UNEP

CBD

Convention on Biological Diversity

Distr.
GENERAL

UNEP/CBD/COP/11/INF/8
17 August 2012

ENGLISH ONLY

CONFERENCE OF THE PARTIES TO THE
CONVENTION ON BIOLOGICAL DIVERSITY
Eleventh meeting
Hyderabad, India, 8-19 October 2012

A REVIEW OF BARRIERS TO THE SHARING OF BIODIVERSITY DATA AND INFORMATION, WITH RECOMMENDATIONS FOR ELIMINATING THEM

Note by the Executive Secretary

1. The Executive Secretary is pleased to circulate herewith the document entitled "A review of barriers to the sharing of biodiversity data and information, with recommendations for eliminating them". This document, whose preparation was led by UNEP-WCMC in its capacity of the secretariat of the Friends of the Conservation Commons, is a contribution to the following decisions highlighting the importance of sharing biodiversity data and information for the implementation of the Convention:

(a) In paragraph 2 of decision X/7 on examination of outcome-oriented goals and targets and associated indicators, the Conference of the Parties recognized the need to continue strengthening the ability to monitor biodiversity at all levels, including through "identifying and addressing barriers that limit the availability of data, including through the work of the Conservation Commons";

(b) In paragraph 5 (c) of decision X/15 on scientific and technical cooperation and the clearing-house mechanism, the Conference of the Parties requested the Executive Secretary to "explore, in collaboration with Parties, other Governments, relevant partners and members of the Conservation Commons, ways to promote free and open access to data and information for conservation purposes, and report back on progress at the next meeting of the Conference of the Parties";

(c) In decision X/2 on the Strategic Plan for Biodiversity 2011-2020, Aichi Biodiversity Target 19 states that "by 2020, knowledge, the science base and technologies relating to biodiversity, its values, functioning, status and trends, and the consequences of its loss, are improved, widely shared and transferred, and applied".

2. A draft of this paper was previously circulated for comments by Parties through notification SCBD/ITS/ODM/79013 (2012-030) issued on 1 February 2012. Feedback received was taken into account when preparing document UNEP/CBD/WG-RI/4/INF/13 which was submitted to the fourth meeting of the Ad Hoc Open-ended Working Group on Review of Implementation of the Convention. The present version of this document only contains minor updates. It is provided in the form and the languages in which it was received by the Secretariat and does not necessarily represent the views of the Executive Secretary.

/...

A review of barriers to the sharing of biodiversity data and information, with recommendations for eliminating them

First submitted 10 April 2012, updated 17 August 2012

Prepared by the UNEP World Conservation Monitoring Centre in its capacity as Secretariat of the Friends of the Conservation Commons as an input to planning for the future of the Conservation Commons, and as a contribution to both the Convention on Biological Diversity and the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services

Contents

Executive summary

Discussions in various intergovernmental fora, including both the tenth Conference of the Parties to the Convention on Biological Diversity and various meetings concerning establishment of the intergovernmental science-policy platform on biodiversity and ecosystem services, have recognised the importance of using data and information effectively in decision making processes at all levels, and the need to promote increased access to relevant data, information and knowledge.

This paper presents an overview of the major barriers to sharing data, information and knowledge relevant to the conservation and sustainable use of biodiversity, presenting this in the context of a number of initiatives already being taken to address these barriers. These barriers include:

a) ***Psychological and behavioural barriers***, which range from unwillingness to share for commercial reasons to barriers resulting from concerns over how data, information and knowledge might be used, as well as including legal barriers.

b) ***Barriers relating to describing information and data***, which range from lack of widely agreed classification systems for some types of biodiversity, to insufficient use of contextual and explanatory information linked to datasets.

c) ***Practical barriers***, which range from knowing how to make data, information and knowledge you hold available in meaningful ways, to locating the information that you need from amongst the plethora of data, information and knowledge available.

d) ***Inadequate strategies and resources*** that result in data, information and knowledge often being made available in an opportunistic manner rather than being focused on need, often without sufficient resources being made available.

While this review is not comprehensive or exhaustive, it is sufficient to indicate a wide range of concerns, and some of the ways in which these might be addressed in order to promote and facilitate increased access to data, information and knowledge relating to biodiversity and ecosystem services. The following recommended actions aim to address these concerns, but at the same time they also serve to illustrate that there is no one single solution to increasing access to data, information and knowledge, but a range of steps to be taken by a range of stakeholders.

Actions focused on governments

Action 1 - Further develop national and international policies for sharing data, information and knowledge, and facilitate their implementation.

Action 2 - Encourage continuation of funding for information resources and maintenance of infrastructures.

Action 3 - Consider making research permissions dependent on those applying providing access to any data and information generated.

Actions focused on donors and others providing support

Action 4 - Ensure that each funded project provides appropriate access to the data, information and knowledge that it generates.

Action 5 - Promote longer-term investment in maintenance of data and information resources, and funding of knowledge management and dissemination.

Actions focused on publishers and publications

Action 6 - Increase open access to publications and reports, including archives.

Action 7 - Facilitate access to data and information on which publications are based.

Action 8 - Support and encourage the publication of 'data papers' that describe available datasets.

Actions focused on academics and research institutions

Action 9 - Create and promote incentives for academics and research institutions to increase access to data and information.

Actions focused on 'knowledge brokers'

Action 10 - Establish and maintain easy-to-use electronic information infrastructures or repositories for ensuring long-term access to data, information and knowledge.

Action 11 - Develop and promote means for improving access to the 'long tail' of results from small and medium-sized research projects.

Action 12 - Review and further promote the use of common vocabularies, classification systems, and standards.

Action 13 - Create better road maps to useful knowledge.

Potential future role of the Conservation Commons

Action 14 - Signatories to the Principles of the Conservation Commons should set an example in increasing access to data, information and knowledge that they hold.

Action 15 - Improve communication of the value of sharing data, information and knowledge, and the potential costs of not doing so.

Action 16 - Help to ensure that those willing to increase access to data, information and knowledge have access to guidance on how to do so.

1 Introduction

1.1 Context for this review

During 2010, two intergovernmental processes relating to the conservation and sustainable use of biodiversity concluded lengthy discussions. The Conference of the Parties (COP) to the Convention on Biological Diversity (CBD) adopted a new strategic plan for 2011-20 incorporating the Aichi Biodiversity Targets, and a new protocol on access and benefit sharing, as well as reaching agreement on a number of other key issues. Meanwhile, the third intergovernmental and multi-stakeholder meeting on an intergovernmental science-policy platform on biodiversity and ecosystem services (IPBES) agreed that the new platform should be established, and identified a number of characteristics of the future platform.

Discussions in both fora recognised the importance of using data and information effectively in decision making processes at all levels, with interventions from governments addressing issues such as data access, interoperability, and the adequacy of data for addressing particular issues. Essentially governments were making clear the importance of ensuring the effective use of existing data, information and knowledge.

During the IPBES intergovernmental and multi-stakeholder meeting in October 2009, the Group of Latin American and Caribbean countries submitted an information document identifying key capacity building needs.[1] Prominent amongst these was the need to "*consolidate and expand, in partnership with existing initiatives, access to biodiversity and ecosystem service information*", identifying a number of mechanisms by which this could be achieved.

During the International Expert Meeting on IPBES and Capacity Building convened by the Norwegian and Brazilian Governments in May 2011,[2] further prominence was given to the need to improve access to data, information and knowledge that already exist. It was recognised that many of the barriers to data access were surmountable with appropriate will and support, and that IPBES could give added impetus to both ongoing and new work aiming to address these barriers.

During the CBD Conference of the Parties in October 2010:[3]

- In decision X/7 on examination of outcome-oriented goals and targets and associated indicators, the CBD COP recognised the need to continue strengthening the ability to monitor biodiversity at all levels, including through "*identifying and addressing barriers that limit the availability of data, including through the work of the Conservation Commons*".
- In decision X/15 on scientific and technical cooperation, the CBD COP requested the CBD Executive Secretary to "*explore, in collaboration with Parties, other Governments, relevant partners and members of the Conservation Commons, ways to promote free and open access to data and information for conservation purposes, and report back on progress at the next meeting of the Conference of the Parties*".
- Finally, in the Strategic Plan for Biodiversity 2011-2020 (decision X/2) target 19 states that "*by 2020, knowledge, the science base and technologies relating to biodiversity, its values, functioning, status and trends, and the consequences of its loss, are improved, widely shared and transferred, and applied* ".

It is intended that the present analysis inform both the CBD and IPBES processes on the main barriers to the sharing of biodiversity data and information, and how these barriers might be further addressed.

1.2 Role of the Conservation Commons

The Conservation Commons was created at the 3rd IUCN World Conservation Congress in 2004, with the aim of encouraging the release of biodiversity data in order to facilitate biodiversity conservation. It established three fundamental Principles which relevant organizations are invited to endorse by formal signature. The Principles of the Conservation Commons are:

> *Open Access: The Conservation Commons promotes free and open access to data, information and knowledge for conservation purposes.*
>
> *Mutual Benefit: The Conservation Commons welcomes and encourages participants both to use resources and to contribute data, information and knowledge.*
>
> *Rights and Responsibilities: Contributors to the Conservation Commons have full right to attribution for any uses of their data, information, or knowledge, and the right to ensure that the original integrity of their contribution to the Commons is preserved. Users of the Conservation Commons are expected to comply, in good faith, with terms of uses specified by contributors and in accordance with these Principles.*

The Conservation Commons recognises the fundamental contribution which open access to data and information can provide to bridging the science-policy divide. It is an association of like-minded organisations and individuals, a 'community of practice' which supports and promotes the Principles and Objectives of the Conservation Commons. To date, the Conservation Commons has focused primarily on raising awareness around the issue of open access to data and information of importance to the conservation community.

