Biodiversity Data Publishing: Importance, Barriers, data types, publishing workflow

Prof. L. Jimu

Biodiversity data publishing

- In the GBIF context, data publishing means sharing biodiversity data using a **standardized** format, making them **freely accessible**.
- “Free and open access to primary biodiversity data is essential for informed decision-making to achieve conservation of biodiversity and sustainable development.

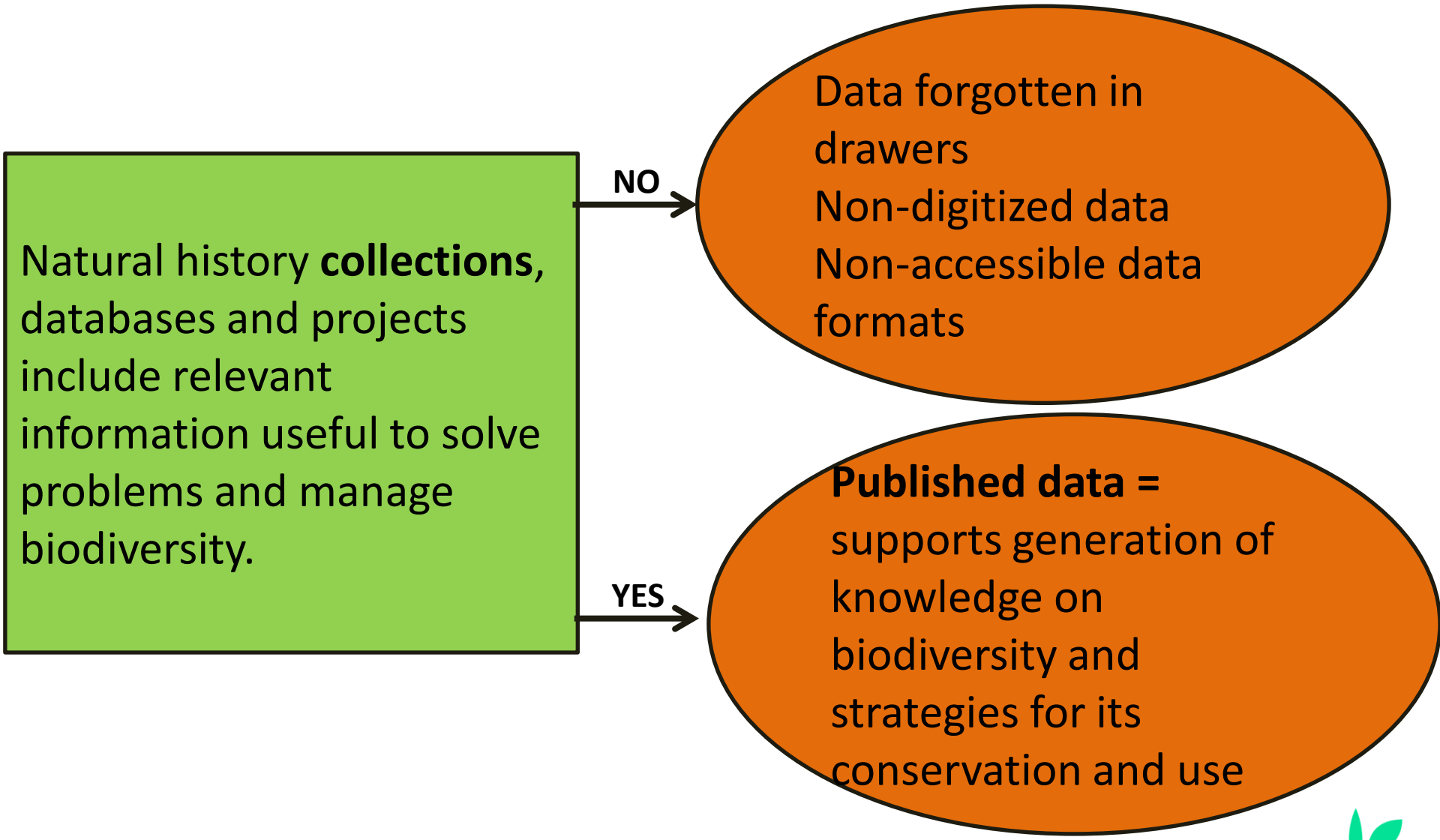
Publishing and discovery of biodiversity data: the constraints and challenges

- GBIF has facilitated mobilisation of over **1,388,248,840** occurrence records, **50,268** datasets, published by **1,561** institutions.
- **4,172** peer-reviewed papers using data.
- Partnerships with e.g. the Catalogue of Life, Biodiversity Information Standards (TDWG), the Consortium for the Barcode of Life(CBOL), the Encyclopaedia of Life(EoL), and Integrated Taxonomic Information System (ITIS).

