Tools and workflows for building a collaborative, library-based data management service

Stacy Konkiel
Science Data Management Librarian
IU Libraries-Bloomington

Eric Snajdr
Science Librarian
University Library-IUPUI

Brianna Marshall
Science Data Curation Assistant
IU Libraries-Bloomington

INDIANA UNIVERSITY

Data Management & Curation as "Grand Challenge"

Library roles in data curation

- Historically, data preservation the domain of scientists & archives (Witt, 2008)
- In past 20 years, libraries have taken an increasingly active role in data management & curation as:

  Educators (Johnston *et al*, 2012; Shorish, 2012; Carlson *et al*, 2011)

  Evangelists (Adamus *et al*, 2013; Ward *et al*, 2011)

  Connectors (Steinhart *et al*, 2008)

  Curators & Preservationists (Ball, 2010; Kuipers & Koeven, 2009; Witt, 2008)

  Data quality hubs (Giarlo, 2013)

IU Bloomington
Data Management Service

Coordinated by Science Data Management Librarian

- **Education & Gatekeeper**
  Workshops, Documentation, DMP Consults
- **Evangelism**
  OA Week, Workshops
- **Connector**
  DMP Consults
- **Preservationist**
  DMP Consults, IR Access

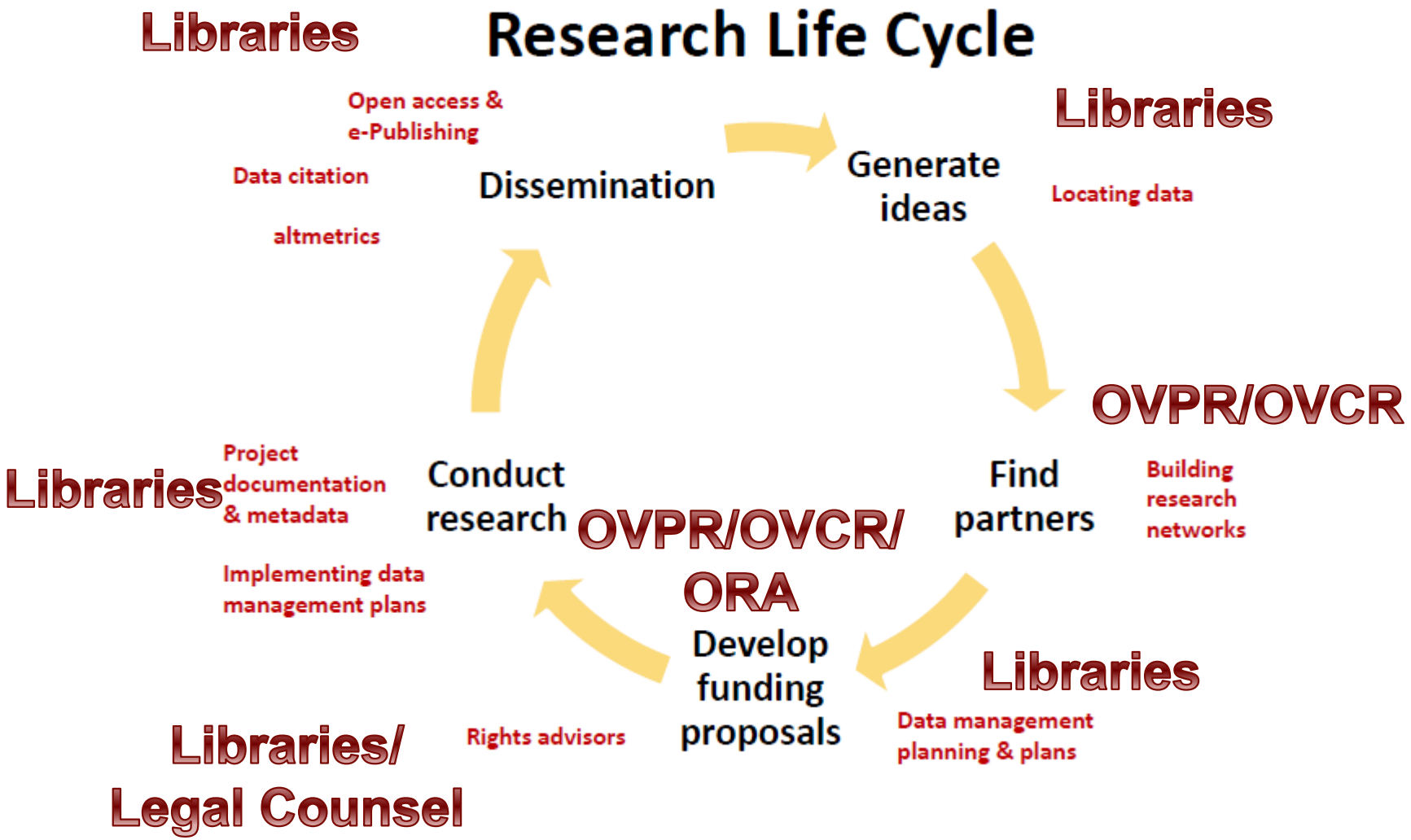
Not yet a **Curator**,
likely never a **Data Quality Hub**