The ultimate aim of the Conservation Commons is to ensure improved action for the conservation and sustainable use of biological resources. The presumption is that good data, information or knowledge in the right hands will result in good decisions, and hence the achievement of this aim. A lack of good data, information and knowledge may arise for a number of reasons:

- the data, information or knowledge do not exist
- they exist (or have existed) but are not accessible to the potential user
- they are accessible but not in a form that the potential user finds useful
- they are accessible and in a potentially useful form, but the potential user does not know of, and/or cannot discover, their existence

The challenge for the Conservation Commons is therefore twofold:

a) identifying how to get institutions and organizations to provide and encourage access to relevant and reliable data, information and knowledge; and

b) facilitating the broader use of the resulting data, information and knowledge by encouraging the development and use of tools, mechanisms and processes that promote its discovery and access

1.3 Origin of the paper

This paper has been prepared as an input to planning for the future of the Conservation Commons, and as a contribution to the CBD and the future IPBES. It will also be relevant to a range of other data-related initiatives, including Global Earth Observation System of Systems, and to other assessment processes. The paper was prepared by the UNEP World Conservation Monitoring Centre in its capacity as the Secretariat of the Friends of the Conservation Commons, with the support of staff from the secretariat of the Global Biodiversity Information Facility. It has been made available

for review by organizations and individuals involved with the Conservation Commons, and to CBD Parties through CBD Notification 1012-030.

2 Scope: what are we talking about?

2.1 Defining data, information and knowledge

The principles of the Conservation Commons refer to 'data, information and knowledge', these three terms being taken to reflect somewhat different aspects of ways that we can understand the world.

a) ***Data*** can be thought of as sets of information points, often quantitative or numerical, and usually coded in a systematic way. In a scientific context, raw data are generally measurements from experiments or observations.

b) ***Information*** has a much looser meaning but may generally be thought of as implying descriptions of the world intelligible to others but not necessarily embodying a powerful conceptual or analytic framework.

c) ***Knowledge*** is taken to entail an understanding of processes, concepts and contexts, and as a result may be specific to particular user communities who are familiar with those processes, concepts and contexts.

We tend to think of these three categories as representing discrete steps from observation to understanding, often entailing transformation from the numerical and quantitative to the descriptive, to the interpretive. For example, data might be in the form of a polygonal 'shape file' representing protected area boundaries, information might be a wide-ranging description of the area, including lists of species, description of vegetation and so forth, while knowledge might be an analysis of the importance of that area for biodiversity and of the relative economic costs and benefits of maintaining it. We also tend to think of knowledge as more value-laden and potentially contentious, with people being much more likely to agree on where the boundaries of a protected area are, or which bird species occur there, than they are on the value of the area.

In reality, data, information and knowledge cannot be easily separated, practically or conceptually. Data cannot exist without a conceptual framework, as a sheet of numbers is just a sheet of numbers unless we have some understanding of what those numbers refer to. In the same way, any kind of information embodies some understanding or conceptual framework about the world, which is a more or less implicit body of knowledge. This is important to recognise because we often make the erroneous assumption that other people (potential users) share our understanding, and that data (in particular) can speak for themselves.

2.2 Kinds of knowledge

The principles of the Conservation Commons refer to "*data, information and knowledge for conservation purposes*". It could be assumed that this refers essentially or primarily to data, information and knowledge about biodiversity, but this is not necessarily the case. Effective conservation requires access to many other kinds of data, information and knowledge, and in particular that relating to human behaviour. Knowledge relevant to conservation can be about:

- non-human organisms at genetic, individual, population and community levels
- the physical environment and processes
- humans and human constructs
- the interactions between the above

While reference is only made to "conservation", it is assumed here to embrace data, information and knowledge relevant to all three objectives of the Convention on Biological Diversity: the conservation of biological diversity; the sustainable use of its components; and the fair and equitable sharing of the benefits arising out of the utilization of genetic resources.

With respect to this issue, it should also be borne in mind that our knowledge about the natural world comes from information gathered in the light of - and refracted through - a theoretical framework which itself changes over time in the light of new observations and new thinking. Understanding changes over time, bringing with it the need to consider new areas of information and knowledge, and the need for additional data and types of data.

2.3 Capturing and transforming data, information and knowledge

It is important to recognise the wide range of potential formats, as these often have different use constraints and legal protections. These might range from observation records to song recordings, and from photographs to sets of metrics. Data and information come from multiple sources, but they are essentially based on four major categories:

- observation, measurement or collection of natural phenomena *in-situ* (in the field)
- observation or measurement in *ex-situ* conditions (laboratories, zoos, herbaria etc)
- remote sensing
- observation or measurement of human interventions and constructs

The results of observation and measurement are subject to different kinds of transformation, they are analysed and (in classical understanding of the empirical scientific method) used to test and refine various hypotheses about the world.

Increasingly, hypotheses and representations of the world, particularly those concerning ecological and earth-systems processes, are derived from models and scenarios. There are two main reasons for this. Firstly enormous advances in computing power in the past few decades have meant that it is relatively easy to construct models of complex processes, while it is hard (and sometimes impossible) to study these processes as a whole in real life, because of the spatial and temporal scales involved and their sheer complexity. Secondly, models can have considerable explanatory power and can be persuasive communications tools. The results gained from models and scenarios are certainly a form of knowledge, but it is important not to overlook how dependent they are on baselines and assumptions.

Capture of knowledge is an even broader concept, potentially including: publications; records of interviews; sharing of case studies, best practice and lessons learnt; monitoring impact of decision made; and so on.

2.4 Why knowledge is held, and how it can be conveyed to others

Data, information and knowledge are held by many different kinds of people (academics, managers, amateurs) in different capacities and for different purposes which may include:

- personal interest
- research
- resource management
- because of a mandate or request

- as a public good
- as a by-product of other processes

The boundaries between these may be blurred, so that a person or organization may be gathering data, processing information and generating knowledge for a variety of reasons, and sharing all three with other individuals or organisations who are using them for a different purpose.

Data, information and knowledge can be conveyed – communicated – to others through each of the following mechanisms, although the predominant focus these days is on the last of these:

- through person-to-person communication, chiefly orally
- through print and other physical media
- electronically

It is important to realise that knowledge exists in many different forms - as words (in many different languages), numbers, symbols, images or artefacts – and can be stored in different places: in people's heads, on paper, in specimen cabinets and in various electronic forms - hard drives, CDs, memory sticks, websites and so forth. Meanwhile the dominance of different forms of storage has changed over time as new means evolve, and as practices change.

2.5 Existing levels of access to knowledge

Finally, it is important to recognise the range of levels of access to data, information and knowledge, as these are potential factors in understanding and overcoming barriers to access. Data, information and knowledge may in theory be available to:

- the person or people who originated or collected it
- their immediate colleagues or peers
- more widely within a particular organisation or institution
- those outside the holding institution or organisation under potentially variable restrictions (license agreements, fees, conditions of use etc)
- freely, to anybody

The extent to which it is actually accessible (and useful) to anyone other than its originator depends on a range of factors, including:

- whether a potential user knows that it actually exists
- whether it is in a form, and at a place, that they can access it
- whether it is incorporated into tools that increase its accessibility and value
- whether the potential user fulfils any conditions placed on access by the originator

2.6 Barriers to access, and barriers to use

This paper is intended to address those barriers which reduce access to data, information and knowledge relevant to the conservation and sustainable use of biodiversity and ecosystem services. In this regard the paper addresses both barriers to making the data, information and knowledge available, and those barriers with might make it difficult to locate and access. What the paper does not address is why data, information and knowledge are not used where they are both available and accessible.

3 Barriers to sharing of biodiversity information and data

This assessment was based on a desktop scan of current experience, lessons learned and recent initiatives, and was augmented through peer review. The barriers identified here are not necessarily new,[4,5,6] but it is hoped that by building on this assessment with an understanding of some of the initiatives that are already underway to address known barriers, we can then identify opportunities for action to address these barriers further. In this respect the review has not been exhaustive or comprehensive, but we believe it has been sufficient to identify future direction and needs.

Impediments or barriers to the free interchange of data, information and knowledge can be loosely categorised as:

- psychological and behavioural barriers (including legal barriers)
- barriers relating to describing information and data
- practical barriers
- inadequate strategies and resources

In reality, none of these impediments or barriers can be looked at in isolation, as the extent to which some specific data, information or knowledge is accessible to a given person is the result of a series of decisions and actions by one or more individuals or organisations. All the above factors are likely to have had some influence in that decision.

Impediments will also be, or appear, different to different constituencies. For a potential user, as noted above, a major initial impediment may be ignorance of the existence of something useful. For a potential provider, it may be the absence of (or ignorance of) an effective way of letting other people know that they have something that might be useful. Meanwhile, barriers to effective data sharing and preservation can be deeply rooted in the practices and culture of a particular constituency such as, for example, the research community.[7]

There are, of course, many existing initiatives concerned with addressing barriers to accessing data, information and knowledge, and it is important that any new initiatives to address barriers ensure coordination and collaboration with existing activities, supporting and augmenting what is already being done.. For example, the Sistema de Información sobre Biodiversidad (SiB) in Colombia,[8] the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) in Mexico,[9] the South African National Biodiversity Institute (SANBI),[10] the Norwegian Biodiversity Information Centre,[11] the National Biodiversity Network in the United Kingdom,[12] the *species*Link network in Brazil,[13] and the Atlas of Living Australia[14] all demonstrate what can be achieved at the national level with respect to improving access to data. Meanwhile, the Global Biodiversity Information Facility (GBIF) was established in 2001 with the explicit objective of enabling free and open access to biodiversity data online, and has been working very actively in this area since. In the paragraphs that follow, GBIF is frequently used as an example because of its nature and mandate, and because of its practical experience in addressing many of the barriers referred to.

