

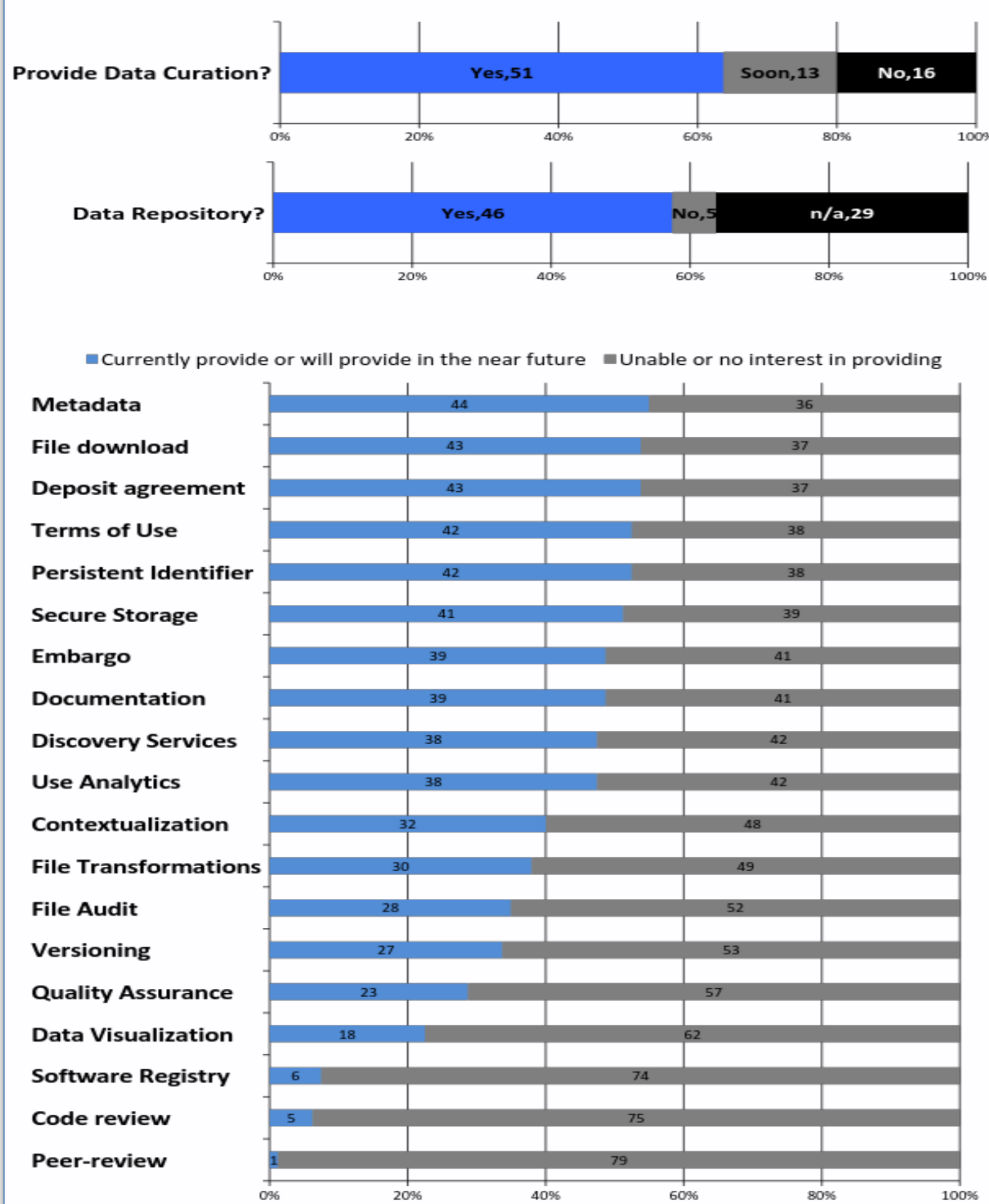
# Data Curation Network

A network of expertise model for curating research data in digital repositories

Robert Olendorf, Pennsylvania State University  
 Jake Carlson, University of Michigan  
 Cynthia Hudson-Vitale, Washington University in St. Louis  
 Heidi Imker, University of Illinois at Urbana—Champaign  
 Wendy Kozlowski, Cornell University  
 Lisa R. Johnston, (PI) University of Minnesota  
 Claire Stewart, University of Minnesota

