

Biodiversity Informatics 101

Setting the Stage

James Macklin

Agriculture and Agri-Food Canada

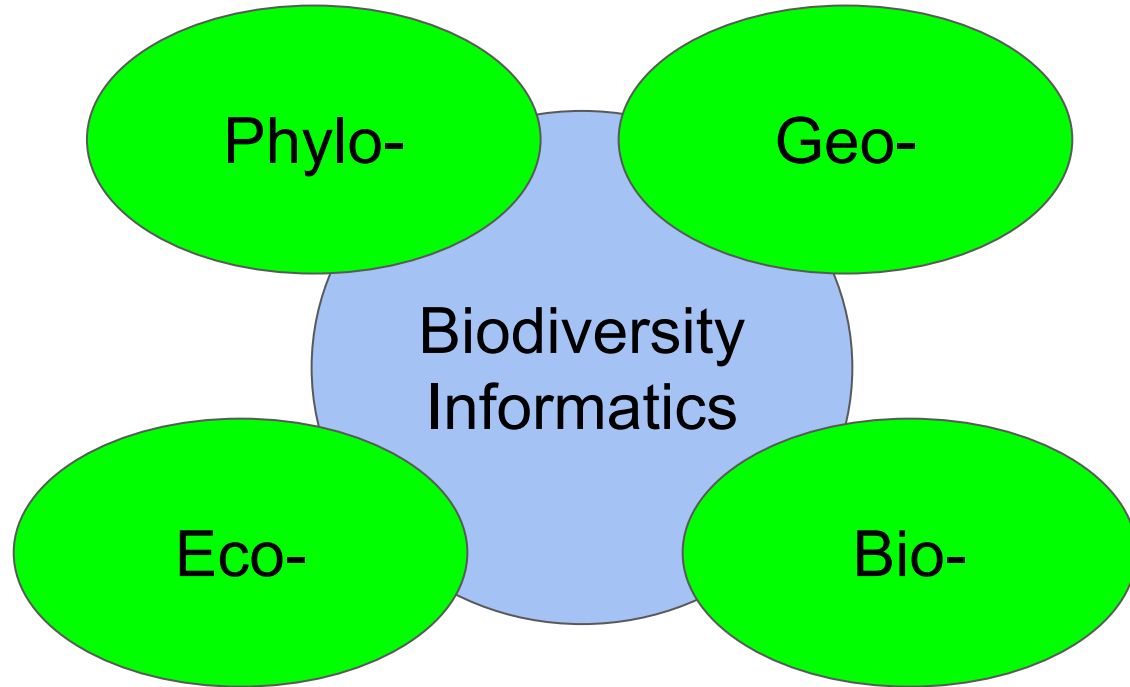
Outline

What is Biodiversity Informatics?

