Natural Science Collections: Matching Lessons to *a new* Community

Applications

Collection Catalog

Researcher

from the field and / or from the specimen

Manage Data

encoding
standardization
outlier detection
taxonomy
duplicate detection
annotation
georeferencing

Possibilities

species ranges, outlier discovery, new species, gaps in collecting, traits, relationships, predictive niche models, collector maps,...
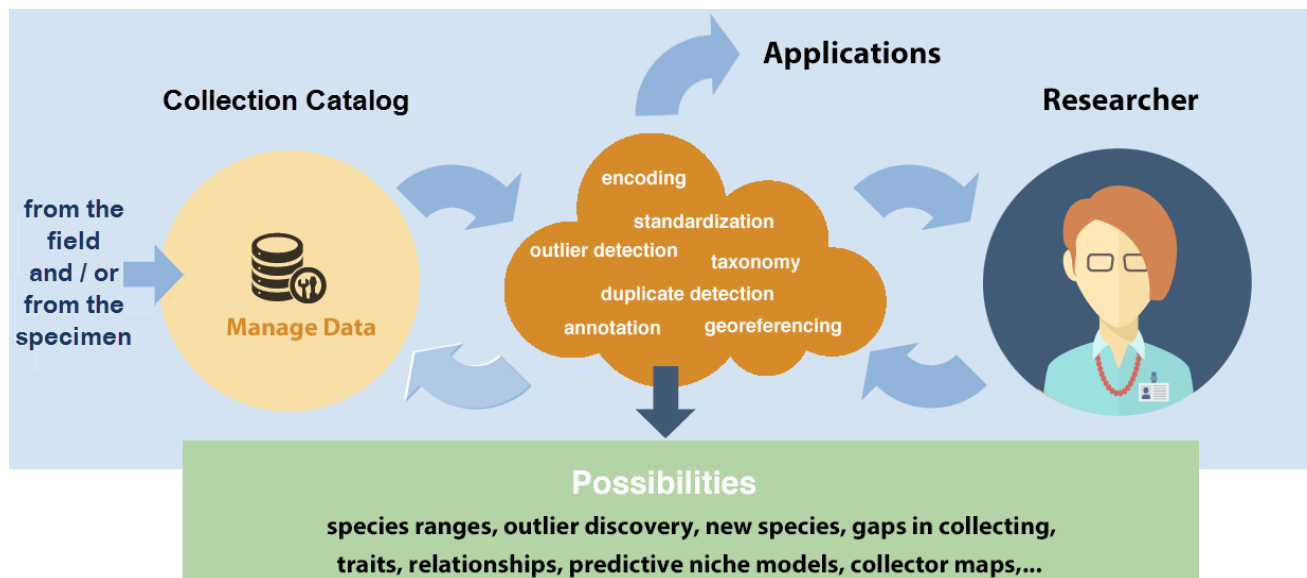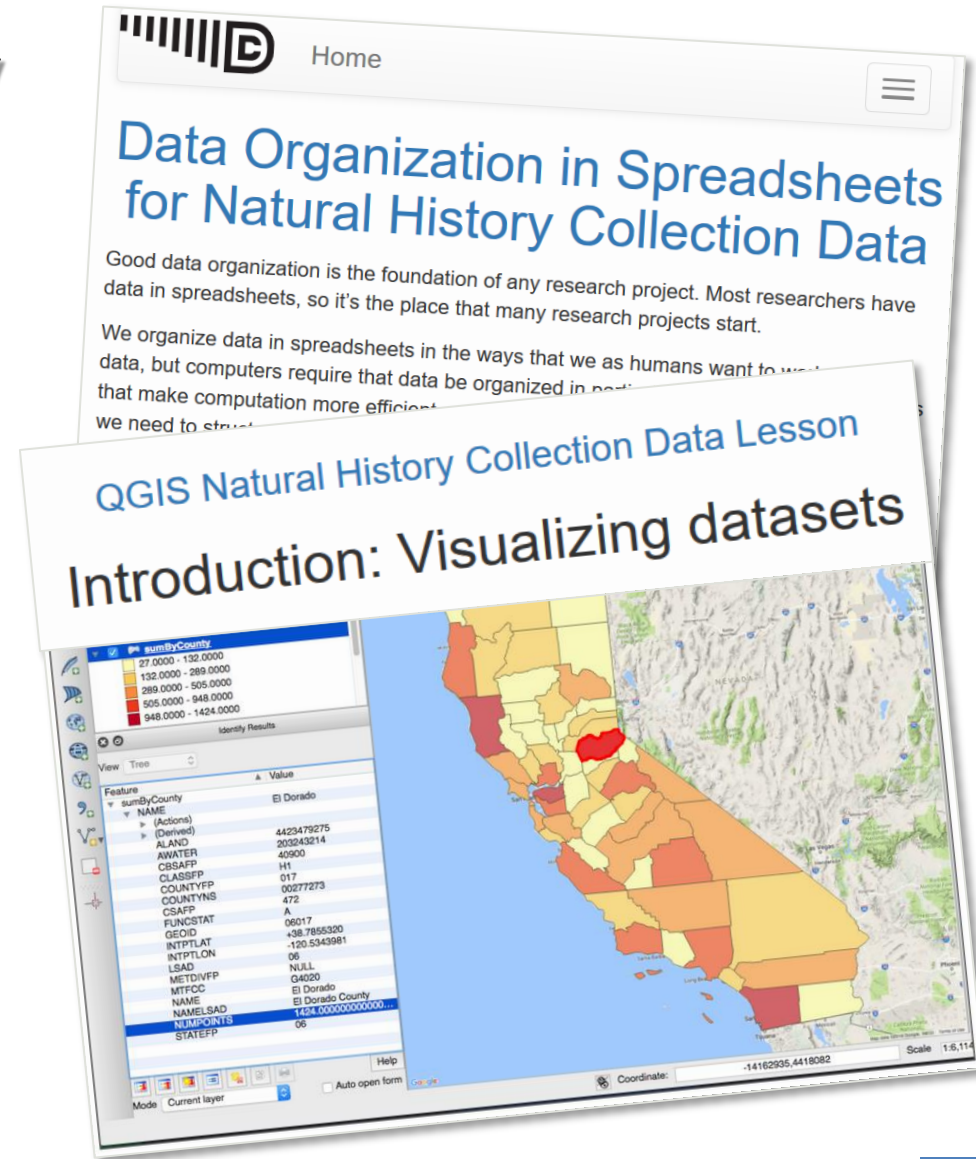
@idbdeb