IU Bloomington's approach to data curation

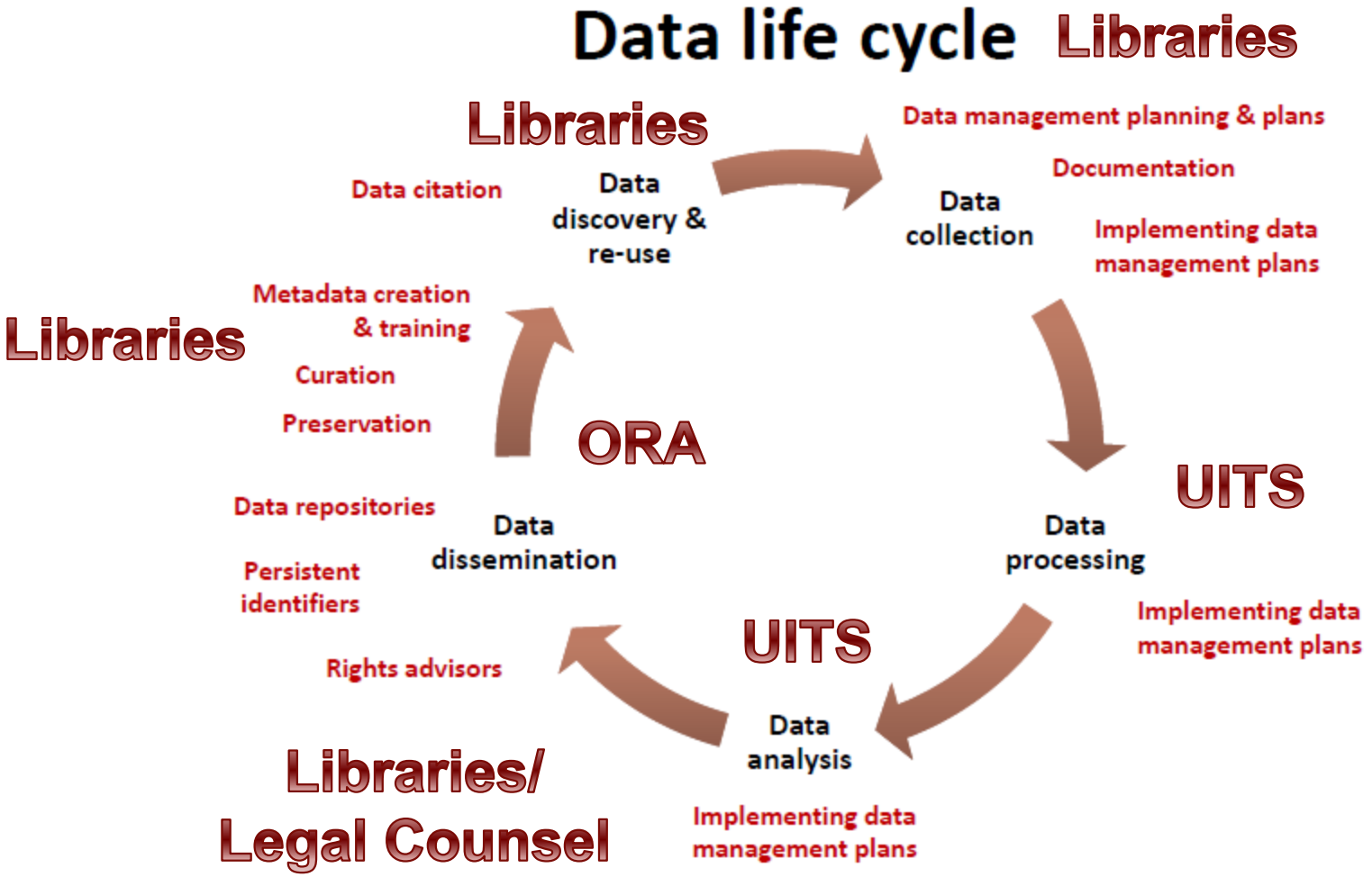
A "HIGH-TOUCH" DATA CURATION SERVICE PILOT

Indiana University-Bloomington

- Carnegie Research classified as “R1”
- Large, suburban, Midwestern school
- Grants 100+ different doctoral degrees
- 6 science-related branch libraries
- 500+ full-time faculty across…
  - 35+ institutes
  - 7 schools
  - 28 STM departments



