



## Where we are now...

## “Bridging the Gaps of Providing and Scaling Research Data Curation Services”

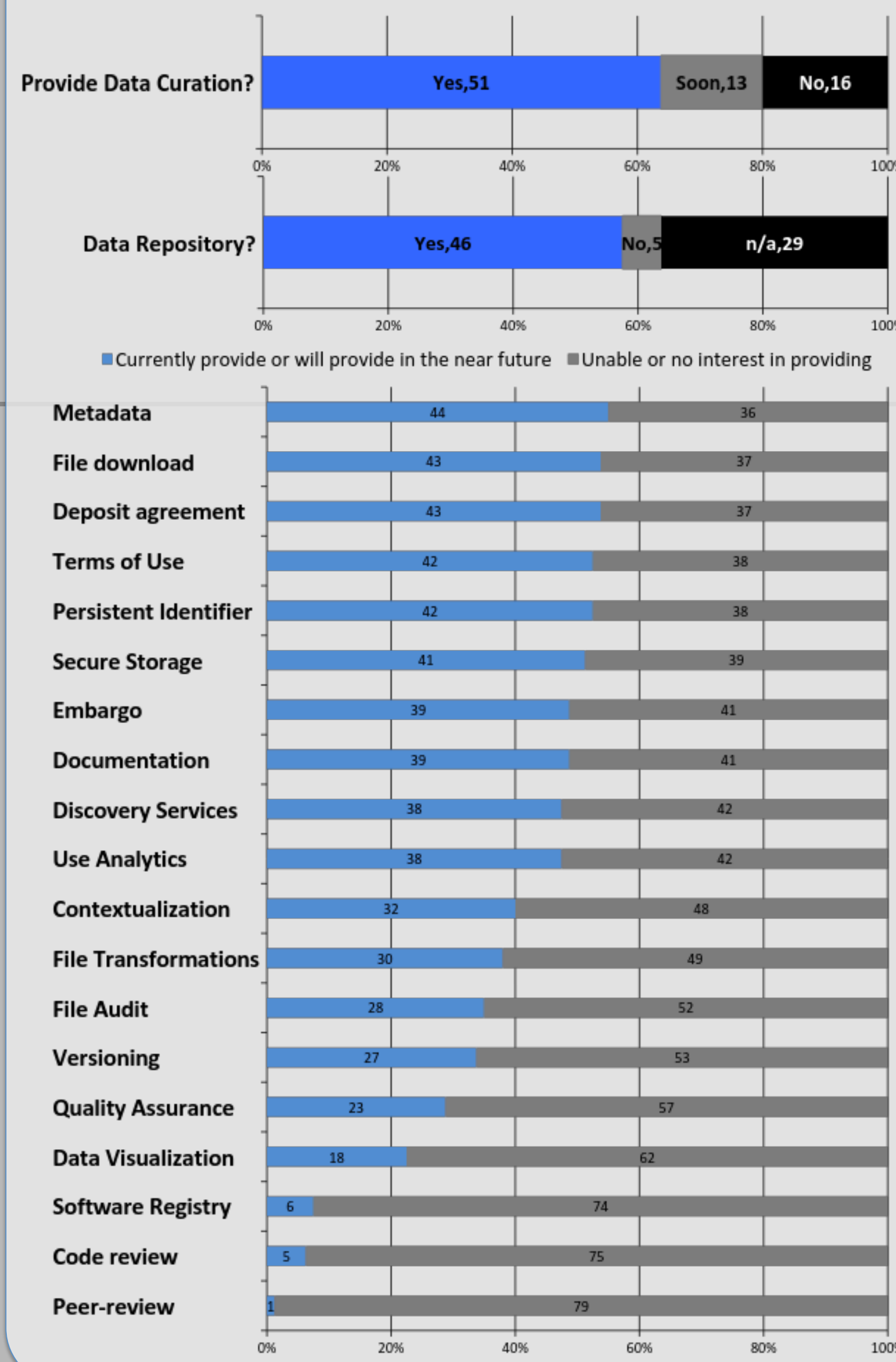
## Where we need to go in the future

### Libraries offer data curation services

#### Data Curation Activities<sup>++</sup>

- Method: Ran in-depth assessments of our 6 partner institutions and a 2017 ARL Spec Kit Survey<sup>++</sup> to gauge current level of data curation support.

