big data collection initiatives, this can include data generated by individual scientists that want to share their data. Although, it may happen that LifeWatch is requested to provide storage services when no external service is available, processed and published data is stored by LifeWatch with third parties. All data assets acquire a persistent identifier on first entry to the infrastructure. This will be used for tracking purposes.

3.1.7.4 Services and tools

LifeWatch is a Service Network. It offers a range of services to its user community, from the foundational (such as authentication and authorization) that are needed for the overall operation of the infrastructure to those that are highly specialised for specific biodiversity research tasks. All services are Web services, conforming to relevant standards of W3C, OASIS, TDWG, ISO, etc. that can be composed by the user into workflows.

Most services available today, from various institutes, are not yet offered through a single portal together with data and sufficient computing power. LifeWatch (and its contributor projects such as the national LifeWatch initiatives, BioVeL and ViBRANT) aims to bring these together through the LifeWatch catalogue of services[76] and to make them available for composing in workflows[77] and virtual laboratories.

The LifeWatch portal will offer various pre-constructed virtual labs, and with services allow to build new dedicated virtual labs. LifeWatch defines an access policy (draft) that aims to offer open access to its resources and capabilities to all users, but which at the same time controls access to data and resources.

LifeWatch offers: Services catalogue; Datasets catalogue; Annotations repository; Applications servers; Virtual laboratories; Provenance and citation tracking repository; Security (AAA) services; access to computational resources; and portal framework.

3.1.7.5 Typical applications

LifeWatch is intended to be a generic ICT environment for biodiversity research, allowing users to build their preferred virtual laboratories (including their preferred data and software selections) and to easily create workflows. Typical applications offered by LifeWatch may include:

- Population biology of migrating birds
- Understanding the role of marine wetlands on biodiversity migration patterns
- Impacts of invading alien species
- Assessment of regional biodiversity
- Preservation of ecosystem services (curing habitat destruction)
- ICT support for human observations of biodiversity

3.1.7.6 Standards

LifeWatch relies on and conforms to published standards whenever feasible. Several standards and best practices related to standards provide a guideline for the LifeWatch ICT conceptualization and should be adopted whenever feasible and appropriate. The LifeWatch

Reference Model document provides more specific information. Relevant Standardization organizations are:

- Biodiversity Information Standards (TDWG)
- Open Geospatial Consortium (OGC)
- Organization for the Advancement of Structured Information Standards (OASIS)
- Open Grid Forum (OGF)
- World Wide Web Consortium (W3C).

Of particular interest are the following standards sets: ODP, SOAP, REST, INSPIRE metadata standards, OGC abstract and implementation specifications, DarwinCore, DarwinCoreArchive, Access to Biological Collection Data (ABCD), Structured Descriptive Data (SDD), Taxonomic Concept Transfer Schema (TCS), TAPIR, EML, etc.

3.1.8 The South Africa National Biodiversity Institute (SANBI) infrastructure

URL: http://www.sanbi.org/, http://www.sabif.ac.za/
URL: http://biodiversityadvisor.sanbi.org/

3.1.8.1 Overview and purpose

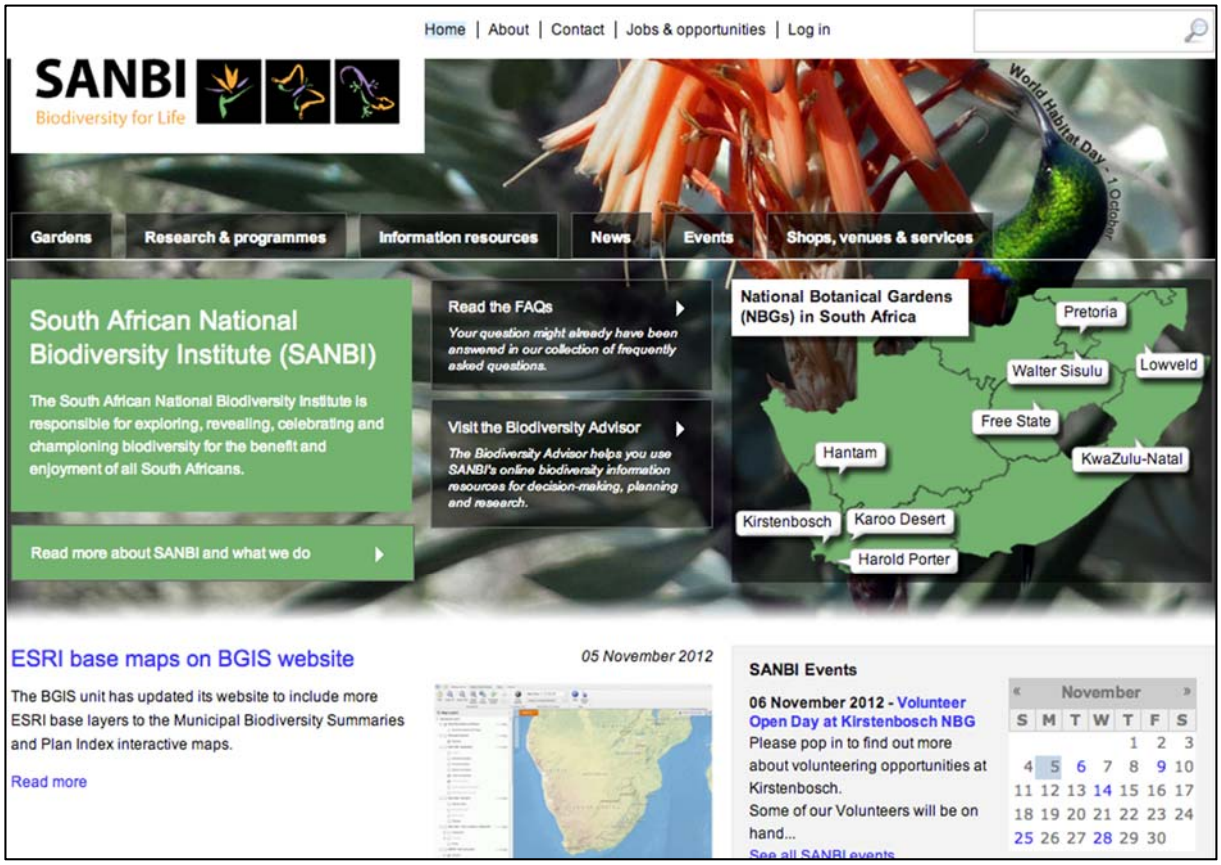


Figure 20. The SANBI web portal

The South African National Biodiversity Institute (SANBI) leads and coordinates research, and monitors and reports on the state of biodiversity in South Africa. The institute provides knowledge and information, gives planning and policy advice and pilots best-practice management models in partnership with stakeholders. SANBI engages in ecosystem restoration and rehabilitation, leads the human capital development strategy of the sector and manages the National Botanical Gardens as 'windows' to South Africa's biodiversity for enjoyment and education. SANBI maintains and offers multiple information systems to support a range of biodiversity-related activities, from environmental impact assessments to species identification and taxonomy. These include:

- Biodiversity Advisor system – a single point of entry, making it easy to navigate to the right information and tools among the range of resources provided by SANBI.
- SIBIS is the gateway to SANBI's biodiversity information, and information shared by SANBI partners. SIBIS provides access to over 1.6 million species occurrence records and comprehensive coverage of South Africa's 22 000+ plant species. Animal species

are being added as they become available. SIBIS focuses on species and specimen information and offers:

- Threatened species information,
- Distribution maps,
- Area checklists,
- General species details.

- The Biodiversity GIS system offers information for landscape and mapping information.

3.1.8.2 Architecture

SANBI is a managed network of 19 institutions, supervising 7 programmes, amongst which 1 programme addresses biodiversity information management, i.e. application policy, human capital (training), infrastructures, content, journal, monitoring etc. The following figure 21 gives an overview of the SANBI baseline architecture.