3.1 Psychological and behavioural barriers

Organizations and individuals may be reluctant to share knowledge or disclose data and information relevant to biodiversity for a number of reasons, generally to do with concern that someone else will benefit or profit 'unfairly', or will misuse it in some other way. As a result, they may control access to the data, information and knowledge that they hold in various ways. In particular they may:

- refuse to share it at all
- share it once they have extracted maximum value from it themselves

- share it under various restrictions (e.g. providing it only on request, for non-commercial use only, on condition that they are involved in its further use etc)
- charge for accessing/using it and, usually, apply use restrictions

The attitude of any one individual or organisation to the sharing of information is likely to be complicated, often not fully thought through, potentially inconsistent and sometimes contradictory. In fact, many holders of information are in principle prepared to share their information, either freely or universally, or under certain caveats or restrictions. This may be for one of a number of reasons:

- because they think it will contribute to a public good
- because they receive acclaim or praise for doing so
- because they are mandated to do so
- because they are paid to do so

However, putting this willingness into practice is never entirely effort-free. Generally, of course, the less effort that is required then the smaller the incentive has to be. Conversely, the more difficulty entailed, the greater the incentive needed to act. Information technology has in fact not undermined the notion that data, information and knowledge are intrinsically valuable – indeed it has if anything strengthened it - but it has created great problems in how to protect or realise any value it might have. It has also exacerbated the related, but different problem, of how to recover the costs of knowledge management, from origination to dissemination.

A further issue of potential concern, but the effect of which has not really been assessed, is that for many of the larger institutions and organizations which could share data the internal communications department has significant control over what can appear on the institutional website and hence, often, how data get exposed to the web. Particularly for images, etc., the communications department may see data resources as part of the branding and profile for the institution and may resist their wider use.

Many of the psychological and behavioural barriers to sharing data, information and knowledge can be addressed at least in part by increasing understanding of its value when shared more widely, and by simultaneously trying to address some of the more practical concerns. Some of the ways in which this is currently being done are addressed further below. However, many of these advances are still relatively new, and there is much still to be done to build on the existing initiatives in order to increase access to data, information and knowledge that is either not being released, or is being released in a manner which restricts access in some manner.

Scientific publication, and academic and research careers: For academics and researchers, knowledge is the fundamental currency of their livelihoods. They advance in their careers by disseminating it, usually through teaching and publication, with the latter having precedence in career terms over the former. Sharing knowledge resources, without at first having extracted as much personal value as possible from it (usually through publications) can be seen as unproductive, and may even be counterproductive in career terms.

However, there are signs that this situation is gradually changing. A recent article in the Wall Street Journal on "The new Einsteins will be scientists who share"[15] discussed the benefits of open sharing of knowledge, drawing on the benefits of networked science experienced in a number of areas. These include both approaches for wider interaction and sharing such as the online academic social network Mendeley,[16] and tools for increasing the value of content such as the PLoS Biodiversity Hub.[17,18] These efforts are reinforced by the findings of a recent Royal Society review of *Science as an*

open enterprise, which recognised the importance of collaboration, and the benefits of open data policies.[19]

There is certainly an increased interest in making primary data from published research available more widely, although there is still some way to go in achieving this consistently.[20] In addition it has been argued that data publishing should be recognised on a par with scholarly publishing in peer-reviewed journals, and the case has also been made for academic journals to make it mandatory for the data on which studies are based to be made accessible.[21] This is increasingly happening, particularly with online journals. To help address some of the barriers to data publishing, a task group was set up by GBIF to develop a data publishing framework. They made recommendations for overcoming a range of barriers or impediments affecting the discovery and publishing of biodiversity data,[22] three of which concern incentive mechanisms for addressing the behavioural barriers identified here:

a) Data Paper: A new mechanism is being developed to enable article manuscripts to be generated automatically from enriched metadata documents. The first 'data paper', describing a dataset of Indian birds compiled from a century of literature references, was accepted for publication in November 2011 following peer review.[23,24,25,26]

b) Data Citation: In close consultation with professional societies and expert groups (including the CODATA Data Citation Standards and Practices Task Group[27] and DataCite[28]), a style for data citation and a deep data citation service has been proposed, with the intention of creating a system of data citation that gives visibility to data publishers and contributors, as well as showing characteristics of the data and how to access it.

c) Data Usage Index: Algorithms for a Data Usage Index (DUI) have been developed, and GBIF intends to commission a prototype in the near future.[29,30] The aim of this is to provide incentives through recognition of the efforts required for publishing biodiversity data.

Commercial barriers to accessing the scientific literature: The publishing of journals has been a commercial exercise for many years, but with the explosion in the number of journals, and their increasing cost, it has increasingly become for many a barrier to accessing published knowledge, rather than the means of wide dissemination it was originally intended to be. This is the case even through the journals are often now available online. Various initiatives are focused on increasing access to the published literature. For example OARE (Online Access to Research in the Environment) provides online access to around 3,000 peer reviewed titles to registered institutions in around 100 low income countries,[31] the Scientific Electronic Library Online (SciELO) based in Brazil, but involving many countries in Latin America and beyond, similarly provides access to the published literature,[32] and the Biodiversity Heritage Library[33] provides access to more than 34 million digitized pages in over 47,000 titles, many of which were not previously digitized. Meanwhile, a range of online journals have become established in recent years (several of which are referenced in this paper), and many of the more traditional journals have now posted their archival copies (or a considerable proportion of them) online, although there is often still a cost involved in accessing them.

Sensitivity: Data, information and knowledge may also be withheld because those in possession of them think that risks will be incurred if they share them with others. Within the biodiversity conservation community, it is quite common practice to restrict access to some forms of sensitive information, and in particular information relating to locations of breeding sites of rare birds or of populations of 'covetable' orchids. In order to address these sorts of issues, best practices in generalising such data were investigated by GBIF, leading to development of a 'best practice' guide for generalising occurrence data [34] which provides a step-by-step process for: determining sensitivity; categorising sensitive data; generalising textual and spatial information; authentication and authorisation of potential users. The Atlas of Living Australia commissioned a review of how

difficulties associated with sensitive data act as a barrier to data sharing, and subsequently implemented a 'Sensitive Data Service' infrastructure to try and address some of these problem.[35] Using these sorts of approaches, it is possible to address sensitivity issues in a consistent and acceptable manner for species occurrence data at least, so as to be able to share such sensitive data through information infrastructures.

Legal barriers and intellectual property regime: The main legal constraints on the free flow of data, information and knowledge are laws and regulations protecting copyright and other forms of intellectual property right – that is laws that support those who choose not to share information, for one reason or another. In fact the relationship between intellectual property laws and electronic media is rapidly changing and fiercely contested ground. Different legislatures are adopting different approaches, so that what is legal in one jurisdiction at any one time in terms of file-sharing, for example, may be illegal in another. It is therefore urgent to develop copyright legislation that keeps pace with technological change. Additionally, as many countries follow the Berne Convention for the Protection of Literary and Artistic Works,[36] and national law within the Member States of the European Union have to respect the Directive on harmonization of certain aspects of copyright and related rights in the information society,[37] it may be necessary to also consider revising these legal instruments in order to include exemptions for the transfer of protected works for certain research and scientific purposes. Clear international law in this area would be helpful, so that the lack of a stable, uniform global regime on intellectual property does not impact on willingness to share such property electronically. Meanwhile communities of practice are growing, and currently Creative Commons licenses are the most frequently used legal tool enabling sharing of created work,[38] with Wikipedia probably being the largest and most notable internet-based user of Creative Commons licences, but with many other users including the Encyclopedia of Life,[39] the Atlas of Living Australia and the Sistema de Información sobre Biodiversidad in Colombia referred to earlier.

Perceived commercial value: Data, information and knowledge may have direct commercial value. For example, the location of stocks of valuable living resources, such as exploitable fish species, is clearly of commercial importance to those that hold it. It may also be thought to have potential value. In a conservation context the assumed potential value of genetic resources has considerable bearing on the willingness of governments to share knowledge of those resources without some form of compensation, notably through some agreed intergovernmental access and benefit-sharing regime. Knowledge, and potentially also data and information, may also be of value because they provide opportunity for consultancies and contracts when an organization has unique assets or holdings.

Privacy, cultural and tradition as barriers: Reluctance to share data, information and/or knowledge may stem from a fear that they might be misinterpreted, or used or altered in ways with which the holder does not agree. For example, the invasion of privacy and potential abuse of personal information is a growing concern for many people in their interactions with the Internet, and there are somewhat similar concerns amongst indigenous peoples and traditional communities about the ways in which their knowledge might be used. If individuals, societies and organizations do not have complete confidence in security controls in any system they are considering contributing to, or they do not agree with the policies of that system in these areas, or they are concerned about the ways in which their contributions are likely to be used, then they are unlikely to contribute.

Possible 'side effects' of communities of practice: As has noted during the workshop series on data sharing and archiving hosted by the Ecological Society of America,[40] the first level of cooperation for facilitating sharing of information is a 'community of practice'. In the international conservation world, the botanical taxonomic community, centred on herbaria, botanic gardens and natural history museums, was one of the first to organise itself around the subject of computer databases as a mechanism for information sharing, establishing the Taxonomic Databases Working Group, now

TDWG – Biodiversity Information Standards, in 1985.[41] Such communities can be very successful at developing their own ways of sharing information amongst themselves but, paradoxically, this very success, by reinforcing a sense of community, can serve to exclude others unless addressed carefully. Potential reasons for this include use of different sets of terms, or different approaches to addressing problems and challenges, and so on. Successful sharing of biodiversity data, information and knowledge requires engagement from a wide range of individuals in different communities – requiring connections and cooperation across potentially wide gaps in understanding, perspective and behaviour.

3.2 Barriers relating to describing information and data

The transmission of data, information and knowledge is not a transparent or unmediated process. It rests on a series of assumptions about shared understanding of words and concepts. The wider the intended community of users of some piece of data, information or knowledge, the less likely it is that these assumptions will be fully met, and the greater the need to work to ensure that understanding.

Languages: The use of many different languages worldwide is the most obvious barrier to universal understanding. English is now the *lingua franca* of international discourse, and dominates the World Wide Web. Its prevalence has led many to assume that there is no longer a real language problem, but this is far from the case. Information in other languages has limited penetration, while people with no English or poor understanding of English have limited access to information. These are issues that have been flagged as being of significant concern in a number of the IPBES-related discussions with respect to access to data and information.

Accurate and reliable translation and interpretation incurs considerable costs and is undoubtedly underused in the conservation community, except where mandated (essentially within the UN system). On the other hand, automated translation programmes are becoming increasingly accurate, and at least provide an opportunity to gain some understanding of written materials available in electronic format. In addition, controlled vocabularies and multilingual thesauri are increasingly supporting discovery of relevant data, information and knowledge in other languages where it is available (see below).