Figure 1: Data curation activities offered at ARL Institutions (n=80 from January 2017)



### But, researchers value this...

#### Researcher Engagements Fall 2016<sup>++</sup>

- Method: Held focus groups at partner institutions with 91 researchers – they rated the “Importance” of 35 data curation activities and supplied feedback on their data curation challenges.

Very Important + Well Supported!

- Documentation
- Secure Storage
- Persistent Identifier
- Metadata

Figure 2: Importance of Data Curation Activities

Quality Assurance  
Software Registry  
Data Visualization  
File Audit  
Versioning  
Contextualization  
Code Review  
File Transformations  
Importance Rating by Researchers

Key Finding: Researchers value data curation activities that go beyond services traditionally provided by institutional repositories.

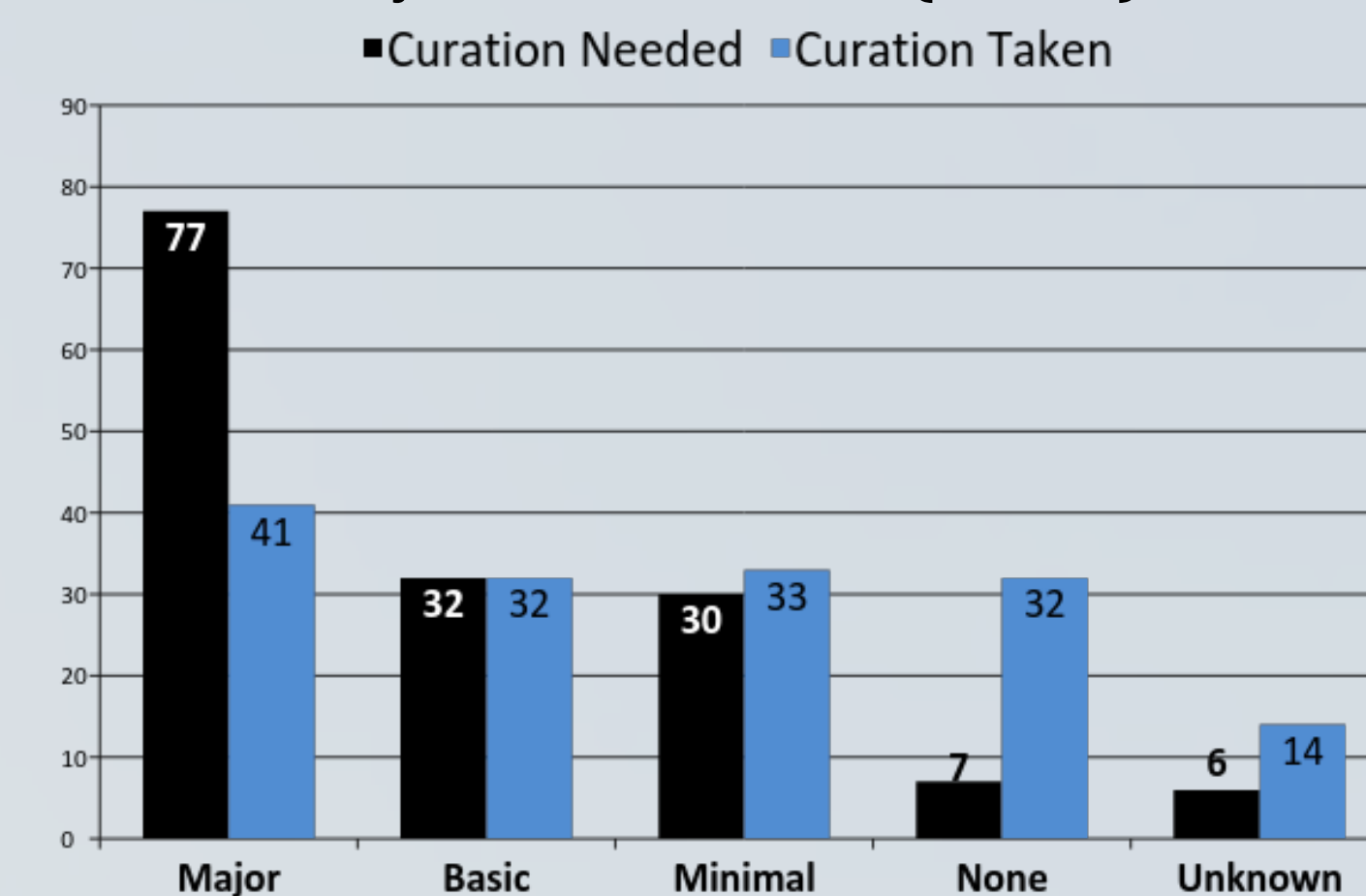
### ...and their data are more varied ...

