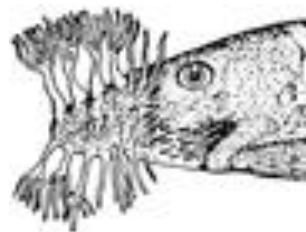


Biodiversity science: A new, unique cyberinfrastructure challenge: Or familiar, generic problem space?

Reed Beaman

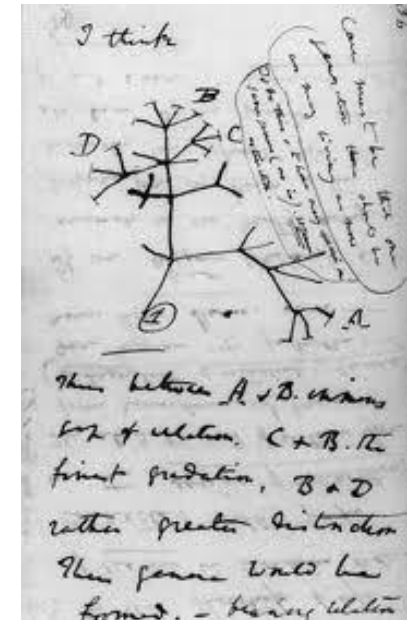
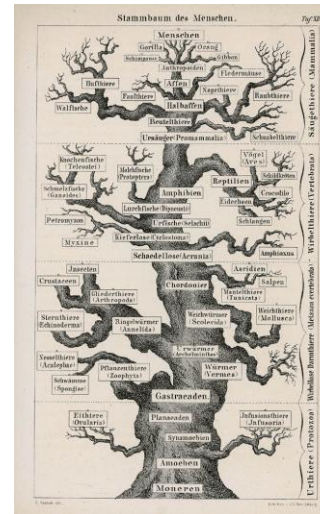
Florida Museum of Natural History

University of Florida



Biological Diversity

- Biodiversity: the variety of all forms of life, from genes to species, through to the broad scale of ecosystems.

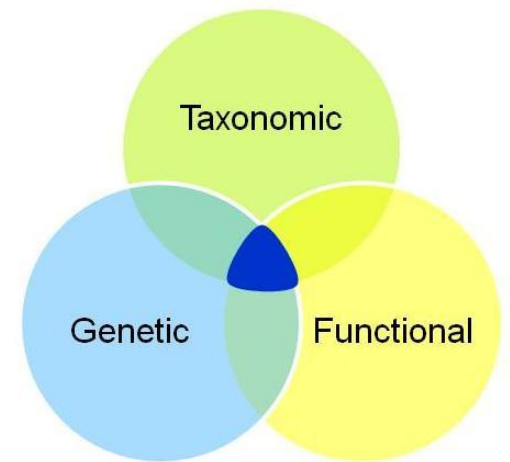


Elements ->> Process

- "atomistic" bias of western culture towards objects.
- Has biodiversity science been overly focussed on "inventory" of species, genes, ecosystems and has neglected processes that create and maintain natural values.
- Cf. reductionist focus on model species (Arabidopsis, mouse, human, etc.)

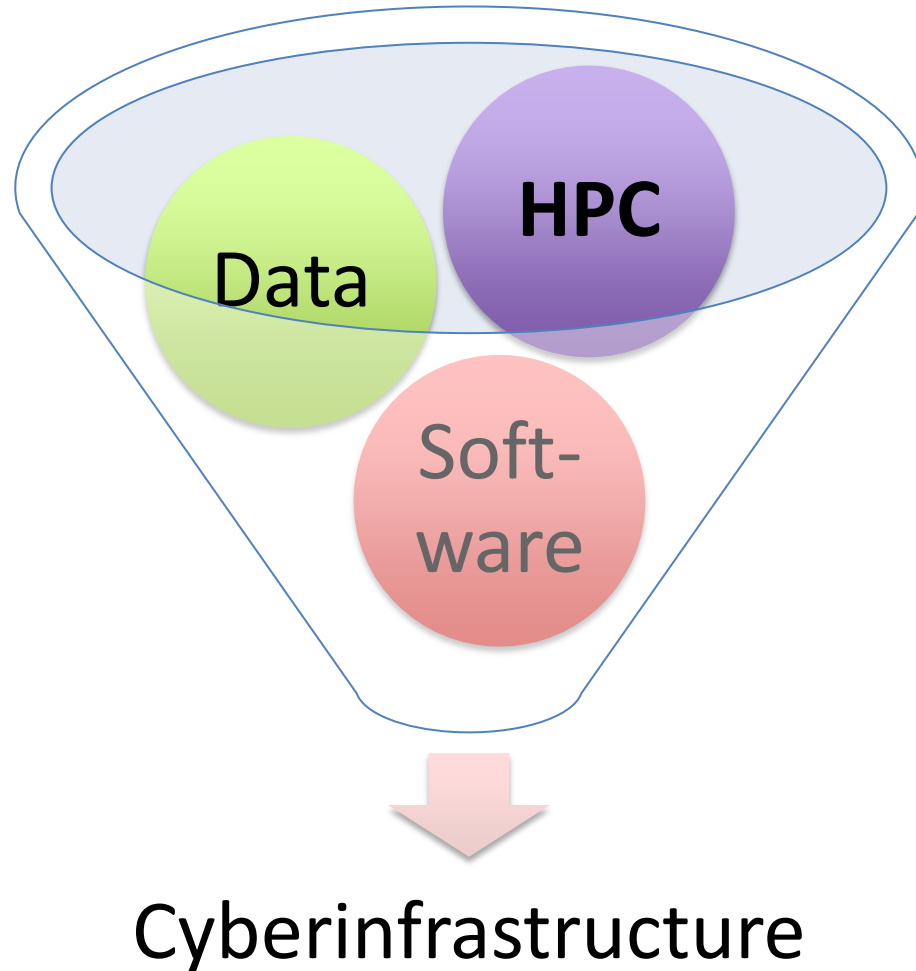
Biodiversity: the broader view

- NSF Dimensions of Biodiversity Program (cross-directorate)
- How many dimensions are there?