(Coates & Konkiel, 2013)



(Coates & Konkiel, 2013)

Shared Cyberinfrastructure

box

Dropbox

Local Data

Dept. Server

Lab Server

Local Analysis & Short-Term Storage

IUScholarWorks Repository

IUPUIScholarWorks Repository

IUPUIDataWorks Repository

Scholarly Data Archive/ MDSS

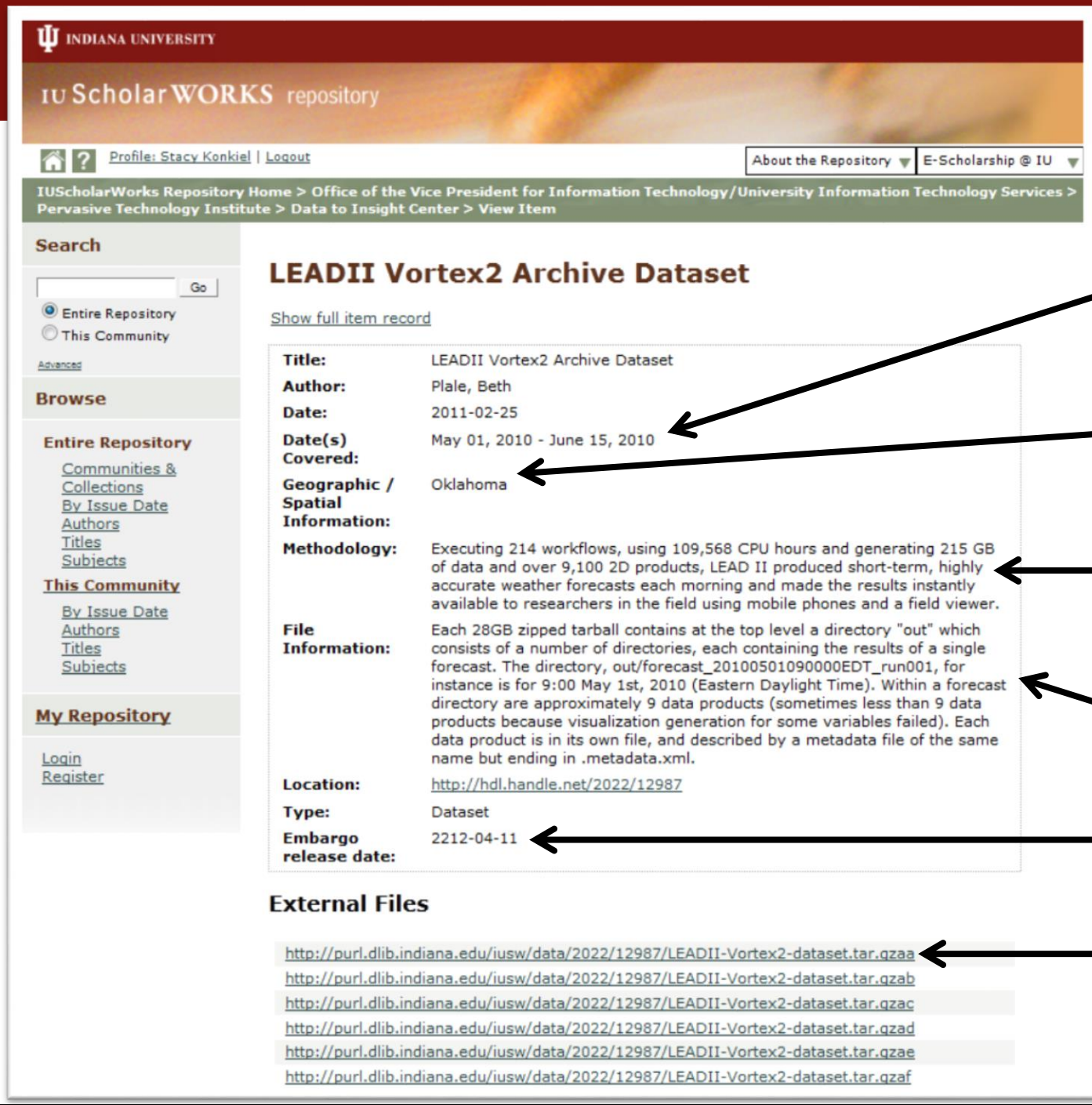
Data Analysis

Big Red II

Data Capacitor II

Analysis & Short-Term Storage

Research File System



Coverage dates

Geographic info

Methodology

File info

Embargo info

SDA link(s)

Data Curation Service Pilot

Basic

- Set IR collections and admin rights for self-upload
- “Data enable” collections
- Consult with depositor on metadata, licensing
- BIG DATA: Instruct depositor on how to push data from personal to IUSW SDA account
- Serve as a liaison between depositor and library and UITS staff

