Deliverable 4.1 Legal input discussion document for workshop 1

Project acronym: *GLOBIS-B*
Project full title: "GLOBal Infrastructures for Supporting Biodiversity research"
Grant agreement no.: 654003

Due-Date:	M6
Actual Delivery:	M8
Lead Partner:	UAH
Dissemination Level:	PU
Status:	Final
Version:	1

DOCUMENT INFO

Date and version no.	Author	Comments/Changes
1	**Enrique Alonso**	

TABLE OF CONTENTS

1   Executive summary

The GLOBIS-B project aims to facilitate the global cooperation of world-class research infrastructures with a focus on supporting frontier research on biodiversity. The project aims to contribute to developing global indicators and other key measurements that are required to study, report and manage biodiversity change (Pereira et al., 2013). More specifically, the project is focusing on potential infrastructure services supporting research on measuring biodiversity change, specifically Essential Biodiversity Variables (EBVs). Thereby GLOBIS-B serves a major aim of the Group on Earth Observations Biodiversity Observation Network (GEO BON). More details about the GLOBIS-B project can be found in Kissling et al. (2015) and on the project homepage (www.globis-b.eu).

This document is supplementary to the document "Background and program for the 1st Workshop in Leipzig, 29 February – 2 March 2016" that contains a detailed description of the technical and scientific issues of the project and of the 1st Workshop, and which summarizes legal and policy issues to be analyzed as: "Views on associated policy and legal implications, e.g. a.- Preliminary observations on legal issues; b.- Potential implications for contributing to policy requirements and to inform policy bodies".

The Globis-B project itself identifies its main objective as follows: "The objective of GLOBIS-B is to facilitate the multi-lateral cooperation among these research infrastructures with the aim to advance the implementation of EBVs. This includes to: (1) specify users' requirements for extracting, handling, visualizing and analyzing the required biodiversity data; (2) identify where these research infrastructures can enhance existing capabilities or develop new ones; (3) agree on realistic solutions for supporting user requirements so that the research infrastructures can offer targeted services to calculate selected EBVs; (4) draw up best practices for infrastructure support; and (5) address legal implications with respect to licensing, intellectual property rights, and sharing of resources."

It identifies the policy and legal aspects around a broad notion of "legal interoperability" that focuses on the legal side of technical interoperability: "The other challenge is related to legal interoperability, conditional for the automatic processing of data as supported by 'machine–machine' interactions across political and institutional boundaries. This is often difficult due to varying provenance of authorship and ownership of data and requires the identification of legal and policy bottlenecks. Comparing mutual agreements between research infrastructures and data providers may avoid potential problems with respect to sharing of data and resources, re-licensing of services and adoption of open/public access principles. Such topics are addressed internationally by the Research Data Alliance and its RDA-CODATA legal interoperability interest group (https://rd-alliance.org/group/rdacodata-legal-interoperability-ig.html), but issues specific to EBV-relevant biodiversity data need to be assessed separately". But it is also true that more recent developments under the GEO process have also created an additional more specific framework that not only is it applicable to one of the main partners of the Globis-B project (GEO BON) but also but that was agreed, in the kick-off meeting, as an essential policy-making framework within which Globis-B should try to work, in particular since both its Data Sharing and Data Management Principles have been agreed and are ready for final sponsorship by the GEO community worldwide (and they have been "partially" incorporated in the U.S. Common Framework for Earth Observation on December 22, 2015, under review until January 15th 2016).

A final introductory remark: many of the main obstacles and bottlenecks created by the lack of adequate legal interoperability of biodiversity-related data were already identified in a previous EU project. See point 6 of the attached document: "Flock together with CReATIVE-B: A roadmap of global research data infrastructures supporting biodiversity and ecosystem science"

2 Contributors

Enrique Alonso as WP leader, with input from various other legal and data policy experts within the context of RDA, GEO and CODATA.