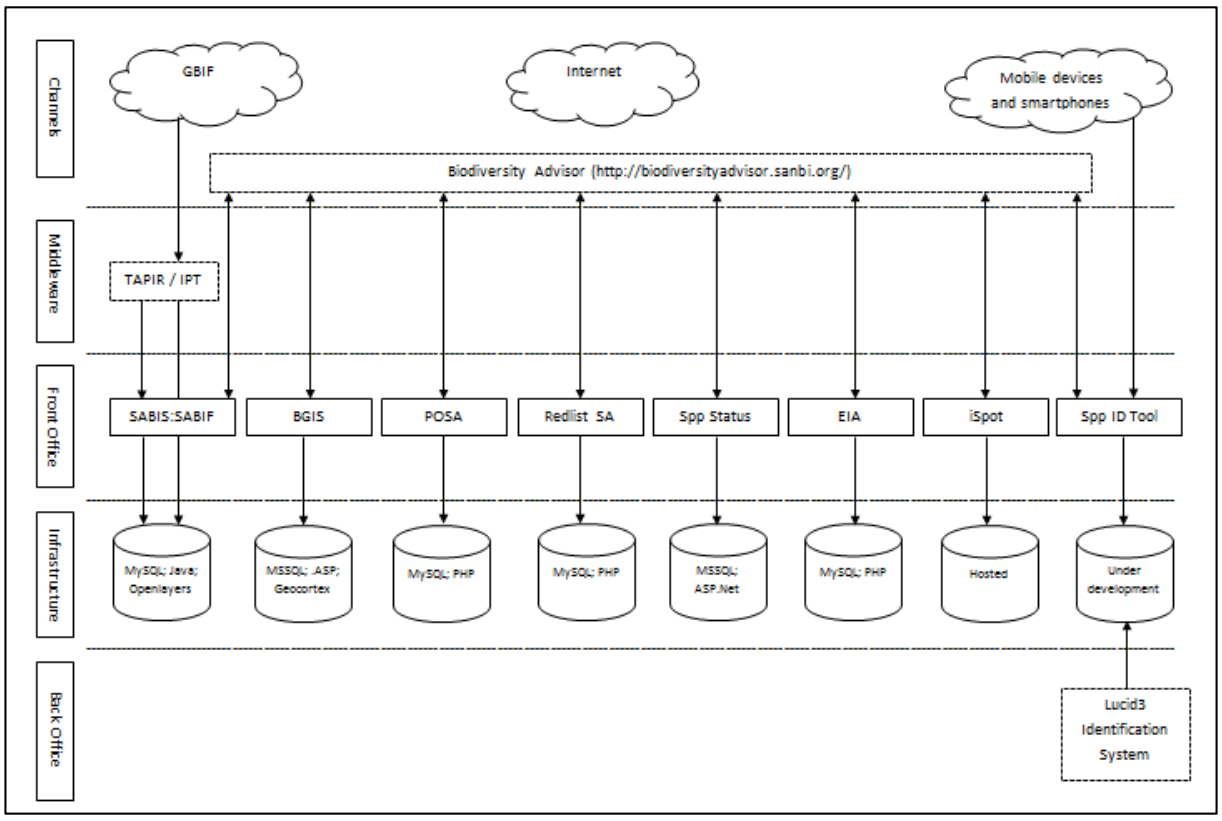


Figure 21. SANBI baseline architecture

As is illustrated in Figure 21, SANBI offers several tools to assist biodiversity stakeholders:

- SIBIS[78]: SABIF[79] a framework for disseminating species and specimen data.
  SIBIS aggregates information from a number of SANBI databases:
  - Acocks (plant species observations),
  - CREW (threatened plant species localities),
  - DNA laboratories (plant and reptile DNA accessions),
  - Garden Accessions (plant collection records),
  - Millennium Seed Bank(plant seed collection records),
  - PRECIS (taxonomy and herbarium specimens),
  - Protea Atlas (Proteaceae species occurrence records),
  - Species Status (NEMBA-listed species),
  - TSP (threatened plant species).
- BGIS[80]: a framework for disseminating landscape and map data.
- POSA[81]: an online checklist providing access to plant names and floristic details for southern African plant species.
- Redlist SA[82]: provides up to date information on the national conservation status of South Africa's indigenous plants.
- Species Status[83]: a database that stores all the TOPS- (a threatened or protected species, listed under the National Environmental Management Biodiversity Act (NEMBA) regulations on Threatened and Protected Species) and CITES- (a CITES listed species, listed on one of the Appendices of the Convention on International Trade in Endangered Species (CITES) to which South Africa is a signatory) listed species.
- EIA[84]: an Online Repository of EIA Data and Reports.
- iSpot[85]: citizen science platform, aimed at helping the public identify anything in nature.

Species Identification Tool: a tool aimed at assisting customs officials, law enforcement officers, border police and Environmental Management Inspectorate's with the identification of threatened species so placing them in a better position to regulate and monitor the trade in South African TOPS and CITES listed species and potentially other traded non-indigenous species without needing the assistance of provincial nature conservation staff.

3.1.8.3 Data

SANBI's major data types and sources of data:

- Taxonomic names/checklists,
- Occurrence records (presence only),
- Occurrence records (including absence records),
- Population density/dynamic data,
- Species information (descriptive data),
- Multimedia resources.

Data management resources:

- Metadata,
- Integrated Publishing Toolkit (IPT),
- TAPIR.

3.1.8.4 Services and tools

SIBIS offers capabilities to search for and retrieve species related information for South Africa, including distribution maps and to create custom checklists of species data for user specified geographical areas. The Biodiversity GIS system offers several tools to support land use and planning activities.

3.1.8.5 Typical applications

Typical application areas supported by SANBI information systems include:

- Environmental assessments,
- Land-use planning,
- Systematic biodiversity planning,
- Species distribution modeling,
- Research and taxonomy.

3.1.8.6 Standards

Information standards:

- DwC;
- TAPIR;
- National Spatial Information Framework (NSIF) – Metadata Standards;
- BGIS Data Submission Guidelines.

Publication standards:

- Publication Guidelines for Bothalia;
- ISI Rating criteria from Thomson and Reuters.

Cataloguing standards:

- Universal Decimal Classification;
- Dewey Decimal Classification;
- Anglo American Cataloguing Rules (AACR);
- Library of Congress Subject Headings).

3.2 Interoperability definition and analysis

3.2.1 Research infrastructure definition

As can be appreciated from research infrastructures' descriptions in former sections, 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[86]. 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.

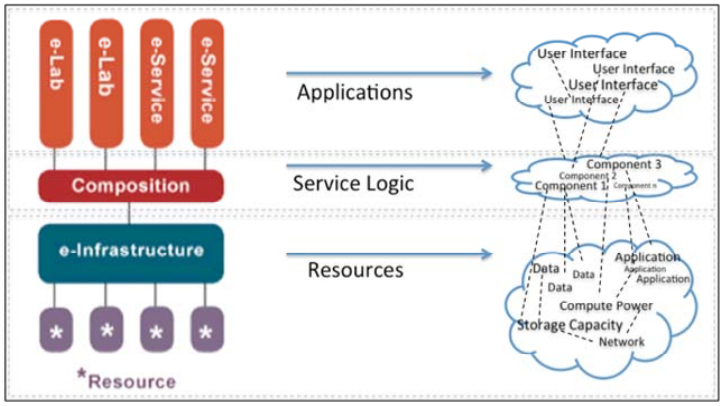


Figure 22. Research infrastructure first-class objects

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 rather 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. Such a convergence could open the door to unprecedented innovative new decision support and multimodal application scenarios to assist biodiversity stakeholders in their daily work.

With the aim of establishing a common understanding of interoperability between participating infrastructure stakeholders, CReATIVE-B thus started from a simple layered representation of what a Research Infrastructure (RI) is[87], and progressively rendered the definition more accurate to the point where it can be used as a standard descriptive pattern.

In Figure 22, a RI is thus 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 a 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.

In the remainder of this document, these first-class concepts, i.e. resources, service logic and applications, are used to support and structure a more elaborated definition of RIs, in particular to elicitate further elements on which to base a quantitative comparison of participating infrastructures useful to identify key differences and similarities.

Note: the following technical definitions are formalized in a standard representation using the Unified Modeling Language (UML[88]). UML is a general-purpose modeling language coming from the field of object-oriented software engineering. It includes a set of graphic notation techniques to create visual models of object-oriented software-intensive systems. UML is usually utilized to specify, visualize, modify, construct and document the artifacts of a system under development. However, beyond software engineering, UML is particularly useful to define conceptual objects and to explicit their interrelations. This is why it is used in subsequent definitions of this report.

3.2.2 Analysis scoping

According to the practical guidelines report "*Legal Framework for a ERIC*" of the European Commission[89], the term "research infrastructure" can be defined as follows:

> *"**research infrastructure** means facilities, resources and related services that are used by the scientific community to conduct top-level research in their respective fields and covers major scientific equipment or sets of instruments; knowledge-based resources such as collections, archives or structures for scientific information; enabling Information and Communications Technology-based infrastructures such as Grid, computing, software and communication, or any other entity of a unique nature essential to achieve excellence in research. Such infrastructures may be "single-sited" or "distributed" (an organised network of resources)."*

Transposed into UML notation, the model extract presented in Figure 23 recalls identified first-class objects (highlighted in orange) and expands on their constitution. In particular, RI applications (top-left) are described as sets of user interfaces exposing services logic(s). A good example can be that of a web portal integrating various (i.e. portlet[90]) applications.

Service logics wrap functional software components, which themselves expose computing resources of various types. Typically service logics are application servers and commercial-On-The-Shelf (COTS[91]) business logic. A good example can be that of JBOSS[92].

Exposed resources can for instance be computational power (i.e. CPU) made available to applications for processing purposes, data sources such as databases (i.e. MySQL registry of specimen metadata) or more basic and raw information (e.g. specimen images), or even software applications (e.g. Matlab[93] toolkit etc), as is described at the bottom with Computational, Data and Software Resources objects. Computing resources may also be assigned with an unique identifier (i.e. GUID[94]) and comply with a Service Level Agreement (SLA[95]), guaranteeing traceability and agreed quality of service in terms of response times, security, reliability etc.

