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

GBIF

Established in 2001

Response to OECD recommendation

Open to all countries

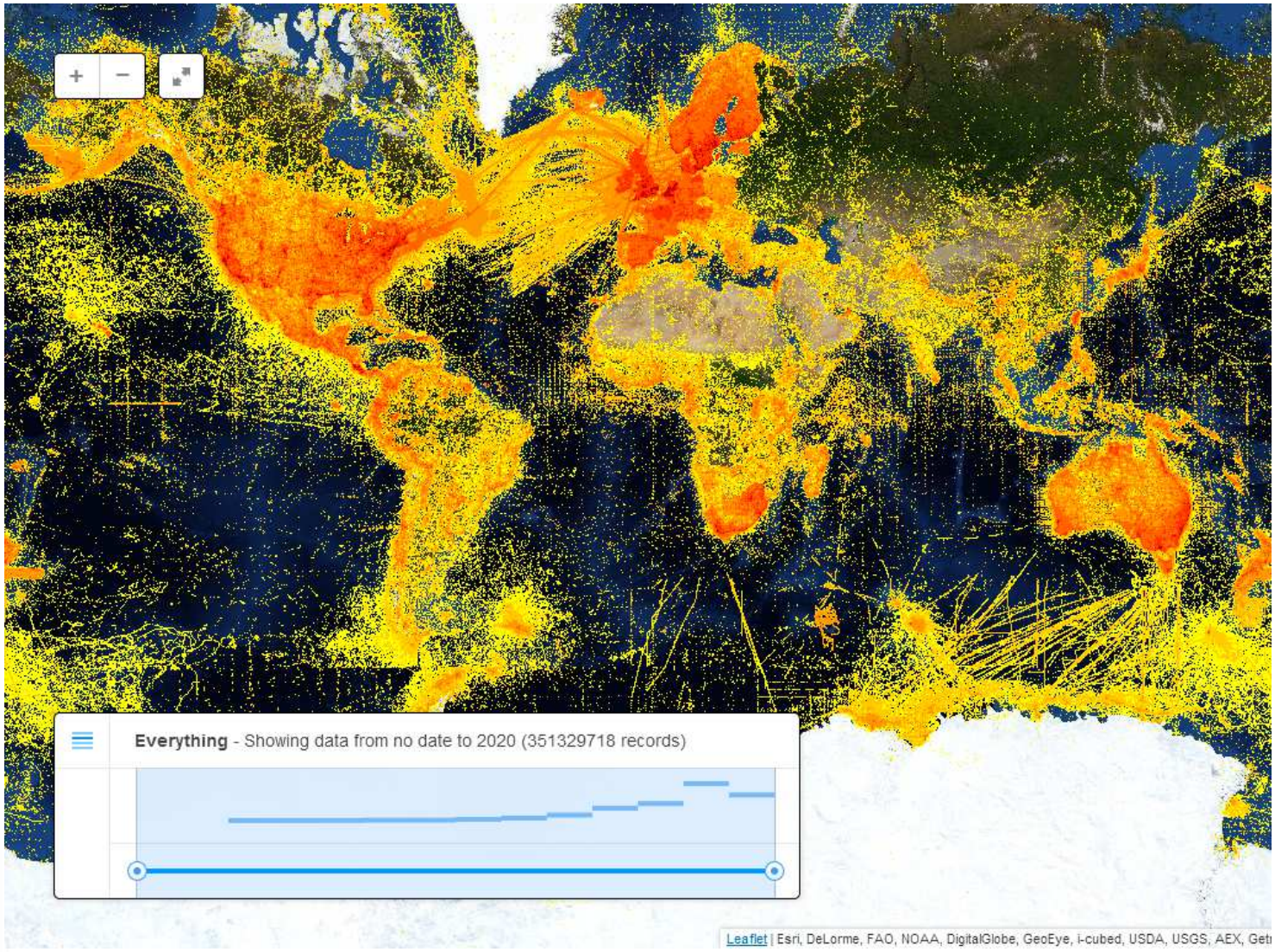
Voluntary memorandum of understanding (MoU)

Vision - a world in which:

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

GBIF
www.gbif.org

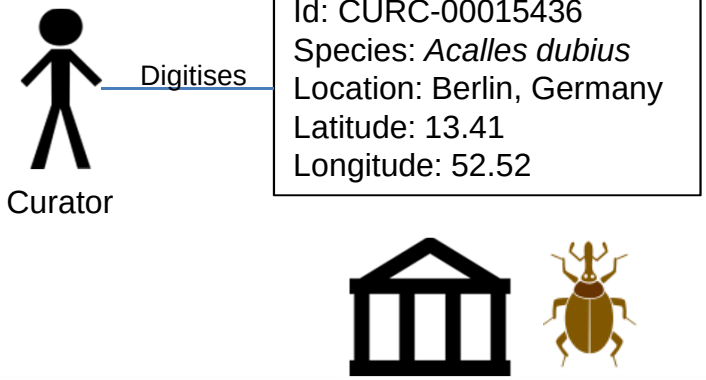
12 years – 417 million records



GBIF
www.gbif.org

Weaknesses with distributed data

GBIF
www.gbif.org



Weaknesses with distributed data

Distribution Modeller

Uses

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

Environment Agency

Rejects

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

Climate Change Researcher

Corrects

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

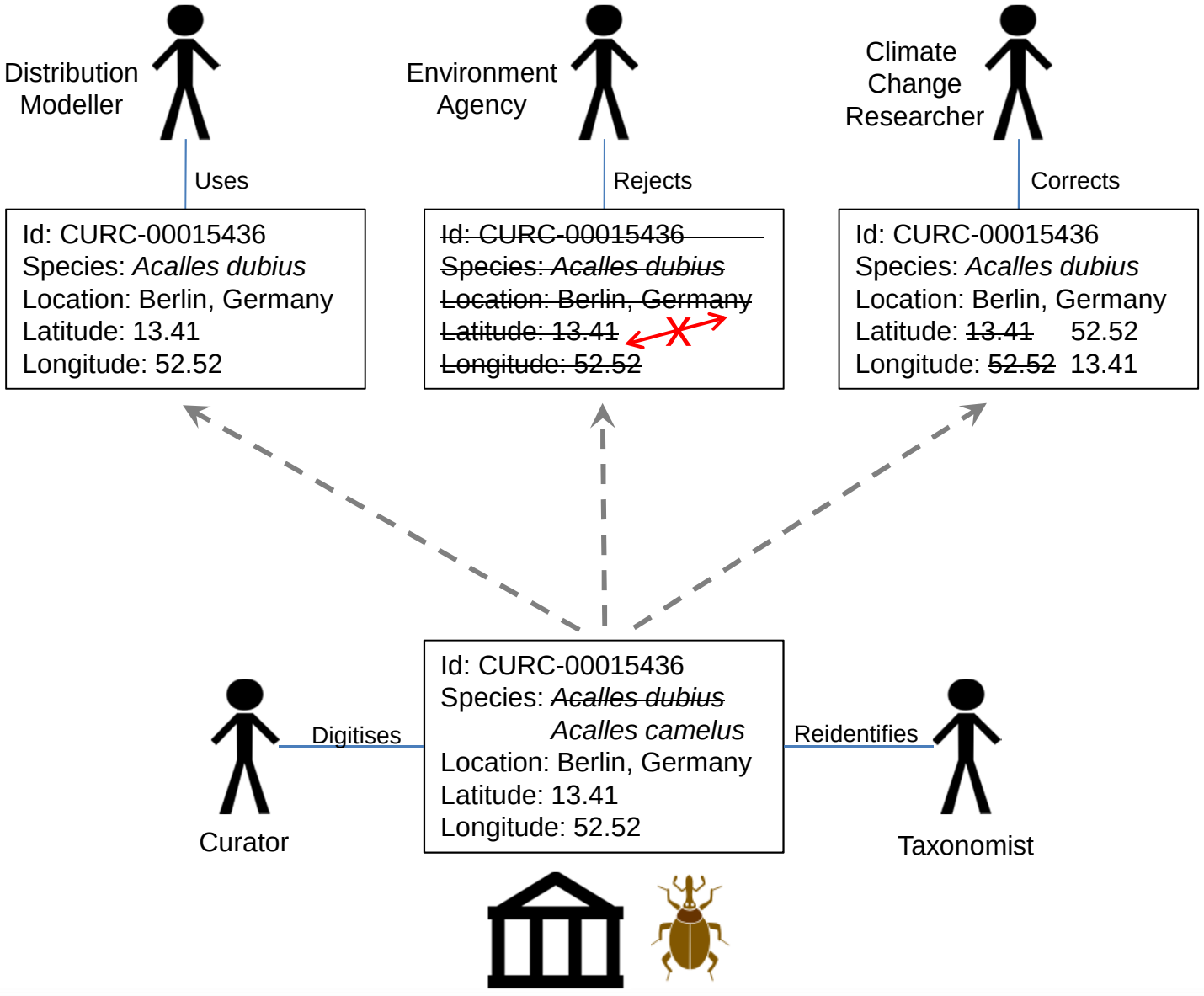
Curator

Digitises

