Designing a Digitisation Framework

Gil Nelson
Institute for Digital Information and Scientific Communication
Integrated Digitized Biocollections
Florida State University

Bristol, England
9 March 2018

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (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. All images used with permission or are free from copyright.

ZooKeys 209: 19–45 (2012)
doi: 10.3897/zookeys.209.3135
www.zookeys.org

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

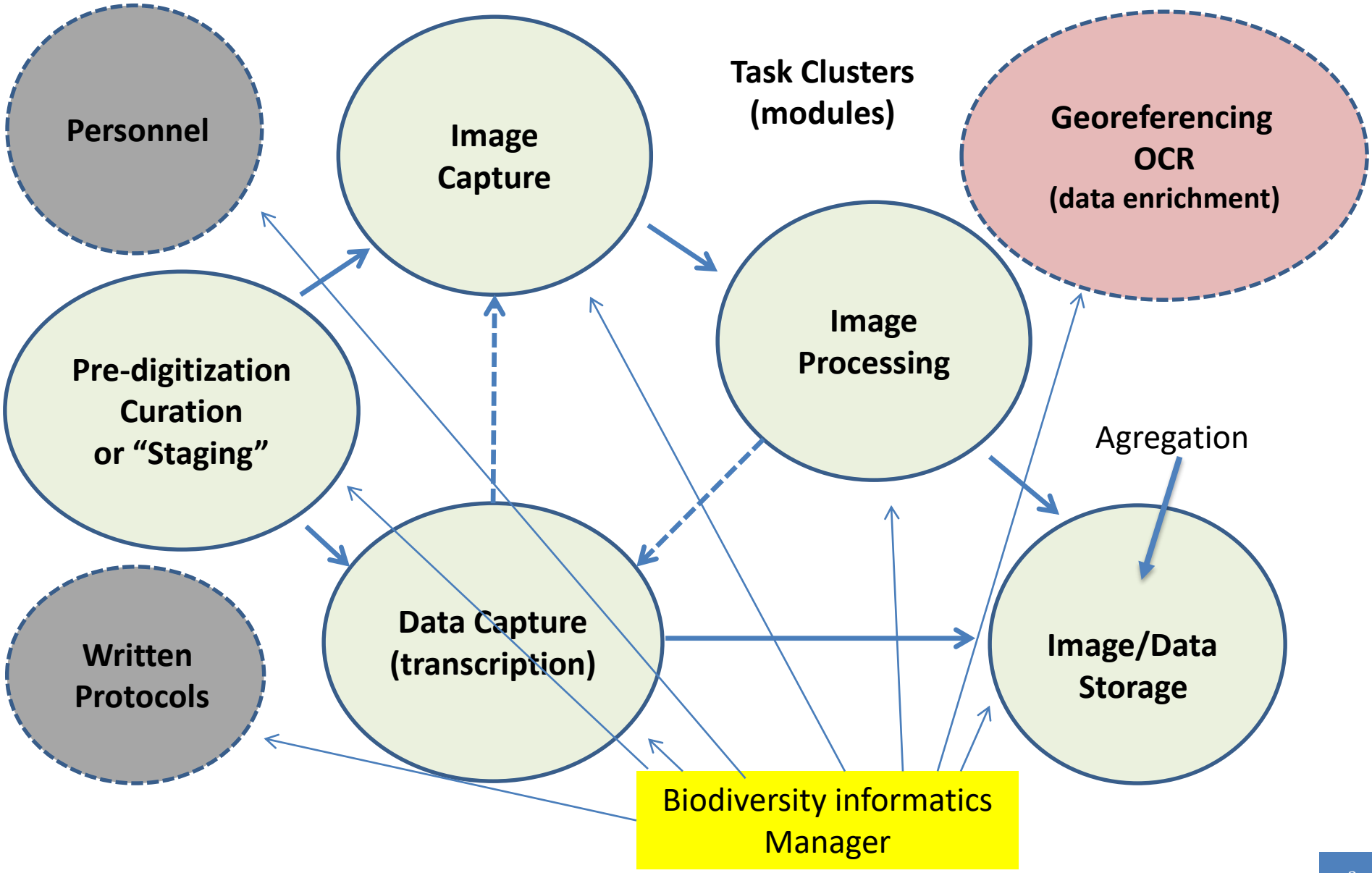
Five task clusters that enable efficient and effective digitization of biological collections

