

Biodiversity data sharing through GBIF



Dr. Lizanne Roxburgh Conservation Science Unit, Endangered Wildlife Trust

This webinar covers the following topics:

- 1. What is GBIF?
- 2. What is data mobilization?
- 3. The biodiversity data life cycle
- 4. Where to find online data and how to share it?
- 5. What are the barriers to data sharing
- 6. Why you should share your data



WHAT IS GBIF?

 An international network and research infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.



 The GBIF network of participating countries and organizations is coordinated through its Secretariat, and provides data-holding institutions around the world with common standards and open-source tools that enable them to share information about where and when species have been recorded.

THE ORIGINS OF GBIF



Organisation for Economic Co-operation and Development

The OECD promotes policies that will improve the economic and social well-being of people around the world, and it provides a forum in which governments can work together to share experiences and seek solutions to common problems.

In 1999, a recommendation of the OECD Global Science Forum was that:

"An international mechanism is needed to make biodiversity data and information accessible worldwide"

Following this recommendation, GBIF was established two years later

MEMORANDUM OF UNDERSTANDING (MOU)

In 2001: the Global Biodiversity Information Facility (GBIF) Memorandum of Understanding was opened for signature by countries and institutions around the world

In 2003: the GBIF Secretariat was established in Copenhagen under a country host agreement with Denmark



THE GBIF NETWORK TODAY



- Participation in GBIF is through 'nodes' that coordinate data mobilization from national institutions/networks
- Collaboration occurs at regional and global level

WHAT IS DATA MOBILIZATION?

Mobilization – making data available for use/ re-use by others.

In the past, this meant publishing a print document that could be read by many people. Today, mobilization usually means making something available on the web – truly mobile data is published to a **freely accessible web site**.

Once mobilized, data can be downloaded and imported for use in other analyses, possibly in combination with data from other data sources, so that they can be used to create new insights.

By encouraging and helping institutions to publish data according to common standards, **GBIF** enables research not possible before, and informs better decisions to conserve and sustainably use the biological resources of the planet.

What is Biodiversity Data?

Biodiversity data – bits of information about different kinds of organisms that have been observed somewhere in space and time.

Individuals of 1 Occurrence data

species occurring in a place and time

Sampling events

Individuals of 1 species occurring in a place and time, recorded with a particular effort

Vegetation Type

Density Classes
 No trees (0 trees)
 Low density (1-100 trees/ha)
 Medium density (101-500 trees/ha)
 High density (501-1000 trees/ha)
 Sich density (>1000 trees/ha)

ja 30.

Checklists

A list of species occurring in a particular place

Biodiversity Data Life Cycle



Typically, biodiversity data in a research project would go through the cycle of collection, processing, analysis, publication of analysed results, storage on a computer or hard drive, and ultimately loss after some years or decades



Biodiversity Data Life Cycle



A much preferable lifecycle would not end with data loss, but rather with data sharing. These data could then be combined with other data sources, and reenter the cycle of processing, analysis, publication of analysed results, and further sharing.

GBIF is arguably one of the most important avenues for sharing and reuse of biodiversity data.



GBIF.ORG - MOBILISED DATA



This map represents the biodiversity data available/ mobilised on GBIF. Each dot represents evidence of species occurrence with standardized information on what was observed, where, by whom, when, and based on what evidence?

The GBIF website – www.gbif.org



Where else can data be shared?

A sample of online data-sharing platforms:

- GBIF for organismal occurrence data
- Dryad for article-related data
- Encyclopedia of Life has a page for each species
- DataONE Data Observation Network for Earth environmental science data
- figshare for sharing of institutional scientific data and outputs



