





**Publishing primary biodiversity data:
Information overload or conservation necessity?**

Brenda Daly & Harriet Davies-Mostert (Endangered Wildlife Trust, South Africa)
& Nick King (Global Biodiversity Information Facility, Denmark)

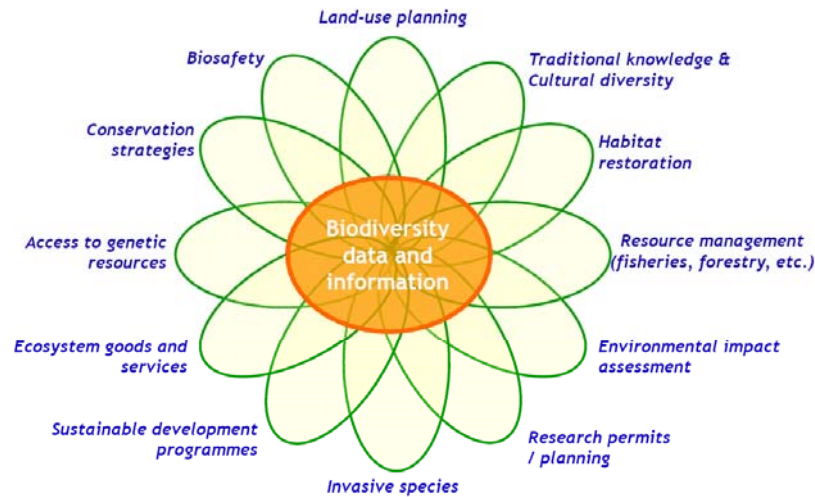
Setting the scene



- ▶ The world's biodiversity is in crisis (Jenkins 2003)
- ▶ Biodiversity is the cornerstone of ecosystem functioning, good and services
 - ▶ Food
 - ▶ Fibre
 - ▶ Medicines
 - ▶ Fresh water
 - ▶ Crop pollination
 - ▶ Pollutant filtration
 - ▶ Buffer from natural disasters
 - ▶ Etc...
- ▶ Multiple causes of biodiversity loss, but anthropogenic causes are ultimately due to conflict (direct and indirect)
- ▶ Reversing the decline in species and biodiversity requires access to credible information



Essential uses of primary biodiversity data



▶ Source: Bello, J.C. 2008. Roles and importance of GBIF participant nodes. II Course on GBIF Node Management, Arusha Tanzania.



The challenge

- ▶ Biodiversity data and information are scattered, heterogeneous resources
 - ▶ Individuals
 - ▶ Organisations
 - ▶ Institutions
 - ▶ Government agencies, etc
- ▶ Therefore difficult to access and use!



Outline of talk

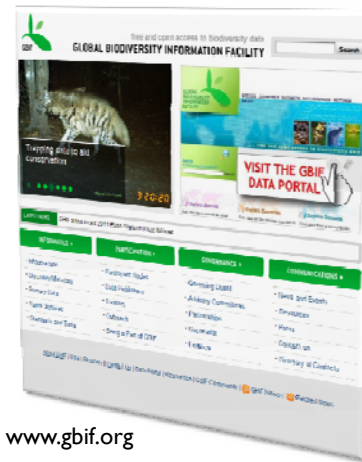
1. Initiatives to promote publishing of biodiversity data
 - ▶ International
 - ▶ National
2. Overcoming impediments to data publishing
 - ▶ Data Papers
3. Preliminary results of a survey to understand the data needs of users



Initiatives to promote data
publishing

Promoting data publishing:
 The Global Biodiversity Information Facility (GBIF)

- ▶ Global mega-science initiative to harness knowledge of the Earth's biological diversity
 - ▶ for science
 - ▶ for society
 - ▶ for sustainable future
- ▶ **MISSION:** "To promote and enable free and open access to biodiversity data worldwide via the Internet to underpin science, conservation and sustainable development".