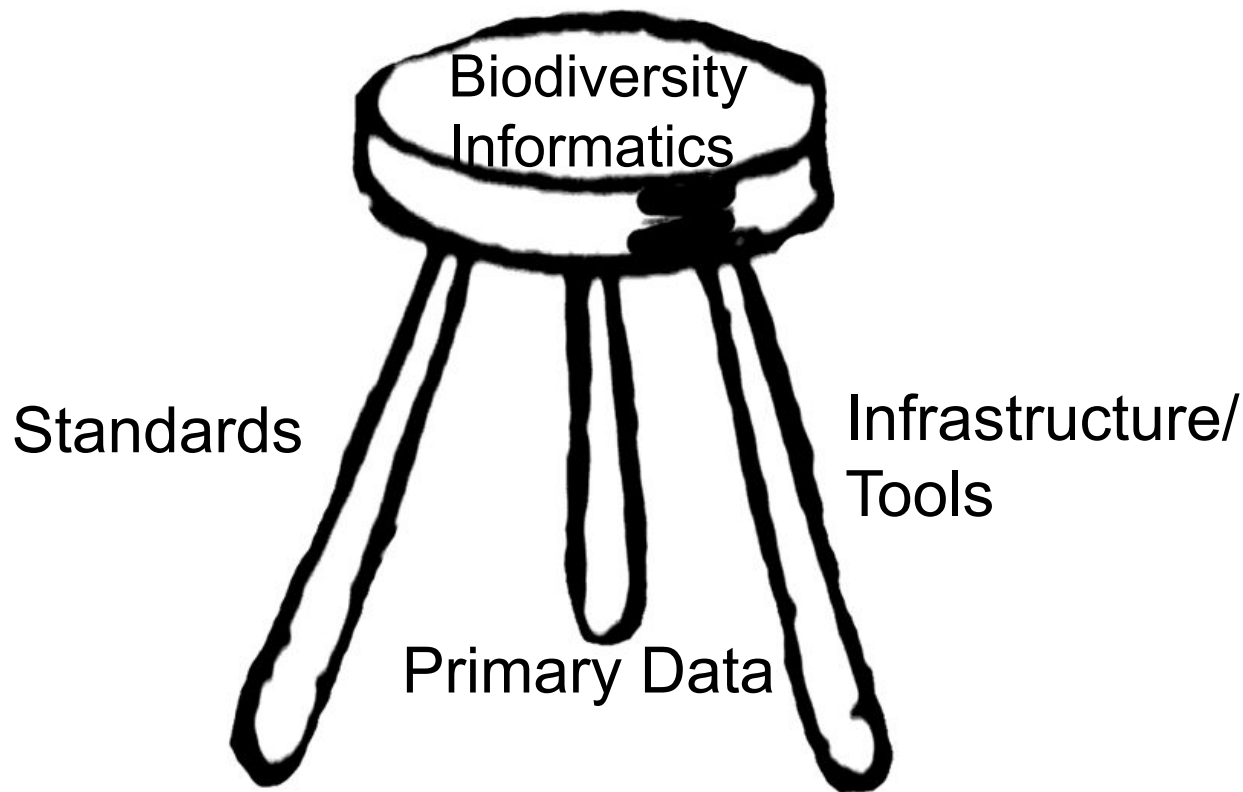
Role of data standards.

What are some of the challenges?

