



Digitization in a global environment

Donald Hobern
GBIF Director
Global Biodiversity Information Facility (GBIF)

Biological Collections Digitization in the Pacific Workshop

Honolulu, 25 March 2014

GBIF: origins and principles



Established in 2001

Response to OECD recommendation

Open to all countries

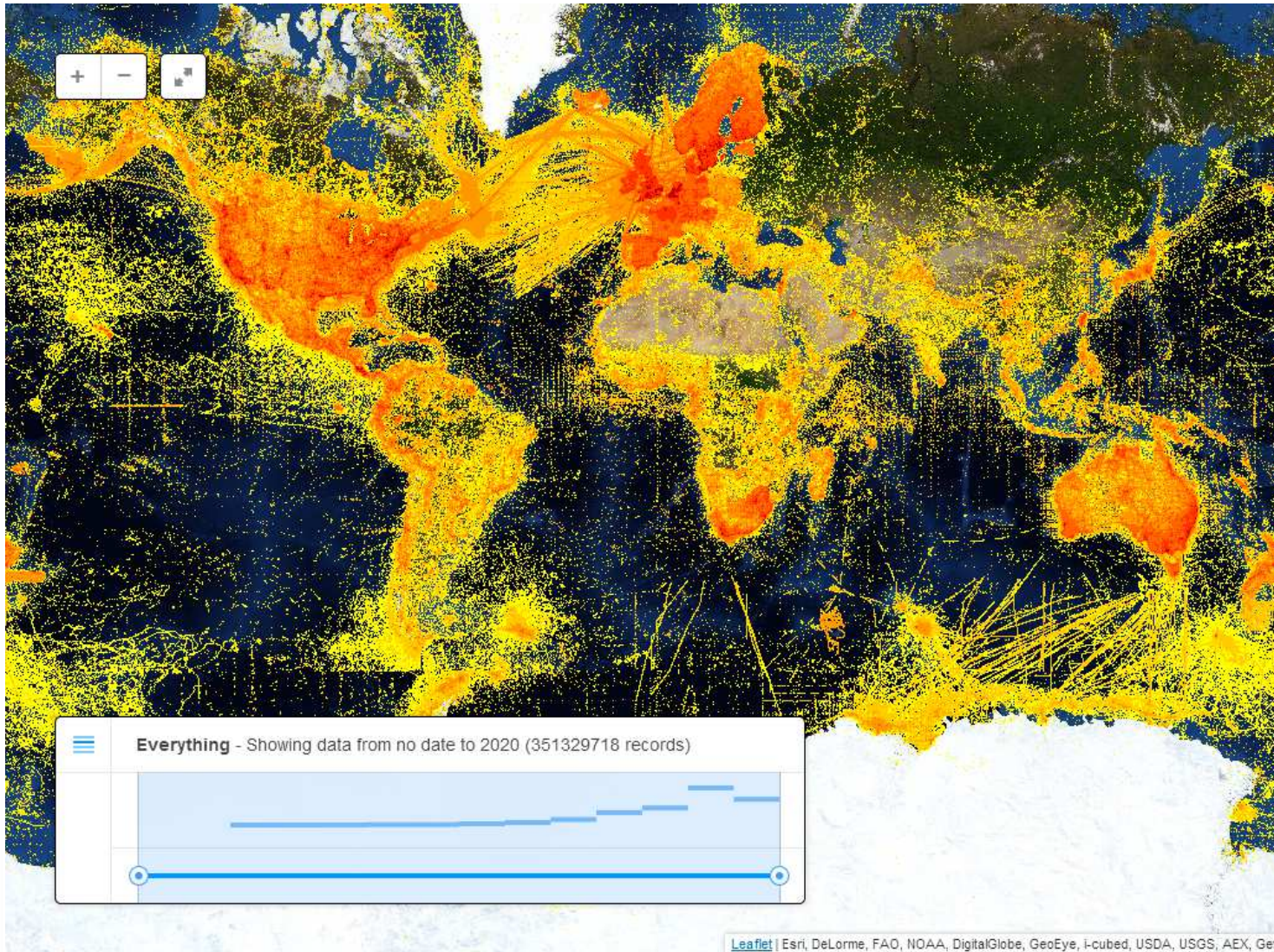
Voluntary memorandum of understanding (MoU)

Vision - a world in which:

Biodiversity information is freely and universally available for science, society and a sustainable future