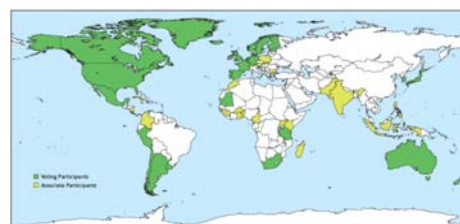


www.gbif.org



Promoting data publishing:
 GBIF (cont.)

- ▶ Founded and funded by governments in 2001
- ▶ Free access via the Internet
- ▶ Secretariat in Denmark
- ▶ Network of participants (BIFs)
 - ▶ National → SABIF: National (SANBI)
 - ▶ Thematic → ICLEI: Thematic – Local governments for Sustainability (Cape Town)
 - ▶ Regional → EWT: Regional – trans-boundary data
- ▶ Facilitating access to
 - ▶ >300 million biodiversity records
 - ▶ 18,000+ data resources
 - ▶ 342 Data Publishers



Promoting data publishing: The South African Biodiversity Information Facility

▶ Mission

- ▶ Developing and maintaining databases
- ▶ Establishing and maintaining web portal
- ▶ Providing funding for digitisation
- ▶ Providing funding for data provider interface
- ▶ Facilitating the development of application tools for data use
- ▶ Encouraging networking and coordination of activities



www.sabif.ac.za



Promoting data publishing: Migrating biodiversity knows no boundaries



EWT / GBIF Partnership:

- ▶ Associate Participant Node since 5 August 2008
- ▶ Create links between existing trans-boundary datasets
- ▶ Increase participation in GBIF by sub-Saharan African countries
- ▶ Help build greater capacity in sub-Saharan Africa
- ▶ Mobilise data for greater access and utilisation



Promoting data publishing: The EWT's Knowledge Management System

- ▶ Network of monitoring projects
 - ▶ Independent
 - ▶ Multiple geographic scales
- ▶ With partners, creates vast amount of data
 - ▶ Species-specific
 - ▶ Location
 - ▶ Breeding
 - ▶ Mortality and its causes
 - ▶ Interactions between wildlife and people
 - ▶ Conflict
 - ▶ Persecution
 - ▶ Poisoning events
- ▶ Available on the internet



www.kms.ewt.org.za



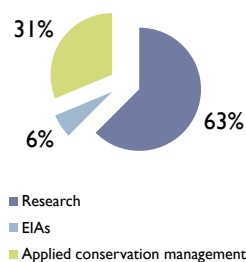
Promoting data publishing

Core data types

- Resource (or dataset) METADATA
 - ▶ Metadata are data records that provide descriptive information about dataset (i.e. **data about data**).
 - Occurrence data
 - ▶ Refers to a data or information relating to a specific instance of a taxon
 - ▶ Digital text or multimedia data record detailing facts about the instance of occurrence of an organism, (i.e. on the what, where, when, how and by whom of the occurrence and the recording)
 - Taxonomic data
 - ▶ Data or information relating to a taxon and not necessarily to a specific instance (occurrence) of an individual within that taxon
- Archive data

 - ▶ Data that do not fit into the above categories (behavioural, genetic, social...etc)

Uses of EWT data in 2010



- EXAMPLES:**
- Crane breeding locations
 - Crane sightings
 - Powerline mortalities
 - Poisoning events
 - Oribi sites in KZN
 - Blue Swallow nesting locations



Impediments to data publishing?

Impediments to data publishing

- ▶ Concerns about inadvertent misuse of data
 - ▶ For example: sensitive data – crane breeding locations, cycad locations, etc
- ▶ Lack of ownership agreements
- ▶ Competition for academic position and funding
- ▶ Usability
- ▶ Lack of informed consent and confidentiality
- ▶ **Lack of incentives**

▶ Chavan & Ingwersen 2009



Current Practice of Data Citation

Source: GBIF Data Portal, data.gbif.net
 Search string: *Panthera tigris*
 Search results: 696 records, from 37 datasets, published by 31 Data Publishers
 Date: Thursday, 4 November 2010, Time: 10.03.30