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

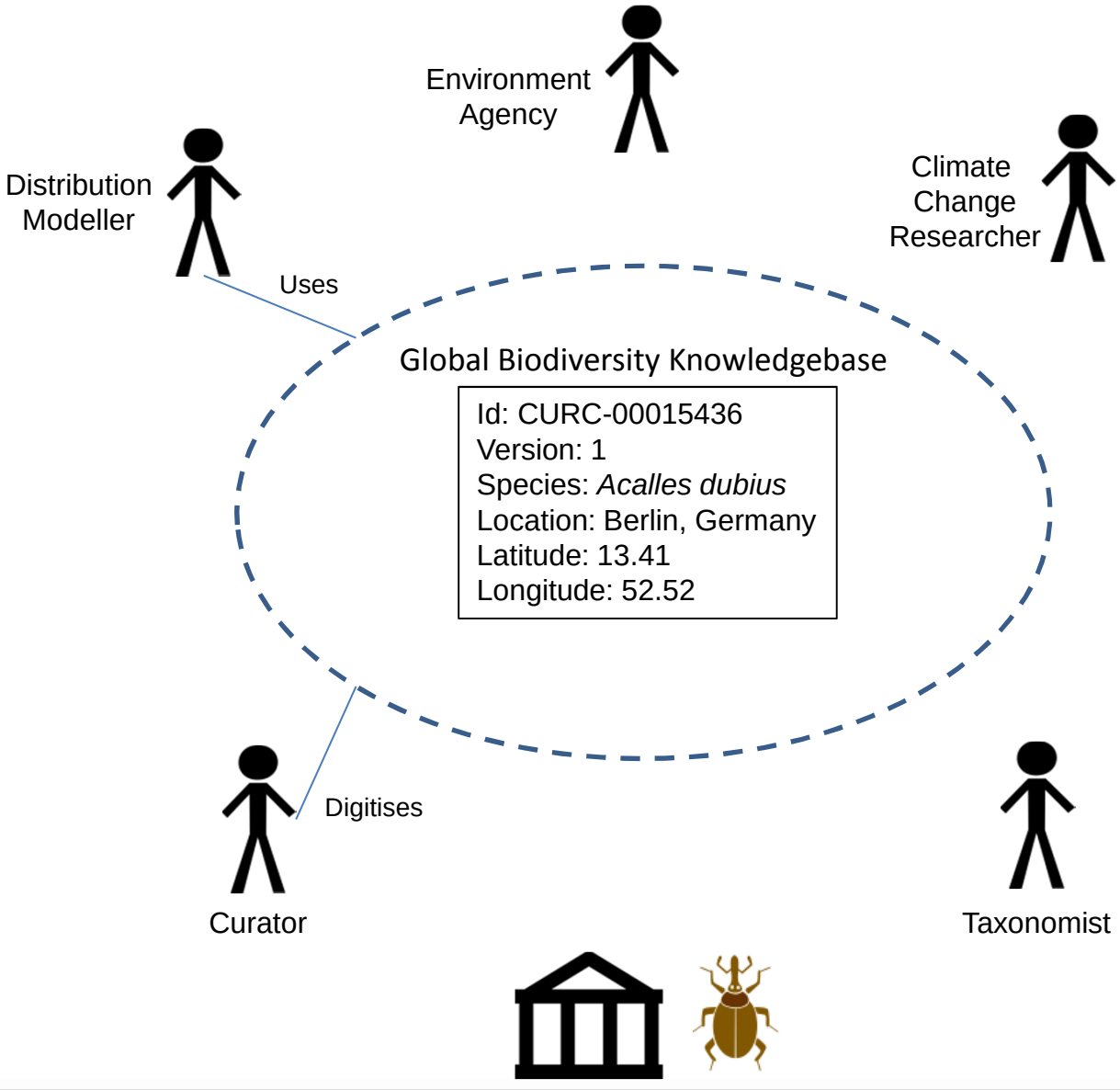
GBIF
www.gbif.org

Weaknesses with distributed data



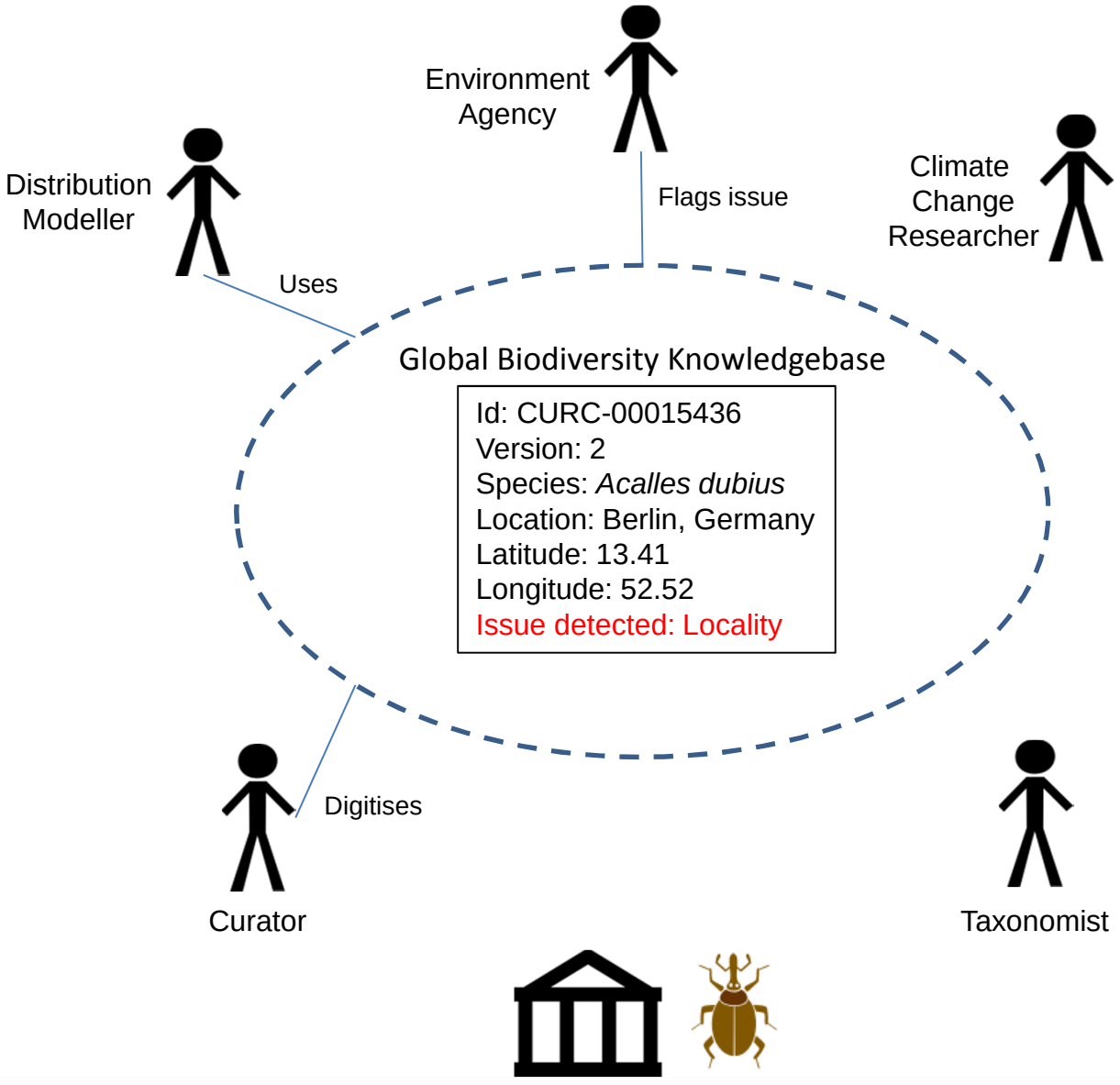
GBIF
www.gbif.org

Toward a shared knowledgebase



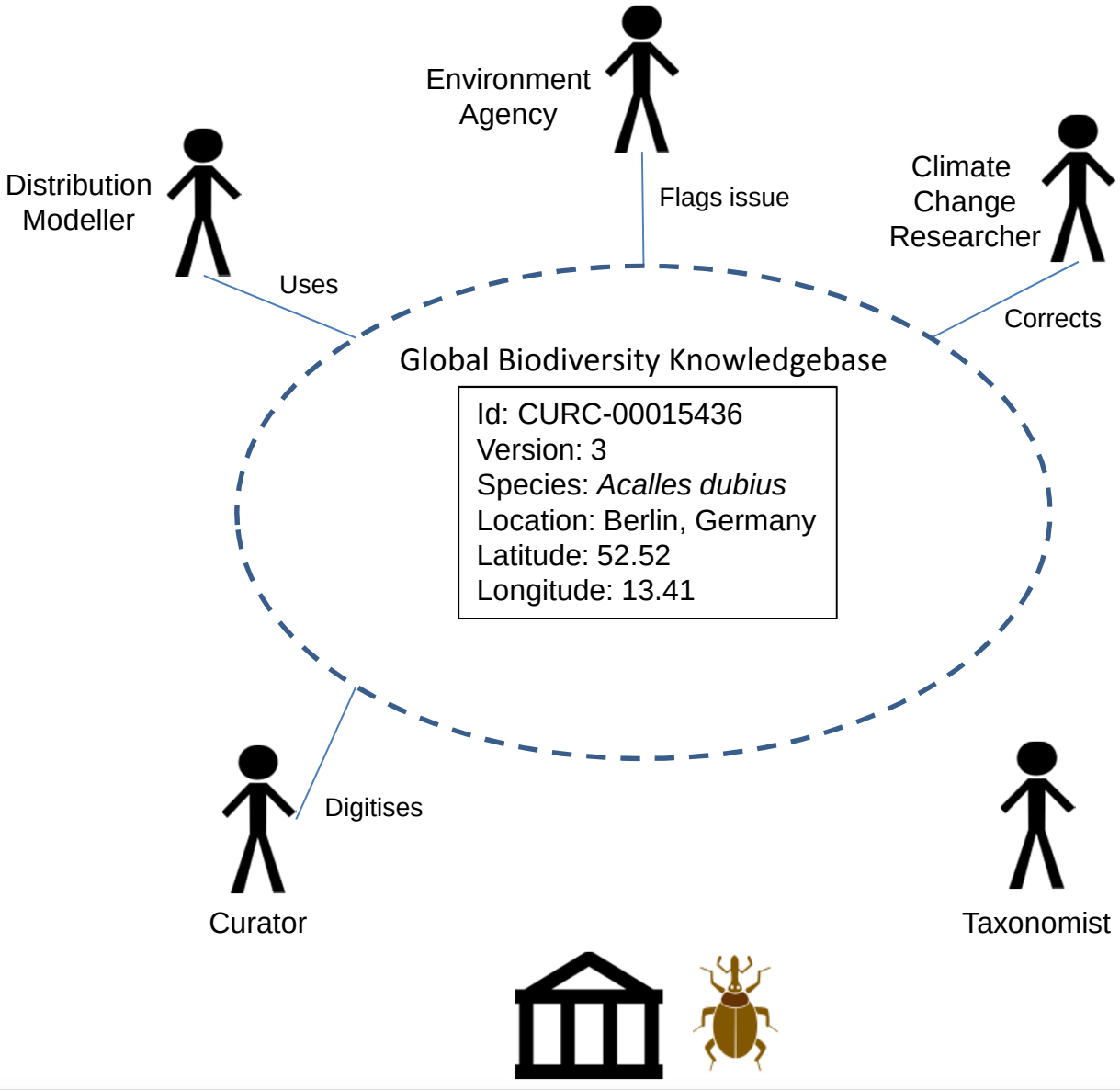
GBIF
www.gbif.org

Toward a shared knowledgebase



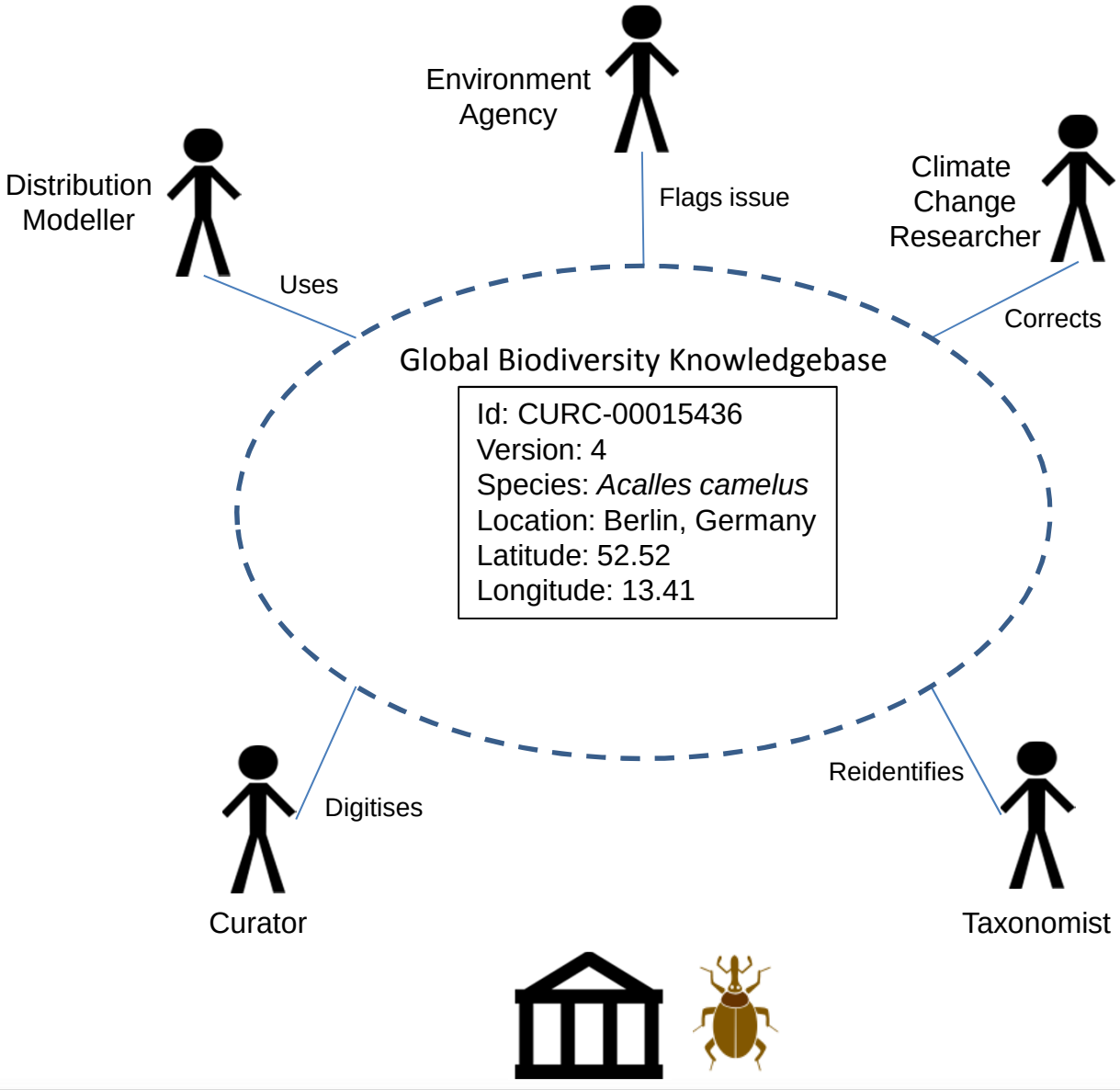
GBIF
www.gbif.org

Toward a shared knowledgebase



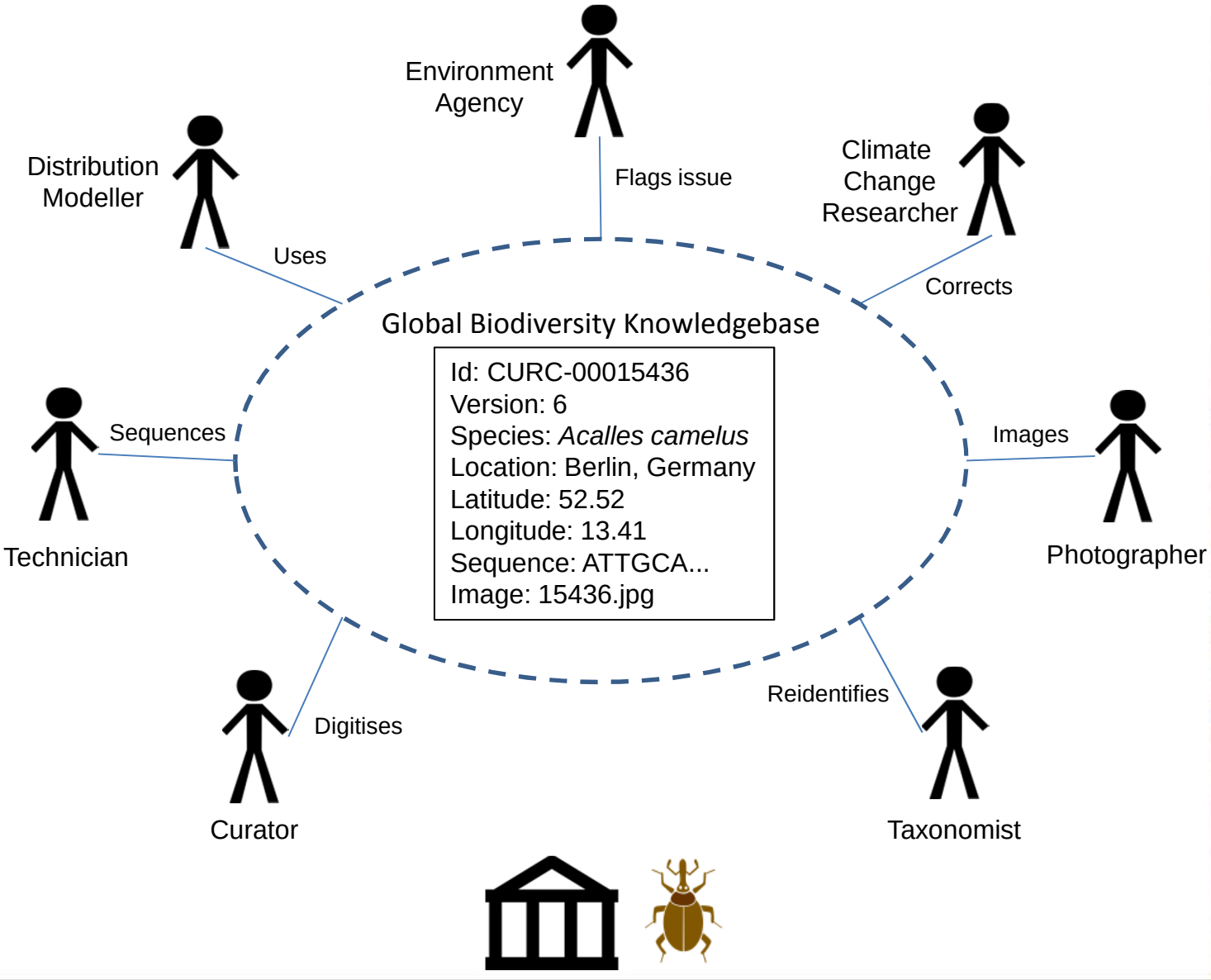
GBIF
www.gbif.org

Toward a shared knowledgebase



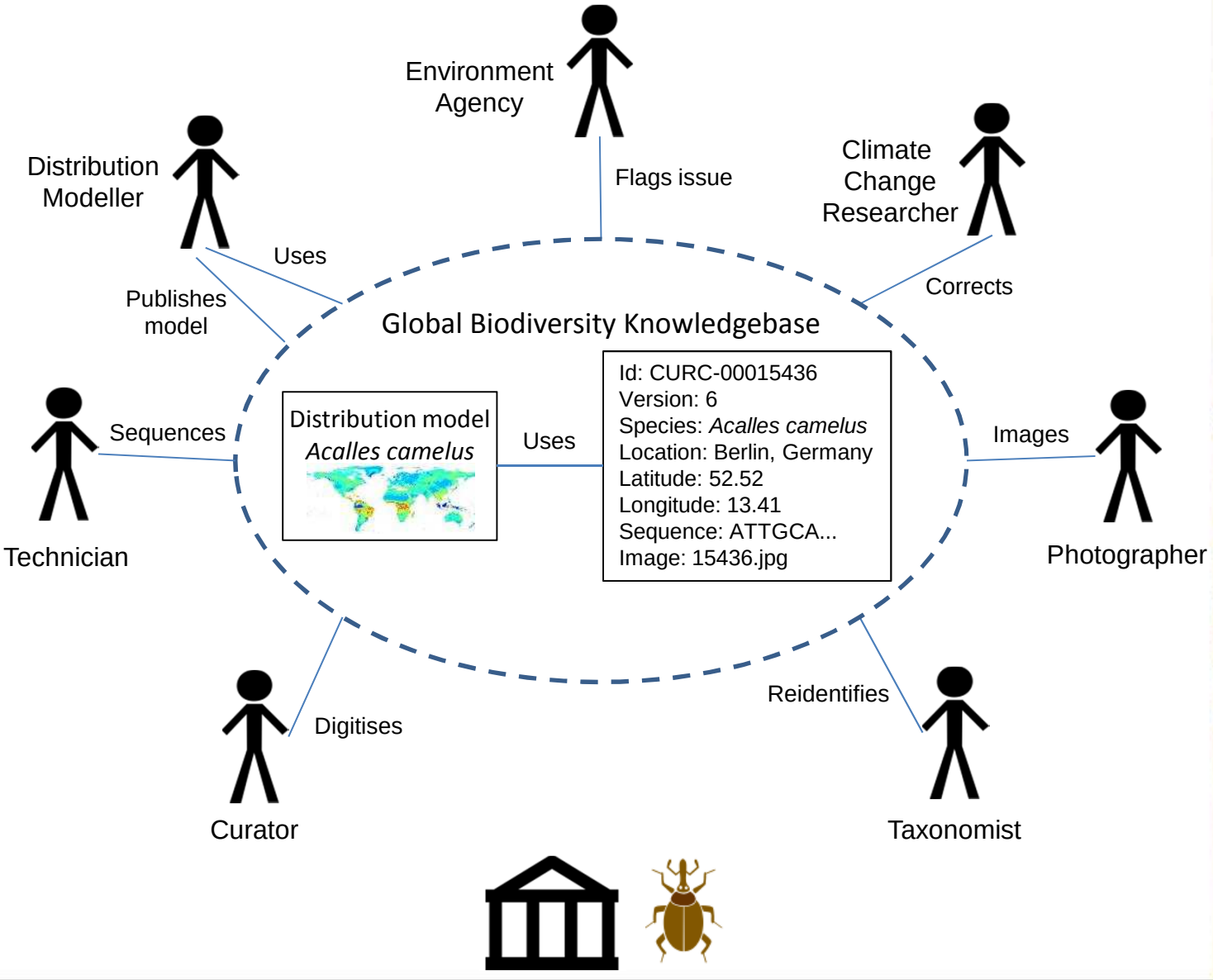
GBIF
www.gbif.org

Toward a shared knowledgebase