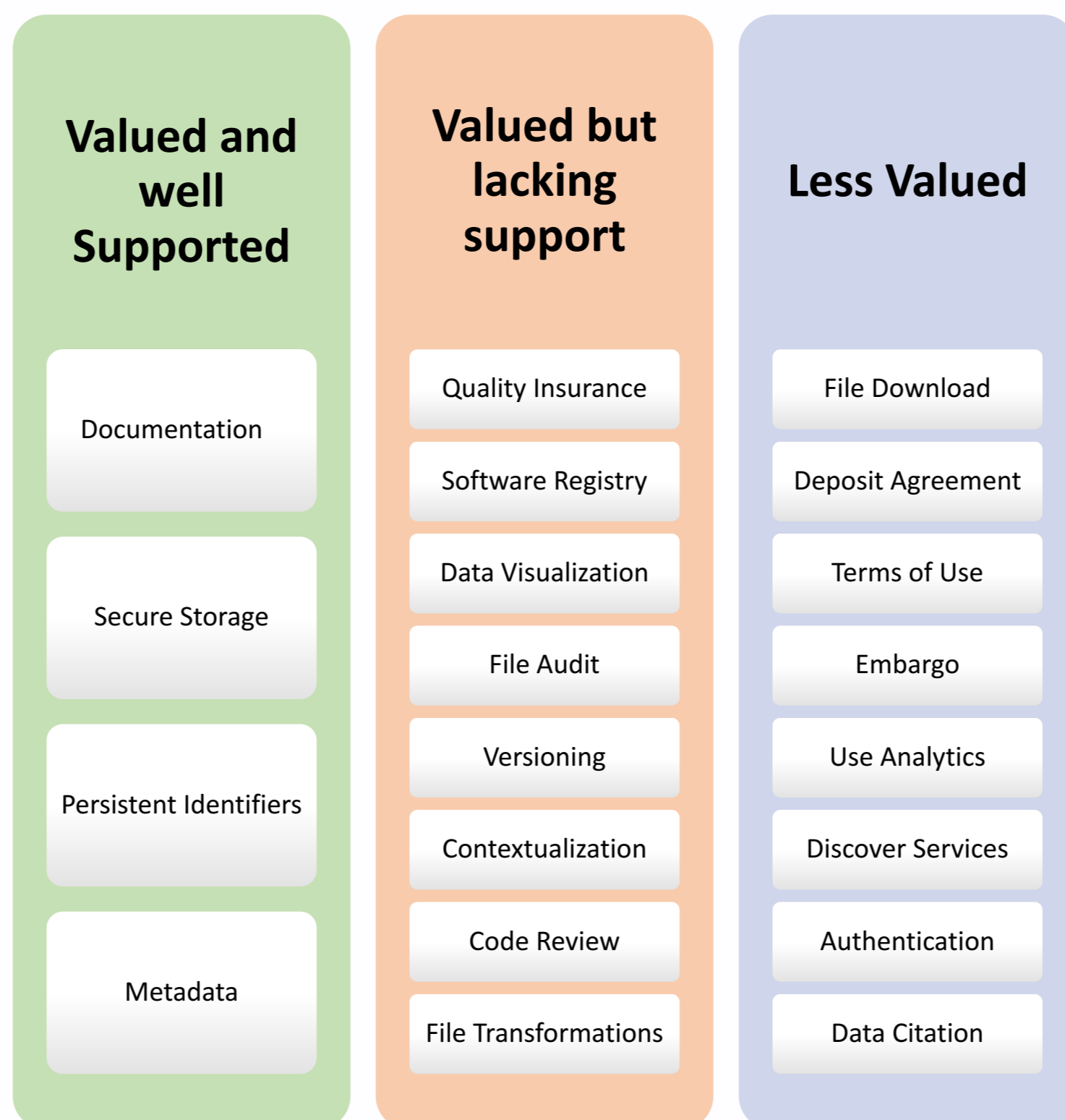
The Data Curation Network will enable academic institutions to better support researchers that are faced with a growing number of requirements to ethically share their research data.