Last but not least, applications, service logics and resources may integrate a secure context requiring authentication and authorization, as is modeled bottom-left, sometimes depending on governance, SLA or even underpinning technologies.

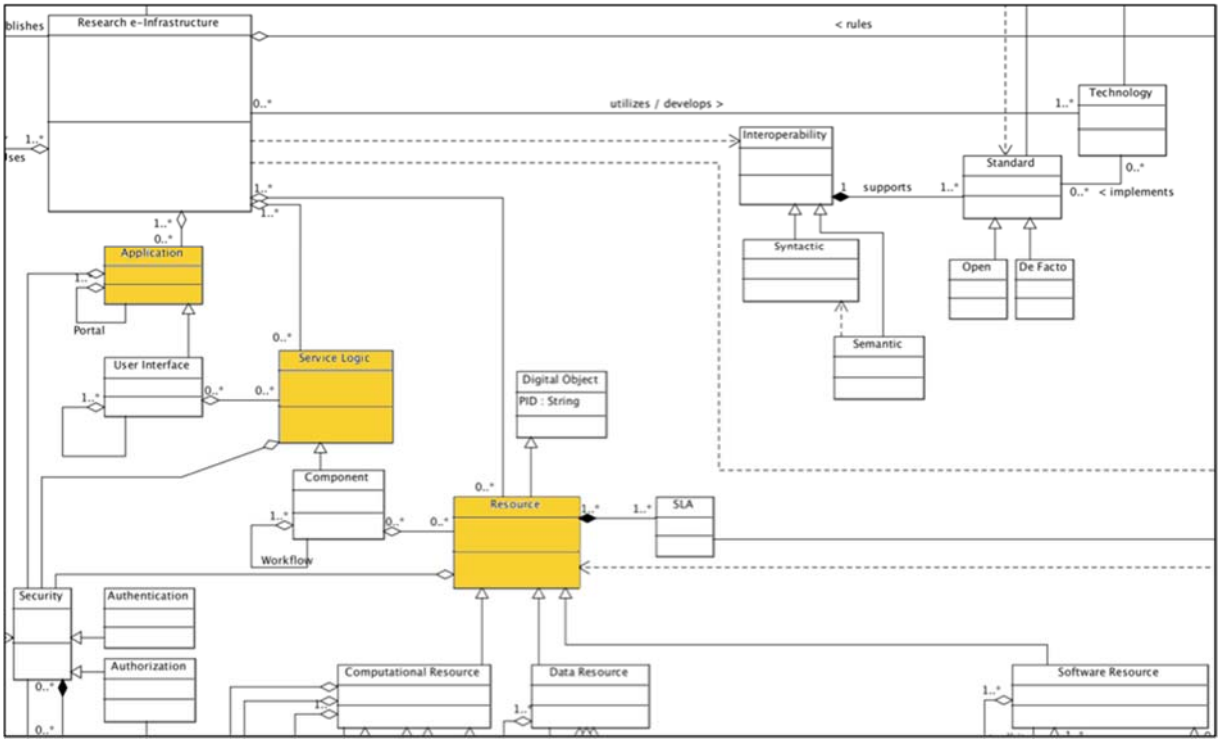


Figure 23. Research infrastructure model extract

Figure 24 gives a more detailed description of resource objects, which may be found and manipulated in biodiversity RIs, although not being exhaustive. Computational resources thus not only encompass processing power of various kinds (such as Clusters, Grids, HPC or Clouds of CPUs and/or GPUs), but also storage capacities (e.g. iSCSI, SATA disk pools etc), see bottom-left.

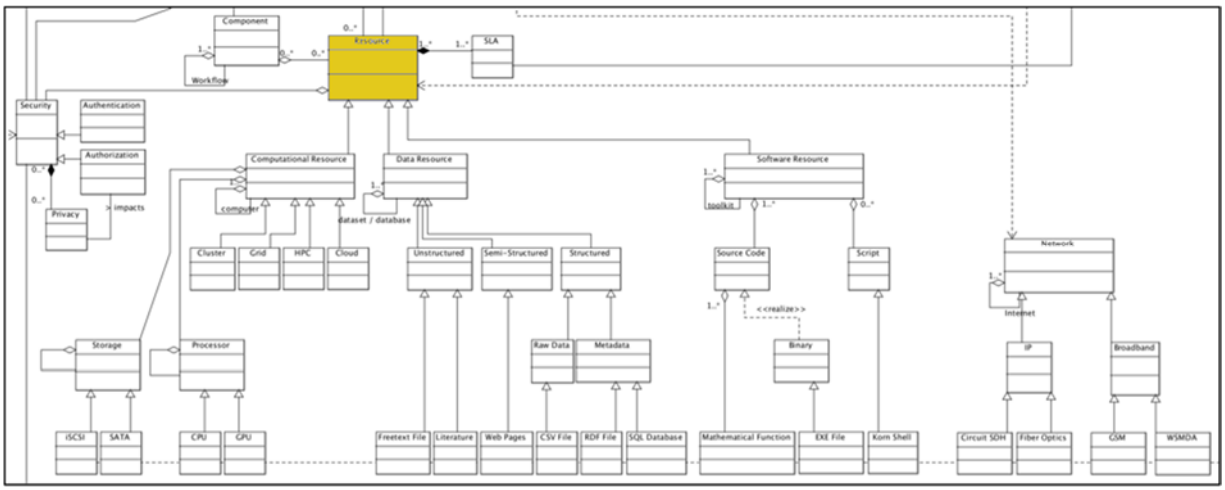


Figure 24. RI resources model extract

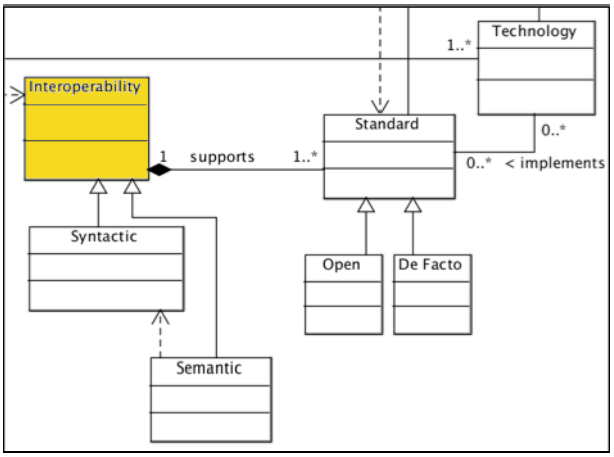


Figure 25. Interoperability model extract

Data resources may be found unstructured, semi and fully structured, from freetext files, literature, web pages, to RDF files or even more complex metadata databases.

Software resources on the other hand, cover scripts, application source codes and binaries being used individually or composed in more elaborated toolkits and workflows, see bottom-right.

Finally, the top-right area of the diagram exposed in Figure 23 recalled here in Figure 25, defines interoperability in relation with research infrastructures.

In our model, interoperability is expressed as a dependency of RIs, which implies the latter are constructed based on standards and thus utilizing/implementing technologies conforming with these. To be more accurate, CReATIVE-B elaborates on the simple definition of interoperability that can be found on Wikipedia[96] (see extract in the box below).

> ***Interoperability*** *is a property referring to the ability of diverse systems and organizations to work together (inter-operate). The term is often used in a technical systems engineering sense, or alternatively in a broad sense, taking into account social, political, and organizational factors that impact system to system performance...*
>
> *... **Syntactic** interoperability*
>
> *If two or more systems are capable of communicating and exchanging data, they are exhibiting syntactic interoperability... Syntactical interoperability is required for any attempts of further interoperability (ie)...*
>
> *... **Semantic** interoperability*
>
> *Beyond the ability of two or more computer systems to exchange information, semantic interoperability is the ability to automatically interpret the information exchanged meaningfully and accurately in order to produce useful results as defined by the end users of both systems...*

This definition distinguishes two major aspects of systems interoperability:

(i) Syntactic interoperability, on the one hand, which is required for infrastructures to communicate with each other. For communicating data, specified data formats, communication protocols, interfaces descriptions and the like are fundamental. XML and SQL standards are good examples of syntactical interoperability. Syntactic interoperability however does not enable systems to understand each others de facto.

(ii) Semantic Interoperability, on the other hand, refers to the understanding and treatment of exchanged data. Two systems that are able to process exchanged data and to produce meaningful results out of them are considered to be semantically interoperable.

The full realization of interoperability therefore lies in the syntactic interconnection of and semantic understanding between systems. While, at a first glance, the former appears to be the simplest form of interoperability to implement, the latter requires more in depth analyses of resource structures, service logics and last but not least applications knowledge.