Language is clearly an issue for any international initiative that aims to increase data access. For example, GBIF has needed to address the issue of language in developing its information infrastructure, and has taken a number of steps to facilitate publishing and discovery of data in multiple languages, including within the Integrated Publishing Toolkit (IPT)[42] and the GBIF Metadata Profile.[43] The use of the Dublin Core language declaration is also recommended in the Darwin Core specification[44,45] which is in widespread use within the GBIF network and elsewhere.

Standard vocabularies and multilingual thesauri: Agreement on use of terms is needed both for categorizing and organising information, and so that different people have confidence that they are talking about the same thing. Controlled vocabularies and multilingual thesauri have the potential both to support increased data integration through an improved common understanding, and to support improved search and browse functions. For example, controlled vocabularies provide optional defined lists of values for some of the Darwin Core and extension terms already discussed, and GBIF makes such vocabularies available through the GBIF Resource Browser[46] which enables search and browse through all terms and concepts served by the repository. Controlled vocabularies also support the search and browse functions of the web-based information service being developed to integrated search across the information resources of many of the Multilateral Environmental Agreements (InforMEA),[47] and other thesauri and controlled vocabularies are increasingly used

within different communities and sectors in order to aid understanding. Other widely used multilingual thesauri include GEMET,[48] AGROVOC,[49] and the USGS Biocomplexity Thesaurus.[50]

GBIF is a good example of a global network serving sub-networks that operate in many different languages, and multilingual thesauri are therefore an essential component of their knowledge organization systems. The creation and translation of vocabularies is achieved through the GBIF vocabulary server[51] that provides an editorial environment for drafting and publishing vocabularies through a set of web services. Terms used in vocabularies may be created *de novo* within the tool or reference an external defined term. GBIF provides an online bionomenclatural glossary[52] of over 2500 terms which may be assembled into controlled vocabularies.

In fact, as a result of the efforts of GBIF in particular, data on species occurrence is an example of an area where significant work has been done on supporting standardized use of terms and concepts. The GBIF network relies upon an array of standard concepts, vocabularies and taxonomies for enabling integration of species occurrence data from over 8,000 databases. At its core is the Darwin Core set of terms,[53] which combines concepts from the Dublin Core vocabulary with over 160 additional concepts relating to taxa and parameters relating to the occurrence of species in nature. Other supported data structures, such as the Access to Biological Collection Data (ABCD) schema[54] are mapped to the Darwin Core[55] to support integration. GBIF manages the Darwin Core terms, as well as defined extensions to the Darwin Core[56] in a resource repository[57] that provides access to all core concepts in a standard format, enabling them to be embedded within tools and services within the GBIF network infrastructure. GBIF makes available a suite of 'how-to' guides detailing how to prepare and publish data and metadata through the GBIF network which cover many of these issues.[58]

Taxonomies and classification systems: In organizing data and information, some things are widely agreed on and in essence unproblematic, but these are remarkably few: time in the widest sense (i.e. including date and age) and location are perhaps the most important, although in practice the latter is not as straightforward as might seem (geo-referencing of data is quite a complex business[59]) and even the former is not always unambiguous. In geography, the ISO 3166 standard for countries, dependent territories and areas of special geographic interest is very widely but by no means universally used. Almost every other relevant concept remains open to differing interpretations, and it may not be possible to get widespread agreement, let alone consensus, on meanings and taxonomies of terms or categories, because:

a) State of knowledge may be insufficient: The state of knowledge of most of the areas where norms might be valuable changes so that the fixing of norms may risk the loss of scientific integrity. Constant updating, on the other hand, is onerous, particularly if it requires reclassifying or labelling of existing information. The optimum trade-off between stability and keeping up to date will depend on the uses to which the information is intended to be put, so that a consensus is unlikely to be easily reached on exactly where this should lie.[60]

b) People may be unwilling: Even where widespread agreement can be reached, potential suppliers of information may not wish to conform, either because they do not agree scientifically or because the proposals do not fit well with their own information needs.

c) Cost of complying may be too high: Even if willing to agree on meanings and taxonomies, adhering to them may simply be too onerous in terms of time and cost, including time that could be spent working on something else.

This results in the following situation with respect to organisms, habitats or ecosystems, and threats to biodiversity and ecosystem services, for example.

a) Organismal taxonomies: Information on organisms is at the core of much biodiversity information. The classification and naming of organisms is a sophisticated endeavour which in many ways is tailor-made for the organisation of large amounts of knowledge. It is thanks to the existence of the pretty much universally accepted Linnaean system, and of long-standing codes of practice, such as the International Code for Botanical Nomenclature (ICBN) and its zoological equivalent, that so much progress has been made in the development of systems for disseminating information on organisms. For example, the Catalogue of Life[61] holds information on more than 1.4 million species and nearly 850,000 synonyms. However, even organising information at the level of species is a complex and often contentious process, as there is no one universally accepted classification of organisms, nor is one ever likely to be settled on.

b) Habitat and ecosystem terms and systems of classification: It is difficult to get agreement on the meaning of even basic terms that refer to habitats, biomes or ecosystems (forest, wetland) let alone to reach consensus on, say, habitat classification systems or land-use categories. A great deal of energy and time has been expended on the development of many such systems at various scales over the past few decades. There is little sign that any one system in any of these areas will prevail universally, nor reason to think that this would necessarily be a desirable goal (the value of any one system is very much purpose dependent). This difficulty is compounded further in attempts to describe and categorise habitat condition, an area fraught with difficulties of both interpretation and measurement.

c) Taxonomies of threats and impacts: These are even more contentious and uncertain than those of habitats and ecosystems, for a number of reasons. There are many different ways of looking at them ('drivers', 'pressure', 'impacts' all have differing meanings in the minds of different people); it is rare for there to be detailed, agreed understanding of exactly what the factors are that affect different components of biodiversity; and despite best efforts, these measures almost invariably have some component of subjective value built into them, so that adoption of any particular terminology implies some acceptance of the value-system underlying it.

Despite many existing initiatives, there is much still to be done in order to reduce barriers to data access. This may be because no standard vocabularies for a particular subject area exist, but it may also be because standard vocabularies have been developed but not fully adopted, or even as a result of different (essentially competing) standards operating within different communities of use.

Contextual and explanatory information: Some kind of context is invariably crucial to the understanding of any given piece of data or information. Such context is usually referred to as 'metadata', a term that embraces a range of attributes. It may refer to:

- the form in which the data or information is held (whether physical or semantic)
- the attributes or characteristics of the data or information
- ways in which the data or information have been collected and potentially transformed
- wider information necessary or helpful to interpreting the data or information itself

In all forms of knowledge management the first two kinds of metadata attributes tend to be quite well characterised, with more or less widely accepted standards, notably the Dublin Core metadata elements[62] and ISO/IEC 11179 Metadata Registry Standard.[63] Similarly, many Latin American countries are using the Colombian SiB 2.0 standard, and the ISO 19139 standard for building metadata of biodiversity-related datasets. Considerable work has also been done in the arena of biodiversity or conservation information in the Darwin Core,[64] and, for multi-media resources, the Audubon Core,[65] both under the auspices of Biodiversity Information Standards (TDWG).[66]

However, the last kind of metadata attribute, the contextual information, is often neglected or poorly addressed. This represents a major barrier to effective sharing of data and information. There are good reasons why this is the case: it is generally onerous, presents an imaginative challenge to

providers of information (people in general tend to assume that what they understand is what everyone else understands) and there is no widely accepted way of going about it. Content providers and information managers or brokers are often interested in providing as much information as possible, and in organising it systematically. This is where their energies tend to be directed. Providing contextual information is seen as a peripheral task, yet without it the value of the data and information can be seriously compromised.

One very common difficulty is finding out what is not in a database, which may be as important to understanding as knowing what is. A database, for example, listing species recorded in a set of protected areas might be used to assess which areas were most important from the point of view of biodiversity, but without a clear statement of how complete each list is, a straightforward ranking would be highly misleading. Often some assessment of this is provided in documentation associated with databases, but equally often it is hidden in a set of technical specifications and can only be found if actively searched for. Many users of such databases will make an assumption that the information is more or less complete, and may well therefore not even search for such information unless there is very clear encouragement to do so.

Metadata standards: There are several metadata standards available for biodiversity data and information. However, information managers and information providers largely do not use them because it takes extra time and effort to add metadata elements and keywords to information and thus make them ready for effective search and sharing. If a minimum set of metadata elements with controlled keywords could be agreed for biodiversity data, information managers would use them and users could find information more effectively.

If a standardized metadata description and controlled keywords were available, it would be possible to build a data and information searching system taking advantage of the semantic web technologies. These technologies include a variety of data interchange formats and notations intended to provide a machine-understandable description of concepts, terms and relationships within a knowledge domain. Ontology based semantic searching is a modern method to raise data searching to a more intelligent level.

The GBIF Task Group on Metadata Implementation Framework looked into these issues and recommended a set of actions to encourage authoring of enriched metadata documents that can adequately describe the data resources published through GBIF network.[67] The GBIF Metadata Profile has been implemented through the recently-released second version of the Integrated Publishing Toolkit , and the GBIF Metadata catalogue.[68] To facilitate uptake of this infrastructure, GBIF has also released a set of best practice guides.[69]

Metadata and discoverability: It is perhaps worth noting that over time the publishing of metadata has been a significant step in increasing the discoverability of data, and hence the potential for accessing it. In this regard, the sharing of metadata can take place without the sharing of the data itself. While this is not ideal, where there are barriers to making the data themselves publically available the sharing of metadata can at least allow for potentially relevant datasets to be discovered, which then makes it possible for them to be asked for – you do not even know to ask for them if you do not know that they exist. Ideally the sharing of metadata and data take place together, but even just making the metadata alone accessible is a useful step forward.