12 years – 417 million records



Weaknesses with distributed data

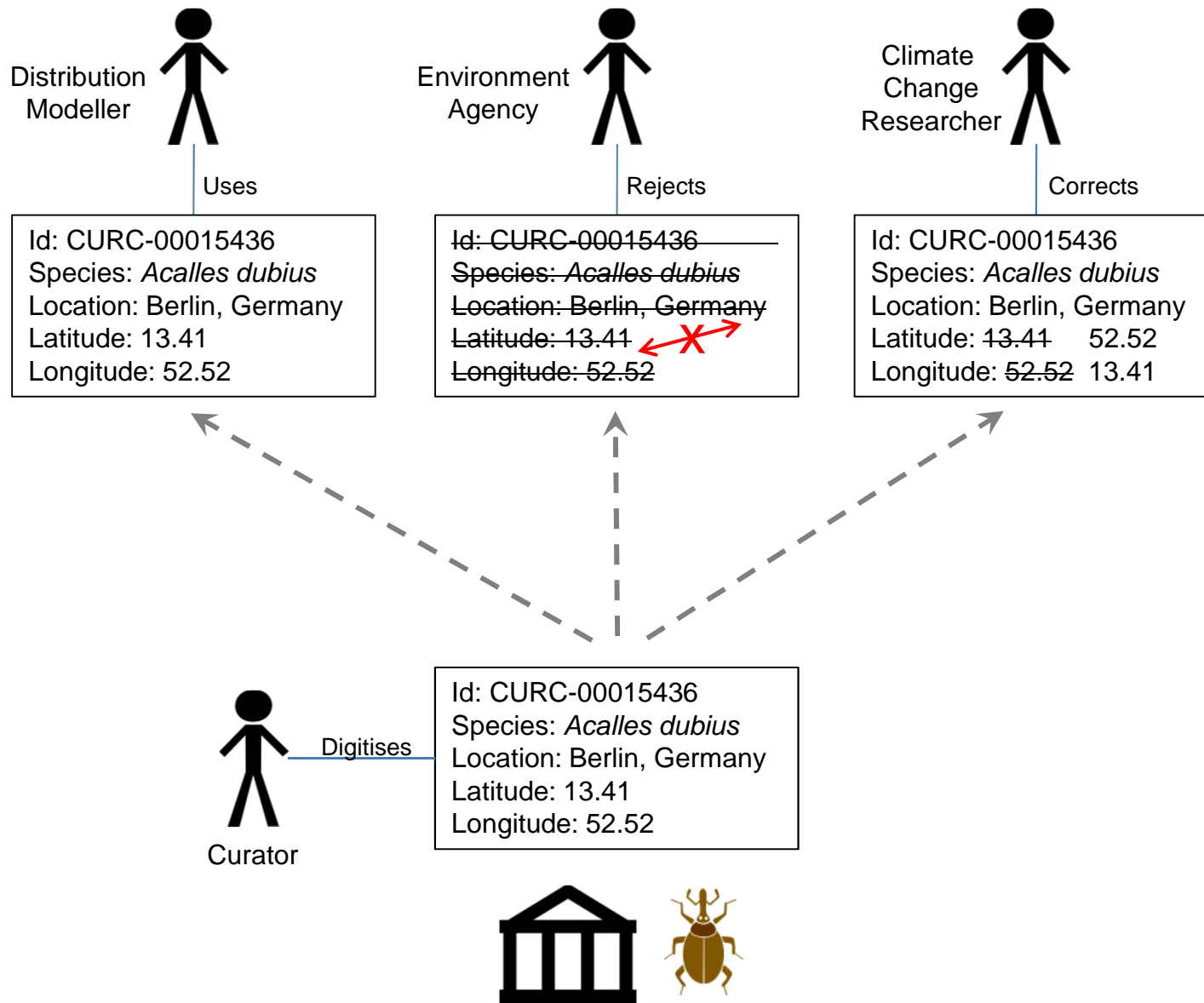


Digitises

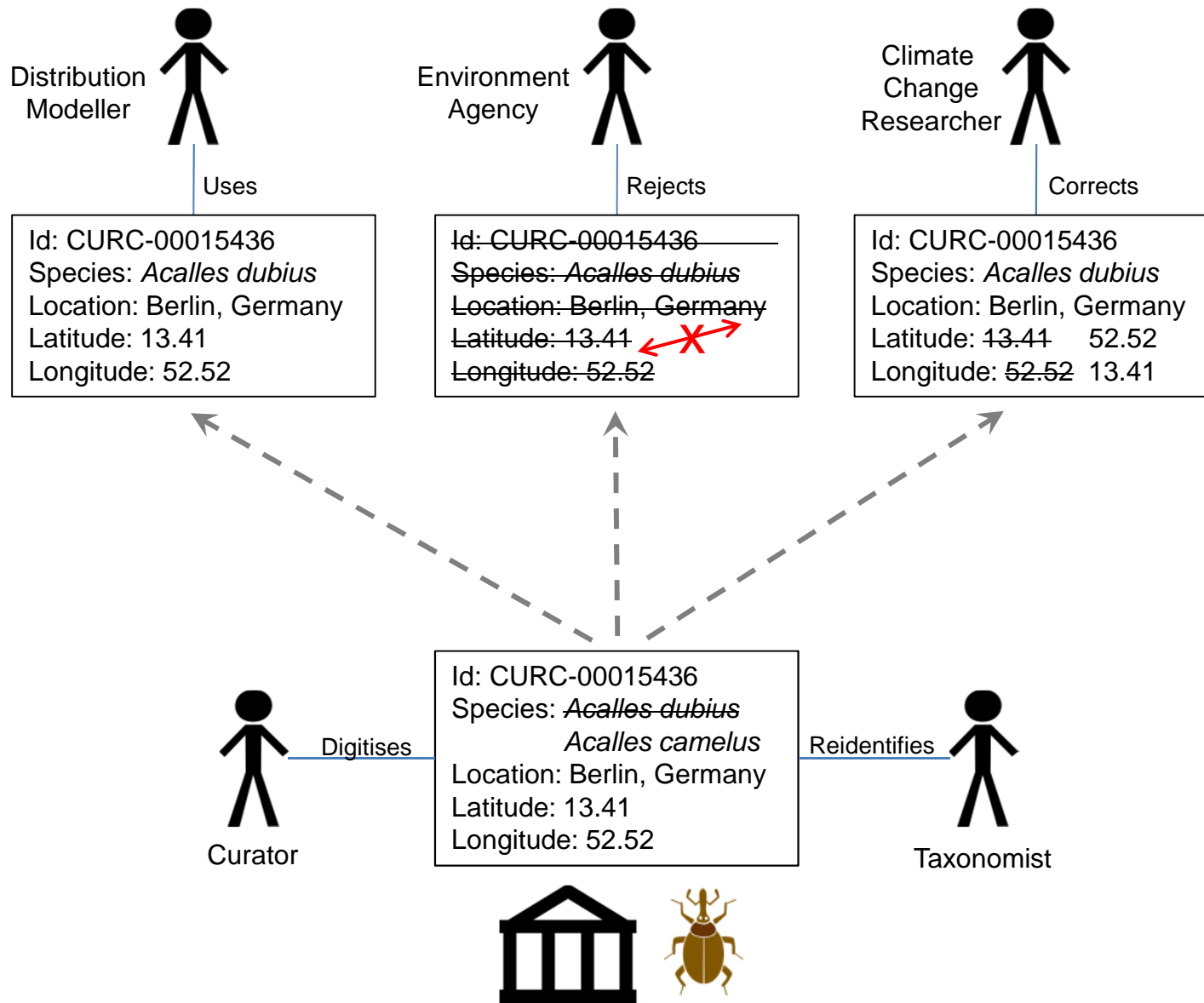
Id: CURC-00015436
Species: *Acalles dubius*
Location: Berlin, Germany
Latitude: 13.41
Longitude: 52.52