Publishing biodiversity data: constraints and challenges

- lack of easy-to-use tools and related guidelines;
- the difficulty of dealing with heterogeneity and diversity of standards;
- the cost of creation and maintenance of infrastructure by small-and medium-scale data publishers; and
- the lack of professional reward structures or incentives.

Why publish biodiversity data?



Why publish biodiversity data?– Other benefits

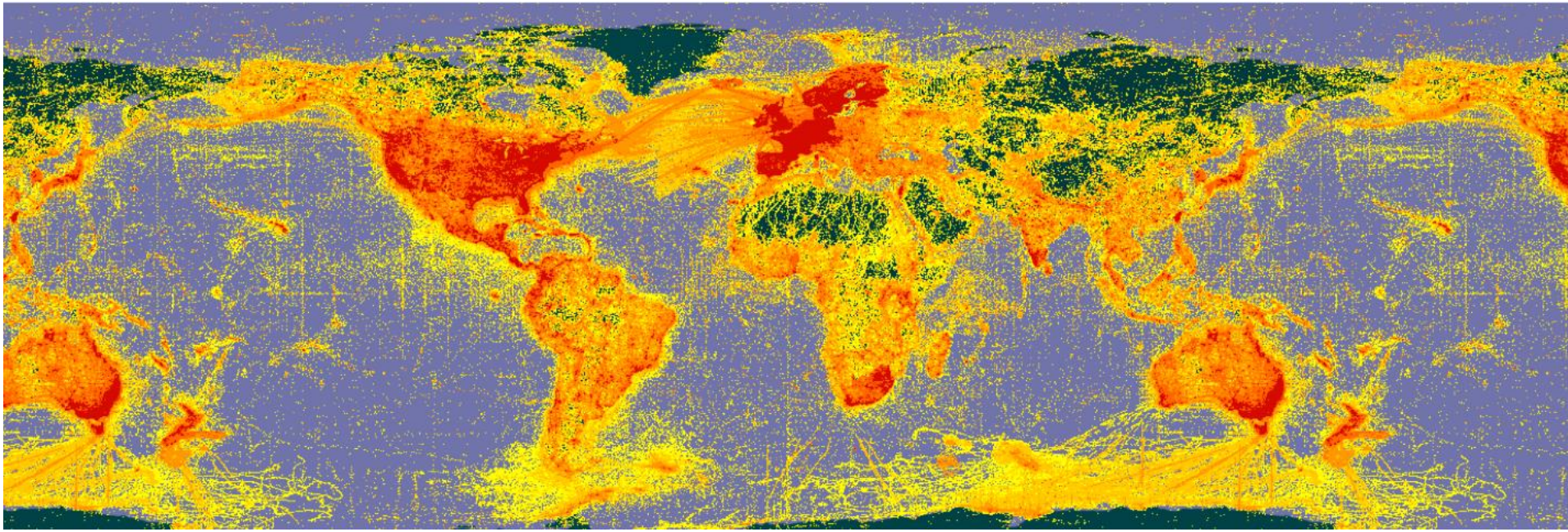
- Our work gets more visibility and recognition
- Increased chances for collaboration with other institutions.
- Increase of professional networks and relations with experts related to our fields of expertise
- High quality publications in high impact journals