GBIF
www.gbif.org

Toward a shared knowledgebase

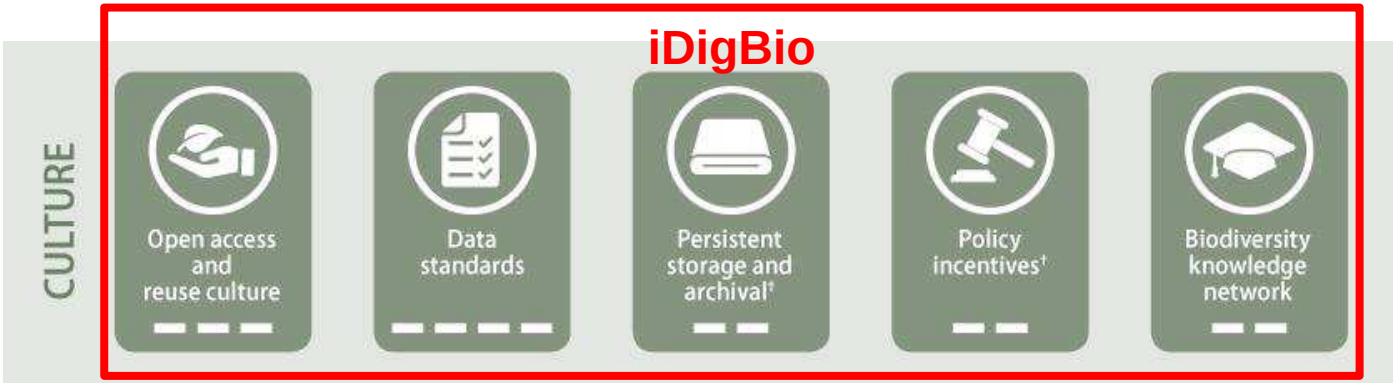


GBIF
www.gbif.org

Global Biodiversity Informatics Outlook

GBIF
www.gbif.org

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

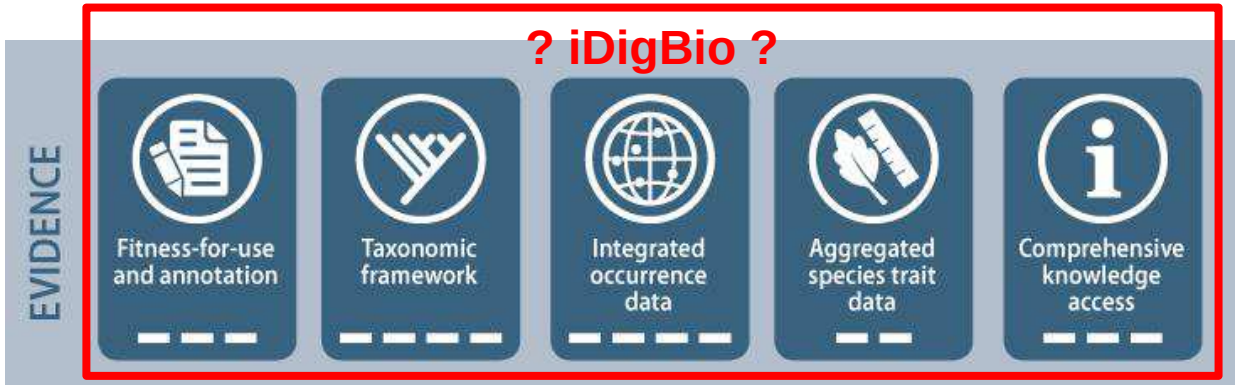
GBIF
www.gbif.org

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

GBIF
www.gbif.org

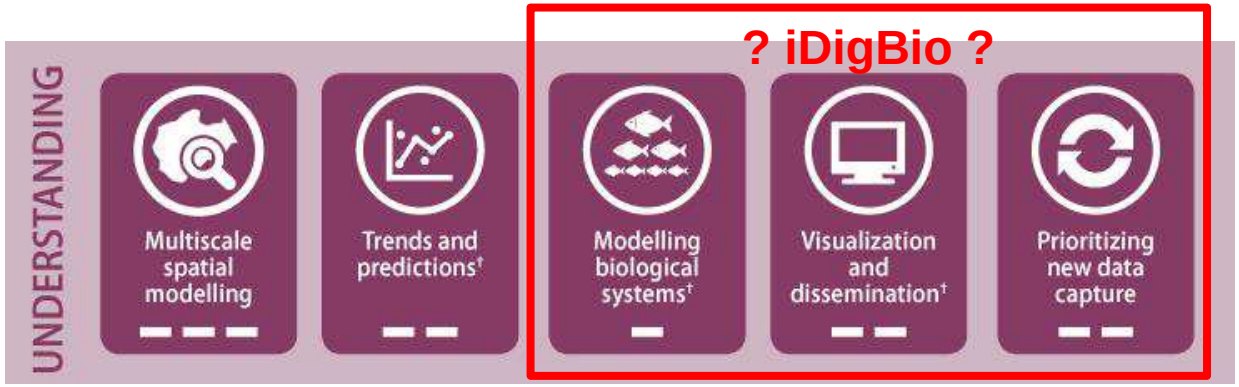
Focus Area: Evidence



- 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

GBIF
www.gbif.org

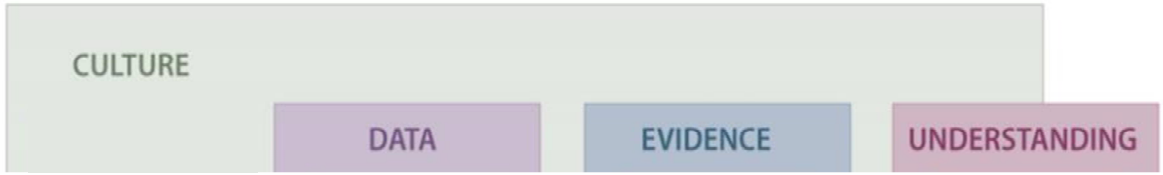
Focus Area: Understanding



- 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

GBIF
www.gbif.org

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

GBIF
www.gbif.org

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: • Identification history • Collection identifiers • Locality • Images, sequences • Morphometric data	Primary data elements: • Scientific name • Coordinates • Date • Confidence/evidence level
Critical linkages: • Nomenclators • Biodiversity Heritage Library • Barcode of Life Database • Phylogeny data	Critical linkages: • Climate and environment • Politics and land use • Stable classification