## Libraries Offer Data Curation Activities ...



Data from in depth assessments of partner institutions and a 2017 ARL Spec Kit Survey. \*

## ... But, Researchers Value These Activities.



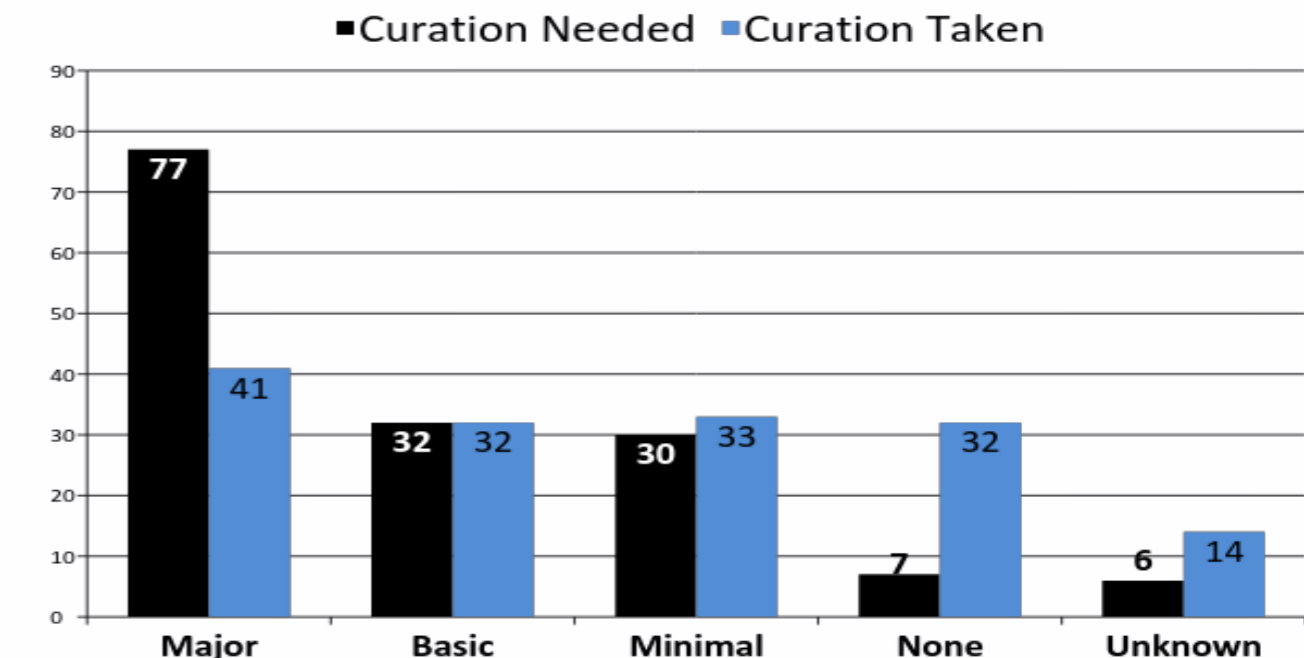
Researchers value data curation activities that go beyond services traditionally provided by institutional repositories.

Based on 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. \*

## The Data Are More Varied...

- 152 Data sets curated
- 51 File types (~30% tabular)
- 1 Hour Time to curate majority of datasets.
- 47% Percent of datasets lacking documentation.
- 31% Included a README file
- 11 Most datasets received in one day.
- 77 Datasets needed Major Curation. Only 41 received this level.

## Observed Levels of Curation in the DCN



Taken from data sets ingested into partner institution repositories over one year. May 2016 – March 2017 (n= 152) \*