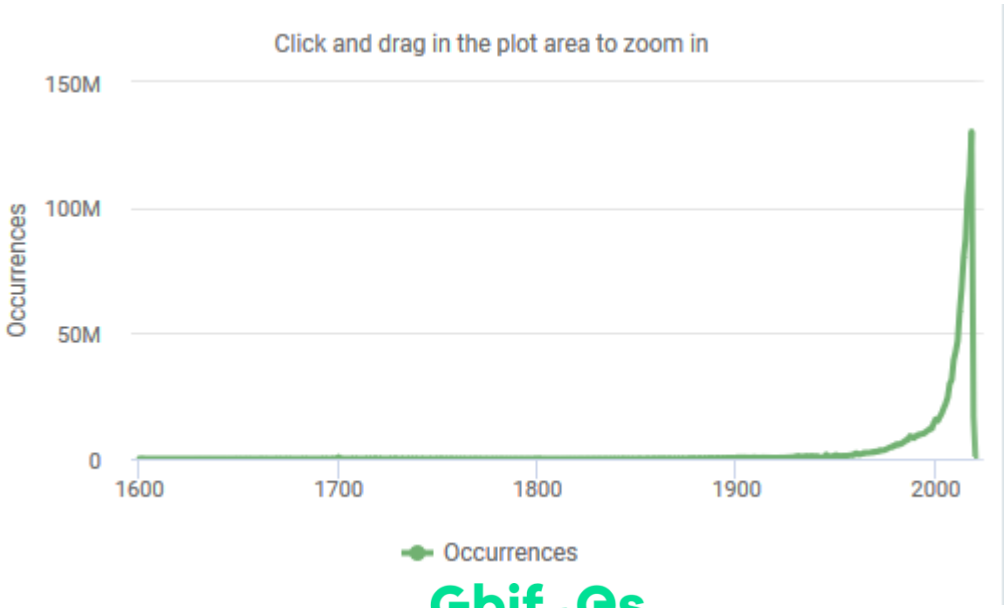
Why publish biodiversity data?– Data uses

- **Traditional uses:** Taxonomic studies, specie distribution studies, red lists development, threatened and invasive species identification, etc.
- **Modern uses:** Potential distribution models, models related to climate change, etc.

- Chapman, A.D., 2005, Uses of Primary Species-Ocurrences Data: https://www.gbif.org/document/80545/uses-of-primary-species-occurrence-data

Why publish biodiversity data?





GBIF

Gbif.es

GBIF Zimbabwe

Barriers to biodiversity data publication

Cultural and psychological barriers	➢ Lack of knowledge ➢ Lack of commitment ➢ Data quality
Institutional barriers	➢ Privacy concerns ➢ Lack of authorisation
Capacity and resource barriers	➢ Lack of time ➢ Lack of capacity
Practical barriers	➢ Lack of funds ➢ Lack of infrastructure

Barriers to biodiversity data publication- Restrictions

- ~~Economical compensation for data sharing.~~
- Threatened species.
- Inadequate data quality.

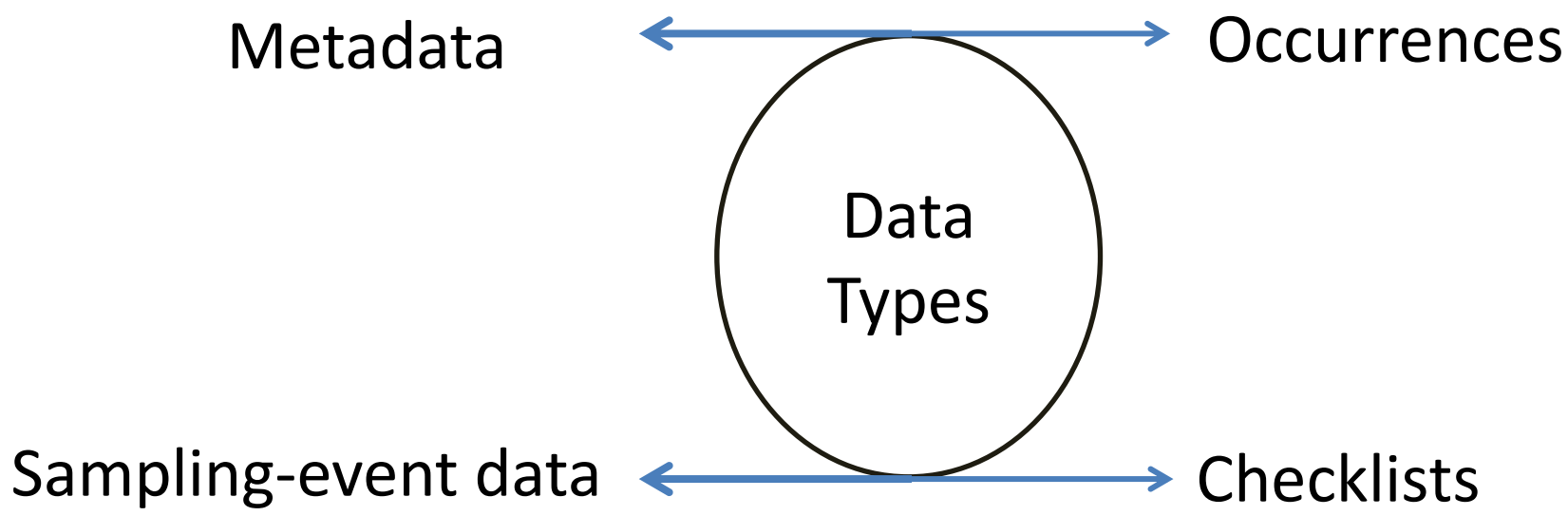
Barriers to biodiversity data publication

