

# Implementing a Cross-Institutional Staffing Model for Curating Research Data

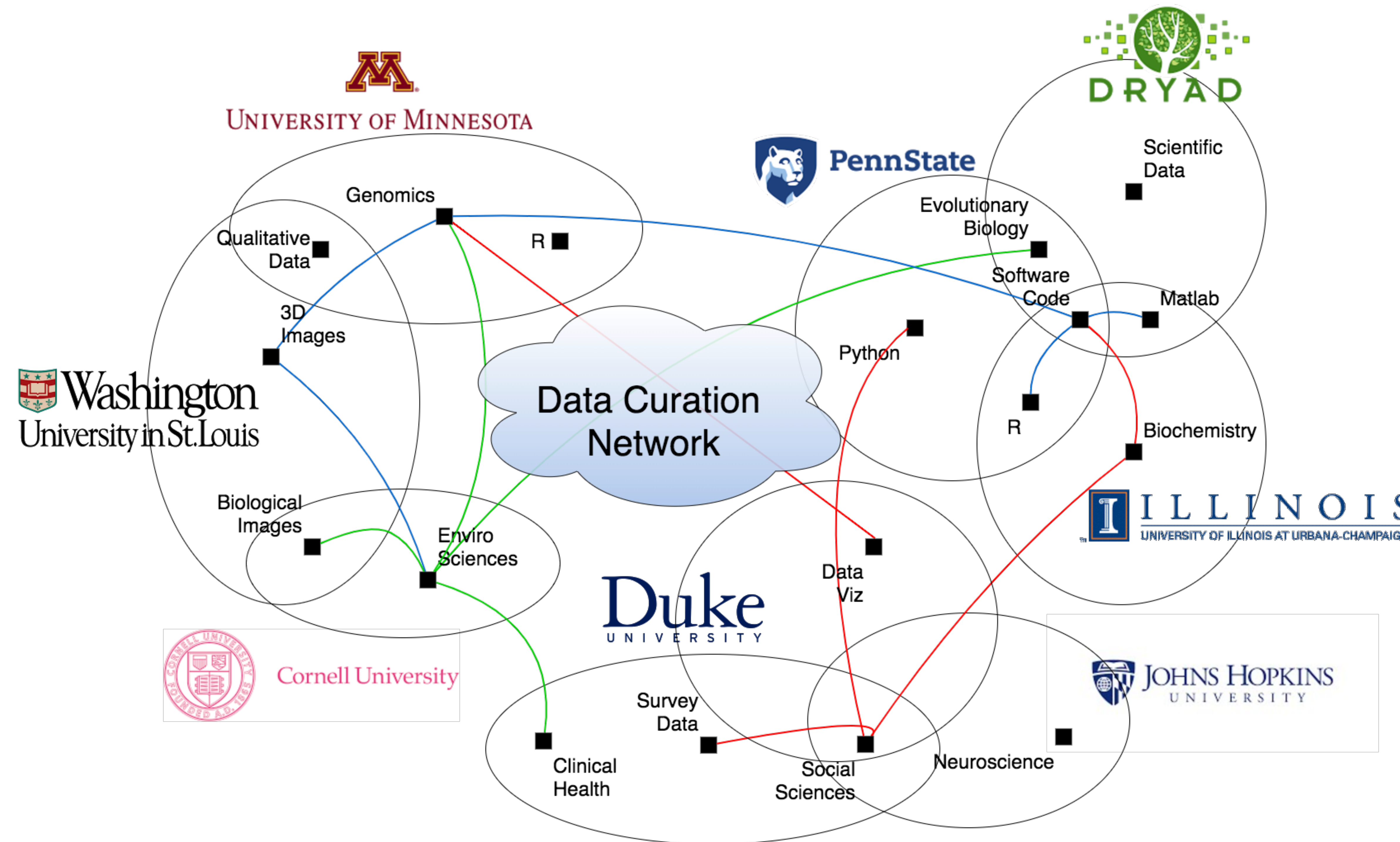
The **skills and expertise** required to curate data cannot be fully automated nor reasonably provided by a few experts siloed at a single institution.



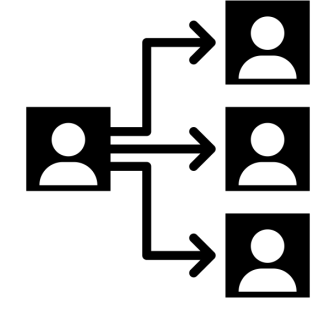
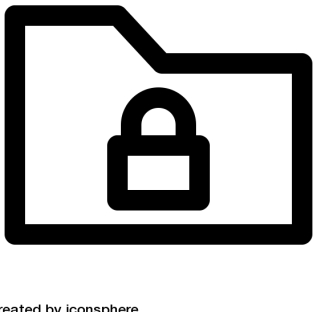
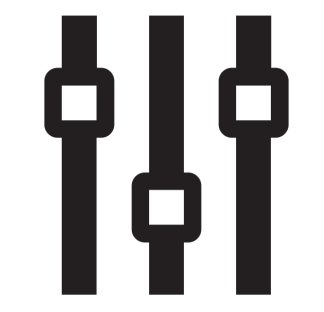
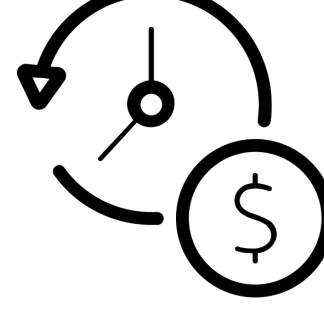


