Improving EIA practice: Best Practice Guide for publishing primary biodiversity data

Version 1.0

August 2011

Suggested citation:

GBIF (2011). Promoting biodiversity data inclusive EIA: Best Practice Guide for publishing primary biodiversity data, (contributed by Cadman, M., Chavan, V., King, N., Willoughby, S., Rajvanshi, A., Mathur, V.B., Roberts, R. and Hirsch, T.) Copenhagen: Global Biodiversity Information Facility, pp 50. ISBN: 87-92020-35-6. Accessible at http://links.gbif.org/eia_biodiversity_data_publishing_guide_en_v1

ISBN: 87-92929-35-6

Persistent URI: http://links.gbif.org/eia_biodiversity_data_publishing_guide_en_v1

Language: English

Project Partners:

GBIF www.gbif.org	IAIA www.iaia.org	भारतीय वन्यजीव संस्थान Wildlife Institute of India www.wii.org.in
	SANBI Biodiversity for Life www.sanbi.org.za	

Document Control:

Version	Description	Date of release	Author(s)
1.0	Review, edits and final styling	August 2011	Cadman, M., Chavan, V., King, N., Willoughby, S., Rajvanshi, A., Mathur, V.B., Roberts, R., Hirsch, T.

A summarised version of this document titled, 'Publishing EIA-Related Primary Biodiversity Data: GBIF-IAIA Best Practice Guide' is accessible at http://www.iaia.org/publicdocuments/special-publications/sp7_web.pdf

Cover photo credits: Tigers: Wildlife Institute of India; Sunbird on *Protea*: SANBI; Plant , Vinod Mathur

About GBIF and the project partners

*GBIF:*The Global Biodiversity Information Facility (GBIF) was established by countries as a global mega-science initiative to address one of the great challenges of the 21st century – harnessing knowledge of the Earth's biological diversity. GBIF envisions 'a world in which biodiversity information is freely and universally available for science, society, and a sustainable future'. GBIF's mission is to be the foremost global resource for biodiversity information, and engender smart solutions for environmental and human well-being (GBIF 2011a). To achieve this mission, GBIF encourages a wide variety of biodiversity data holders, generators and users across the globe to discover and publish (make discoverable) data through the GBIF network.

IAIA: The International Association for Impact Assessment, established in 1980, provides an international forum for advancing innovation and communication of best practice in all forms of impact assessment, to further the development of local, regional and global capacity in impact assessment. IAIA members number more than 2,500, representing over 100 countries. IAIA seeks constantly to improve impact assessment procedures and practices around the world through development of Best Practice Guides, training, professional quality assurance and creating opportunities for professional networking.

SANBI: The South African National Biodiversity Institute is a public entity under the Department of Environmental Affairs of South Africa. SANBI's purpose is to champion the exploration, conservation, sustainable use, appreciation and enjoyment of South Africa's extremely rich biodiversity, for all people. Part of SANBI's mandate is to collect, process, co-ordinate and disseminate information about biodiversity, and to establish and maintain databases thereof. In addition, SANBI must report to the South African Government on the status and trends of biodiversity. The EIA Biodiversity Data Publishing Facility described in this Guide, is being developed by SANBI with input and support from the SA Department of Environmental Affairs, the IAIA in South Africa (IAIA-sa), various provincial designated authorities and other interested parties. This collaboration will ensure the usefulness of the final product to a wide range of stakeholders, and provide for its long-term support. SANBI hosts SABIF, the South African Node for GBIF activities

WII: The Wildlife Institute of India, established in 1982, is an autonomous organisation under the Ministry of the Environment & Forests of India. WII's mandate is to train government and non-government personnel, conduct research and advice on matters of conservation and management of wildlife resources. WII serves as the co-ordinating node for GBIF activities in India and has played an important role in assessing policy and institutional constraints and opportunities for development and customisation of the EIA Biodiversity Publishing Facility described in the guide.

Table of Contents

List of Figures

List of Boxes

List of Tables

Section 1: Why we need this guide

The importance of biodiversity

Biodiversity is a critical foundation of human well-being and contributes significantly to shaping the development path of a region or country. In addition to its own intrinsic value, biodiversity (Box 1) is an invaluable resource for building sustainable livelihoods, creating jobs and alleviating poverty; enhancing rural development, food security and land-use; delivering water and other critical ecosystem goods and services; and assisting communities with adaptation to the effects of climate change. Good biodiversity management is good for the economy, good for local development and good for business.

Development choices impact on biodiversity, and not all land-uses are compatible with conserving biodiversity. Countries need to find a development path that leads to social and economic upliftment, while reducing biodiversity loss and environmental degradation, and maintaining healthy ecosystems. Globally, biodiversity resources are being eroded at an increasing pace by unsustainable land-use practices, over-extraction, inappropriately-located development, loss of habitat, invasive alien species, pollution and other environmental changes, including those that are attributable to climate change. Environmental Impact Assessment (EIA) is an important tool for generating options to relieve pressures on biodiversity and ecosystems, whilst promoting ecologically sustainable development that is located appropriately.

Box 1: Definition of biodiversity

Biodiversity is defined in the Convention on Biological Diversity as: "the variability amongst living organisms from all sources including, *inter alia,* terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems."

United Nations (1992). Text of the Convention on Biological Diversity, accessible at http://www.cbd.int/doc/legal/cbd-en.pdf

Biodiversity-inclusive environmental impact assessment: the issue

Environmental Impact Assessment provides opportunities for integrating biodiversity values with development objectives (Rajvanshi et al., 2007),and ***biodiversity-inclusive impact assessment***

is promoted by provisions of, *inter alia*, the Convention on Biological Diversity (CBD), the Ramsar Convention and the Convention of Migratory Species (Rajvanshi *et al.,* 2007; IAIA, 2005). For a variety of reasons, however, biodiversity has not always given specific or appropriate consideration in EIAs, despite being a critical component of the broader environment (King *et al*., *in prep*).

Knowledge about the ***identity and occurrence of organisms*** forms the backbone of our understanding of the biological world, and is essential for monitoring the state of natural ecosystems, for developing sound environmental management policies and making ecologically sustainable development decisions. Ideally***, biodiversity-inclusive*** EIA should: (a) use biodiversity information to determine the biological or ecological sensitivity of a development site, and (b) generate new biodiversity records about these sites.

Until recently, however, little of the primary biodiversity information generated as part of impact assessment (IA) work has been published or peer-reviewed, and it remains difficult to access. There has been no easy-to-use mechanism of discovering and accessing digital biodiversity data for use in EIA, or for publishing the biodiversity data that EIA generates (King, *et al.*, in prep.).

These shortcomings can compromise the quality of the EIA findings and limit the extent to which they can meaningfully include biodiversity. This, in turn, reduces the confidence that can be placed in subsequent decisions and the transparency of the EIA process. The lack of easy-to-use mechanisms and standardised formats for publishing the biodiversity data gathered during EIA means that most of these data — which sometimes represent the only knowledge about the biological diversity of a certain site — are 'lost' once the EIA is complete (Rajvanshi *et al*., 2007). Data are seldom captured in a re-usable form and are not accessible for use in subsequent EIAs, or by researchers, planners, decision-makers or society at large.

The solution: best practices for publishing biodiversity data from EIAs

Through the ***Global Biodiversity Information Facility*** (GBIF), digital biodiversity data are being made freely and openly available via the Internet for scientists, researchers, decision-makers and the general public. GBIF promotes a suite of standards and data publishing tools that can be used to capture, publish and discover primary biodiversity data in standardised formats. To date, these tools have been used for publishing biodiversity data gathered from observational records (accounting for about 60% of some 300m records accessible through GBIF as of mid-2011) and biological specimens housed in natural history collections (about 40%) . This Best Practice Guide describes these tools, standards and infrastructure, and explains when and how they should be

used, with specific reference to environmental impact assessment practitioners. It also indicates sources of additional help, should this be required.

Towards development of an "EIA Primary Biodiversity Data Publishing Facility"

GBIF, in co-operation with the International Association of Impact Assessment (IAIA), the South African National Biodiversity Institute (SANBI) and the Wildlife Institute of India (WII), has invested in developing an "EIA Biodiversity Data Publishing Framework".

The aims of this project have been to:

- Develop a suite of best practices, protocols, standards and tools for the capture, archiving, discovery and publishing of EIA-related biodiversity data;
- Develop data-transformation tools to make EIA biodiversity data interoperable and exchangeable according to globally accepted standards;
- Facilitate the discovery and publishing of EIA biodiversity data through local and global information systems, networks and data hosting environments (King *et al*., in prep.).

As part of this process, two pilot projects were initiated, one in South Africa and the other in India.These countries were selected because: (a) they are emerging economies with rich biodiversity assets and long histories of mandated EIA, with numerous EIAs conducted annually; (b) economic and development pressures influence the outcome of EIA processes very strongly, and capturing biodiversity data is currently not prioritised ; (c) biodiversity informatics are currently not a mainstream activity in these countries, so EIA practitioners are not generally aware of the contributions they could make to building up databases of primary biodiversity data, nor how this could benefit future EIA processes; and (d) at the time of inception of the project, South Africa and India co-chaired the Biodiversity and Ecology section of IAIA, thus facilitating rapid endorsement and start-up of the pilot projects (King *et al*., in prep).

The South African pilot project, which has been co-ordinated by SANBI, in conjunction with IAIA-sa, has focussed on facilitating discovery , and access to, biodiversity data in a format that can be easily incorporated into EIA processes. A key output of this pilot has been, *inter alia*, the development of a web-based 'EIA Biodiversity Data Publishing Facility', based on extensive stakeholder consultation and a comprehensive assessment of user needs. The development of this facility forms part of a broader strategy by SANBI to provide reliable and scientifically defensible biodiversity information to support land-use decision-making, through its Biodiversity Advisor Web Portal (www.biodiversityadvisor.sanbi.org).

The Indian pilot project, co-ordinated by the Wildlife Institute of India (WII), builds on the platform established by the South African project, and has focussed on developing policy support and effecting the socio-cultural changes required for uptake and use of the EIA Biodiversity Data Publishing Facility. Future focal areas will include adapting the technical infrastructure developed in South Africa, with appropriate customisation of procedures to the Indian context.

This guide aims to ensure rapid and easy uptake of the framework developed in the pilot projects and to mainstream the publishing of EIA-related biodiversity data through the GBIF network, by explaining the relevant standards, protocols and procedures.

What this guide is, and what it is not

Primary biodiversity data is defined as 'digital text or multimedia data records detailing facts about the instance of occurrence of an organism, i.e. the what, where, when, how and by whom of the occurrence and the recording (GBIF, 2009).' This best practice guide is one of a series of publications developed by GBIF relating to primary biodiversity data, The overall purpose of the guide is to enable EIA practitioners, consultants and other interested and affected parties to discover, capture, manage and publish to common standards, the primary biodiversity data generated during impact assessment processes. If widely adopted, the best practices described here should ensure that:

- EIA practitioners adopt, as best practice, the GBIF tools and infrastructures for discovery and publishing of EIA-related primary biodiversity data;
- Primary biodiversity data from EIAs are archived as part of national or other datasets, freely available in the public domain for further access and use;
- Biodiversity-inclusive EIAs are promoted, enhancing decision-making processes and contributing positively to stemming biodiversity loss.

The document is divided into two main parts. The following two sections (2 and 3), provide the theoretical information on key concepts and principles that EIA practitioners should understand, in order to make effective use of the biodiversity publishing tools and infrastructures that are available to them. The remainder of the guide (sections 4 to 6), is more practically-orientated, providing 'how-to' guidelines on what environmental assessment practitioners should do to publish the primary biodiversity data gathered during EIA, and how they can access data that they may require to carry out EIA.

The guide is not a comprehensive reference work on environmental impact assessment; the Global Biodiversity Information Facility; biodiversity informatics standards, tools,

processes and infrastructure; or biodiversity. Wherever possible, reference is made to other documents that provide in-depth guidance for different steps in the data publishing process, for those users who require more detailed information (see also Appendix 3 for a detailed listing of references and websites where this information can be obtained).

Section 2: Biodiversity data and principles of best practice

This section of the guide describes the levels at which impact on biodiversity can be assessed, the types of biodiversity data that exist, their value in informing decision-making, and the principles that underpin best practice for publishing EIA-related biodiversity data.

To provide an understanding of how biodiversity is likely to respond to a proposed development activity, impacts need to be assessed in terms of ecosystem composition, structure and function (See Figure 1). Functioning ecosystems with intact biodiversity deliver ecosystem services that are essential for survival. Ecosystem functioning is impaired or lost if the composition, structure or ecological processes that maintain an ecosystem are disrupted. Biodiversity loss disrupts ecosystem functioning and compromises sustainable economic development. Disrupted ecosystems are also more vulnerable to shocks and disturbances and are less able to supply society with critical ecosystem services. Reducing biodiversity loss is one of the surest ways of maintaining functioning ecosystems and adapting to the inevitable effects of climate change (Cadman *et al*., 2010).