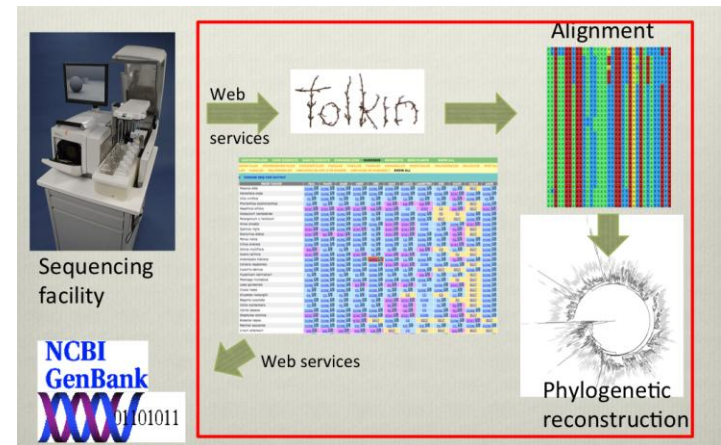
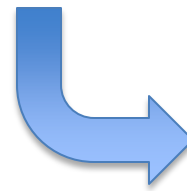
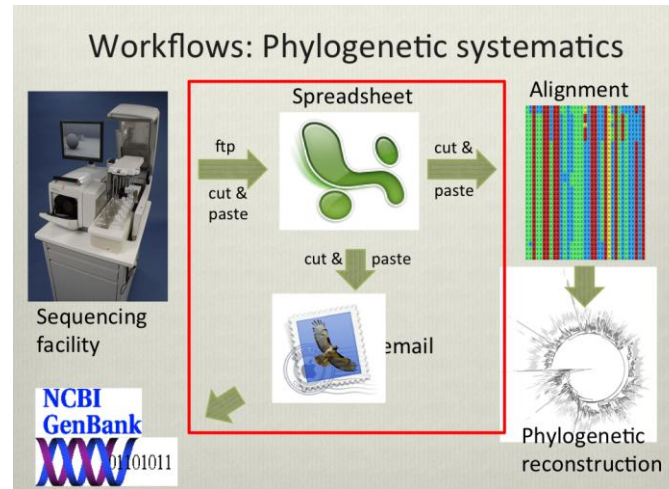
Systematics, taxonomy, evolution, biogeography, ecology, population genetics, genomics, metagenomics....

What does this have to do with computing?



Phylogenetic cyberinfrastructure

- CIPRES
 - AToL (NSF program)
 - iPlant/iPTOL,
 - AToL, AvaToL
-
- High Throughput sequencing
 - Genes to Genomes



Data? Gustav's Problem....



Generates ...

Lots of Data ..

The collage features several logos and images: NCBI (National Center for Biotechnology Information), Smithsonian National Museum of Natural History, CalPhotos (California Photo Library), geneIOUS (a logo with an orange hourglass), morphBank (a logo with a green and white background), and moorea (a logo with a yellow lion). There are also various scientific illustrations and photos of nature.

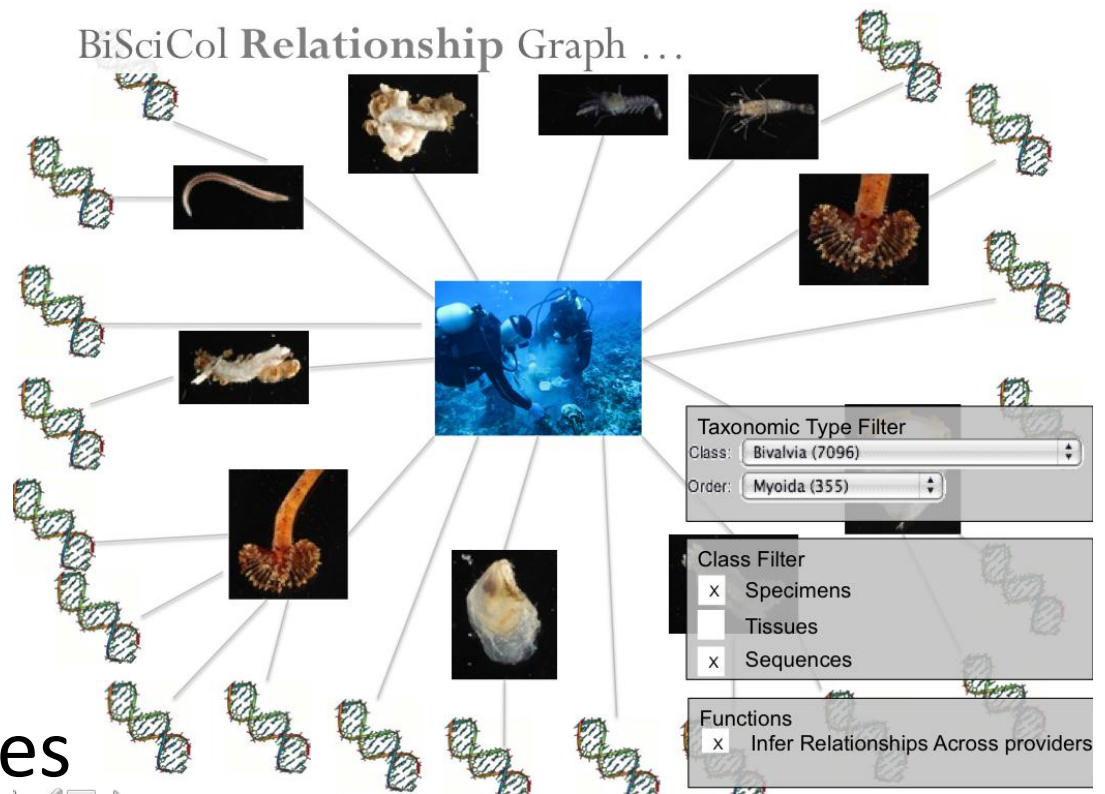
(Prefers to collect stuff)



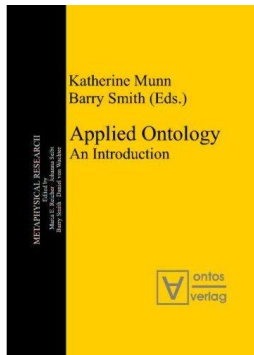
Due to project requirements and integration needs, Gustav is left navigating a plethora of redundant and disconnected distributed Databases. Lots of effort to track objects And their derivatives.

