Opportunities for Change in Crisis:
Human and Infrastructure Evolution for 21st-Century Collections

Deborah L Paul @idbdeb, Erica Krimmel, Dirk Neumann, Megan McCuller, Gabriela Hogue, Laurence Livermore, Shelley James, Christopher Marshall #SPNHC2020 #NATHIST2020

SPNHC and ICOM NatHist Virtual 2020 Symposium: Envisioning Collections Management for the Evolving Biodiversity Data Lifecycle 23 June 2020

Our Symposium Story Overview

Deborah Paul, et al ❏ Human and Infrastructure Evolution for 21st-Century Collections

G Hogue & M McCuller ❏ The Dynamic Role of Collections Managers

Sylvia Orli ❏ Collection Management in a changed digital landscape: progress and challenges for the US Herbarium with a newly digitized collection

Anton Güntsch ❏ Better Together: Merging our knowledge about people, places, collections, and taxonomies with Wikidata

Elycia Wallis ❏ Interpreting data quality, or how collections can approach providing useful and useable data

T Mayfield-Meyer ❏ Data Management Strategies for the Extended Specimen

Advancing the Digitization of Biological Collections
iDigBio Hub and Thematic (Museum) Collection Networks

total: 121,428,342

Digitization
Workflows & Protocols
Dissemination

Research Use
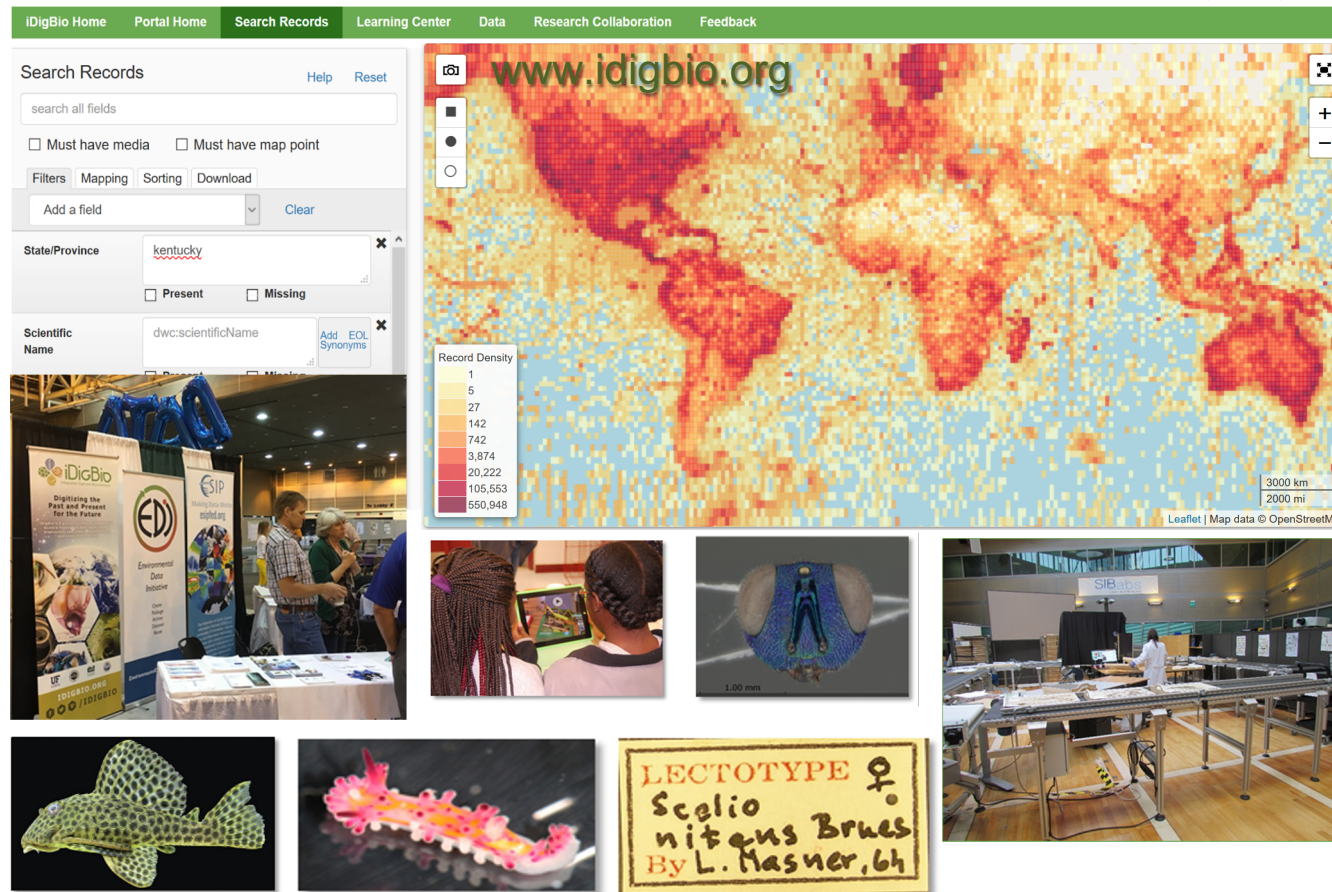Cyberinfrastructure
Tool collaboration
Portal development
ENM workshop
Research focus
Data quality
APIs