Figure 26, on the next page, gives the complete UML representation of a RI, integrating the model extracts exposed in figures 23, 24 and 25. Based on this detailed definition, the next section reports on technical information gathered thus far on each and every important elements of the model. More specifically, three tables corresponding to the first-call objects report on the following aspects for all 9 participating RIs:

General Overview	Research Infrastructures
Geographical coverage	Implicit, to determine geographical complementarities
Topical coverage	Implicit, to determine biodiversity complementarities
Infrastructure topology	To understand how the RIs are deployed and what their key pillars are
Native interoperability & enablers	To gather more information on potentially enabling collaborations or technical convergences
Merging of science & policy needs	To understand how policy needs may be supported in RIs functions, which is key to sustainability
Merging of science & industry needs	To understand how industry needs may be supported in RIs functions, which can further support sustainability
Engagement of citizens	Implicit
Access policy	Implicit, to determine legal compatibilities
Licensing & business model	Implicit, to determine legal compatibilities
Funding	Implicit, to understand short-term sustainability and potential funding synergies
User applications & functions	Implicit, to determine functional complementarities

Service Logic	Research Infrastructures
Software architecture	Implicit, to determine conceptual complementarities
Programming languages	Implicit, to determine software engineering compatibilities
Authentication	Implicit, to determine security compatibilities
Authorization	Implicit, to determine security compatibilities
Middleware	To understand the complexity and significance of developed (proprietary) services
Technology	Implicit, to determine compatibilities
Standards	Implicit, to determine compatibilities
Computing Infrastructure	Implicit, to determine compatibilities

Data	Research Infrastructures
Data sharing & quality control	Implicit, to determine data quality and complementarities
Data information dealt with	Implicit, to determine biodiversity data complementarities
Data source tracking	To understand how the RIs are deployed and what are their key pillars
Data citation tracking	To gather mode information on potentially enabling collaborations or technical convergences
Data integration	To understand how policies may be integrated in RIs functions, which is key to sustainability
Geographic information system	To understand how industries may be integrated in RIs functions, which can support sustainability
Standards	Implicit
Technologies	Implicit

Note 1: Some criteria were not investigated due to the lack of time. The latter may be addressed in an updated version of present contents, to be formalized in deliverable D3.2. The concerned criteria are the following:

General Overview	Research Infrastructures
Sensing network approach	Implicit, to determine data collection networks complementarities
Sensing network standards / technologies	Implicit, to determine data collection networks complementarities
Network infrastructure	Implicit, to determine physical ICT networks complementarities

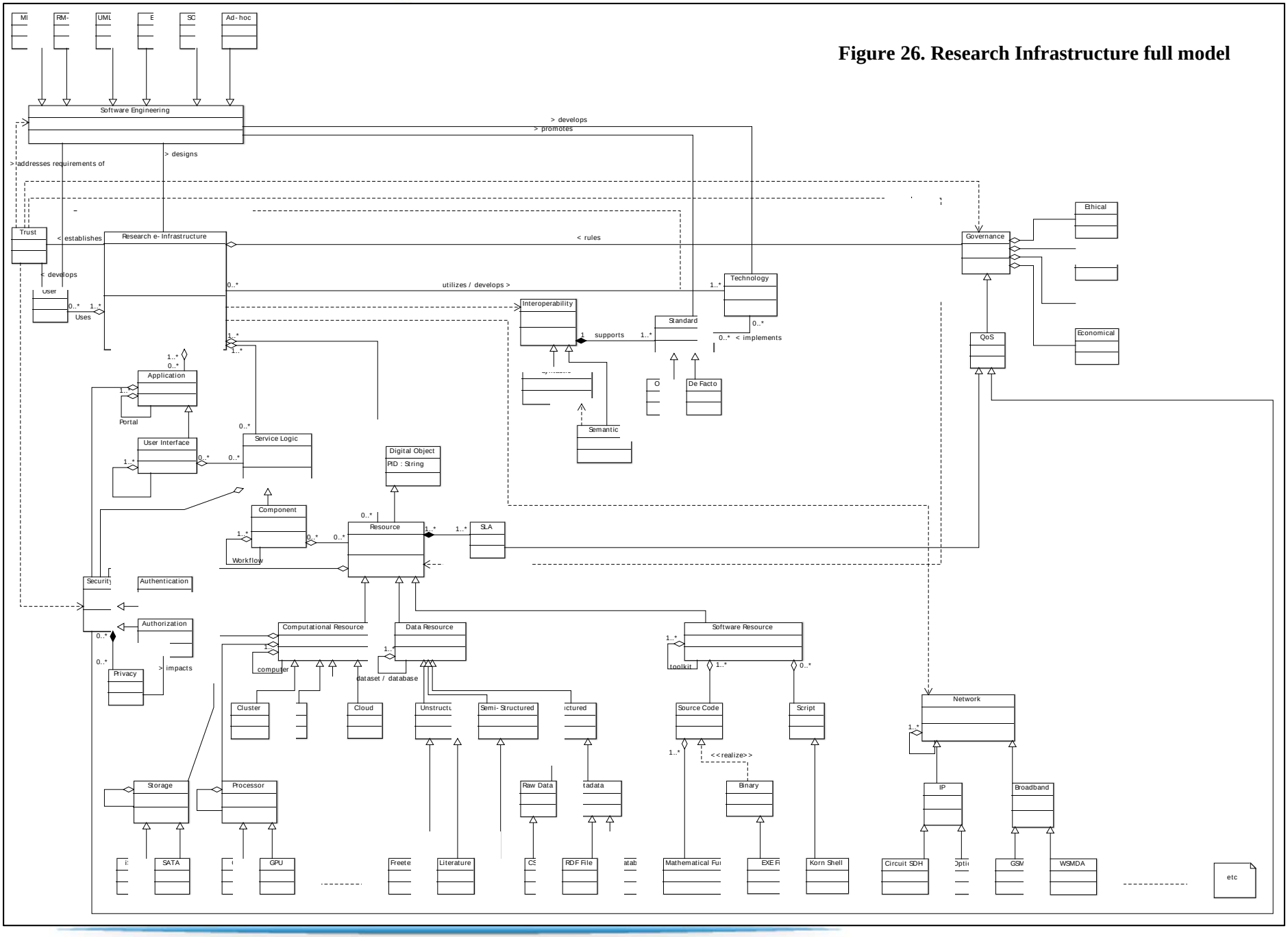
Service Logic	Research Infrastructures
Service infrastructure governance	To understand in place (or future) governing bodies, mechanisms and their translation into ICT non-functional aspects
Scientific workflow management	To understand how RIs allow scientists to combine sets of resources into executable, traceable and archived experiments
Scientific workflow standards / technologies	To understand how RIs allow scientists to combine sets of resources into executable, traceable and archived experiments
Scientific provenance	To understand how experiments' metadata and outputs are stored and used
Scientific provenance standards / technologies	To understand how experiments' metadata and outputs are stored and used

Data	Research Infrastructures
Data privacy approach	Implicit, to determine legal access rights complementarities
Data privacy standards / technologies	Implicit, to determine legal compatibilities

Note 2: Some objects of the model exposed in Figure 26 are not detailed for the sake of clarity, in particular the concepts of Trust and Governance. These notions will be further addressed in deliverables D3.2 and D3.3 as progress will be made with interoperability scenarios, guidelines and final recommendations.

SEVENTH FRAMEWORK PROGRAMME

Figure 26. Research Infrastructure full model



3.2.3 Interoperability analysis

Table 4. Research infrastructures' general overview

General Overview	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI
Geographical coverage	Australia	Brazil	Brazil	China	USA (Global in long term)	Global	Global	Europe (Global in long term)	South Africa (Africa in long term)
Topical coverage	An online directory of all living things in Australia and related environmental data	The entire 'library of life' in Brazil	Species & specimen data (plants, animals, microorganisms)	Botany and zoology of China	Life on earth and the environment sustaining it	Making the world's biodiversity information available	Observations on biodiversity change information / monitoring the biosphere	Broadly supporting research on biodiversity and ecosystems	Knowledge, information and policy advice on biodiversity of South Africa

	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI
Infrastructure topology	Centralized Hub and spoke. Single point of access	Distributed Managed network of data providers with cache nodes and centralized data warehouse. 260 data providers, 24 out of 27 Brazilian states covered	Distributed Managed network of data providers. Data is harvested to a central database. 290 data providers from 26 of 27 Brazilian states; 5.6 million data records (3.1 geo-referenced, 190 thousand images	Centralized Single point of access, centralized database with 300 distributed seed collectors	Distributed Tiered, with a set of coordinating nodes (3) and many member nodes (8 now and a further 12 next year)	Distributed Tiered system of systems. Central registry for schemas, vocabularies etc. 350 nodes	Distributed Tiered system of systems. 8 topical working groups, thematic BONs, national and regional BONs	Distributed Tiered system of systems, with common facilities. Reference Model (LifeWatch-RM[97] based on RM-ODP[98])	Distributed Managed network of separate systems
Native interoperability[99] & enablers	Acts as Australian GBIF node Interaction at regional level with IOCI[100] initiative Integrates the Online Zoological	Collaborates with GBIF on *LifeMapper*[101] development Collaborates with regional institutions on *openModeller*[102]	Web services available: images (Exsiccata[110]), maps, data records (TAPIR[111] provider) tools: openModeller (source forge)	Collaborates with UK's Royal Botanic Gardens KEW[112] Collaborates with the World Agroforestry Centre[113]	Integrates access to XSEDE[114] Allows authentication from ALA IdP[115] + 150 IdPs	Encyclopedy of Life (EOL[116]), *LifeMapper*, Map of Life[117], IABIN[118], Wallace initiative[119], VertNet, IUCN Red List[120], IUCN	OGC[122] leads core task to develop GEOSS Architecture GEO BON promotes interoperability across biodiversity initiatives, and	Acts as European GEOSS / GEO BON Implements RM-ODP recommendations Aims for mediation at	Use GBIF Integrated Publishing Toolkit (IPT[129]), looking to use more of GBIF Provides infrastructure for partners to share