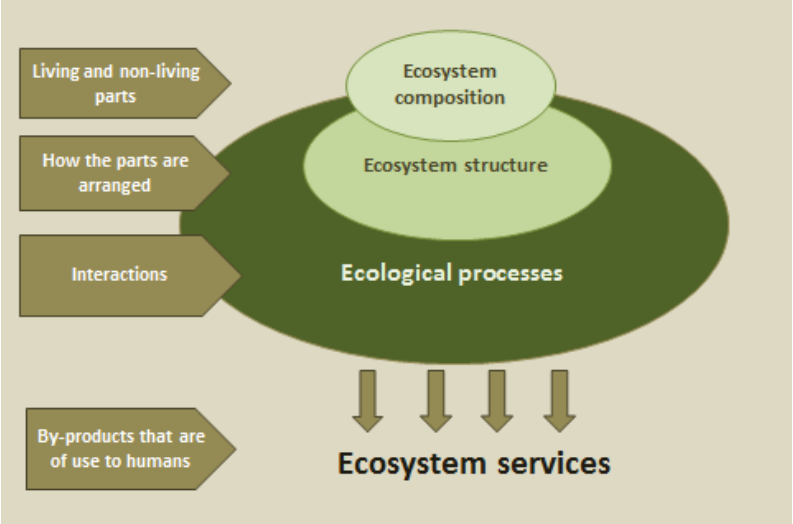


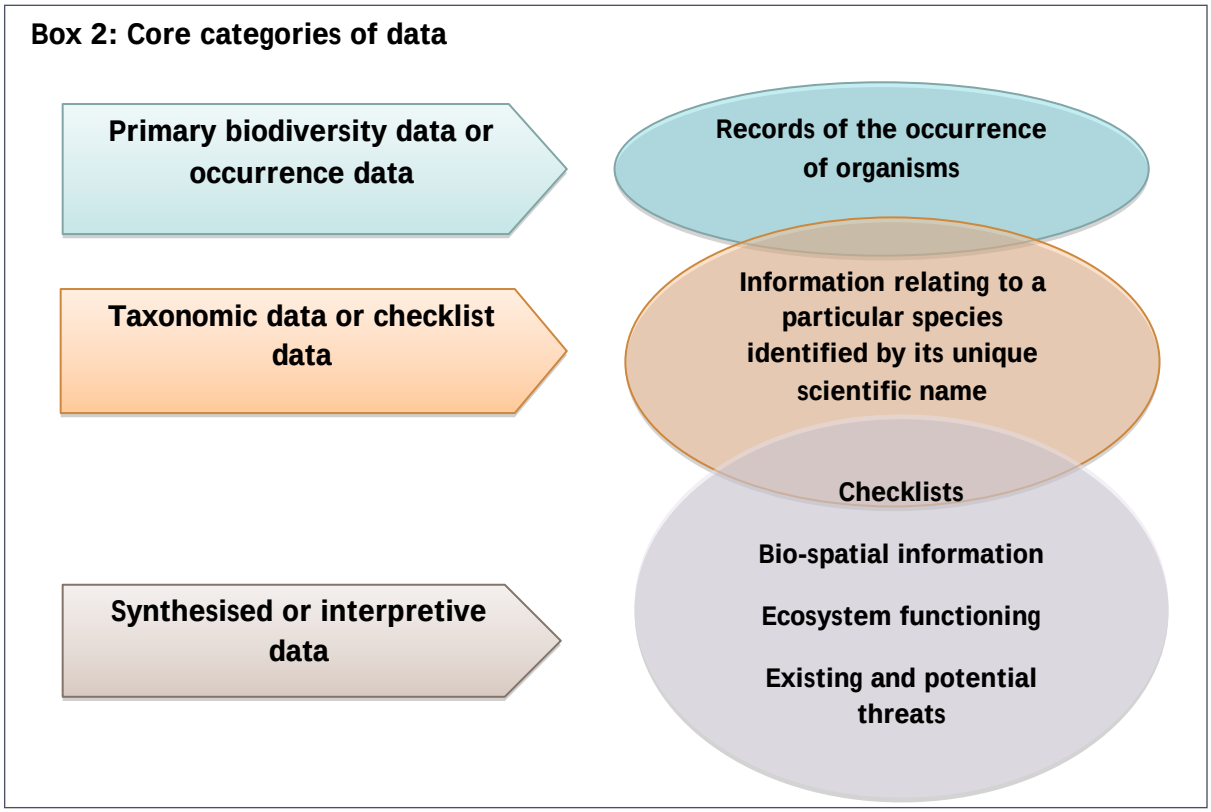
Figure 1: Components of functioning ecosystems, showing the different levels at which impacts on biodiversity can be assessed (from Cadman *et al*., 2010)

Types of biodiversity data

To make meaningful assessments, environmental assessment practitioners need access to biodiversity data that are in a readily usable form, can be accessed using standardised protocols and are verifiable.

There are several categories of biodiversity data, or levels at which data can be gathered and used, and it is important to distinguish between these. The first distinction to be drawn is between ***primary biodiversity data*** (species occurrence data), **taxonomic data** (checklists and information about the identity of organisms), and ***synthesised or interpretive (secondary) data*** (see Box 2).

Box 2: Core categories of data



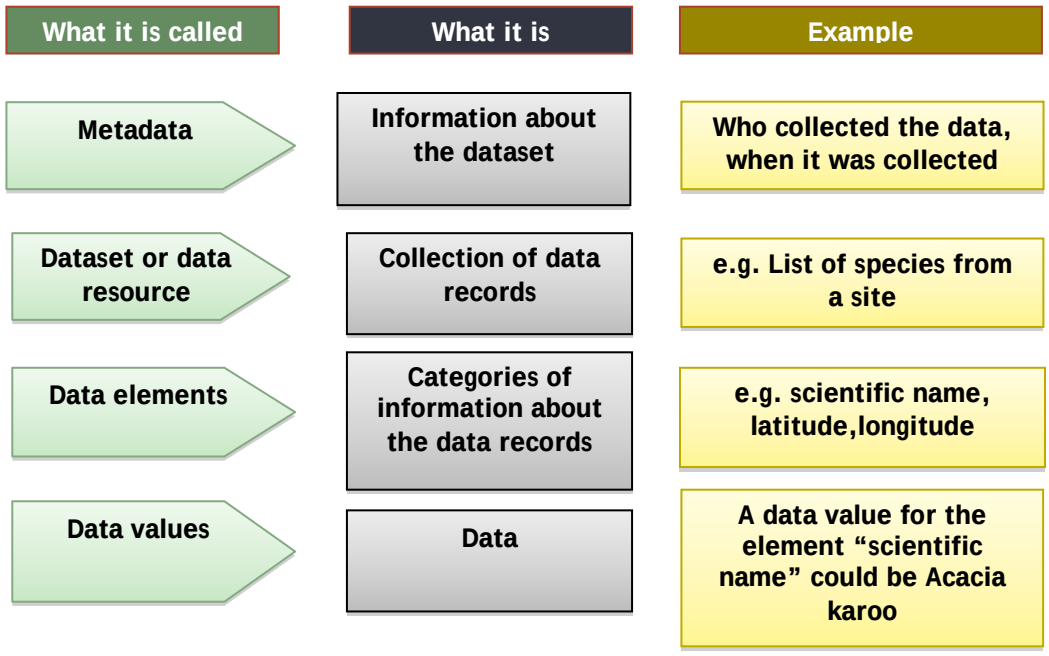
Primary biodiversity data are the digital text or multimedia data records that detail the ***occurrence*** of organisms. These data provide information about what organisms are present, where they occur (specified by geographic co-ordinates), when they are found and who collects the data (GBIF, 2010; GBIF, 2011b). Records of the occurrence of organisms form the backbone of our knowledge about biological systems and are essential for monitoring the state of biodiversity and natural systems, and for predicting the likely impact that a development may have at a site.

The best practices described in this guide for publishing EIA-related biodiversity data apply to primary biodiversity data only.

From a data publication perspective, GBIF makes the distinction between several terms relating to biodiversity data, and it is essential to use these precisely to avoid confusion. These terms, which are described in Box 3, as well as in related GBIF publications (GBIF, 2011b), include: data resources or datasets, data elements, data values and metadata.

Metadata refer to the descriptive information that accompanies a dataset – they are the data about data. Metadata are required for all datasets published through the GBIF network and can describe both digital and non-digital datasets (GBIF 2011b). Metadata improve dataset discovery, and provide potential users with details on the fitness for use of the data they describe.

Box 3: Terminology associated with datasets.



Primary biodiversity data, taxonomic data and metadata are each supported by a different data publishing option within the GBIF network (See Section 5).

Data publishing

GBIF provides a means of sharing biodiversity data, through a process known as 'publishing' (Box 4), that makes it universally accessible through the use of standard procedures and protocols.

This guide will help environmental assessment practitioners, consultants and other interested and affected parties to choose the most suitable option or tool for publishing the primary biodiversity data they have gathered, enabling it to be accessed through the GBIF network.

Box 4: What is data publishing?

In this context, the term '***publish***' refers to making biodiversity datasets publicly accessible in a standardised form, via an online (Internet) access point, which is typically a web address (a URL). This access point is recorded in a GBIF Registry, which then serves to make the virtual location of the dataset freely and openly available. The original data published through the GBIF network is never 'handed over' to GBIF, but instead remains under the authorship of the originator of the data. GBIF maintains a Data Portal (http://data.gbif.org) which facilitates discovery and access to data indexed from published datasets in the requisite formats.

Data publishing through the GBIF network follows a series of clear steps, shown in Figure 2. Each of these steps is described in more detail in the subsequent sections of this document.

Principles of best practice for publishing EIA-related biodiversity data

The best practice guidelines described in this document aim to enhance biodiversity-inclusive EIA and improve access to, and availability of, EIA-related primary biodiversity data.In using the term 'best practice', it is acknowledged that the principles and procedures outlined here represent the best possible practice based on current knowledge and technology, and that future improvements may be possible as this field of work develops further.

There are six key principles that must be applied at all stages of the data publishing workflow to underpin best practice: accuracy, precision, fitness for use, effectiveness, efficiency and transparency (Chapman, 2005a).

Accuracy: refers to how correct the data are. For example, is the organism correctly identified? Or, does the locality information match with the known distribution of the organism? If the organism is incorrectly identified, then the accuracy of this information is low. If it is correctly identified, the accuracy of the data is high.

Precision or resolution: refers to the exactness or level of detail of the data. For example, if an organism is identified only to the level of family, the precision of this record is low, even if it is

accurate (correct). If the organism is identified to subspecies level (or the lowest possible taxonomic rank), then the precision of the data is high. Similarly, for occurrence data, the precision is low if only the broad area of occurrence is given, but the precision of the data would be high if exact co-ordinates are supplied.

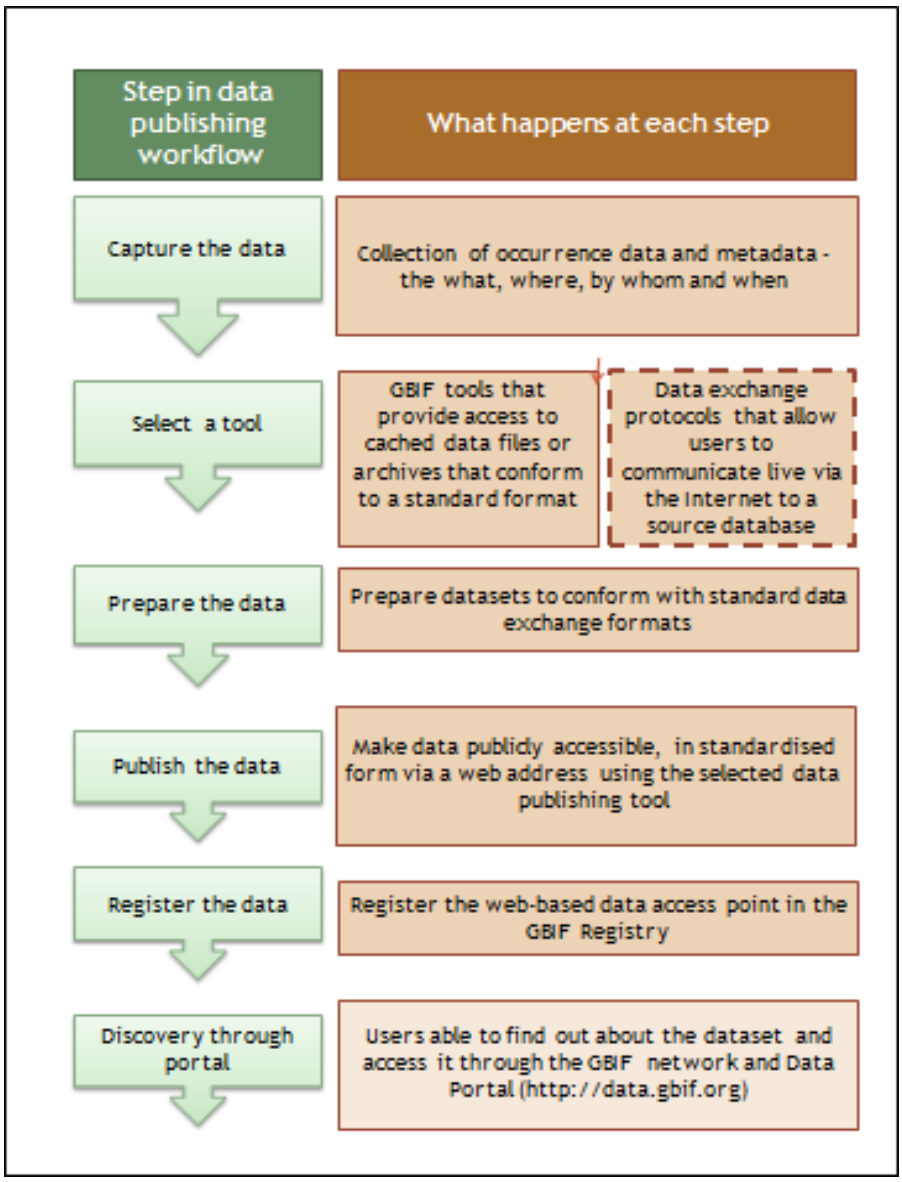


Figure 2: The data publishing workflow (adapted from GBIF, 2011b)

