

Evaluating data usage

Reed Beaman
University of Florida
27 March 2014

What happens with your data?

- Aggregation
- Globally unique identifiers
- Taxonomy and names
- Linkages to related resources
- Data standards and modeling data
- Attribution

Is your data making it into iDigBio?

[About iDigBio](#) | [Portal](#) | [Technical Information](#) | [Education](#)

[iDigBio Home](#) | [Portal Home](#) | **Search Records** | [Tutorial](#) | [Publishers](#) | [Feedback](#)

Search Records

Full Text Search:

only records with images [Hide Advanced Search](#)

Current Results
 Query: Family = urticaceae & Recor
 Records: 798
 Approx. Download Time: 0hrs 0mins 0secs
 Email: Enter Email to Download

12,999,602
 Specimen Records
 1,637,767
 Media Records
 200
 Recordsets

Advanced Search

Family: urticaceae
 Scientific Name:
 Genus:
 Country:

Present Missing

Add a field: please select
 Sort by: Scientific Name
 Direction: Ascending

[Tips & Hints](#)

Table view | [Label view](#)

Family	Scientific Name ^	Genus	Country
Urticaceae	Boehmeria	Boehmeria	China

[About iDigBio](#) | [Portal](#) | [Technical Information](#) | [Education](#)

[iDigBio Home](#) | [Portal Home](#) | [Search Records](#) | [Tutorial](#) | **Publishers** | [Feedback](#)

	Record Count	Media Record Count
Total Indexed (all data)	12999602	1637767
Total Published (all data incorporated in new workflow)	12960418	1601258

Publisher Name	Record Count	Media Record Count
CAS-IPT Show Recordsets	1841443	0
Museum of Comparative Zoology, Harvard University Show Recordsets	1708244	0
MyCoPortal Darwin Core Archive rss feed Show Recordsets	1541664	227368
Consortium of North American Bryophyte Herbaria Darwin Core Archive rss feed Show Recordsets	1529688	538101
KU Biodiversity Institute IPT Show Recordsets	1335374	0
CNALH Darwin Core Archive rss feed Show Recordsets	964356	374659
Florida Museum of Natural History IPT Service Show Recordsets	957972	1183

Specimen record search

Value added?

Publishers page

Data usage: What does this mean?

Number of queries:	7783
Number of records queried:	1055175684
Number of downloads:	816
Number of records downloaded:	21218617

