Eliciting Faculty Requirements for Research Data Repositories

Michael Witt
Interdisciplinary Research Librarian & Assistant Professor
Purdue Libraries & Distributed Data Curation Center (D2C2)
mwitt@purdue.edu

Open Repositories 2009, Georgia Tech: May 18, 2009

Ten Questions to Begin a Conversation With Your Faculty About Data Curation

1. What is the story of your data?
2. What form and format are the data in?
3. What is the expected lifespan of your data?
4. How could your data be used, reused, and repurposed?
5. How large is your dataset, and what is its rate of growth?
6. Who are potential audiences for your data?
7. Who owns the data?
8. Does the dataset include any sensitive information?
9. What publications or discoveries have resulted from the data?
10. How should the data be made accessible?

Witt, M. & Carlson, J. (2007). Conducting a data interview. http://docs.lib.purdue.edu/lib_research/81/.

PURDUE UNIVERSITY Libraries Access. Knowledge. Success.

Investigating Data Curation Profiles Across Multiple Research Disciplines

Investigators in the Distributed Data Curation Center in the Libraries at Purdue University, and the University of Illinois, Urbana-Champaign will address the question *"which researchers are willing to share data, when, with whom, and under what conditions?"* The team will produce case studies of researcher data/metadata workflow, data curation profiles describing policies for archiving and making available research data, a matrix to compare parameters across disciplines, system requirements for managing data in a repository, and recommendations for implementing results under diverse systems. The project will describe the roles of librarians and identify the skill sets they need to facilitate scholarly communication and data sharing. Supported by IMLS LG-06-07-0032-07.

Investigators

D. Scott Brandt (PI) – Purdue University

Jacob Carlson – Purdue University

Melissa Cragin – University of Illinois

P. Bryan Heidorn – University of Illinois

Carole Palmer – University of Illinois

Sarah Shreeves – University of Illinois

Michael Witt – Purdue University

Two-year research project began on 11/15/2007.

Project Activities

- Two interviews each with 20 faculty who produce data in a variety of research domains
- Transcription, coding, and analysis (NVivo)
- Creation of “data curation profiles” and wiki
- Developing two case studies in Agronomy and Geology
- Two focus groups with subject-specialist librarians who acted as liaisons

- Distinguish and map needs expressed by faculty to repository functionality
- Assess current capabilities of repository systems and related technologies
- Experiment using institutional repositories for data curation in practical terms

Subjects

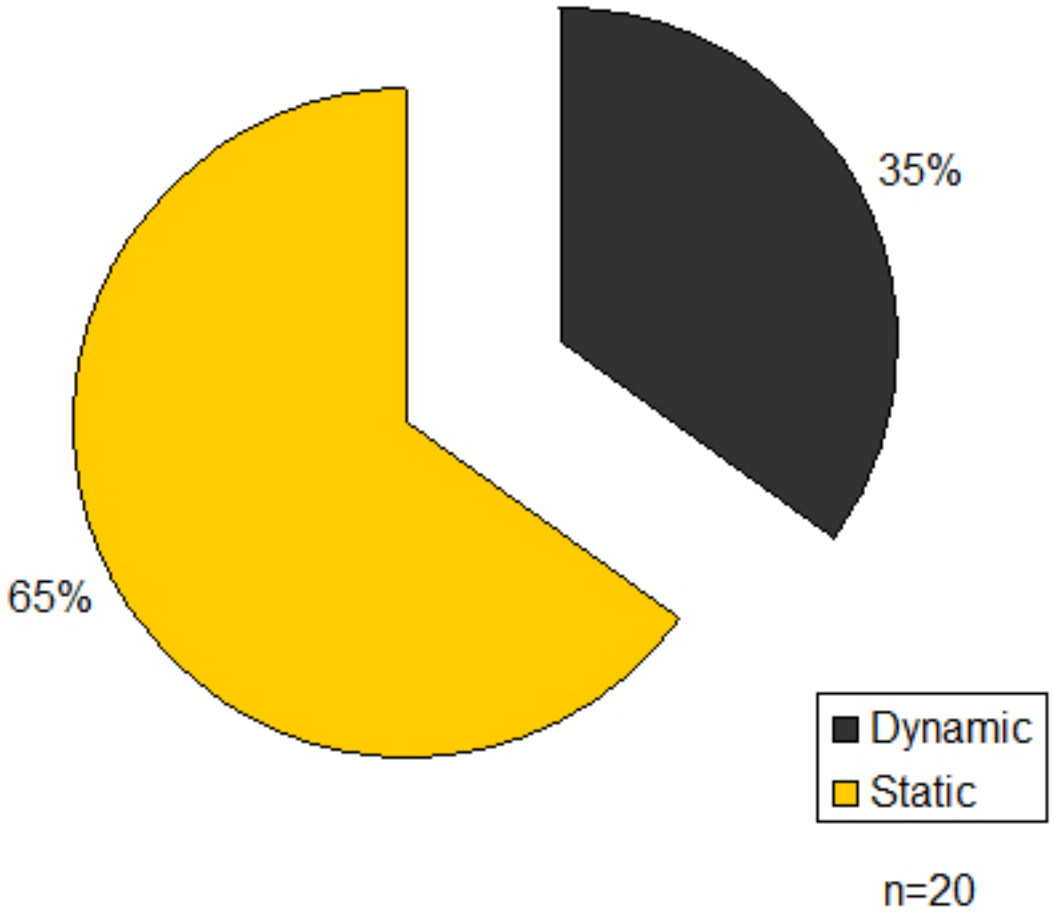
Purdue	Illinois
• Biology	• Kinesiology
• Horticulture	• Atmospheric Sciences
• Civil Engineering	• Speech & Hearing
• Electrical & Computer Engineering	• Soil Science
• Biochemistry	• Anthropology
• Food Science	• Anthropology
• Earth & Atmospheric Science	• Anthropology
• Agronomy	• Geology
• Agronomy	• Geology
• Agronomy	• Geology
• Agronomy	