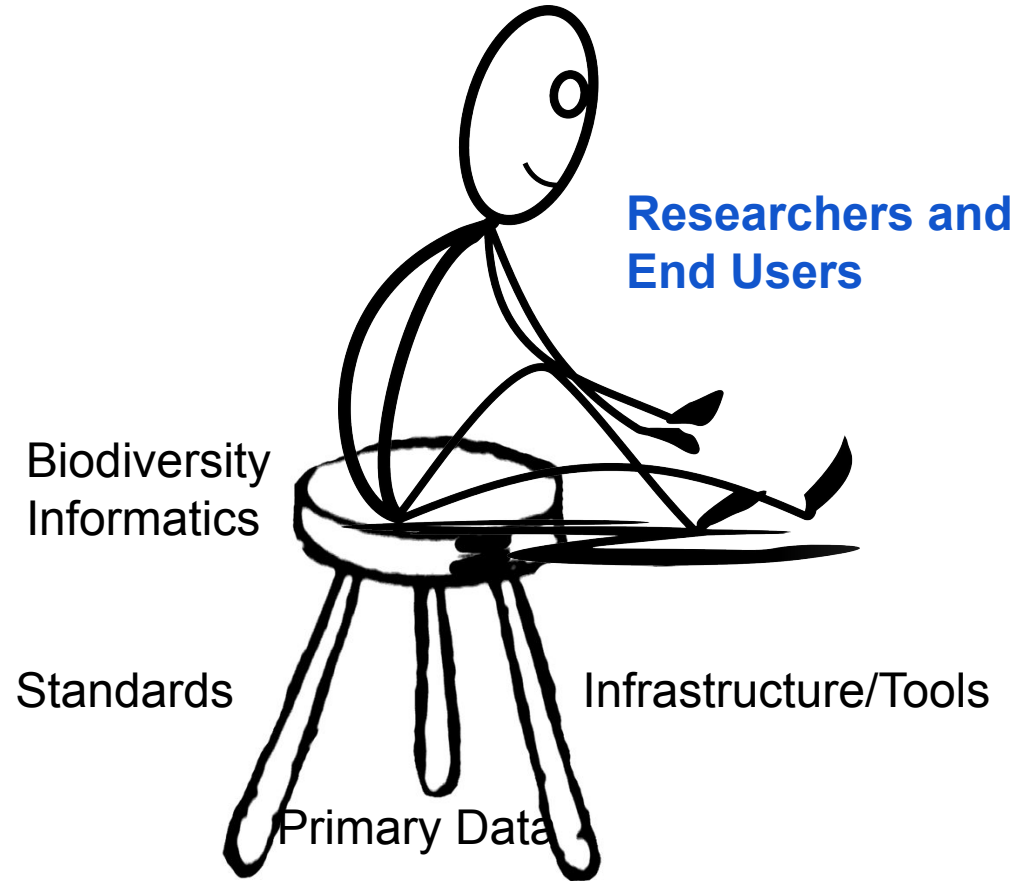
Biological Informatics



Biodiversity Informatics



Biodiversity Informatics



Scope of Biodiversity Informatics

- Data Acquisition
 - Data capture tools (text, images, sounds, videos, etc.);
- Data Management/Curation
 - Collection Management Software; multimedia/document repositories; quality control tools;
- Data Discovery
 - Websites/Portals; Globally Unique Identifiers (GUIDS); Ontologies;
- Data Analysis
 - Morphometrics; distribution modelling; molecular identification/function.

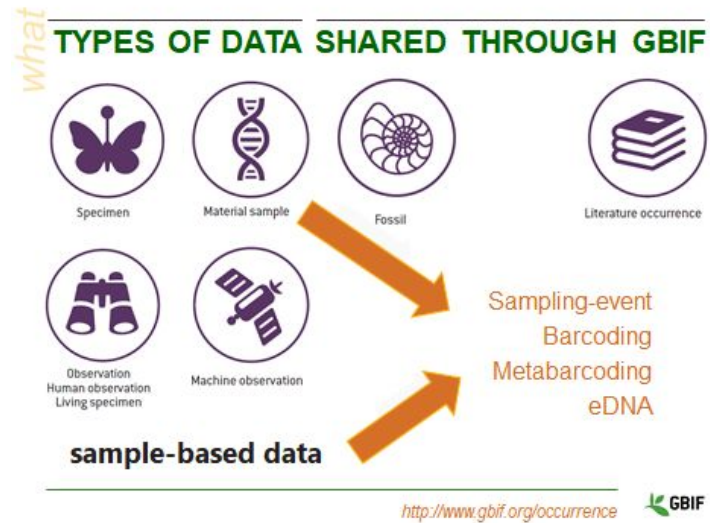
How do I become a Biodiversity Informatician?

There is no direct path...

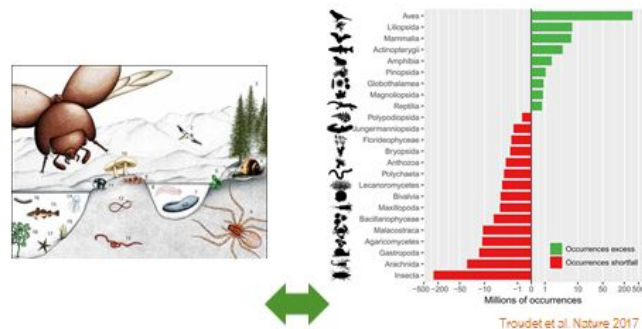
- Biologist who likes to script/code;
- Information/Library Scientist who likes biological “big data”;
- Computer Scientist who likes developing code to address biological challenges;
- Other...

Observation

- Traditionally, the “Observation” has driven our work
 - Who, What, Where, When, How?
 - Closely tied to the Natural Science/History Collections
- However, technological advances have opened the door to new forms of observation:
 - Molecular-based: DNA barcoding; e-DNA; metagenomics;
 - Sensor-based: remote sensing; biologging.