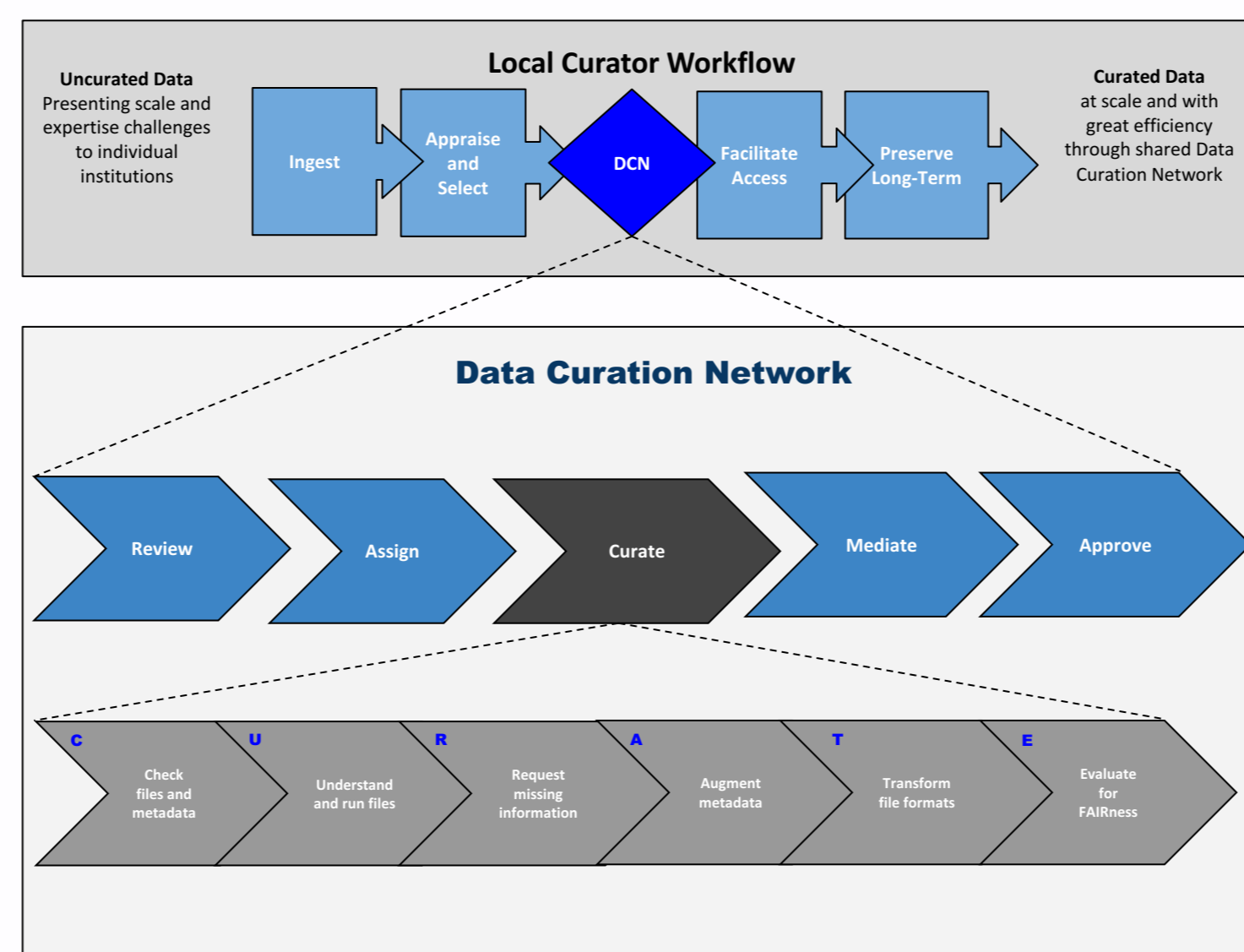
## ... Than our individual staffing can support.

Repository	Curator Staff Full/Part time	Expertise
LIBRARIES Data Repository for the U of M (DRUM)	0/7	Excel, GIS, Survey, R, STSS, SASS, genomics, public health
Cornell University Library eCommons @ Cornell	0/2	Excel, biology, environmental, earth sciences, EML
Illinois Data Bank	0/3	Excel, Survey, R, STSS, SAS, biochemistry, audio/visual, TEI
LIBRARY Deep Blue Data	1/3	Excel, biology, genomics, health informatics
PennState University Libraries ScholarSphere	0/2	Excel, GIS, R, Software, Private, EML, bioinformatics
Washington University in St. Louis UNIVERSITY LIBRARIES Digital Research Materials Repository	0/3	Excel, GIS, MatLab, Mathematica, RDF/XML, Health