Training
Biodiversity informatics
Data skills and literacy
Collections software
Imaging
Project Management

Education Outreach
Diversity Inclusion
Citizen Science
K-12 materials
Undergraduate
Fossil Clubs
Mentor teachers

Methods
Workshops
Webinars
Symposia
Conferences
Working Groups
Short Courses
Adobe Connect
Listservs
Publications
Social Media

@idigbio

Human and Infrastructure Insights

Data and Technology Leadership

Data Fitness, Data Loss
fitness, management, attribution

Communication
product management, sharing horizontally

People and Capacity Development

Duplication of Effort
people, taxon names, localities, …

Collection and Digitization Metrics
what to measure?

Vision Evolution
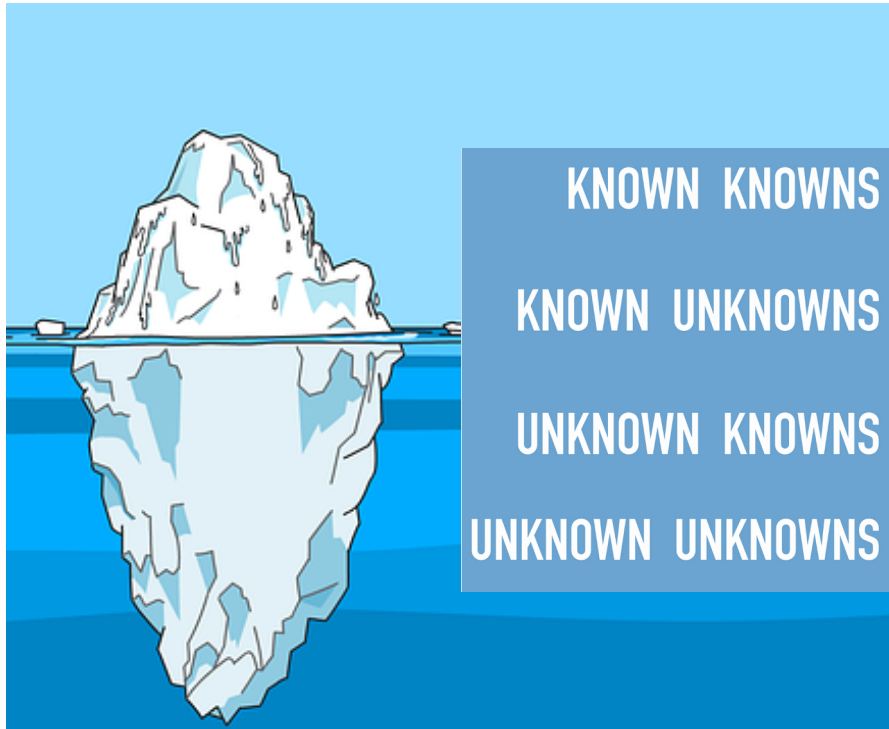
KNOWN KNOWNS	analogies, lateral thinking
KNOWN UNKNOWNS	build hypothesis, measure, iterate
UNKNOWN KNOWNS	brainstorming, group sketching
UNKNOWN UNKNOWNS	research, exploration