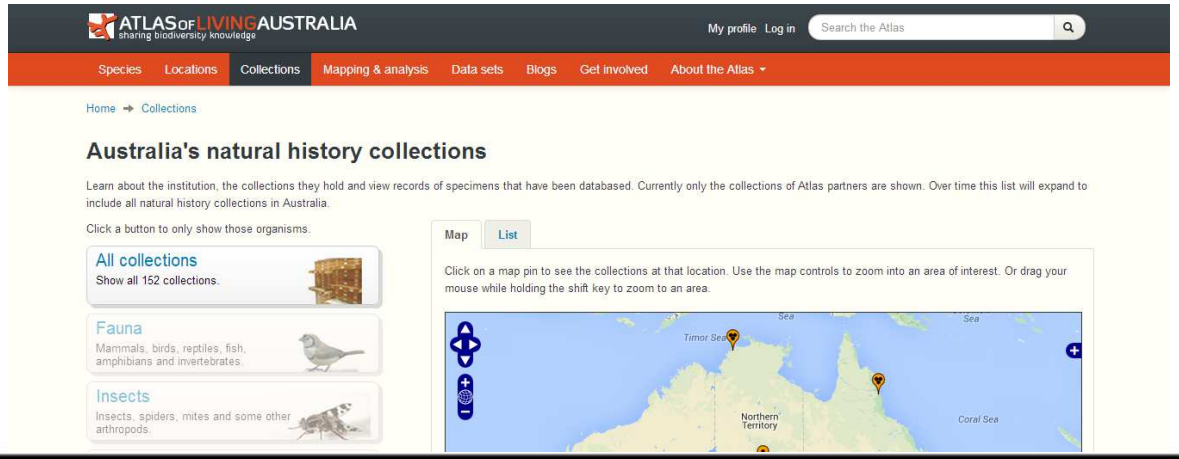
GBIF
www.gbif.org

Two major applications

Planning digitization

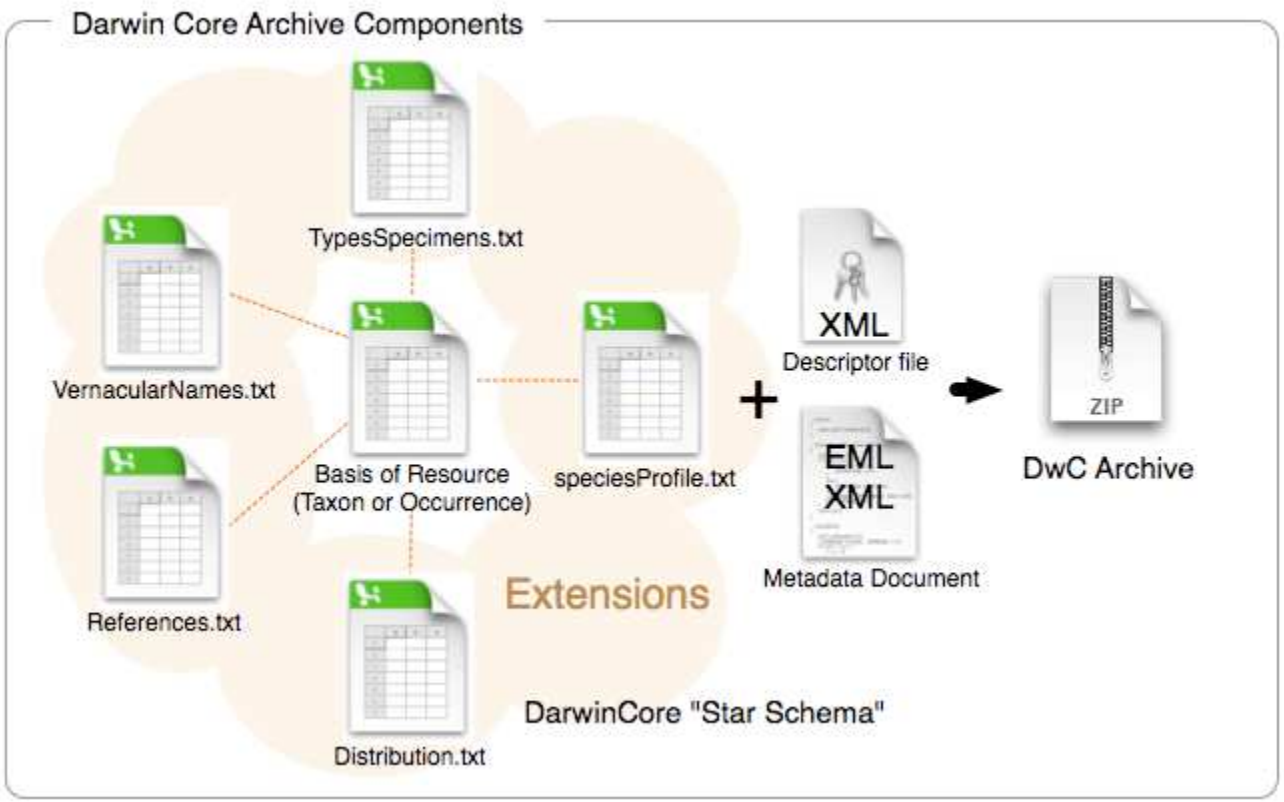
- Understand content of collections
  - Collection type: wet, dry, pinned, tissue, living, ...
  - Factual: size, taxa, geography, types, curation status, subcollections, ...
  - Relevance: time series, protected areas, threatened species, indicator taxa, ...
  - Planning: digitisation options, costs/benefits

→ Metadata for discovery and publicity
→ Inputs to institutional/national strategy
→ Opportunity for collaborative funding or labour