Case Studies

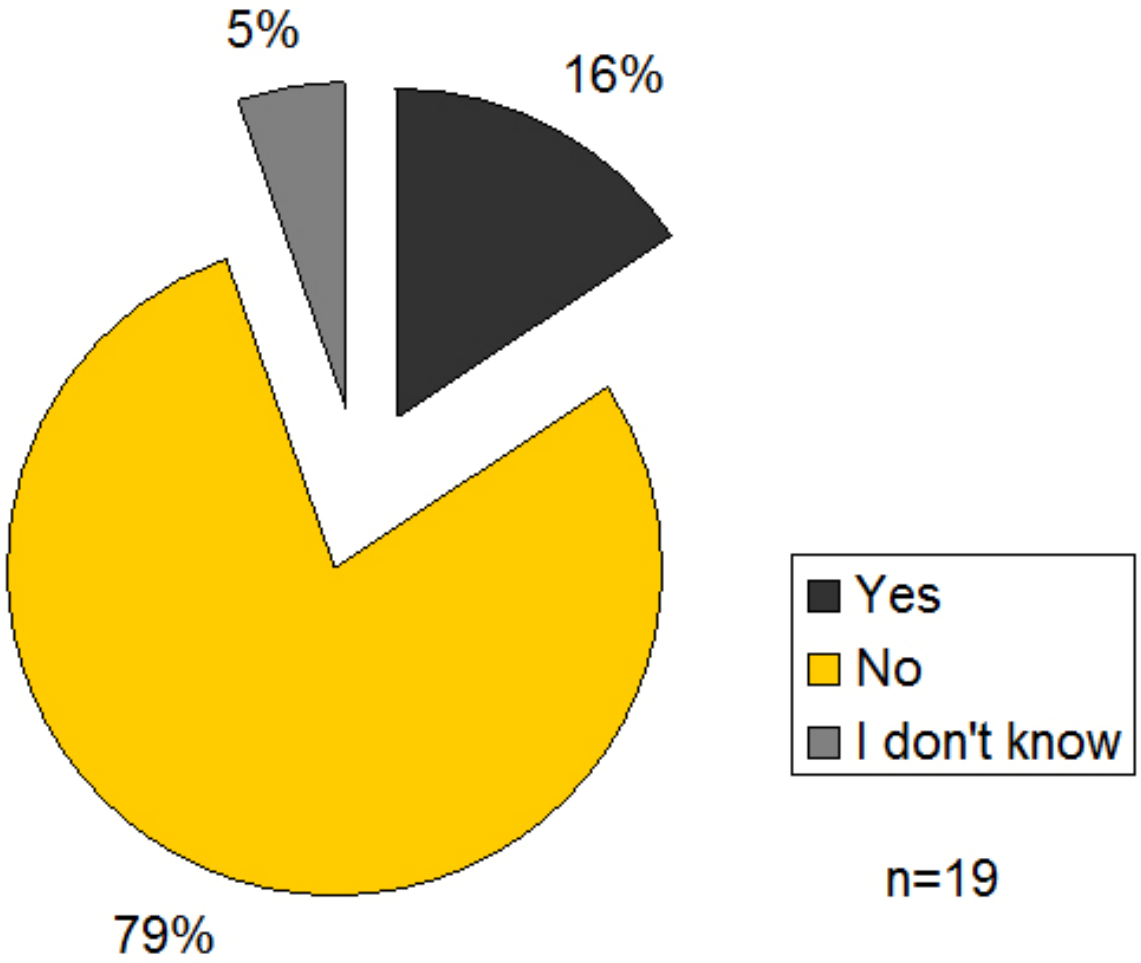
Caveat audiens

- Preliminary findings (project is not yet complete)
- Convenience sample, not statistical
- Exploratory, qualitative study
- The subjects provide much more context and information in the interview transcripts, which are still being coded

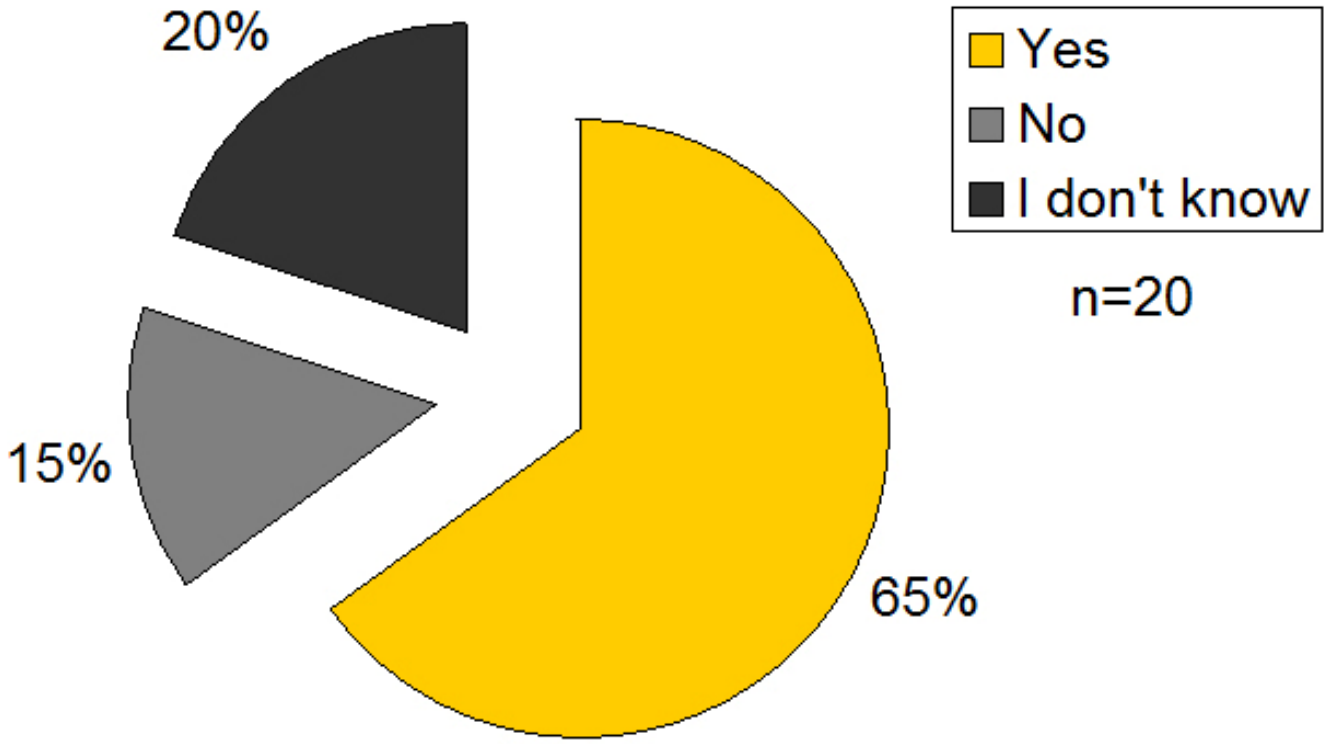
Dataset static or dynamic?



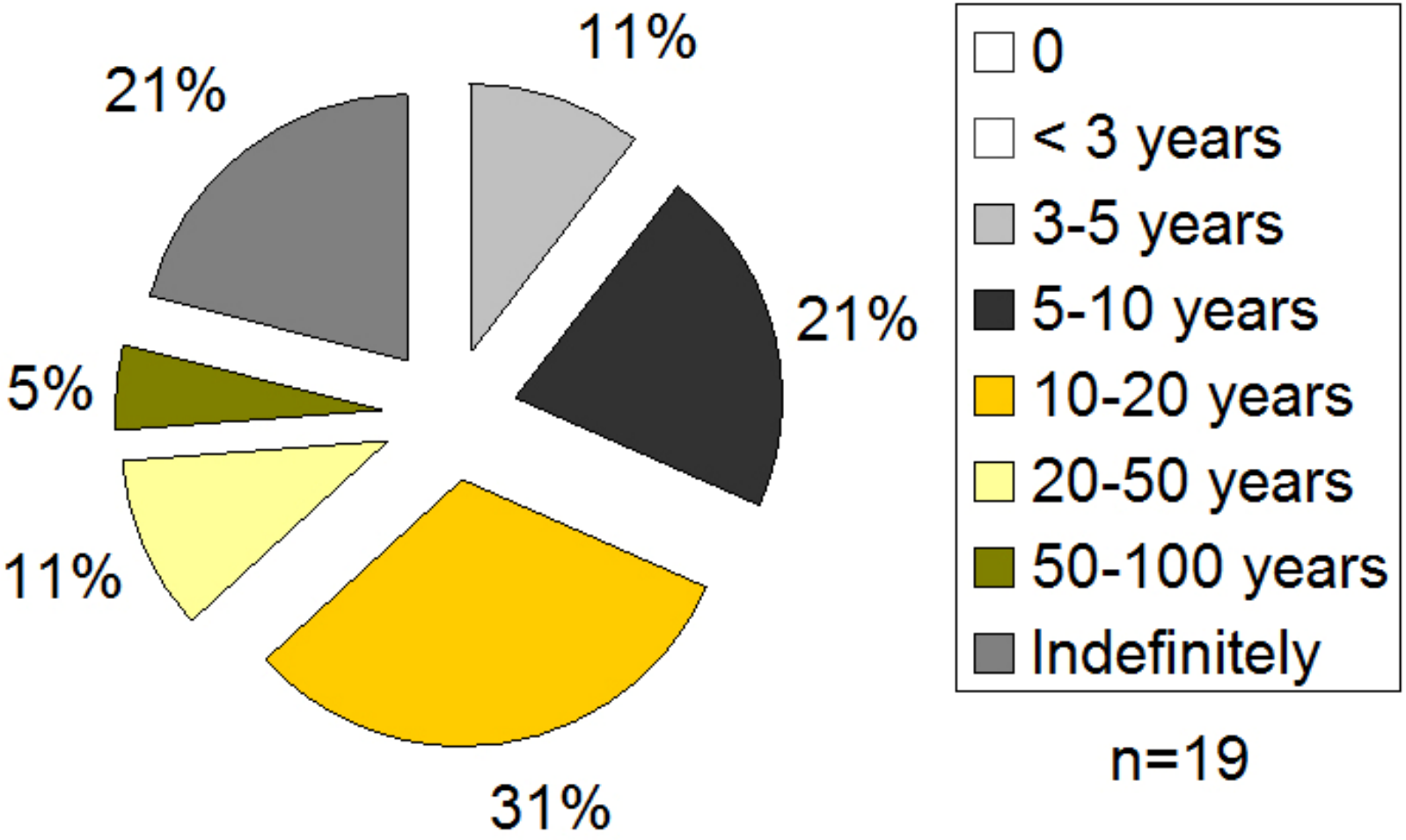
Data bound by confidentiality or privacy concerns?