- Insufficient data exploitation → publishing data already used.
- Threatened species → set diffuse locations (e.g. 10 km radius).
- Inadequate data quality → consider alternative data types.

Data types published-Data sources

- Specimens from natural history collections, herbarium samples, tissues, DNA, etc.
- Checklists/ Red lists.
- Thesis, Literature, Reports, etc.
- Satellite data, audio and video records, etc.
- Environmental impact studies, inventories, citizen science projects, etc.

Data types published-Data sources



Data types-Checklists

- lists of scientific names of organisms grouped into taxonomic hierarchies,
- provide taxonomic 'backbones' around which species information can be organized.

Data types-Occurrences

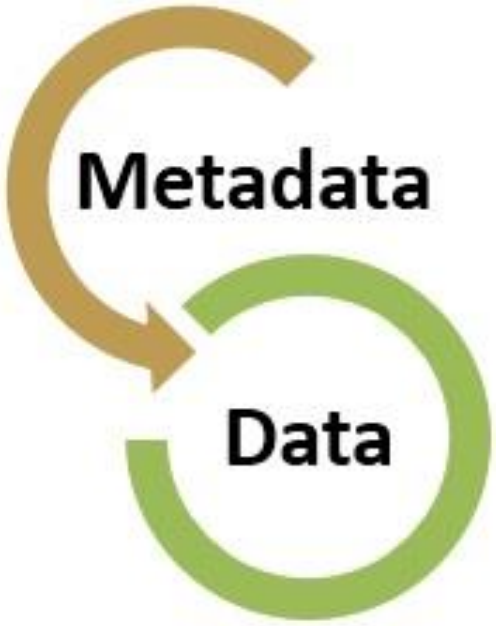
Occurrences (observations, specimens etc)

- collection event
- an observation in the field, vouchered (labeled) specimen in a museum or herbarium, or other evidence.

Data types-Metadata

Metadata (data about data)

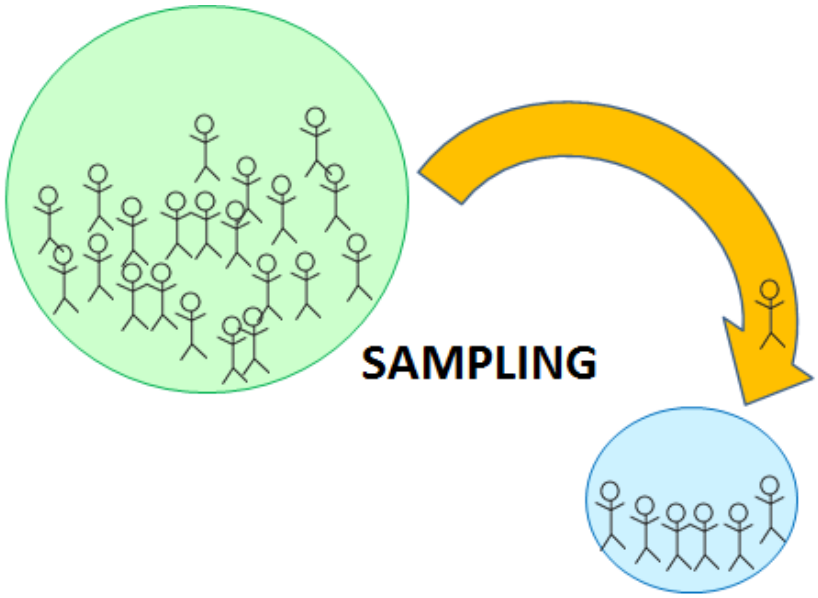
- structured descriptions of datasets
- help to give context to datasets and enable users to assess whether data are fit for use in a particular research project or application