@GoodStickLtd

https://uxdesign.cc/the-knowns-and-unknowns-framework-for-design-thinking-6537787de2c5

… a Skilled *Inclusive and Diverse* Workforce
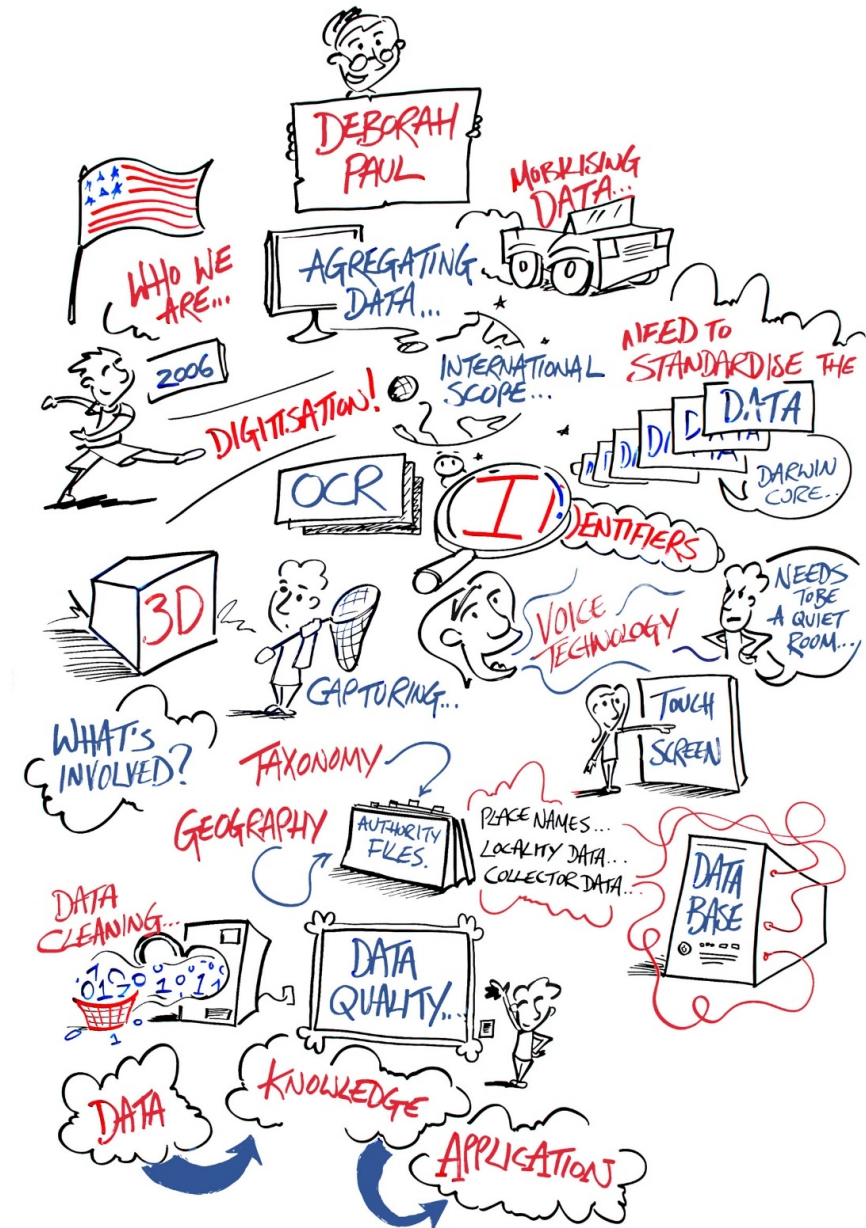
- is critical to all national research infrastructure *and for multidisciplinary research*

- Ongoing commitment to training and career progression, not only by the facilities and projects but also by the universities and research institutions that harness them, is essential.

graphic by Dorothy Allard

PEOPLE IN THE LOOP

What is your professional development experience personally? in your country? in your organization? got metrics?