3.3 Practical barriers

Information storage: Explosive growth in the generation of all kinds of information, particularly electronic information, is creating problems of storage for many organisations. The problems of

storage of non-electronic forms of information are generally more evident to specialists and non-specialists alike: the physical space required for storage; the conditions necessary to prevent deterioration; the desirability of some kind of organising principle in its arrangement so that in theory at least it can be accessed systematically. All this has been evident since the development of physical record-keeping millennia ago. The problems of storage of electronic information are in many ways less clear, at least to non-specialists. The widespread perception among the latter that there is no real cost involved in storing electronic information means that more and more of it accumulates, often in more or less disorganised form, stretching the storage capacity of many organisations to the limit and beyond. There is also a widespread view that electronic information is somehow permanent and unlikely to be lost; perhaps reinforced by the way in which back-ups are increasingly carried out automatically, removing any sense of responsibility from individual users and making people less careful about custodianship of information.[70]

There are two main responses to this: in-house archiving and out-sourcing of storage. In-house archiving has the potential to remove information from ready access, both within the organisation and outside it. Out-sourcing of storage, which is increasing rapidly, may present longer-term issues of security and ownership. More generally, the electronic information age is too young for us to understand fully the long- or even medium-term fate of much electronic information. Given the ease with which electronic data can be over-written, the difficulty of keeping track of much of it and the fact that some forms of storage may be less stable than generally thought, there is at least a possibility that much of it will prove more ephemeral than we think and less durable than some forms of non-electronic information.

Recognising that much biological data is at risk of either being lost or orphaned, organizations such as GBIF have been promoting the concept of 'Data Hosting Centres' and other electronic information infrastructures. A recent study suggests that use of data standards, public domain licensing, establishment of data hosting and archival services, and tools for data discovery are essential to discover, rescue and archive data at risk,[71] and several data hosting centres have already been established for species occurrence data. The ICSU Committee on Data for Science and Technology (CODATA) also has a Data at Risk Task Group[72] addressing these sorts of issues.

Non-digital data: While it is true that new data often become available in digital form, there are still large quantities of data that are difficult to access because they are only available in non-digital format, or information that is in digital forms such as non-searchable PDFs. Examples range from the very large number of biological specimens held in collections globally, to reports and publications that have never been digitized. There is a real danger of losing these valuable resources because of basic issues like people moving post, or as a result of accidents such as fire or flooding. A Content Needs Assessment Task Group set up by GBIF has made several recommendations aimed at bridging the data gaps through digitisation of non-digital data. One of the significant recommendations is to undertake frequent local to global scale data gap analysis exercises. As an early uptake of these recommendations, GBIF undertook a data gap analysis exercise and made a set of recommendations to address the issue of digitisation of natural history collections data.[73] Meanwhile, the Biodiversity Heritage Library (see below) has made a substantial effort to increase access to reports and publications that had not previously being digitized or readily available, and now provides access to more than 34 million digitized pages in over 47,000 titles.

Finding information: For data, information and knowledge to be accessible to potential users, they need to know that it is available, and they need to be able to home in on the particular pieces that they are interested in. This requires a combination of effective searching, good organisation of the material, and good communication of its availability.

Web-based search engines, which can be very powerful, are also often frustrating when looking for technical or scientific information. The algorithms they use are not based on quality or primacy, and as a rule they do not reveal information stored in databases online or do so unpredictably. These are often key information sources. Perhaps even more significant is the inability of web-based search engines to access systematically those thematic and geographic networks based on integrating data from distributed data providers, or to access in a systematic way the many local and national data providers/systems that underpin global databases and information sources.

Increasingly, Wikipedia serves as a portal to a number of key databases, with standardised tags for, for example, IUCN threatened species and protected area categories. However, the nature of Wikipedia means that linkage is not systematic, nor is coverage complete, although it is growing rapidly. The main concept and objective behind Wiki is "quick collaboration on the Web". The user becomes a correspondent for sharing her/his own information and expertise. Wiki information is useful in collaboration, but because of many contributors this information can be unstructured and subject to frequent changes. However, the use of wiki style approaches on web-based information services is increasing, and becoming more organized, and this is therefore becoming a more significant tool in accessing biodiversity data, information and knowledge.

Organisation of information is a principal focus of the work of managers of electronic and other data. It is not surprising that this is often a strength of web-accessed databases and other systematic sources of information. Once people know of the existence of such a source, and if they know exactly what they are looking for, and are familiar with the database in question, they can usually extract what they want in a clear, well-organised format. However, many of these databases are not particularly easy for novices to make use of, having, for example, inflexible or apparently weak search functions (typically in not recognising variant names or slight misspellings, something which commercial search engines such as Google have made considerable strides in), or rather opaque ways of creating subsets, or inflexible outputs. They are very often lacking in adequate contextual and explanatory information, as discussed above.[74]

The GBIF Registry is a core component of a ‘global information infrastructure’, which holds information about data publishers and datasets and plays a significant role in discovery of primary biodiversity data resources. [75] A stakeholder workshop held in 2009 envisaged that this might evolve into a more widely shared system, facilitating global discovery of all biodiversity information resources, both digital and non-digital.[76,77] While the work of GBIF is significant, there is a significant number of other catalogues, portals and web-based information services that aim to increase access to data, information and knowledge, including helping potential users to find what they need. These tools are improving with time and experience, and the evolution of the technology.

Data integration: There has been great progress in this field, with the increased use of common standards and protocols, and the development of new web services and new applications. To a significant extent the emerging practices are less expensive, less time consuming, and are increasingly able to ensure control over the data is maintained by the data provider rather than the data integrator. However, data integration remains a significant challenge that can cause bottlenecks in the process, as there are still significant amounts of data that are not integrated in useful ways, and a number of significant initiatives working on data integration remain under funded. There are potential concerns that the ‘control’ of data integration is in the hands of the integrator rather than driven by the decision maker, but this is a matter of working to ensure improved practice in the future, with products and services based on the needs of decision makers. Our ability to create advanced visualization, analysis and modelling from multiple sources of data is mature, but there is as yet no effective medium to share our transformations. Further the

preservation of original and transformed data, information, and knowledge is irregular and often done by integrator not at the source.

Information transfer: Electronic transfer of information is the bread-and-butter of the Internet. A vast amount of work has been carried out in this area – in terms of data exchange protocols, metadata formats and standards. These are both generic – applying to all kinds of electronic information, such as the Z39.50 client-server protocol for searching for and retrieving information from remote computer databases, or may be more specific to the environmental or biological sciences, such as the Biodiversity Information Standards.[78] In order to facilitate seamless and easy exchange/sharing of data (especially primary biodiversity data and its associated metadata) it is important to adopt and properly use common standards and protocols. For example, GBIF promotes a set of standards and protocols through its network including the GBIF metadata profile, [79] the Darwin Core standard,[80] the ABCD data specification,[81] the BioCASe[82] and DiGIR[83] protocols, and the TAPIR standard,[84] and more recently the Darwin Core Archives.[85] In fact, although there are always improvements to be made, technical issues in data and information transfer are no longer the major barriers that they once were, provided the appropriate steps are taken.

Access to technology: It is important to remember that in reality a huge proportion of the world's population still doesn't have access to the internet, or at least not in the way that the rest of the world takes for granted. Emphasis on increasingly sophisticated internet-based solutions to the problem of organizing and communicating knowledge runs the danger of reinforcing this digital divide and may well be leading to the neglect of other ways of communicating knowledge and building capacity. The important issue in this paper is finding ways to promote and facilitate sharing of data, information and knowledge openly and freely primarily through the Internet. However it is important to also recognize the need to promote stable and fast connectivity to all interested stakeholders, to promote communication to communities without internet access (although the latter may well be on a case-by-case approach to determine the best strategy), and to build capacities to use these technologies. There may be longer term problems in information transfer with regard to the capacity of the infrastructure of the internet to go on dealing with the explosive growth of information transfer that takes place using it, and of the possibility of the Internet becoming fractured or no longer serving as a neutral platform for information transfer. There is a real possibility of a multiple-tier Internet developing, with those able and willing to pay more having access to faster information transfer. If this were to become reality it is very probable that most conservation and academic organisations would find themselves in the slow lane.

3.4 Inadequate strategies and resources

Access is also a function of available resources, and the ways in which those resources are used. All aspects of the availability and transfer of data, information and knowledge entail costs of some sort for someone, and this can become a real issue when resources are limited. Meanwhile the lack of strategies for addressing the demands of stakeholder communities for authentic, adequate and up-to-date data in particular can be an important reason for non-availability of 'fit-for-use' biodiversity data. A number of issues are important with regard to access.

Under-estimation of real costs: The costs of managing data and information and making it available to others – and particularly those of maintaining information 'live' – have often been seriously underestimated, leading to the abandonment or slow decay of many database or information management projects or programmes. To this can be added those data and information projects that are started but then come to a halt as there is no sustainable model for their future implementation after the initial project has been completed.