GBIF
www.gbif.org

Publishing data

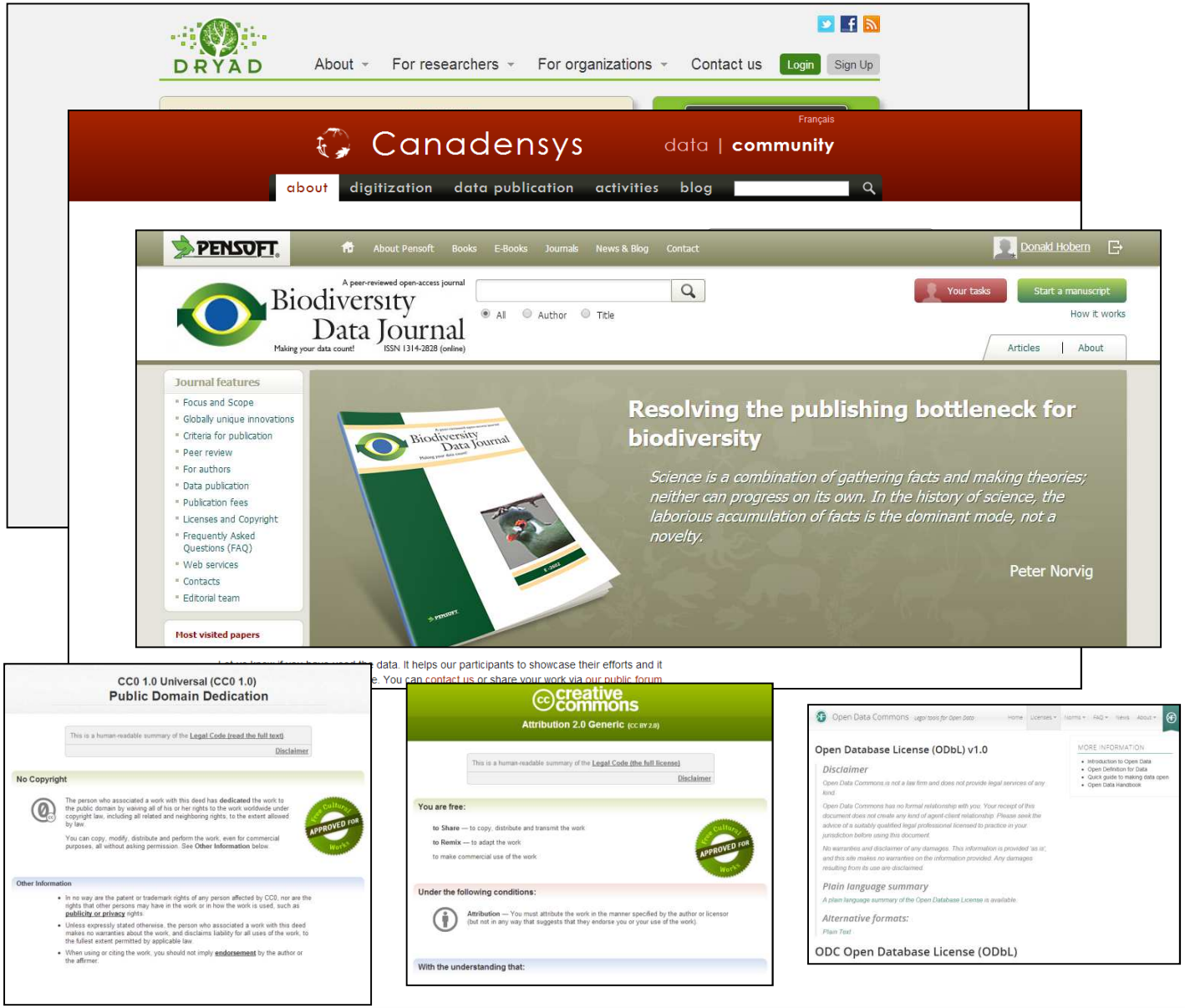


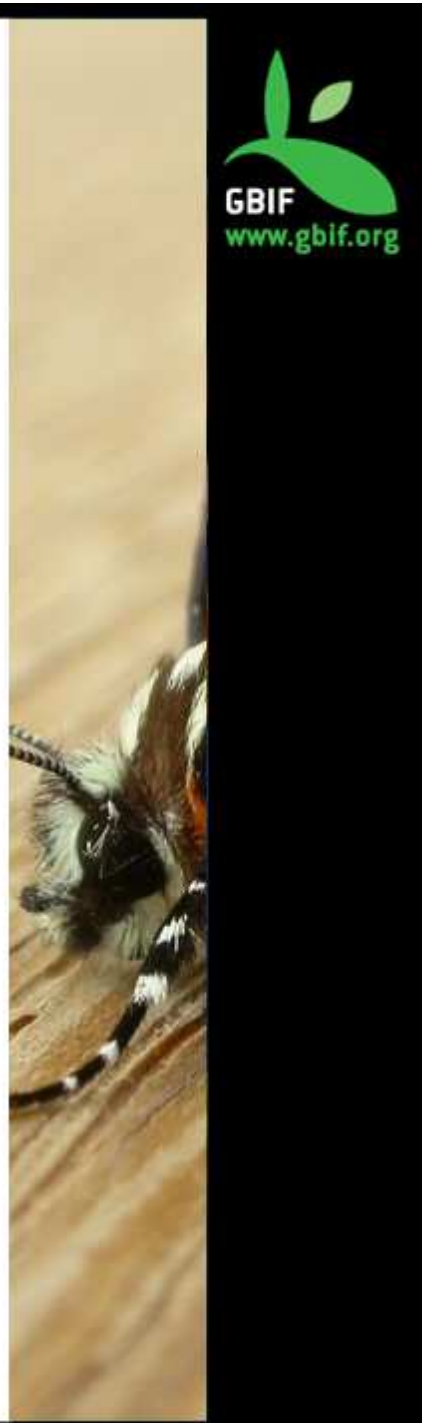
Darwin Core Archive

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

GBIF
www.gbif.org

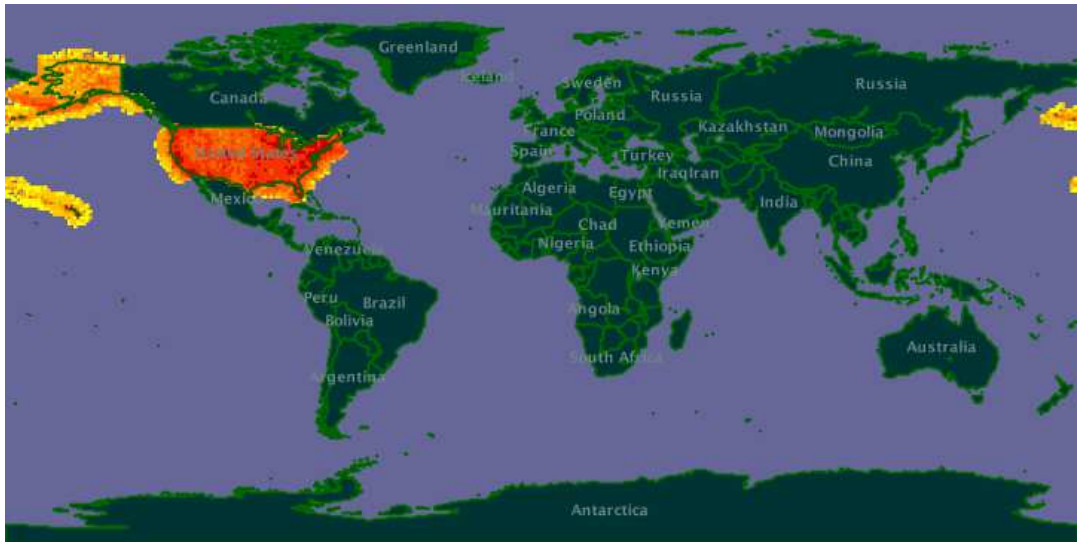
Supporting data reuse



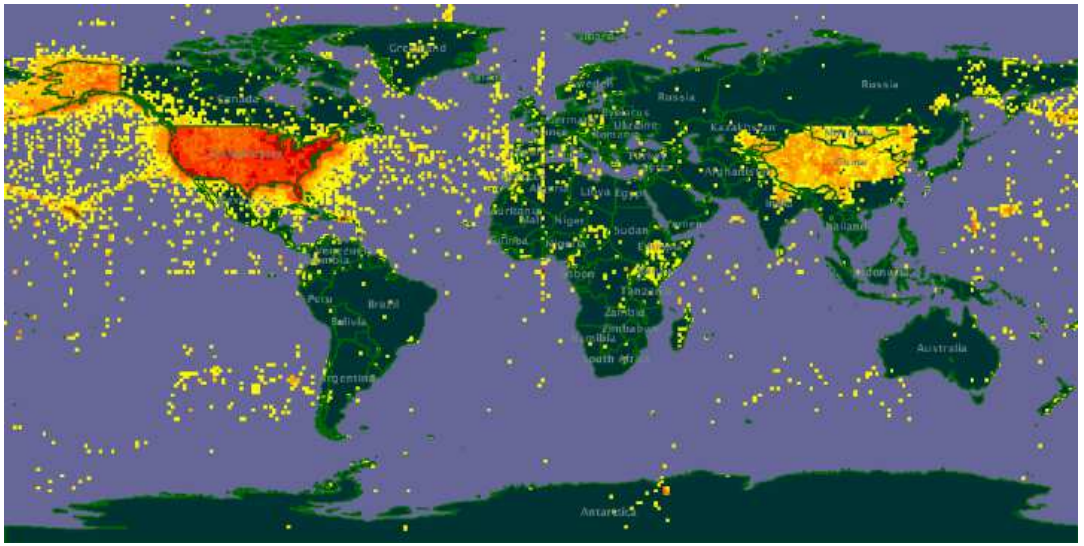


GBIF
www.gbif.org

Integrating data



GBIF data for “United States” intersecting USA



All GBIF data for “United States”

GBIF
www.gbif.org

Collaborative curation

Great Crested Newt (Denmark)

next »

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.

iSpot

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**

GBIF
www.gbif.org

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*.

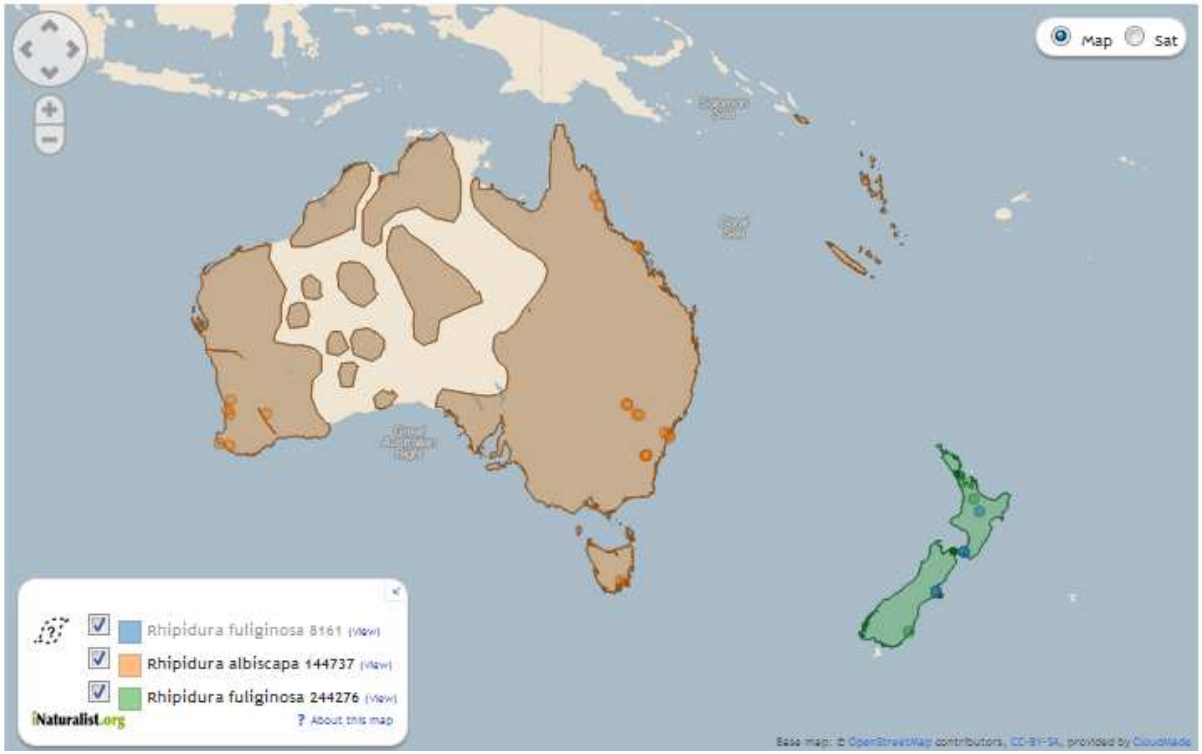
Source: Christidis, Les; Boles, Walter (2008)... (Citation)
Added by loarie on 2013-04-08 | Committed by loarie on 2013-04-10