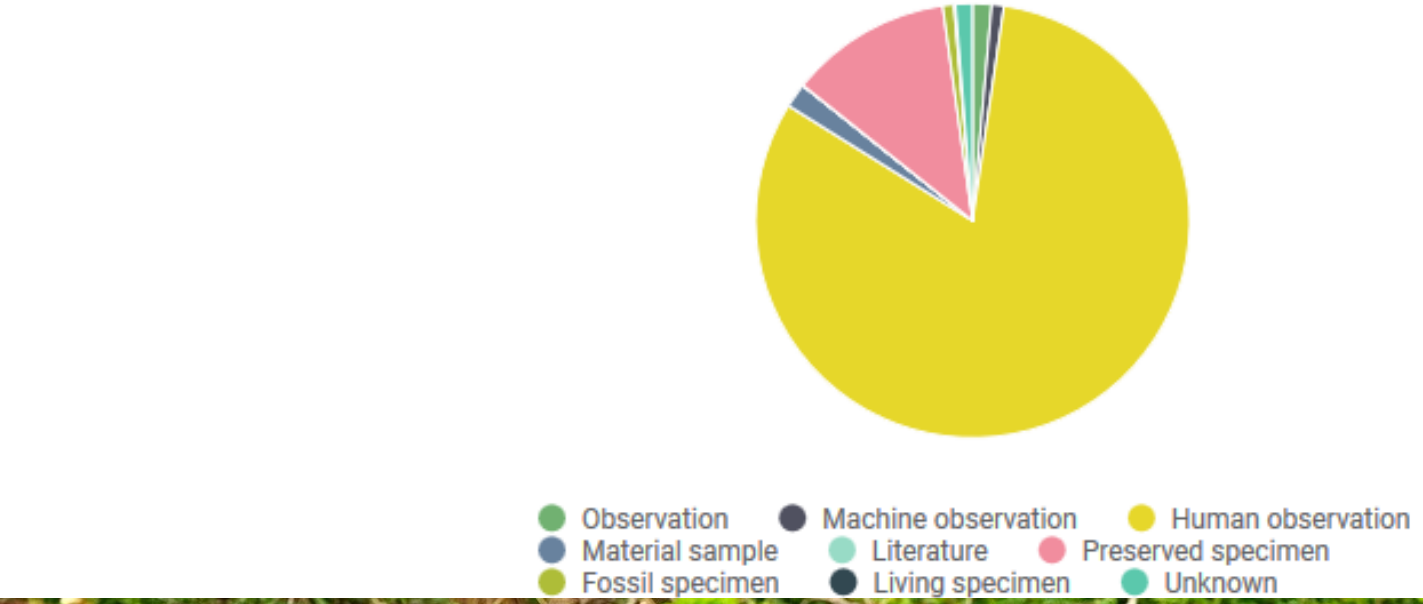


Data types: Sampling-event

Sampling-event (quantitative information)

- records from thousands of different kinds of environmental, ecological, and natural resource monitoring and assessment investigations




Observation
Machine observation
Human observation
Material sample
Literature
Preserved specimen
Fossil specimen
Living specimen
Unknown