3 Main body of the report

3.1 The legal and policy issues of the GLOBIS-B project

The previously listed documents encompass the main issues clustered around the world-wide framed concept of legal interoperability. The main idea of the legal experts participating in the Leipzig meeting is to be aware of the issues and be ready to understand which of them may imply problems in the access and use of data needed to quantify EBVs. The listed topics are simply an array of the main issues that the management of research data raises in general. Some of them might not be present at all in the realm of biodiversity-related data or, in particular, in the data and software needed for the specific EBVs addressed by the project, starting with species abundance in the 1st and 2d workshops.

1. Are there top-down and/or bottom-up approaches to harmonized policies (or mandates) on open access and data sharing of all the types of data needed to measure EBVs (including time and spatial changes) in particular of species abundance? The role of GBIF and other international datasets as well as national biodiversity inventories as adequate harmonized sources of data.

2. Common standards and protocols mandated by legal instruments; either by hard law mandates (national, supranational or international treaty-based mandates), by legal conditioning of publicly funded research, or by contractual or customarily based agreements of the different scientific communities and the IT web services that link together data providers and data users. The effects of recommended v. totally open standards (e.g. the US US Group on Earth Observations (USGEO) Common Framework for EO Data (CFEOD) is list of recommended standards).

3. Common notions of what is implied by dedication of biodiversity data to the public domain. Assessment of how many of the data sets and which of them are dedicated to the public domain worldwide.

4. Legitimate limitations of biodiversity data restricting open access and data sharing in general (e.g. genetic resources, location of endangered species…)

5. In particular: science-based limits in the different fields of biodiversity : a) confidentiality of data (& trade secrets) and embargo periods; b) "it´s my data syndrome" & professional careers based on data secreccy; different philosophies of the different scientific communities of relevance for the provision of the data needed to measure biodiversity change, EBVs in particular and specifically species abundance.

6. Use by biodiversity researchers of CC licenses and their limits (in particular CC0 licenses for data)

7. Management of IPRd expressions of data and of data itself by existing databases and web services.

8. Attribution management of biodiversity-related data, management of aggregation of attributions in digitized biodiversity data.

9. Metadata addressing rights statements issues. Is it a useful approach?

10. Software licenses limits to data retrieval or use; in particular GIS software and virtual environments software.

11. Biodiversity-related data are of multiple scientific origin since not only all sorts of biotic data but also of abiotic data are of essence to understand biodiversity at all levels (from the genetic to the landscape level). Biodiversity is interdisciplinary by essence. Should norms (instead of rules) be used to facilitate the exchange of data among different distinctive communities? Should biodiversity web services and e-infrastructures work on a normative (instead of legal) basis? Is it realistic to create "a-legal" operative frameworks in which IPRs and other limits to data access and use are simply set aside through normative schemes based on a common "scientific culture"?

12. In the era of the knowledge economy based on digitized technologies, data has become a valuable business asset that can be exploited through licensing or a sale to third parties. Data protection due to the increase of the value of data as an object that can be the focus of transactions per se or as an ancillary consideration of technology services arrangements has not only spurred a legal movement to increase data protection but a whole legal culture of data management as a business service subject to its own particular rules. The fact that open access asserts itself as a the opposite legal principle due to the need to advance science and public policy (when data is being produced via public funding) cannot obscure the fact that this legal culture on how data needs to be managed may be applicable independently of the fact that it is managed as a public service or as a private business (or non-profit) venture. What could be called "data licensing agreements" is a subsector of IT digital law which operates under an umbrella of statutory, customary or case law governing different aspects of the process. So not only are there new rules regulating "data ownership" or "data licensing" (the main difference being that the first set of rules are enforceable against third parties while the later are only enforceable vis a vis the other contractual parties but with the caveat that the simple interaction with a web customer can create zillions of contracts per year either by simple clickings or even by accessing a web page via internet such it is the case with the so-called clickwrap or browsewrap agreements). Are biodiversity related web services or e-infrastructures bound by such rules?

3.2 The broader context: GEO

Research data are a particular set of data for the sake of the issues listed in point 2.