	Collections of Austalian Museums (OZCAM), AVH (Australia's Virtual Herbarium), AMRiN, Seedbanks	development Collaboration with Species List of the Brazilian Flora[103]; INCT Herbario Virtual de Flora e dos Fungos[104]; TAPIR provider Connected to RedCLARA [105] (part of GEANT[106]), eduGAIN[107] compatible security framework CIPO[108] experimental service Member of the GLIF[109] initiative				species ranges GBIF partnerships [121]	works with the GEOSS Standards and Interoperability Forum (SIF[123])	the semantic level. Recommends harmonization at the syntactic level but does not mandate it Terrestrial LTER[124], marine reference and focal sites, natural science collections Collaborates with Flanders observatories: EuroBIS[125], FLORAweb[126], FlaWet[127], WATERvogels Integrates EUDAT[128]	information Long-term aim is to integrate all African countries
	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI

Merging of science & policy needs	Yes	Yes Towards conservation and sustainable use of Brazil resources	Yes As end-users; ecological niche modeling tools & data gaps analysis	Yes Towards conservation and sustainable use of China resources	Yes	Yes	Yes	Yes	Yes
Merging of science & industry needs	Yes	Yes Industry users involved	Yes As end-users and catalogue of microbial strains	Yes Floriculture and biomedicines	No	No But planned with MTRG[130]	Yes	No But planned	Yes
Engagement of citizens	Yes As data collectors E.g. *BDRS*[131] mobile app, web portal	No Except as end-users	No Except as end-users	No	Yes PPSR[132] projects and working group	Yes Through regional / national work of GBIF participants, data collectors publish content	Yes. Crowd sourcing E.g. *GeoWiki*[133] portal from EuroGEOSS[134]	Yes As data collectors (human sensors and crowd sourcing) and users	Yes As data collectors E.g. *iSpot*[135] citizen science application
	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI

Access policy	Open and free access	Open and free access	Open and free access	Open and free access	Open and free access	Open and free access	Open and free access	Open and free access	Open and free access
Licensing & business model	Open source support Creative Commons, Google code environment[136]	Open source support	Open source support Creative commons	*NA* *Not available at the time of writing*	Open source support Business model under discussion, sponsorship, taxation etc	Open source support Google code environment [137]	Open source support Voluntary contributions, pilot projects	Open source support But not mandatory	Open source support Based on GBIF
Funding	Australian government funding (National Collaborative Research Infrastructure, Education)	Brazil government funding Next round of funding under discussion	Project based	China central government	NSF funding for now	IGO MoU agreements with GBIF members	UN (GEO) + EC (EuroGEOSS, EU BON, LifeWatch)	EC + member states funding for now (30% reached)	Public entity overseen by Ministry of Water and Environment
	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI

User applications & functions	*TRIN* wiki[138] annotation services, taxonomic name service, data validation, sensitive data service, taxonomic support tools *BRDS* Web portal for Citizen Science Mobile app (IOS/Android) *Australian Biodiversity Heritage Library* (BHL[139]) *Australian Morphbank*[140] database of images *Sandbox*[141] for data uploads to ALA	*LifeMapper* online geospatial species occurrence data to create distribution maps and, predict where an individual species should exist *SpeciesLink* data search, filter and retrieval, maps, images, indicators, integrated information (endemic, endangered, etc), data information gaps for plants *openModeller* fundamental	data search, filter and retrieval, maps, charts, images, indicators, integrated information (endemic, endangered, etc), data information gaps for plants, data cleaning reports	*Genobank*[145] online database, management and sharing of germplasm information, as 37,124 relevant data and 43,821 images are now open to public users	*Investigator Toolkit (ITK*[146]*)* to interact with DataONE services *ONEMercury*[147] google like discovery interface, for searching resources *Data Management Plan* (DMP[148]) tool for managing data collection lifecycle *ONE-R*[149] *client* for R users to access DataONE data;	*GBRDS*[155] global broker, explore species, explore countries, explore datasets, analytics and visualization *IPT* publish and share biodiversity datasets through GBIF network	*GEO Web Portal*[156]; GEOSS Clearinghouses [157] ; GEOSS Components and Services Registry; browse and search available datasets, application, services GEOSS Standards and Interoperability Registry [158]; provides integrated and cross-cutting information on standards and interoperability	User facilities and functions under development, some of which being implemented by BioVeL[159], ENVRI[160], ViBRANT[161] and others Expected data integration, modeling and publishing, as well as dedicated virtual laboratories with support to developers	*BGIS*[162] online GIS mapping, Land-use decision support, Species distribution modeling, municipal profile analysis, redlist, iSpot Biodiversity Advisor[163], portal
	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI

	Australian Barcode of Life Network (ABOLN[142]) *Species Interactions of Australia Database* (SIAD[143])	niche modeling experiment *Virtual Herbarium* [144] categorization of species and identification for producing decision models			*ONEDrive* file space mount for users; Workflow management, statistics and visualization interfaces: *Vistrails*[150], *Kepler*[151], *Taverna*[152], *myExperiment*[153], *Matlab*[154]				
	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI

Table 5. Research infrastructures' service logic

Service Logic	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI
Software architecture	SOA	SOA	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	SOA	SOA	SOA	SOA	SOA
Programming languages	Mainly Java, Groovy/Grails, Scala	Mainly PHP	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	Mainly JAVA and Python	Mainly Java	Not relevant	Not relevant	Mainly JAVA
Authentication	Single Sign On	Single Sign on eduGAIN based	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	Single Sign On 150 IdPs + ALA IdP	Open currently, Adopting optional single sign on	GEOSS Registry sign on	UAA[164] (includes Single Sign On)	*NA* *Not available at the time of writing*
Authorization	Credential delegation, CAS[165]	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	Credential delegation, Shibboleth [166]	To become CAS	*NA* *Not available at the time of writing*	UAA	*NA* *Not available at the time of writing*
Middleware	Proprietary layer of REST	Proprietary sets of	*NA*	*NA*	Proprietary sets of	Proprietary suite of	446 service interfaces of	Proprietary middleware	GBIF based

	JSON web services Services mediate multiple independent data sources to multiple application front-ends	libraries	*Not available at the time of writing*	*Not available at the time of writing*	libraries for each sub-system (i.e. CN, MN, ITK) (proprietary libs not required for web services communications and authentication)	tools: *IPT*, *GBRDS*, Checklist Bank, etc. Central facilities: Taxon data service, occurrence records, occurrence density, dataset metadata, data publisher metadata, data network metadata Nodes: DiGIR, TAPIR, BioCASE, IPT, Darwin Core Archive By 2012-13 Service layer of RESTful JSON services	participating systems listed in GEOSS Components and Services Registry GEONETCast for disseminating spatial data	presently under development Expected species occurrence, name validation, mapping, gazetteer and catalogue services	
	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI

	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI
Technology	Based on XML/JSON Cassandra/SOLR/MySQL/PostGIS OGC[167] (Web Map Service (WMS), Web Feature Standard (WFS)) JSON	PHP + PostgreSQL [168] OGC (Web Map Service (WMS), Web Feature Standard (WFS)) JSON	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	XML message encoding, Apache, Tomcat[169], SOLR, Hazelcast[170] (inter-CN messaging)	Based on XML/JSON (for central repository)	Based on XML/JSON OGC	Various, according to need OGC (many)	Based on XML/JSON OGC
Standards	Web services RESTful, XML	Web services RESTful, XML	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	Web services RESTful XML	Web services RESTful XML	Web services RESTful XML GEOSS SIF arrangements	Web services RESTful XML W3C[171], OASIS[172]	Web services RESTful XML
Computing infrastructure	Offline	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	XSEDE	Offline, (HADOOP[173], HBase)	Grid, HPC	Grid (EGI[174]), HPC[175], Cloud	HPC (in house)

Table 6. Research infrastructures' data

Data	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI
Data sharing[176] & data quality control	Yes Uses a data quality model and process for assessing shared data	Yes Quality control and data cleaning reports produced and shared online	*NA* *Not available at the time of writing*	Yes Quality control process formalized for data collectors	Yes Provides reference documentation on improving data quality, driven by member nodes policies and practices	Yes Covers all levels of data management lifecycle Applies processing techniques to registered data to screen out obvious inaccuracies; provides training and reference materials on improving data quality	No But, brokering model of mediation and mapping. No direct data sharing nor quality control	No But, data provider admission scheme and certified data provider scheme (Service Level Agreements) allow access to data for those that need it	Yes Provides standards and implemented quality control processes, also available to all partners