Existing Data Citation style

Please cite this data as follows:

(accessed through GBIF data portal, Mammal specimens, <http://data.gbif.org/datasets/resource/559>)
 (accessed through GBIF data portal, Vertebrate specimens, <http://data.gbif.org/datasets/resource/541>)
 (accessed through GBIF data portal, Natural History Museum Rotterdam, <http://data.gbif.org/datasets/resource/693>)
 (accessed through GBIF data portal, Database Schema for UC Davis Wildlife museum, <http://data.gbif.org/datasets/resource/736>)
 (accessed through GBIF data portal, UNSM Vertebrate Specimens, <http://data.gbif.org/datasets/resource/812>)

Un-answered facts

What was the search string?

How many records were retrieved?

How many Data Publishers contributed to the data?

When search was carried out?

Who is the original contributor of the data?

Who played what role from collection to publishing?

How can I retrieve the same result?



Proposed solution: Incentivise data discovery

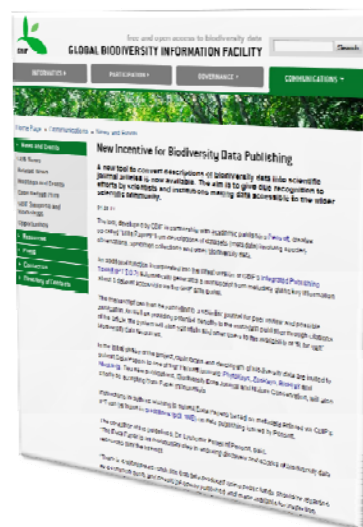
DATA PUBLISHING
 FRAMEWORK



DATA USAGE INDEX
 (DUI)

Due recognition given to multiple
 players in the data
 collection/creation, management and
 publishing cycle

Data publishing = Scholarly publishing



Incentivising data discovery

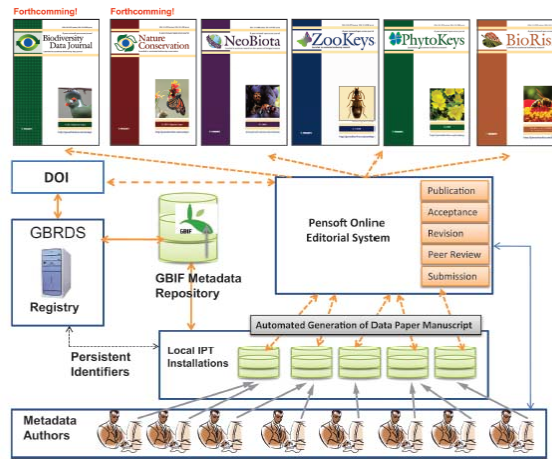


Figure 1. The GBIF/Pensoft workflow of data publishing and automated generation of Data Paper manuscripts



Data papers

- ▶ Promote the creation of robust peer-reviewed metadatasets
- ▶ Requires that data are published to the Internet
- * Early-bird special - first 100 papers have no page charges!!

ZooKeys is an ORCID iD
www.pensoft.net

DATA PAPER



IndFauna, electronic catalogue of known Indian fauna

Jyendra Galkwad¹, Rebecca James², Monica Paterni¹, David Robertson¹,
Ton Geisen³, S. Krishnan⁴

¹ National Chemical Laboratory, 411007, Pune, India; ² Belgian Academy of Sciences, 2000, Sable, Belgium; ³ National Natural History Museum, 7722, Leiden, The Netherlands; ⁴ 1888, 75 South Sheshadri Street, 50004, Coimbatore, India; ⁵ California Academy of Sciences, 1111, San Francisco, USA

Corresponding author: Jyendra Galkwad (galkwad@ncl.ac.in), Monica Paterni (mpaterni@ncl.ac.in)

Received: 11 June 2011 | Accepted: 13 July 2011 | Published: 20 July 2011
Citation: Galkwad J, James R, Paterni M, Robertson D, Geisen T, Krishnan S (2011) IndFauna, electronic catalogue of known Indian fauna. ZooKeys 10: 1-10. doi:10.3897/zookeys.10.1