Gil Nelson[1], Deborah Paul[1], Gregory Riccardi[1], Austin R. Mast[2]

1 *Institute for Digital Information, Florida State University, Tallahassee, FL 32306-2100, United States* **2** *Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, United States*

Corresponding author: *Gil Nelson* (gnelson@bio.fsu.edu)

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (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. All images used with permission or are free from copyright.


Personnel
Image Capture
Task Clusters (modules)
Georeferencing OCR (data enrichment)
Pre-digitization Curation or "Staging"
Image Processing
Agregation
Written Protocols
Data Capture (transcription)
Image/Data Storage
Biodiversity informatics Manager

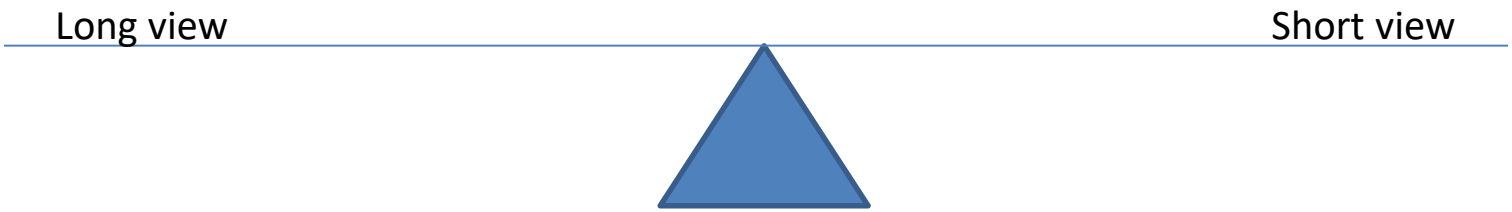
The Biodiversity Informatics Function

Essential to successful digitisation programmes
Better with broad institutional support
Oriented toward service to staff
Better if based on strong knowledge of the biological sciences
Strong technology skill set
Often “home-grown”
Timely assistance and decision-making
Open to new and/or enhanced protocols and techniques
Continuously seek ways to increase technological efficiency and productivity
Map and mobilize data to aggregators

Tracks to Digitisation Implementation

• **Taking the inside track [short view]** is often based on stretching the institution's resources. Decisions are made to maximize resources available for user-initiated digitisation by using solid baseline practices. The primary focus on the inside track is to get the job done quickly and to fill the user's request.
• **Taking the middle track** has the widest range of options, standards, and results. This is the most flexible of the tracks, where decisions often fall in gray areas.
• **Taking the outside track [long view]** focuses on the collections themselves. While users may initiate digitisation, it is undertaken to deliver materials to a greater public. These decisions may lead to comprehensive digitisation, such as an entire book, series, or collection. The goal is to create maximum access to special collections, using preservation and archival standards. This track usually involves a level of thought and planning that is more in-depth than the fulfillment of day-to-day digitisation requests.