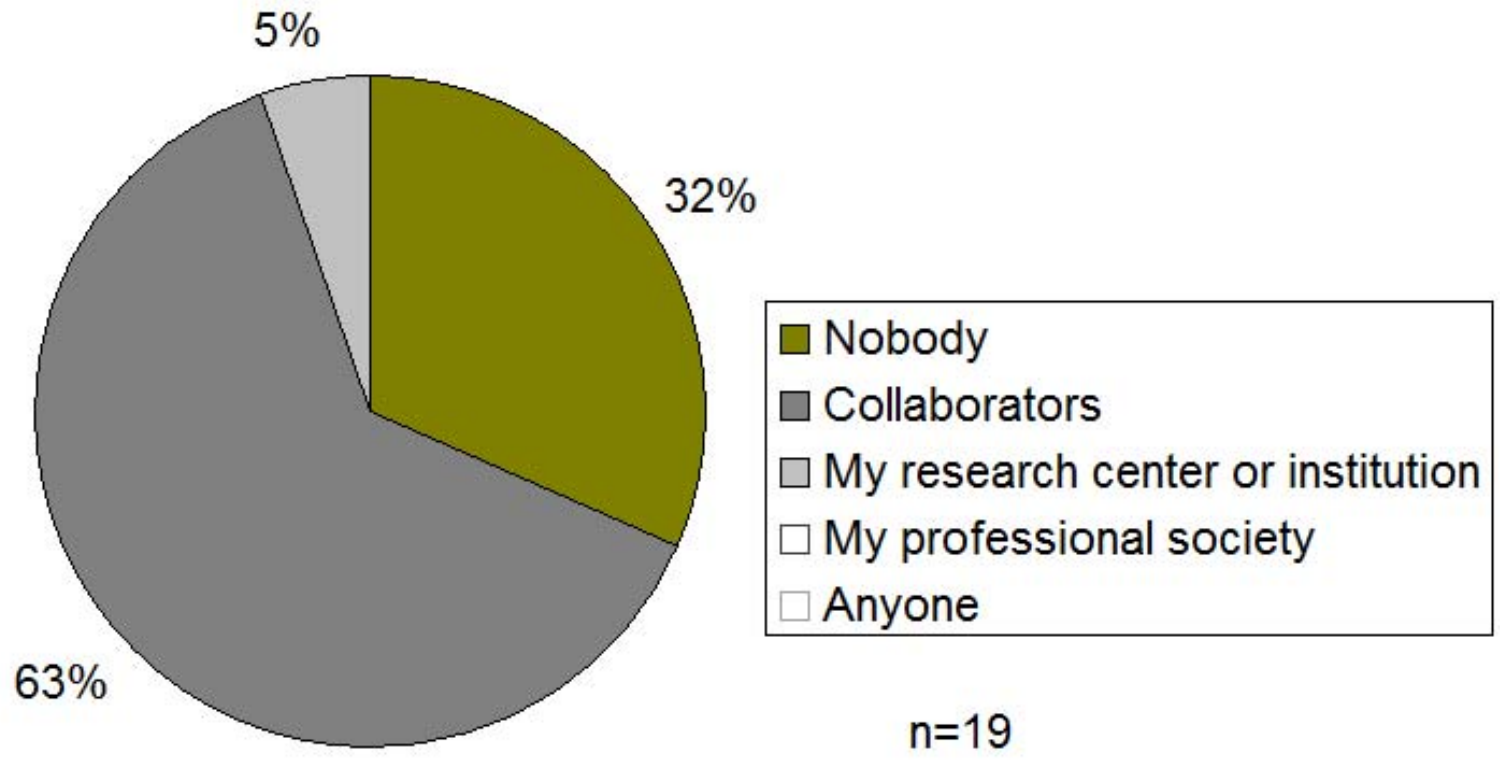
Is your manner of organization/description sufficient for another person with similar expertise to be able to understand and properly use the data?



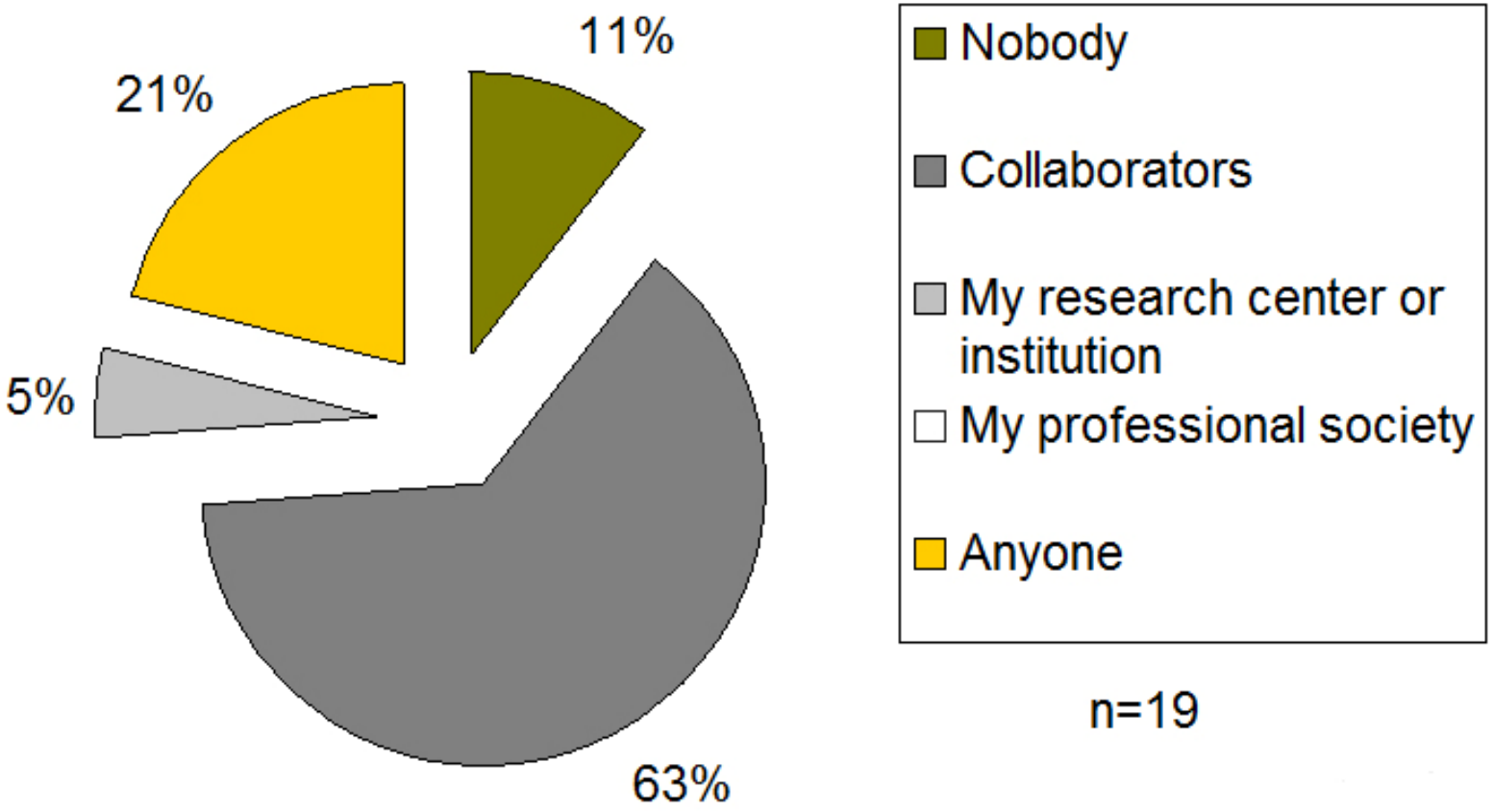
How long to preserve your data?