Global biodiversity vs. digitally available data

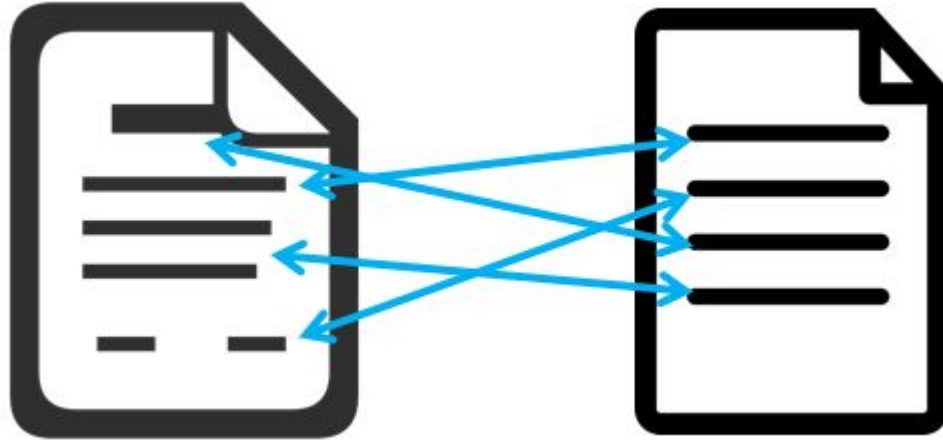


Primary use case...



Clients and Services need to combine data from different sources to use it.

Primary Use Case



Mapping documents is difficult.
Combining data is difficult.
Primary use case is difficult!

Why Standards?

- Communication between humans and now machines rely on standards! (e.g., languages)
- Terminology essential (e.g., dictionaries, thesauri)
- Data integration: Connecting data resources (silos) to aid discovery

BUT: How do you know what an apple is and that my “apple” is your “apple”?

- Data Sharing/Exchange (Aggregators, DwC Archives)
- Data Quality: Fitness for use/purpose

Standards



Lots of standard bodies world-wide!

- Aggregator: International Organization for Standardization (ISO)
 - (e.g., Dates, Time, Country Codes, Language Codes...)
- World Wide Web Consortium (W3C)
 - (e.g., HTML, XML, RDF, CSS...)
- Library-based Standards:
 - American Library Association (ALA); National Information Standards Organization...
- Biodiversity-based Standards:
 - Biodiversity Information Standards (TDWG)
 - Genomic Standards Consortium (GSC)
 - Open Geospatial Consortium (OGC)

Biodiversity Information Standards Mission

- To jointly develop global open standards, guidelines and unified protocols for the recording, exchange and re-use of natural history and biodiversity data;
- To promote the use of our products to facilitate biodiversity informatics and to underpin work towards making data FAIR (Findable, Accessible, Interoperable, and Re-usable);
- To support the natural history and biodiversity data community by exchanging knowledge, experience and ideas for standardisation through our annual conference, publications and by acting as a platform for discussions.