Balancing the long view with the short view: The local decision

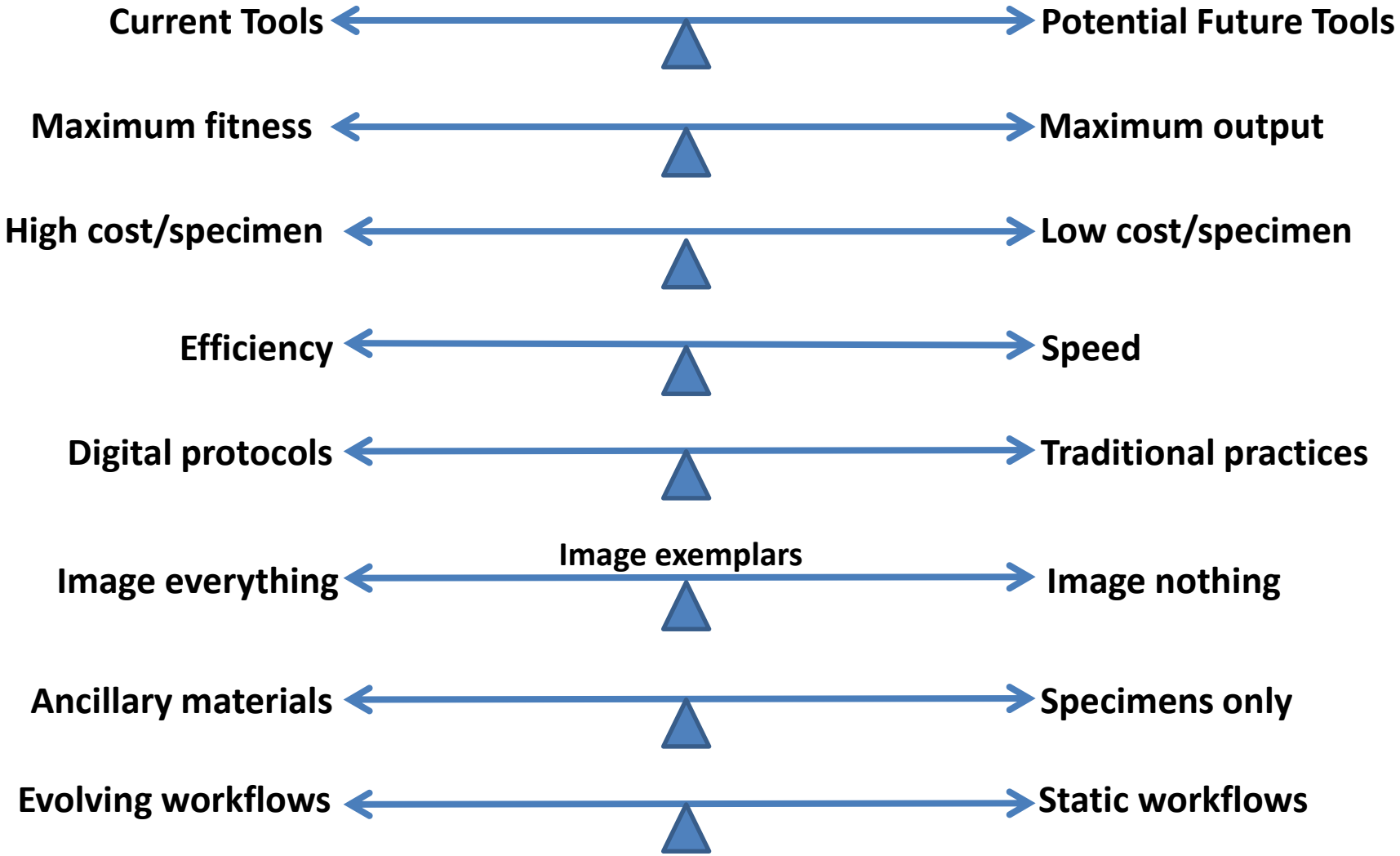


How does an institution develop doable, effective, and sustainable strategies for balancing long term goals with short term constraints, including commitments to future enhancements?

Pressures mitigating the long view

So much data, so little time.
Collections are not getting smaller (proactive vs. legacy).
Funders often have high output/low cost expectations.
We only have 3 years to get this done (sustainable models?).
How to choose: all of our data and all of our specimens are important.
Let's just use the images!
We'll do the minimum now and enhance it later (inside track).