Quality, or 'fitness for use': In the context of this guide, data are described as 'fit-for-use' or 'potential use' (Chapman, 2005a), if they are suitable for the intended use in EIA and subsequent decision-making about development. Data of low accuracy and low precision are poor quality data that will not be fit for use. High quality data are both accurate and precise, and should be

comprehensive, complete, up to date, easy to access and interpret and consistent with other sources. GBIF (2009) describes 'fitness for use' as the suitability, effectiveness or usefulness of GBIF-mediated data in delivering accurate, authenticated, replicable and scientifically valid data for analysis and forecasting in conservation and management of natural resources. Whilst GBIF promotes the use of the highest quality data possible, it is recognised that the quality of data may vary from one application to the next, and that some data may be adequate in one context but not in others.

Effectiveness: this is the likelihood that the data, or a method, might have of achieving the intended outcomes.

Efficiency: relates to the ratio of output to input.

Transparency: relates to how good the metadata are that describe the dataset (enhancing accessibility), and also has bearing on the fitness for use of the data.

Each of these principles can be applied to the primary biodiversity data themselves, and to the tools, protocols and practices that are employed at each step of the data publishing workflow.

Suggestions for further reading on these topics is provided in Box 5.

Box 5: Further reading on this topic

Cadman, M.J; Petersen, C.; Driver, M.D.; Sekhran, N; Maze, K., and Munshedzi, S. 2010. *Biodiversity for Development: South Africa's landscape approach to conserving biodiversity and promoting ecosystem resilience.* SANBI, Pretoria

Chapman, A. (2005a). Principles of Data Quality, version 1.0. Copenhagen: Global Biodiversity Information Facility. 58 pp. ISBN: 87-92020-03-8. Accessible at http://www.gbif.org/orc/?doc_id=1229&l=en

GBIF website: www.gbif.org

GBIF (2010). Best practice guide for 'Data Discovery and Publishing Strategy and Action Plans' version 1.0. Authored by Chavan, V. S., Sood, R. K., and A. H. Arino. Copenhagen: Global Biodiversity Information Facility, 29 pp. ISBN: 87-92020-12-7. Accessible online at http://www.gbif.org/orc/?doc_id=2755&l=en

GBIF (2011b). 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. Accessible at http://www.gbif.org/orc/?doc_id=2815

IAIA Special Publications Series No.3: Biodiversity in Impact Assessment. Accessible at: www.iaia.org/publications

King, N; Rajvanshi, A; Willoughby, S; Roberts, R; Mathur, V;Chavan, V; and Cadman M. *Improving access to biodiversity for, and from, EIAs – a data publishing framework built to global standards.* In prep.

Rajvanshi, A.; Mathur, V; and Iftikhar, U.A. (2007). Best-practice guidance for biodiversity-inclusive impact assessment: a manual for practitioners and reviewers in South East Asia. CBBIA-IAIA Guidance Series, Capacity Building in Biodiversity and Impact Assessment (CBBIA) Project, International Association for Impact Assessment, North Dakota, USA.

Section 3: EIA Processes and Biodiversity Data

This section of the guide provides a brief explanation of the generic EIA process, the types of baseline data used and generated through EIA, and the sources from which these data can be obtained. It provides an overview of the challenges facing EIA-related data publishing and outlines the incentives for, and benefits of, publishing the primary biodiversity data gathered during EIA.

The EIA process

Environmental Impact Assessment is a pro-active planning and decision-support tool used by governments and other stakeholders to assist them in assessing the viability of proposed development activities. It is the principal tool through which the economic, social and ecological impacts of a development proposal can be assessed, enabling a planning decision on whether the proposed development should be approved or not. If approved, the EIA can help identify measures for improving the economic, social and ecological viability and sustainability of the proposed development.

Since this tool first came into legislated use in the USA in 1969, the number of EIAs conducted across the world has risen steadily. In India and South Africa, both developing countries with pressing needs for economic growth, there are currently over 2,000 and 5,000 EIAs, respectively, carried out per year (King *et al.*, *in prep.*). Although the legislation governing EIA may vary from one country to another, and details of the process or names given to reports may differ, there is a universally accepted, generic step-by-step process that each EIA must follow (see Figure 3).

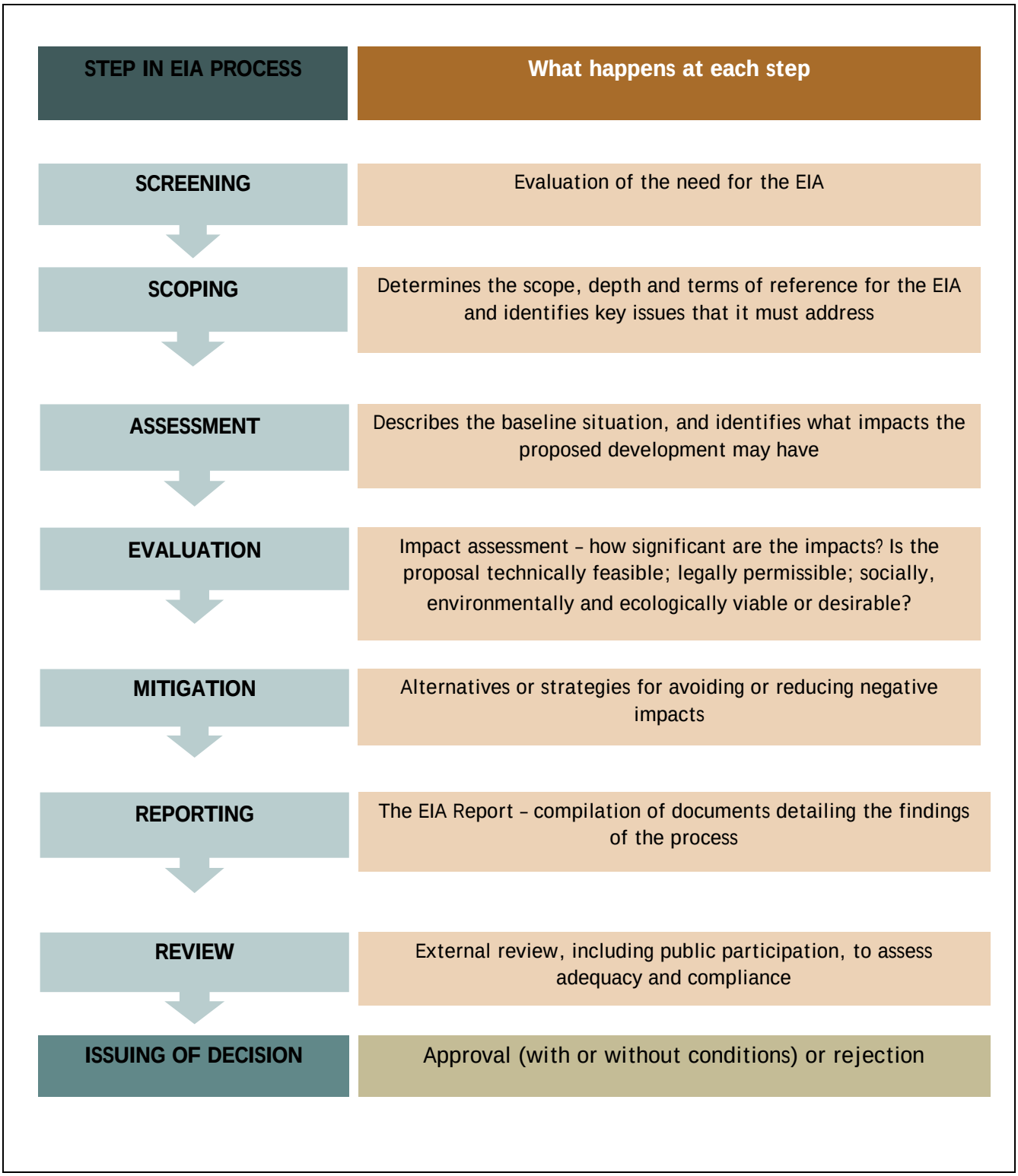


Figure 3: The generic Environmental Impact Assessment process

Impact assessment holds great potential to ensure that biodiversity values are recognised and taken into account in decision-making about land-use and development, and almost every EIA uses and generates biodiversity records of some sort. Historically, however, global experience

reflects that the treatment of biodiversity in EIAs has generally been inadequate (King *et al*., *in prep*.). To help implement the provisions of the CBD (Box 6), the International Association for Impact Assessment (IAIA) has developed a set of guiding and operating principles for integrating biodiversity into EIA (IAIA website: www.iaia.org) and a best practice manual for biodiversity-inclusive EIA has been published for practitioners and reviewers in South East Asia (Rajvanshi, *et al*., 2007).

Box 6: CBD provisions for biodiversity-inclusive environmental impact assessment

Each Contracting Party, as far as possible and as appropriate, shall:
(a) Introduce appropriate procedures requiring environmental impact assessment of its proposed projects that are likely to have significant adverse effects on biological diversity, with a view to avoiding or minimizing such effects and, where appropriate, allow for public participation in such procedures;
(b) Introduce appropriate arrangements to ensure that the environmental consequences of its programmes and policies that are likely to have significant adverse impacts on biological diversity are duly taken into account; […]. (**CBD, Article 14.1**)

Types of biodiversity data used and generated in EIA

The types of baseline biodiversity data required during the screening, scoping and assessment phases of EIA include: ***taxonomic data*** (the correct scientific names of the organisms that occur at the site); ***occurrence and distribution data*** (precise records of occurrence, distribution and abundance); and ***ecological data*** *(*including information regarding habitats and ecosystems, the role of particular biodiversity elements, rarity and conservation status).

These data are used to generate a dataset specific to the project site. Much of the biodiversity information contained in final EIA reports is synthesized or interpretive biodiversity information (King, *et al*., in prep). This includes:

- species checklists;
- bio-spatial data about the project site;
- indigenous knowledge of biological resources;
- ecosystem functions, services and benefits;
- existing and potential pressures on biodiversity at the site (e.g. invasive alien species).

All of the interpretive data, which are used to assess the sensitivity of the site and its likely response to the proposed development, are based on large volumes of primary biodiversity data.

Sources of biodiversity data for EIA

In general, the collection of biodiversity data for EIA tends to be *ad hoc* and opportunistic, making use of a variety of sources that happen to be available to the practitioner. These include: literature (scientific or popular); floral and faunal inventories; consultation with experts (data based on expert knowledge of a site or species); and specimens in natural history collections. Observational data are also gathered through field studies, but these tend to be 'once-off' and of limited duration, due to the time constraints under which most EIAs are expected to be completed. Environmental assessment practitioners may make use of existing local, national, regional and global data resources where they are aware of them, but these provide patchy geographic and taxonomic coverage and do not serve their data using consistent formats.

GBIF and its participants (see Box 7) provide a rich source of biodiversity data, served in consistent formats that are freely and openly accessible. Since GBIF has 'proven the concept' for global biodiversity data publishing, access and discovery in support of policy-making, it should be a relatively short step to secure uptake of the tools and protocols that are available for publishing EIA-related data. Primary biodiversity data that is published and accessible can be used in other EIAs, and can contribute significantly to the state of knowledge of biodiversity.

Challenges facing EIA data publishing

There are three main sets of challenges facing data publishing: those related to the nature of the data themselves; challenges relating to capacity and technology, and issues of principle and mindset.