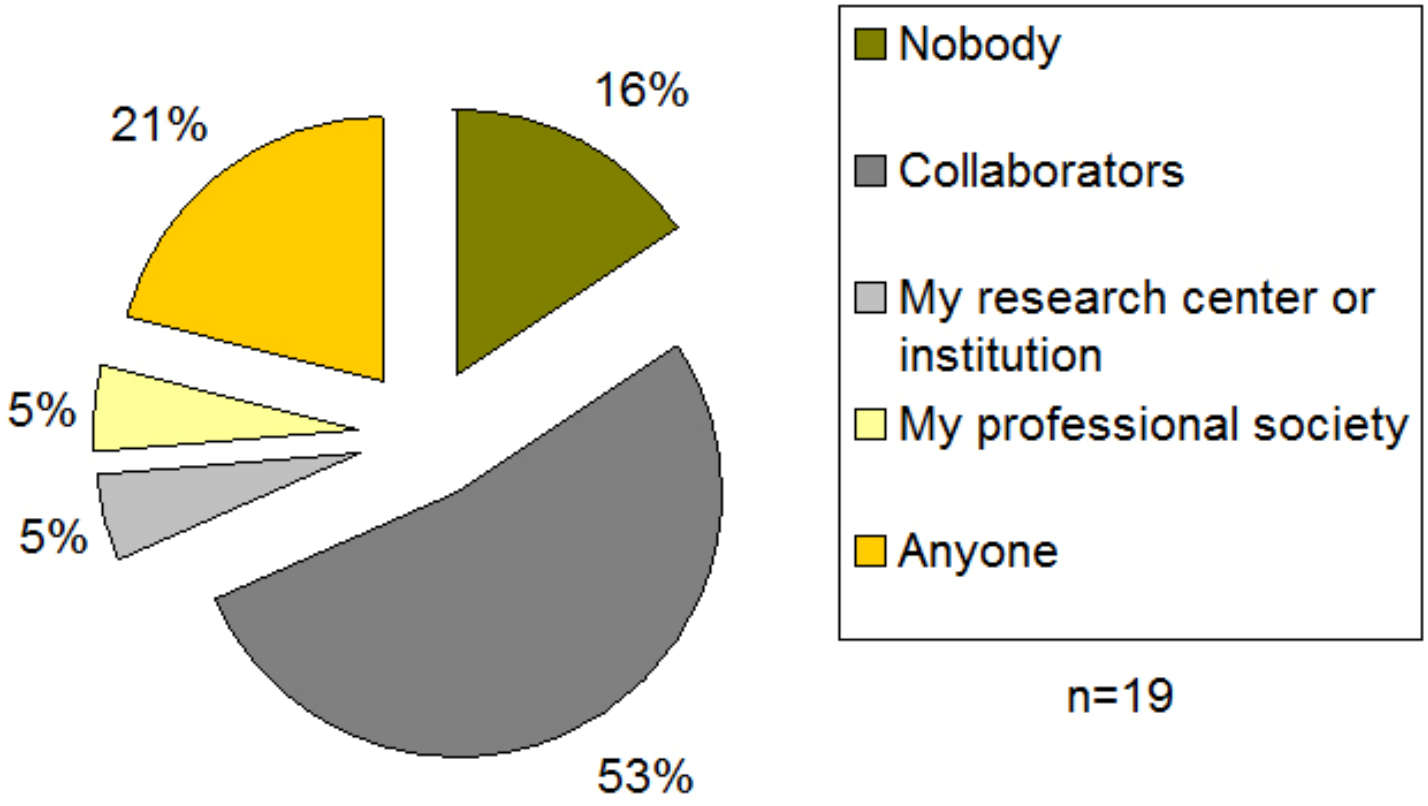
With whom would you share your data *immediately after the data were generated?*



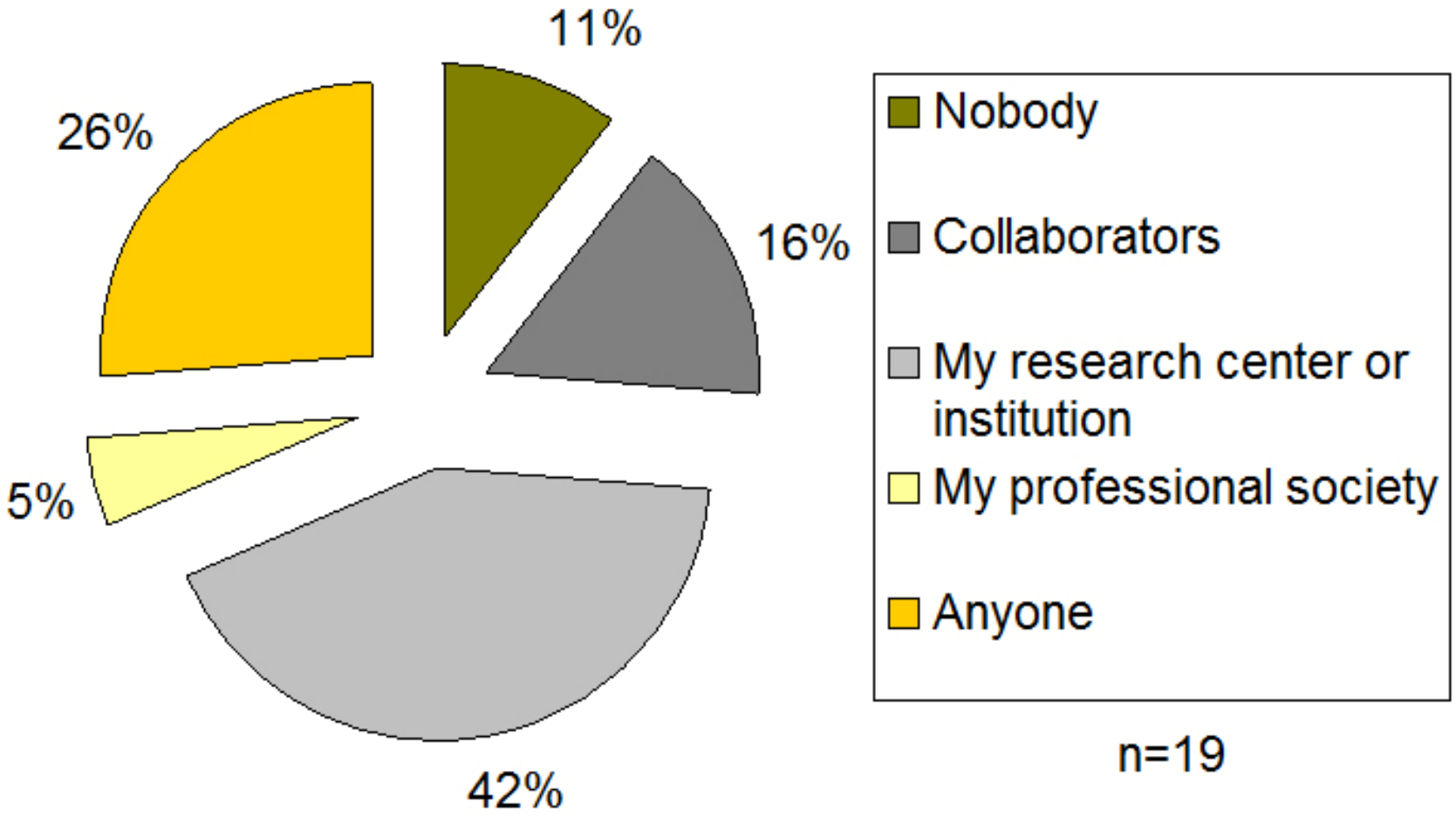
With whom would you share your data *after the data were normalized and/or corrected?*