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



GBIF
www.gbif.org

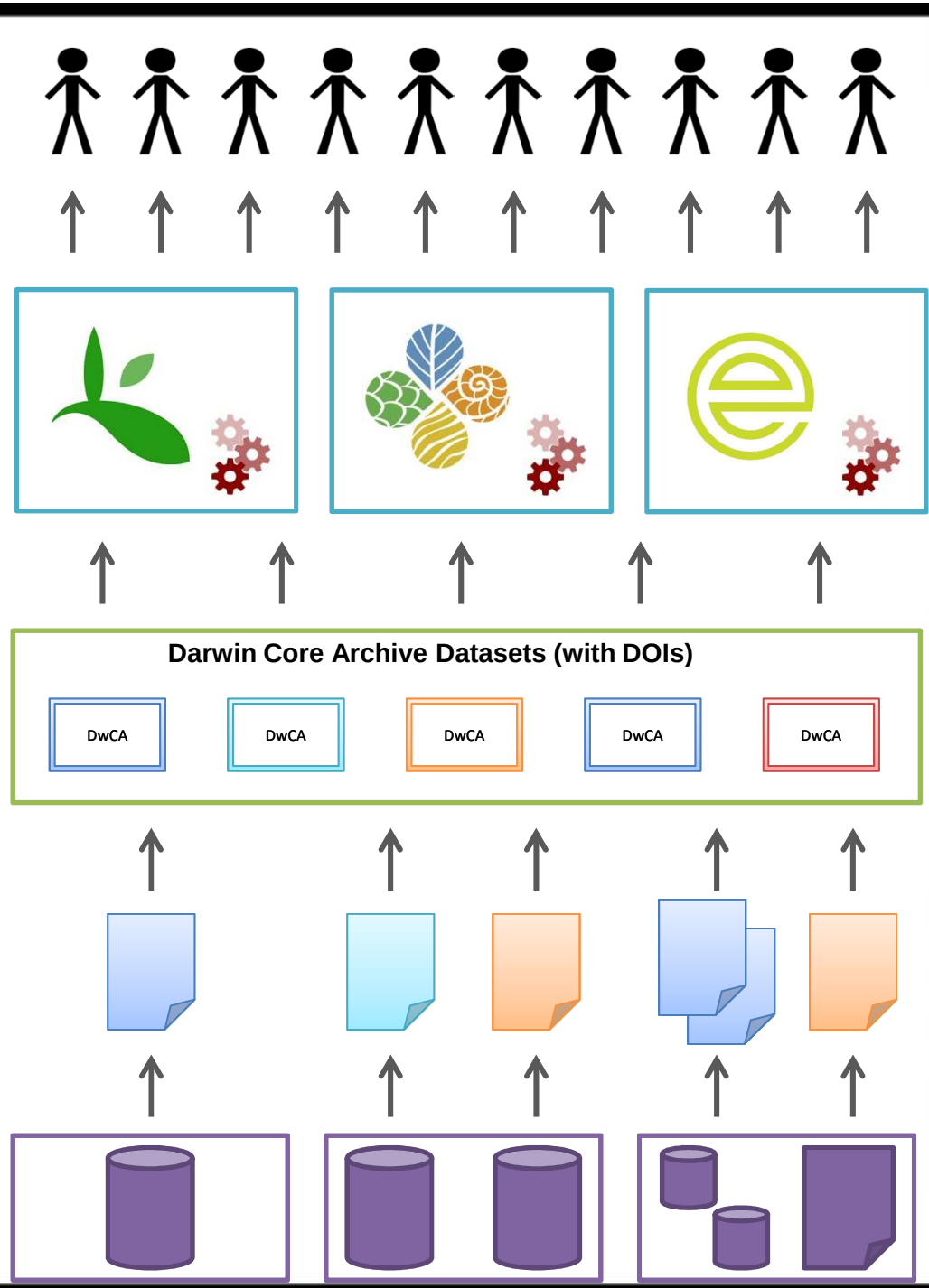
Users

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"

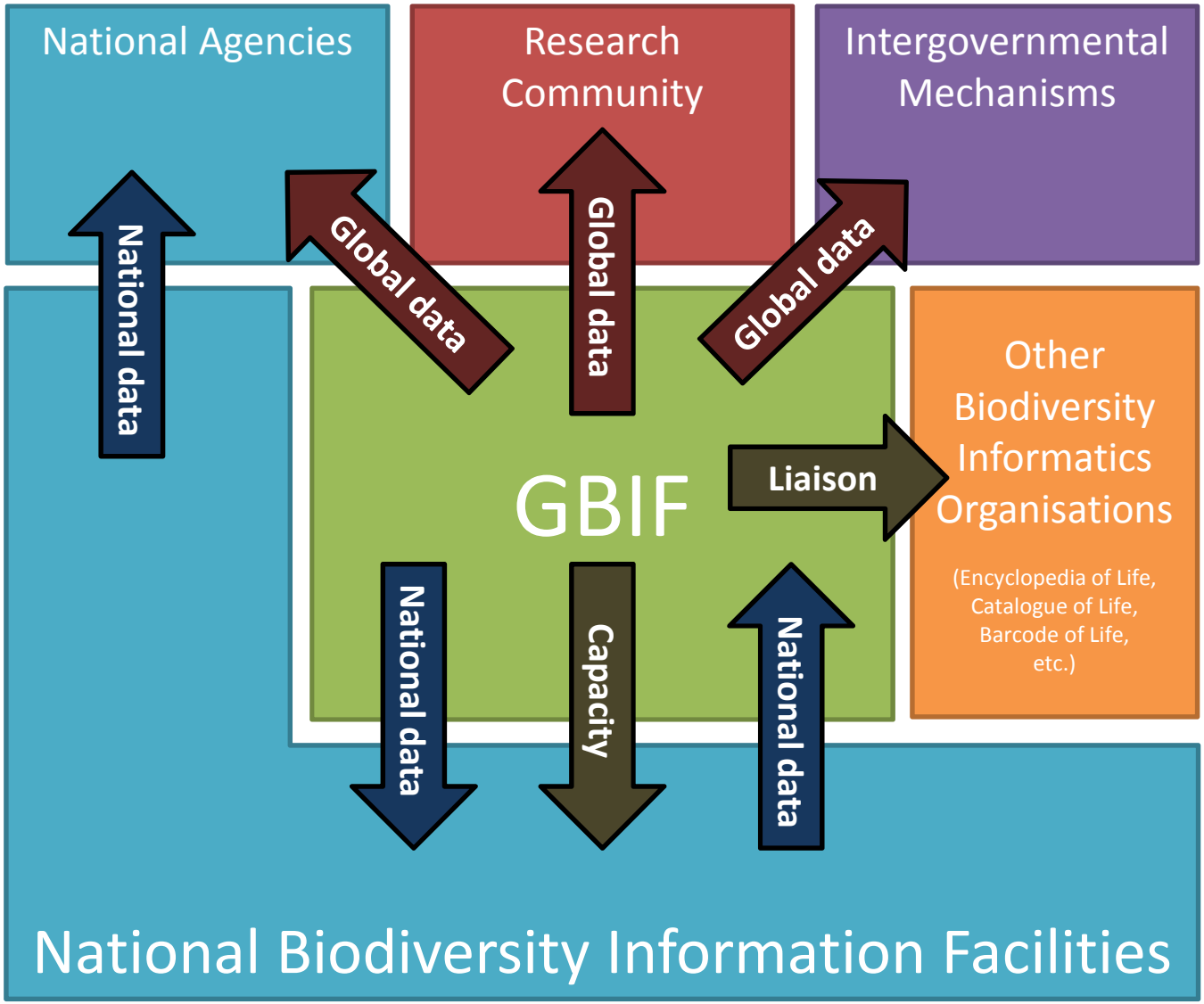
Spreadsheet Data
(DwC-A / CSV / Excel / XML)
Data and metadata for free and open use
Standards compliance

Data Publishers
Biodiversity data in institutional databases
Stable record identifiers



GBIF
www.gbif.org

GBIF in the landscape



GBIF
www.gbif.org

Thank you

Donald Hobern
GBIF Director
Global Biodiversity Information Facility (GBIF)

Biological Collections Digitization in the Pacific Workshop

Honolulu, 25 March 2014