Data / information dealt with	Species pages, species lists, occurrence records; Environmental layers; Citizens' data submission	Species checklists, red lists, occurrence data, citizens' observations ; Invasive species management ; taxonomy ; publications, maps, images and sounds	*NA* *Not available at the time of writing*	Species pages, species lists, occurrence records, images; environmental layers ;	Ecological, Biogeochemical, biodiversity, biosciences, environmental data	Primary biodiversity occurrence network[177] (museum collection records, observations, etc). Catalogue of names with associated data, species oriented, supplementary data such as IUCN RL, dataset descriptive such as ILTER[178]	All types of biodiversity information GEOSS Data Core of coordinated priority data sets	All kinds of biodiversity and ecological data, and supporting environmental sciences, socio-economic data, etc	Species lists, red lists, specimen records, pictures, guides, plans, map layers
	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI

	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI
Data source tracking	Yes Source of data is captured	Yes Source of data is captured, attribution given	Yes Source of data is captured, attribution given	*NA* *Not available at the time of writing*	Yes Source of data is captured. D-OPM (being replaced by D-PROV) Use PIDs (DOI[179] provided by DataCite[180])	Yes Source of data is captured	No	No But planned	Yes Use PIDs
Data citation tracking	Yes Tracks numbers of records downloaded, purpose and download frequency	Yes Track downloaded records per source	Yes Track downloaded records per source	*NA* *Not available at the time of writing*	Yes Scientists can keep track of who has re-used their data Email associated to user profile	Yes Tracks, at a coarse level, the number of times GBIF data has been cited in academic publications Monitoring of pubs referencing use of GBIF enabled data	No	No But planned	Yes Track all downloads and on-line assessment

	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI
Data integration	Aggregation of data from a wide range of providers. Catalogue of datasets TDWG Darwin Core + 400 ad-hoc columns from environmental/ contextual data layers	Aggregation of data from multiple sources	Yes Red lists, CoL	Aggregation of data from multiple sources	Federation of metadata in many places; data deposition; replication; identifier resolution, search and discovery, and federated identity	Aggregation of data from multiple sources Digitization , publishing, discovery, indexing Crawling and indexing (over 25 Darwin core fields) of networked data Being extended with monitoring, annotation s and notification	Federation of metadata, no semantics Mediation of services using a broker approach. Catalogue of components	Aggregation of data from multiple sources Publishing, discovery, retrieval, annotation, modeling, workflows execution. Catalogue of services, catalogue of datasets,	Aggregation of data from multiple sources Search and retrieve; distribution maps and checklists
Geographic Information System (GIS)	postGIS[181], GeoServer[182], GeoNetwork[183]	mapCRIA[184], LifeMapper, postGIS	postGIS	*NA* *Not available at the time of*	May be provided by third party services (e.g	PostGIS, Geoserver, Custom tile servers	Spatial data infrastructure linked to the GEO Portal	Spatial Data Infrastructure (SDI) supporting various GIS	Biodiversity GIS (BGIS), based on SAGDAD[185]

				writing	Member Nodes)			applications	taxonomy
Standards	DwC[186] (+ 400 ad-hoc columns), DwC-A[187], EML[188], TAPIR, DIGiR[189], BioCASE[190], OAI-PMH[191], OGC, RIF-CIS, EOL schemas, KML[192]	DwC, TAPIR, DIGiR	DwC, TAPIR, DIGiR	*NA* *Not available at the time of writing*	OAI-ORE, EML, RDFa, ISO 19115, DwC, DwC-A, COinS[193]	DwC (extended), DwC-A, EML, TAPIR, DIGiR, BioCASE, ISO19119, ABCD, CSV, Avro, RDF	ISO, OGC, and others listed in GEOSS Standards and Interoperability Registry	ISO, TDWG, INSPIRE	TDWG, DwC
Technology	*CASSANDRA*[194] data management system (column-based indexing) + *SOLR* Indexing of one very large table (35 M records)	*Species LINK* regional replication server	*NA* *Not available at the time of writing*	*Genobank* database 190.000 plants, 190.000 species	*NA* *Not available at the time of writing*	*HADOOP* + *HBASE*[195] *SOLR*[196] *PostGIS, MySQL* Central registry for datasets, news and endpoints with schemas and vocabularies 388M records mobilized	*NA* *Not available at the time of writing*	*NA* *Not available at the time of writing*	*BRAHMS* for plants taxonomy SIBIS for specimen data
	ALA	SIBBr	CRIA SpeciesLINK	GBoWS (CAS)	DataONE	GBIF	GEOSS (GEO BON)	LifeWatch	SANBI

3.2.4 Interoperability differences and similarities

Table 7. Similarities and differences between biodiversity RIs

Table 7, on the left, graphically highlights identified similarities / differences between the 9 participating RIs, along the 3 first-class concepts and associated criteria, as introduced in former sections.

The approach taken to obtain this visual output lies in assigning values from 0 (in white), to 0.5 (in yellow) or 1 (in green) to each and every aspect, where the colour coding corresponds to the following assertions:

- Green = a good level of interoperability seems to be achievable,
- Yellow = significant differences are felt that make interoperability more difficult to achieve,
- White = out of scope or does not meet the criterion, thus making interoperability nonsensical for this category,
- Grey = insufficient or no information is available.

Thus, the greener the table or row is, the more likely it is to achieve interoperability. On the contrary, the yellower or the whiter, the more differences there are. For all 3 inner tables a radar map is then proposed on the right, which recalls all aspects of the category and processes an “average interoperability” percentage based on the formerly assigned values. The radar maps also 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.

4 CONCLUSIONS & OUTLOOK

4.1 Preliminary analysis outcome

The first conclusion which can be drawn from this preliminary analysis is that generally speaking, all 9 RIs seem to exhibit a good level of potential interoperability in particular in the way they offer access to biodiversity data, applications and related resources. They seem to pursue similar objectives in terms of business models, industry and policy involvement, and overall sustainability plans, thus potentially facilitating the setup of a future international IVE and accompanying governance. Participating RIs seem to have complementary geographical and topical coverage while differing on actual implementations, starting from the physical topologies of their networks.

The latter becomes more obvious with the service logic aspects, where despite similar software architectures and standards being adopted, proprietary middleware services were developed with different security infrastructures, programming languages and technologies. The service logic is actually the place where the most differences can be found and thus where one can expect the most work will be needed to make systems syntactically and semantically compatible.

On the data side, the observation is more optimistic. A set of specific standards seems to have emerged in the biodiversity community, which address the needs faced with data integration and organisation. Similar sharing and quality control processes are thus in place for initiatives dealing with data collection, and traceability seems to be a shared concern for scientific citations and raw data tracking.

4.2 Additional considerations

From the various fora and exchange opportunities created by CReATIVE-B, additional feedback was gathered from RI stakeholders and biodiversity researchers on actual needs felt unaddressed in the community. Thus, three particular aspects revealed, as are detailed in the following of this section.

A) Data integration

There is a clear need to promote the use of consistent vocabularies and eventually to develop a biodiversity informatics ontology to facilitate the integration of data and to enable their better understanding thanks to concepts alignment and agreed (meta)structures. Additionally, data and metadata need to be qualified in terms of quality, i.e. whether they were quality controlled and following what protocol. Finally, the granularity between data and metadata also requires a subtle and well-balanced thinking, to be turned into meaningful information for users.

SEVENTH FRAMEWORK PROGRAMME

B) Infrastructures development

RIs have to be developed in a way that ensures their financial and technical sustainability. Indeed, RIs have to deliver robust services operating beyond initial project/funding lifetimes. There is thus a need for continuous support and increased technical capacity over time. In this sense, software sustainability (and in particular open source) shall be considered, taking into consideration important aspects such as trustability of software repositories and their potential certification.

C) Benefits of interoperability to users

The concrete benefits of interoperability to users should be emphasized and promoted. More specifically, it will be important for initiatives such as CReATIVE-B to formalize and simply explain the major obstacles/enablers to interoperability, what can be achieved time-wise and to define new use-case scenarios illustrating these aspects.

Supporting these conclusions and extending them with more specific considerations on actual data integration requirements, the following extract can be noted from the preliminary report of the 2012 GBIC Conference[197]: "*These discussions clarified the complexity and interconnections between these components and reinforced the recognition that many other activities which are already under way are also essential to the successful delivery of integrated infrastructure for managing biodiversity data in support of science and policy".*

4.3 Future works

This preliminary analysis will be presented at the next CReATIVE-B workshop, to be held in Kunming, China, next April 2013. Authors will use this report as a solid knowledge base to investigate further aspects of RIs' interoperability (as partly anticipated in section 3.2.1), in particular to identify interesting new use-case scenarios involving the participating stakeholders. Such use-cases will help formulating a prioritized roadmap of technical actions to be carried out towards the definition of interoperability guidelines in deliverable D3.2.