Traditionally, these issues have been tackled with in different ways by the distinctive scientific communities. Some of them are used to very open access and use frameworks, including society in general; some others are only open to the registered members of the community; and some others are used to normative systems in which data are considered ownership of the holder (or first "discoverer" or author of its first expression). Although publicly funded research is progressively moving toward open access there is still a very long way to go.

Being biodiversity-related data multidisciplinary by essence, only broader cooperative frameworks mat be efficient enough to provide common standards, understanding of statements of rights (or absence of them, e.g. public domain), legitimate limits to data sharing or openness, or the utility of working under common notions such as fair use/fair dealings, norms v. rules, public domain status, free software licenses or other common clauses or licenses/waivers/terms of use…

The Globis-B project , not entirely disregarding the possibility of working on particular recommended or prescribed schemes particular to biodiversity-data, is committed to approach the treatment of the data needed to measure EBVs under the broader legal interoperability schemes of two of the most recent developments: the GEO and the RDA processes.

During the year 2015, GEO has produced two main sets of principles: the Data Sharing and the Data Management Principles. Globis-B, not only because of the participation of GEO BON as a partner, is supposed to be able to test whether both sets of principles can be implemented when developing the strategy for EBVs measurement. GEO still, notwithstanding its advancements (see Barbara Ryan, Director. GEO Secretariat. "Unleashing the Power of Earth Observations –Together", September 2015), has still to be convinced of the importance of data sharing (see P. Uhlir, "*The Value of Open Data Sharing: A White Paper for the Group on Earth Observations*", November 2015 https://www.earthobservations.org/documents/geo_xii/GEO-XII_09_The%20Value%20of%20Open%20Data%20Sharing.pdf).

The 3 GEO Data Sharing Principles are the following:

1. data, metadata and products will be shared as Open Data by default, by making them available as part of the GEOSS Data Collection of Open Resources for Everyone (Data-CORE) without charge or restrictions on reuse, subject to the conditions of registration and attribution when the data are reused;

2. where international instruments, national policies or legislation preclude the sharing of data as Open Data, data should be made available with minimal restrictions on use and at no more than the cost of reproduction and distribution; and

3. all shared data, products and metadata will be made available with minimum time delay. Their post-2015 status can be examined in http://www.earthobservations.org/dswg.php

The 10 Data Management Principles and theirs Implementation Guidelines status is the following:

Data Management Principles Strategy

The value of Earth observations are maximized through data life-cycle management based on ten Principles supporting five themes.

DISCOVERABILITY
DMP-1: *Data and metadata* will be discoverable

ACCESSIBILITY
DMP-2: Data will be accessible via online services

USABILITY
DMP-3: Encoding   DMP-4: Documentation
DMP-5: Traceability   DMP-6: Quality

PRESERVATION
DMP-7: Preservation   DMP-8: Verification

CURATION
DMP 9: Review and reprocessing
DMP 10: Persistent and resolvable identifiers

DMPs Implementation Guidelines

- Description / explanation of each principle and implementation
- Guidance on implementation, with examples
- Resources for implementing each principle
- Metrics measuring adherence to implement each principle

http://digitalliteracy.us/wp-content/uploads/2013/09/Guideline_Icon.jpg

Implementation Guidelines
Review Process

- GEO-XII Plenary (11-12 Nov 2015): Present preliminary draft for review
  - Comments at http://www.earthobservations.org/survey_dmp_ig
  - Deadline: 15 March 2016
- GEO WP Symposium (~ May): Review updated draft
- Living Planet Symposium (9-13 May): Review updated draft
- GEO Program Board (~ July): Review near-final draft
- GEO-XIII Plenary (~ Nov): Present final draft version for adoption

http://www.qmed.com/files/ck_images/File-abstract.jpg

https://www.earthobservations.org/documents/geo_xii/GEO-XII_10_Data%20Management%20Principles%20Implementation%20Guidelines.pdf