Data Standards

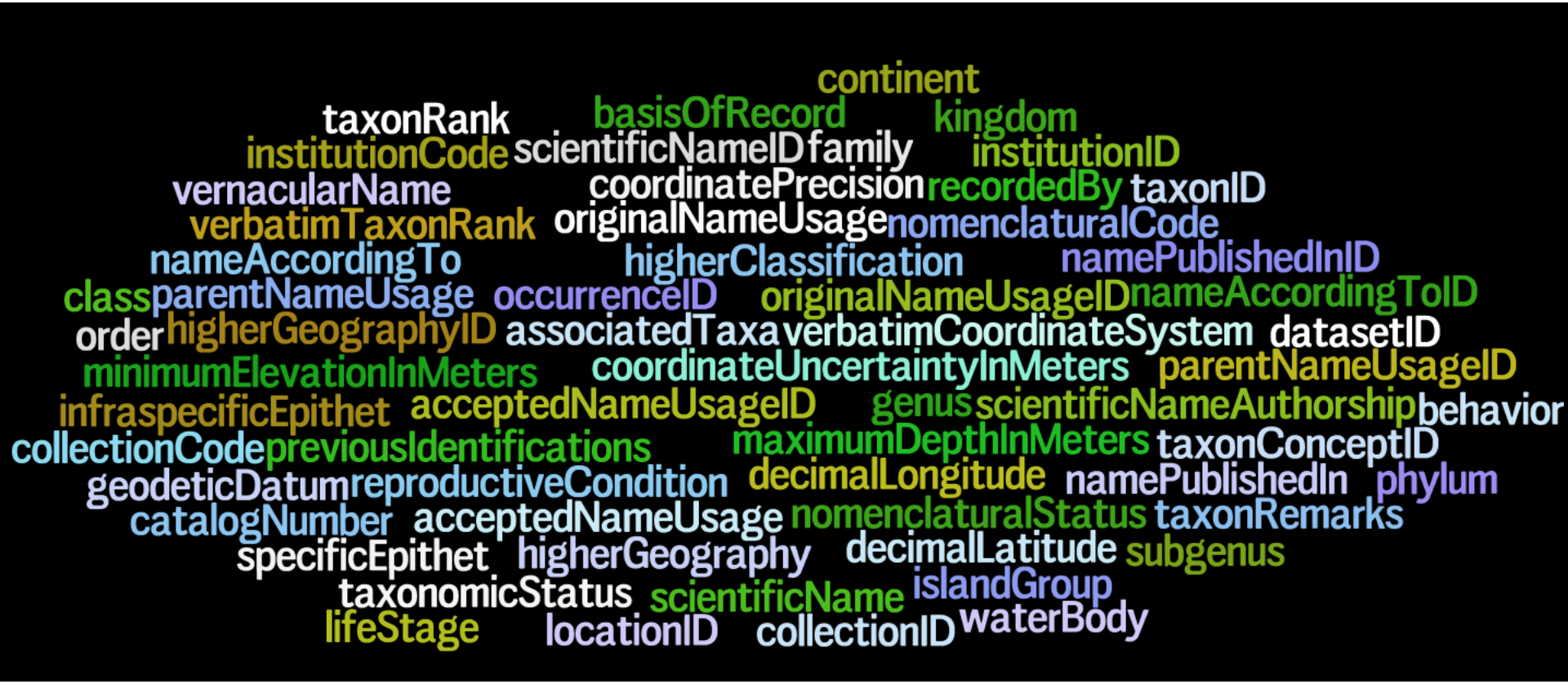
- **ABCD** Access to Biological Collection Data (2005)
- **DwC** Darwin Core (2009)
- **AC** Audubon Core Multimedia Resources Metadata Schema (2013)
- **NCD** Natural Collection Descriptions (Draft)