SEVENTH FRAMEWORK PROGRAMME

5 APPENDIXES

5.1 List of figures

5.2 List of tables

5.3 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 e-Biosphere conference http://www.e-biosphere09.org/
[8] The EC Horizon 2020 http://ec.europa.eu/europe2020/index_en.htm
[9] The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services http://www.ipbes.net/
[10] The GEO Biodiversity Observation Network (BON) http://www.earthobservations.org/geobon.shtml
[11] The Atlas of Living Australia (ALA) http://www.ala.org.au/
[12] The Centro de Referência em Informação Ambiental (CRIA) http://www.cria.org.br/
[13] The CRIA SiBBr Project www.sibbr.gov.br
[14] The CRIA speciesLINK Tool http://www.splink.cria.org.br/
[15] The China Germplasm Bank of Wild Species (GBoWS) http://www.genobank.org/
[16] The DataONE initiative http://www.dataone.org/
[17] The Global Biodiversity Information Facility (GBIF) http://www.gbif.org/
[18] The South African National Biodiversity Institute (SANBI) http://www.sanbi.org/
[19] The outGRID Project http://www.outgrid.eu/
[20] The "neuGRID for you" Project https://neugrid4you.eu/
[21] The Laboratory of NeuroImaging (LONI) at UCLA http://www.loni.ucla.edu/
[22] The CBRAIN Project https://cbrain.mcgill.ca/
[23] The CReATIVE-B Project Website and Background Materials http://creative-b.eu/background
[24] The LifeWatch Reference Model http://subs.emis.de/LNI/Proceedings/Proceedings154/gi-proc-154-15.pdf
[25] The GEO BON Information Architecture http://www.earthobservations.org/documents/cop/bi_geobon/geobon_information_architecture_principles.pdf
[26] The Convention on Biological Diversity (CBD) http://www.cbd.int/
[27] The CBD AICHI Biodiversity Targets http://www.cbd.int/sp/targets/
[28] The Global Biodiversity Informatics Outlook http://imsgbif.gbif.org/CMS_ORC/?doc_id=4937&download=1
[29] The outGRID Interoperability Cookbook http://www.outgrid.eu/public/outgrid/download/deliverables/D2.4.pdf
[30] The outGRID Interoperability Analysis http://www.outgrid.eu/public/outgrid/download/deliverables/D2.3.pdf
[31] The 4+1 Architectural View Model http://en.wikipedia.org/wiki/4%2B1_architectural_view_model
[32] The Service Oriented Modeling Framework (SOMF) http://www.modelingconcepts.com/pdf/SOMF_2.1_Conceptualization_Model_Language_Specifications.pdf
[33] Virtual imaging laboratories for marker discovery in neurodegenerative diseases. G. B. Frisoni, A. Redolfi, D. Manset, M-É. Rousseau, A. Toga & A. Evans. Nature Reviews: Neurology August 2011 5; 7(8) pp 429-38. doi:10.1038/nrneurol.2011.99
[34] The Taxonomy Research and Information Network (TRIN) http://wiki.trin.org.au/
[35] The Online Zoological Collections of Australian Museums (OZCAM) http://ozcam.ala.org.au

[36] The ALA Spatial Portal http://spatial.ala.org.au
[37] The ALA Spatial Portal Layers http://spatial.ala.org.au/layers
[38] The ALA Web Services http://spatial.ala.org.au/ws, http://biocache.ala.org.au/ws, http://http://spatial.ala.org.au/ws/examples/
[39] The ALA GeoNetwork http://spatial.ala.org.au/geonetwork
[40] The ALA Email Alert Service http://alerts.ala.org.au
[41] The ALA Sandbox Service http://sandbox.ala.org.au
[42] The ALA Single Sign On Service http://auth.ala.org.au
[43] The GeoServer Software http://geoserver.org/
[44] The ALA Dashboard http://dashboard.ala.org.au/
[45] The Australian Biodiversity Heritage Library (BHL) http://bhl.ala.org.au
[46] The ALA Data Quality Assurance http://www.ala.org.au/about-the-atlas/how-we-integrate-data/data-quality-assurance/
[47] The ALA Sandbox Service for Data Uploads http://sandbox.ala.org.au
[48] The ALA Biodiversity Volunteer Portal http://volunteer.ala.org.au
[49] Announced 17 April 2012; http://www.mct.gov.br/index.php/content/view/337439.html
[50] The Chinese Academy of Sciences (CAS), Institute of Botany http://english.ib.cas.cn/
[51] The Chinese Nature Museum www.cfh.ac.cn
[52] The CAS, Institute of Zoology http://english.ioz.cas.cn/
[53] The CAS, Catalogue of Life http://test.pilot.4d4life.animal.net.cn/
[54] The CAS, Taxonomic Tree Tool http://ttt.biodinfo.org/indexen.asp
[55] The CAS, Kunming Institute of Botany http://english.kib.cas.cn/
[56] The CAS, Kunming Institute of Botany MitoTool http://www.mitotool.org/
[57] The Chinese Animal Scientific Database http://www.zoology.csdb.cn/
[58] The Database of Animal Resources in South-West China http://www.swanimal.csdb.cn/
[59] The CAS, Kunming Institute of Botany, Antimicrobial Peptide Database http://159.226.149.45/other1/kizapd/
[60] The DataONE e-infrastructure http://mule1.dataone.org/ArchitectureDocs-current/overview.html
[61] Currently 3, as of Spring 2012, located at University of New Mexico, University of California Santa Barbara and University of Tennesse (in collaboration with Oak Ridge National Laboratory)
[62] The DataONE Architecture http://mule1.dataone.org/ArchitectureDocs-current/
[63] The DataONE Data http://mule1.dataone.org/ArchitectureDocs-current/design/WhatIsData.html
[64] The DataONE Tools http://www.dataone.org/all-software-tools
[65] The DataONE Best Practives http://www.dataone.org/best-practices
[66] The GBIF GBRD Application http://gbrds.gbif.org/
[67] The GBIF Data http://data.gbif.org/
[68] The GBIF Services Tutorials http://data.gbif.org/tutorial/services
[69] The Biodiversity Information Standard, Taxonomic Databases Working Group (TDWG) http://www.tdwg.org/
[70] The GEO BON Implementation Overview http://www.earthobservations.org/documents/geo_v/07_%20GEO%20Bon%20-%20Implementation%20Overview.pdf
[71] The GEO BON Detailed Implementation Plan http://earthobservations.org/documents/cop/bi_geobon/geobon_detailed_imp_plan.pdf

[72] Much of the description in this section is extracted / summarised from the document: Ó Tuama, Saarenmaa, et al. Principles of the GEO BON Information Architecture, version 1, 14th June 2010.

[73] May 2012. Negotiations are presently underway between the EU BON consortium and EC FP7 funding programme to obtain funding for a European contribution to GEO BON. EU BON's deliverables include a comprehensive "European Biodiversity Portal" for all stakeholder communities, and strategies for a global implementation of GEO BON and for supporting IPBES.

[74] The GEOSS Registry System http://geossregistries.info/

[75] http://www.lifewatch.eu/images/stories/PDFs/WP5_Deliverables/rm-v1.0.pdf

[76] The Biodiversity Sciences Web Services Registry www.biodiversitycatalogue.org

[77] The myExperiment Web Portal www.myexperiment.org

[78] The SANBI Biodiversity Information Online (SIBIS) http://sibis.sanbi.org/

[79] The South African Node of the Global Biodiversity Information Facility (SABIF) http://www.sabif.ac.za

[80] The SANBI Biodiversity GIS http://bgis.sanbi.org/

[81] The SANBI Online Checklist of Plants of Southern Africa http://posa.sanbi.org/

[82] The SANBI Redlist of South African Plants http://redlist.sanbi.org/

[83] The SANBI Species Status Database http://www.speciesstatus.sanbi.org/default.aspx

[84] The SANBI Environmental Impact Assessment Toolkit http://www.eiatoolkit.ewt.org.za/

[85] The SANBI iSPOT Application http://www.ispot.org.za/

[86] Reference under embargo. Paper recently submitted to BMC Ecology.

[87] Simple RI representation as initially introduced in LifeWatch.