Knowledge networking challenge

- Ontologies and semantics
- Data curation
- Genotype to phenotype
- Gene expression and function
- Ecotypes
- Ecosystem services



250 years of Biological Ontologies

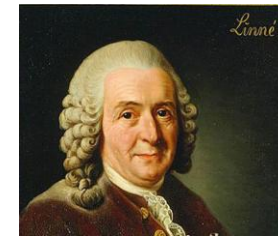
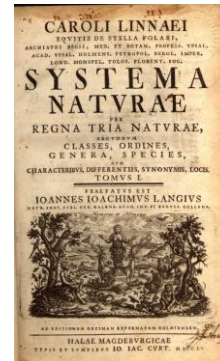


Chapter 1: Philosophy and Biomedical Information Systems

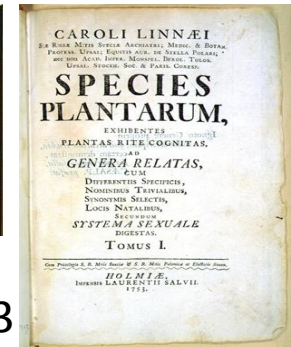
Barry Smith and Bert Klages

1. *The New Applied Ontology*

2008



1735



1753

Each partial category system will divide its domain into classes, types, groupings, or kinds, in a manner analogous to the way in which Linnaeus's taxonomies divided the domain of organisms into various upper-level categories (kingdom, phylum, class, species, and so forth), now codified in works such as the *International Code of Zoological Nomenclature* and the *International Code of Nomenclature of Bacteria*.

Storage and Archiving: 250 years of documenting biodiversity

- Scientific collections (Museums and Herbaria)
- ca. 2.5 billion collections objects (specimens) world-wide
- Documents 2 – 10 million species
- NSF ADBC program (Advancing Digitization of Scientific Collections): National HUB and Thematic Networks

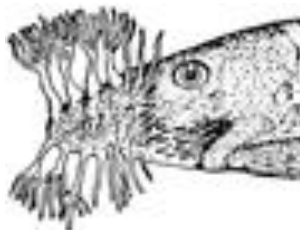




iDigBio
Integrated Digitized Biocollections

National Resource: Making collections accessible

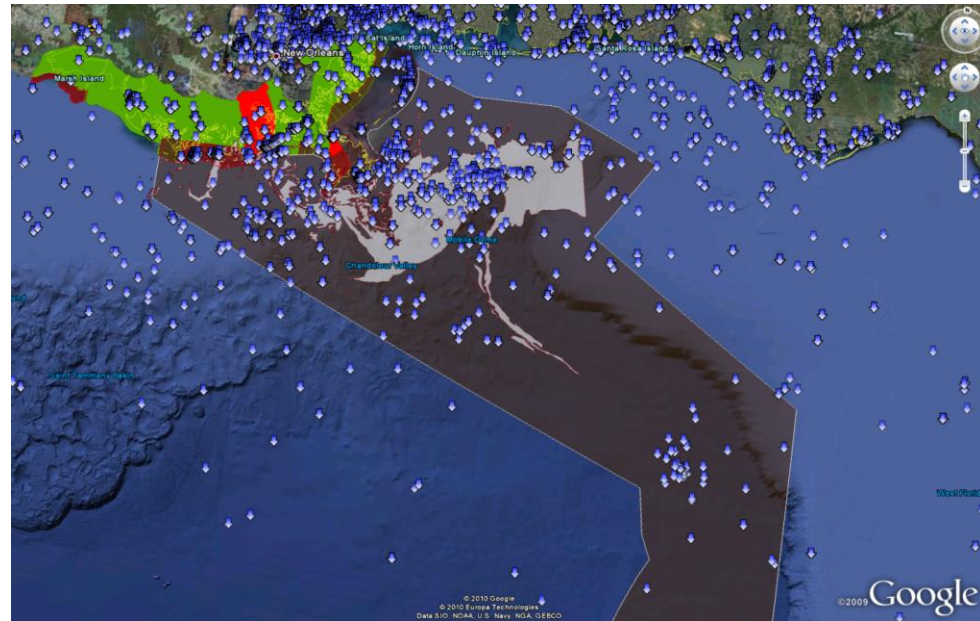
- Enabling answers to scientific questions
- Information integration across collections and across domains
- Adding value to collection data
- Digital data capture and
- management