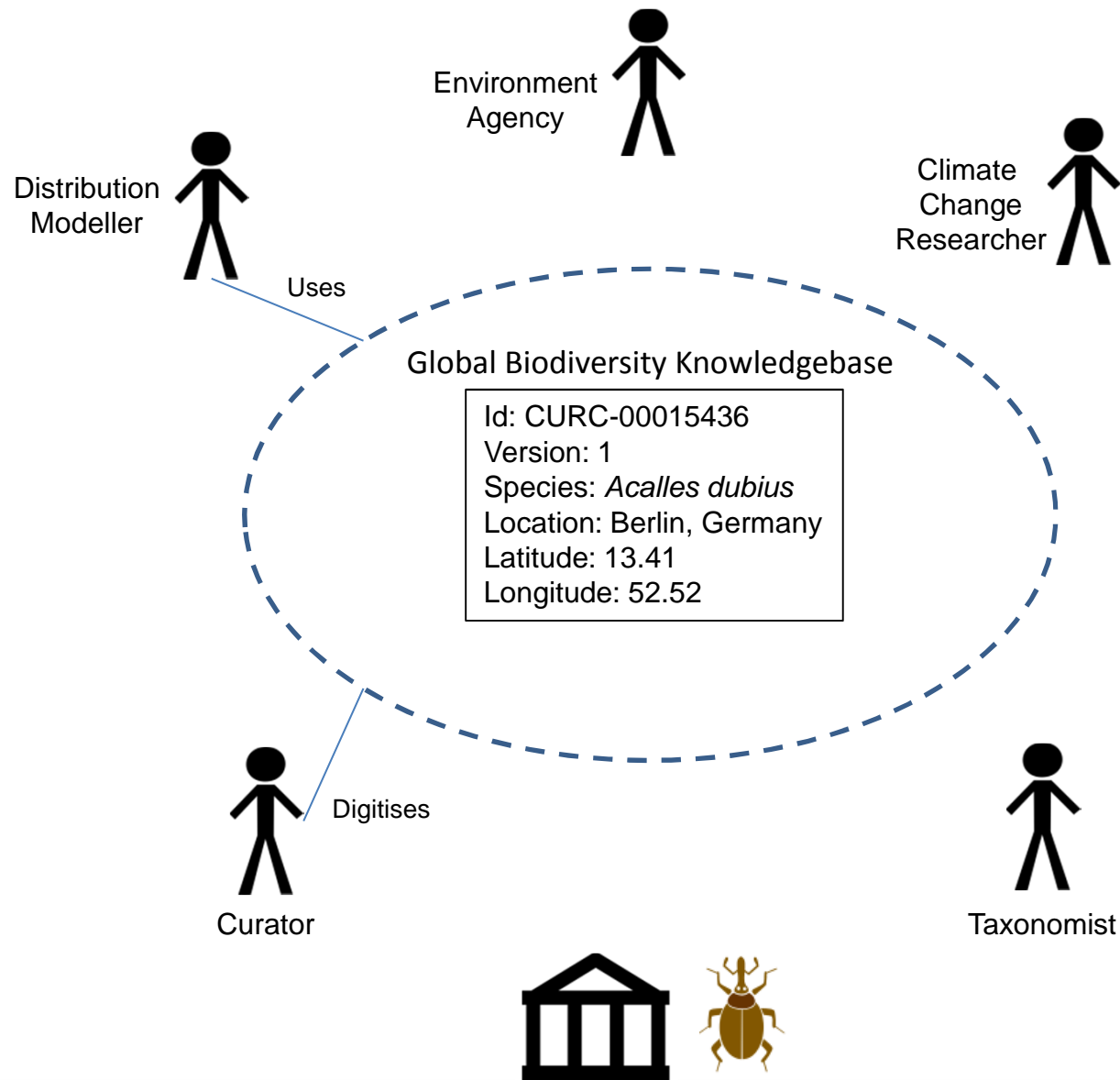
Weaknesses with distributed data



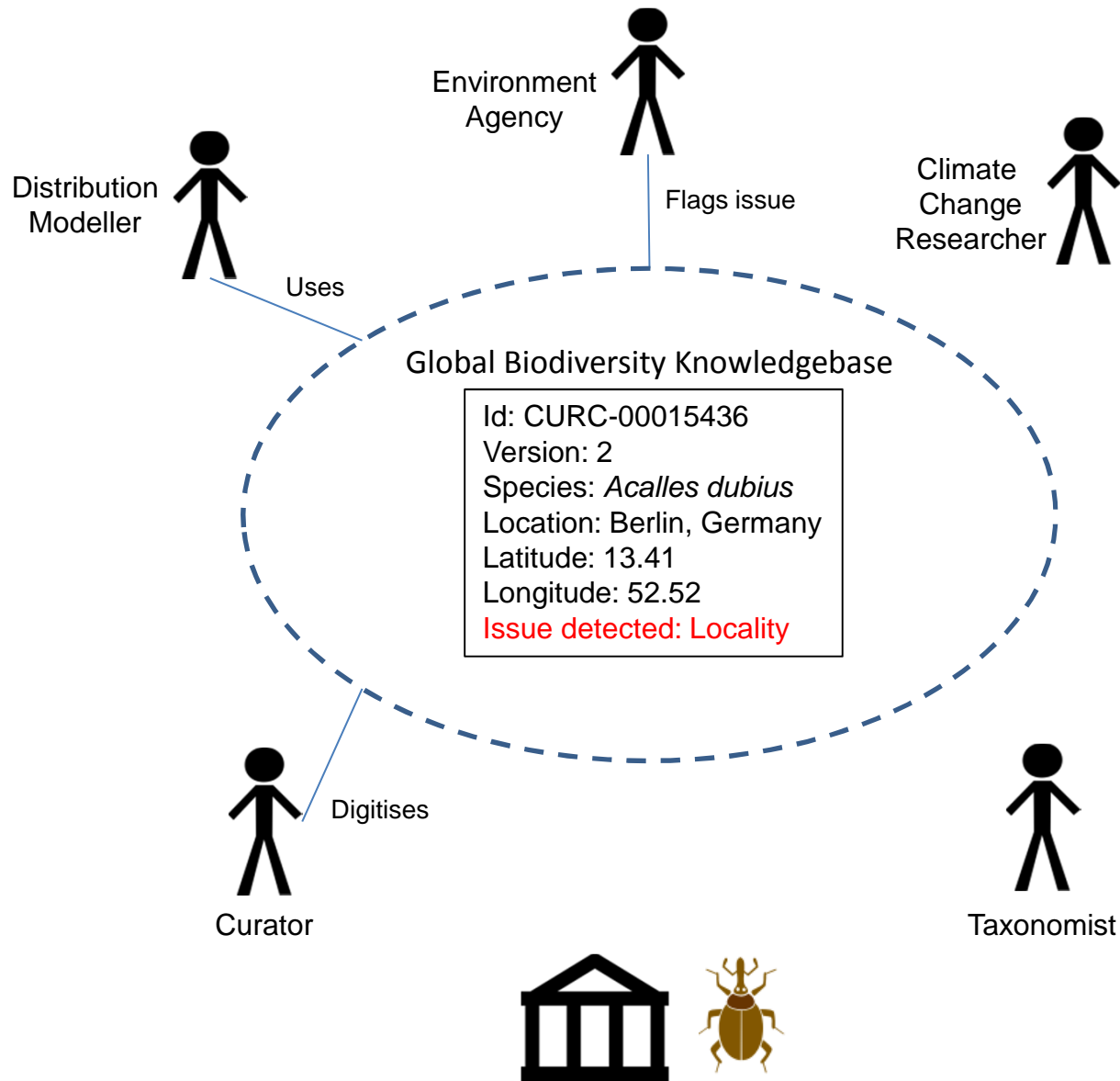
Weaknesses with distributed data



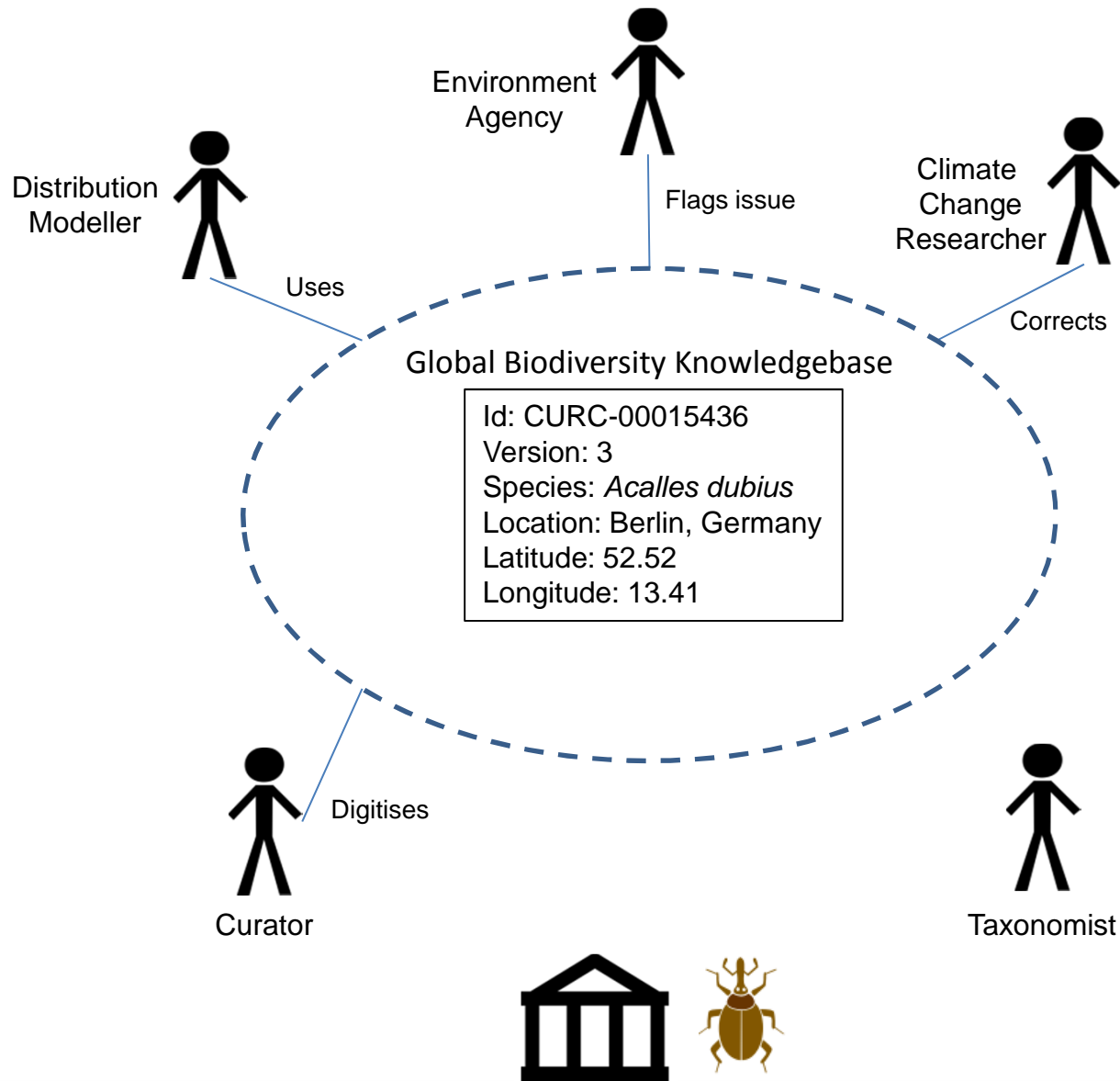
Toward a shared knowledgebase



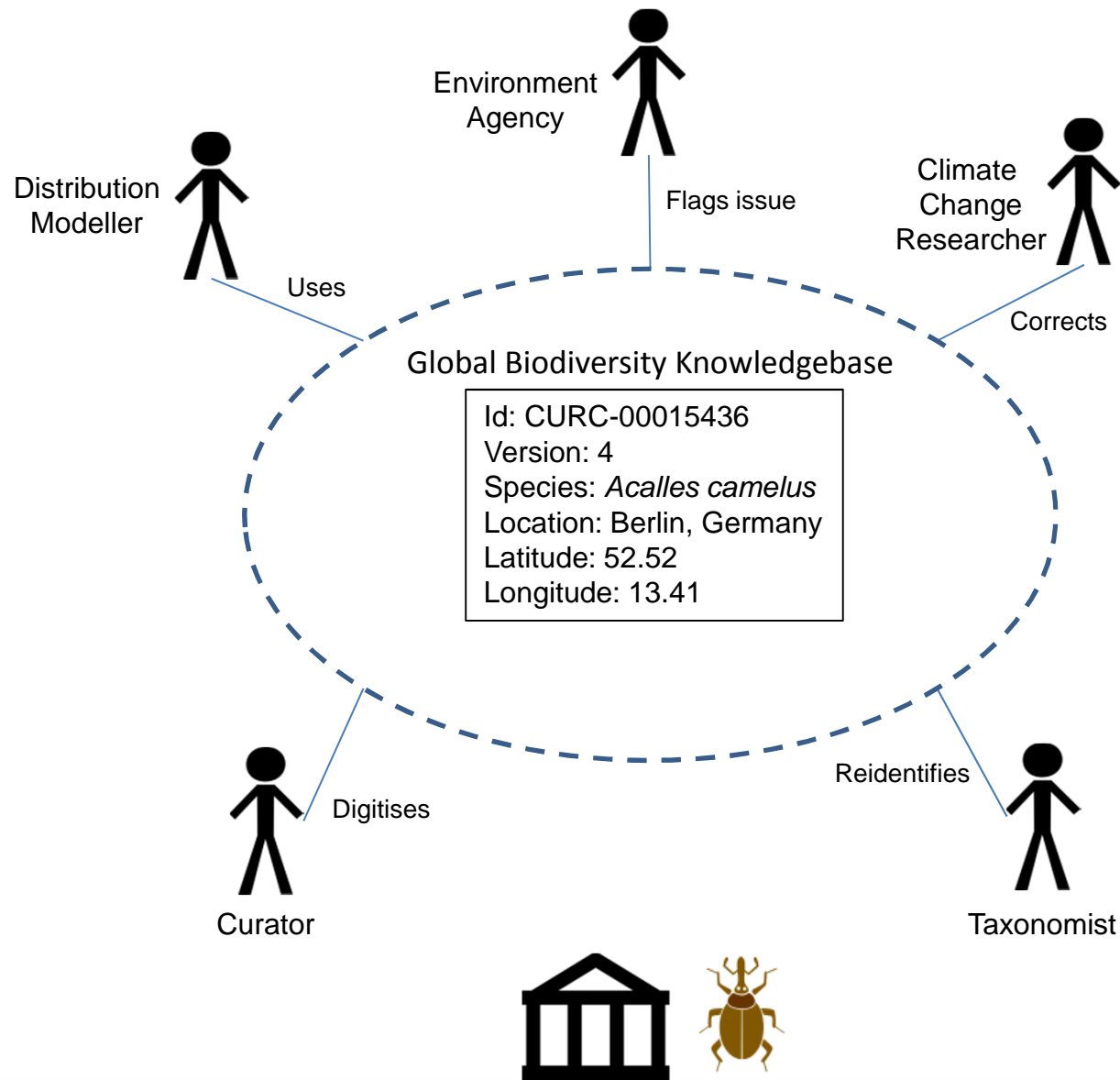
Toward a shared knowledgebase



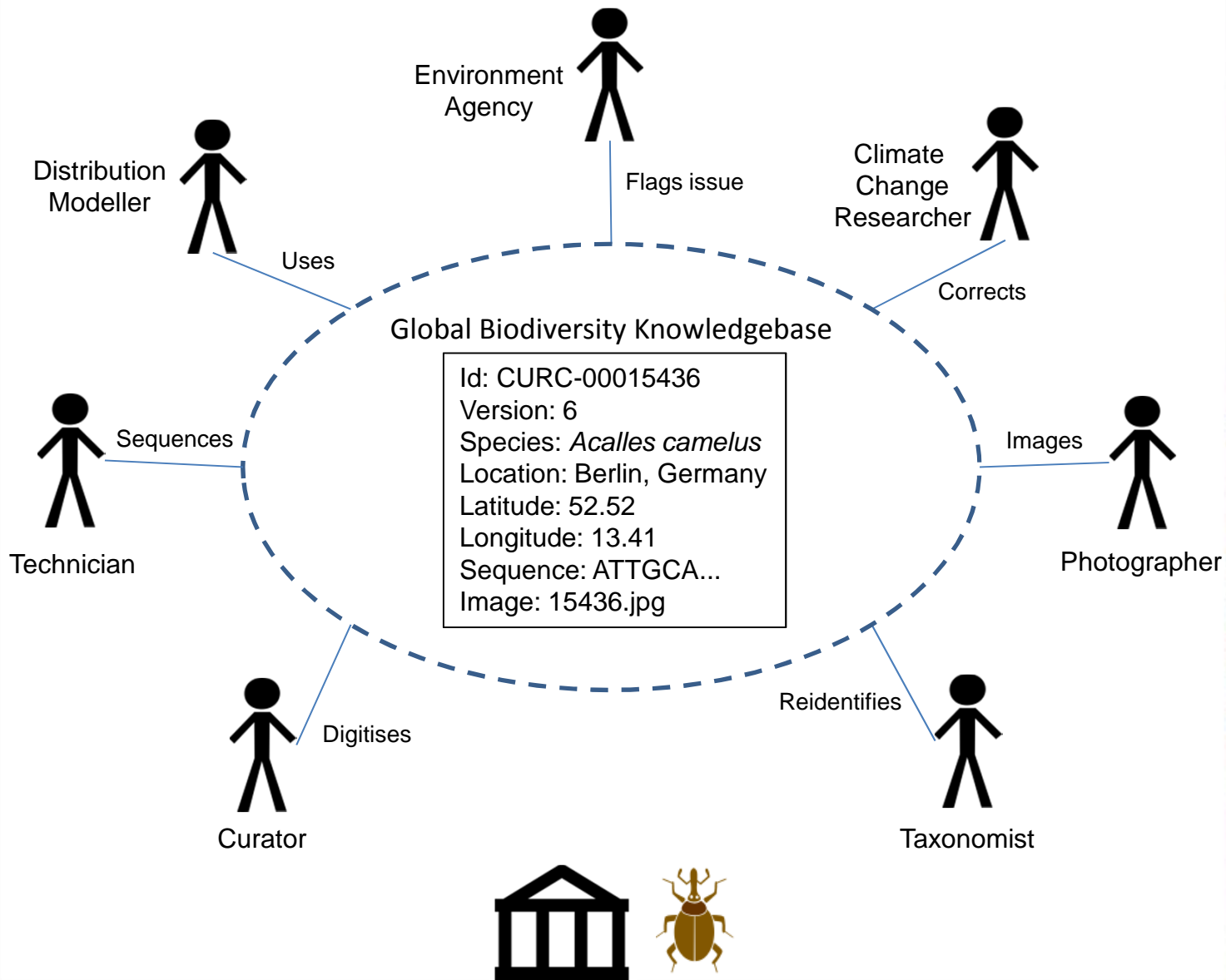
Toward a shared knowledgebase



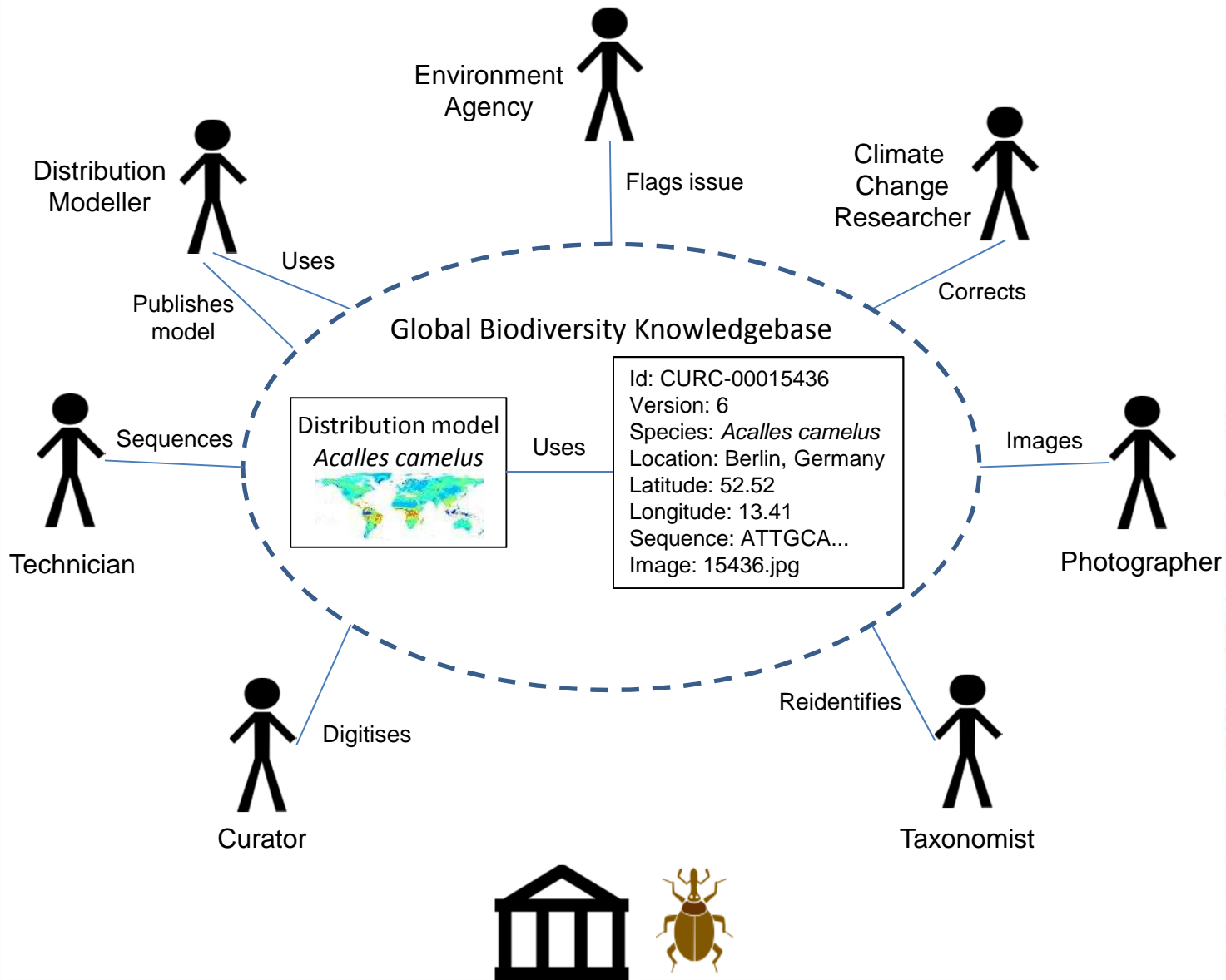
Toward a shared knowledgebase



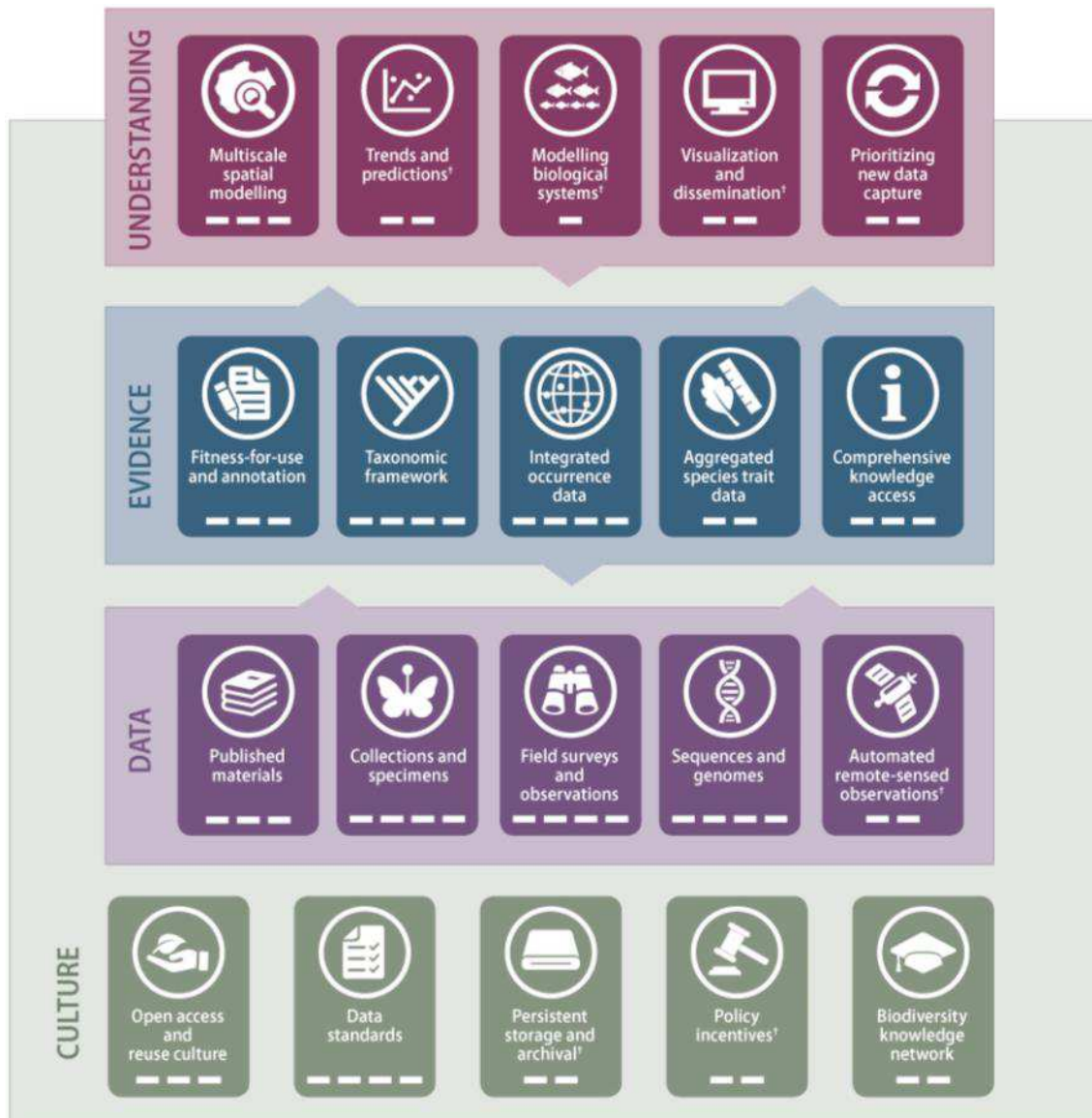
Toward a shared knowledgebase



Toward a shared knowledgebase



Global Biodiversity Informatics Outlook



Focus Area: Culture



- The context for sharing digital knowledge
 - Data must be **available for reuse**
 - Data must **follow standards** to support discovery and use
 - Data must be **preserved for future uses**
 - **Policies and practices** must reinforce open use
 - The whole community should collaborate to **curate data**