Abstract
The article describes the development and features of IndFauna, electronic catalogue of known Indian fauna. Available at <http://www.ncl.ac.in>, this catalogue raises several issues concerned with taxonomy or systematic and information technology in biodiversity information management. Besides information on more than 93% of the 59,000 known faunal species in India has been documented in IndFauna, which demonstrates a model of collaboration between domain experts and IT managers. It is our belief that such ICTs would be effective in encouraging economic investments as well as better sustainable use and conservation of our biotic resources.

Keywords
Biodiversity informatics, IndFauna, data publishing, electronic catalogue

Taxonomic coverage
General taxonomic coverage description: The coverage of this database spans whole of Kingdom: Animalia. Database contains occurrences of over 90000 species belonging to 2222 genera.

Taxonomic ranks: Kingdom: Animalia, Phylum: Acarothorapha, Arachnida, Arthropoda, Mollusca, Chordata, Rotifera, Class: Amphibia, Aves, Chondrichthyes, Mammalia, Reptalia, Order: Monotremata, Amura, Caudata, Gymnophiona, Family:

Copyright Jyendra Galkwad et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



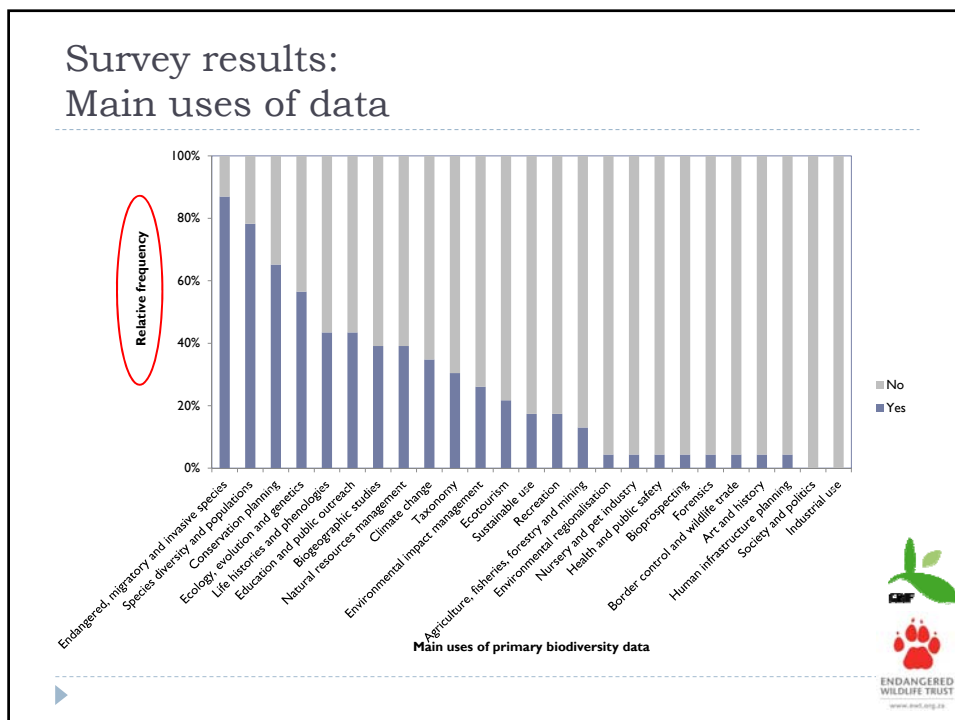
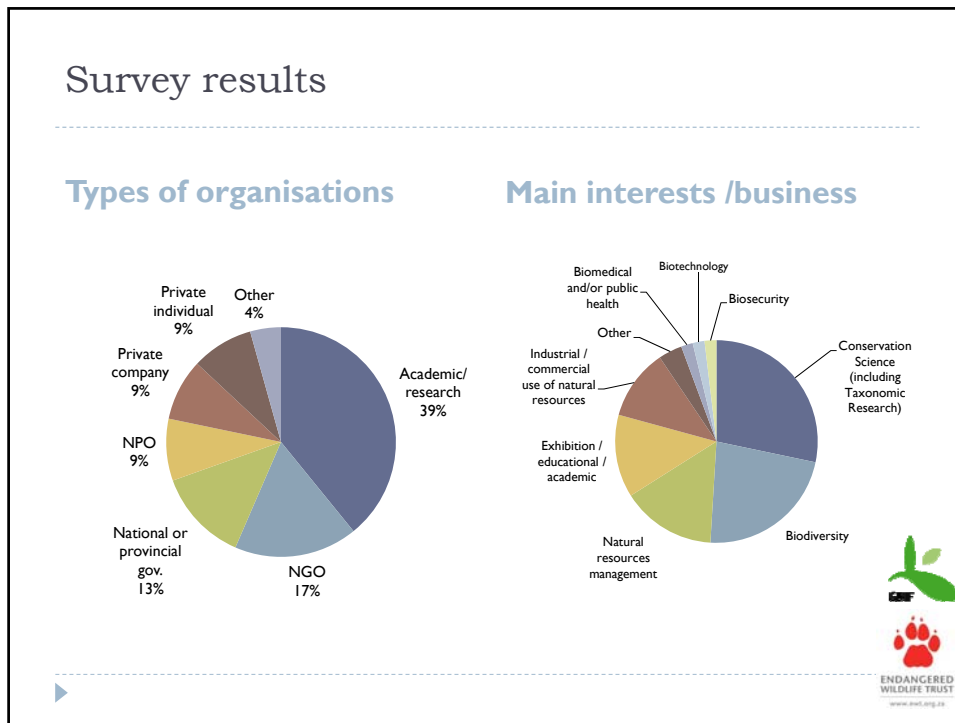
Survey of data needs