They have already started to be taken under consideration by the USGEO Data Management Working Group, an independent GEO group co-chaired along with NASA and USGS, has already produced a first

draft of the US Group on Earth Observations (USGEO) Common Framework for EO Data (CFEOD), which is essentially a list of recommended standards to enable data discoverability, accessibility, and usability. Implementing these recommendations would directly support several (but not all) of the GEO DM Principles developed by DMP-TF, and the CFEOD document explicitly indicates which of the DMP are supported by the standards in each area. CFEOD was developed by USGEO Data Management Working Group. All of the CFEOD recommended standards are listed in the GEOSS Standards Registry (https://www.earthobservations.org/gci_sr.shtml)

 The difference is that GEOSS does not favor any of those Standards more than the others, treating every one equally, whereas USGEO CFEOD explicitly recommends a smaller list that are in broader use to better promote interoperability.

For the US CFEOD document which under public review until January 15th, see
https://www.whitehouse.gov/administration/eop/ostp/library/shareyourinput

Whether these DMPs are approached in a selective way, as the US has proposed, or in a broader all-encompassing manner by the experts looking for data available for measuring the selective EBVs, should be a policy issue that the legal experts should assess, Ideally only data appraised and managed following these principles should be of enough quality to ensure an adequate EBV indicator reliable system. But the GEO DMPs are on the making as seen in the figure above, so they are open for debate.

3.3   The broader context: the RDA legal interoperability principles

More focused on legal interoperability, the RDA.CODATA legal interoperability IG has been working on the main issues implied in all research data.

The first draft of the 10 Principles and their Implementation Guidelines has just been finalized by the mid-December 2015.

Notwithstanding their provisional status (just submitted to review by CODATA and the RDA) they are going to be presented in the next RDA plenary in Tokyo, coinciding with the dates of the Leipzig meeting. Although the documents are still under limited distribution, the legal experts attending the 1st Workshop of Globis-B are supposed to expose the scientific and IT experts to the Principles and Guidelines in order to double check if the data needed to measure the EBVs follows these principles, which would ensure the legal interoperability of the data thus preventing the EVs from being developed with limited data or with data that cannot be used in a way that would make the EBVs workable at a global scale because of the bottlenecks and imitations in data access or data flows.

Since they are still under limited distribution they are attached as documents 3 and 4.

The RDA-CODATA Legal interoperability principles have been developed as a way to deal with a first layer of basic legal problems that arise when managing digitized data. What matters is the legal interoperability (not semantic or technological or any other interoperability) of the data with other data, assuming that data are "something" used by researchers and scientists irrespective of where they come from and how they are stored, managed and accessed. In a computing sense, they are a response to a query. What happens behind the scene to deliver that response is out of scope for this IG for the moment being since if legal interoperability is established and promoted at this level, that will already perform a very valuable service. E.g. It doesn't matter how the data are obtained, since data come in a multiple formats and modes, what legal interoperability intends to provide is legal clarity for its use since some results will be copyrightable and others not, some will have certain other restrictions and others not. The key is clarity so that interoperability can be maximized.

3.4 Interactive databases and e-infrastructures data management issues: software.

Limits to legal interoperability might also stem from the data itself but from the software where they are embedded, if such software needs to be access or used in order to access or simply reorganize or mix the data with other data sets organized in a different way.

The CReATIVE-B Roadmap (attached Doc 1) example of the Biovel enquiry (figure3 page 18) shows that many of the bottlenecks in data flows are software-license related.

Open source then becomes the key to open access. With the exception of GIS related software, they tend to be expressed in open source formats and applications. But that might not be true in all instances. The legal experts attending the 1st Workshop should pay attention to potential limitations of the farther use of data because of the need to obtain licenses for their processing (see e.g. TeresaGomez-Diaz. CNRS. Laboratoire d'informatique Gaspard-Monge. "Free software, Open source software, licenses: A short presentation including a procedure for research software and data dissemination", September 14th 2014).

Virtual environments usually imply new apps and new software. Is there a clear policy of free access provision of algorithms, formats and applications/software needed for farther reuse of the data when provided by the holders of the main data sets considered to be of essence for the measurement of the EBVs?