#### Metrics Tracking in the DCN

- Method: Monitor the data types, frequency, and levels of curation applied to datasets ingested at our 6 partner institutions over one year.

- 152 Data sets curated across 6 institutions Mar 2016-Mar 2017
- 51 Different file types with ~30% Tabular/Spreadsheet files
- 1 hour Time it took to curate the majority of datasets. 70% took less than 2 hours.
- 47% Percent of datasets lacking documentation. 31% did include a readme.txt file.
- 77 Datasets that needed “Major” curation (e.g. documentation, transform file formats), but only 41 received this level.
- 11 Most datasets received in one day

Figure 3: Observed Levels of Curation in DCN May 2016-Mar 2017 (n=152)



### ...than our individual staffing can support.

#### Data Curation Pilots Fall 2016<sup>++</sup>

- Method: Ran two rounds of pilots with 15 Data Curation Staff across our 6 partner institutions to see what we *actually* do.

#### CURATE Steps Needed for all Datasets

- C - Check data files and read documentation
- U - Understand the data (or try to), if not...
- R - Request missing information or changes
- A - Augment the submission with metadata
- T - Transform file formats for long-term reuse.
- E - Evaluate and rate the overall submission for FAIRness.

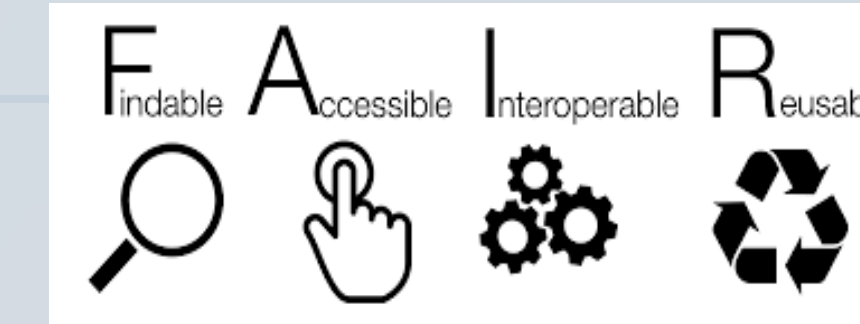


Figure 4: Local Curation Staffing in the DCN

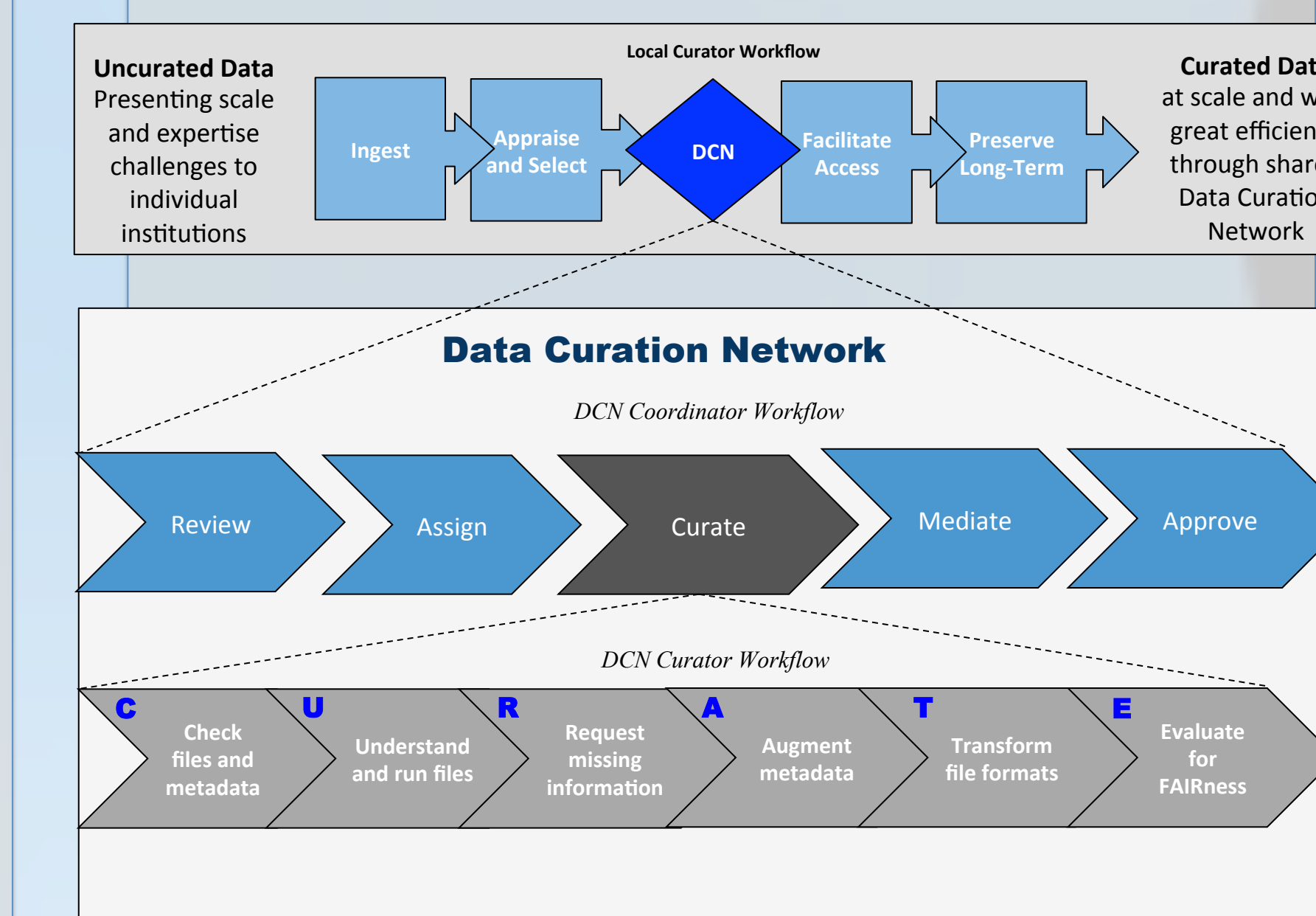
| Repository  | Curator Staff Full/Part time | Subject/Format Expertise  |
|---|------------------------------|---|
| LIBRARIES UNIVERSITY OF MINNESOTA                       | 0 / 7                        | Excel, GIS, stat/survey (R, STSS, SAS), genomics, public health   |
