Sharing Information Online

Greg Riccardi
Florida State University
iDigBio
griccardi@fsu.edu

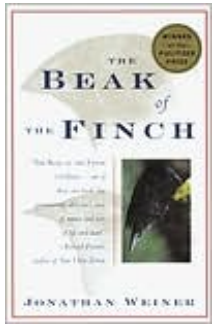
This material is based upon work supported by the National Science Foundation under Cooperative Agreement EF-1115210. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Topics in this presentation

- Moving from private to shared data
- Identifiers for objects and properties
- Representing relationship among objects
- Adding more detail with annotations

Imagining data from *The Beak of the Finch*

- Information about individual birds



tag	sex	mother	father	birthdate
1154	M			
1158	F			
1160	F			
1188	F			
1207	M	1158	1154	1/1/1975
1008	M	1188		2/12/1976
1330	F	1188	1207	3/15/1978

- Measurements

tag	date	beak1	beak2	wt	notes
1158	12/15/1974	25	12	30.5	
1158	1/15/1975	25	12	28.3	
1158	5/21/1978	25	12	32.1	
1207	1/1/1975			2.3	newborn
1207	2/5/1975	5	3	6	

Information for a single bird

- Example bird

Finch 1207		
id	**property**	**value**
1207	sex	M
1207	mother	1158
1207	father	1154
1207	birthdate	1/1/1975

- Example measurements of the bird

Measurement		
id	**property**	**value**
	bird id	1207
	date	2/5/1975
	beak1	5
	beak2	3
	wt	6

  - No identifier for the measurement!

Exporting information

- Step 1: improve identifiers and add fields

Finches							
Id	tag	sex	mother	father	birthdate	latitude	longitude
http://beaks.org/finch/1154	1154	M				30.2564	-88.1253
http://beaks.org/finch/1158	1158	F				30.2564	-88.1253
http://beaks.org/finch/1160	1160	F				30.2564	-88.1253
http://beaks.org/finch/1188	1188	F				30.2564	-88.1253
http://beaks.org/finch/1207	1207	M	1158	1154	1/1/1975	30.2564	-88.1253
http://beaks.org/finch/1008	1008	M	1188		2/12/1976	30.2564	-88.1253
http://beaks.org/finch/1330	1330	F	1188	1207	3/15/1978	30.2564	-88.1253

id	finch	date	beak1	beak2	wt	notes
http://beaks.org/measurement/113	http://beaks.org/finch/1158	12/15/1974	25	12	30.5	
http://beaks.org/measurement/114	http://beaks.org/finch/1158	1/15/1975	25	12	28.3	
http://beaks.org/measurement/115	http://beaks.org/finch/1158	5/21/1978	25	12	32.1	
http://beaks.org/measurement/116	http://beaks.org/finch/1207	1/1/1975			2.3	newborn
http://beaks.org/measurement/123	http://beaks.org/finch/1207	2/5/1975	5	3	6	

Exporting information

- Step 2: Improve properties
  - Find specific properties to use
    - *sex* replaced with *dwc:sex*
    - *latitude* with *dwc:decimalLatitude*
    - *mother* replaced with ?
  - Define new properties as needed
    - E.g. *beak1* and *beak2* are specific to this measurement
    - Create definitions and register with some repository

Better Identifiers

- Step 1: improve identifiers and add fields

Finches

Id	tag	sex	mother	father	birthdate	latitude	longitude
urn:uuid:40c842c9-c04c-489a-b20e-d84bfc16dedd6	1154	M				-0.4197	-90.37
urn:uuid:143ce8bc-ba5d-432b-be31-a5d40b691a63	1158	F				-0.4197	-90.37
urn:uuid:dc0ea8ef-9f87-457e-9989-4232e37fb0fd	1160	F				-0.4197	-90.37
urn:uuid:38554297-2144-413d-81db-7ea3e89790a2	1188	F				-0.4197	-90.37
urn:uuid:cb521b8d-dc14-4691-a9d4-80e4b7899d95	1207	M	1158	1154	1/1/1975	-0.4197	-90.37
urn:uuid:e0868656-c398-4dbf-907b-e1fbf0b0a1ee	1008	M	1188		2/12/1976	-0.4197	-90.37
urn:uuid:c3dd7903-2665-4749-a649-54dadb6a4999	1330	F	1188	1207	3/15/1978	-0.4197	-90.37

Measurements

id	isAbout	date	beak1	beak2	wt	notes
urn:uuid:025a42e4-4dd7-4b98-9ddc-0b13d664c205	urn:uuid:143ce8bc-ba5d-432b-be31-a5d40b691a63	12/15/1974	25	12	30.5	
urn:uuid:3b7a6f1b-b0e3-4d3b-a68c-a380fea3f0ca	urn:uuid:143ce8bc-ba5d-432b-be31-a5d40b691a63	1/15/1975	25	12	28.3	
urn:uuid:a3ca802a-93ed-459a-91ef-4593a4e43455	urn:uuid:143ce8bc-ba5d-432b-be31-a5d40b691a63	5/21/1978	25	12	32.1	
urn:uuid:432c0a0e-7013-4ee2-ab5a-6ef711f13708	urn:uuid:cb521b8d-dc14-4691-a9d4-80e4b7899d95	1/1/1975			2.3	newborn
urn:uuid:ff1deac1-23c8-4e14-a163-efd06f88c50e	urn:uuid:cb521b8d-dc14-4691-a9d4-80e4b7899d95	2/5/1975	5	3	6	