3.5 Interactive databases and e-infrastructures data management issues: policies and legal interoperability.

But most web services and all e-infrastructures need farther policies and agreements in order to ensure that they prevent getting into legal problems when managing data produced or "owned" by third parties or by parties with whom they engage in sharing agreements.

This is usually not referred to as legal interoperability but as a broader set of issues that experience has shown that such services have to live with in order to prevent eventual legal problems when managing digitized data from multiple interactive actors. The main legal principles on which web services that manage data base their activities have been developed outside the open access movement as part of the "private data" management subsector of the technology-based knowledge economy (also called the 4th Industrial Revolution).

Thus, since "data" is a very valuable business asset, the list of potential ways to protect data as IP are increasing in the last years, the classic copyright being only one of many other ways to obtain such protection: trade secrets, data as intangible property (protected by tort law such as trespass to chattels or conversion), computer fraud and abuse, sui generis data bases protection, or novel forms of patent protection have been developed to ensure "ownership of data" (see, e.g., Michael W. Carroll, "*Sharing Research Data and Intellectual Property Law: A Primer*", PLoS Biology 13(8): e1002235. doi:10.1371/journal.pbio.1002235, August 27, 2015). These services are provided under multiple types of licenses, conditions, terms of use, waivers, representations, warranties … with contents about access or usage (exclusivity, sublicensing, limits governed by the purpose, premises, personnel, geography, or types of devices, decompilation, alteration or obscuring of the data…, or the typical SaaS ("software as a service") data hosting or data service agreements, with usage conditions - understanding by "Usage Data" , for example, "any data reflecting the access or use of the Services by or on behalf of Customer or any Authorized User, including any end user profile-, visit-, session-, impression-, click through- or click stream- data and any statistical or other analysis, information or data based on or derived from any of the foregoing"-.

Other clauses or conditions have as main purpose to allocate the risk of damage to third parties and not only to the parties involved in the contractual (or license/terms of use) relationship, so disclaimers, representations and warranties as well a clauses on indemnification and limitations on liability such as those that require each party to defend and indemnify the other party against third-party claims resulting from the other party's breach of the agreement, with or without liability caps or exclusion of damages for lost data, profits, business and reputation, and any indirect, special, punitive or consequential damages (SeeDaniel Glazer, Henry Lebowitz & Jason Greenberg, Fried, Frank, Harris, Shriver & Jacobson LLP and Practical Law Intellectual Property & Technology, "*Data as IP and Data License Agreements*", Practical Law Company, 2013; Enrique Alonso, Elizabeth Kirby, Chris Morris, Irene Schluender & Stephanie Suhr, "*Open Web Ser4vices and the Law*", Pending publication, 2016).

Sometimes they are simply referred to as "Infrastructure policies". Infrastructures enable researchers in different locations across the world to collaborate in the context of their institutions or in national or multinational scientific initiatives, which leads to the need to combine multiple software components into a coherent and centrally managed system of hardware, software, and data resources which may not be bound to a specific community; may be operated by third-party providers; may serve uncoordinated data providers and data consumers; may need to support heterogeneous and multi-disciplinary communities; and may be built by combining other infrastructures.

So, the combination of web services that e-infrastructures imply needs a set of coherent exploitation policies to regulate copyright, workspace service, email service, social service, notification service, Virtual Research Environments, sharing data, derivative work …etc

In the case of biodiversity related data, web services and e-infrastructures (in particular those consisting on virtual labs) the management of data has shown that not only legal interoperability in the strict sense and open access to data and software is needed but that other sets of norms or agreements are also needed.

As Pasquale Pagano has described, VREs need a distributed and dynamically created environment , where subset of resources (data, services, computational, and storage resources) - regulated by tailored policies (e.g. data encryption with VRE specific key, quota on service calls and storage usage, …)-, are assigned to a subset of users via interfaces for a limited timeframe at little or no cost for the providers of the participatory data e-infrastructures (Pasquale Pagano. CNR. "*Infrastructure Policies*". EGI Forum, Bari, November 2015; see also http://www.slideshare.net/PasqualePagano/virtual-research-environments-as-aservice).