The Data Curation Network (DCN) addresses this challenge by **collaboratively sharing data curation staff** across a network of partner institutions and data repositories.

Method: In 2016-2017 six DCN partner institutions performed baseline assessments of local services, held focus groups with faculty researchers, ran controlled data curation pilots, and surveyed the library curator community to understand existing support and future plans for services in these areas. The results showed that multiple data curation experts are needed to effectively curate the diverse data types a repository typically receives and to keep up with changing trends and emerging tools that support research data best practice.

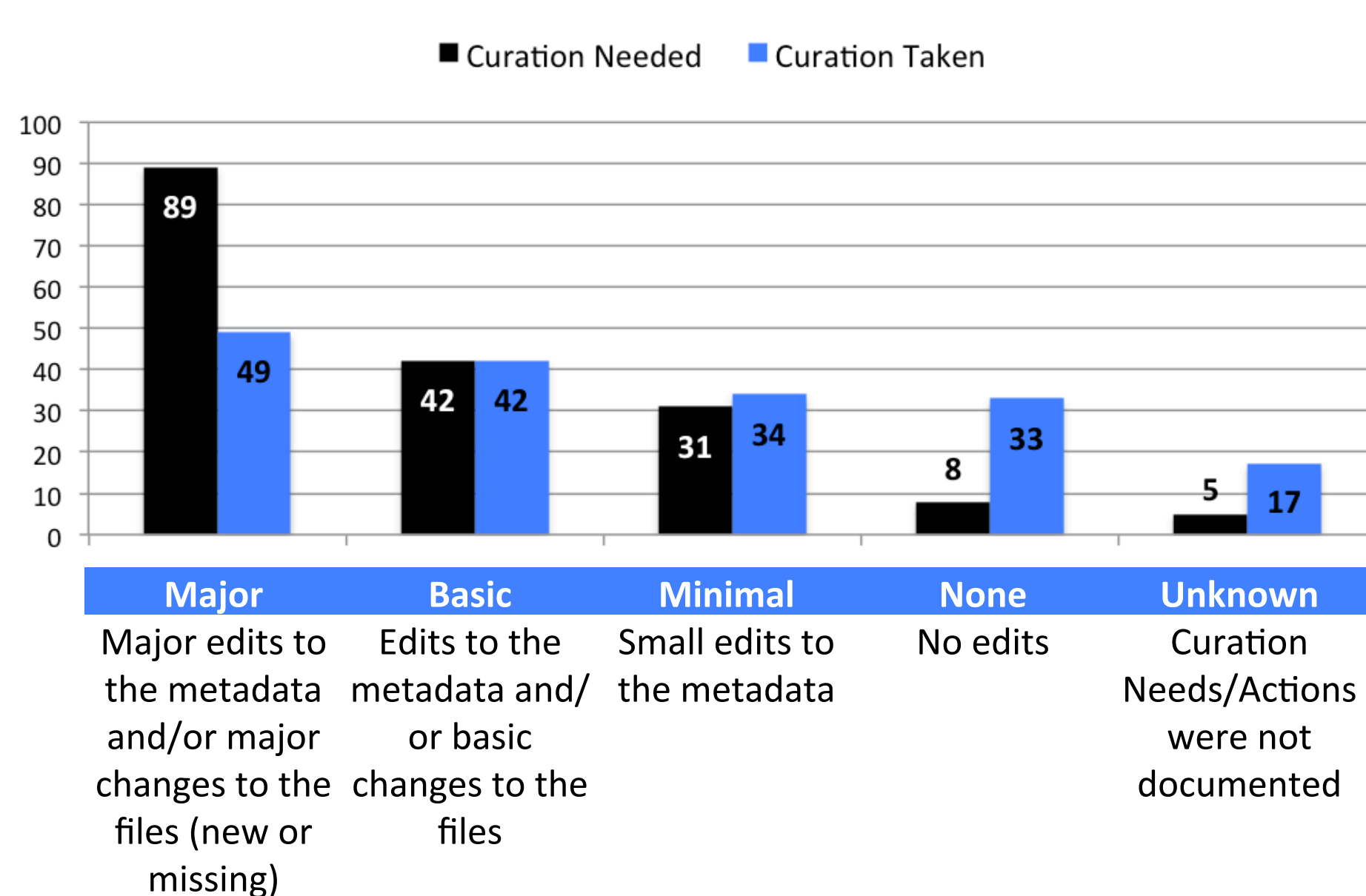
- 175** Data sets observed across 6 DCN institutions in 2016-2017.
- 77%** Representation from scientific disciplines including agricultural, biological, engineering/applied, and physical sciences.
- 51** Different file types with ~30% tabular/spreadsheet files.
- 1 hour** Average time spent curators spent processing the majority of datasets. 70% took less than 2 hours.
- 47%** Percent of datasets lacking documentation.
- 77** Datasets that needed "Major" curation (e.g., documentation, transform file formats). Only 41 received this level.
- 11** Most datasets received in one day.