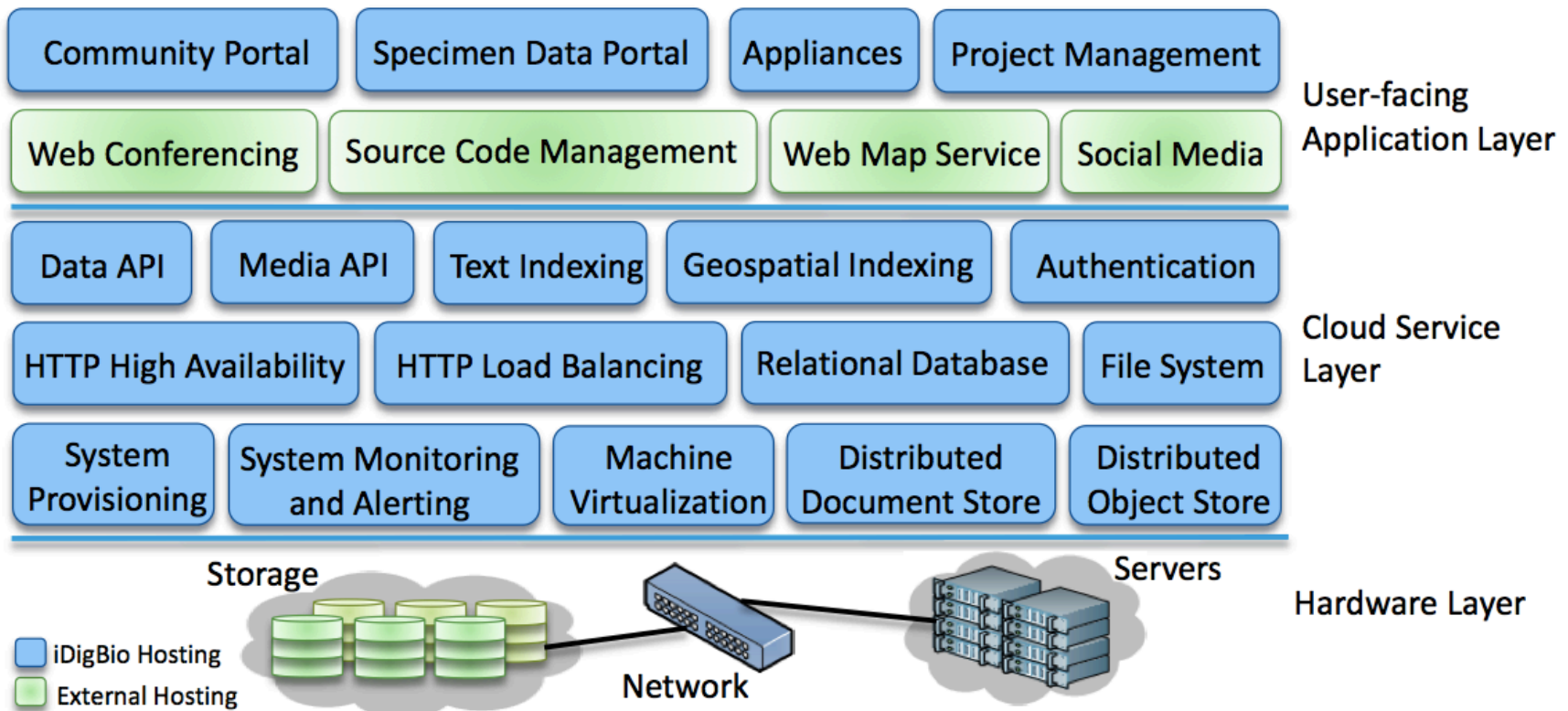
Attribution

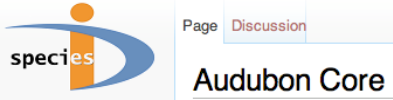
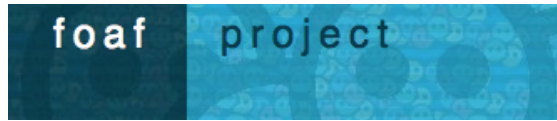
Are researchers and other downstream users, funded or not, providing sufficient level of attribution?

- Publishing data sets?
- Metadata Data providers, aggregators, source collections and institutions
- Object origin (collectors, geographic, political, cultural entities)

iDigBio Cyberinfrastructure

- Developed in consultation with stakeholders
- Implementation determined internally
- Feedback on prototypes solicited from community





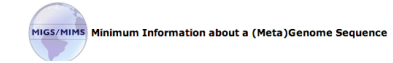
Access to Biological Collections Data - ABCD

XKCD on Standards



darwin-sw

Semantic web terms for biodiversity data, ba



Darwin Core

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)



Ecological Metadata Language (EML)



HISPID

Herbarium Information Standards and Protocols for Interchange of Data



RDF

Resource Description Framework (RDF)

Guidelines for authors

TAXON 63 (1) • February 2014: 231–234

Sequence data

DNA or protein sequences must be deposited in public data bases (GenBank, EMBL, etc.), and accession numbers must be

Voucher specimens are to be cited and deposited in a public herbarium. Cite vouchers and order voucher data as follows: Taxon name including authority, country, largest political subdivision (where applicable; locality and date optional, depending on type of study), *collector(s)* and *collector number*, herbarium acronym (plus herbarium accession number or barcode if available). Names

• • •

in Appendix format (see recent issue of *Taxon* for style). Lengthy specimen lists will be provided as Electronic Supplement to the online version, but not published in print due to space limitation.

Data publishing

Providers need attribution for data usage

- Should journal publishers maintain the dataset
- Should the authors post datasets to on-line, accessible resources (e.g., Dryad, DataONE, iDigBio, GBIF)?
- Identifier usage
 - global, unique, persistent, resolvable
 - Dataset level (DOIs)
 - Record level (Arks, UUIDs?)
- What resources can be considered as long-term digital archives?
- The call for attribution needs to emanate from collections community, political entities, and cultural groups.