Decision Continua



Long view

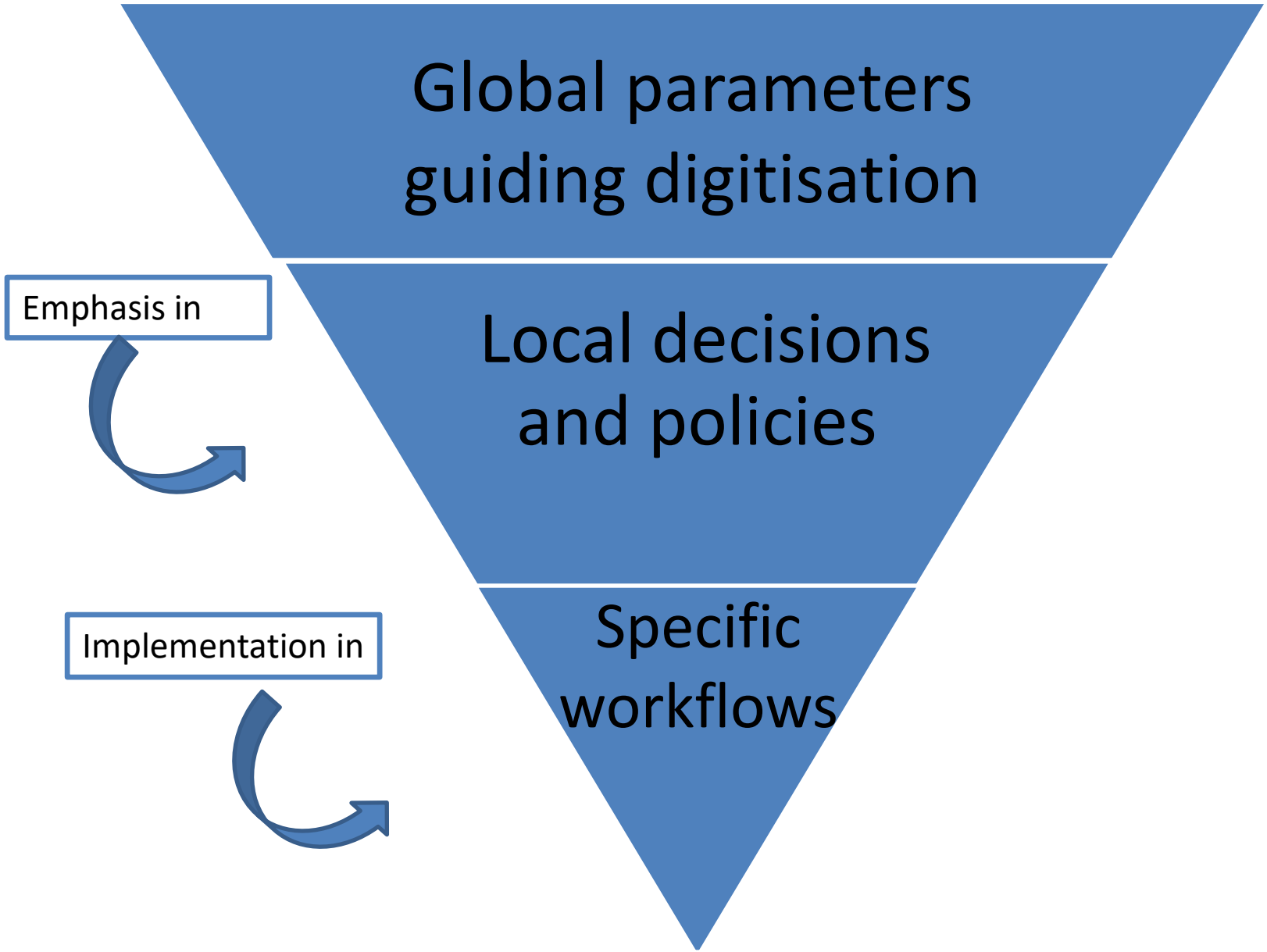
Short View

Facilitators

- Emphasize fitness for use
- Robust data
- Data validation/cleaning
- Integrated quality control
- Integrated georeferencing
- Intensive curation
- Record historical annotations
- Staff specialization
- Small collection
- Emphasize images
- High quality images