Challenges relating to data:

Key challenges relating to the nature of the primary biodiversity data collected during EIAs include:

- *Data standards*: compatibility, integration;
- *Data types* – what data types are needed, and by whom?
- *Data quality* – fitness for use, and by whom?
- *Data volumes* – how many data are enough?

Box 7: GBIF Participants

GBIF works with many participants to mobilise data and put it to use, to improve search mechanisms, data and metadata standards, web services and other components of an internet-based information infrastructure for biodiversity. A GBIF participant can be any country or organisation that has signed the GBIF Memorandum of Understanding (MoU) and that has expressed its intention to abide by its provisions. Voting members are countries that contribute financially to the GBIF budget, with voting rights on the GBIF Governing Board, and associate members are countries or organisations that do not contribute to the GBIF budget, but that are willing to observe the terms of the MoU. Associate members may participate in the deliberations of the Governing Board, but do not have voting rights. As of mid-2011, GBIF had 33 voting countries, and 24 associate countries as Participants. In addition to this, 47 international and regional organizations are GBIF associate participants. The lists of these participants are available at http://www.gbif.org/governance/governing-board/voting-participants/, http://www.gbif.org/governance/governing-board/associate-country-participants/, and http://www.gbif.org/governance/governing-board/other-associate-participants/

Primary biodiversity data collected during EIAs tend to be gathered 'once off', and vary greatly in accuracy, precision and type. The methods used to collect, present, store and archive them are varied and inconsistent, with no standards provided or utilised. Primary biodiversity data often do not even appear in EIA reports, and are seldom captured in any forms that are accessible and re-usable. Specialist reports (which usually contain the biodiversity data) are often submitted in summarised form and may not include full datasets or any supporting field research. This means that the biodiversity findings of many EIAs are often not verifiable. The suite of tools presented in this guide provides ways of overcoming these challenges through:

- requiring the use of standardised templates for data collection and management (see Section 4);
- requiring the inclusion of supporting metadata (see Section 5), that makes it possible to authenticate the data and assess its quality (see Box 8).

Box 8: Additional guidance on issues of data quality

Practitioners seeking more guidance on issues of data quality are referred to theGBIF Training Manual, or to the following GBIF publications:

1. Principles of Data Quality (Chapman, 2005a) – accessible at http://www.gbif.org/orc/?doc_id=1229&l=en
2. Data Cleaning (Chapman, 2005b) – accessible at http://www.gbif.org/orc/?doc_id=1262&l=en
3. Georeferencing – (Chapman and Wieczorek, 2006) – accessible at http://www.gbif.org/orc/?doc_id=1288&l=en
4. Generalising Sensitive data (Chapman and Grafton, 2008)- accessible at http://www.gbif.org/orc/?doc id=1233&l=en

Challenges relating to capacity and technology:

Challenges relating to capacity and technology:

There may be misconceptions regarding the cost of the technology or professional time required to publish EIA-related data. All of the GBIF publishing tools are freely and openly available via the Internet, and they do not require expensive software or hardware to operate, so technology costs should not be a barrier to their use by most environmental assessment professionals.

There may also be concerns that publishing biodiversity data will result in an EIA taking longer and, potentially, costing more in professional fees, but this should not be the case. Environmental assessment professionals are already collecting large volumes of occurrence data, but currently use heterogeneous methods of data collection and variable reporting formats. Using GBIF tools for data collection and publishing will help them capture the data more efficiently, in formats that make it possible to exchange data, without significant investment of time or resources.

It is important to acknowledge that capacity for data management and information technology varies considerably between institutions and amongst individuals. A variety of tools and procedures have been developed specifically to accommodate these variable levels of capacity.

Challenges relating to mindset:

Potential barriers to publishing biodiversity data from EIA include concerns about intellectual property rights and data security.

- *Intellectual property rights (IPR)**:*** There is an urgent need to eradicate the misconceptions around intellectual property rights and data 'ownership', often put forward as a reason by consultants or developers for not sharing the biodiversity data generated in EIA. The GBIF suite of tools for publishing biodiversity data allow for free and open access to data, with a mechanism for due attribution and credits to data publishers. This means that data ownership rights are respected, whilst promoting free and open sharing of data (King, *et al.,* in prep).

- *Data security*: There are legitimate concerns around making the exact locality data of rare, threatened, endangered or otherwise sensitive species openly available. The practices outlined in this guide include mechanisms for restricting or managing access to sensitive data, without compromising the utility and transparency of the data publishing system (Chapman and Grafton, 2008).

Overcoming the barriers described above requires a change of mindset. Demonstrating the benefits that EIA data publishing holds for improving the effectiveness of the EIA process will help bring about that change.

Incentives for, and benefits of, EIA biodiversity data publishing

The most compelling arguments for publishing EIA-related biodiversity data through the GBIF network include:

- Ready access to free primary biodiversity data in standardised formats will improve the quality of EIA and streamline the EIA process (offsetting the few additional steps required to publish the data generated by an EIA);
- Assessing cumulative impacts will be made easier, as comparable data will be available from other EIAs that might have been conducted in the area (or in other areas); this also makes it possible to profile pressures based on conservation status of particular species or habitats;
- The reliability, credibility and transparency of the EIA process will be enhanced;
- Continual contribution to and improvement of the local, national and global biodiversity databases on which EIA depends, will promote biodiversity-inclusive EIA, and should lead to more informed decision-making;
- Use of the publishing infrastructure provided through GBIF and its national and institutional nodes, provides opportunities for practitioners to develop their public profile, as authors of the datasets published through GBIF are acknowledged.

Figure 4 illustrates how EIA practice can be enhanced through the increased availability of published, verifiable biodiversity data.

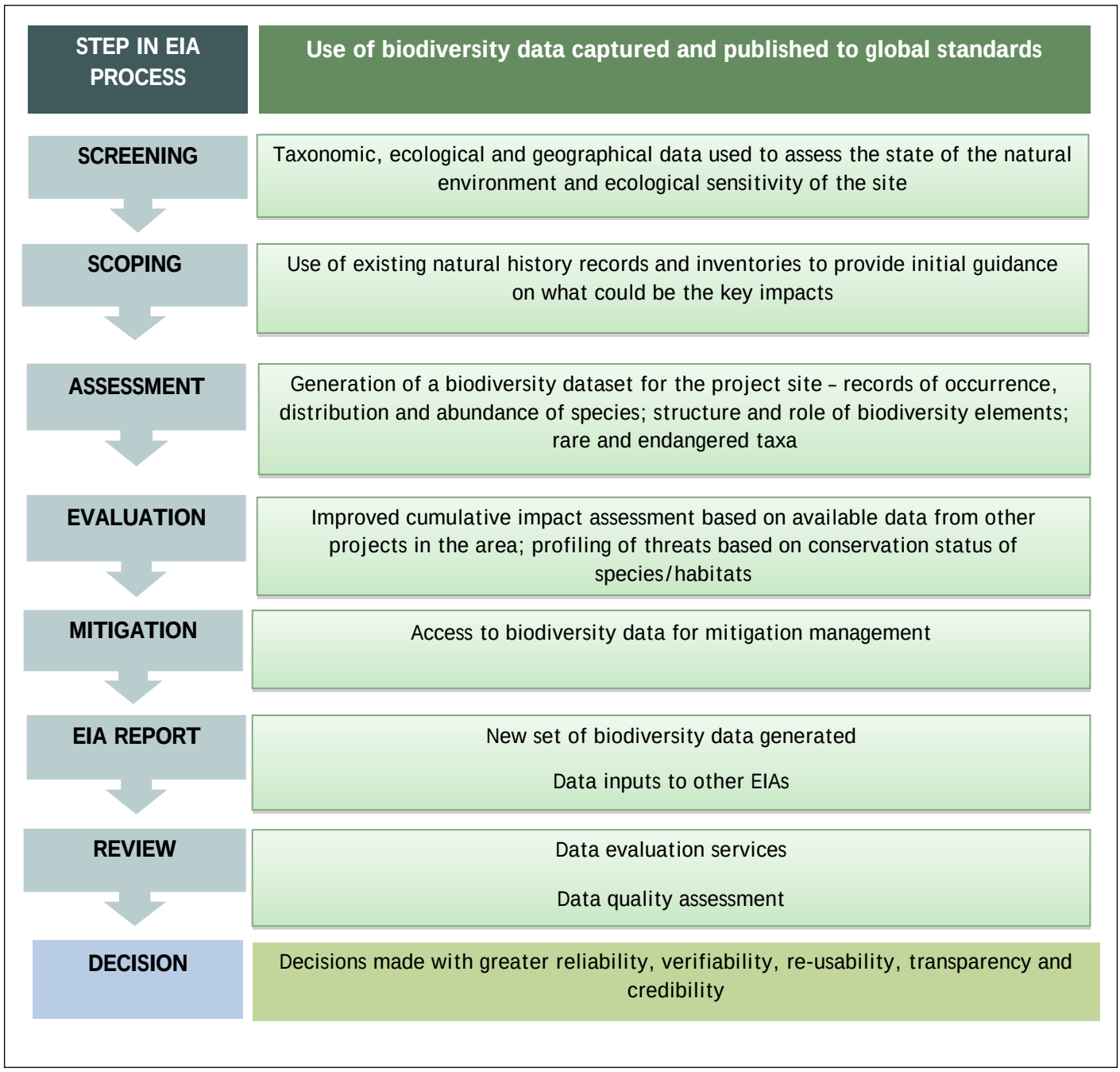


Figure 4: Enhancing EIA practice through the biodiversity data publishing framework, showing anticipated benefits at different steps of the EIA process (from King *et al*., in prep).

Section 4: How to capture or manage primary biodiversity data

This section of the guide describes the current practices used for capturing data in EIA processes, and outlines simple tools that GBIF provides for EIA practitioners to capture data in standard formats, thus making their datasets suitable and available for sharing.

Current practices for capturing biodiversity data during EIA processes

There is no standardised format that EIA practitioners use during data capture, making it very difficult for data to be compared between sites, or to be captured and integrated with national and global biodiversity databases.

Several GBIF tools exist for capturing, publishing and discovering primary biodiversity data, (digitised or otherwise). This is work that is usually carried out by collections managers (in herbaria and museums) or their data management staff, or by specialist researchers who possess the specific technical or technological capacity for doing this work. EIA practitioners, however, are often neither biodiversity experts, nor well versed in the technicalities of biodiversity informatics or the use of technically complex data management tools. GBIF has, therefore, provided some simple tools that: (a) make it easier for EIA practitioners to collect and manage primary biodiversity data; (b) improve the consistency and utility of data collection, and (c) collect the data in a form that is suitable for publishing using GBIF infrastructure.

Best practice for capturing data during EIA: using GBIF Excel Templates

GBIF provides a set of pre-configured Excel spreadsheet files that serve as templates for capturing occurrence data (primary biodiversity data), metadata and simple species checklists. These spreadsheets are simple tools that provide a common format and standard for collecting data, using consistent terminology. The spreadsheets include a large number of data fields (or data elements), described using standarised terms, into which data (or data values) can be captured. Although it is recommended that as many fields as possible are used in order to maximise the quality of the data, there is a minimum set of six compulsory fields that must be filled in. Essential data elements include: taxon name, latitude/longitude, date/time, name of data collector, name of identifier, reference or link to specimen or photograph. Using these data elements means that the data are not only precise, but are easier to authenticate, thus increasing the confidence with which they can be used.

There are three GBIF Excel spreadsheet templates available:

(i) *Metadata template*: suitable for composing a metadata document, i.e. including all the information needed to describe the dataset. This template is available for download and use at http://tools.gbif.org/spreadsheet-processor/templates/metadata/metadata-1_v1.xlsx

(ii) *Occurrence template*: suitable for collating or capturing primary biodiversity (species occurrence) data. These data can be specimens housed in natural history collections or species occurrence observations recorded in the field. This template is available for download and use at http://tools.gbif.org/spreadsheet-processor/templates/occurrence/occurrence-1_v1.xlsx

(iii) *Species (checklist) template*: suitable for recording and storing simple annotated species checklists. Currently, GBIF recommends three types of checklist (species) templates, depending on the taxonomic information that one is using. These templates are available at: http://tools.gbif.org/spreadsheet-processor/templates/checklist/checklist-1_v1.xlsx (Template one, using a 2-column classification); http://tools.gbif.org/spreadsheet-processor/templates/checklist/checklist-2_v1.xlsx (Template 2, allowing use of taxonomic categories outside of the core Linnaean ranks – e.g. Super Class); and http://tools.gbif.org/spreadsheet-processor/templates/checklist/checklist-3_v1.xlsx (Template 3, using the compulsory Linnaean ranks of Kingdom, Phylum, Class, Order, Family, Genus, Species).