- Issues shared in common with all research domains
- Investments here will multiply value of other components



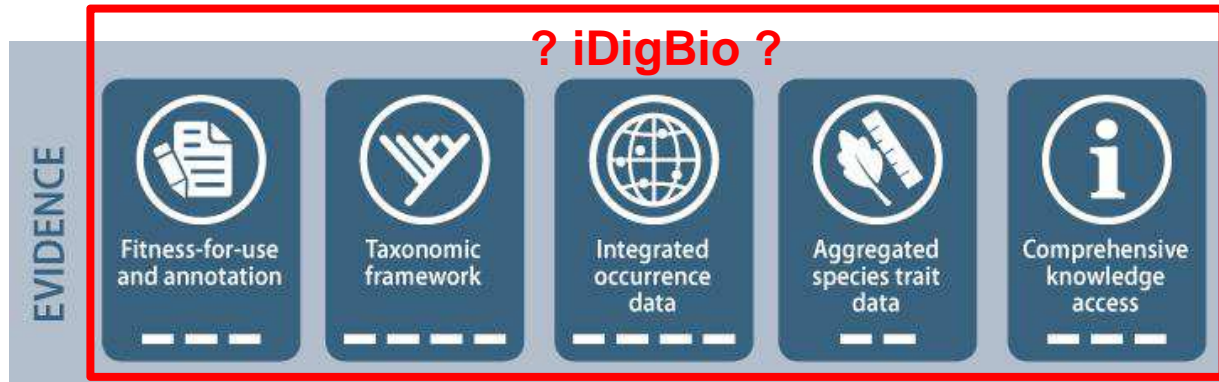
Focus Area: Data



- The streams of primary biodiversity data
 - **Literature** and journals
 - Natural history **collections**
 - Professional and amateur **field observations** and surveys
 - Molecular **sequencing**
 - **Remote sensing** (including camera traps, acoustic monitoring, etc.)
- All deliver fundamental observations and measurements of biodiversity
- Foundations for analysis and understanding



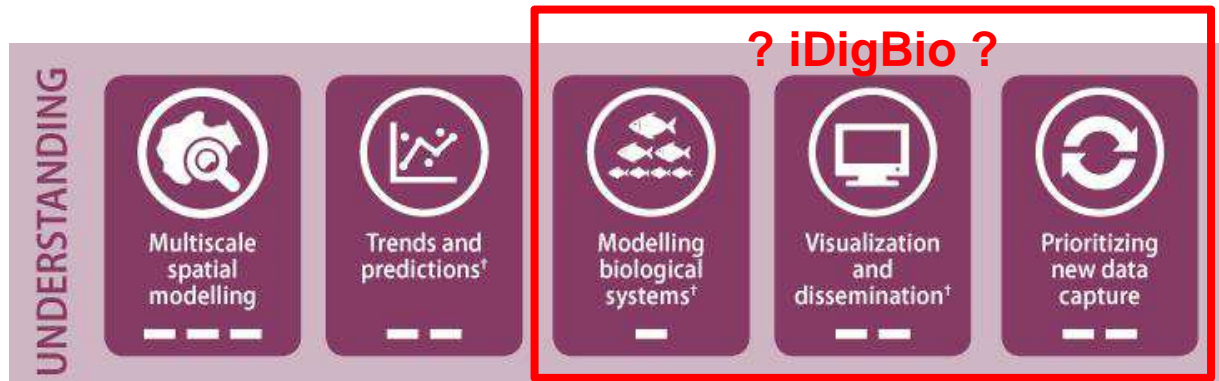
Focus Area: Evidence



- Organised views of biodiversity data
 - Consistent assessment of **quality and fitness-for-use**
 - Comprehensive digital **nomenclature and taxonomy**
 - Access to all evidence for recorded **species occurrence**
 - Access to species **traits, measurements and interactions**
 - **Services and interfaces** to access data as needed
- Provide comprehensive organised views of all relevant data
- Act as a “lens” into primary data



Focus Area: Understanding



- The application of data to address questions
 - Integrate data into **spatial models**
 - Develop **temporal analyses**
 - Incorporate **biological reality** into models
 - Present **compelling representations** of biodiversity
 - Optimise **future investment** in biodiversity informatics
- Data-driven models for science and planning
- Integrate biodiversity with other research and data domains



Data → Evidence → Understanding



Example for occurrence data

Data

Dataset A asserts that species X was recorded at a given locality on a given date



Evidence

Community assessment concludes that species X was recorded at a given set of localities on given dates