Facilitators

- Emphasize output
- Spartan data
- Defer validation/cleaning
- Minimum quality control
- Deferred georeferencing
- Deferred or cursory curation
- Record current determination
- Staff generalization
- Large collection
- Emphasize data
- Low quality images


Global parameters guiding digitisation
Local decisions and policies
Specific workflows
Emphasis in
Implementation in

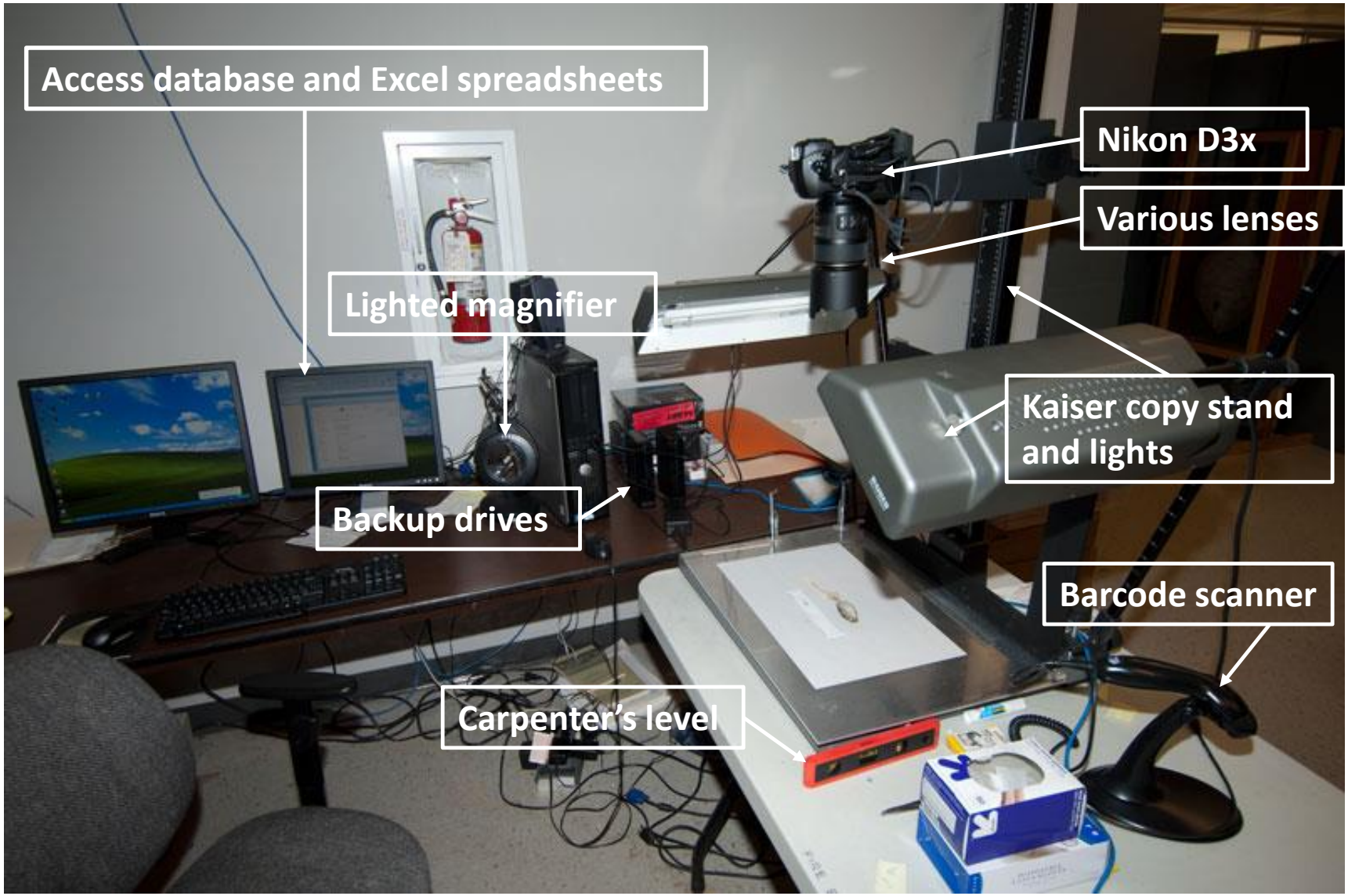
Pre-digitisation Preparation/Staging Considerations

