**Section 1: Data Survey Scope**

**Problem Statement** Data transcription, inventory, and workflows result in many local procedures for the handling of biodiversity collections that have not been standardized. This leads to messy and inconsistent data with biodiversity collections. Creating standardized frameworks that are widely implemented across the biodiversity data collections would improve data quality and increase efficiency. This research will focus on identifying challenges and possible solutions to this issue. These include technology (e.g., software improvements), wider implementation of different tools, such as community adoption and promotion of metadata schemas and controlled vocabularies (e.g., through workshops or other training), or other methods.

**Question:** What developmental frameworks are most efficient in cleaning and refining biodiversity data?

* What functions and technology are helpful in cleaning biodiversity data?
* What are efficient ways to enforce data consistency and quality standards across biodiversity collections?

**Research Design:** Literature review will help assess the frameworks that have been effective to maintain biodiversity data quality across different institutions. A short survey will be sent out to different Listservs within the iDigBio page. Survey respondents have three weeks to complete the short questionnaire. If willing, survey respondents will be followed up on with three in-depth survey interviews based on a selection of biodiversity aggregator and museum staff that will pinpoint the more intricate processes that occur in formulating these frameworks, both from a technology and staff development process. These assessments combined will formulate the future trajectory for museum institutions in assembling biodiversity data. More specifically, the research design will highlight the best approaches to improve data quality. These include:

1 Improved Software

* What software is used in cleaning biodiversity data?
* How satisfied are users with this software?
* What improvements, if any, would be useful?

2 Use of Existing Tools

* What metadata schemas are currently used or preferred?
* What controlled vocabularies are currently used?

3 Training Resources

* What training resources or workshops currently exist?
* How can training be improved or more widely offered and used?
* Are there workflows being used as training resources for staff/volunteers?

**Target Audience:** The following include the Digitization Listserv from iDigBio and other museum professionals. <https://www.idigbio.org/wiki/index.php/IDigBio_Listservs> The target audience includes natural history museum professionals that seek to improve and refine their workflows and methodologies for digitization and museum workflow initiatives.

**Note: By completing this survey or questionnaire, you are consenting to be in this research study. Your participation is voluntary and you can stop at any time.**

**Section 2: Survey Questions**

**Please respond to the following questions as applicably as possible.**

**Institution Type:** Collection HolderData Contributor Data Aggregator

Other (Please Explain):

**Email Address and Respondent Name (if willing to clarify and/or respond to further inquiries by an interview via Zoom, phone, or further email responses):**

**1** What software are you currently using to clean and edit biodiversity data?

OpenRefine

Easy Data Transform

Trifacta

Data Wrangler

ZoomInfo OperationsOS

DemandTools

DataCleaner

Melissa Data Cleansing

Other (Please Explain:)

**1.1** On a scale of 1 to 5, how satisfied are you with using your specific software?

Not Satisfied Extremely Satisfied

1 2 3 4 5

**1.2** What improvements do you desire the most out of your current software system based on what you are using? If you have no desire/ideas for improvements, choose N/A

Better reporting systems (e.g., integrated software to reduce redundant data entry)

Error catching mechanisms (programs to correct hand made errors)

Improved system layout/format (easier usability and/or personalized workflows)

Improved data sharing mechanisms (sharing to online or other institutions easier)

Other (please explain):

N/A

**2** What metadata schemas are currently being used at the institution (check all that apply)?

Dublin Core (DC)

Metadata Object Description Schema (MODS)

Darwin Core

Integrated Taxonomic Information System (ITIS)

Directory Interchange Format (DIF)

Ecological Metadata Language (EML)

Geospatial Interoperability Framework (GIF)

Biological Collection Access Service (BioCASe/ABCD)

Other (please explain):

**2.1** How useful do you perceive the metadata schemas being (can be based on one or multiple schemas being used)?

Not Helpful Extremely Helpful

1 2 3 4 5

**2.2** If below a 4, what are the prevailing challenges with these schemas? (2-3 sentences)

**3** What specific controlled vocabularies have you been using to enforce data standards? (Check all that apply)

Subject heading lists

Controlled lists

Synonym ring lists

Authority files

Taxonomies

Alphanumeric classifications

Thesauri

Ontologies

Folksonomies

Other (please explain):

**3.1** Explain why these forms of controlled vocabularies have been the most helpful (2-3 sentences)?

**4** What quality assurance measures (using the technology/software) have been put in place to ensure data consistency across the museum staff and/or volunteers? (Check all that apply)

Accuracy

Completeness

Consistency

Validity

Uniqueness (not duplications or overlapping of values)

Timeliness (e.g., data available when required)

Other (Please Explain):

**4.1** Which quality assurances measures do you believe are the most important for ensuring data standards and why? (2-3 sentences)

**5** Does your institution use volunteers to fulfill any of this type of work? If so, are there specific challenges with enforcing data standards and quality assurance measures with volunteers? (Check all that apply)

Enforcing data standards on volunteers (quality assurance checks)

Lack of motivation/drive from volunteers

Lack of communication between both staff and Volunteers

Other (please explain):

Institution does not utilize volunteers

**5.1** How (or would) have you addressed these challenges? (If volunteers are used) (2-3 sentences)

**6** Which of the following have been the most helpful in reassessing and improving data management standards? (Check all that apply)

Conferences

Workshops

Meetings (staff only)

Building (and refining) workflow templates

Stakeholder meetings (e.g., staff discussing with financial members and/or volunteers)

Training sessions

Other (please explain):

**6.1** Explain why these methods you selected have been the most effective (2-3 sentences)

**7** Have you been using (or are considering using) any of the following to improve these resource/training standards? (Check all that apply or choose N/A)

Asking for or researching different institutions for advice

Employee one on one meetings or transparency improvement measures

Incorporating surveys post meetings for feedback

Incentivize or reward members for work/accomplishments during these meetings/training sessions

Other (Please Explain):

N/A

**7.1** If you have **not** selected N/A, name two to three benefits these improvement measures have given to the data standard work environment?