Understanding

The best available model presents the probability that species X was present at any locality on any date



Collections and biodiversity data

Virtual natural history collection	Ecoinformatics resource
Focus: collection objects	Focus: geospatial data
Goal: integrated access to all collection materials for any species	Goal: maximise reliable evidence for species occurrence in time and space (collections are core data)
Primary data elements: <ul style="list-style-type: none">• Identification history• Collection identifiers• Locality• Images, sequences• Morphometric data	Primary data elements: <ul style="list-style-type: none">• Scientific name• Coordinates• Date• Confidence/evidence level
Critical linkages: <ul style="list-style-type: none">• Nomenclators• Biodiversity Heritage Library• Barcode of Life Database• Phylogeny data	Critical linkages: <ul style="list-style-type: none">• Climate and environment• Politics and land use• Stable classification

Two major applications



Planning digitization

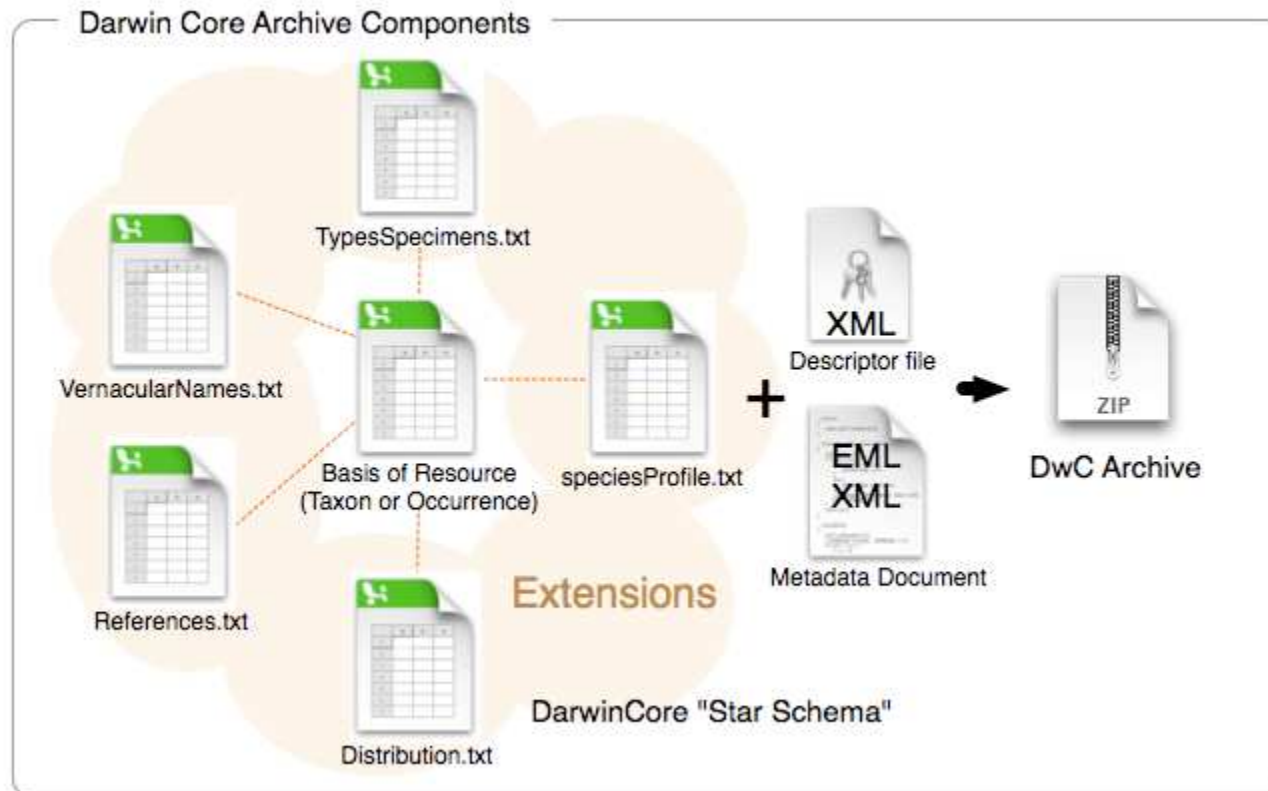
- Understand content of collections
 - Collection type: wet, dry, pinned, tissue, living, ...
 - Factual: size, taxa, geography, types, curation status, subcollections, ...
 - Relevance: time series, protected areas, threatened species, indicator taxa, ...
 - Planning: digitisation options, costs/benefits
- Metadata for discovery and publicity
- Inputs to institutional/national strategy
- Opportunity for collaborative funding or labour



The screenshot shows the ATLAS of Living Australia website. The header includes the logo 'ATLAS of LIVING AUSTRALIA' with the tagline 'sharing biodiversity knowledge', a search bar, and navigation links for 'My profile', 'Log in', and 'Search the Atlas'. The main navigation bar contains 'Species', 'Locations', 'Collections', 'Mapping & analysis', 'Data sets', 'Blogs', 'Get involved', and 'About the Atlas'. The page title is 'Australia's natural history collections'. Below the title, there is a brief description and a note that currently only Atlas partner collections are shown. There are three filter buttons: 'All collections' (152 collections), 'Fauna' (mammals, birds, reptiles, fish, amphibians and invertebrates), and 'Insects' (insects, spiders, mites and some other arthropods). A map of Australia is displayed with several collection locations marked by pins. The map includes labels for 'Timor Sea', 'Northern Territory', and 'Coral Sea'.



Publishing data



Darwin Core Archive

- GBIF's preferred standard for sharing data
- ZIP file with data spreadsheet and metadata



Supporting data reuse



About For researchers For organizations Contact us Login Sign Up



Canadensys

data | community

about digitization data publication activities blog



About Pensoft Books E-Books Journals News & Blog Contact

Donald Hobern



A peer-reviewed open-access journal
Biodiversity Data Journal
 Making your data count! ISSN 1314-2828 (online)

All Author Title

Your tasks Start a manuscript

How it works

Articles About

Journal features

- Focus and Scope
- Globally unique innovations
- Criteria for publication
- Peer review
- For authors
- Data publication
- Publication fees
- Licenses and Copyright
- Frequently Asked Questions (FAQ)
- Web services
- Contacts
- Editorial team



Resolving the publishing bottleneck for biodiversity

Science is a combination of gathering facts and making theories; neither can progress on its own. In the history of science, the laborious accumulation of facts is the dominant mode, not a novelty.

Peter Norvig

CC0 1.0 Universal (CC0 1.0) Public Domain Dedication

This is a human-readable summary of the [Legal Code](#) (read the full text)

Disclaimer

No Copyright



The person who associated a work with this deed has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. See [Other Information](#) below.



Other Information

- In no way are the patent or trademark rights of any person affected by CC0, nor are the rights that other persons may have in the work or in how the work is used, such as [publicity or publicity rights](#).
- Unless expressly stated otherwise, the person who associated a work with this deed makes no warranties about the work, and disclaims liability for all uses of the work, to the fullest extent permitted by applicable law.
- When using or citing the work, you should not imply [endorsement](#) by the author or the affirmer.



Attribution 2.0 Generic (CC BY 2.0)

This is a human-readable summary of the [Legal Code](#) (the full license)

Disclaimer

You are free:

- to **Share** — to copy, distribute and transmit the work
- to **Remix** — to adapt the work
- to make commercial use of the work



Under the following conditions:

- Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

With the understanding that:

Open Data Commons [Log out for Open Data](#) Home Licenses Terms FAQ News About

Open Database License (ODbL) v1.0

Disclaimer

Open Data Commons is not a law firm and does not provide legal services of any kind.

Open Data Commons has no formal relationship with you. Your receipt of this document does not create any kind of agent-client relationship. Please seek the advice of a suitably qualified legal professional licensed to practice in your jurisdiction before using this document.

No warranties and disclaimers of any damages. This information is provided "as is," and this site makes no warranties on the information provided. Any damages resulting from its use are disclaimed.

Plain language summary

A plain language summary of the Open Database License is available.