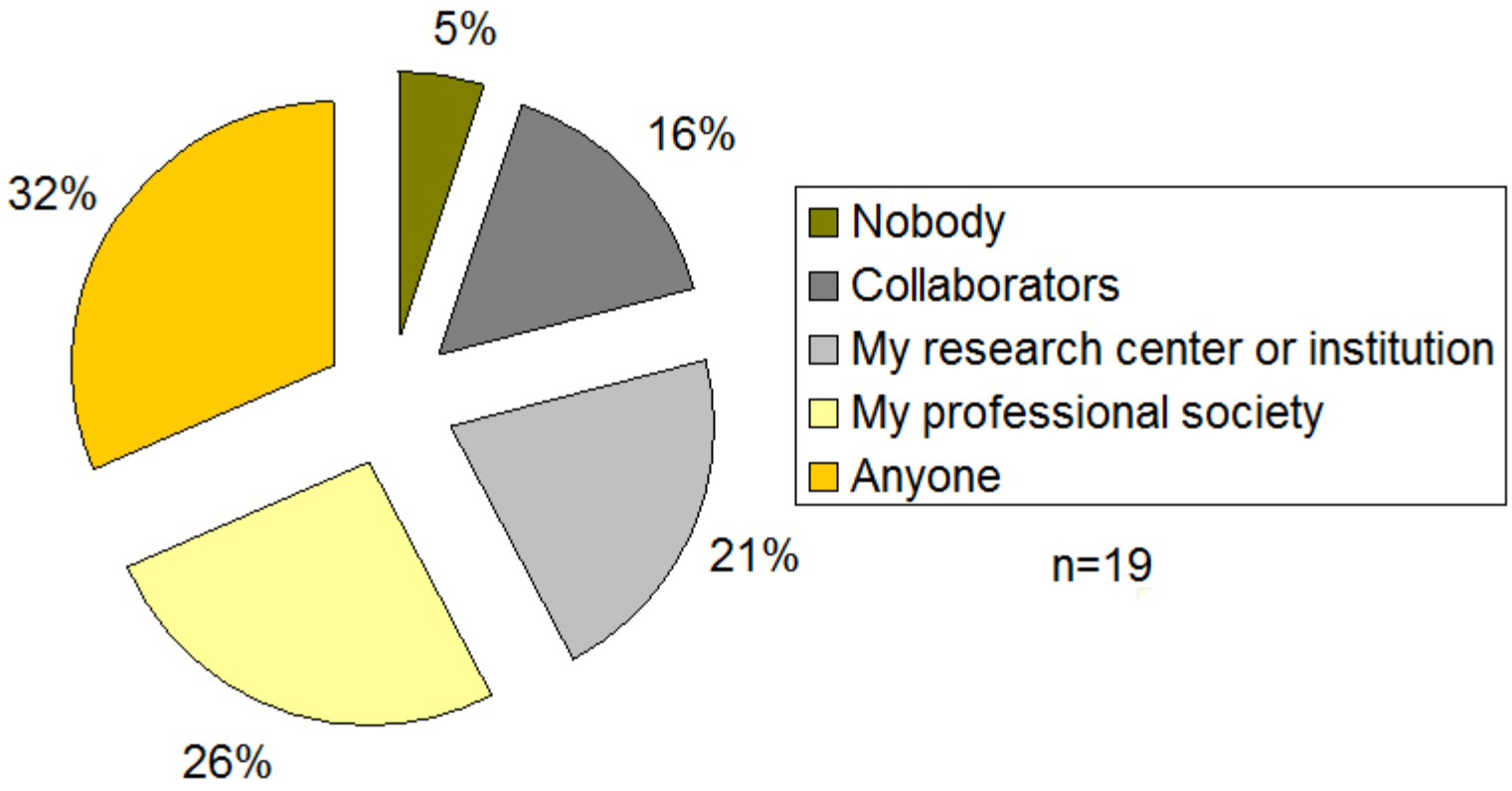
With whom would you share your data *after the data have been processed for analysis?*



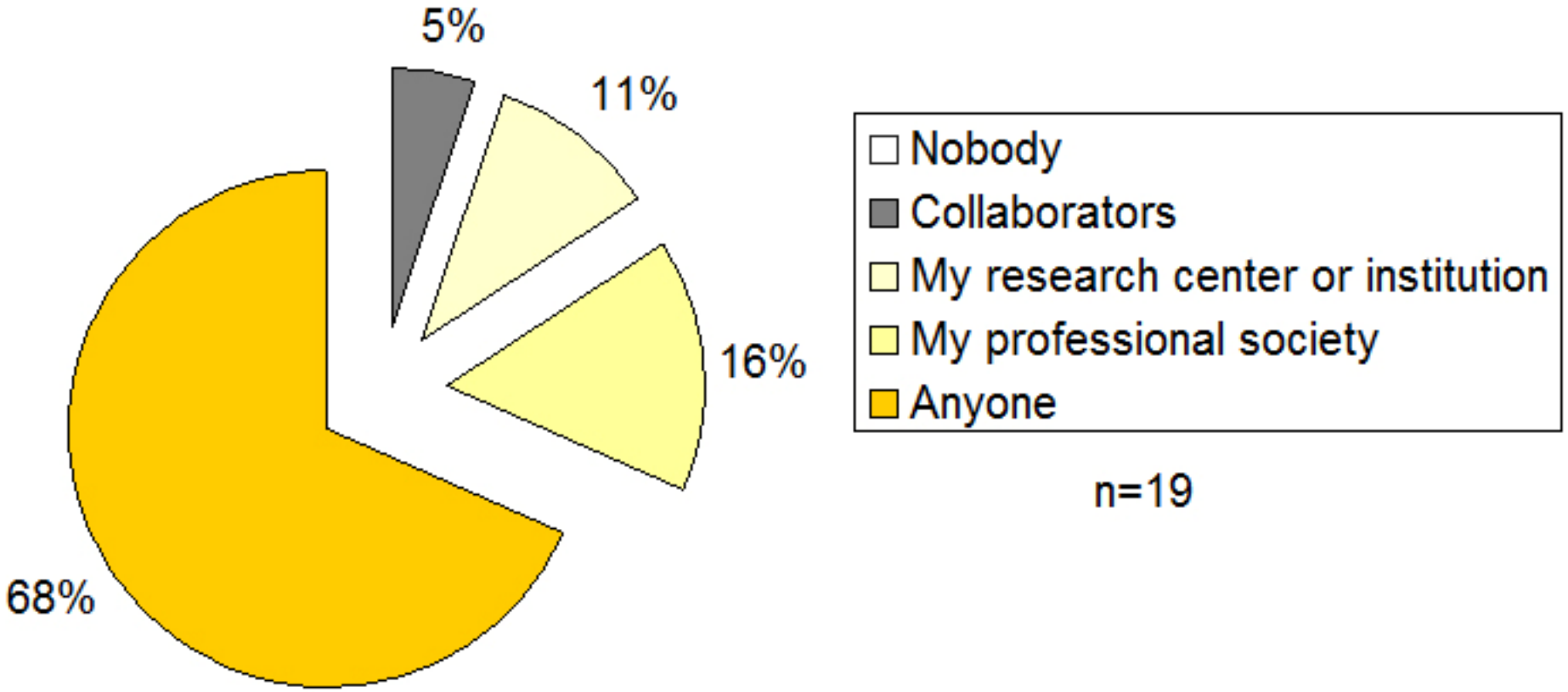
With whom would you share your data *after the data have been analyzed?*