Societal Challenge

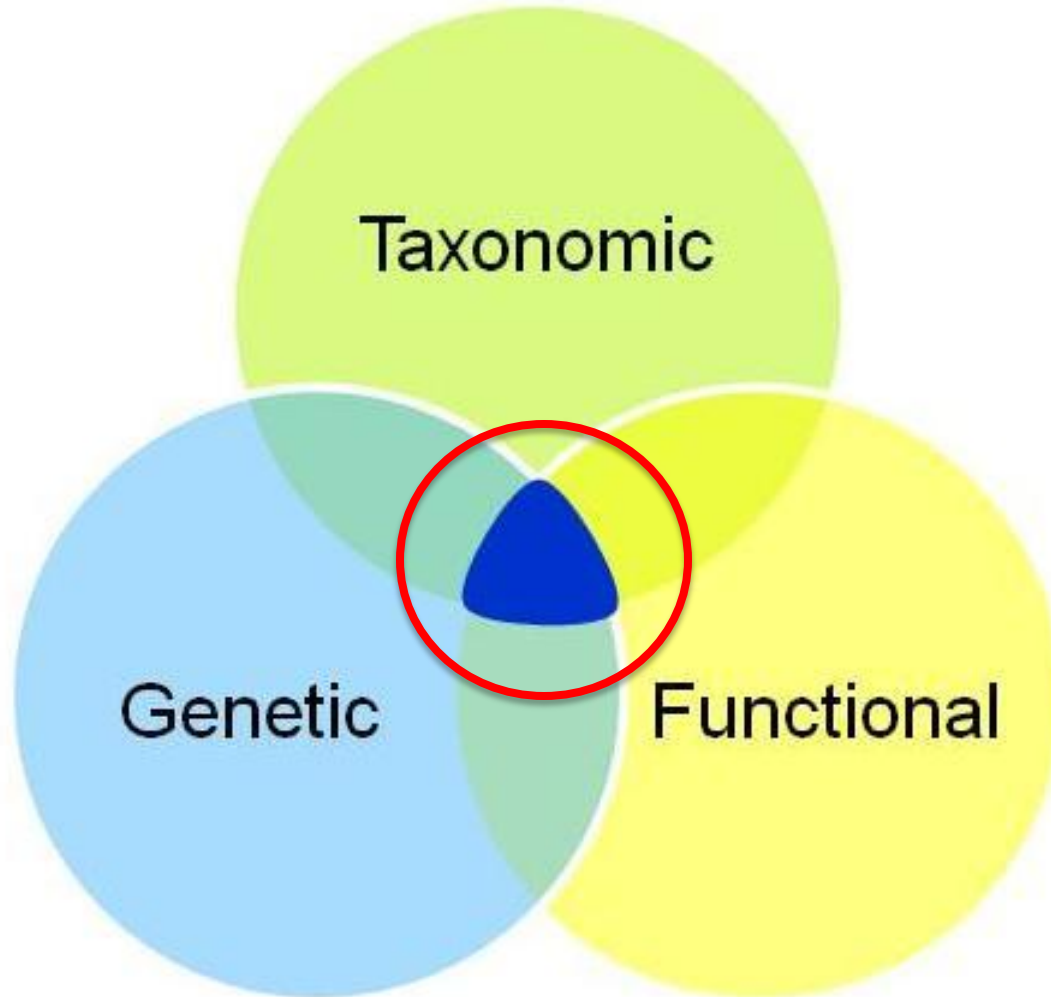
Oil Spill response

- How are we doing in terms of measuring effects on biodiversity?
- What information infrastructure is available for comparison to baselines, modeling, impact, and assessment?
- Data problem: Sensor networks, crowd-sourcing, specimens/observations

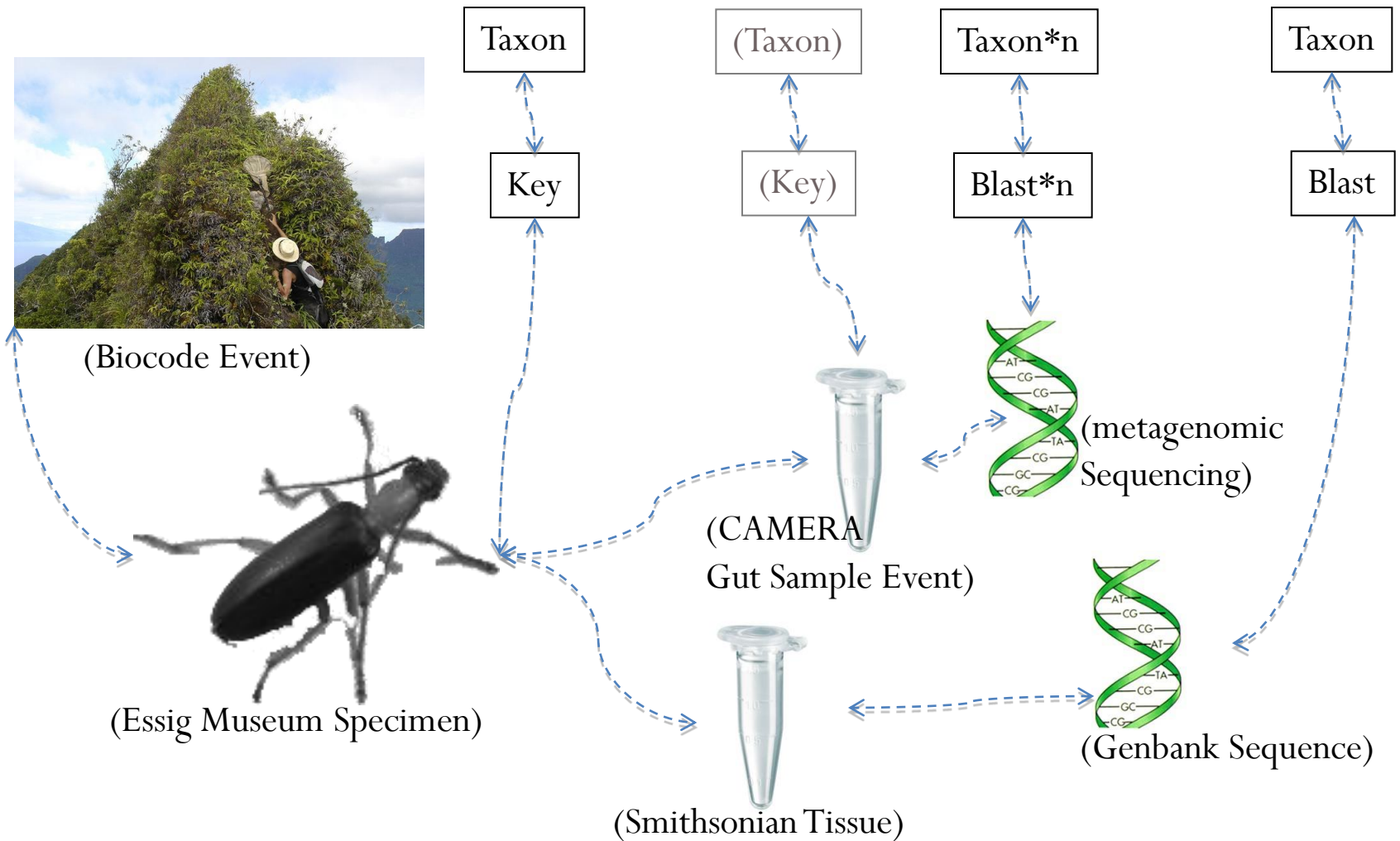


Courtesy Hank Bart, Nelson Rios, Tulane Univ.

Integration challenge



Moorea Biocode Example: Tracking biological material from field collection through analysis, across multiple systems