Alternative formats:

Plain Text

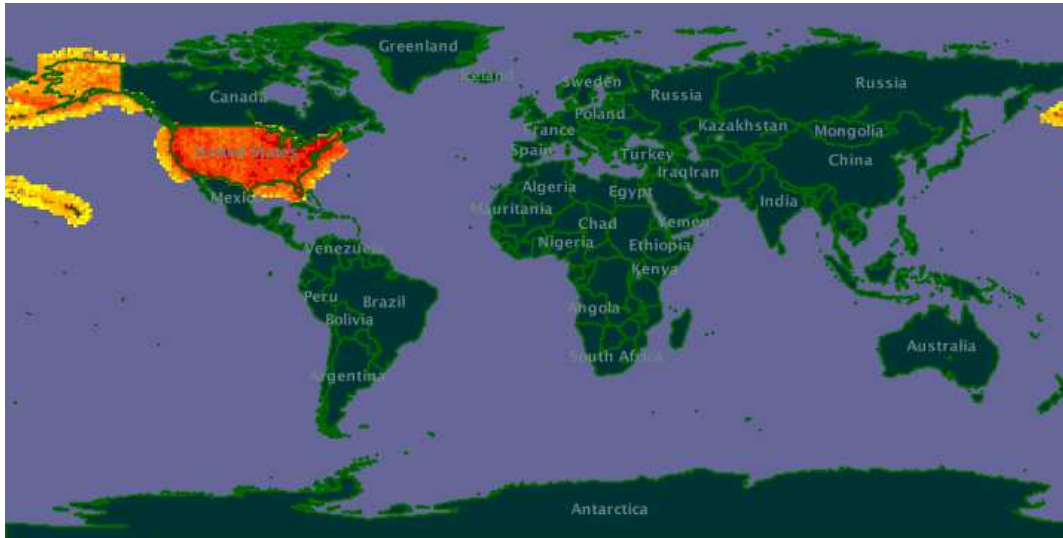
ODC Open Database License (ODbL)

MORE INFORMATION

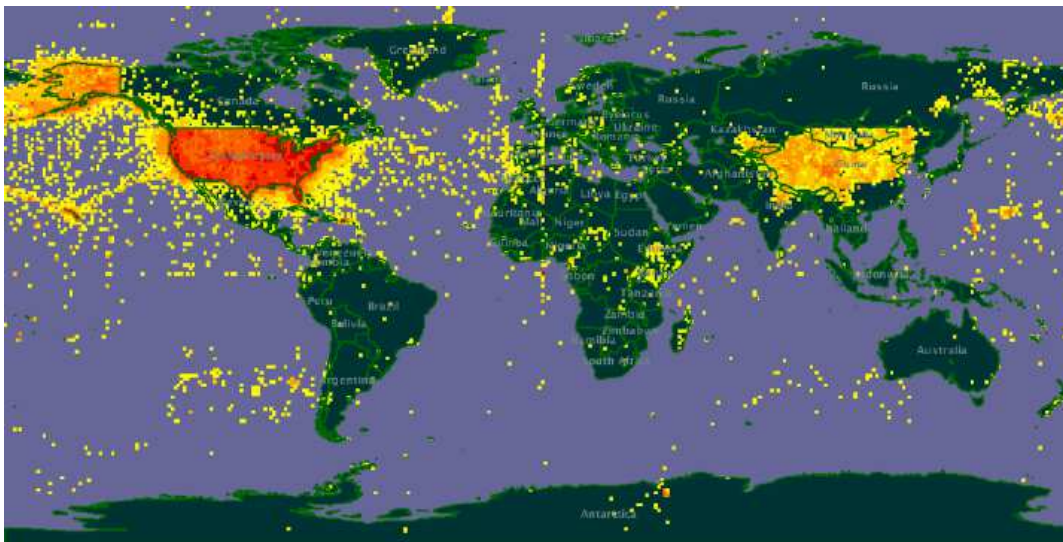
- Introduction to Open Data
- Open Definition for Data
- Quick guide to making data open
- Open Data Handbook



Integrating data



GBIF data for "United States" intersecting USA



All GBIF data for "United States"



Collaborative curation

Great Crested Newt (Denmark)

Observed by [Donald Hobern](#) 🌱🌱🌱 on 14th April 2013

(Added to iSpot on 14th April 2013)



Apparently dried out on gravel path.

Location: 2, Ørehøj Alle

Identifications

Smooth Newt (*Lissotriton vulgaris*) by [Donald Hobern](#) 🌱🌱🌱 at

7:35 pm 14/04/13

Confidence: I'm as sure as I can be.

Search **Encyclopedia of Life** for *Lissotriton vulgaris*

Search **Wikipedia** for *Lissotriton vulgaris*

View **NBN** map for *Lissotriton vulgaris*

Great Crested Newt (*Triturus cristatus*) by [Masked Marvel](#) **likely ID**

at 10:25 pm 14/04/13

Confidence: I'm as sure as I can be.

Notes: The black throat, deep orange belly and orange stripe along the bottom the tail are characteristics of a great crested newt.

I agree!

ID agreements (): 4 people agree with this identification.




Reputation in groups

Group	Reputation	Observations	Identifications	Received	Given
Other organisms	?	0	5	17	15
Birds	👍👍👍	62	125	428	952
Invertebrates	🌱🌱🌱🌱	135	412	546	921
Fish	👍	0	21	19	5
Amphibians and Reptiles	🌱🌱🌱	3	15	50	61
Mammals	👍👍	3	21	74	71
Plants	🌱🌱🌱	30	36	126	65
Fungi and Lichens	🌱🌱	6	14	32	15
totals		239	649	1292	2105



Collaborative curation

iNaturalist.org

 Taxonomic Split 4518 (Committed on 2013-04-10)

[edit taxon change](#)

[Update your content](#)

Following the Systematics and taxonomy of Australian Birds (and also Clements/eBird), *R. fuliginosa* is confined to New Zealand. All else is now considered *R. albiscapa*. If your observations are from New Zealand, we recommend you reID them as the narrower concept of *R. fuliginosa*, otherwise reID them as *R. albiscapa*.

Source: Christidis, Les; Boles, Walter (2008)... (Citation)

Added by loarie on 2013-04-08 | Committed by loarie on 2013-04-10



Rhipidura fuliginosa 8161

6 Obs | LC | Range | Inactive | Flag for curation

Schemes: IUCN Red List of Threatened Species, Version 2012.2, IUCN Red List of Threatened Species, Version 2012.1