These set of policies have recently tended to be clustered or organized in different ways but respond to similar models.

It is unclear at this stage to which extent the management of data needed for EBVs measurements will imply the need for such VREs or equivalent complex web services organization, but the experience that comes from the biomedical realms (including non-human genetic data in biobanks) and some of the examples already existing on data management of biodiversity data (and much more if ecological abiotic data is used) shows it is going to be unavoidable.

Limiting the examples to databases and services in the area of biological resources two examples can be descriptive of the set of data policies needed (and the jargon used to describe the sets or groups of policies). The first one could be the BlueBridge infrastructure (see Pasquale Pagano, op. cit.; or Donatella Castelli, Leonardo Candela, Pasquale Pagano Gianpaolo Coro. CNR. "*Maximizing uptake by opening access to research: The BlueBRIDGE endeavor*", EGI Forum, Bari, November 2015,

http://es.slideshare.net/BlueBridgeVREs/egi-forum-2015-gcoro-et-al-maximising-uptake-by-opening-access-to-research-the-blue-bridge-endeavor-readonly).

The other one is the DataOne analysis of the coordination of policies implied in such VREs, in particularla when biodiversity data is provided by multiple citizen science communities or even openly crowd-sourced (see Anne Browser, Andrea Wiggins & Robert D. Stevenson, "*Data Policies for Public Participation in Scientific Research: A Primer*", DataOne Public Participation in Scientific Research Working Group, August 2013
http://www.birds.cornell.edu/citscitoolkit/toolkit/policy/Bowser%20et%20al%202013%20Data%20Policy%20Guide.pdf).

The main task of the legal experts would consist in flagging out where the solutions to measure the EBVs provided by the technical and scientific experts might be weakened or have constraints based on these types of –or equivalent- legal environments under which the management of data will need to function.
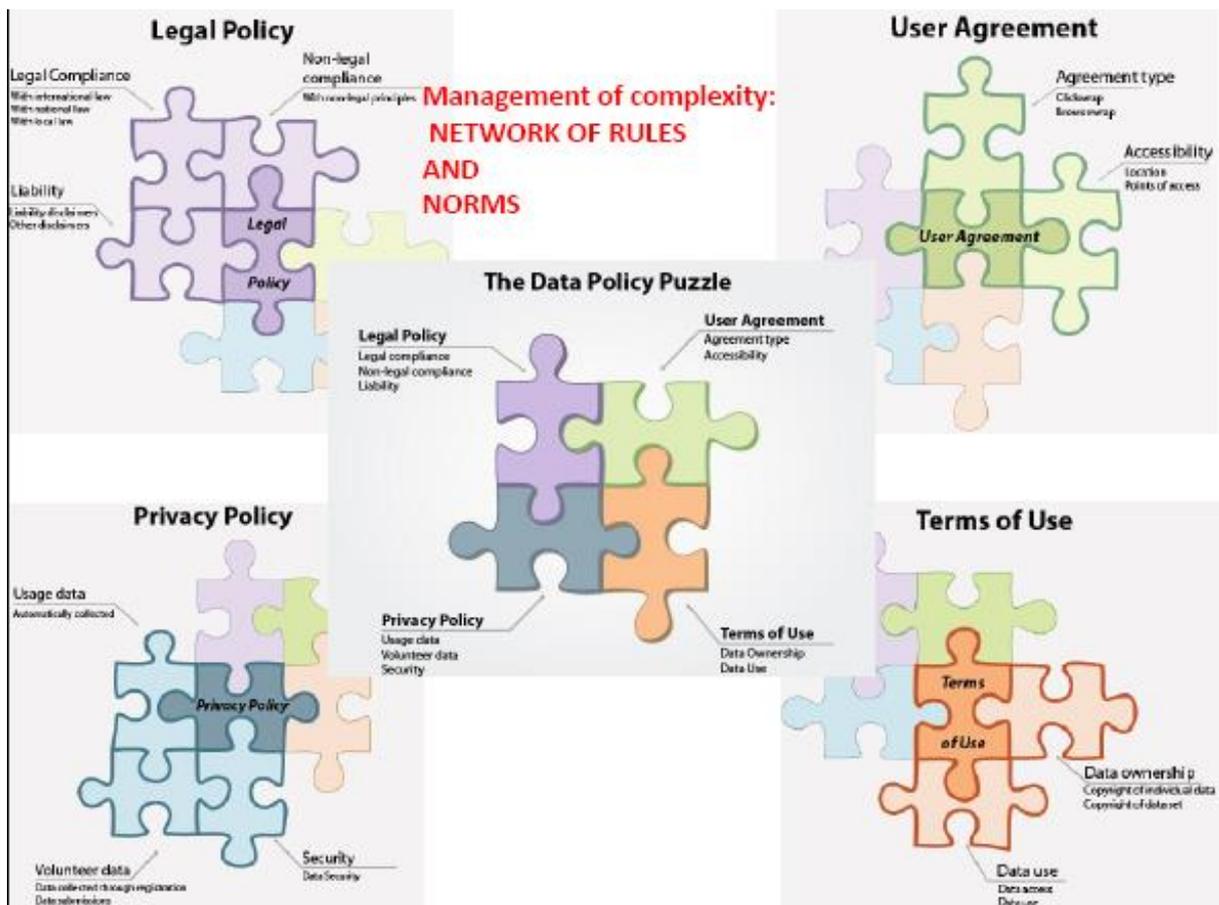
3.6 Other issues