Prepare an institution/collection management and digitisation policy manual to include provisons for:

- **Inspecting for and repairing specimen damage and evaluating collection health,**
- **Re-pinning or remounting specimens, replenishing or replacing preservatives in containers, other conservation,**
- **Treating specimens for pests (IPM),**
- **Attach a locally unique identifier (often a 1- or 2-D barcode) to a specimen, container, or cabinet (unit of digitisation),**
- **Discover important but previously unknown, lost, or dislocated holdings (e.g. those owned by other institutions or the government agencies),**
- **Update nomenclature and taxonomic interpretations on specimens and in database,**
- **Reorganize the contents of cabinets, cases, trays, and containers, especially when these are the units of digitization,**
- **Vet type specimens,**
- **Select exemplars for digitisation, when that approach is appropriate,**
- **Include researchers and users in the mix.**

Imaging

Unequalled opportunity
for completeness and detail



Imaging station

Kaiser Copy Stand RSX with RTX Arm

Nikon D810
36.3 megapixel
Full frame
Live view

Canon EOS 5D Mark IV
Full frame
31.7 megapixel
Live view

Canon T6
17.9 megapixel
Crop Frame
Live view

Photo eBox


The Box
One Step Photo Studio System
Digital Imaging System
MK
PhotoeBox

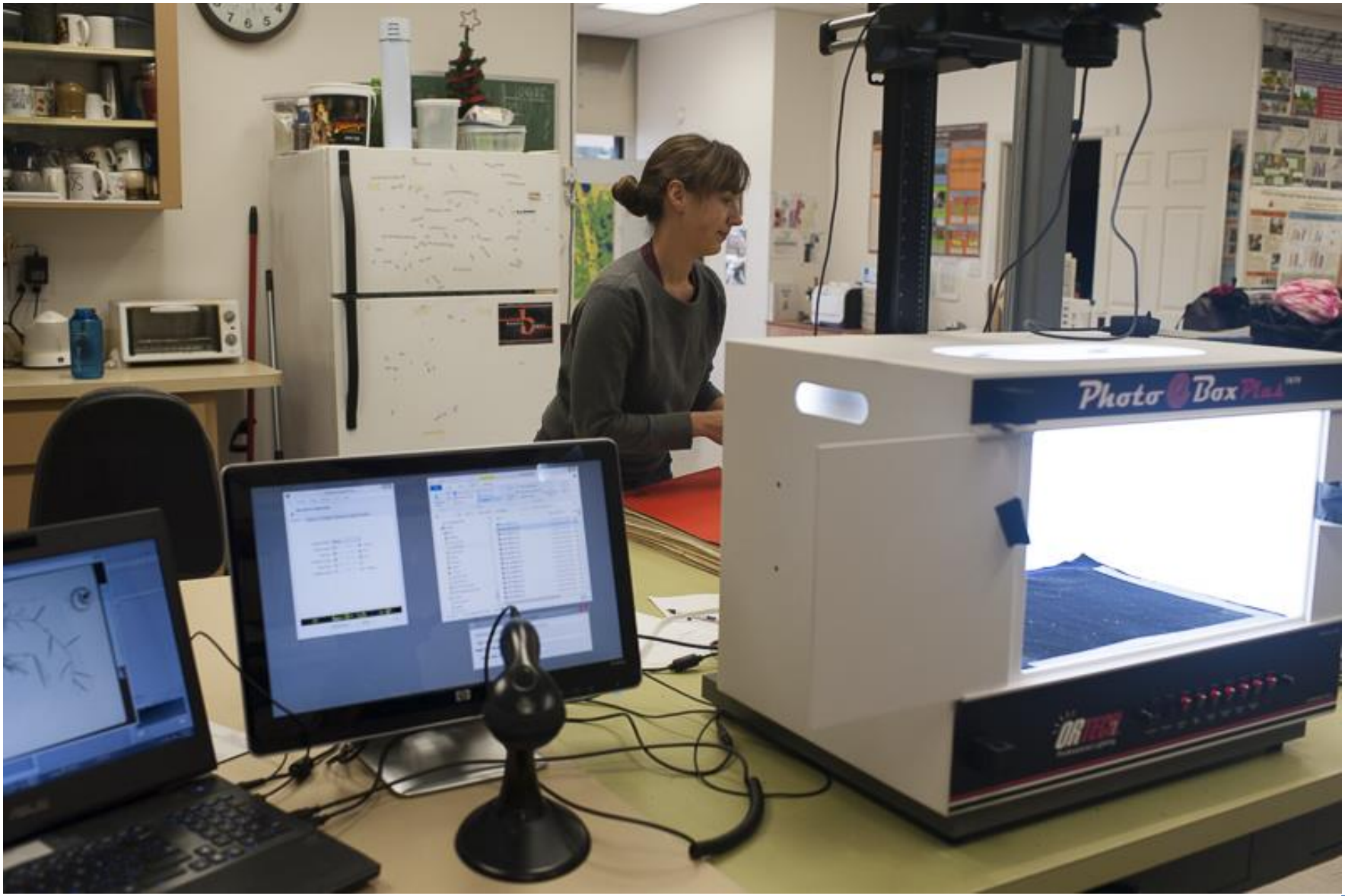
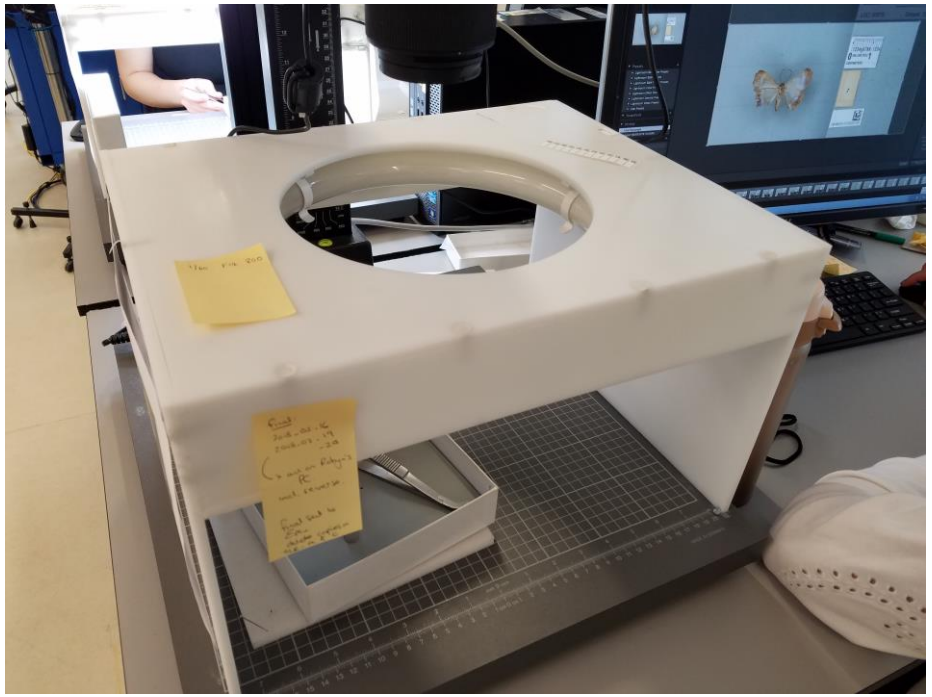
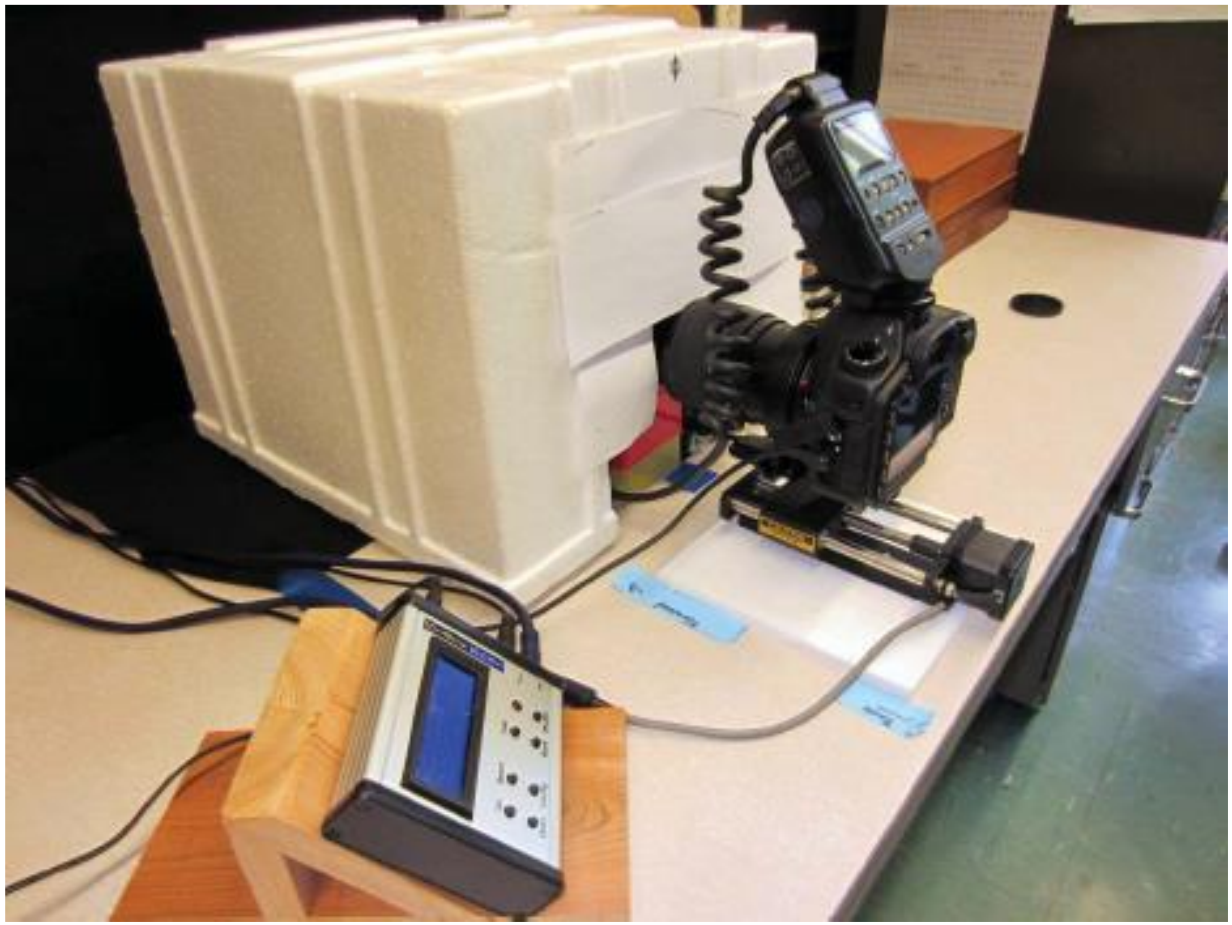


Photo Box Plus

Ortery box



Natural History Museum



Sam Droege, USGS

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

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (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. All images used with permission or are free from copyright.