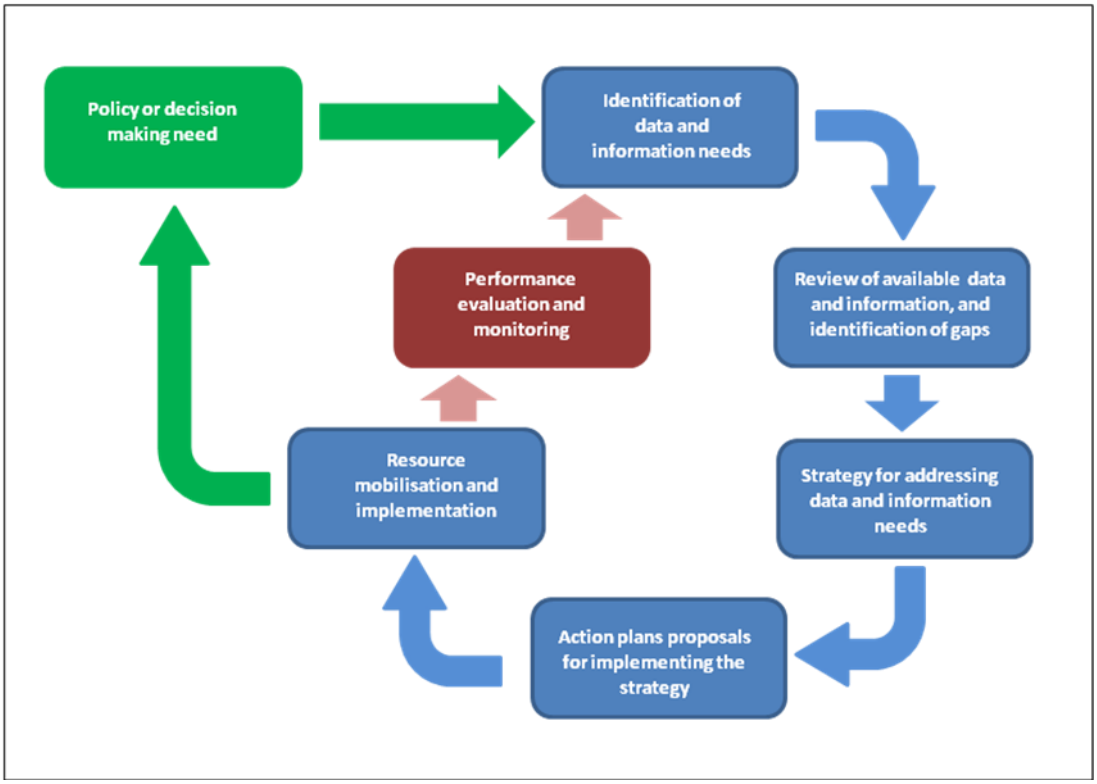
Cost recovery: While it is accepted by many that it is reasonable to recover from users the costs of making data and information available,[86] this can in itself become a barrier to data access for many potential users because they cannot afford to pay. Others may not accept the real cost because their perception of what these costs may be does not match with the reality. Mixed payment models have been explored, in which those that can pay are encouraged or expected to do so, while others have free access (effectively subsidised by the former). Often access may be graded so that those who pay (or who pay more) have privileged access. Cost recovery of this sort is in theory different from paying for the intrinsic value of the data or information, but in practice it is very hard to say where one ends and the other begins. There are two main problems entailed in trying to recover costs effectively, the difficulty of getting people to pay a realistic price (in other words one that pays the real costs of collection, management and delivery), and the practicalities of developing a way of allocating any sums are recovered between different providers that is equitable or acceptable to all those involved. Meanwhile, there is a widely held view that biodiversity data, information and knowledge should be freely and openly available to all interested, without barriers, especially when one considers the long term potential of integrating data through web services. It is probably fair to say that more thought needs to be given to cost recovery models and their implications.

Donor funding: Many organisations that generate data, information and knowledge relevant to conservation rely heavily on donor funding to support their activities. There may be dedicated funding for data and information capture, management and dissemination as stand-alone activities, or alternatively it may be seen as part of a broader activity or set of activities. Whatever the approach, there is potential for data and information management and dissemination to compete with other activities for funding unless they are properly integrated. The problem is exacerbated by the fact that donors are often reluctant to make long-term commitments, and there is a tendency to support development of new infrastructure rather than maintaining existing infrastructures. Meanwhile, there is clearly potential for donors to take steps to promote access to biodiversity data resulting from the projects that they support. In 2006, the GBIF governing board comprising 57 countries and 47 international organisations adopted a statement on open access to biodiversity data that called for those funding proposals to actively promote proposals that included data management and data sharing plans. It further appealed for biodiversity data and metadata generated in funded projects to be made publicly available.[87] The Suwon Declaration[88] further commits to these principles, and pledges to work towards achieving them. A number of funders are already taking steps in this direction, for example, the US National Science Foundation, amongst others, requires all grant applicants to submit data management plans as a part of their applications,[89] and many other public bodies are beginning to expect data to be made available when they are funding research. The UK Medical Research Council not only has a well-developed set of policies related to data access, it also has an active data sharing and preservation initiative covering the substantial amount of research that it funds annually.[90]

Mandatory funding through taxation: The gathering and dissemination of data, information and knowledge may be mandated through legislation implementing international agreements or national policies and may be funded by governments either directly in the case of national activities or through government contributions to international (usually UN) agencies. Government funding of this kind has the great advantage of stability, often generating long-term datasets and other kinds of information gathering and knowledge dissemination activities that can be of immense value. However, governments are often unable to do more than the minimum to meet their obligations, and in the current economic climate this is likely to become worse almost everywhere in the foreseeable future. In becomes increasing important to ensure that adequate feedback is given on the value of the investments being made.

Voluntary contributions: Without doubt, access to a significant amount of data and information results from the willingness and ability of those who generate or hold it to provide it to others. While sometimes people are employed solely for this purpose, very often it forms only part of their paid activities or they are not actually paid to do it all. In recent years wiki and 'citizen science' approaches on the internet have evolved rapidly, drawing on the contributions of professionals and amateurs alike. There are obviously potential problems with these sorts of approaches, relating to issues such as quality control, ensuring focus on priority needs, sustainability, and so on, but where these issues are adequately addressed there is clearly a potential for development of significant information resources, and the new tools certainly facilitate the actions of volunteers making data and information available, and thereby increase access to it.

Demonstrating value, usefulness and fitness for use: Information or knowledge-based activities tend to lose out because it is often hard to demonstrate their impact, and because they have significant recurrent costs which donors are traditionally reluctant to defray. The difficulty in funding knowledge management and dissemination is certainly partially a result of short-termism on the part of donors. On the other hand, over the past thirty years large amounts have been invested in a range of information-based projects and programmes some of which have yielded little useful return. This may arguably be because many such activities have been seen as ends in themselves, with insufficient thought given to what purpose they are intended to serve. It is therefore essential that activities that increase access to data, information and knowledge are developed and implemented with a clear understanding of who is going to be using the outputs, and how. At the same time this value needs to be clearly demonstrated so that it is clearly understood. One of the ways of doing this is through support for scientific research and publication. The informatics infrastructure promoted by GBIF, for example, enables researchers to access and synthesize a wide range of diverse datasets on the occurrence of species, spanning continents and based on decades or even centuries of collection and observation. This access has facilitated a growing body of scientific literature in policy-relevant areas, including invasive alien species, impacts of climate change, crop wild relatives and cultural values of biodiversity.[91,92,93,94]



Lack of strategies and action plans: Careful analysis of existing data discovery and publishing efforts often leads to a conclusion that they are immature and opportunistic, aimed at tapping low hanging fruit. Data discovery and publishing strategies from local to global scale have a vital role in ensuring the availability of the data, information and knowledge required to implement national or institutional biodiversity action plans. Without appropriate data discovery and publishing strategies it is hard to mobilise the available data resources in a systematic way. To help in moving away from a culture of immature and opportunistic data publishing to demand-driven data mobilisation, GBIF on the basis of recommendations of two task groups, has developed the 'Best Practice Guide on Data Discovery and Publishing Strategy and Action Plans',[95] from which the preceding diagram is adapted. It is hoped that this sort of approach will increasingly be used.

4 Recommendations for addressing the barriers

Huge advances in information technology and the explosive growth of the internet in the past twenty years, combined with a strong commitment in many quarters towards provision of open-access resources, has led to a vast amount of data, information and knowledge related to the conservation and sustainable use of biological diversity being widely and often freely available electronically. From the preceding section, it is evident there has been good progress on many of the barriers, but there remains significant ground still to cover.

Addressing the barriers calls for changes in behaviour in a wide range of groups and individuals, including publishers of academic journals, donors, academics, governments and 'knowledge brokers' such as large internationally operating NGOs, in particular those which have endorsed the Principles of the Conservation Commons. Rather than identifying potential actions for addressing each of the barriers separately, this section focuses on actions that could be taken by each of these groups to address the barriers identified.

It should be noted that these recommended actions are in addition to any that may be made by the CBD Conference of the Parties concerning the potential role of the Clearing-House Mechanism in facilitating knowledge sharing and information exchange. They are also complementary to the priorities identified earlier this year at the *Global Biodiversity Informatics Conference* as synthesised in the resulting *Global Biodiversity Informatics Outlook* due to be launched later in 2012.[96]

4.1 Actions focused on governments

Action 1 - Further develop national and international policies for sharing data, information and knowledge, and facilitate their implementation: Through this approach data, information and knowledge can be more effectively used at the national level, and at the same time be more accessible internationally. Without national level enabling frameworks there is potential for duplication of effort in data collection and management, and opportunities for increased synergy may be missed. On the other hand such national policies can lead to more effective use of data, information and knowledge in addressing national needs and international obligations. Where appropriate it is valuable if national policies draw on international frameworks so as to build consistency and coherence. It would also be valuable if, in promoting implementation of such policies at the national level, Governments were to put in place processes for helping to ensure improved access to Government-generated data, information and knowledge.

Action 2 - Encourage continuation of funding for information resources and maintenance of infrastructures: In the current financial climate, it is important that Governments continue to recognise the long-term importance of data, information and knowledge relating to biodiversity and

ecosystem services, and the value of its more effective and efficient dissemination and use. For example, observation and monitoring data compiled over many years is essential in understanding change over time, and for developing the indicators that are increasing being used in assessing progress in achieving targets. Ensuring continuation of funding may require improved understanding of needs, and better communication of the values of the different processes and infrastructures. This may also require review of the different existing initiatives in order to ensure that the most effective investments are being made.

Action 3 - Consider making research permissions dependent on those applying providing access to any data and information generated: Where Governments are in a position to give permission for research to proceed, unless there are good reasons why not they should consider making it a requirement of that permission that electronic copies of all data and information generated are deposited in a timely manner in designated institutions, using appropriate standards. At smaller scale, similar approaches can be adopted by, for example, protected area administrators in granting research permission in any given protected area, who might require submission of data and information to a specific institution.