The spreadsheet templates that will be of most use to EIA practitioners are the metatdata and occurrence templates.These spreadsheets are easy to use and include inline help, which is accessed by hovering the cursor over spreadsheet cells with red upper-right corners .There are also a number of GBIF User Guides (GBIF 2011c, GBIF 2011d) that provide step-by-step assistance for use of the Excel spreadsheets (See Box 9).

Once the spreadsheets have been filled in, a web-based service called the GBIF Spreadsheet Processor is used to check the validity of the data and to convert them into a standardised format suitable for sharing. This is described in Section 5. (The EIA biodiversity data publishing facility described in Section 5 also makes use of the GBIF Excel spreadsheets for uploading data).

An important note on use of the Excel spreadsheet templates: Do not add, delete or move any columns or cells in the templates. This will break the validation process and result in error from the spreadsheet processor described in Section 5.

Box 9: Additional guidance for using GBIF Excel spreadsheet templates

GBIF (2011c). GBIF Spreadsheet Templates: User Guide. http://www.gbif.org/orc/?doc_id=2823&l=en GBIF (2011d). Publishing species checklists: Best Practices. http://www.gbif.org/orc/?doc_id=2814&l=en

Metadata Template: http://tools.gbif.org/spreadsheet-processor/templates/metadata/metadata-1_v1.xls

Species Occurrence Template: http://tools.gbif.org/spreadsheet-processor/templates/occurrence/occurrence-1_v1.xls

Checklist Templates:

http://tools.gbif.org/spreadsheet-processor/templates/checklist/checklist-1_v1.xls ; http://tools.gbif.org/spreadsheet-processor/templates/checklist/checklist-2_v1.xls ; http://tools.gbif.org/spreadsheet-processor/templates/checklist/checklist-3_v1.xls

Section 5: Tools for EIA biodiversity data publishers

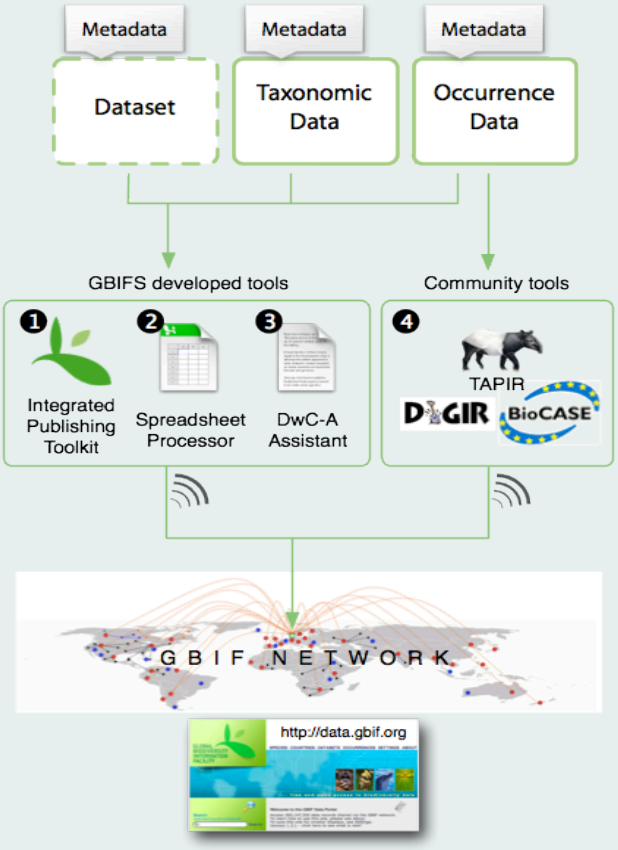
To 'publish' data means to make them publicly available on the Internet, followed by registering the access point (URL or web address) with the GBIF Registry.

Occurrence data may be published through GBIF via two different routes: (a) via access to complete or partial datasets provided as cached data files or ***archives*** that conform to a standard format; or, (b) through the use of biodiversity exchange protocols that allow users to communicate 'live' via the internet to a source database. Exchange protocols are the traditional way data have been published to date, and although it remains an option, it is not the preferred method, and is not well-suited to publishing EIA-related primary biodiversity data (see Box 10).

Before the data can be published, they must first be 'prepared' for publishing – that is, they must be converted into a standardised format that is supported by the GBIF network. GBIF provides a rich array of support and tools for customising data formats and for publishing primary biodiversity data in compliance with global standards (Option (a), above). These standards and tools, and their use, are described briefly in this section of the guide.

Box 10: Tools for publishing biodiversity data

GBIF provides a suite of tools that can be used to publish Standardised data files, including: the GBIF Integrated Publishing Toolkit, the GBIF Spreadsheet Processor and the Darwin Core Archive Assistant. These tools have largely replaced the use of other community tools (data exchange protocols) that were traditionally used for publishing and making data accessible.



A note on data exchange protocols:

Data exchange protocols define a particular process for interacting with a database, and then data are returned to the user in a standardised response format. There are three data exchange protocols that GBIF can accept (TAPIR, DiGIR and BioCASE) but that are no longer the preferred method for publishing biodiversity data to the GBIF network (GBIF, 2011b). These protocols are not suitable for EIA data publishing as they either use older data standards that are no longer supported (TAPIR), or they are no longer under active development (DiGIR), or they focus on data from a restricted geographic area (BioCASE), and are mostly suited to taxonomic data only. They also require the use of specialised software called “wrappers”.

Tools and Infrastructure for EIA Data Publishing

To be published via the GBIF network, primary biodiversity datasets must first be converted to a standardised format, known as Darwin Core Archive (See Box 11). EIA data publishers do not have to generate Darwin Core Archive files themselves, unless they choose to do so, in which case they can make use of a tool called the Darwin Core Archive Assistant (See Appendix 1 or visit http://tools.gbif.org/dwca-assistant/).

GBIF provides a number of simple tools that can be used to generate the Darwin Core Archive files, and publish them via the GBIF network, including:

- GBIF Excel Spreadsheet Templates and the GBIF Spreadsheet Processor;
- the GBIF Integrated Publishing Toolkit.

Two further infrastructures that will be available by the end of 2011 are the EIA Primary Biodiversity Data Publishing Facility (being developed by SANBI-SABIF South Africa), and Data Hosting Centres (DHCs), that will be piloted by DanBIF, in Denmark and EWT South Africa. These are described briefly below.

Box 11: A note on data standards and formats

Adherence to standards is essential to facilitate data sharing, but publishers of EIA-related biodiversity data do not have to become bogged down in the technical detail underpinning these standards. Use of good application tools can ensure compliance with standards whilst providing an interface that allows the data publisher to focus on data content, rather technical details of the system.

There are, however, two standards of which publishers of biodiversity data should be aware, as they underpin the publishing process. The first of these is the **Darwin Core (DwC)** body of standards, which is a standardised glossary of terms for describing and documenting the occurrence of species. GBIF has used the Darwin Core guidelines to develop a standardised format that can be used to publish both species occurrence data and checklist data. This format is called the **Darwin Core Archive,** and is the preferred means for publishing primary biodiversity data. An expanded explanation of these standards is provided in Appendix 1 at the end of this guide, or interested readers can download the relevant user guides from the GBIF website. Key references are provided at the end of this section.

The GBIF Spreadsheet Processor

The Spreadsheet Processor is a web based application that transforms pre-configured Excel spreadsheet files for occurrence data or metadata (described previously, in Section 4) into GBIF-supported formats (GBIF 2011c). The Spreadsheet Processor accepts the completed Excel spreadsheet templates as a web form or as an email attachment. It then performs a series of data checking (validation) and transformation steps, and then returns a validated Darwin Core Archive file to the user, suitable for publishing via GBIF (or other biodiversity networks that support this format). This process is illustrated in Figure 5.

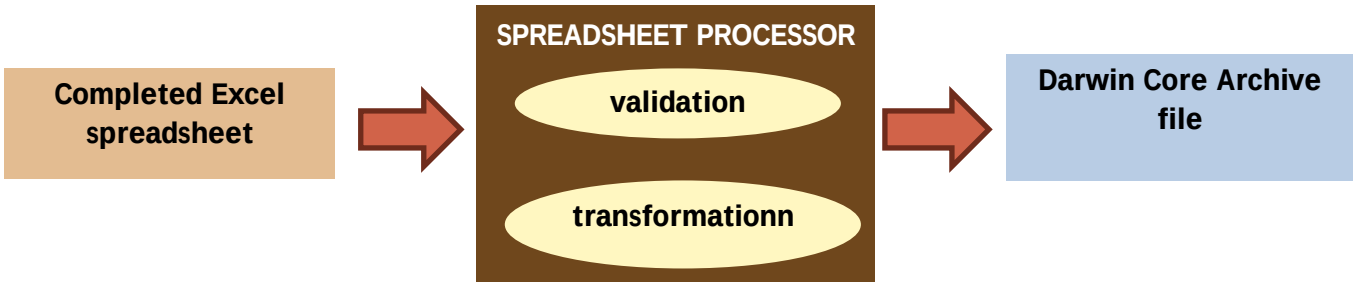


Figure 5: Generating a Darwin Core Archive using the GBIF Spreadsheet Processor

The GBIF Integrated Publishing Toolkit (IPT)

The Integrated Publishing Toolkit (IPT) is a software platform developed by GBIF to facilitate easy and efficient publishing of biodiversity data on the Internet. It can be used to manage and publish primary occurrence data, taxonomic checklists and resource metadata. To use the IPT, data must already be digitised as existing Darwin Core Archives or as any delimited text files (e.g. text files using comma or tab-separated values). The IPT also supports automatic registration of the dataset.

Currently, data publishers wishing to use the GBIF IPT need to install and host a local version of the IPT at their home institution. In future, it will be possible to access the IPT via a GBIF-endorsed Data Hosting Centre (See Box 12), and this will be the easiest option for EIA practitioners to use.

A further data publishing option that will become available in future is the EIA Primary Biodiversity Data Publishing Facility that is being developed in South Africa. This facility will provide a simple, custom-built mechanism for EIA practitioners both to publicise their work and to publish biodiversity-related datasets to GBIF standards. The facility will not only provide a low-cost, easy-to-use pathway for generating Darwin Core Archives for publishing, but will also provide a range of other data and services relevant for EIA (e.g. mapping services, ecological data). It will also provide an easy way for other stakeholders to keep up to date on EIA-related activity through emailed notifications, and will allow users to comment on the reports and datasets that are published on the system. This feedback will help other users to evaluate the quality and relevance of the data, and encourage interaction with biodiversity specialists working in the EIA sector.

Anyone will be able to register as a user of the EIA Primary Biodiversity Data Publishing Facility, which is accessible via the SANBI Biodiversity Advisor Portal (www.biodiversityadvisor.sanbi.org.za). Once registered, users can create projects and project-specific datasets. These are described using Darwin Core compatible metadata, and are spatially linked to the footprint of the proposed development or specific area where the biodiversity data were gathered. Using either the GBIF Excel spreadsheet templates, or plain text files, users may then upload a file containing the report or assessment species list; the facility will transform the latter into a standard Darwin Core Archive file, as well as provide an additional file containing information about the species listed (such as their threat status, and correct taxonomy). These files are then available for any user of the system to download.

This facility will be ready for South African users in late 2011. Once it has been fully tested, it can be customized for use in other countries.

Box 12: Data Hosting Centres

These will, in future, provide a service to scientists and practitioners who generate large volumes of primary biodiversity data, but who are unable to share or publish it directly themselves due to a lack of suitable informatics infrastructure, or the skills-sets required to use this. The Data Hosting Centres will provide an ideal 'one stop shop' through which EIA data publishers can capture, prepare, publish, register and archive their data. Once the practitioner has submitted data in the required format, the rest of the data publishing process can be channelled through the Data Hosting Centre, although the originator of the data always retains authorship and control of the dataset. There are currently GBIF-supported Data Hosting Centres under development in Denmark and South Africa.

Criteria for selecting a data publishing solution

If the GBIF IPT, or community tools such as data exchange protocols (TAPIR, BioCaSe or DiGIR) are used, the publishing functionality is built into the system. For EIA data publishers, who are more likely to use the Spreadsheet Processor (or the EIA Primary Biodiversity Data Publishing Facility, when this comes into full operation), the datasets need to be manually posted on a website. There are different options for doing this.

Important criteria to apply in selecting the most appropriate publishing solution for your data include:

(i) Whether the data have been digitised or not, and how many datasets you need to manage; and,

(ii) The level of technological and data management capacity available to you.

GBIF Excel spreadsheets and the Spreadsheet Processor offer simple solutions for preparing data for publishing in situations where data have not yet been digitised (as is likely to be the case with many EIAs). An attractive feature of using these particular tools is that they do not require any specialised software, and they offer easy and quick options for managing smaller numbers of datasets. Once the DwC-Archive has been generated, it can be published using either the GBIF Integrated Publishing Toolkit (or,in future, via a Data Hosting Centre) (see Figure 6).