Advanced

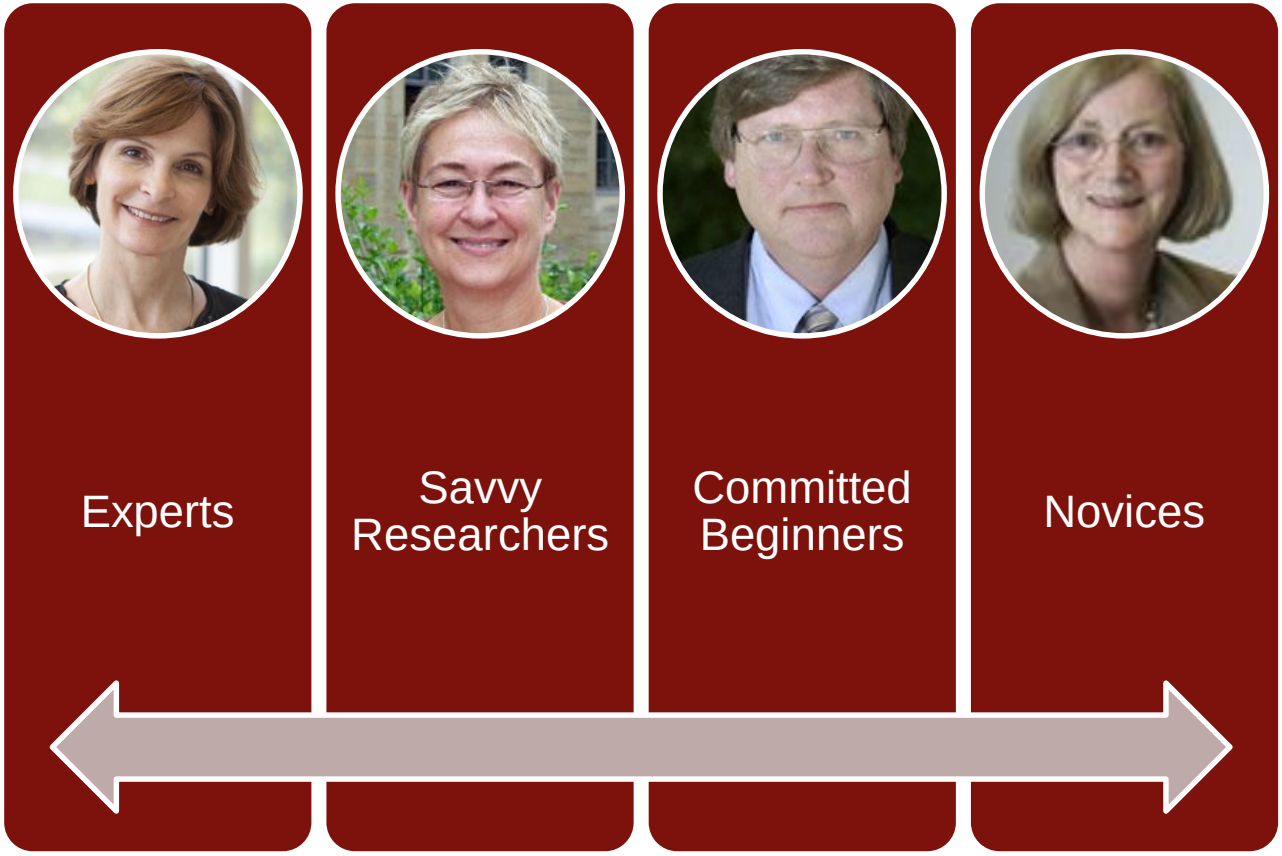
- *All of the Basic Services, plus...*
- Specialized workflows for ongoing fixed data management
- Upload files and create general and disciplinary metadata on behalf of user
- Create reports documenting their work, for researchers to include in funding agency DMP compliance documentation (as needed)