Department of Education and Training. 2016. National Research Infrastructure Roadmap. https://www.education.gov.au/2016-national-research-infrastructure-roadmap [Last accessed 14 February 2019].
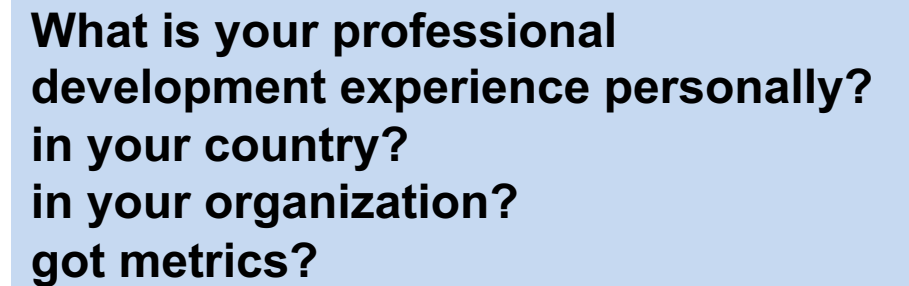
Skills Needs: two vantage points

Researcher point-of-view

- I usually manage data in Excel and it's terrible.
- I'm organizing GIS data and it's a nightmare.
- My advisor insists that we store 50,000 barcodes in a spreadsheet, not good.
- I want to use public data.
- I work with faculty at undergrad institutions and want to teach data practices, but need to learn it myself first.
- I'm interested industry and companies are asking for data analysis experience.
- I'm trying to reboot my lab's workflow to manage data and analysis sustainably.
- I'm re-entering data over and over again by hand – help!
- I have overwhelming amounts of data.
- tired of feeling out of my depth on computation.

BIO Centers IT Staff View

- man pages
- automated file-naming / renaming
- OpenRefine (http://openrefine.org)
- look at large text files
- concatenate files together
- computer directory structure
- write a text file not in Word
- the idea that you can parse a file
- the shell commands, piping, grepping
- which tool do we need to use to do an analysis
- introducing the idea of csv files (delimited) files
- you can formulate questions for your data and computer
- preparing data for analysis (such as formatting, trimming)
- awareness of data and metadata standards, and how to map data to them or make them compliant

Sentiments on data across the NSF BIO Centers
(BEACON, SESYNC, NESCent, iPlant, iDigBio, NIMBioS)

Carpentries Network Challenge!

- What are your (known) skills needs?

- Let's create a sustainable capacity development network

- Together, let's build a worldwide Carpentries community across collections

Data and Technology Leadership

- Museum Administration Support and Vision for Digitization
  - staffing
  - care, linking, and feeding of digital assets
- Leadership recognizing
  - new and changing roles
  - professional development needs
  - power of speaking with one voice

Data and Technology Leadership

Global Benchmark Challenge!