In cases where data have already been digitised or are in a relational database, it is possible to use the Darwin Core Archive Assistant to generate the DwC-A file, if adequate data management and IT capacity are available (See Figures 6 and 7).

The levels of IT and data management capacity available to a practitioner will also influence the choice of publishing pathway. When both the data management and IT capacity are low, the best option is to use the Excel spreadsheet templates and Spreadsheet Processor to generate the Darwin Core Archive, followed by publication via a Data Hosting Centre (once these are operational). This is the situation that is most likely to apply to EIA data publishers (See Figure 9). If the data management capacity is low, but the IT capacity is high, then the institution or practitioner may choose to install the Integrated Publishing Toolkit software itself, and publish the dataset directly using the IPT. If both the data management and IT capacity is high, then the publisher may choose to generate its own Darwin Core Archives and publish them by installing the IPT to its own system. The relationship between IT and data management capacity in determining data publishing solutions is illustrated in Figure 7, below.

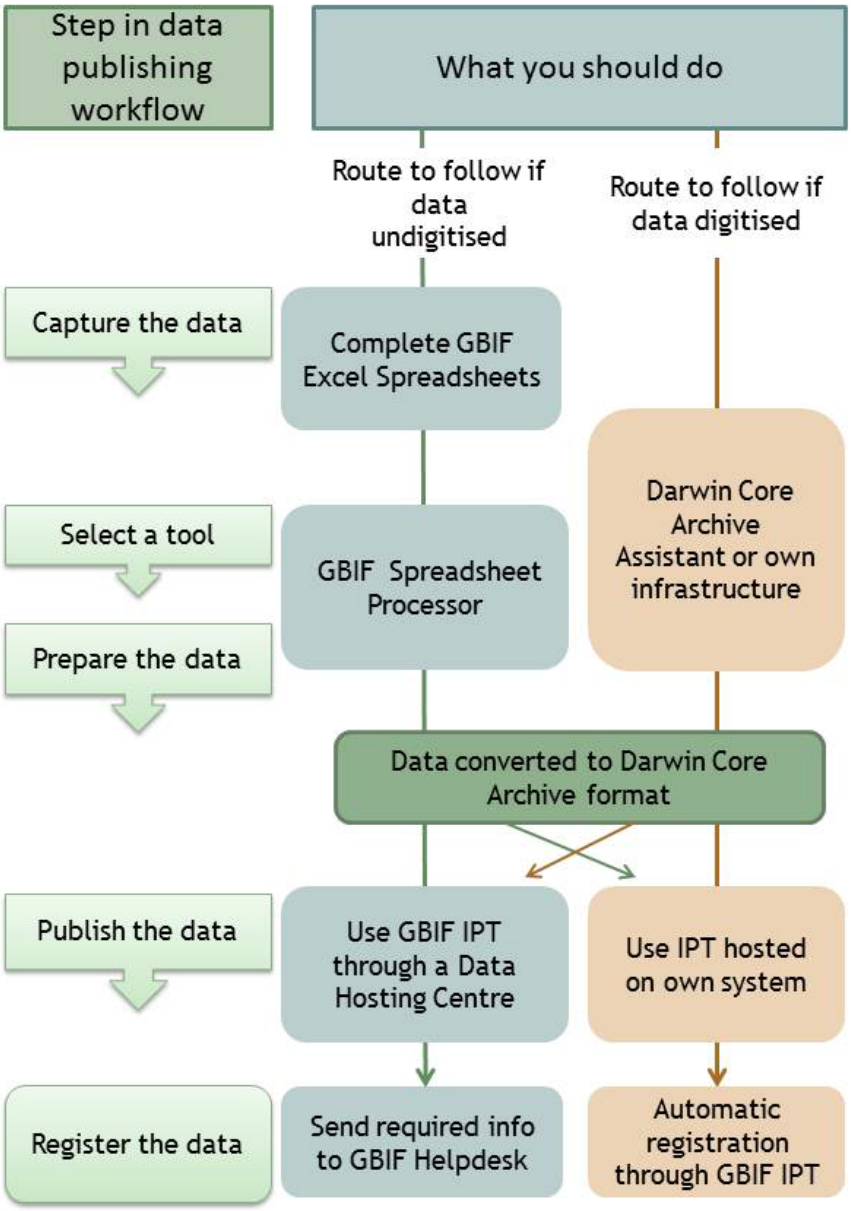


Figure 6: Options for preparing and publishing primary biodiversity data

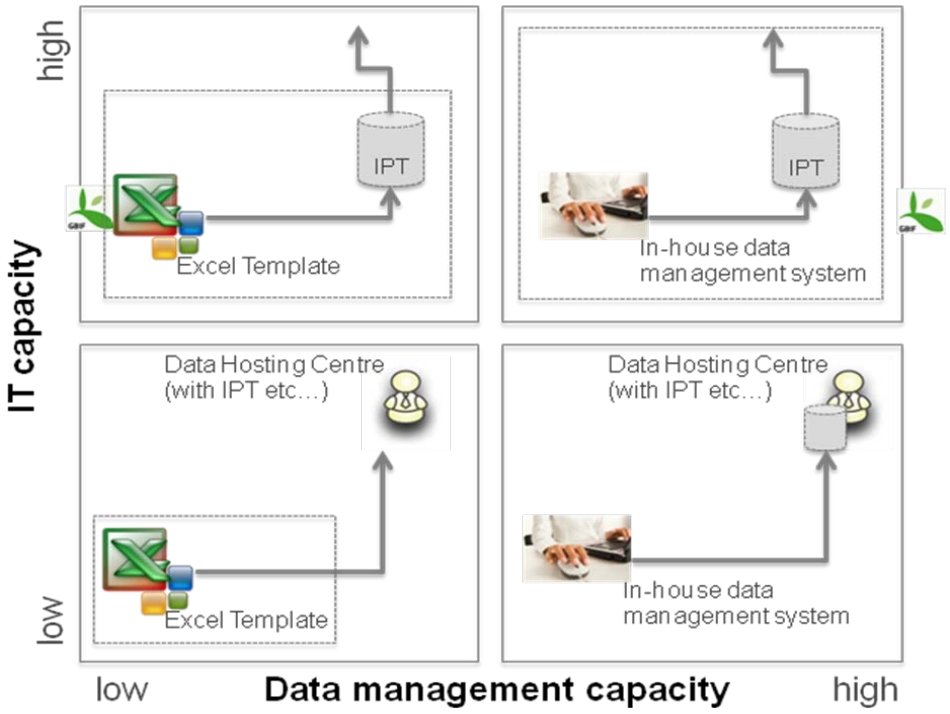


Figure 7: Data publishing options in relation to IT and data management capacity.

Publishing metadata

Documentation describing datasets (resource metadata), is an essential part of any data management system, and allows data users to assess the quality of the dataset and its fitness for use (GBIF, 2011f). It is not possible to publish primary biodiversity data on the GBIF network without the accompanying metadata.

Metadata standards and data elements

As with the publication of ocurrence data, there are clearly defined standards, known as a metadata profile, with which a metadata document must comply. These standards for metadata files are described as a Metadata Profile, which is explained in detail in the GBIF Metadata Profile Reference Guide (GBIF, 2011f).

The types of data elements (or fields) described in a metadata document include, *inter alia:* who collected the data, where and when (basic information); geographical and temporal coverage of the data; sampling methods; physical data; project data; associated parties; where voucher specimens are housed. There are five compulsory data fields, designed to allow a prospective

end-user of the data to discover the name and a brief description of the dataset, details of the key contact person and information regarding data mangement rights. Once the compulsory data fields have been filled in, the file can be saved, then modified and updated at any later time (GBIF, 2011f).

Options for authoring and publishing metadata

As with occurrence data, there are three ways of writing (authoring) a metadata file:

- Using GBIF Excel spreadsheet templates;
- Using the GBIF Integrated Publishing Toolkit;
- Modifying an existing sample document.

If occurrence data are being published using Excel spreadsheets or the GBIF IPT (either directly or via a Data Hosting Centre, when these are available), then there is a built-in metadata authoring function that can be used to write the accompanying metadata document, following the same general procedures as are outlined in Table 1, above. Using the GBIF IPT makes it easier to handle large numbers of metadata documents, but requires the data to be already in a digitised form, and involves loading the IPT software onto your computer or data management system. For EIA practitioners, it is probably quicker and easier to use the GBIF Excel spreadsheet template for metadata (http://tools.gbif.org/spreadsheet-processor/templates/checklist/checklist-1_v1.xls), and the Spreadsheet Processor (which in South Africa will also be accessible via the EIA Primary Biodiversity Data Publishing Facility) and then to publish it through a Data Hosting Centre.

There are three GBIF Guides that provide assistance with the steps and procedures to follow for authoring metadata files. These include:

- GBIF spreadsheet templates: User Guide (http://www.gbif.org/orc/?doc_id=2823&l=en)
- GBIF Metadata Profile: Reference guide (http://www.gbif.org/orc/?doc_id=2820&l=en)
- GBIF Metadata profile: How-to-Guide (http://www.gbif.org/orc/?doc_id=2821&l=en)

Where to find assistance

GBIF has a series of User Guides available online, to assist with the publication of primary biodiversity data and its associated metadata using the GBIF tools.They provide detailed, step-by-step instructions in the use of all the key tools used at different steps in the data publishing process, so this detail is not repeated here. Instead, Figures 8 - 10 provide a summary of the data publishing workflow for metadata (Figure 8), occurrence data (Figure 9) and checklists (Figure 10). These summaries indicate the steps involved in the publishing process and lists the GBIF Guides that provide assistance at each step.

Should further assistance be required, the GBIF Participant Node in your country or region can be contacted (see Box 13).

Table 1 provides a summary of which publishing tools are available, when they should be used and what the EIA data publisher has to do to use the tool.

Table 1: Quick Guide to publishing occurrence data using GBIF-supported tools

Publishing tool	When it should be used	How to use this tool
GBIF Excel Spreadsheet templates	✓ When occurrence or species checklist data is not already digitised and you need to generate a DwC Archive OR when you need to generate a metadata document ✓ When publishers want a quick and easy pubishing solution that does not require high data management or IT capacity ✓ When there are relatively few datasets involved	✓ Access spreadsheets by logging on to the GBIF website and downloading the appropriate template (http://tools.gbif.org/spreadsheet-processor/) ✓ Populate the spreadsheet with your data, using at least the compulsory data fields; make use of the inline help by hovering the cursor over the cells with red upper corners, or use the Guides shown in Fig. 6.3 ✓ Upload the completed template to the Darwin Core Spreadsheet processor (http://tools.gbif.org/spreadsheet-processor/)
GBIF Spreadsheet	✓ When you need to convert a completed GBIF Excel	✓ Access the Spreadsheet processor at http://tools.gbif.org/spreadsheet-

Processor	Spreadsheet Template into a DwC-Archive file that is suitable for publishing	processor/ and follow the instructions for uploading and processing the filled in Excel Spreadsheet ✓ A DwC archive file will be returned to you and saved in the same folder as your template ✓ Publish the DwC-A yourself by posting it on a web server and registering the URL with GBIF OR send it by FTP or email to a Data Hosting Centre for publication via the GBIF IPT
GBIF Integrated Publishing Toolkit	✓ When you need to publish occurrence data, taxon data and associated metadata that are already digitised ✓ When you have an already-created Darwin Core Archive (e.g. that has been created using the Spreadsheet Processor), OR when you need to generate the Darwin Core Archive from pre-digitised data ✓ When you need to validate, publish and register DwC-A files ✓ when large numbers of datasets are being managed	✓ To work directly with the IPT you must install the IPT software on your computer; information on installing and operating the IPT can be found in the IPT User Manual or on the IPT project website (visit http://code.google.com/p/gbif-providertoolkit/ ✓ To generate a DwC-A file using the IPT, follow the instructions provided in the GBIF Darwin Core Archive How-to Guide (GBIF 2011e) ✓ Publishing (and registration) are built in automatically
EIA Primary Biodiversity Data Publishing Facility	✓ A facility that allows you to generate DwC Archives and that provides access to biodiversity data, mapping services and other functions that are geared specifically for EIA practitioners	✓ Register as a user (many functions on this system are only available to registered users) ✓ Create a user profile and create a project ✓ Within a project, create your dataset(s) using the GBIF Excel Spreadsheets. (The Facility will also accept plain and delimited text files)

	✓ Soon to be available in South Africa (SANBI)	✓ Once the DwC Archive is returned to you, either post the file on a website yourself, or do so using the GBIF IPT, OR send the dataset to a Data Hosting Centre for publsihing and registration.
Data Hosting Centres	✓ A one-stop shop at which you can download spreadsheet templates, publish DwC Archives and deposit data, if you do not have the time, resources or skills sets to install and work with the GBIF IPT yourself	✓ Access the IPT either at DanBiF Data Hosting Centre; the Endangered Wildlife Trust (EWT) Data hosting Centre, or SANBI (www.biodiversityadvisor.sanbi.org.za) ✓ Contact helpdesk@gbif.org for further information, or contact your Particant Node (see Box 13) DanBiF: Currently under revision EWT: available at www.ewt.org.za/foryou/datasharing.aspx…..