Deborah L. Paul, Florida State University, @iDigBio
CarpentryCon 2018
University College Dublin, Ireland

Prototype using Ecology Lesson as Foundation

- better spreadsheet skills
- Open Refine
- intro SQL
- intro R and data visualization using R
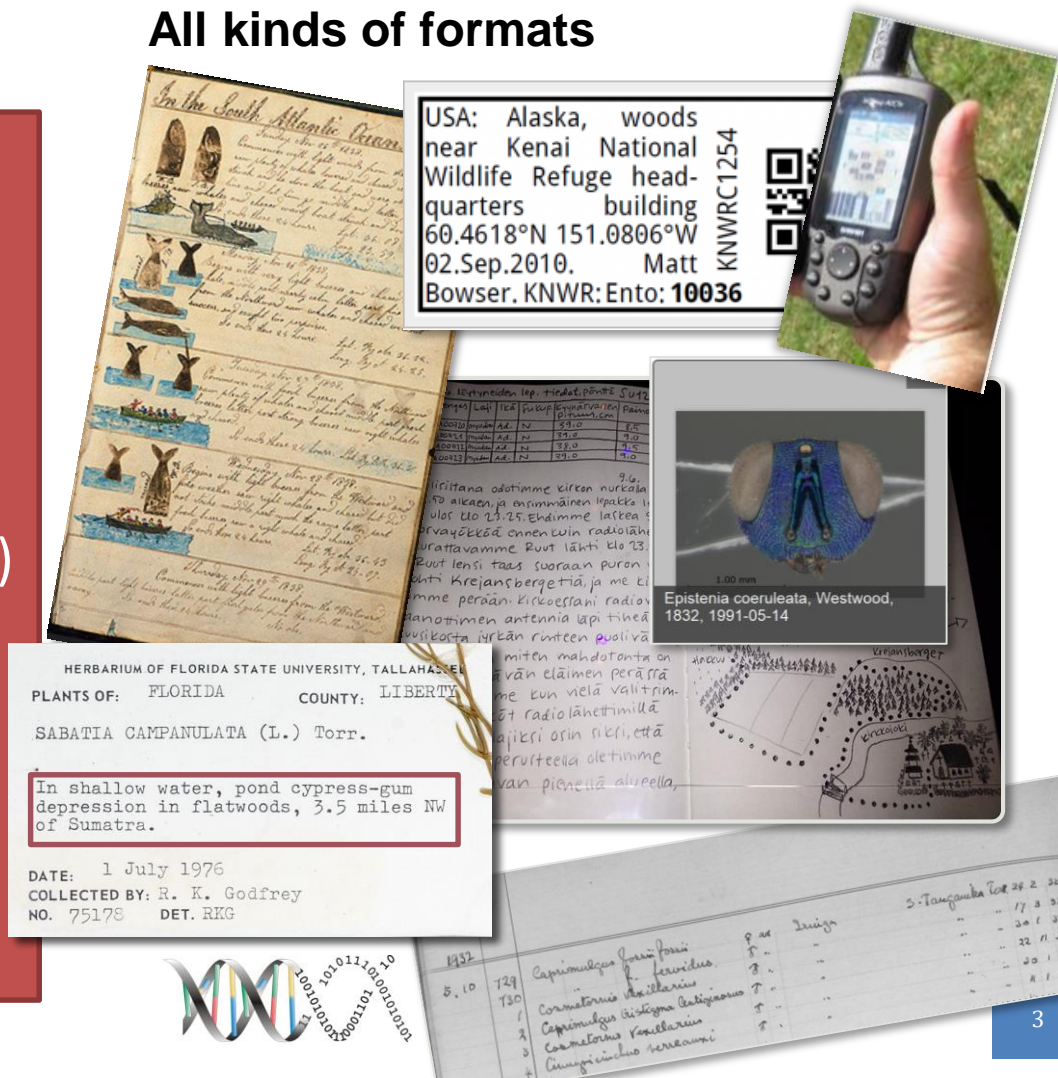- *QGIS lesson*
  - *UCSB graduate student Sara Lafia*

Data Organization in Spreadsheets for Natural History Collection Data

Good data organization is the foundation of any research project. Most researchers have data in spreadsheets, so it's the place that many research projects start.

We organize data in spreadsheets in the ways that we as humans want to work with data, but computers require that data be organized in particular ways that make computation more efficient.

we need to structure

QGIS Natural History Collection Data Lesson

Introduction: Visualizing datasets

https://data-lessons.github.io/QGIS-nhcdata-lesson/

What do our collections data look like?

	A	B	C	D	E	F
1	coreid	dwc:scientificName	dwc:locality	dwc:decimalLa	dwc:decimalLongit	dwc:eventDat
50	40996d8c-6b6e-425a-99ed-184826829e90	Carabidae	Wildcat Canyon, lakeside	32.913257	-116.927522	1963-05-05
51	4613bb37-a507-4882-a109-f6b8fb39aa45	Akephorus marinus	Ocean Beach, San Diego Co., CA	32.7508	-117.2517	
52	472f8f06-c9e4-4a01-bc1e-d814db70e7aa	Tetragonoderus pallidus	Short Canyon, 1 mi W jct Hwys 14 and	35.707267	-117.893225	2004-03-29
53	49				-122.0333	1913-05-06
54	4b				-117.157255	1914-04-08