USE CASES

Spectrums of researchers'…

- Ability
- Interests
- Openness
- Support

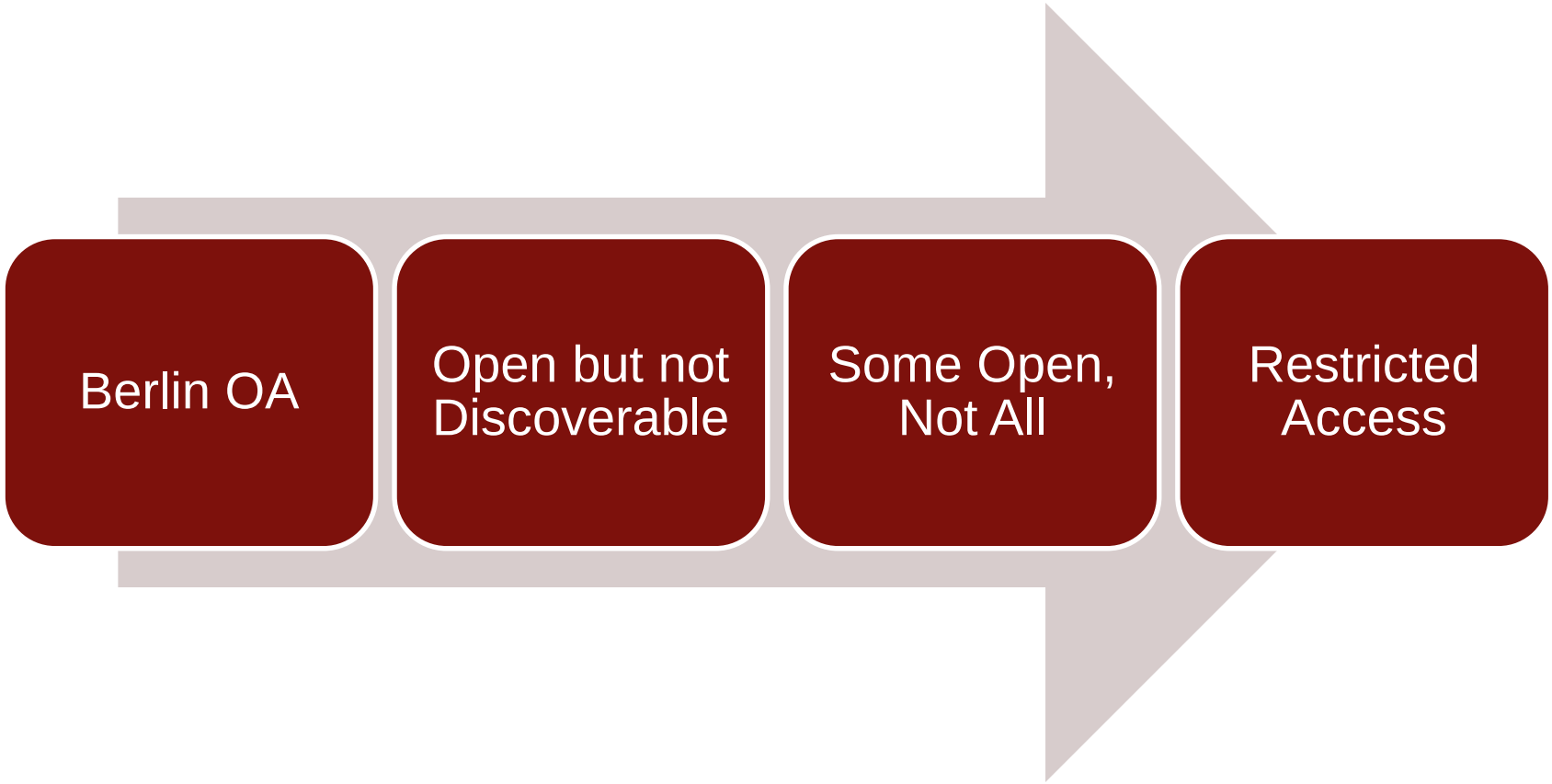
A spectrum of ability



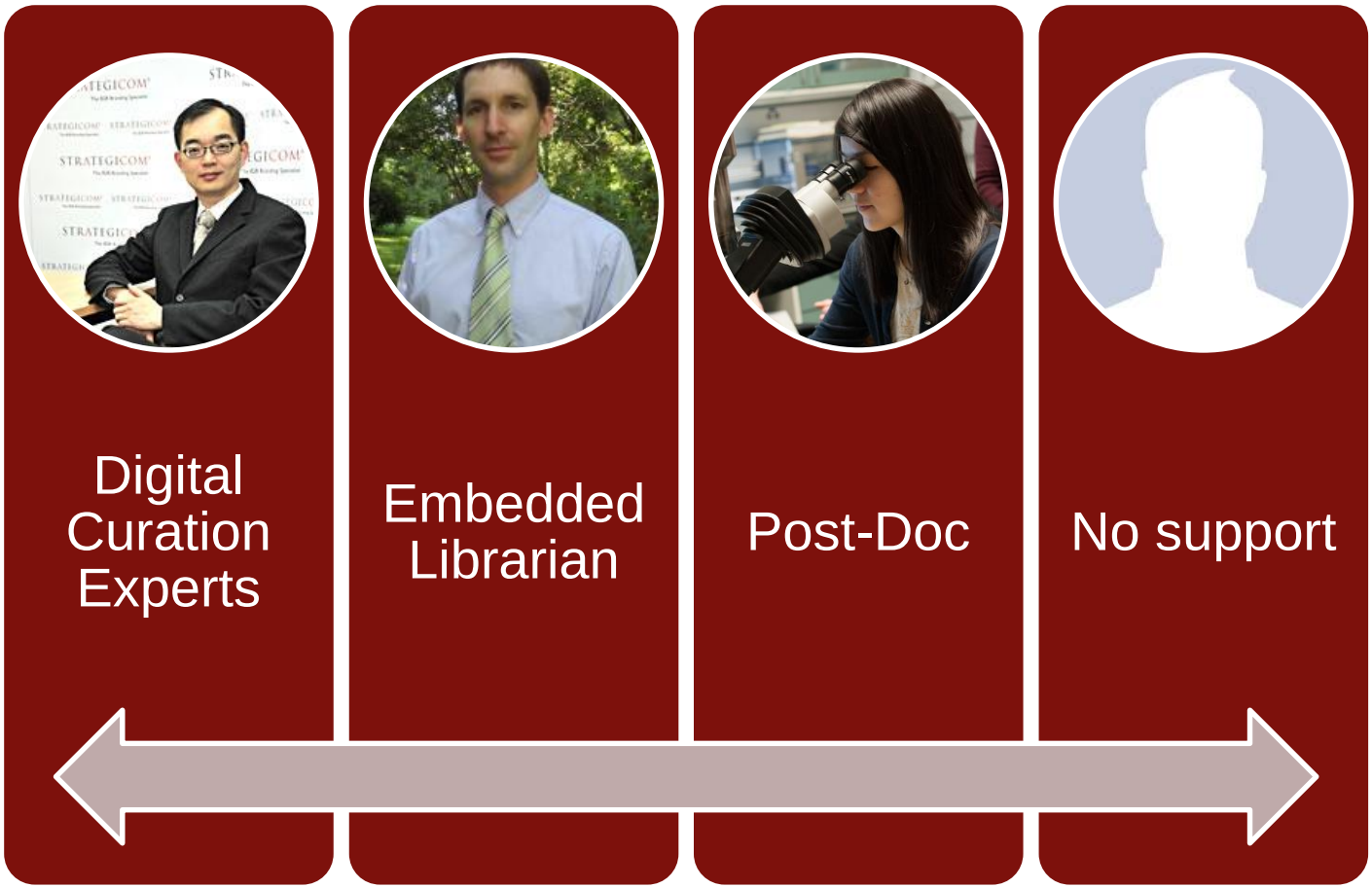
A variety of interests

- Grant Compliance
- Legacy
- Open Data
- Discoverability

A spectrum of “openness”



A spectrum of support



Experts

- Informatics & Atmospheric Science
- NSF & NOAA Funded
- Team interested in Grant compliance, Open Data
- Disciplinary metadata
- Big Data, mixed formats

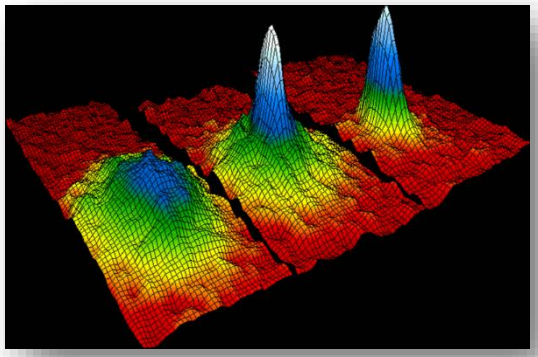
Savvy Researchers

- Evolutionary Biology
- NSF, some NIH Funded

- PI interested in Grant compliance, Discoverability, and Legacy
- 25+ years of data includes specimen information stored in relational database, photos, video, field notes
- Homebrewed metadata schema
- Embedded Librarian affiliated with team

Committed Beginners

- Condensed Matter Physics (Experimental)
- NSF & DOE Funded
- Commitment to general understandability
- Interested in Grant compliance only
- Data is small, in a mix of proprietary and open formats
- No metadata, little documentation