Data Shervation Network for Earth





GBIF

Google Dataset search – provides a single place to search all data platforms

https://toolbox.google.com/datasetsearch

figshare.com

Google Dataset Search	Q african lion X About III O L
Data from: A molecular analysis of African lion (Panthera leo) mating datadryad.org Published Feb 6, 2013	Data from: A molecular analysis of African lion (Panthera leo) mating structure and extra-group paternity in Etosha National Park G Related Article G datadryad.org
PLOS Model outputs of eight African lion incidence function models. figshare.com Updated Dec 2, 2015	23 scholarly articles cite this dataset (View in Google Scholar) Dataset published Feb 6, 2013
PLOS Mean mass of African Lion skeletons for individuals of varying sex, size, figshare.com Updated Dec 3, 2015	Dataset provided by Dryad Digital Repository Authors Martha M. Lyke; Jean Dubach; Michael B. Briggs Description
Data from: Bayesian estimates of male and female African lion mortality for datadryad.org Published Feb 22, 2016	The recent incorporation of molecular methods into analyses of social and mating systems has provided evidence that mating patterns often differ from those predicted by group social organization. Based on field studies and paternity analyses at a limited number of sites, African lions are predicted to exhibit a strict within-pride mating system. Extra-group paternity has not been previously reported in African lions; however, observations of extra-group associations among lions inhabiting Etosha National Park in Namibia suggest deviation from the predicted within-pride mating pattern. We analyzed variation in 14 microsatellite loci in a population of 164 African lions in Etosha National Park. Genetic analysis was coupled with demographic and observational data to examine pride structure, relatedness, and extra-group paternity (EGP). EGP was found to occur in 57% of prides where paternity was analyzed (n = 7) and the overall rate of EGP in this population was 41% (n = 34). Group sex ratio had a significant effect on the occurrence of EGP (p < 0.05), indicating that variation in pride-level social structure may explain intergroup variation in EGP. Prides with a lower male to female ratio were significantly more likely to
Actions: Predictors of Lion Killing by Maasai	provide evidence that social structure may be an important indicator of EGP in some social mammals.

Why share data?

"Data reuse is the fundamental goal of data sharing"

Examples of reuses: Red Lists, climate change modelling, national biodiversity assessments, identification of critical biodiversity areas, environmental impact assessments, alien invasive species research



ANNUAL NUMBER OF PEER-REVIEWED ARTICLES USING GBIF-MEDIATED DATA



GBIF Secretariat. (2018). GBIF Science Review 2018. https://doi.org/10.15468/VA9B-3048

An example of data published by an African country: GBIF data for Namibia



Why doesn't everyone share their data? Examples of barriers to data publishing

Psychological & cultural barriers Institutional barriers Capacity barriers **Practical** barriers GBIF

Global Biodiversity Information Facility

- Lack of knowledge
- Lack of understanding
- Lack of will
- Perceived data value
- Privacy concerns
- Lack of authorization
- Lack of time / planning
- Lack of capacity
- Lack of funding
- Lack of infrastructure

Data holders are possessive about their data, and not aware of the value of sharing

Data holders believe their data has financial value and the potential to make them rich

Data holders may not have the authorization from their institution to share data, even if they personally think they should share

Data holders do not build enough time into their projects for data management and sharing

Data holders work for institutions that do not provide sufficient IT infrastructure

Risks associated with sharing vs hoarding of data



Benefits associated with sharing of data

- catalyses new collaborations
- increases confidence in findings
- generates goodwill among researchers

Sharing

- allows datasets to be cited (DOIs) and publishers credited
- improves publishers' CVs, and job, tenure, and promotion applications
- improves access to funding, as some funders require that data is freely accessible



The benefits of sharing outweigh the risks (which can be minimised)

Why publish your data online?



- Established lecturers and researchers leave a legacy, allowing future researchers to reuse your data and acknowledge your contribution to science
- Students and young researchers this is the future: research funders and journal publishers will no longer allow you to hoard your data; there are opportunities for you to make use of others' data and collaborate on research projects





How can you publish your data?

- Become a recognised data publisher by being endorsed by a GBIF node
- Prepare your data according to DarwinCore standards in a spreadsheet or data table (or other data management software)
- Prepare metadata to accompany the dataset (Metadata is data about data)
- Have the data published through an IPT (Integrated Publishing Toolkit) hosted by any GBIF node



How do you obtain the skills you need to prepare your data for publication?

Online GBIF tools and other resources available at https://www.gbif.org/resource/search?contentType=tool

Funding for training and resources available through GBIF https://www.gbif.org/programme/82219/capacity-enhancementsupport-programme

Online training in biodiversity informatics available at http://biodiversity-informatics-training.org/

GBIF community site https://discourse.gbif.org