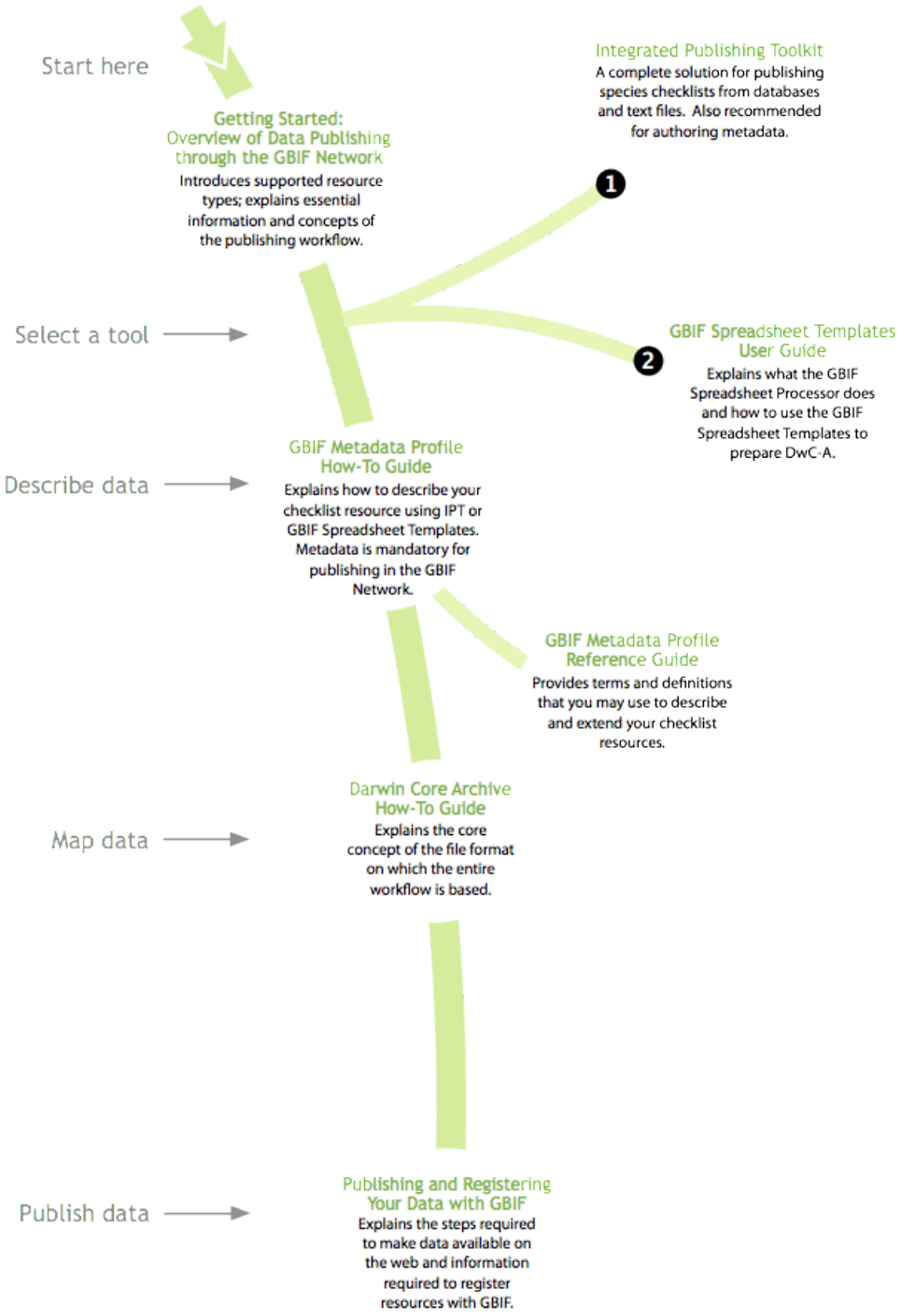


Figure 8: The metadata publication workflow together with links to relevant guides.

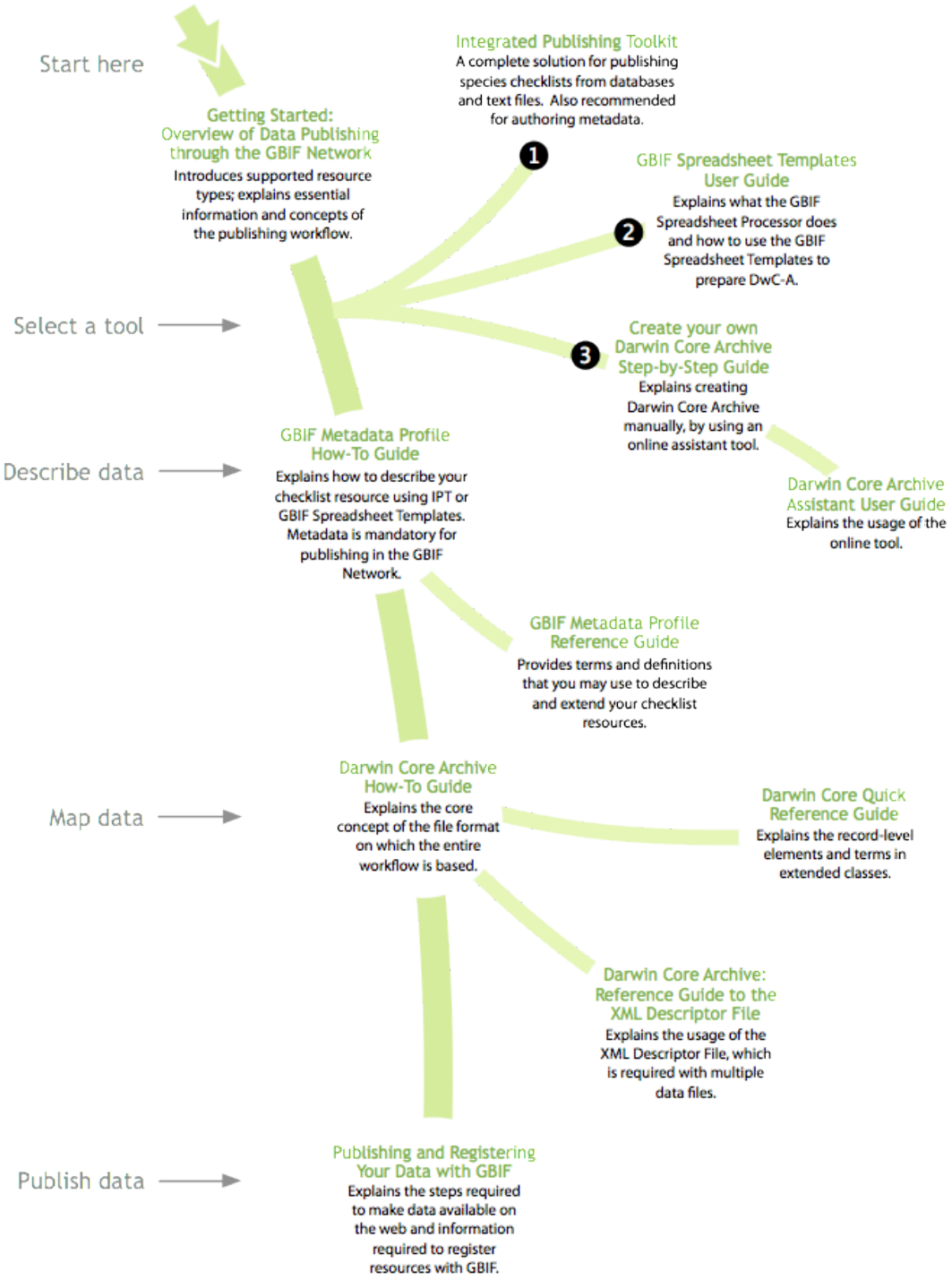


Figure 9: The occurrence data publishing workflow together with links to relevant guides.

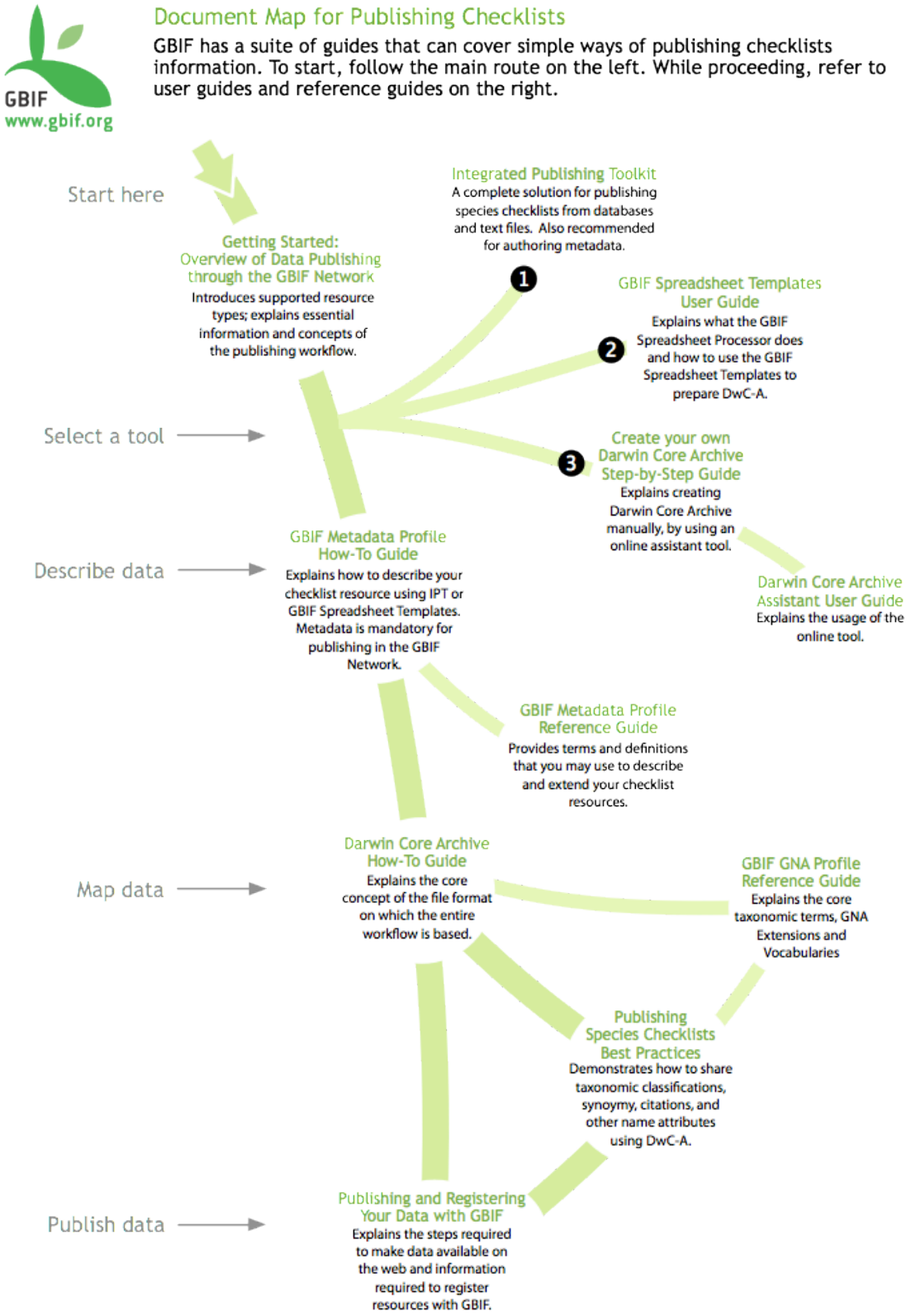


Figure 10: The checklist publication workflow together with links to relevant guides..

Section 6: Registering your data

Registration is the final step in the data publication process. An entry for the dataset URL is made in the GBIF registry that enables the resource to be discoverable and accessible. There are three options for registration of datasets:

(a) Using the GBIF Integrated Publishing Toolkit;
(b) Using the Spreadsheet Processor; and,
(c) Using other tools.

How to register your dataset

The GBIF IPT supports automatic registration in the GBIF network (see the online manual for the IPT). Using the Spreadsheet Processor or other tools there is no automatic registration. An email must be sent to helpdesk@gbif.org with the following information:

1. Dataset title
2. Dataset description
3. Technical contact (the person to be contacted in matters regarding technical availability or resource configuration issues on the side of the dataset or data publisher)
4. Administrative contact (the person to be contacted in all matters regarding scientific data content and usage of a specific dataset or data publisher)
5. Institution name
6. Your relation to this Institution
7. The name of the GBIF Participant Node (see Box 13) that can endorse the publishing institution
8. The dataset URL: either the access point URL (if you are publishing using one of the provider softwares), or the DwC-Archive URL (if you are publishing via a zipped DwC-Archive)
9. The metadata document URL.