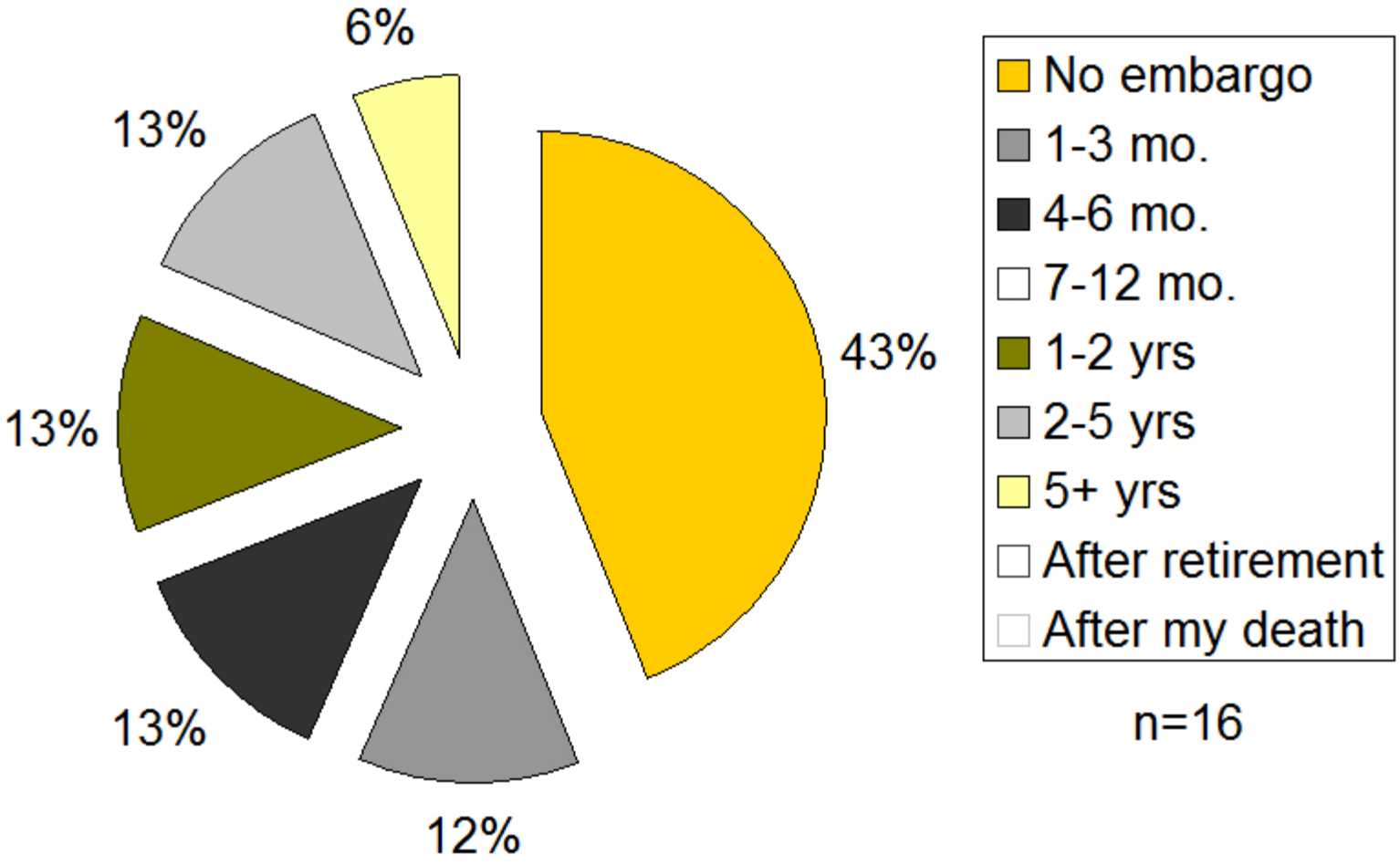
With whom would you share your data *immediately before publication of findings?*



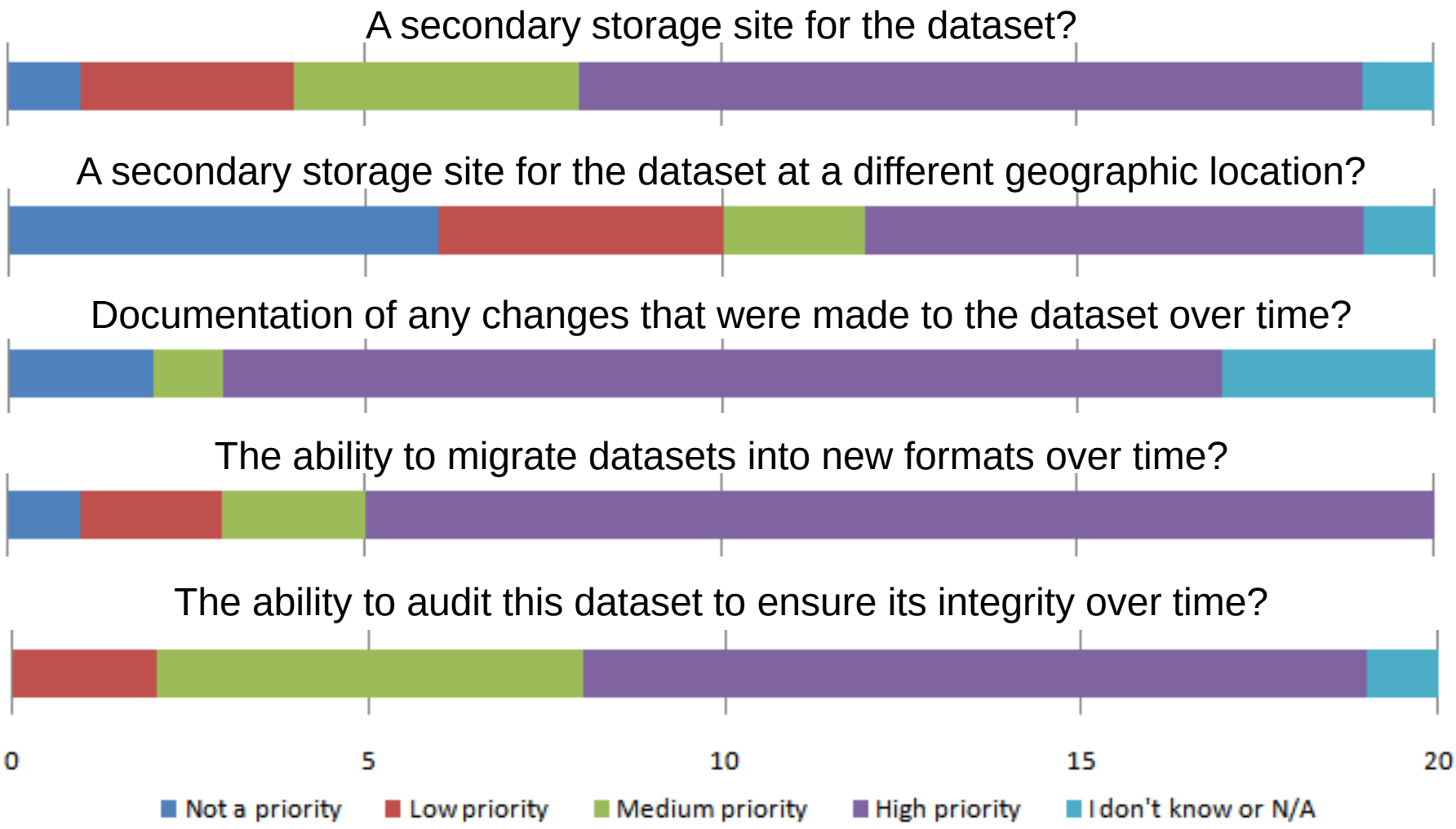
With whom would you share your data *immediately after publication of findings?*