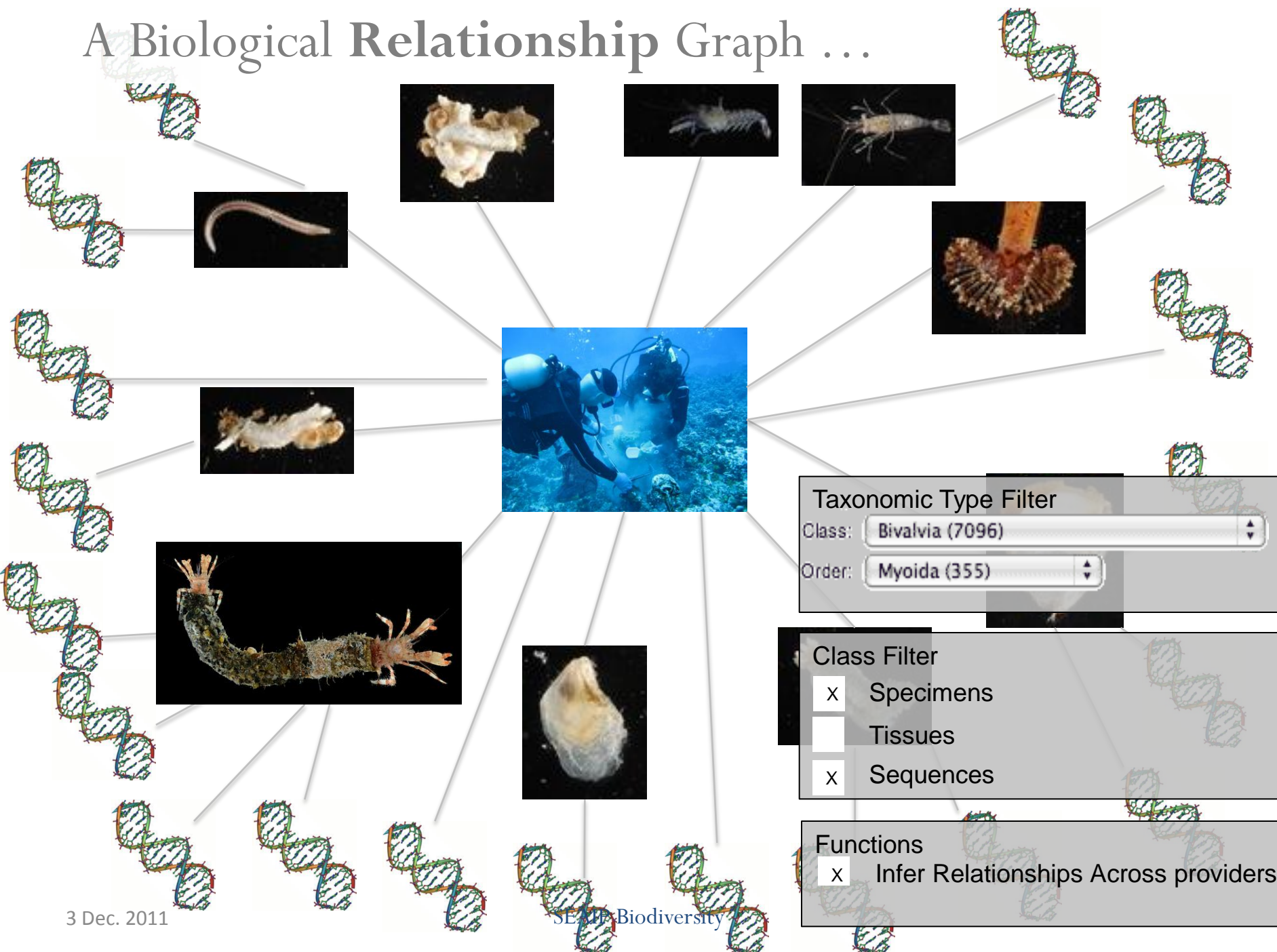
Linked queries

- List all participating institutions that house sequenced DNA samples for marine organism species Y
- For plant species X, indicate if an image and/or a tissue sample is available.
- List crab taxonomy updates by species and acceptance status by participating institution.
- For sponge species N, find updated determinations and list institutions holding the reference specimens
- List all species in publication Y that have determinations and locality updated in date range *d1-d2*.
- Find all plants of Species N collected within date range *d1-d2* and have digital images.

What relationships exist that haven't been explicitly expressed?

What can you discover with linked relationships that you can't do easily with distributed data (e.g., DiGIR over Darwin Core) or standard relational database queries?

A Biological Relationship Graph ...



Taxonomic Type Filter

Class:

Order:

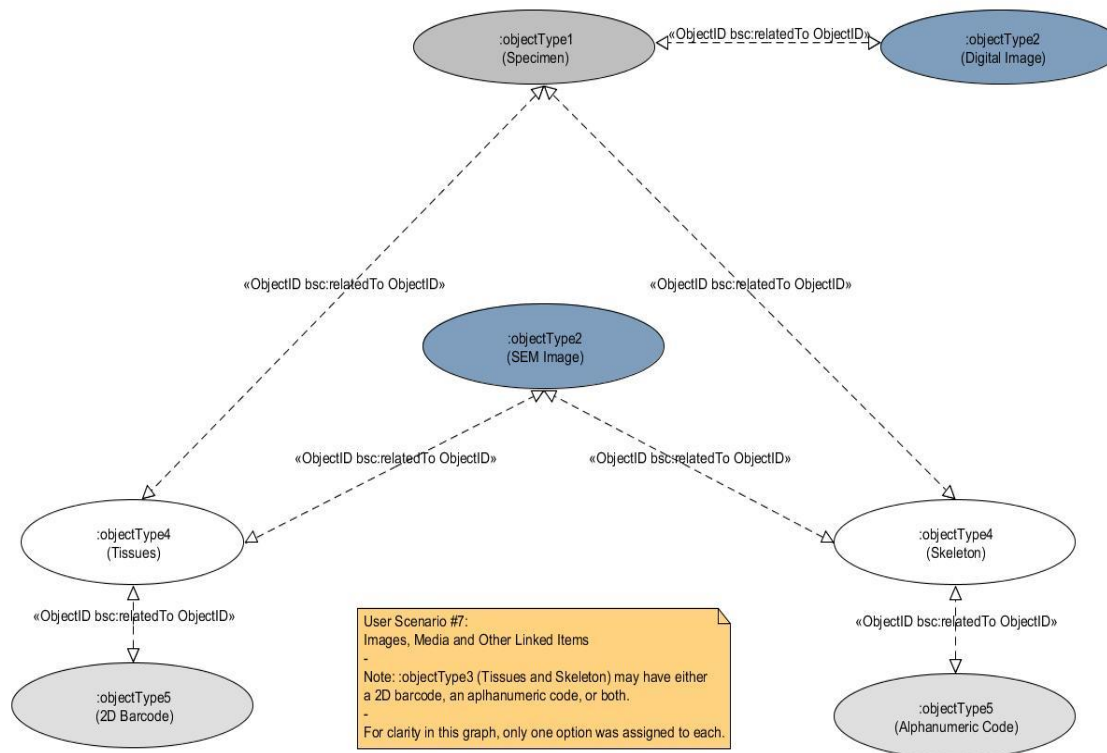
Class Filter

- Specimens
- Tissues
- Sequences

Functions

- Infer Relationships Across providers

How do we Track Biological Objects and their **Relations** Across Distributed, Heterogeneous systems?



