

Discussion 3

Maximizing scientific impact

- Metadata quality
- Tools and incentives to facilitate archiving
- Crowd-sourcing approaches
- Assessment of quality and fitness for use



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.

Metadata quality

- What vocabularies are available and appropriate?
 - E.g. Dublin, Darwin and Audubon cores
- How do we divide metadata between media object and media subject?
- What are effective practices for connecting metadata to media objects?
- Consider Audubon Core vocabulary for biodiversity media
 - http://terms.tdwg.org/wiki/Audubon_Core_Term_List
 - AC describes the characteristics of a media object

Audubon Core Terms

- Vocabularies (example terms)
 - *Management*: identifier, type, title, modified
 - *Attribution*: rights, logo, credit, license logo
 - *Agents*: creator, provider
 - *Content coverage*: description, caption, physical setting
 - *Geography*: location shown, latitude, city
 - *Temporal coverage*: temporal, creation date,
 - *Taxonomic coverage*: scientific name, taxon coverage
 - *Resource creation*: digitization date, capture device
 - *Related Resources*: associated specimen, associated observation
 - *Service Access point* (for each version): URI, format

Tools and incentives to facilitate archiving

- DataONE website
 - [Metadata tools](#)
 - [Data Management Planning](#)
- Dryad repository for published data
 - Tools and how-tos
- What are the effective practices?
 - Storage
 - Retrieval
 - Archiving
 - Discovery
 - Embedded metadata

Assessment of quality and fitness for use

- What are the use cases?
- What are the minimal data standards for each use case?
- What tools are available to evaluate data standards and fitness
 - E.g. what do Vertnet and GBIF do with data cleaning?
- How can data cleaning results be incorporated in collection databases?

Crowd-sourcing approaches

- What can crowds do?
 - List of use cases
- What crowd-sourcing tools are appropriate for each use case?
 - E.g. NotesFromNature is used for transcribing labels