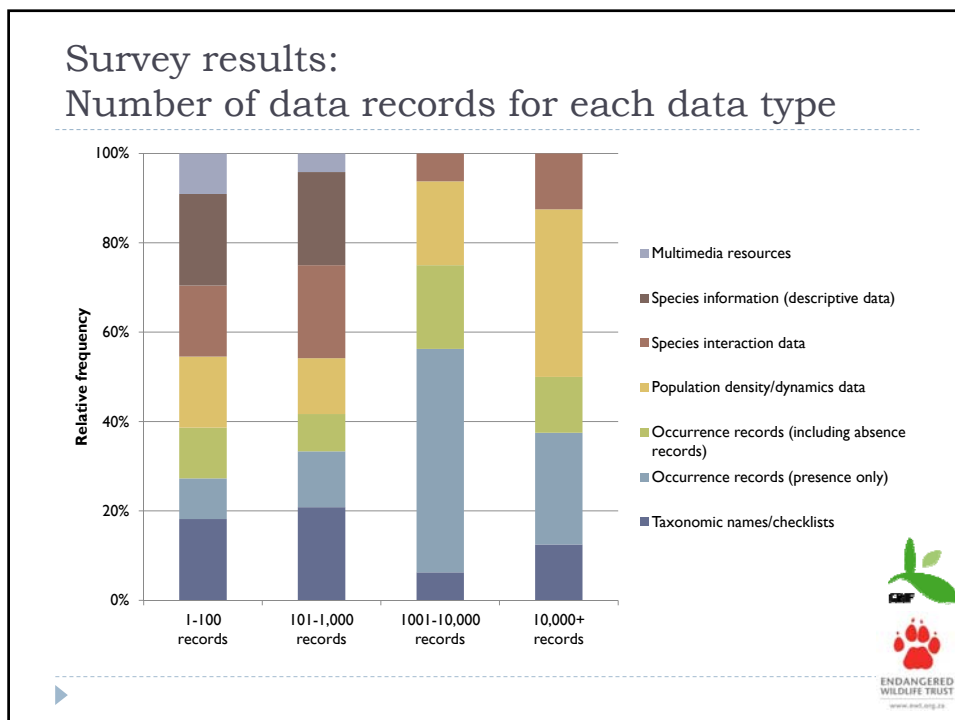
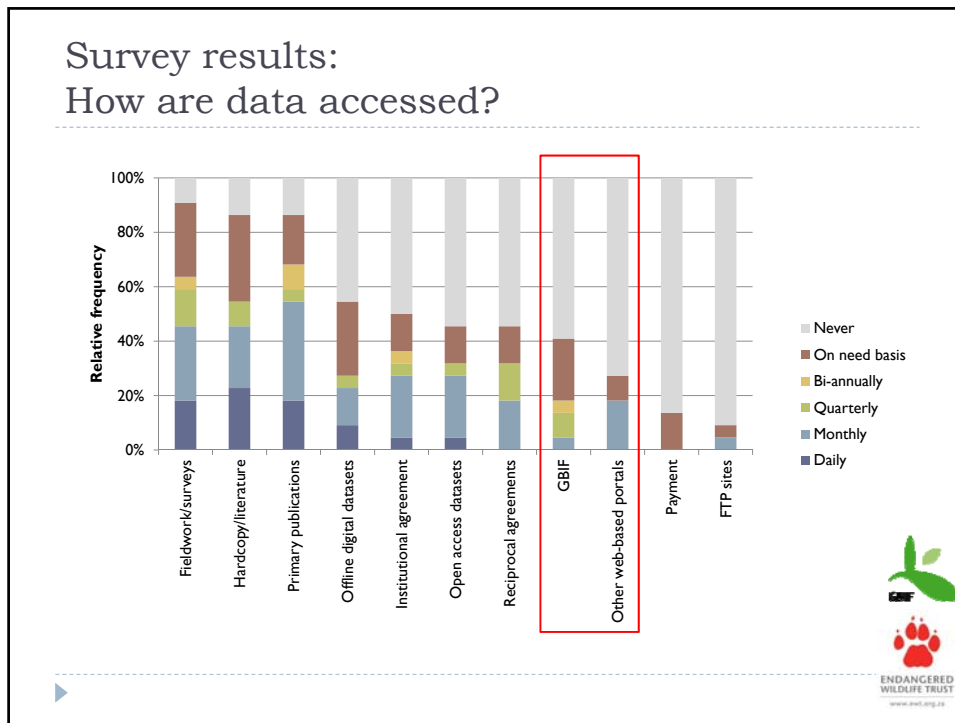
Preliminary results