4.2 Actions focused on donors and others providing support

Action 4 - Ensure that each funded project provides appropriate access to the data, information and knowledge that it generates: A minimum condition of funding of research or monitoring activities should be that data generated are made publicly available unless there are compelling reasons to the contrary. In addition, donors have the opportunity to make it a condition of funding that all recipients have plans in place for the effective management and dissemination of all data, information and knowledge arising from the project that they fund. Demonstration that data have been made available and that these plans have been implemented could be made a condition of release of final tranches of funding, and non-compliance could potentially lead to denial of further funding. This is not to say that each project should set up and maintain in perpetuity its own information infrastructures, but that they should, for example, indicate what standards and protocols will be used, and which existing infrastructures they will use to deposit data and information, and to communicate knowledge gained.

Action 5 - Promote longer-term investment in maintenance of data and information resources, and funding of knowledge management and dissemination: Donors are in a position to support long-term programmes both through their own investments, and through the influence that they can have on the projects that they fund. They also potentially have influence through collaboration with other donors, and with their stakeholders. A longer-term approach, and a focus on building on existing initiatives rather than de-novo development might have less ‘visibility’ and be less ‘sexy’ to fund, but should in the long term be more cost-effective in delivering data, information and knowledge that supports decision making. For their part, those organizations involved in key data, information and knowledge management initiatives should perhaps be more realistic and focused in their goals, and strive harder to find ways of demonstrating the impact of this aspect of their work.

4.3 Actions focused on publishers and publications

Action 6 - Increase open access to publications and reports, including archives: All those that publish papers and reports of potential importance for the conservation and sustainable use of biodiversity should consider ways of making them as widely available as possible so that the information and knowledge they contain can be used in decision making and where appropriate acted upon. In fact in recent years there has already been a considerable shift towards more open

access, including free-access web publication of academic articles, with some online journals entirely free and many others having partially free access. Where charging is regarded as necessary, a mixed model, including partial open access, tiered subscription rates, and subsidies or subventions for disadvantaged groups (such as many institutions and individuals in developing countries) could be investigated. In addition, even journals for which a charge is made could adopt a free-access policy for volumes more than a year or two old, which could open up a considerable amount of historical information at relatively little cost. Meanwhile, initiatives working to increase access to publications, particularly for developing countries, should be supported wherever possible, including through making publications available.

Action 7 - Facilitate access to data and information on which publications are based: Many journals already make it a requirement of publication that original (raw) data are made available, either as subsidiary files and documents on the website, or by request from the authors. This should be further encouraged, with the former option generally preferable as it avoids the problems associated with authors changing jobs and contact details and otherwise becoming uncontactable.

Action 8 - Support and encourage the publication of 'data papers' that describe available datasets: As discussed earlier, one of the potential mechanisms for creating incentives for academics to release data would be the creation of a data publishing framework that would help to ensure that investments and efforts made by institutions and individuals on management and publishing of primary scientific data was appropriately recognised alongside the recognition already given to scholarly publications. One of the possible means of achieving this is through the publication of 'data papers' which describe datasets that academics have worked to compile and make available.

4.4 Actions focused on academics and research institutions

Action 9 - Create and promote incentives for academics and research institutions to increase access to data and information: As noted earlier, creating incentives to make academics more willing to share data and information more readily can be challenging. Ideally, dissemination of basic data and information should be something that becomes itself rewarded professionally, but this runs counter to the major thrust of academia and its reward systems which focus strongly on relatively narrowly defined notions of excellence and impact. However, a number of innovative solutions are already being developed and put into practice, and it seems appropriate to build on these further by:

- promoting cascading or 'deep data' citing mechanisms so that those providing information would be confident that their contribution would be publicly acknowledged
- making it as easy as possible for academics to release data and information once they have extracted value from it
- supporting the development of impact indices based on use and citation of data, as well as of 'data papers' discussed above
- engaging academics in collaborative enterprises which make use of their skills while still ensuring that they can extract their own value through publishing papers
- appealing to notions of a common good, and demonstrating the value of releasing data, information and knowledge

4.5 Actions focused on 'knowledge brokers'

Action 10 -Establish and maintain easy-to-use electronic information infrastructures or repositories for ensuring long-term access to data, information and knowledge: Build on and expand current efforts to establish a network of data hosting centres, to encourage data custodians to publish their data with the minimum of effort and cost. The network of such data deposits could be expanded

relatively easily, and cheaply, if the latest developments in cloud computing are exploited. A system could be established whereby any users could add further labels if they found this useful. To provide optimal access to the resources, the use of semantic web technology ("web 2.0", and beyond), would be vital – allowing a combination of computer and human interaction with data and information throughout the internet.

Action 11 -Develop and promote means for improving access to the 'long tail' of results from small and medium-sized research projects: It has been observed that a large proportion of the data and information generated in scientific research is produced as a result of small or medium-sized projects with limited resources.[97] This is the kind of data and information that is most likely to be stored in *ad hoc* forms (often on spreadsheets, in word-processing files or in notebooks), and is therefore some of the most difficult data and information to discover. Means should be developed for increasing access to this in the future. Potentially useful mechanisms include the development of data registries where researchers could record what information they have generated, the development of low-cost open platform data-sharing systems, and the further promotion of the use of such systems by researchers.

Action 12 -Review and further promote the use of common vocabularies, classification systems, and standards: A critical review of existing knowledge management standards and approaches could be undertaken, building on the work already being done by a number of organizations. The aim would be both to increase the use of existing standards and approaches, and to identify gaps and potential overlaps, and suggest ways in which these could be addressed. This would entail a realistic assessment of the likelihood of being able to develop widely agreed standards in some areas, such as ecosystem and threat classification schemes, and identification of how this might also be addressed so that there is joint understanding of how different systems and interpretations relate . These sorts of issues are already being considered in the standards community, and their work needs to be supported and built upon.

Action 13 -Create better road maps to useful knowledge: As was recognised earlier, many sources of data, information and knowledge would be better used if it was easier to find out that they existed in the first place, and then to discover more easily what was and was not in them, what their origin and quality was, and so on. In this regard, there is a useful trend towards more integrated information systems that access data, information and knowledge from multiple sources, and use different web services to produce new insights and new products. This trend should be encouraged, although at the same time it is important to increase collaboration between providers of these services to help ensure that there is improvement in services available, rather than simply a proliferation in number of information systems. Improved information tools associated with integrated information services would allow further data, information and knowledge to be linked or added, review and improvement of existing material, and so on.

4.6 Potential future role of the Conservation Commons

Action 14 -Signatories to the Principles of the Conservation Commons should set an example in increasing access to data, information and knowledge that they hold: Those who have signed up to the Principles of the Conservation Commons should seriously consider adopting policies where by default all data, information and knowledge that they manage is made publicly available unless an explicit decision to the contrary is made in specific cases. They should also consider encouraging a similar approach in their partners and members, and working together take the lead in encouraging other groups to adopt and implement the Principles of the Conservation Commons. In addition, those who have signed up to the Principles of the Conservation Commons might consider agreeing to an auditing process of some sort to determine how well they were adhering to the Principles.

Action 15 -Improve communication of the value of sharing data, information and knowledge, and the potential costs of not doing so: The Conservation Commons should take a lead in more effectively communicating the many values of sharing data, information and knowledge, as a means of convincing those holding biodiversity-relevant data, information and knowledge that it is important to make it more widely available. This may be most effectively done by considering a series of examples, where a strategy to increase access to data on a specific issue or set of issues can be shown to be beneficial in addressing key policy or research questions.

Action 16 -Help to ensure that those willing to increase access to data, information and knowledge have access to guidance on how to do so: The Conservation Commons could prepare a guide or series of guides setting out how to put the Principles of the Conservation Commons into practice. Such guidance would build on the many guidelines and manuals that already exist, and draw attention to existing documents and tools for those not already familiar with them. Such guidelines could, for example, follow the model of the best practice guidance recently released by the UK Data Archive,[98] supplement the series of guides and manuals currently available in the GBIF Online Resource Centre,[99] and drawn on lessons from the UK Medical Research Council's policies and practice related to data sharing.[100] This could include supporting the development of guidelines for donors on how and why to increase access to data, information and knowledge.

5 Notes and references

1 UNEP/IPBES/2/INF/6 Capacity-building in an intergovernmental science-policy platform on biodiversity and ecosystem services

2 *International Expert Meeting in IPBES and Capacity Building*, Trondheim, Norway, 25-27 May 2011. Meeting Report: Chairmen's report with summaries and findings. Downloaded from www.dirnat.no/ipbes

3 UNEP/CBD/COP/10/27 Report of the Tenth Meeting of the Conference of the Parties to the Convention on Biological Diversity

4 A 2007 meeting of the Ecological Society of America reviewed the obstacles to sharing data, see: www.esa.org/science_resources/programs/DocumentFiles/DataObstacles_Wkshp_notesFinal.pdf

5 Chavan VS, Ingwersen P (2009). Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. *BMC Bioinformatics*, 10(*Suppl* 14): S2, www.biomedcentral.com/1471-2105/10/S14/S2

6 Piwowar HA (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. PLoS ONE 6(7): e18657, doi: 10.1371/journal.pone.0018657.

7 Tenopir C, Allard S, Doughlass A, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M (2011). Data sharing by scientists: practise and perceptions. PLoS ONE, 6(6): e21101, doi: 10.1317/journal.pone.0021101.

8 See www.siac.net.co/web/sib

9 See www.conabio.gob.mx

10 See www.sanbi.org

11 See www.biodiversity.no

12 See www.nbn.org.uk

13 See splink.cria.org.br

14 See www.ala.org.au

15 Nielsen M (2011). The new Einsteins will be scientists who share. The Wall Street Journal, October 29, 2011. Accessible at online.wsj.com/article/SB10001424052970204644504576653573191370088.html

16 See www.mendeley.com

17 See http://hubs.plos.org/web/biodiversity/home

18 Mindell DP, Fisher BL, Roopnarine P, Eisen J, Mace GM, Page RDM, Pyle RL (2011). Aggregating, tagging and integrating biodiversity research. PLoS ONE 6(8): e19491. Doi:10.1371/journal.pone.0019491

[19] The Royal Society (2012). *Science as an open enterprise.* Available from http://royalsociety.org/policy/projects/science-public-enterprise/report http://royalsociety.org/policy/projects/science-public-enterprise/report

[20] Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA (2011). Public availability of published research data in high-impact journals. PLos ONE 6(9):e24357. Doi:10.1371/journal/pone.0024357

[21] Chavan VS, Ingwersen P (2009). Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. *BMC Bioinformatics*, 10(*Suppl* 14): S2, www.biomedcentral.com/1471-2105/10/S14/S2; Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA (2011). Public availability of published research data in high-impact journals. PLoS ONE 6(9): e24357. Doi: 10.1317/journal/pone.0024357; Corbyn Z (2011). Researchers failing to make raw data public. Nature, 14 September 2011, doi: 10.1038/news.2011.536.