split into



Rhipidura albiscapa 144737

20 Obs | Range | Active | Flag for curation

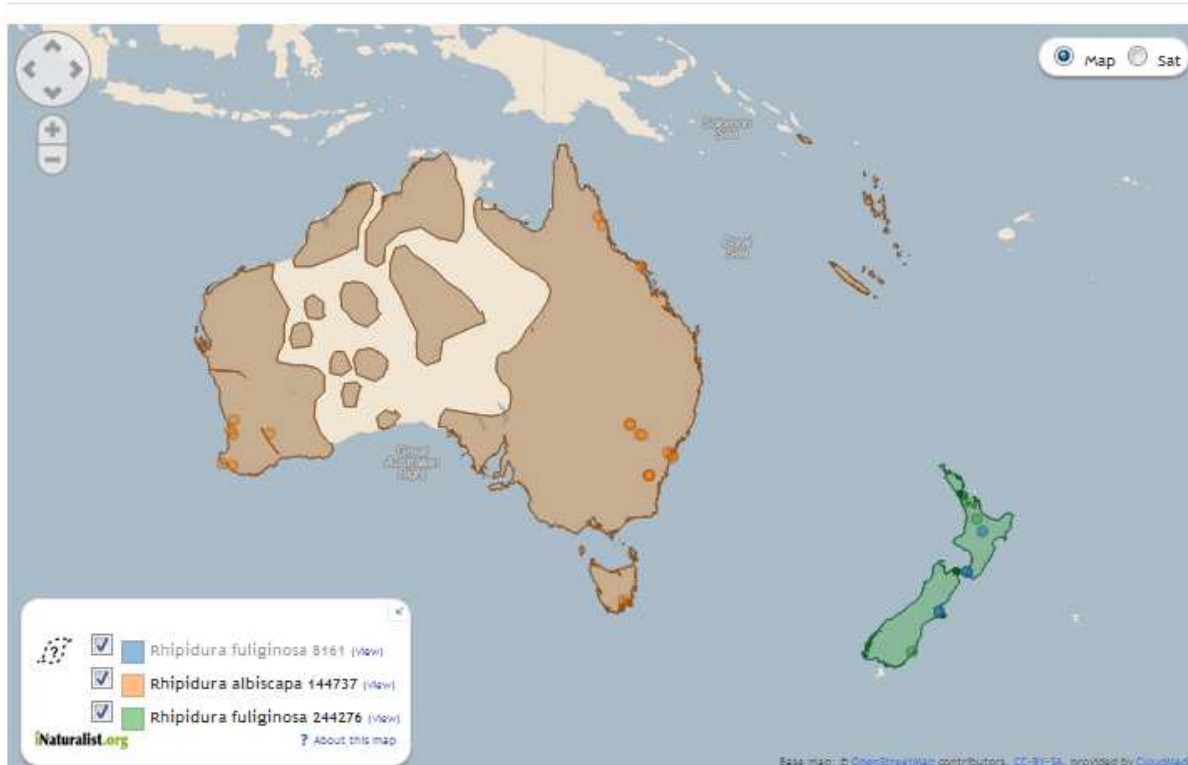
Schemes: eBird/Clements Checklist 6.7



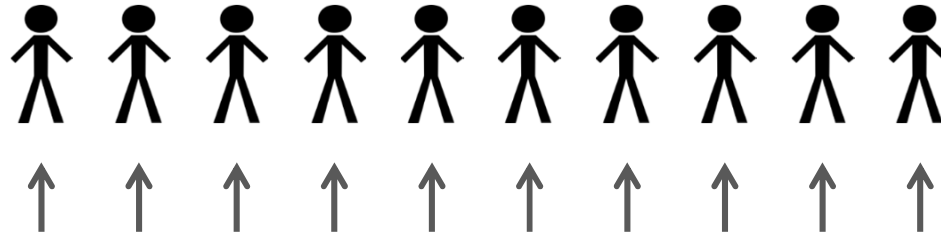
Rhipidura fuliginosa 244276

3 Obs | Range | Active | Flag for curation

Schemes: eBird/Clements Checklist 6.7

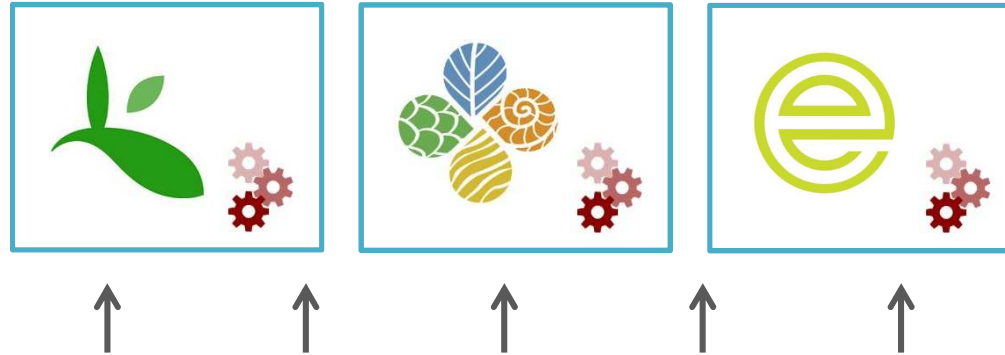


Users



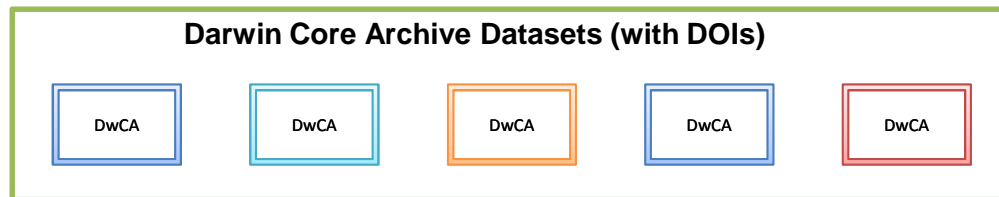
Access Portals

iDigBio, GBIF, BISON
Normalised and verbatim data
Validation and annotation
Integrated access
"EVIDENCE"



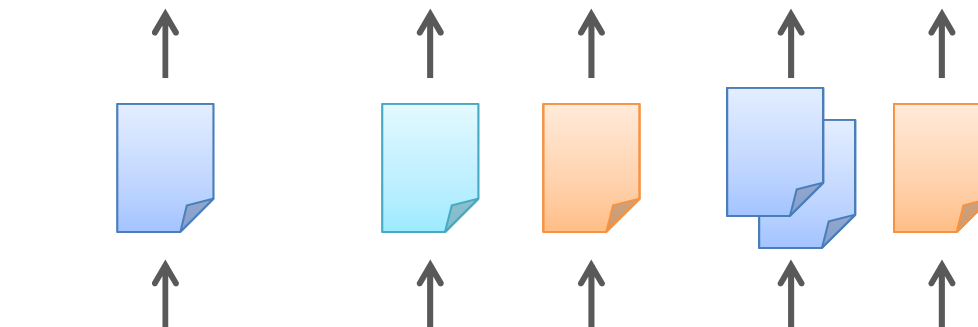
Data Repository

Dryad, Data ONE, GBIF IPT
Stable location, digital licence,
annotation tools, data papers
Stable dataset identifiers
"DATA"



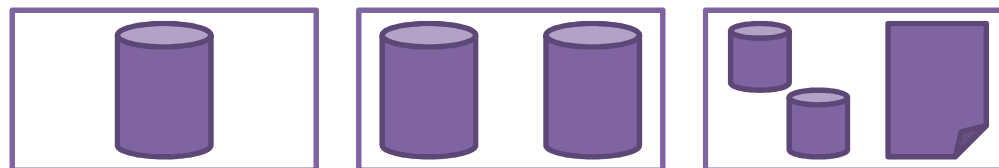
Spreadsheet Data

(DwC-A / CSV / Excel / XML)
Data and metadata for free and
open use
Standards compliance

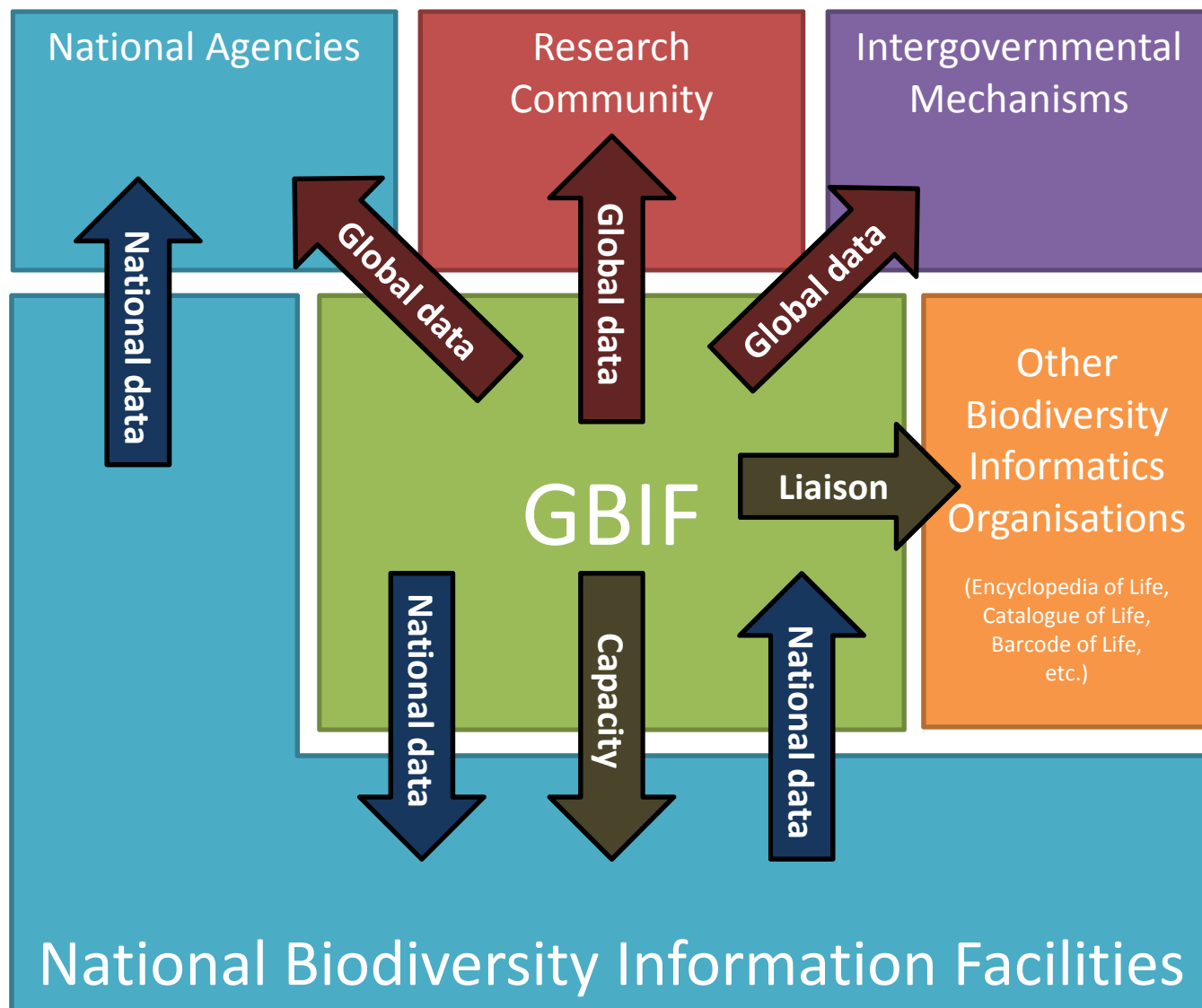


Data Publishers

Biodiversity data in
institutional databases
Stable record identifiers



GBIF in the landscape





Thank you

Donald Hobern
GBIF Director
Global Biodiversity Information Facility (GBIF)

Biological Collections Digitization in the Pacific Workshop

Honolulu, 25 March 2014