



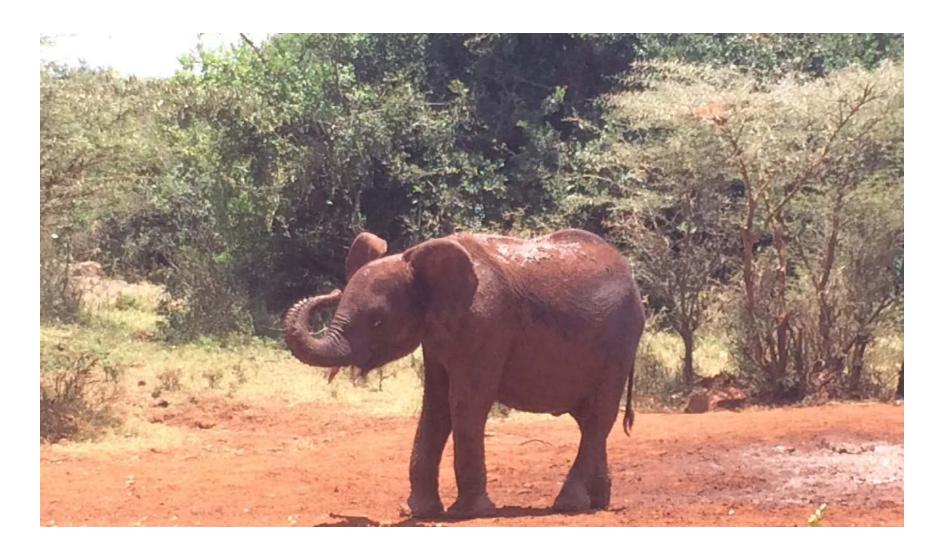
iDigBio: Image aggregation and standardizing metadata

Kevin J. Love and Alex Thompson Collection Imaging Techniques Workshop @ FLMNH November 2016











How we organize these data (and ourselves)

- Collections Management Databases
 - (Specify, Symbiota, Arctos, EMU, etc)
- Digital Asset Managers(DAMS)
 - (commercial and custom built examples)

Often, these are the mechanisms by which we publish data!



Why publish data? The 4 biggies of data aggregation

ACCESSIBILITY

Data Use

Data Quality

Attribution





What do we mean by publishing data?

making biodiversity data publicly accessible & discoverable, in a standardized form, via a URL that is reproducible and automated.







The relevant verb: to publish

- No sending, No pushing
- No action occurs <u>except</u> when you first send an email to say 'my data are ready and you can find it <u>here</u>'.
- Once we know the 'here', we know to pick up any updates AS LONG AS YOU RE-PUBLISH



3 ways to get media to iDigBio:

 Use Audubon Core extension in IPT

Via Symbiota

Media appliance



Required Media Metadata

- id (coreid of the specimen it links to)
- identifier (its own GUID)
- format (image/jpeg)
- accessURI (public path to your best quality jpg)



IDENTIFIERS

Every media record needs an identifier.

• We like UUIDs with a prefix:

urn:uuid:2d5d3a8f-7a18-4825-a129-4a32b4ae58b8



DATASET INFO: info about the provider (metadata)

Include your dataset metadata with your provider information (eml.xml):

- responsible parties (name, address, email, role)
- institution name, institution code, logo
- URL to the data at your institution
- descriptive paragraph about the institution, collection, and the dataset



DATASET INFO: update collections list

iDigBio Collections

https://www.idigbio.org/portal/collections

Index Herbariorum (Botany)

http://sweetgum.nybg.org/ih/

GRBio.org Repositories:

http://grbio.org/find-biorepositories

Do you know what your institutionCode is?



Data Quality: Consider searchability in the aggregate

Discoverability of rights and ownership:

dc:rights, xmpRights:Owner, xmpRights:WebStatement

Non-record descriptive metadata:

ac:tag and a controlled vocabulary for views (dorsal, ventral) or morphology (leg, wing, thorax)

Tics: * [] {} ?

Use the verbatim and remarks fields for things that do not fit the definitions.



Media with field "dcterms:creator"





Media with field "dwc:scientificName"





When is my work done?

- Digitization is never done
 - Label data
 - Georeferenced

- Not until *your data are in iDigBio*.
 - It is not enough to get images into Specify
 - Publish, re-publish with updates



Get involved!



idigbio.org/wiki



facebook.com/iDigBio



twitter.com/iDigBio



vimeo.com/iDigBio



idigbio.org/rss-feed.xml



idigbio.org/events-calendar/export.ics