Identifier types

- Content-rich identifiers contain information
  - Simple identifier, attached to specimen
    - Number stamped on band: 1154
    - Catalog number
  - Compound identifier: globally unique
    - Darwin core triple (institution, collection, catalog number)
    - (BOTF, finch, 1154)
  - URI (uniform resource identifier): internet friendly
    - http://beaks.org/finch/1154
- Content-free identifiers
  - Cannot be parsed, remembered or typed
  - urn:uuid:40c842c9-c04c-489a-b20e-d84bfc16dedd6

Moving to standard vocabularies

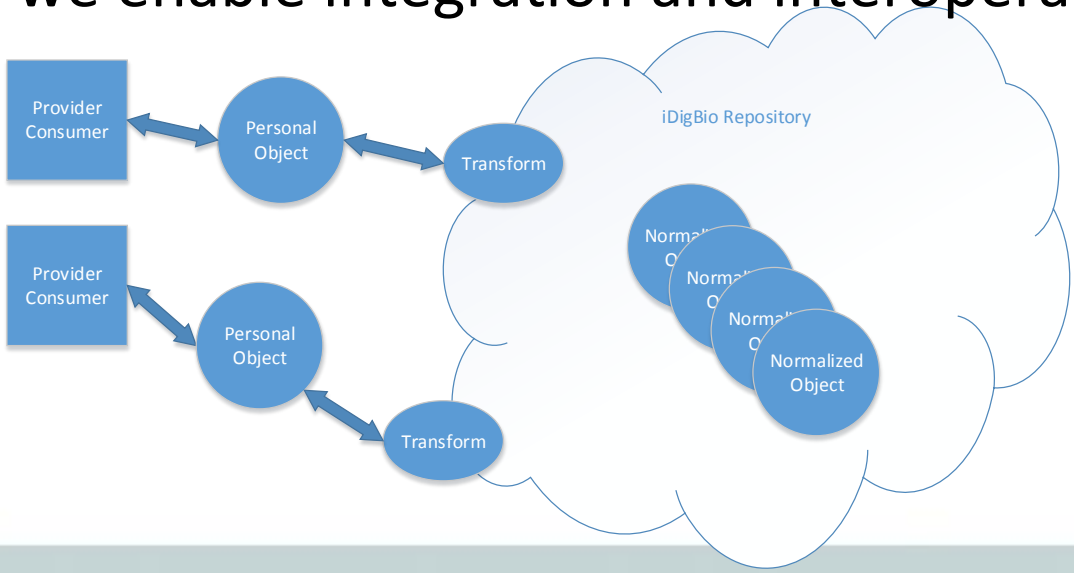
- Formal processes for defining properties
  - A property is an identified thing
    - http://rs.tdwg.org/dwc/terms/sex
    - dwc:sex is abbreviation
    - Property object has a definition
  - Resource-valued property is a type of relationship
  - Suppose *beak1* is a measurement of beak height
    - Properties defined include
      - Units of measure
      - Morphological feature measured
- Social processes for agreeing on properties and values

Media Vocabulary: Audubon Core

- Properties of a media object
  - http://terms.tdwg.org/wiki/Audubon_Core_Term_List
- Examples
  - dcterms:identifier
    - Unique code of the media object
  - dc:type
    - Recommended terms are Collection, StillImage, Sound, MovingImage, InteractiveResource, Text
  - xmpRights:UsageTerms
    - The license statement defining how resources may be used
  - ac:associatedSpecimenReference
    - A reference to a specimen associated with this resource.

Moving to common properties

- iDigBio is an aggregator without a fixed schema
  - People send information to the repository using their own schemas
  - iDigBio import tools transform into common structure
  - Providers do not always use standard property vocabularies
- How can we enable integration and interoperability?



Integration of properties

- iDigBio keeps a list of properties
- Users agree on similarity of properties
  - Different names for same property
  - Different properties that are similar

Relationships as annotations

- A relationship that needs its own properties
  - When, who, why, what evidence
- An annotation (*url1*) is an assertion of properties for objects
  - **On** 4 October 2013, Joe **claims** that specimen (*url2)* is of species *url3* **because** he disagrees with the determination on the label, and for **evidence**, he offers a set of image annotations (*url4*) showing morphological features that can be seen in the photograph (*url5*).
- Many relationships are expressed in the annotation

Identifiers

- Lots of uncertainty in community
  - What form of identifiers, what services to provided, etc.
- We need to
  - Emphasize identification of specimens and other objects
  - Help providers to see value of specimen identifiers
  - Remove obstacles to adoption
    - E.g. validate and advocate standard practice in collections managers
  - Move forward in spite of problems
- Current suggestion
  - UUID as basis of identifier
  - URI with embedded UUID
    - urn:uuid:f47ac10b-58cc-4372-a567-0e02b2c3d47
    - ark:/87286/B2/f47ac10b-58cc-4372-a567-0e02b2c3d479

Conclusions

- Must have identifiers for objects
  - Especially occurrences
- Must have agreement on properties
- Must have strategy for representing relationships
  - In provider databases
  - In repositories
  - In transit