Linking Identifiers Using Relationship Terms

An RDF
Statement:



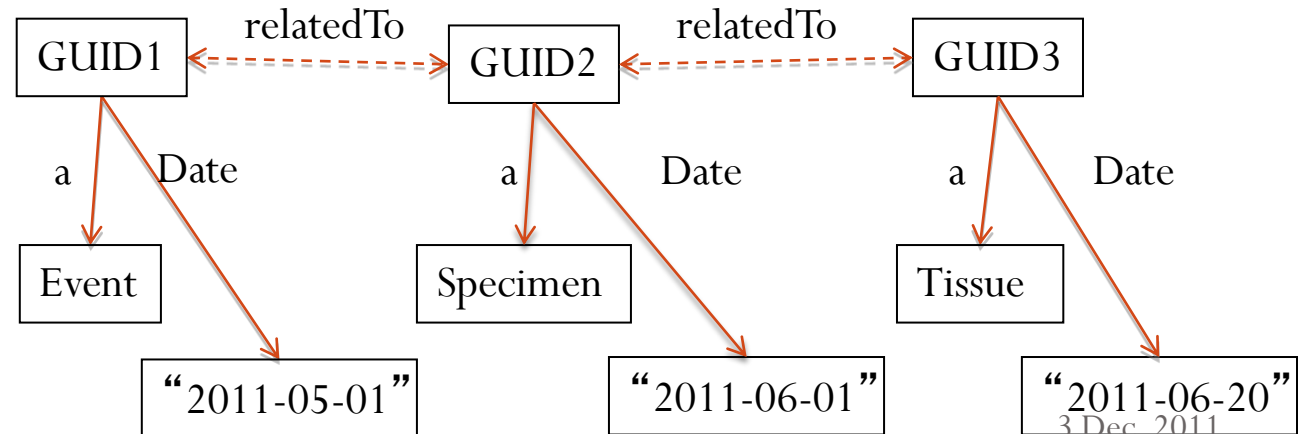
OR



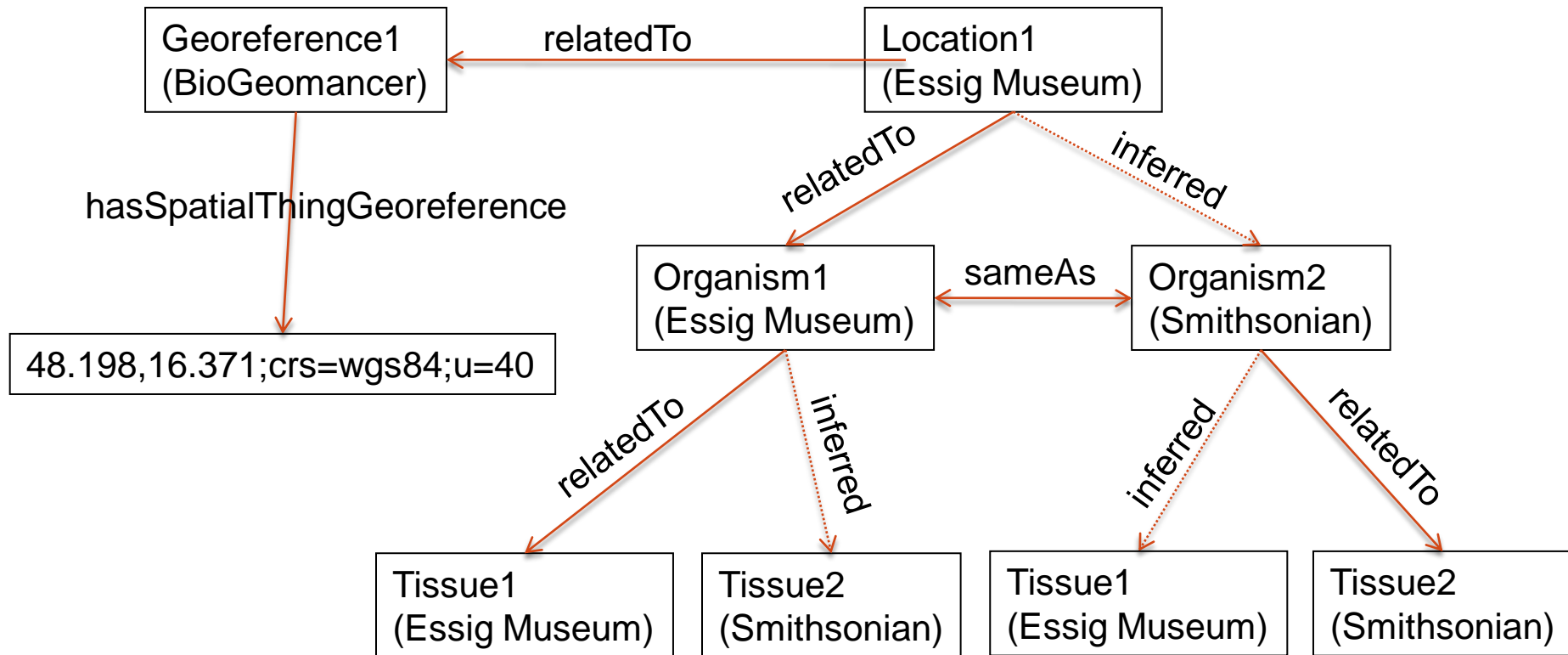
relatedTo
(Transitive):



A Simple
BiSciCol Graph
(graph=set of RDF
Statements):