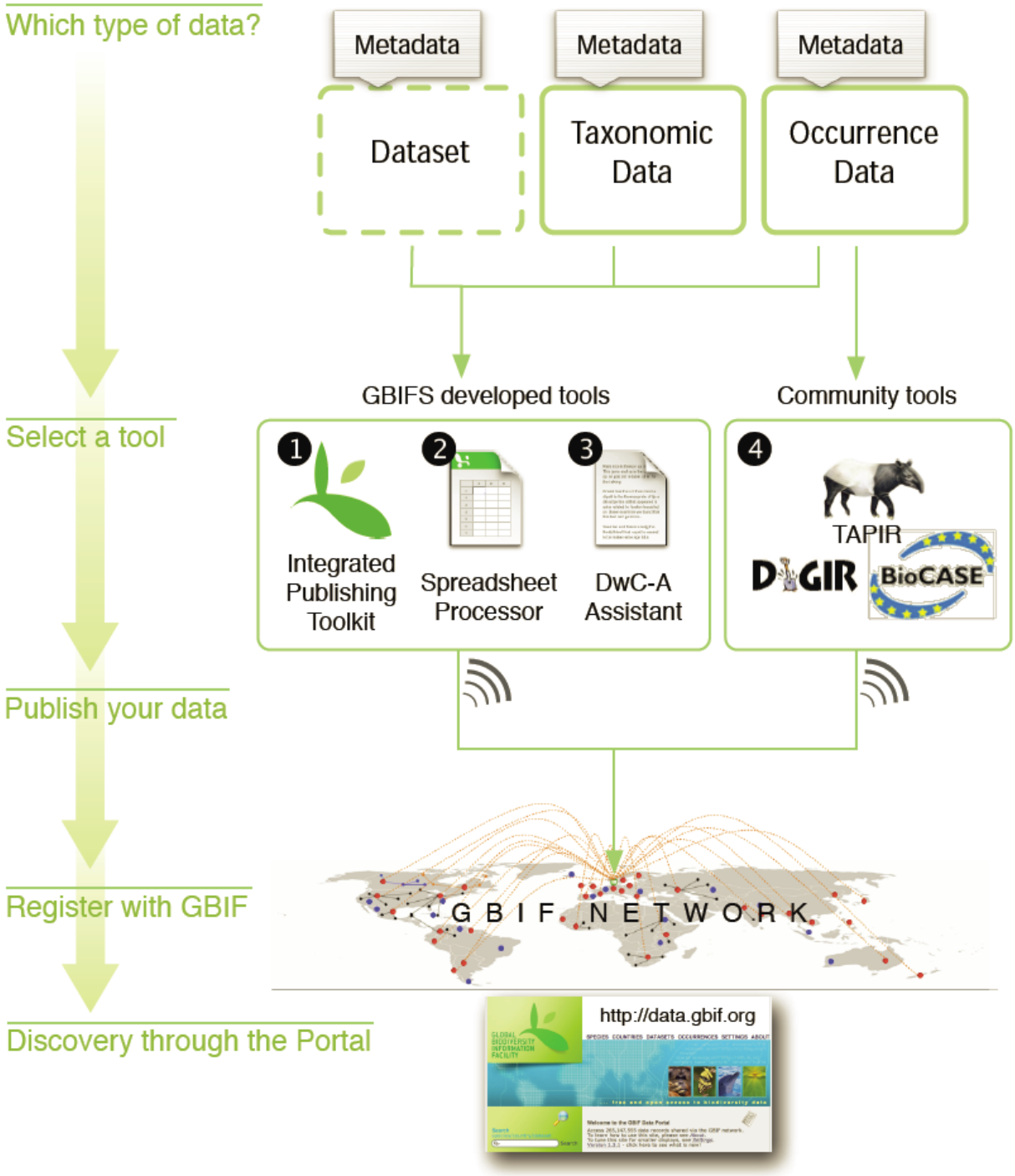
Biodiversity Information Standards

TDWG

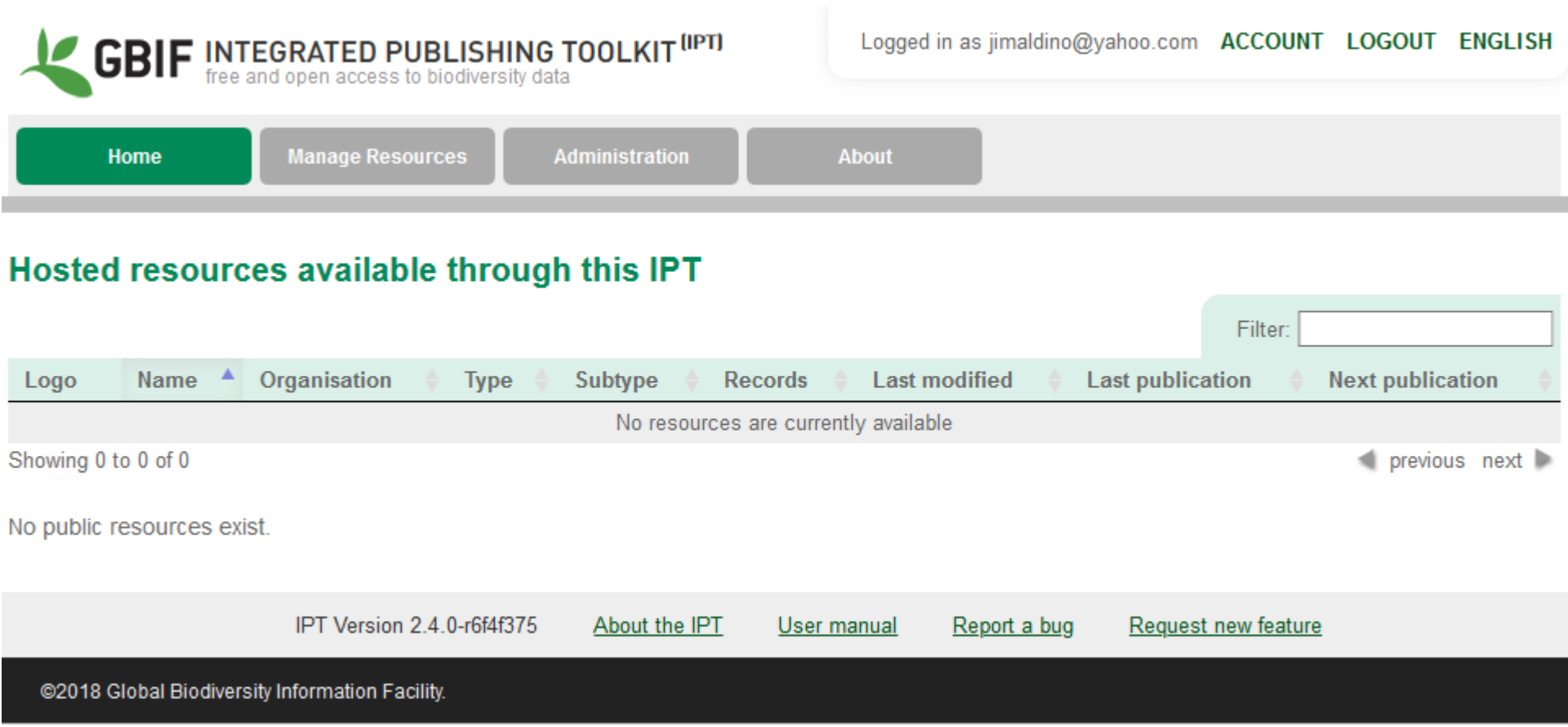
http://www.tdwg.org

Darwin Core


continent
taxonRank
basisOfRecord
kingdom
institutionCode
scientificNameID
family
institutionID
vernacularName
coordinatePrecision
recordedBy
taxonID
verbatimTaxonRank
originalNameUsage
nomenclaturalCode
nameAccordingTo
higherClassification
namePublishedInID
class
parentNameUsage
occurrenceID
originalNameUsageID
nameAccordingToID
order
higherGeographyID
associatedTaxa
verbatimCoordinateSystem
datasetID
minimumElevationInMeters
coordinateUncertaintyInMeters
parentNameUsageID
infraspecificEpithet
acceptedNameUsageID
genus
scientificNameAuthorship
behavior
collectionCode
previousIdentifications
maximumDepthInMeters
taxonConceptID
geodeticDatum
reproductiveCondition
decimalLongitude
namePublishedIn
phylum
catalogNumber
acceptedNameUsage
nomenclaturalStatus
taxonRemarks
specificEpithet
higherGeography
decimalLatitude
subgenus
taxonomicStatus
scientificName
islandGroup
lifeStage
locationID
collectionID
waterBody


Which type of data?
Metadata
Metadata
Metadata
Dataset
Taxonomic Data
Occurrence Data
GBIFS developed tools
Community tools
Select a tool
1
2
3
4
Integrated Publishing Toolkit
Spreadsheet Processor
DwC-A Assistant
TAPIR
DiGIR
BioCASE
Publish your data
Register with GBIF
GBIF NETWORK
Discovery through the Portal
http://data.gbif.org

https://ipt.buse.ac.zw/ipt/



Steps to follow when publishing data through GBIF

- Registration with GBIF
- Data Adaptation to Darwin Core Standards
- Data Quality Improvement
- Online Data Upload
- Dataset Publication