Challenges

Taxonomy

Linked Data

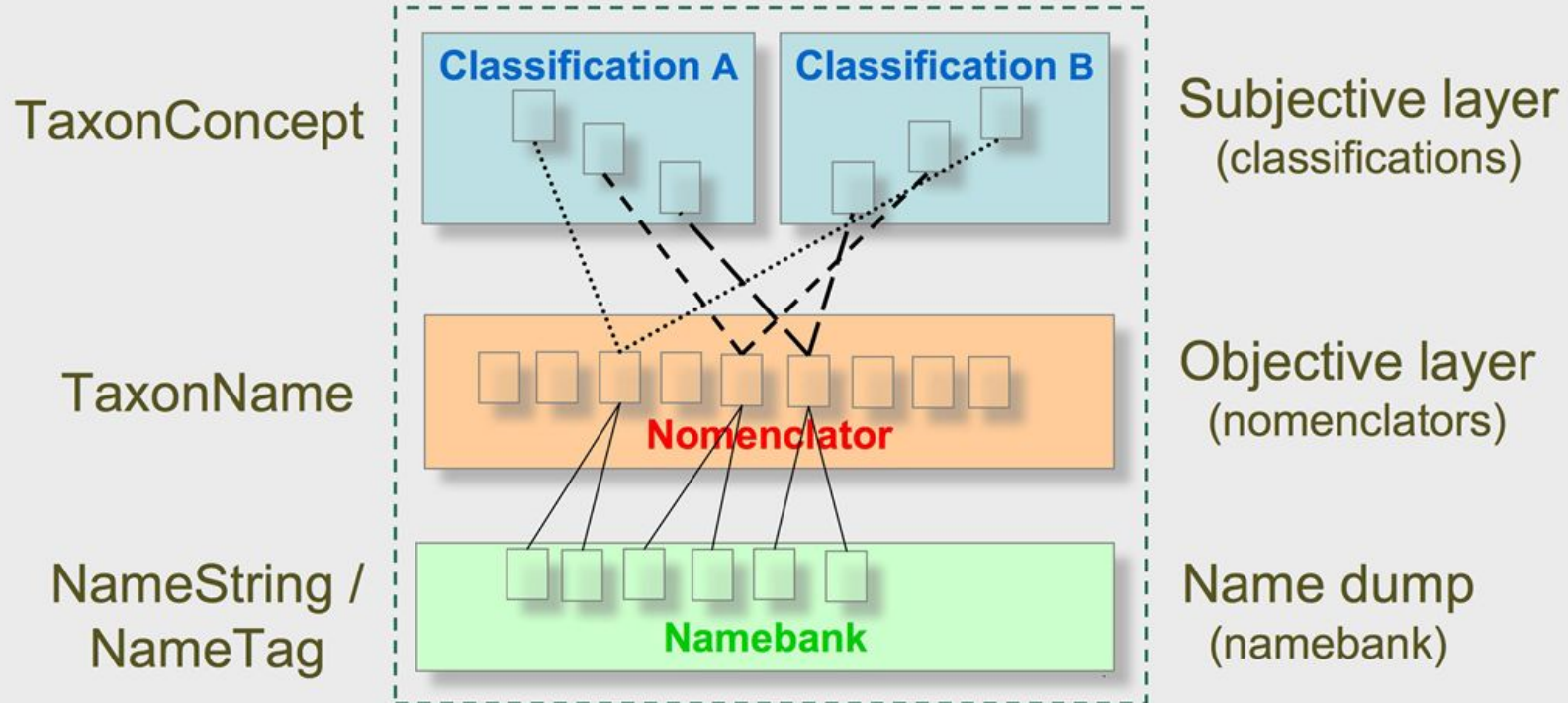
Globally Distributed Systems

Social

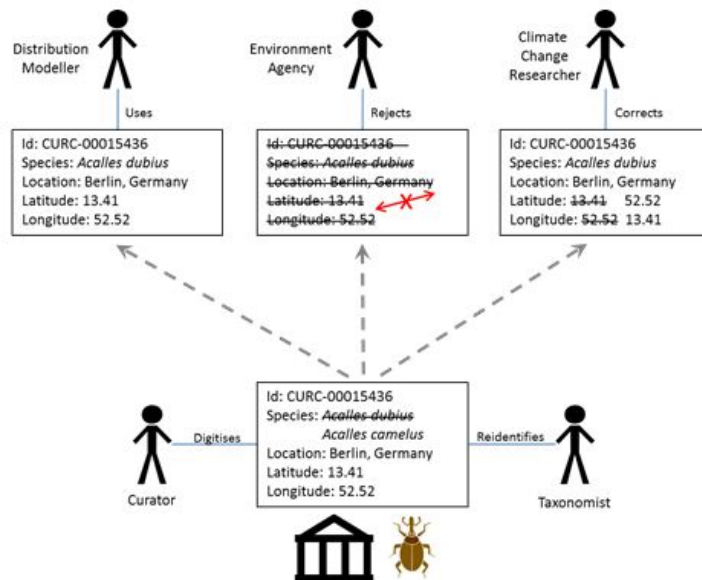
Capacity Building

The Global Names Architecture (GNA)