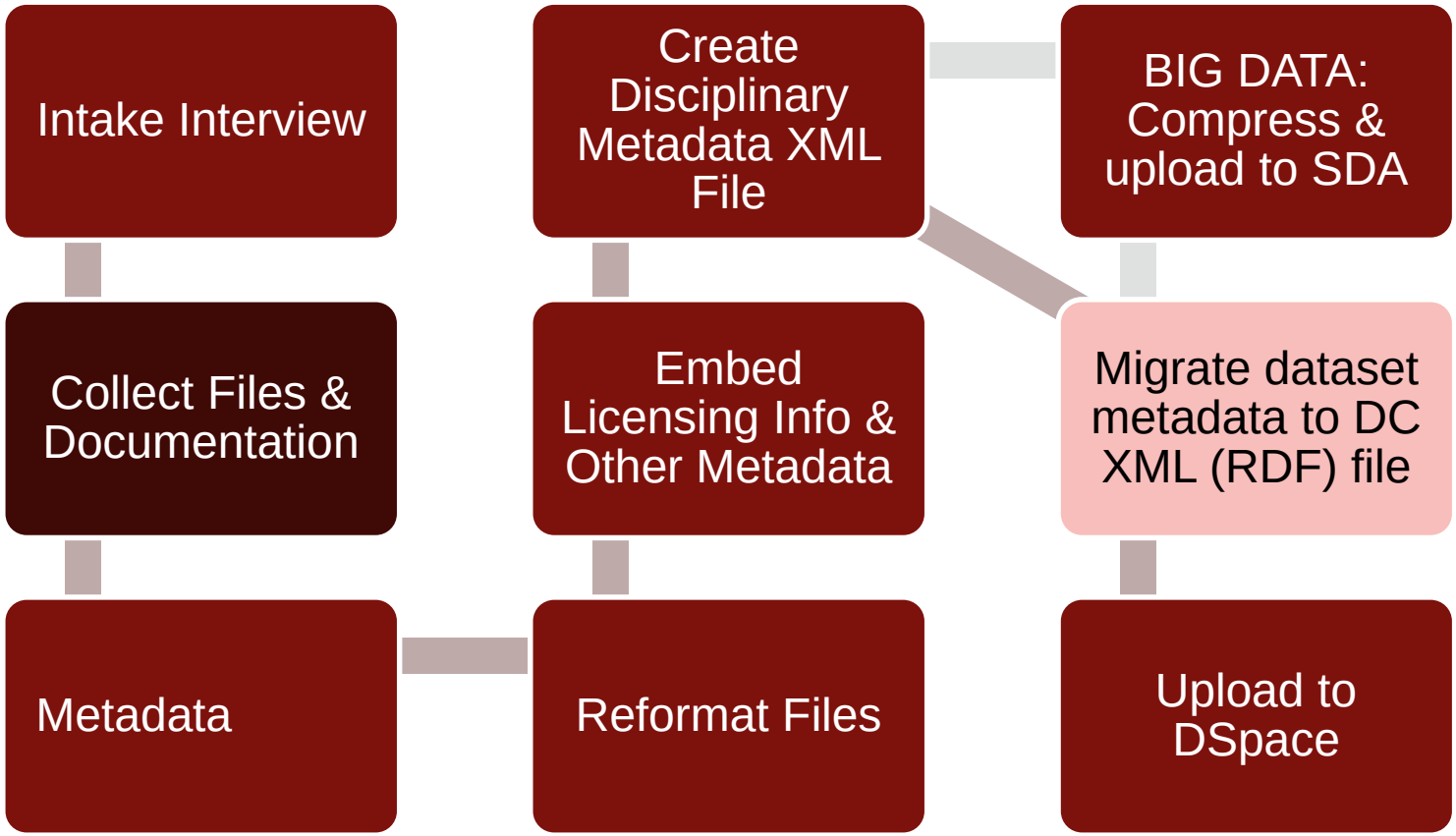


Novices

- Public Health (Emerita)
- Interested in Legacy, Discoverability, Open Data
- Little external funding
- Mix of obsolete, proprietary formats and open formats; 50% analog & 50% digital
- No metadata, little documentation