- Use ALA Digitization Maturity Model for Organization-level **Metrics**

  – organization-level digitization guidance

  – *compare across organizations*
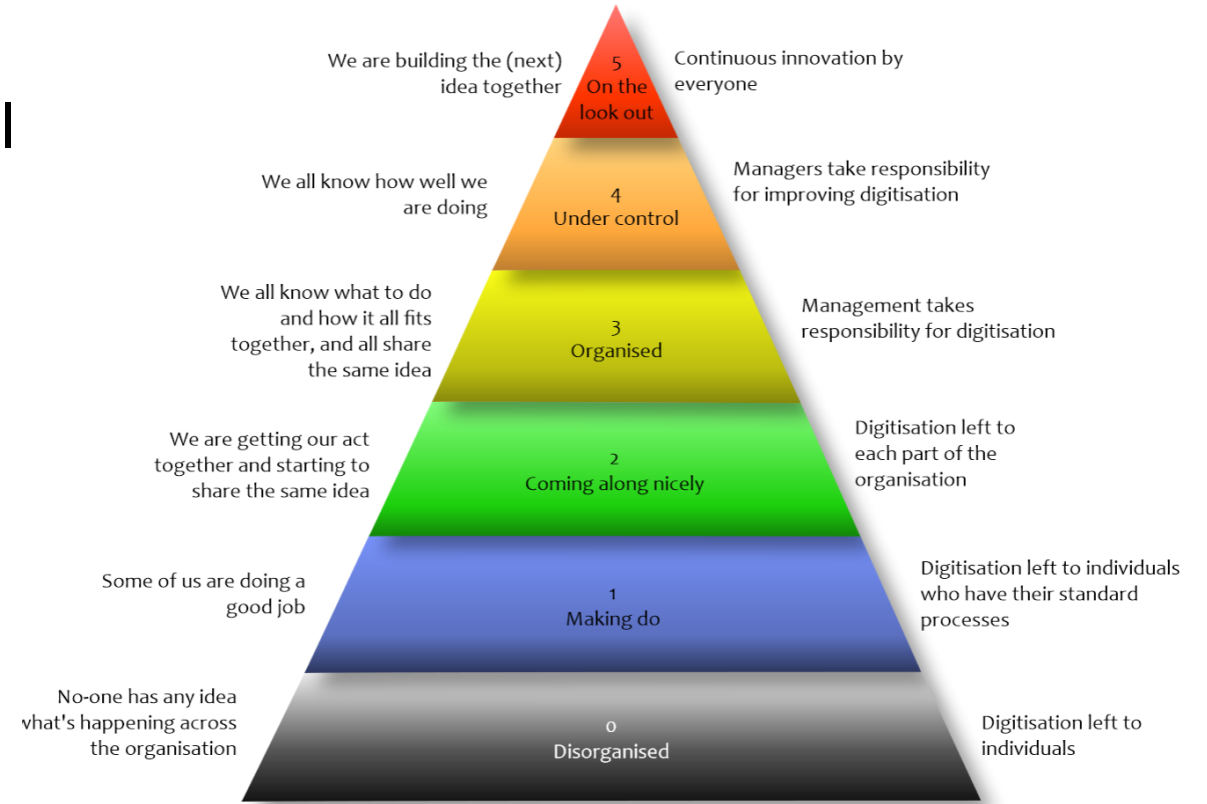
  – opportunity to evaluate collectively

We are building the (next) idea together — **5 On the look out** — Continuous innovation by everyone

We all know how well we are doing — **4 Under control** — Managers take responsibility for improving digitisation

We all know what to do and how it all fits together, and all share the same idea — **3 Organised** — Management takes responsibility for digitisation

We are getting our act together and starting to share the same idea — **2 Coming along nicely** — Digitisation left to each part of the organisation

Some of us are doing a good job — **1 Making do** — Digitisation left to individuals who have their standard processes

No-one has any idea what's happening across the organisation — **0 Disorganised** — Digitisation left to individuals

Figure 4 Digitisation maturity model

https://www.ala.org.au/who-we-are-3/digitisation-guidance/

Metrics and Benchmarking in Demand

- 90% of the respondents want to learn more about what metrics other institutions are using

- Does your collection have a strategic plan with clearly defined benchmarks and goals?

- Metrics Interest Group?
- Metrics Workshop?

#	Answer	%	Count
1	No	26.87%	18
2	Yes, our collection has a formal strategic plan.	17.91%	12
3	Our collection is encompassed into a larger strategic plan involving multiple collections.	13.43%	9
4	Our collection is subsumed under our institution's strategic plan.	34.33%	23
5	Other. Please explain.	7.46%	5
	Total	100%	67
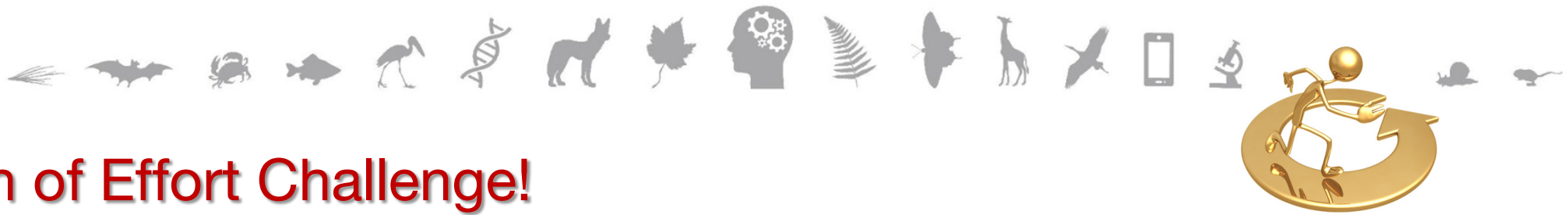
Duplication of Effort Challenge!
the collection's point-of-view

Deborah Paul
@idbdeb

1/ when #digitizing #biocollections we note much #duplication of effort. Everyone's got a #people table in their database. Hmm. How can we share more effectively? @AllianceforBio @iDigBio @arctosdb @taxonbytes

Does Biodiversity Informatics 💘 Wikidata?
Does Biodiversity Informatics love Wikidata?
🔗 data-blog.gbif.org

- taxon names
- people names
- localities
- publications
- trait data

Data Fitness, the aggregator point-of-view

- time (date)
- locality
- place
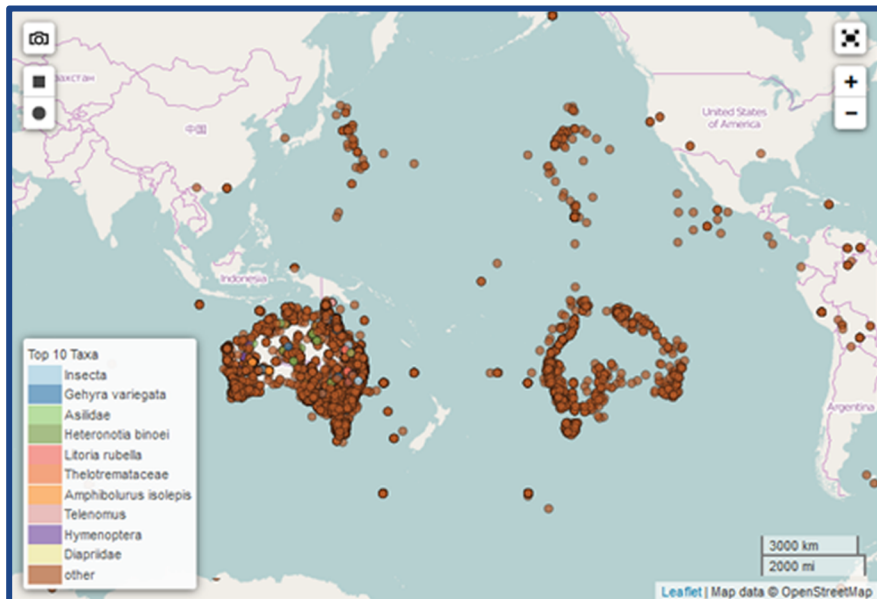
Hannah Frost
@feefifofannah
Following

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?

Data Fitness Standard Tests

- Tests / fitness measures developed
- Aggregators ALA, GBIF, iDigBio, and DiSSCo agreed to implement

Collection Management Software Developer Challenge!

Let's embed them in local collection management software

Figure 1.

Developing Standards for Improved Data Quality and for Selecting Fit for Use Biodiversity Data. ~ Biodiversity Information Science and Standards. Chapman, A.D., Belbin, L., Zermoglio, P.F., Wieczorek, J., Morris, P.J., Nicholls, M., Rees, E.Rr, Veiga, A.K., Thompson, A., Saraiva, A.M., James, S.A., Gendreau, C., Benson, A., Schigel, D. (2020). 4: e50889. https://doi.org/10.3897/biss.4.50889

https://github.com/tdwg/bdq

Take the Tweezer Challenge!
Communication and Cognitive Surplus
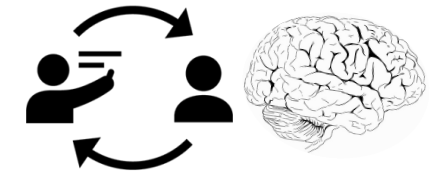
- Sharing Discoveries and Lessons Learned
  - dissemination
  - documentation and tracking

Summarizing some key needs for Human and Infrastructure Evolution for 21st-Century Collections

- Carpentries Network Challenge!

- Global Benchmark Challenge!

- Got metrics?!

- Duplication of Effort Challenge!

- Collection Management Software Developer Challenge!

- Take the Tweezer Challenge!

- *What's your sphere of influence?*

Thank you

Erica Krimmel, Dirk Neumann, Megan McCuller,
Gabriela Hogue, Laurence Livermore, Shelley James,
Christopher Marshall #SPNHC2020 #NATHIST2020

facebook.com/iDigBio

twitter.com/iDigBio

vimeo.com/idigbio

idigbio.org/rss-feed.xml

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

www.idigbio.org

@idbdeb