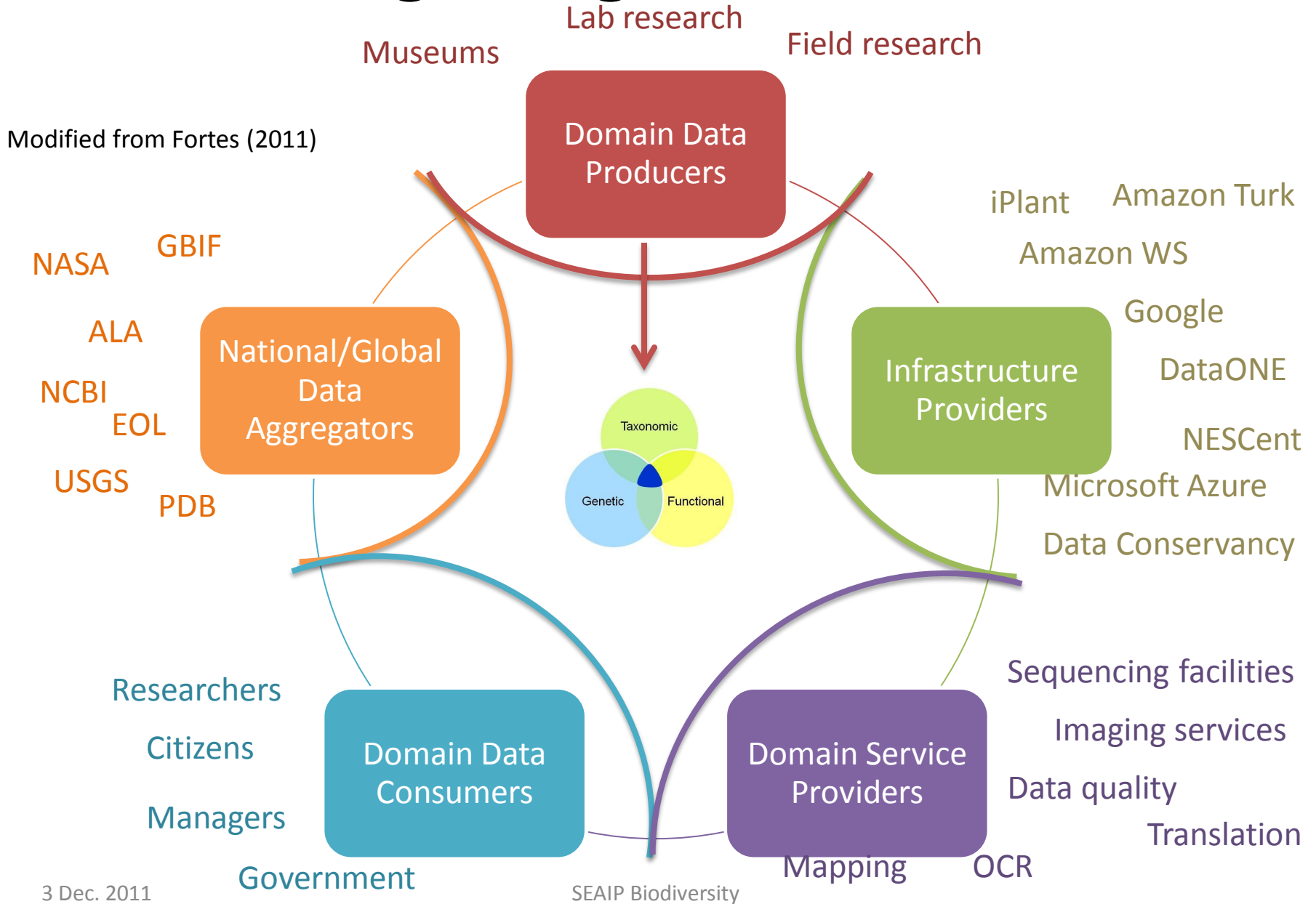


Inferred Relationship Chains



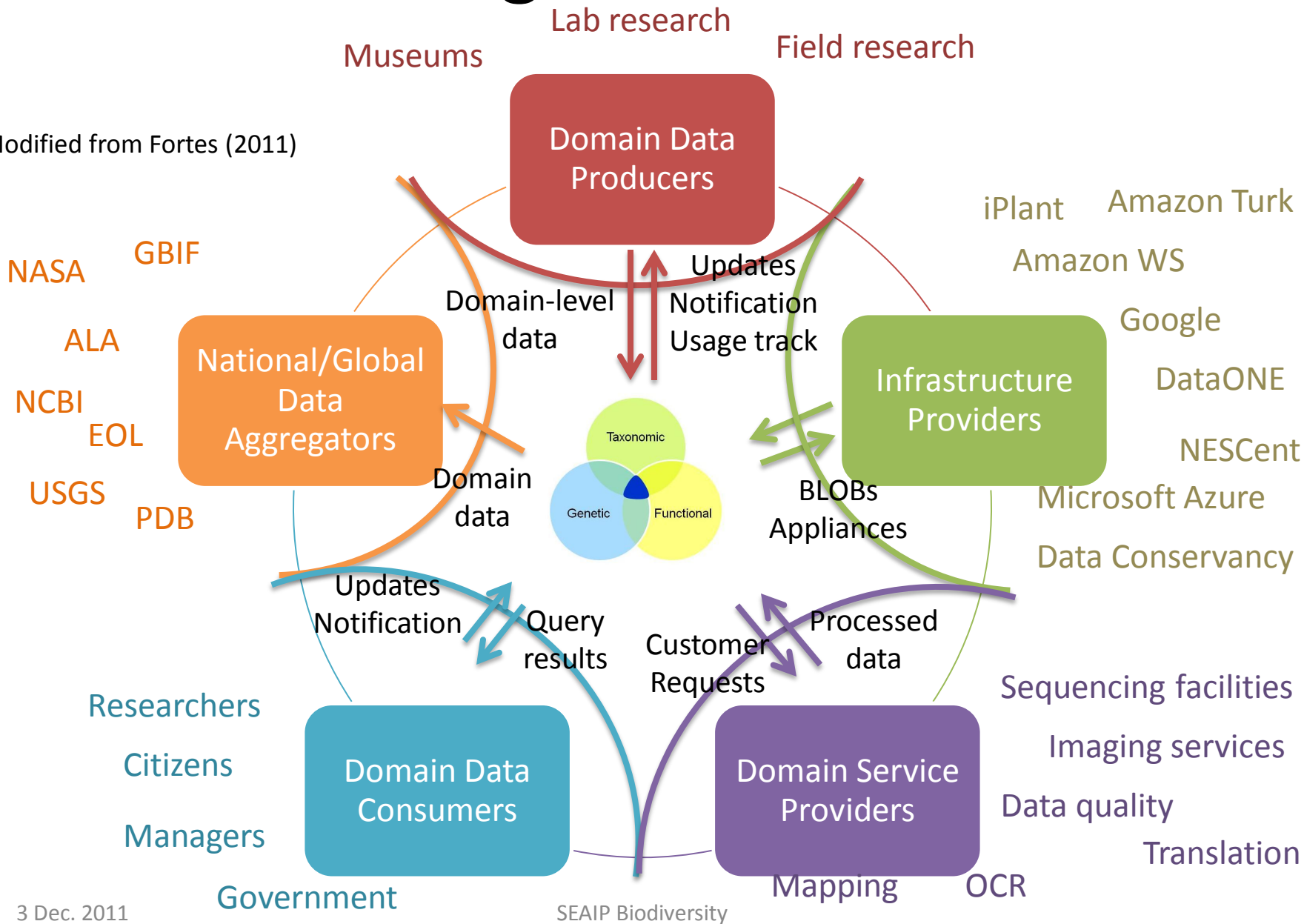
Even though Tissue #2 is not directly related to Location1, we can still infer its relationship through Organism1 and Organism2 being the same as each other.