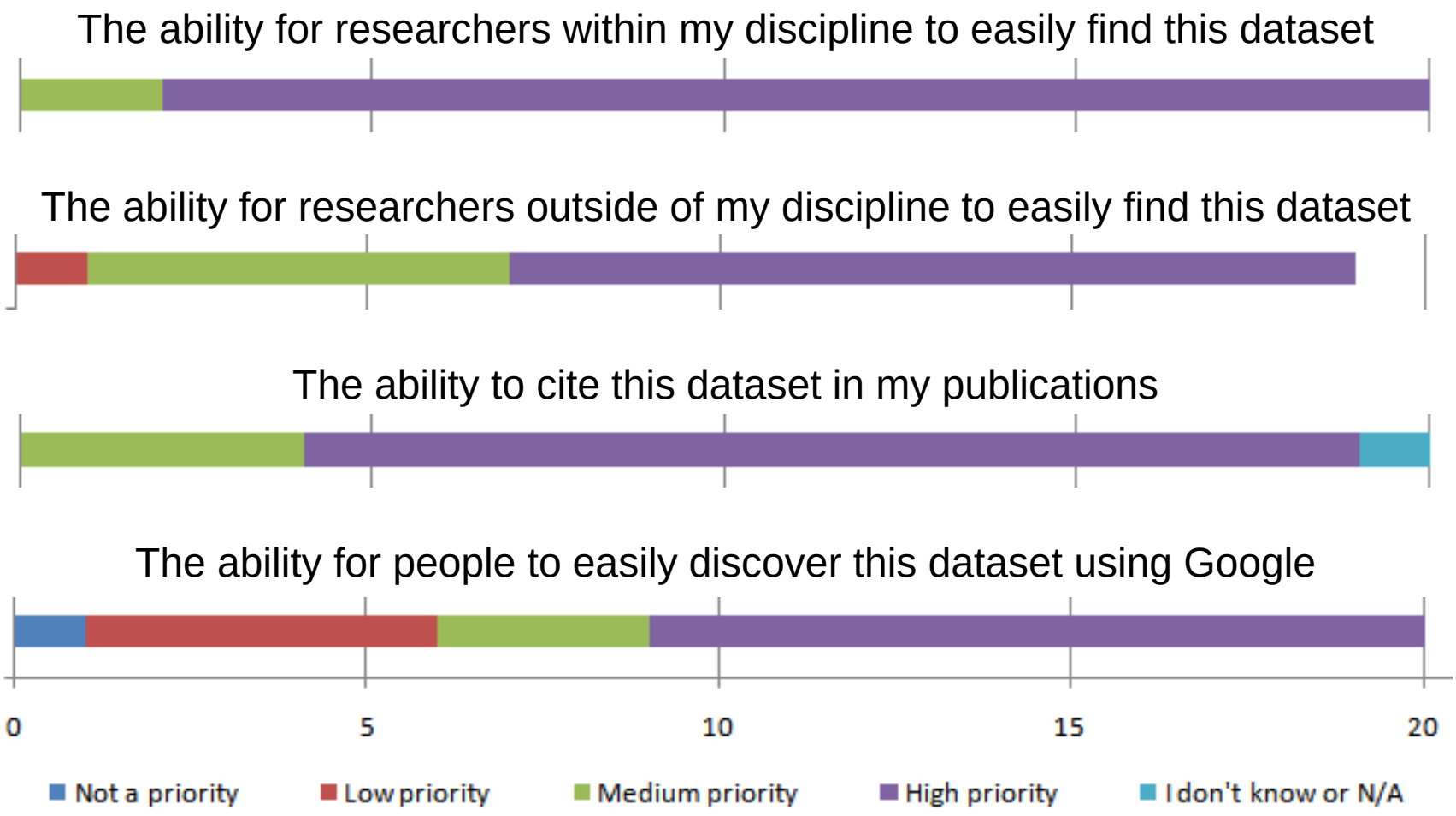
Embargo?


6%
13%
43%
13%
13%
12%
No embargo
1-3 mo.
4-6 mo.
7-12 mo.
1-2 yrs
2-5 yrs
5+ yrs
After retirement
After my death
n=16

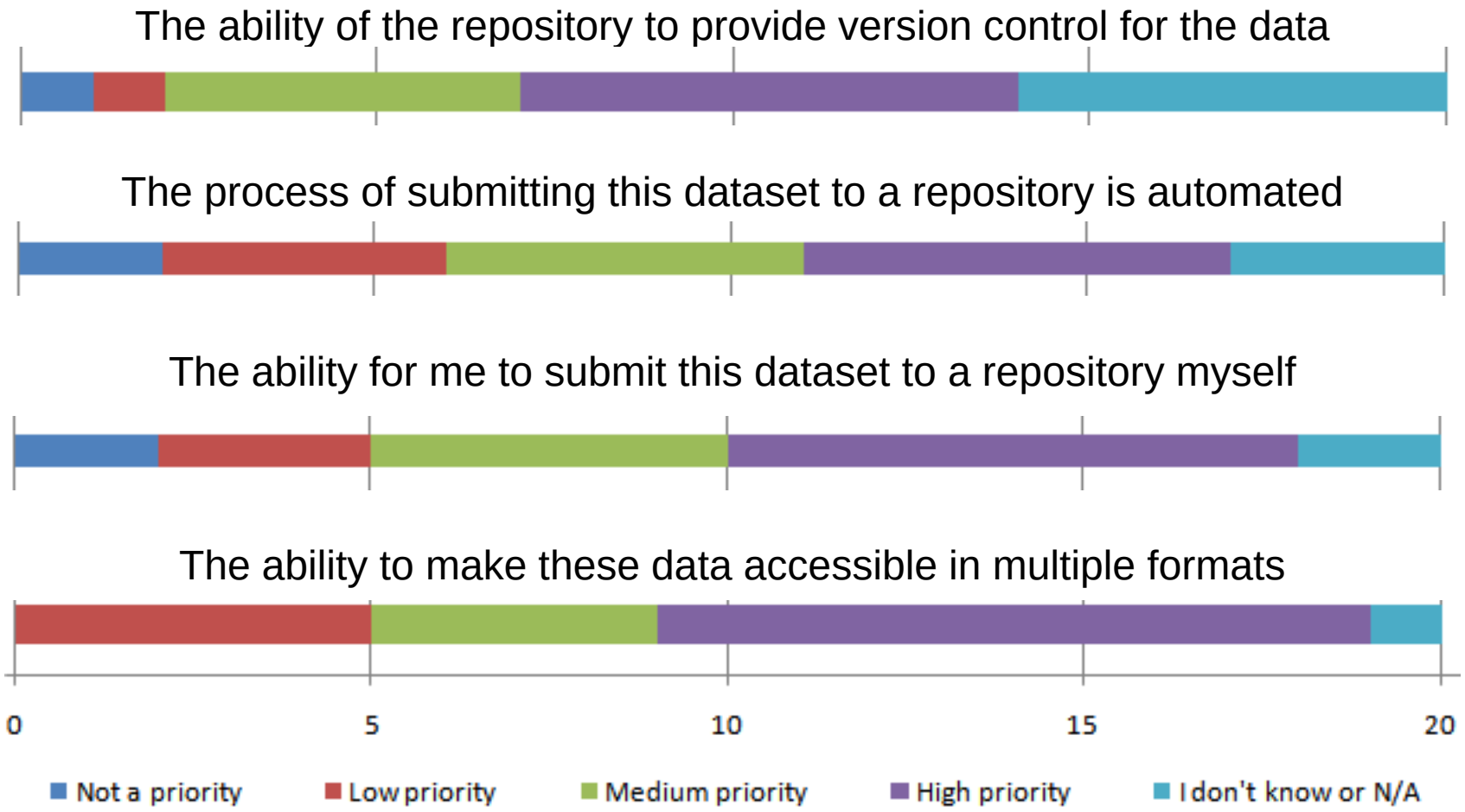
Prioritize your needs for the following types of services