The GBIF Helpdesk will attend to your registration request as quickly as possible. The Helpdesk will first contact the GBIF Node selected with the registration, and ask them whether they want to endorse the new data publisher installation in their domain. Each new registration needs formal endorsement from a GBIF Participant Node manager (who best knows the institutions and databases in their country/organisation) before it is

allowed into the GBIF Registry. This is a simple quality control step required by the GBIF Participant Node Managers Committee.

Once endorsement has been received and the registration is completed, the registered dataset can be found on the GBIF Registry website[1], through searching by institution name or dataset title.

Indexing of datasets by GBIF

Following registration, the GBIF Helpdesk will queue the newly registered dataset for indexing. Depending on the size of the dataset, indexing can take anywhere from minutes to weeks. If problems are encountered during indexing, the GBIF Helpdesk work with you to resolve them as quickly as possible.

When indexing is successful, the new dataset will become publicly available (or discoverable) in the GBIF Data Portal (http://data.gbif.org). At present, GBIF attempts to update each registered dataset at least once every three months. During indexing, a set of core data elements is retrieved from your dataset and is stored in the GBIF index, so that the dataset will become accessible for searches.

GBIF has published a manual (GBIF, 2011g) describing the step-by-step process for publishing and registering data.

Box 13: GBIF Participant Nodes

A GBIF Participant Node is the team that co-ordinates the network of data holders, users and other stakeholders for each GBIF Participant (see Box 7) and co-ordinates data-sharing activities within its domain. The Participant Node is a conduit by which the GBIF Participants meet their own biodiversity information needs within their respective countries or organisations, while benefitting from and contributing to the GBIF network's mission and goals on making biodiversity data globally accessible. To find out who your Participant Node is, check on the GBIF website http://www.gbif.org/participation/participant-nodes/who-we-are/.

[1] http://gbrds.gbif.org

Section 7: Concluding remarks

This best practice guide describes a suite of simple, inexpensive tools and procedures that can be used by the impact assessment community to capture, publish and discover EIA-related primary biodiversity data. Publishing this data using consistent, internationally standardised formats is a relatively quick and easy procedure that can be easily adopted as an integral part or step of the EIA process. Uptake of the tools and processes described in this BPG will:

- enable free and open access to the biodiversity data which is essential for biodiversity-inclusive environmental assessments;
- facilitate the ongoing expansion and improvement of the local, national and global biodiversity databases on which EIAs and other areas of scientific work and land-use management frequently rely, improving baseline knowledge of the ecosystems of a particular site, region or country;
- help impact assessment practitioners to gain recognition for their work by enabling them to be cited in future uses of their data;
- enhance the quality, predictive value, verifiability and transparency of EIAs, thus improving the land-use decisions that they inform and the confidence civil society can place in these decisions.

References:

Convention on Biological Diversity. (1992). United Nations Environment Programme, Publication no.: Na 92-7807

Cadman, M. J., Petersen, C., Driver, M. D., Sekhran, N., Maze, K., Munshedzi, S. (2010). Biodiversity for development: South Africa's landscape approach to conserving biodiversity and promoting ecosystem resilience. Pretoria: South African National Biodiversity Institute, 171 pp. ISBN: 978-1-919976-59-4. Accessible at http://cmsdata.iucn.org/downloads/primer_11_2_mb.pdf

Chapman, A. (2005a). Principles of Data Quality, version 1.0. Copenhagen: Global Biodiversity Information Facility. 58 pp. ISBN: 87-92020-03-8. Accessible at http://www.gbif.org/orc/?doc_id=1229&l=en

Chapman, A. (2005b). Principles and methods of Data Cleaning, version 1.0. Copenhagen: Global Biodiversity Information Facility. 77 pp. ISBN: 87092020-04-6. Accessible at http://www.gbif.org/orc/?doc_id=1262&l=en

Chapman, A., Wieczorek, J. (2006). Guide to best practice for georeferencing. Copenhagen: Global Biodiversity Information Facility. 84 pp. ISBN: 87-92020-00-3. Accessible at http://www.gbif.org/orc/?doc_id=1288&l=en

Chapman, A., Grafton, O. (2008). Guide to best practices for generalising primary species occurrence data, version 1.0. Copenhagen: Global Biodiversity Information Facility. 27 pp. ISBN: 87-92020-06-2. Accessible at http://www.gbif.org/orc/?doc_id=1233&l=en

GBIF (2008). GBIF Training Manual 1: Digitisation of natural history collections data, version 1.0. Copenhagen: Global Biodiversity Information Facility. ISBN: 87-920-07-0. Accessible at http://www.gbif.org.

GBIF (2009). GBIF Work Programme 2009-2010. Copenhagen: Global Biodiversity Information Facility, 59pp. Accessible at http://imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=976c6c3015d7d2505d5ac1d45357c8

GBIF (2010). Best practice guide for 'Data Discovery and Publishing Strategy and Action Plans' version 1.0. Authored by Chavan, V. S., Sood, R. K., and A. H. Arino. Copenhagen: Global Biodiversity Information Facility, 29 pp. ISBN: 87-92020-12-7. Accessible online at http://www.gbif.org/orc/?doc_id=2755&l=en

GBIF (2011a). GBIF Strategic Plan 2012-2016: Seizing the Future. Copenhagen: Global Biodiversity Information Facility, 7pp. ISBN: 87-92020-18-6. Accessible at http://links.gbif.org/sp2012_2016.pdf

GBIF (2011b). 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. Accessible at http://www.gbif.org/orc/?doc_id=2815

GBIF (2011c). GBIF Spreadsheet templates: User Guide, version 1.0. (contributed by Remsen, D., Doring, M., Robertson, T.), Copenhagen: Global Biodiversity Information Facility, 20 pp., IBSN: 87-920-27-5. Accessible at http://www.gbif.org/orc/?doc_id=2823&l=en

GBIF (2011d). Publishing Species Checklists: Best Practices, version 1.0 (contributed by Remsen, D., Doring, M., Robertson, T.), Copenhagen: Global Biodiversity Information Facility, 20 pp., ISBN: 87-92020-26-7. Accessible at http://www.gbif.org/orc/?doc_id=2814&l=en.

GBIF (2011e). Darwin Core Archives: How to Guide, version 1.0. (contributed by Remsen, D., Braak, K., Doring, M., Robertson, T.). Copenhagen: Global Biodiversity Information Facility, 21 pp. Accessible at http://www.gbif.org/orc/?doc_id=2816&l=en

GBIF (2011f). GBIF Metadata Profile: Reference Guide. (contributed by O Tuama, E., Braak, K.), Copenhagen: Global Biodiversity Information Facility, 19 pp. Accessible at http://www.gbif.org/orc/?doc_id=2820&l=en

GBIF (2011g). Publishing and registering data with GBIF, version 1.0. Copenhagen: Global Biodiversity Information Facility, 8 pp. ISBN: 87-92020-29-1. Accessible at http://www.gbif.org/orc/?doc_id=2824&l=en

IAIA (2005). Biodiversity in impact assessment: IAIA Special Publication Series no. 3, July 2005, accessible at http://www.iaia.org/publicdocuments/special-publications/SP3.pdf

King, N; Rajvanshi, A; Willoughby, S; Roberts, R; Mathur, V; Chavan V; and Cadman M.(in prep). Improving access to biodiversity for, and from, EIAs – a data publishing framework built to global standards. (In prep).

Rajvanshi, A., Mathur, V., Iftikhar, U. A. (2007). Best-practice guidance for biodiversity inclusive impact assessment: a manual for practitioners and reviewers in South East Asia. CBBIA-IAIA Guidance Series, Capacity Building in Biodiversity and Impact Assessment (CBBIA) Project, International Association of Impact Assessment, North Dakota, USA.

United Nations (1992). Text of the Convention on Biological Diversity. Accessible at http://www.cbd.int/doc/legal/cbd-en.pdf.

Appendix 1: Darwin Core, Darwin Core Archives and the Darwin Core Archive Assistant

The Darwin Core body of standards

Darwin Core (DwC) is a body of standards that provides a stable set of terms for sharing information about biological diversity (Remsen, Braak, Döring and Robertson, 2011). It is used primarily to describe taxa and their occurrence in nature, as documented by specimens in natural history collections (physical or digital), observations and other related information (http://rs.tdwg.org/dwc/). The Darwin Core glossary includes a standardised set of terms, each accompanied by definitions and examples, and guidelines for use of the terms. The terms defined in the Darwin Core facilitate data sharing by providing a well-defined, standard vocabulary in a flexible framework that minimises barriers to use and maximises re-usability (Wieczorek, Doring, de Giovanni, Robertson and Veglias, 2009). Although Darwin Core was conceived to facilitate the discovery, retrieval and integration of information about modern biological specimens and their occurrence in space and time, its application today is somewhat broader, and it can be used to build enriched data exchange formats (Remsen *et al*., 2011). At this stage, it is considered that the Darwin Core glossary is suitable for describing primary biodiversity data from EIAs, but could be extended in future, if the need arises.

Darwin Core Archive

Darwin Core Archive (DwC-A) is an internationally recognised, biodiversity informatics standard that is the preferred format for publishing primary biodiversity data on the GBIF network. It makes use of Darwin Core terms to produce a single, self-contained dataset for sharing species-level (taxonomic) and species-occurrence data (Remsen *et al.,* 2011). The Darwin Core Archive consists of a series of one or more plain text files written in a particular format, including a compulsory **core data file** and a number of optional **extension files.** The core data file includes occurrence or taxonomic data about species (e.g. a list of species of mammals from a particular location), whereas the extension files include additional information about the core data file (such as the common names of the animals, habitat information, and so on). Each file is accompanied by a descriptor (or ***metafile***) explaining how the files are organised. The core data file, extension files,

descriptor (metafile) and metadata file (file describing the data) are all zipped together to make up the Darwin Core Archive for a particular dataset.

The Darwin Core Archive Assistant

Darwin Core Archive Assistant (DwC-A Assistant) is a web application that presents a simple interface for describing the data elements a publisher wishes to serve to the GBIF network. The data must already be digitised and in the form of delimited text files, or in a relational database management system that is supported by the GBIF network. The DwC-A Assistant is used when a publisher wishes to create and validate their own DwC-Archive file manually. This option is only recommended where both the technological and data management capacity is high. Further information on use of this option can be obtained by visiting http://tools.gbif.org/dwca-assistant/

Appendix 2: Acronyms used in this publication

CBD: Convention on Biological Diversity

DanBIF: Danish Biodiversity Inforamtion Facility

DwC: Darwin Core

DwC-A: Darwin Core Archive

GBIF: Global Biodiversity Information Facility

EAP: Environmental assessment practitioner

EIA: Environmntal Impact Assessment

IA: Impact Assessment

IAIA: International Association for Impact Assessment

IAIA-sa: South African chapter of IAIA

SANBI: South African National Biodiversity Institute

WII: Wildlife Institute of India

Appendix 3: Useful web addresses and References

Key References:

Getting Started:

1. Getting started: overview of data publishing in the GBIF network - http://www.gbif.org/orc/?doc_id=2815

Capturing data (Spreadsheets):

GBIF Spreadsheet templates: User Guide - http://www.gbif.org/orc/?doc_id=2823&l=en

Metadata

1. GBIF Metadata Profile: Reference Guide - http://www.gbif.org/orc/?doc_id=2820&l=en
2. GBIF Metadata Profile: How-to-Guide - http://www.gbif.org/orc/?doc_id=2821&l=en

Checklists:

1. Publishing Species Checklist: Best Practices - http://www.gbif.org/orc/?doc_id=2814&l=en
2. Publishing Species Checklists: Step-by-Step Guide - http://www.gbif.org/orc/?doc_id=2869&l=en

Darwin Core:

1. Create your own Darwin Core Archive: Step-by-Step Guide - http://www.gbif.org/orc/?doc_id=2818&l=en
2. Darwin Core Archive Assistant: User Guide - http://www.gbif.org/orc/?doc_id=2817&l=en
3. Darwin Core Archive Format: Reference Guide to the XML Descriptor File - http://www.gbif.org/orc/?doc_id=2819&l=en
4. Darwin Core Quick Reference Guide - http://www.gbif.org/orc/?doc_id=2803&l=en
5. Darwin Core Archive: How-to-Guide http://www.gbif.org/orc/?doc_id=2816&l=en

Registering/Discovering/Publishing data:

1. Publishing and Registering data with GBIF - http://www.gbif.org/orc/?doc_id=2824&l=en

General websites:

GBIF: www.gbig.org

IAIA: www.iaia.org

SANBI: www.sanbi.org.za

WII: www.wii.gov. in