<http://globalnames.org/>

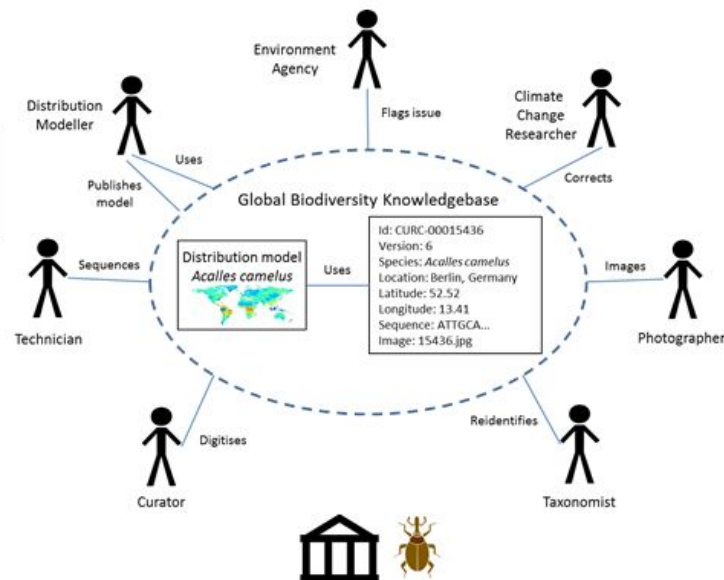


Linked Data



Today

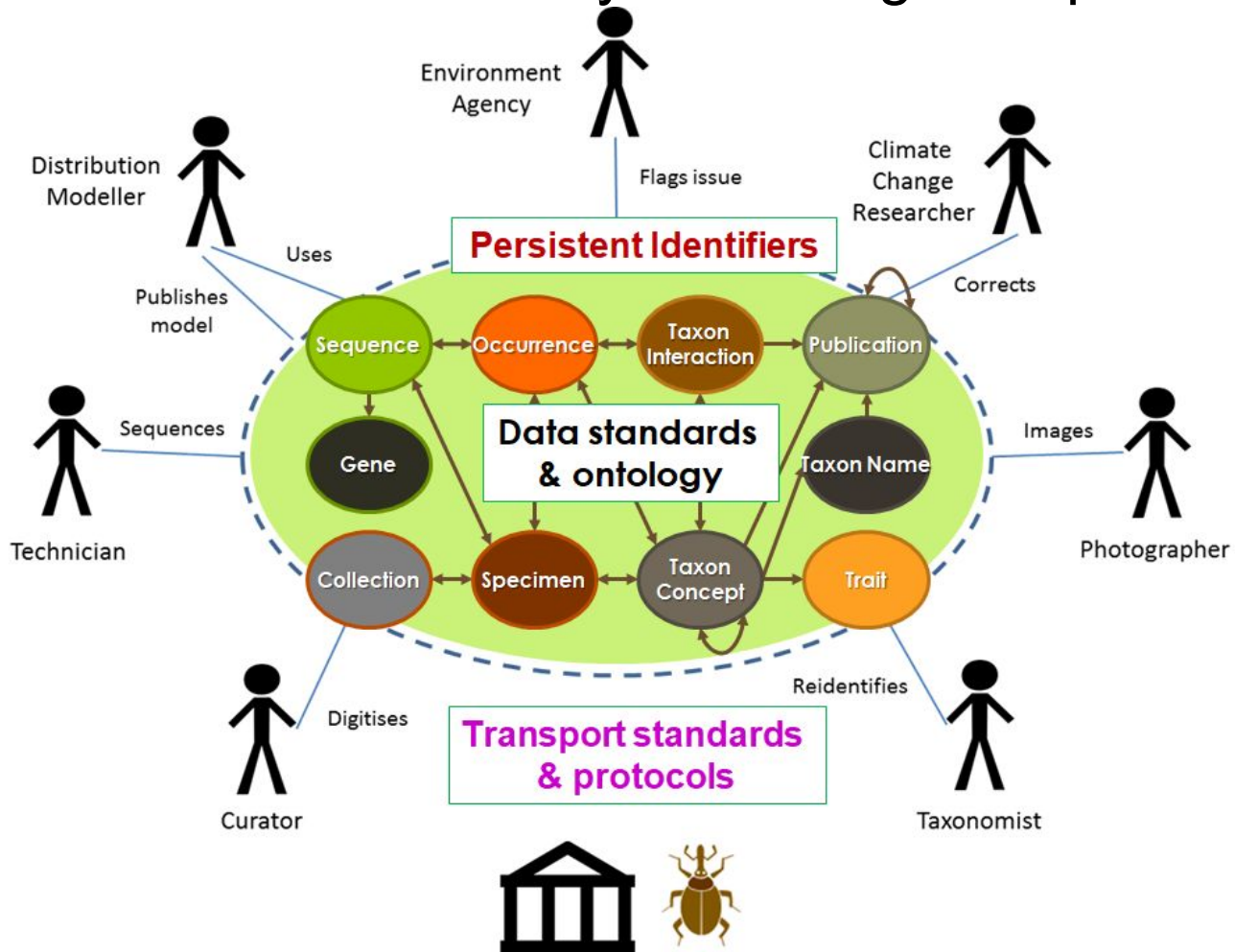
- Disconnected publishing and use
- Limited feedback
- Poor ability to reference data