| Cornell University Library                              | 0 / 2                        | Excel, bio, environmental and earth sciences, EML                 |
| Illinois Data Bank                                      | 0 / 3                        | Excel, stat/survey (R, STSS, SAS), biochem, audio/visual, TEI     |
| Deep Blue Data  | 1 / 3                        | Excel, bio, genomics, health informatics                          |
| PennState University Libraries                          | 0 / 3                        | Excel, GIS, R, Python, PII/sensitive, ecology/EML, bioinformatics |
| Washington University in St. Louis UNIVERSITY LIBRARIES | 0 / 3                        | Excel, GIS, Matlab, Mathematica, RDF/XML, Qualitative Health      |

### So, let's bridge this gap and partner.

#### Vision for the Data Curation Network:

- Develop standards-driven data curation techniques for all types of repository infrastructure.
- Expand into a sustainable entity that grows beyond our initial six partner institutions.
- Datasets curated by the Data Curation Network will be used to advance research and education in ways that are measurably of greater reuse value than non-curated data.
- Build an innovative community that enriches capacities for data curation writ large.

Figure 5: Model for the Data Curation Network



### A network of expertise model for Data Curation

#### Data Curation Network Governance:

- Membership Model: Institutions contribute curation staff time and small \$\$ to sustain DCN post grant.
- Benefits: Training/prof dev for curators (annual), local control over which data get curated by network,

Figure 6: Workflow for the Data Curation Network