Integrating Dimensions



Integration APIs

Modified from Fortes (2011)



Geospatial knowledge from primary sources

Not all is legacy or retrospective but based on imagery, GPS, and sensor nets

Projek Etnobotani Kinabalu (PEK)

- NEON/LTER
- Planetary biodiversity surveys (PBI)
- new plot and census data



Biodiversity valuation

Standing forests have value other than just timber

$$PV_B = \frac{\int_0^T P_b B(t) \cdot e^{-rt} dt}{1 - e^{-rT}}$$

Hartman (1976) modified by Caparros (2003)

- PV_B is the present value for biodiversity values;
- B the biodiversity function; and
- P_b is the biodiversity shadow price

P_b and B are not easy to value in current practice

Making collections relevant to economics

- Biological collections are the ultimate documentation of what, where, and why for vouchering biodiversity research
- Communicate value of collections based data through education and outreach
- Economic modelers can only use information readily available. Biodiversity and informatics collaborators must
 - Provide raw and synthesized data
 - Provide [more] complete data sets
 - Provide a basis for trust

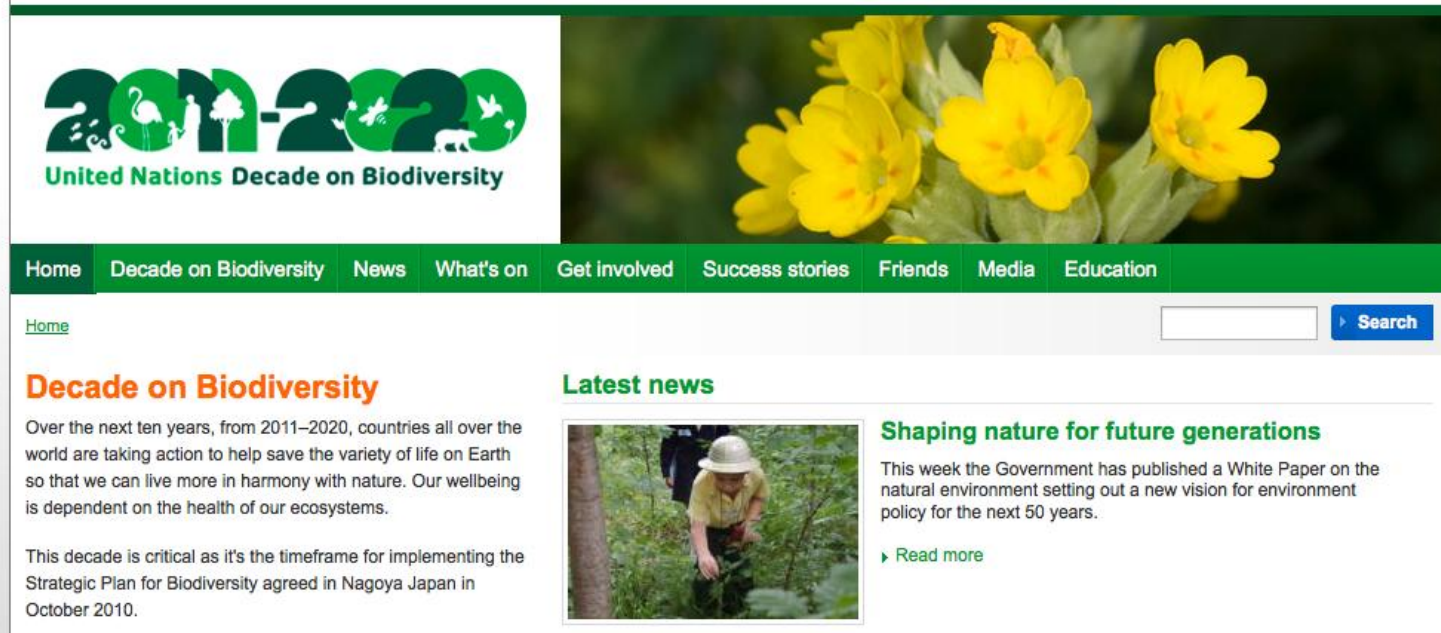
Relevant biological diversity metrics

based on vouchered biological collections

- Identifying hotspots
- Organism occurrence and distribution modeling
- Geochronological change
- Endangered/threatened/invasive species mapping
- Phylogenetic and genetic diversity
- Diversity of economically useful organisms

$$PV_B = \frac{\int_0^T P_b B(t) \cdot e^{-rt} dt}{1 - e^{-rT}}$$

Decade on Biodiversity: 2011-2020



The screenshot shows the homepage of the United Nations Decade on Biodiversity website. At the top left is the logo for the 2011-2020 Decade on Biodiversity, featuring icons of a swan, a person, a tree, a bird, and a cow. To the right is a large image of yellow flowers. Below the logo and image is a green navigation bar with links: Home, Decade on Biodiversity, News, What's on, Get involved, Success stories, Friends, Media, and Education. A search bar is located on the right side of the page. The main content area is divided into two columns. The left column has a heading "Decade on Biodiversity" and a paragraph: "Over the next ten years, from 2011–2020, countries all over the world are taking action to help save the variety of life on Earth so that we can live more in harmony with nature. Our wellbeing is dependent on the health of our ecosystems." Below this is another paragraph: "This decade is critical as it's the timeframe for implementing the Strategic Plan for Biodiversity agreed in Nagoya Japan in October 2010." The right column has a heading "Latest news" and a news item titled "Shaping nature for future generations" with a sub-image of a person in a hat working in a field. The text of the news item reads: "This week the Government has published a White Paper on the natural environment setting out a new vision for environment policy for the next 50 years." A "Read more" link is provided below the text.

- Projects mentioned were funded by NSF and the Moore Foundation. Material in slides provided by Jose Fortes and John Deck. Thanks!