Typical data

- **what, where, when, who**
  - **and …**
- taxonomic data
- geolocation data
- associated taxa
- habitat, soil type
- abiotic data (temperature, …)
- trait data (genomic, morphometric, behavioral, phenological, …)
- paleontologic/geologic data (stratigraphy, isotope, …)
- media

All kinds of formats

USA: Alaska, woods near Kenai National Wildlife Refuge head-quarters building 60.4618°N 151.0806°W 02.Sep.2010. Matt Bowser. KNWR: Ento: **10036**
KNWRC1254

Epistenia coeruleata, Westwood, 1832, 1991-05-14

HERBARIUM OF FLORIDA STATE UNIVERSITY, TALLAHASSE
PLANTS OF: FLORIDA   COUNTY: LIBERTY
SABATIA CAMPANULATA (L.) Torr.
In shallow water, pond cypress-gum depression in flatwoods, 3.5 miles NW of Sumatra.
DATE: 1 July 1976
COLLECTED BY: R. K. Godfrey
NO. 75178   DET. RKG

What possible skills to include?
Skills needed (not exhaustive, not definitive)

some known common skill needs

- data evaluation (fitness for use)
- better spreadsheet skills
- (meta) data standards
- data publishing and reproducible research (FAIR)
- using authority files (taxonomy, geography, people)
- georeferencing, geospatial data skills
- data visualization
- data formatting and transformation
- scripting and apis
- specific tools/software: GeoLocate, Open Refine, GBIF tools, Kurator, GNR, collection management software,...

- data and media: management
- data quality (messy data, cleaning, validation, automation), visualization, data life cycle

repetitive tasks

- data cleaning
  - transcription
  - editing
  - updating
  - mapping
- batch editing
- data import, export

collection data issues

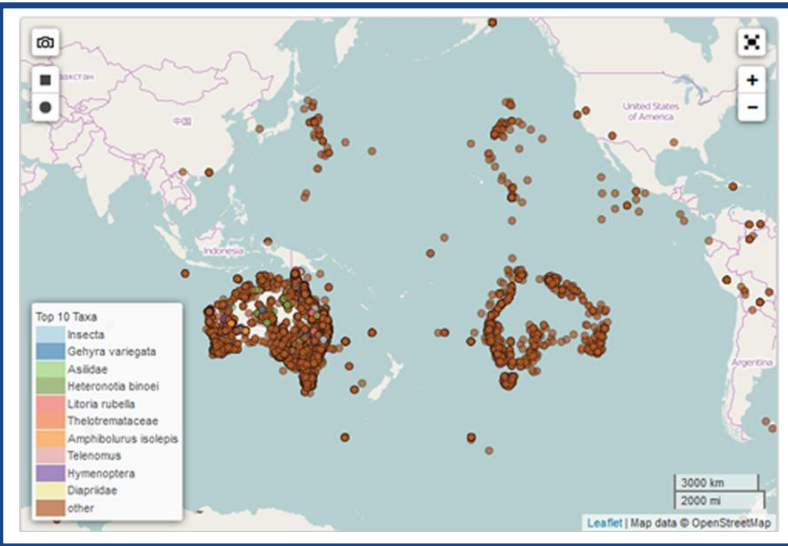Hannah Frost
@feefifofannah
Following

From a @HydraInABox interview: "People will put anything and their dog in the date field. It's absolutely astonishing."

Country: united k

united kindgom
united king
united kingdom
united kingdom (england)
united kingdom (scotland)
united kingdom (wales)
united kingdom [?]
united kingdom of great b
united kingdom?

Next steps

- Models for discipline-specific lesson development
- Upcoming outreach
  - CarpentryCon 2018
  - Community input at Digital Data II
  - Instructor training/mentoring in our community
    - develop a renewable resource – supply to meet demand
- A New Data Carpentry Community

Go raibh maith agat
Thank you!

www.idigbio.org

facebook.com/iDigBio

twitter.com/iDigBio

vimeo.com/idigbio

idigbio.org/rss-feed.xml

webcal://www.idigbio.org/events-calendar/export.ics