[88] The Unified Modeling Language (UML) http://www.uml.org/

[89] The Legal Framework for a ERIC http://ec.europa.eu/research/infrastructures/pdf/eric_en.pdf

[90] The Portlet standard http://en.wikipedia.org/wiki/Portlet

[91] Commercial-Off-The-Shelf (COTS) http://en.wikipedia.org/wiki/Commercial_off-the-shelf

[92] The JBoss Application Server http://www.jboss.org/

[93] The Matlab Toolkit http://www.mathworks.fr/products/matlab/

[94] Global Unique Identifier (GUID) http://en.wikipedia.org/wiki/Globally_unique_identifier

[95] Service Level Agreement (SLA) http://en.wikipedia.org/wiki/Service-level_agreement

[96] The Wikipedia Definition of Interoperability http://en.wikipedia.org/wiki/Interoperability

[97] The Lifewatch Reference Model http://www.lifewatch.eu/images/stories/PDFs/WP5_Deliverables/rm-v1.0.pdf

[98] The Open Distributed Processing (ODP) Reference Model http://www.itu.int/ITU-T/recommendations/index.aspx?ser=X

[99] Interoperability already in place or being developed with other Research Infrastructures (RI)

[100] The Indian Ocean Climate Initiative (IOCI) http://www.ioci.org.au/

[101] The LifeMapper Project http://www.lifemapper.org/

[102] The SIBBr/CRIA OpenModeller Tool http://openmodeller.sourceforge.net/

[103] The List of Brazilian Flora http://floradobrasil.jbrj.gov.br/

[104] The INCT Herbario Virtual de Flora e dos Fungos http://inct.florabrasil.net/en

[105] The RedCLARA Network http://www.redclara.net/

[106] The GEANT Network for Education and Research http://www.geant.net/

[107] The GEANT eduGAIN Service Portal http://www.geant.net/service/edugain/pages/home.aspx

[108] The CIPO Portal http://www.cipo.org.br/portal/

[109] The Global Lambda Integrated Facility http://www.glif.is/

[110] The Dried Specimen Herbarium http://en.wiktionary.org/wiki/exsiccata

[111] TDWG Access Protocol for information Retrieval (TAPIR) http://www.tdwg.org/dav/subgroups/tapir/1.0/docs/TAPIRNetworkBuildersGuide_2010-05-05.html
[112] The UK's Royal Botanic Gardens KEW http://www.kew.org/
[113] The World Agroforestry Centre http://www.worldagroforestrycentre.org/
[114] The Extreme Science and Engineering Environment https://www.xsede.org/
[115] Identity Provider (IdP) http://en.wikipedia.org/wiki/Identity_provider
[116] The Encyclopedia of Life http://eol.org/
[117] The Map of Life http://www.mapoflife.org/
[118] The Inter-American Biodiversity Information Network (IABIN) http://www.oas.org/en/sedi/dsd/iabin/
[119] The Wallace Initiative http://wallaceinitiative.org/
[120] The IUCN Red List of Threatened Species http://www.iucnredlist.org/
[121] GBIF MoU partnerships http://www.gbif.org/governance/partnerships/
[122] The Open Geospatial Consortium (OGC) http://www.opengeospatial.org/
[123] The Group on Earth Observation Standards and Interoperability Forum http://seabass.ieee.org/groups/geoss/index.php?option=com_content&task=view&id=17&Itemid=61
[124] The Long Term Ecological Research network http://www.lternet.edu/
[125] The European Ocean Biogeographic Information System (EOBIS) http://www.marbef.org/data/eurobis.php
[126] The FLORAweb portal http://www.floraweb.de/
[127] The Flanders Wetland Sites http://www.inbo.be/content/page.asp?pid=EN_ENV_ECO_Flawet
[128] The European Data project http://www.eudat.eu/
[129] The GBIF Integrated Publishing Toolkit http://code.google.com/p/gbif-providertoolkit/ , http://www.gbif.org/informatics/infrastructure/publishing/
[130] The GBIF Multimedia Resources Task Group (MTRG) http://www.gbif.org/informatics/primary-data/task-groups/mrtg/
[131] The Biological Data Recording System http://code.google.com/p/ala-citizenscience/ , https://m.ala.org.au/
[132] The DataONE Public Participation in Science and Research Working Group and Projects http://www.dataone.org/sites/all/documents/PPSR_Charter.pdf
[133] The GeoWiki Project http://geo-wiki.org/
[134] The EuroGEOSS Project http://www.eurogeoss.eu/
[135] SANBI's iSpot Citizen Science Portal http://www.ispot.org.uk/
[136] ALA Google Code Environment http://code.google.com/p/ala-portal/
[137] GBIF Google Code Sites http://www.gbif.org/communications/resources/platforms/gbif-google-code-sites/
[138] ALA TRIN Wiki http://wiki.trin.org.au/
[139] ALA Biodiversity Heritage Library http://bhl.ala.org.au/
[140] ALA Morphbank Database of Images http://morphbank.ala.org.au/
[141] ALA Sandbox http://sandbox.ala.org.au/datacheck/
[142] The Australian Barcode of Life Network (ABOLN) http://www.ala.org.au/about-the-atlas/our-data-providers/aboln/ , http://ibol.org/australia/
[143] The Species Interaction of Australia Database (SIAD) http://www.discoverlife.org/siad/
[144] The INCT Virtual Herbarium http://inct.splink.org.br/
[145] The GBoWS Genobank http://www.genobank.org/

[146] The DataONE Investigator Toolkit (ITK) http://mule1.dataone.org/ArchitectureDocs-current/design/itk-overview.html
[147] DataONE ONEMercury Tool https://cn.dataone.org/onemercury/
[148] DataONE DMP Tool http://www.dataone.org/data-management-planning
[149] DataONE ONE R Client Tool http://mule1.dataone.org/ArchitectureDocs-current/design/itk-d1r.html
[150] The Vistrails Workflow Interface http://www.vistrails.org/
[151] The Kepler Workflow Interface https://kepler-project.org/
[152] The Taverna Workflow Interface http://www.taverna.org.uk/
[153] The myExperiment Platform http://www.myexperiment.org/
[154] The Matlab Programming Language and Environment http://www.mathworks.fr/products/matlab/
[155] The Global Biodiversity Resources Discovery System http://www.gbif.org/informatics/infrastructure/discovering/
[156] The Group on Earth Observation GEO portal http://www.geoportal.org/
[157] The Group on Earth Observation Clearinghouses http://www.earthobservations.org/gci_gci.shtml
[158] The Group on Earth Observation Standards and Interoperability Registry http://seabass.ieee.org/groups/geoss/index.php?option=com_content&task=view&id=17&Itemid=61
[159] The Biodiversity Virtual e-Laboratory (BioVeL) http://www.biovel.eu/
[160] Common Operations of Environmental Research Infrastructures http://envri.eu/
[161] The Virtual Biodiversity Research and Access Network for Taxonomy (ViBRANT) http://vbrant.eu/
[162] The SANBI Biodiversity GIS portal http://bgis.sanbi.org/
[163] The SANBI Biodiversity Advisor http://biodiversityadvisor.sanbi.org/
[164] User Management, Authentication and Authorisation (UAA), according to OGC Best Practice 07-097 http://portal.opengeospatial.org/files/?artifact_id=23286
[165] The Central Authentication Service (CAS) http://www.jasig.org/cas
[166] The Shibboleth federated identity framework http://shibboleth.net/
[167] The Open Geospatial Consortium http://www.opengeospatial.org/standards/common
[168] The postgreSQL Open Source Database http://www.postgresql.org/
[169] The Apache TOMCAT http://tomcat.apache.org/
[170] Hazelcast http://www.hazelcast.com/
[171] The World Wide Web Consortium (W3C) http://www.w3.org/
[172] The Open Standards for Information Society (OASIS) https://www.oasis-open.org/standards
[173] The Apache HADOOP http://hadoop.apache.org/
[174] The European Grid Initiative (EGI) http://www.egi.eu/
[175] The Partnership for Advanced Computing in Europe (PRACE) http://www.prace-project.eu/
[176] Data sharing means making provision to allow data owners to share (upload) their data through the infrastructure so that it may be used by others under some kind of licensing terms.
[177] Primary Biodiversity Data is defined by GBIF as: Digital text or multimedia data record(s) detailing facts about the instance of occurrence of an organism, i.e. on the what, where, when, how and by whom of the occurrence and the recording.
[178] International Long Term Ecological Research http://www.ilternet.edu/
[179] The Digital Object Identifier http://www.doi.org/
[180] The DataCite Portal http://www.datacite.org/
[181] The potGIS Portal http://postgis.refractions.net/

SEVENTH FRAMEWORK PROGRAMME

[182] The GeoServer Portal http://geoserver.org/
[183] The GeoNetwork Portal http://geonetwork-opensource.org/
[184] The CRIA mapCRIA Web Service http://www.cria.org.br/mapcria/doc/
[185] The South African Geospatial Dictionary (SAGDAD)
[186] The Darwin Core (DwC) http://rs.tdwg.org/dwc/
[187] The Darwin Core Archive (DwC-A) http://www.gbif.org/informatics/standards-and-tools/publishing-data/data-standards/darwin-core-archives/
[188] The Ecological Metadata Language (EML) http://knb.ecoinformatics.org/software/eml/
[189] The Distributed Generic Information Retrieval (DiGRI) http://digir.sourceforge.net/
[190] The Biological Collection Access Services (BioCASE) http://www.biocase.org/
[191] The Open Archive Initiative Protocal for Metadata Harvesting (OAI-PMH) http://www.openarchives.org/pmh/
[192] The Keyhole Markup Language (KML) https://developers.google.com/kml/
[193] The Context Objects in Spans (COinS) http://ocoins.info/
[194] The Apache CASSANDRA http://cassandra.apache.org/
[195] The Apache HBASE http://hbase.apache.org/
[196] The Apache SOLR http://lucene.apache.org/solr/
[197] The GBIC Conference Initial Report http://imsgbif.gbif.org/CMS_ORC/?doc_id=4711&download=1