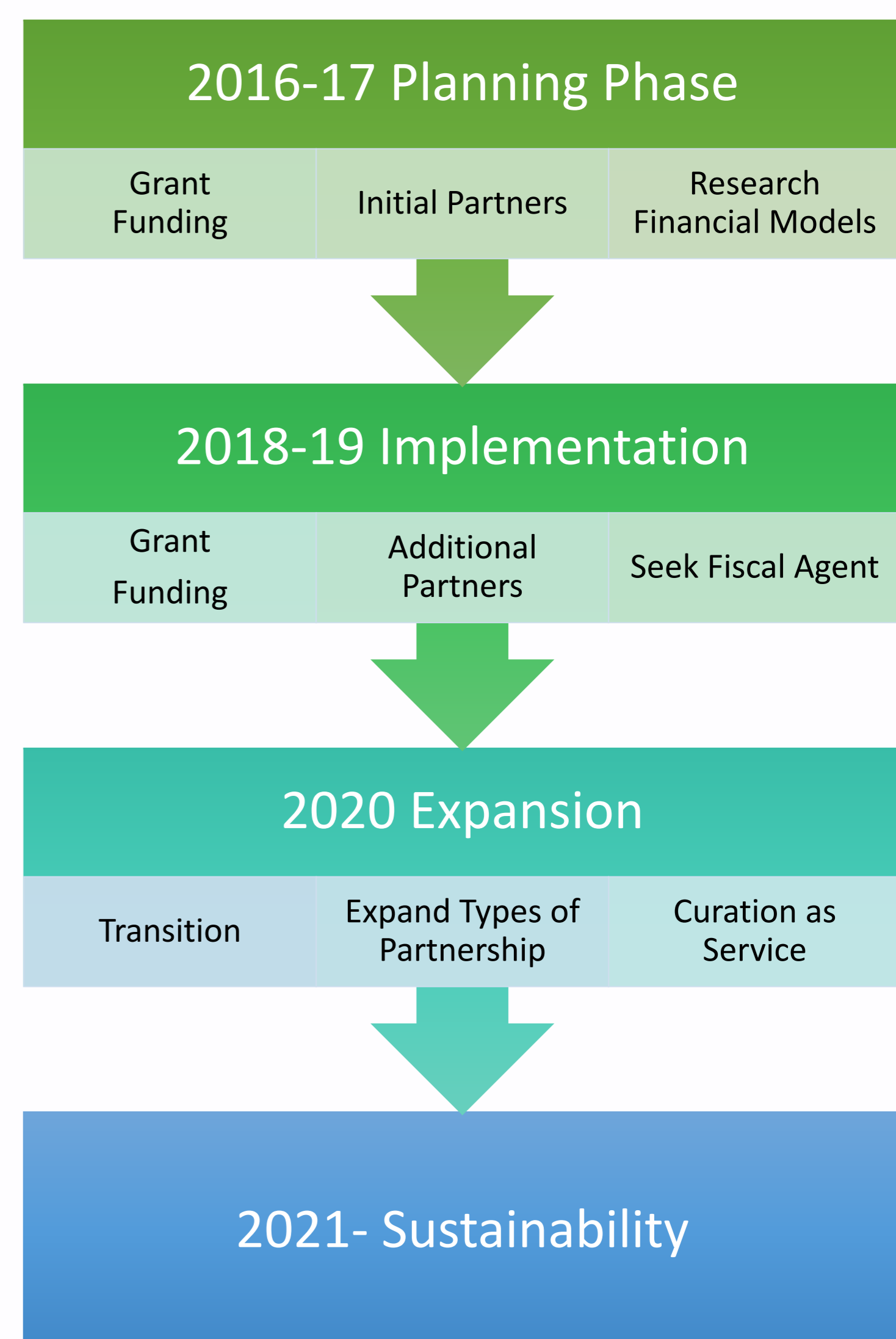
Partner institutions only partially cover needed expertise, and most staffing is part time.

## Shared Curatorial Staffing With the Data Curation Network

- Develop standards-driven data curation techniques.
- Repository Agnostic
- A sustainable entity that grows beyond our initial six partner institutions.
- Advance research and education with quality, curated data.
- Build an innovative community that enriches capacities for data curation.



## The Future of the Data Curation Network



Find more at

<https://sites.google.com/site/datacurationnetwork>

#DataCurationNetwork