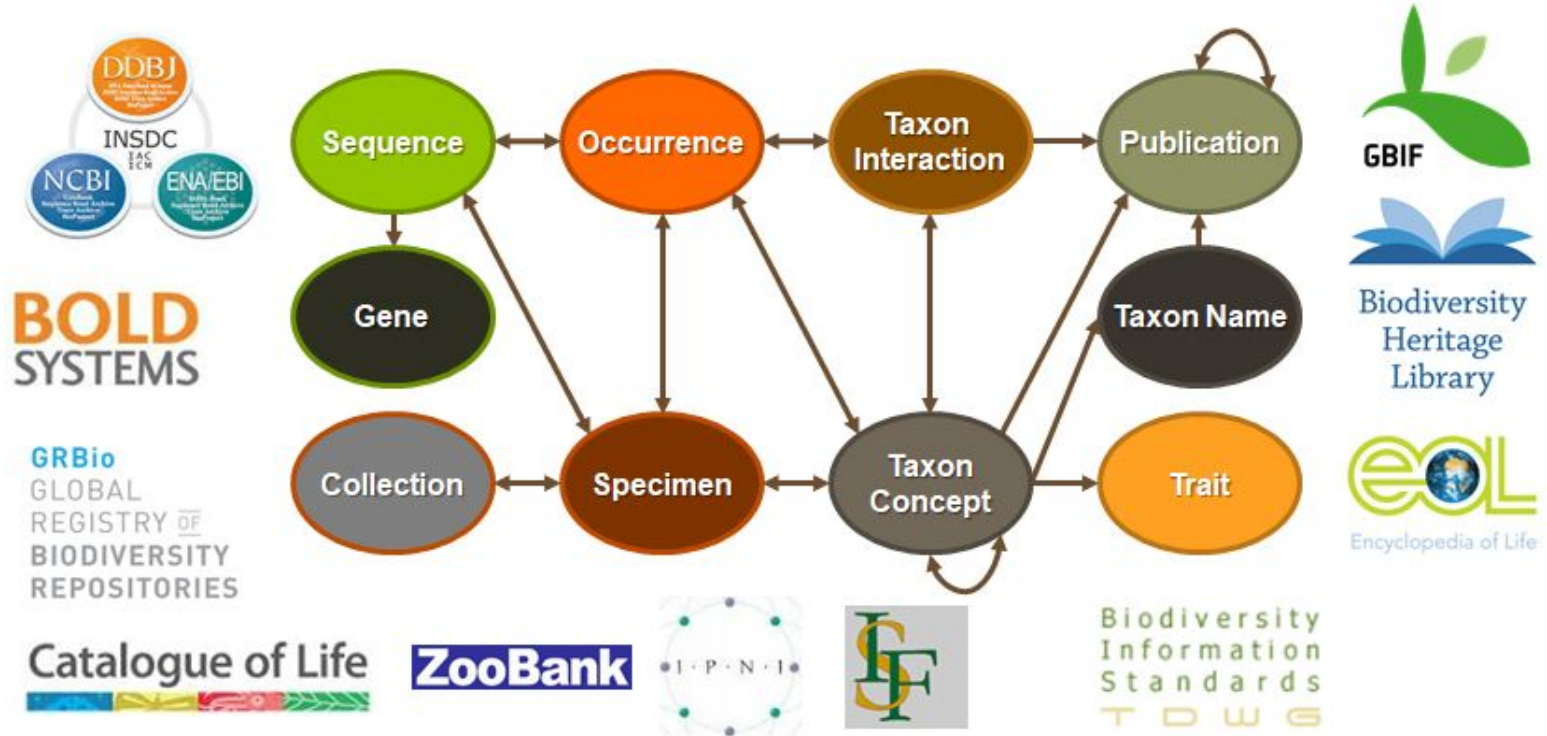
Future?