What biodiversity data do our partners need?

- ▶ Three main objectives:
 - To understand the biodiversity data needs of the users
 - To determine what format data are in, and need to be in, for optimum conservation benefit
 - To assess if and how data are made available to partners
- ▶ Methods
 - Time-bound internet-based survey (Survey Monkey)
 - Sent to 88 data partners
- ▶ Results to date
 - 23 respondents (26% return rate)
 - Results are very preliminary!
 - Bias towards organisations working with terrestrial vertebrates
 - Re-mail planned before the end of the year

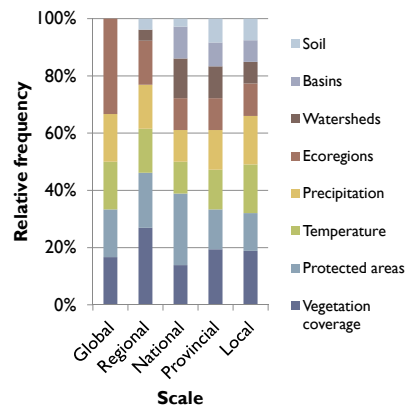




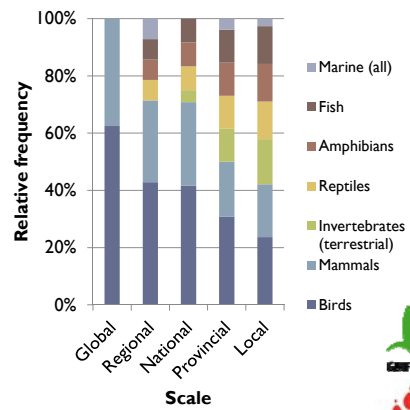


Survey results: Data types needed at different scales

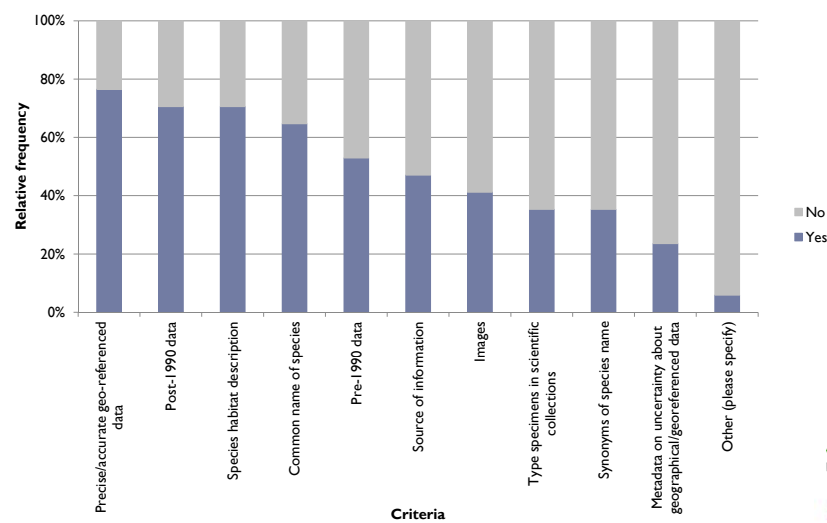
Ecosystem data



Species data

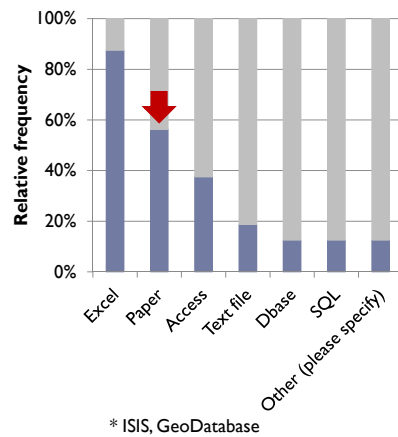


Survey results: Most important characteristics of species data

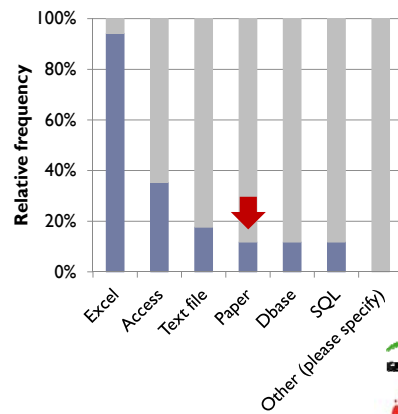


Survey results: Data formats?

How are data currently stored?



What format for external sources?



Survey results: Summary

- ▶ Results are very preliminary but align with results of other global and national surveys to look at data needs
- ▶ Way forward:
 - ▶ Obtain more responses from the remaining respondents and finalise national report
 - ▶ Critical comparison with global trends



OVERALL CONCLUSIONS

1. Mechanisms exist to mobilise biodiversity data for decision-making
2. Rate of uptake in publishing and discovery of data is sluggish
3. New incentives exist to mobilise data

DATA PUBLISHING IS A CONSERVATION NECESSITY
FOR INFORMED DECISION-MAKING



Acknowledgements

- ▶ EWT staff
- ▶ EWT's Information Technology for Conservation Programme (IT4CP)
- ▶ Brenda Daly & Nick King
- ▶ Vishwas Chavan (GBIF)

Brenda Daly
brendad@ewt.org.za

www.kms.ewt.org.za

Thank you!