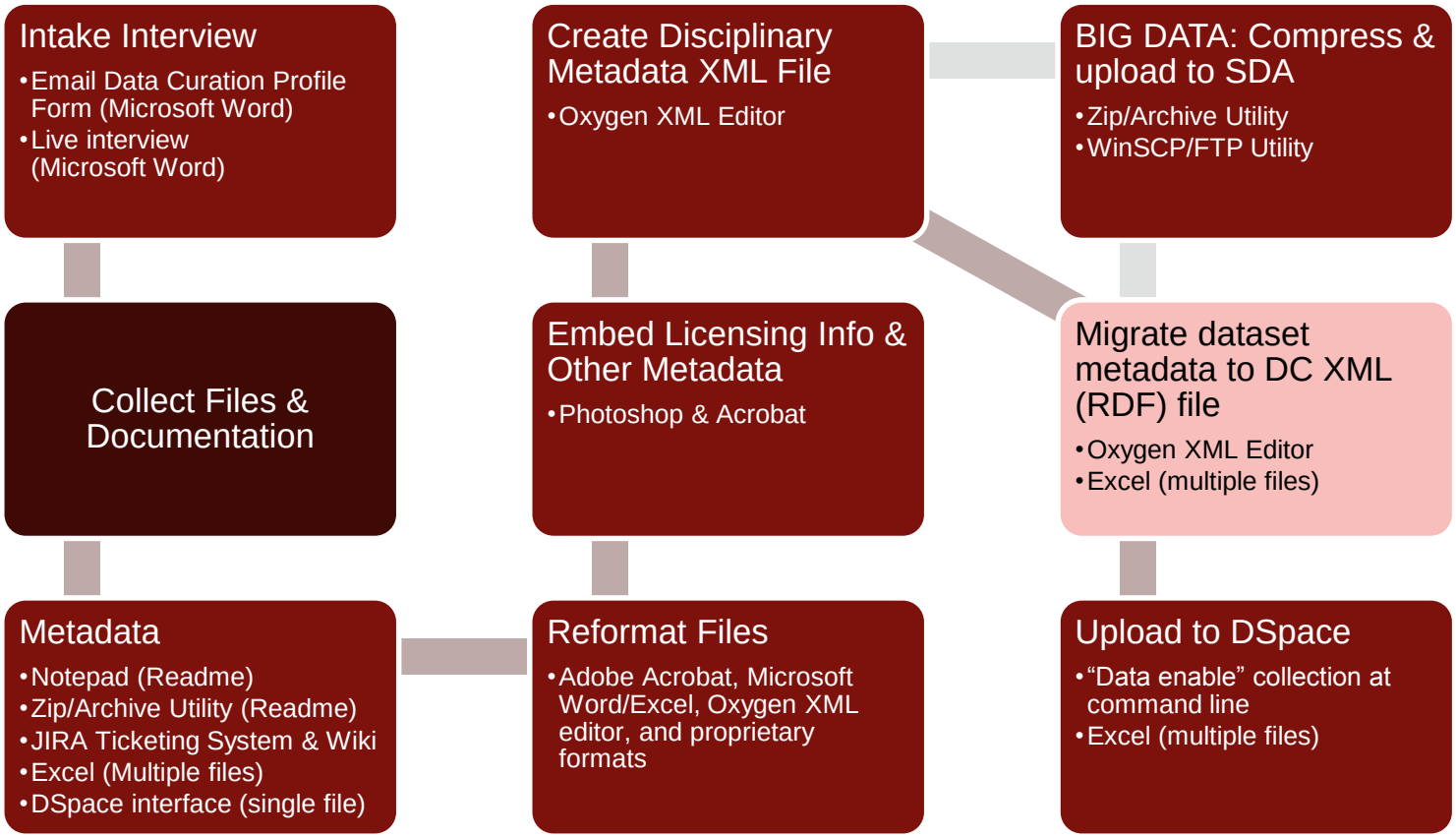
DATA CURATION TECHNOLOGIES, WORKFLOWS & PARTNERSHIPS

General Workflow



Legend: *Pop-out* *Not yet deployed*

General Technologies



General Technologies

Collect Files & Documentation

- Digitization
  - Flatbed scanner, high-speed photocopier/scanner, Adobe Acrobat, Box
- Reorganization
  - Box, Dropbox, network servers
- Description
  - Shared Microsoft Excel spreadsheet
- Weeding
  - Box, Dropbox
- Checksums
  - HSI on SDA

Key Partnerships & Roles

- Researcher/Research Team/PI
- Technology Experts
- Subject Expert
- Licensing
- Metadata
- Hourly Worker(s)

View from a Subject Expert

For the researcher, data management requires

- Time
- Basic understanding of the processes involved
- Expertise (technology and other)

A Subject Expert can help to ease burden on Researcher and Data Librarian

View from a Hourly Student Worker

- Access to data curation readings and resources before jumping into project helped
- “Sink or swim” experience with case study
- Tools like IU Box and JIRA greatly improved collaboration and communication
- Face-to-face meetings are important, too!

Limitations

- Version control
- Permissions for ease of collaboration
- File size limitations
- Differences in awareness/knowledge of tools
- Faculty resistance to new technology
- Newer technologies (ELNs, sci. workflow software) not yet integrated into this model

TAKEAWAYS FOR YOUR INSTITUTION

Staffing & Expertise

- Data librarians nice but not required
- Researcher buy-in absolutely required
- Leveraged expertise wherever possible
- (Good) Hourly student workers key

University Cyberinfrastructure

- On-campus resources for data storage, data preservation, and data access necessary
  - University-wide
  - Campus-wide
  - Departmentally-supported
- Third-party options (non-sensitive data)
  - Cloud storage & backup
  - Repositories (subject & ‘non-denominational’

Scope out the data curation needs & resources

- Data Curation Profiles a great resource
- Must have faculty investment in:
  - Keeping workflow consistent
  - Applying metadata standards & keeping good documentation
  - Using naming conventions
- What is the realistic level of investment your campus library and IT unit can offer?

Thank you!

Stacy: skonkiel@indiana.edu

Eric: esnajdr@iupui.edu

Brianna: bhmarsha@indiana.edu

Download this presentation:

[slideshare url] // [iuscholarworks handle]

Most references and photo sources embedded as links