- Stable, connected data from all sources
- Collaborative management
- Continuous step-wise improvement

Linked Data & the Biodiversity Knowledge Graph

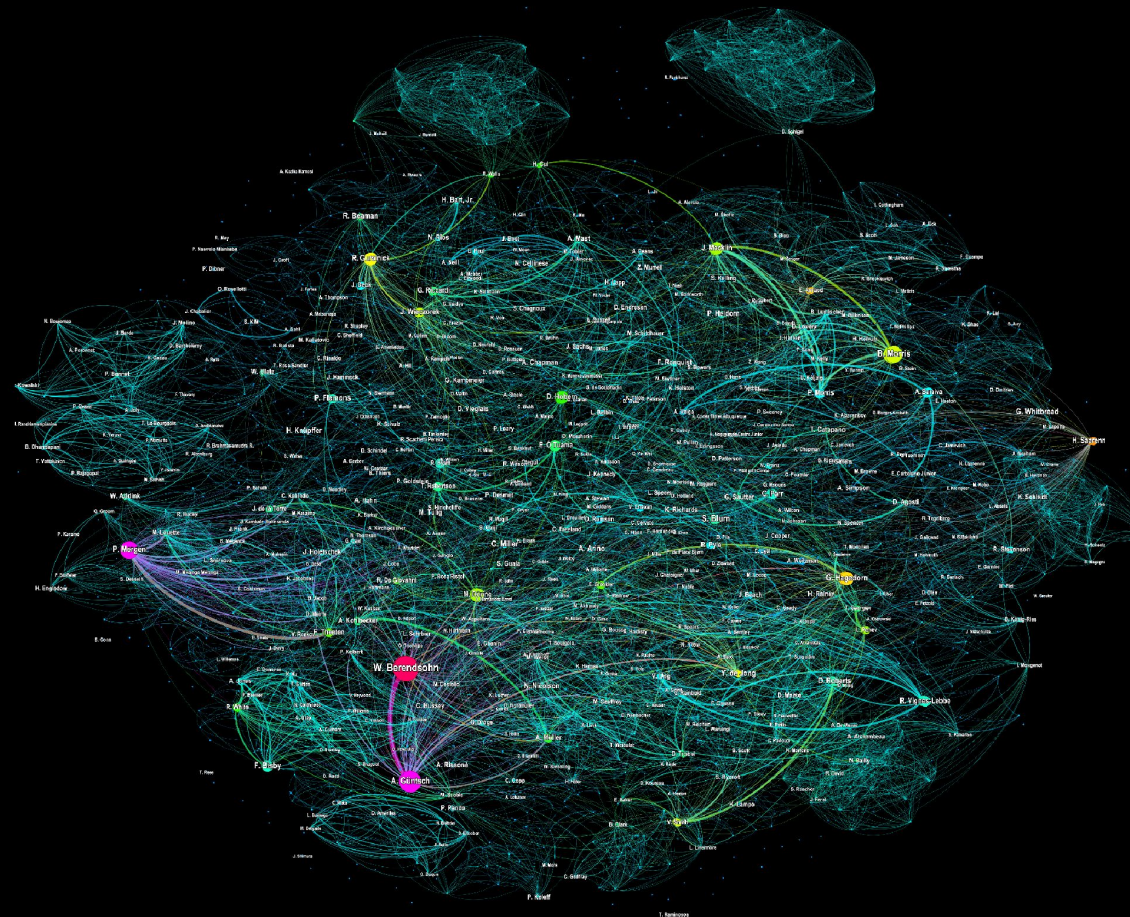


Globally Distributed Systems



Social

- Openly sharing data (no more silos!)
 - FAIR (Findable, Accessible, Interoperable, and Re-usable)
- Professional respect
- Trust
- Engaging stakeholders and citizens about the importance/relevance of our science:
 - Era of “Fake/Exaggerated News”
 - Support evidence/science-based decision making!



Capacity Building

- We live in a globalized world...
- Training and Education: researchers, professionals, educators, citizens
 - Global network of expertise
 - Data Curators/Scientists (current gap)
- Common infrastructure: less duplication; more sustainable
- Common tools: open source; more sustainable.

An Alliance for Biodiversity Knowledge

Outcome of the Global Biodiversity Informatics Conference
in July, 2018.



“propose a coordination mechanism for developing shared roadmaps for biodiversity informatics”

“These initiatives provide models for multiple stakeholders with decentralised funding and independent governance to combine resources and develop sustainable solutions that address common needs.”

www.biodiversityinformatics.org

ASSESSMENTS AND INDICATORS