The experts should also pay attention to any other policy or legal problems concerning data management that could prevent the measurement of EBVs and that might stem from the debates, such as e.g.

- the preference of normative v. legal models (see John Wilbanks "*Public domain, copyright licenses and the freedom to integrate science*", Journal of Science Communication - JCOM 7 (2), June 2008);

- whether marine species data of relevance for EBVs –in particular GIS georeferenced Ocean data- have intrinsic different problems (see Douglas J. McCauley, Malin L. Pinsky, Stephen R. Palumbi, James A. Estes, Francis H. Joyce & Robert Wagner, "*Marine Defaunation: Animal loss on the global ocean*", in Science 347, 1255641 (2015). DOI: 10.1126/science.1255641; David Penman, Andrew Pearce & Missy Morton, "*The Census of Marine Life: review of lessons learned*", June 2011, or Antonio Machado, Integrated Marine Data Repository for Macaronesia -REDMIC-, in International Coastal Atlas Network Newsletter, Volume Two Number 1, March 2013, pgs 12 ff, http://icoastalatlas.net);

- what legal problems might arise from incorrect deidentification techniques when species are associated to humans (see Robert Gellman. "*The Deidentification Dilemma: A Legislative and Contractual Proposal*", in Fordham Intellectual Property, Media and Entertainment Law JournalVolume 21, Issue 1, 2011. Article 2,Vol. XXI Book 1 –only of such species are selected as an example, e.g. microorganisms;

- even if the analysis of the RDA-CODATA Principle 5 already addresses the issue, if it would be possible to agree to a common rights statement metadata that at least distinguishes between "public domain" data, by that implying "that data may be used by anyone, anywhere, anytime without permission, license or royalty payment without prejudice to its conditioning to requirements of user registration and/or proper attribution", and "in Copyright", "Licensed ref." or equivalent or similar metadata expressing that there are some rights attached to the data (see Europeana-DPLA, May 2015 "*Recommendations for the Technical Infrastructure for Standardized International Rights Statements*");

- how to treat aggregation of attribution data or changes in the original data via addition on data to datasets, or of additional data to single specific data, whether taxonomic or otherwise (see Lee Belbin, Joanne Daly, Tim Hirsch, Donald Hobern & John La Salle, "*A specialist's audit of aggregated occurrence records: An 'aggregator's' perspective*", ZooKeys 305: 67–76. 2013) etc..