[22] Moritz T, Krishnan S, Roberts D, Ingwersen P, Agosti D, Penev L, Cockerill M, Chavan V (2011). Towards mainstreaming of biodiversity data publishing: recommendations of the GBIF data publishing framework task group. *BMC Bioinformatics* 12(Suppl 15):S1, www.biomedcentral.com/1471-2105/12/S15/S1

[23] GBIF (2011). New incentives for biodiversity data publishing. Copenhagen: Global Biodiversity Information Facility, accessible at www.gbif.org/communications/news-and-events/showsingle/article/new-incentive-for-biodiversity-data-publishing.

[24] Chavan V, Penev L (2011). Data Paper: A mechanism to incentivise data publishing in biodiversity science. *BMC Bioinformatics* 12(Suppl 15):S2, www.biomedcentral.com/1471-2105/12/S15/S2

[25] See www.gbif.org/communications/news-and-events/showsingle/article/first-database-derived-data-paper-published-in-journal

[26] Peven L, Mietchen D, Chvan V, Hagedorn G, Remsen D, Smith V, Shotton D (2011). Pensoft Data Publishing Policies and Guidelines for Biodiversity Data. Pensoft Publishers. Accessible at www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf

[27] See www.codata.org/taskgroups/TGdatacitation

[28] See www.datacite.org

[29] Ingwersen P, Chavan V (2011). Indicators for Data Usage Index (DUI): an incentive for publishing primary biodiversity data through global information infrastructure. *BMC Bioinformatics* 12(Suppl 15):S3, www.biomedcentral.com/1471-2105/12/S15/S3

[30] Chavan VS, Ingwersen P (2009). Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. *BMC Bioinformatics*, 10(*Suppl* 14): S2, www.biomedcentral.com/1471-2105/10/S14/S2

[31] See www.oaresciences.org. OARE is linked to HINARI (www.who.int/hinari/en) launched by the World Health Organization and Yale University which supports the medical and health community, and AGORA (www.aginternetwork.org/en) launched by the Food and Agriculture Organization of the United Nations and Cornell University which supports the agricultural community.

[32] See www.scielo.org

[33] See www.biodiversitylibrary.org

[34] Chapman AD, Grafton O (2008). *Guide to best practice for generalising primary species occurrence data*. Global Biodiversity Information Facility, Copenhagen. 27pp. Accessible at www.gbif.org/orc/?doc_id=1233

[35] See www.ala.org.au/faq/data-sensitivity and the report linked from this page

[36] Berne Convention for the Protection of Literary and Artistic Works, www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html

[37] Directive 2001/29/EC of the European Parliament and of the Council of 22 may 2001 on the harmonization of certain aspects of copyright and related rights in the Information Society, see http://europa.eu/legislation_summaries/information_society/data_protection/l26053_en.htm

[38] See www.creativecommons.org

[39] See eol.org

[40] See www.esa.org/science_resources/programs/DocumentFiles/DataObstacles_Wkshp_notesFinal.pdf

[41] See www.tdwg.org

[42] See code.google.com/p/gbif-providertoolkit

[43] See www.gbif.org/orc/?doc_id=2820

[44] See dublincore.org/documents/dcmi-terms

[45] See rs.gbif.org/extension/gbif/1.0/vernacularname.xml

[46] See tools.gbif.org/resource-browser
[47] See www.informea.org
[48] See www.eionet.europa.eu/gemet
[49] See aims.fao.org/website/AGROVOC-Thesaurus/sub
[50] See www.usgs.gov/core_science_systems/csas/biocomplexity_thesaurus
[51] See vocabularies.gbif.org
[52] See bionomenclature-glossary.gbif.org
[53] See rs.tdwg.org/dwc
[54] See www.tdwg.org/activities/abcd
[55] See rs.tdwg.org/dwc/terms/history/dwctoabcd
[56] See rs.gbif.org/extension
[57] See rs.gbif.org
[58] GBIF (2011). Getting started: an overview of data publishing in the GBIF network (contributed by Remsen D, Ko B, Chavan V, Raymond M) Copenhagen: Global Biodiversity Information Facility, 16 pp. ISBN: 87-92020-28-3.
[59] Much of what we do in the conservation community is place based, so the management and coding of spatial data is fundamental to access, archiving, retrieval, and analysis. The Data Sharing Task Force of the Group on Earth Observations is a good reference point in this respect.
[60] For example, effective implementation of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) requires a rather conservative approach to taxonomy, with the use of standard references where these exist, and changes approved following due process under the Convention. The IUCN Red List on the other hand is under no such constraint and can adopt changes more rapidly, reflecting the concerns of and retaining credibility with its core constituency.
[61] See www.catalogueoflife.org
[62] See dublincore.org/documents/usageguide
[63] See metadata-stds.org/11179
[64] See rs.tdwg.org/dwc
[65] See species-id.net/wiki/Audubon_Core
[66] See www.tdwg.org
[67] See the report of the GBIF Metadata Implementation Framework Task Group available at imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=2d85d0e8c76408129024c09aa072d6
[68] See the GBIF Metadata Catalogue available at links.gbif.org/gbif_metadata_catalogue_specification.pdf
[69] See imsgbif.gbif.org/CMS_ORC/?doc_id=2821
[70] There is almost certainly something age-related in this. People who used computers in the 1980s and early 1990s are, as a result of their experiences, likely to be considerably more cautious.
[71] Goddard A, Wilson N, Cryer P, Yamashita G (2011). Data hosting infrastructure for primary biodiversity data. *BMC Bioinformatics*: 12(Suppl 15):S5.
[72] See www.codata.org/taskgroups/TGdataatrisk
[73] Berendsohn WB, Chavan V, Macklin JA (2010). Recommendations of the GBIF Task Group on the Global Strategy and Action Plan for the mobilisation of natural history collections data. *Biodiversity Informatics* 7(2): 67-71.
[74] Approaches to tackling some of these problems are discussed in papers in the 2009 supplement to *BMC Bioinformatics* 10(supp. 14) available at www.springerlink.com/content/1471-2105/10/s1, see for example Page on bioGuid, and Smith *et al.* on scratchpads.
[75] See registry.gbif.org
[76] GBIF (2009). Report of the GBIF GBRDS Stakeholders Planning Workshop. Copenhagen: Global Biodiversity Information Facility, November 2009, 24pp, accessible at www2.gbif.org/GBRDSWorkshopReport-Final.pdf
[77] GBIF (2009). GBIF summary paper: The Global Biodiversity Resources Discovery System (GBRDS). Authored by Gaiji S, Robertson T, Chavan V and Cuarda J. Copenhagen: Global Biodiversity Information Facility, July 2009, 5pp.
[78] The GBIF website (www.gbif.org) provides access to guidance and discussion on many issues relevant to biodiversity in this regard.
[79] See www.gbif.org/orc/?doc_id=2820
[80] See rs.tdwg.org/dwc
[81] See www.tdwg.org/standards/115

[82] See www.biocase.org
[83] See digir.net
[84] See www.tdwg.org/standards/449
[85] See www.gbif.org/informatics/standards-and-tools/publishing-data/data-standards/darwin-core-archives
[86] It should be noted that, with respect to the Principles of the Conservation Commons, "free and open access" does not necessarily mean free of cost, as there may be circumstances in which it is appropriate to recover costs of making data and information available.
[87] See imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=20c9b71635072e05c00950f32120e9
[88] See imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=e0810602ceee5fe16eccddf5d4667d
[89] National Science Foundation (2011). Dissemination and sharing of research results. Retrieved from www.nsf.gov/bfa/dias/policy/dmp.jsp
[90] See www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative
[91] Soberón J, Peterson A (2009), AMBIO: A Journal of the Human Environment, 38(1):29-34. 2009, Monitoring Biodiversity Loss with Primary Species-occurrence Data: Toward National-level Indicators for the 2010 Target of the Convention on Biological Diversity www.bioone.org/doi/full/10.1579/0044-7447-38.1.29
[92] Ramírez-Villegas J *et al* (2010), PLoS ONE, 5(10), e13497. A Gap Analysis Methodology for Collecting Crop Genepools: A Case Study with Phaseolus Beans dx.doi.org/10.1371/journal.pone.0013497
[93] Hillyer R, Silman M (2010), Global Change Biology. Changes in species interactions across a 2.5 km elevation gradient: effects on plant migration in response to climate change dx.doi.org/10.1111/j.1365-2486.2010.02268.x
[94] Gaikwad J *et al* (2011), Ecological Modelling. Ecological niche modeling of customary medicinal plant species used by Australian Aborigines to identify species-rich and culturally valuable areas for conservation doi:10.1016/j.ecolmodel.2011.07.005
[95] GBIF (2010). Best practice guide for 'data discovery and publishing strategy and action plans' version 1.0. Authored by Chavan VS, Sood RK, Arino AH (2010). Copenhagen: Global Biodiversity Information Facility, 29 pp. ISBN: 87-92020-12-7; Berendsohn WB, Chavan V, Macklin JA (2010). Recommendations of the GBIF Task Group on the Global Strategy and Action Plan for the mobilisation of natural history collections data. Biodiversity Informatics, 7(2): 67-71.
[96] See www.gbic2012.org
[97] Heidorn PB (2008). Shedding light on the dark data in the long tail of science. *Library Trends* 57(2): 280-299
[98] UK Data Archive. (2011). Managing and Sharing Data: Best Practice for Researchers. Downloaded from www.data-archive.ac.uk/media/2894/managingsharing.pdf
[99] See www.gbif.org/orc
[100] See www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative