BiodivBank | A Global Repository & Portal for Structured Biodiversity Data

Anahita Kazem[1,2], Birgitta König-Ries[2], Henrique Pereira[1,3] & Jonathan Chase[1,3]

In a nutshell

The BiodivBank repository aims to create a one-stop access point, dedicated to species abundance & assemblage data

Comprehensive: All realms, any taxon, both legacy & current monitoring data, global in scope

Unlocking the potential hidden in grey literature, including non-English material

Global in access: emphasis on providing open & FAIR data

Directed at researchers, but providing user-friendly content for diverse stakeholders
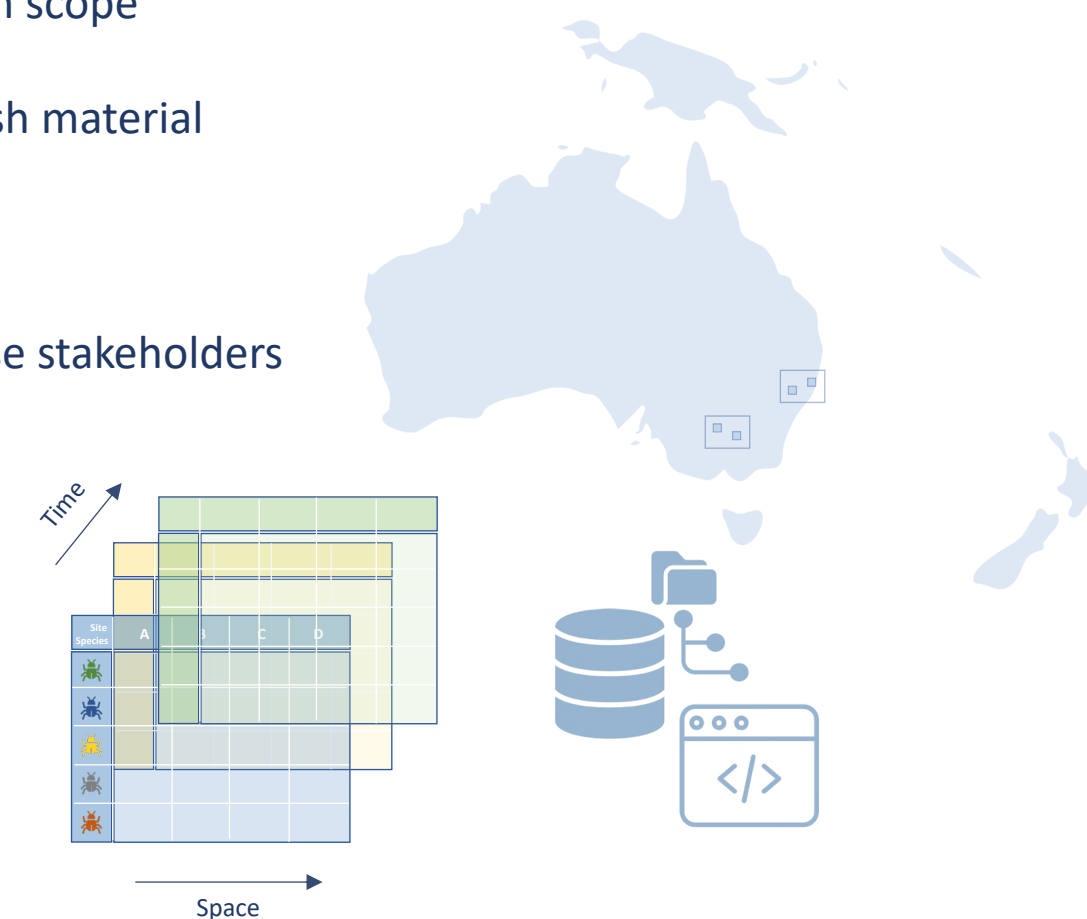
Data discovery

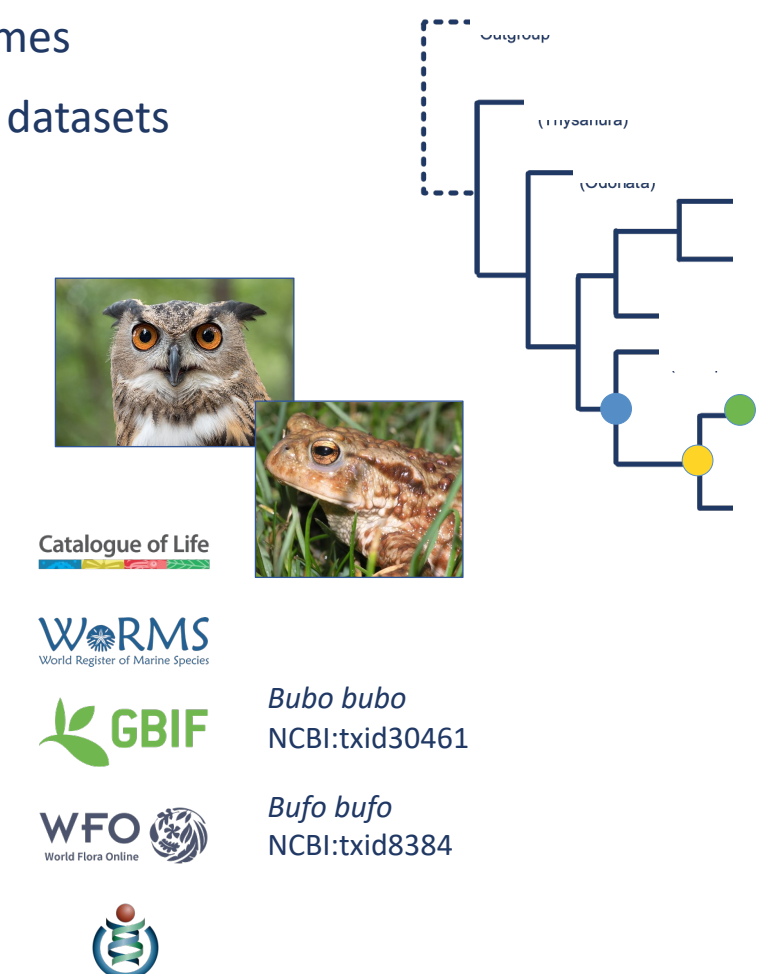Use schema.org markup to aid discovery by search engines & display more informative search results

Retrieval and storage of globally unique, persistent & resolvable identifiers (PIDs) for taxon names allows enrichment of search index & search by concepts (e.g. higher ranks) not present within datasets

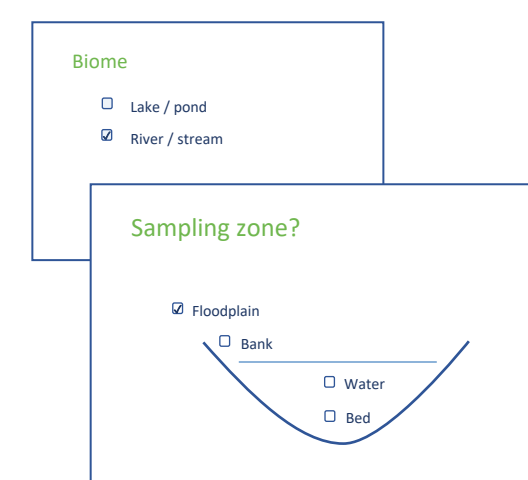Likewise, location names & coordinates can be annotated with regions & political entities

PIDs for authors, organizations, taxa & locations: facilitate future efforts to link open data within online knowledge networks

Bioschemas

FAIR + Open + Linked

Catalogue of Life
WoRMS
GBIF
WFO

Bubo bubo
NCBI:txid30461

Bufo bufo
NCBI:txid8384

Metadata editor

Biome
  Lake / pond
  River / stream

Sampling zone?
  Floodplain
    Bank
    Water
    Bed

Dynamic multi-level form

To be genuinely reusable, such datasets require detailed information on sampling methodology & effort

We are developing a graphical interface to help data providers describe datasets quickly & more accurately

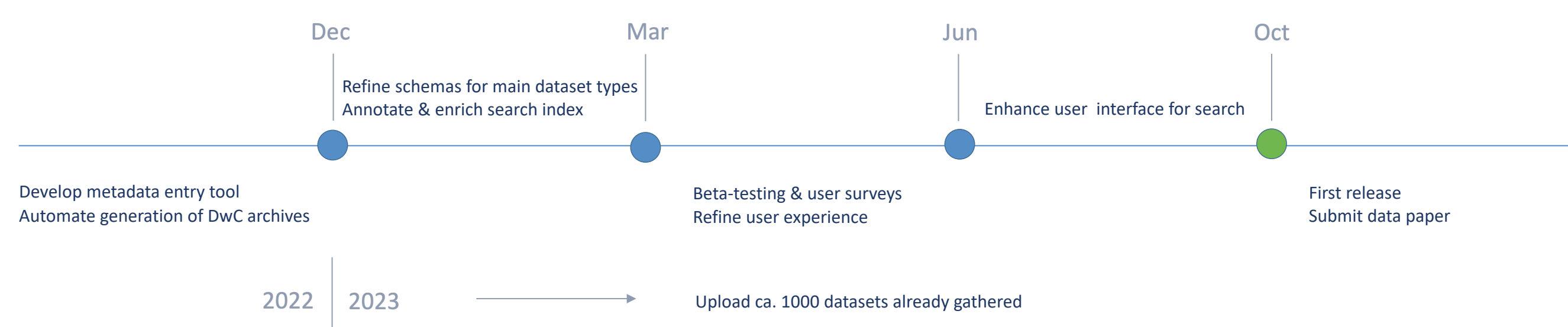Minimizing free text responses also reduces curation load for repository administrators

The pre-defined metadata attributes & values are mapped in advance to standard terms (e.g. DwC, ABCD, GGBN & their extensions) and ontologies - facilitating interoperability with other online resources

Connections with external services

o   Serialization of all datasets as DwC- or ABCD archives, enhancing dataset visibility via aggregators such as GBIF

o   Public API for script-based retrieval of datasets & support for multiple export formats: CSV, JSON, XML (metadata)

o   DOI assignment, with automated transfer of metadata to DataCite

o   Ability to export detailed metadata to data journals, facilitating production of data papers

GBIF   doi   PENSOFT

Current roadmap

Dec   Mar   Jun   Oct

Refine schemas for main dataset types
Annotate & enrich search index

Enhance user interface for search

Develop metadata entry tool
Automate generation of DwC archives

Beta-testing & user surveys
Refine user experience

First release
Submit data paper

2022 | 2023

Upload ca. 1000 datasets already gathered

Future extensions?

o   Community data space - enable feedback from registered users, to highlight dataset quality & flag errors

o   Stories - a rotating series of featured datasets to increase engagement & recommendation engine for users

o   Environmental layers - overlay with external data e.g. vegetation maps, hydrology, climate, or land use

o   Virtual integration - dynamic retrieval & reformatting of relevant datasets held in other repositories