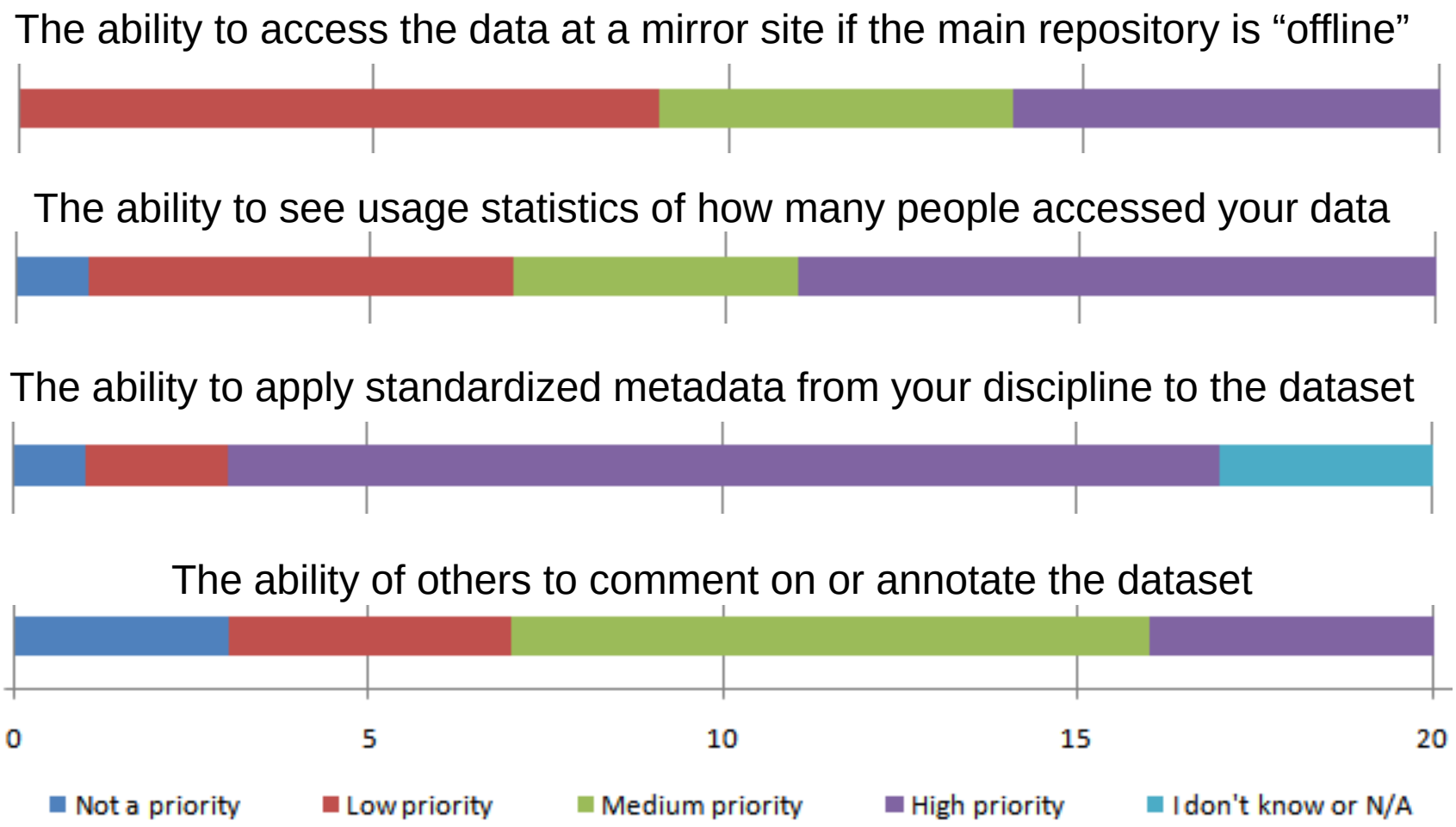
Prioritize your needs for the following types of services



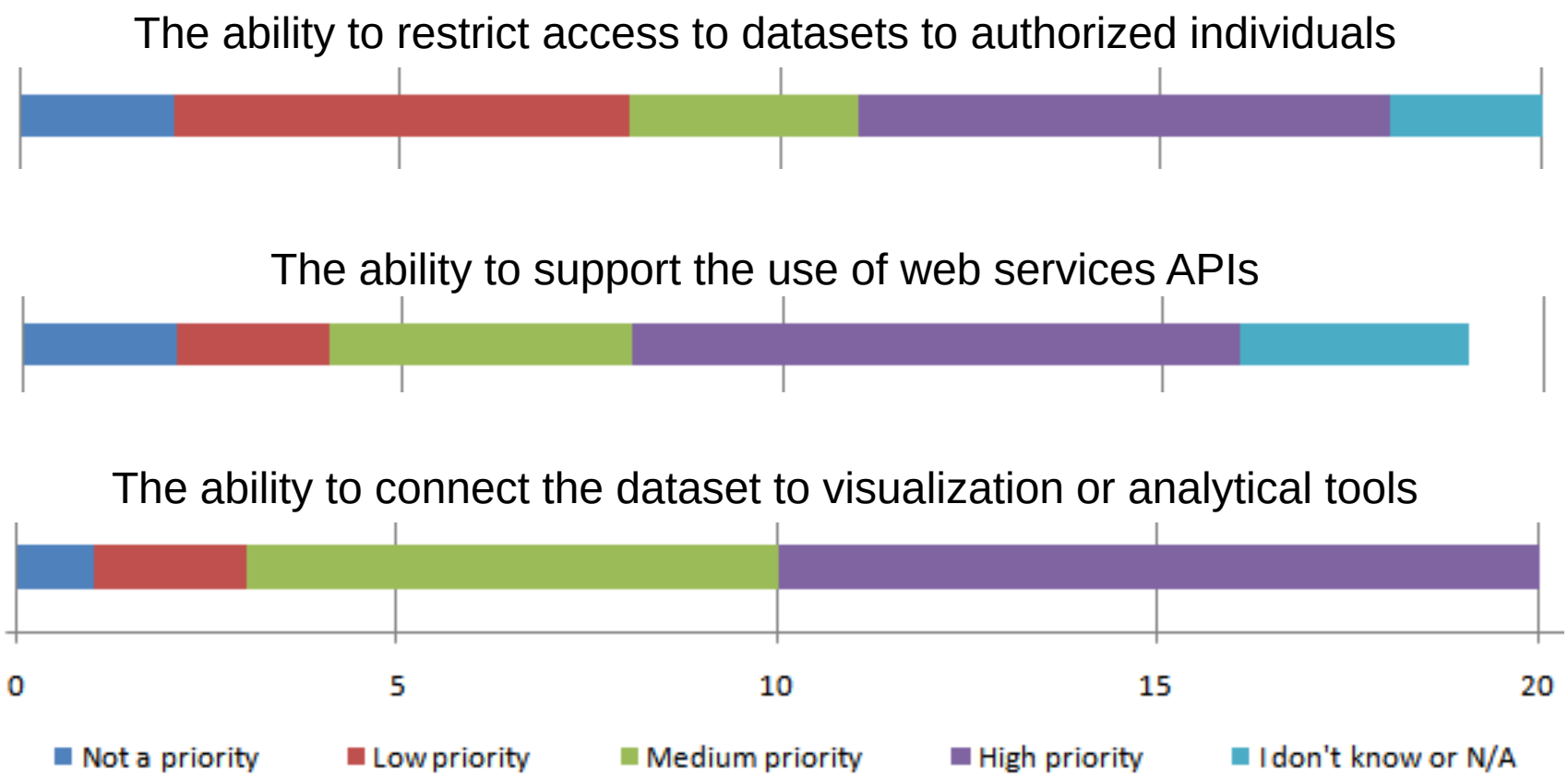
Prioritize your needs for the following types of services



Prioritize your needs for the following types of services



Prioritize your needs for the following types of services



Summary