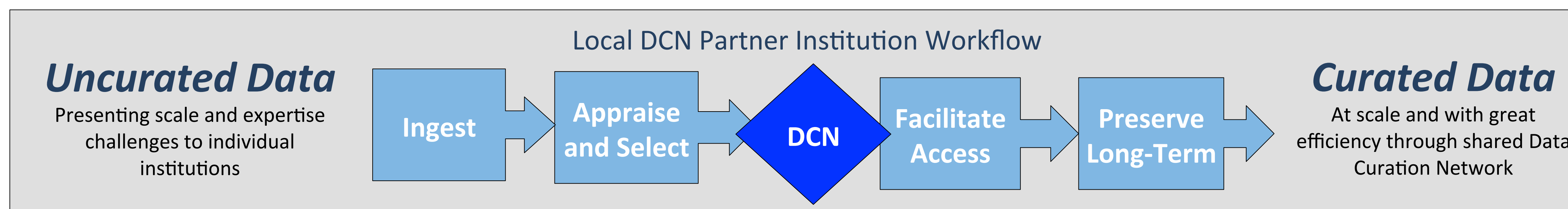
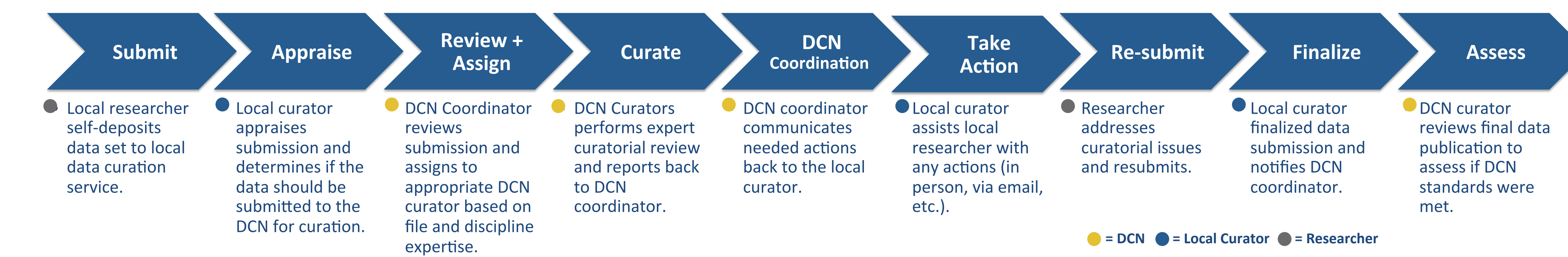
The DCN brings the perspectives of researchers, librarians, administrators, and data curation subject experts in a cross-institutional network led by the University of Minnesota and including Cornell University, Penn State University, the University of Illinois, the University of Michigan, Washington University in St. Louis, Duke University, Johns Hopkins University, and the Dryad Data Repository. The DCN functions as:

-  a well-coordinated and tiered staffing model that incorporates data curator expertise across a wide-variety of domains
-  a technology-agnostic submission workflow that accommodates the various repository technologies in use (e.g. Samvera/Hydra, DataVerse, DSpace, BePress, etc.)
-  standardized minimum levels of curation (CURATE Steps) that enable DCN Curators to prioritize their work and promote FAIR standards
-  a sustainable, non-profit financial plan to support the DCN beyond the grant-supported implementation phase
-  an assessment plan to evaluate how a networked approach to curating research data is more efficient and scalable
-  a professional development program that enables DCN curators to stay up-to-date with data best practices across domains, communities, nations, and beyond.

Levels of Curation Needed vs. Taken for Datasets Observed in 2016-2017 Planning Phase (n=175)



## The Data Curation Network (DCN) Submission Workflow



**CURATE Steps 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.

F Findable, A Accessible, I Interoperable, R Reusable

A written version of our DCN model is available with detailed curation workflows, staffing roles, draft MOUs, and tool tracking functional requirements as "Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data" originally published July 27, 2017 online at <https://sites.google.com/site/datacurationnetwork/results>.