



LANDSCAPE OF CURRENT SCIENTIFIC DATA POLICY AND PLANS IN FEDERAL AGENCIES: RESULTS OF THE PARTICIPANT SURVEY

Bonnie C. Carroll
CENDI Executive Director
J.R. Candlish, Technical Assistant

June 29, 2010

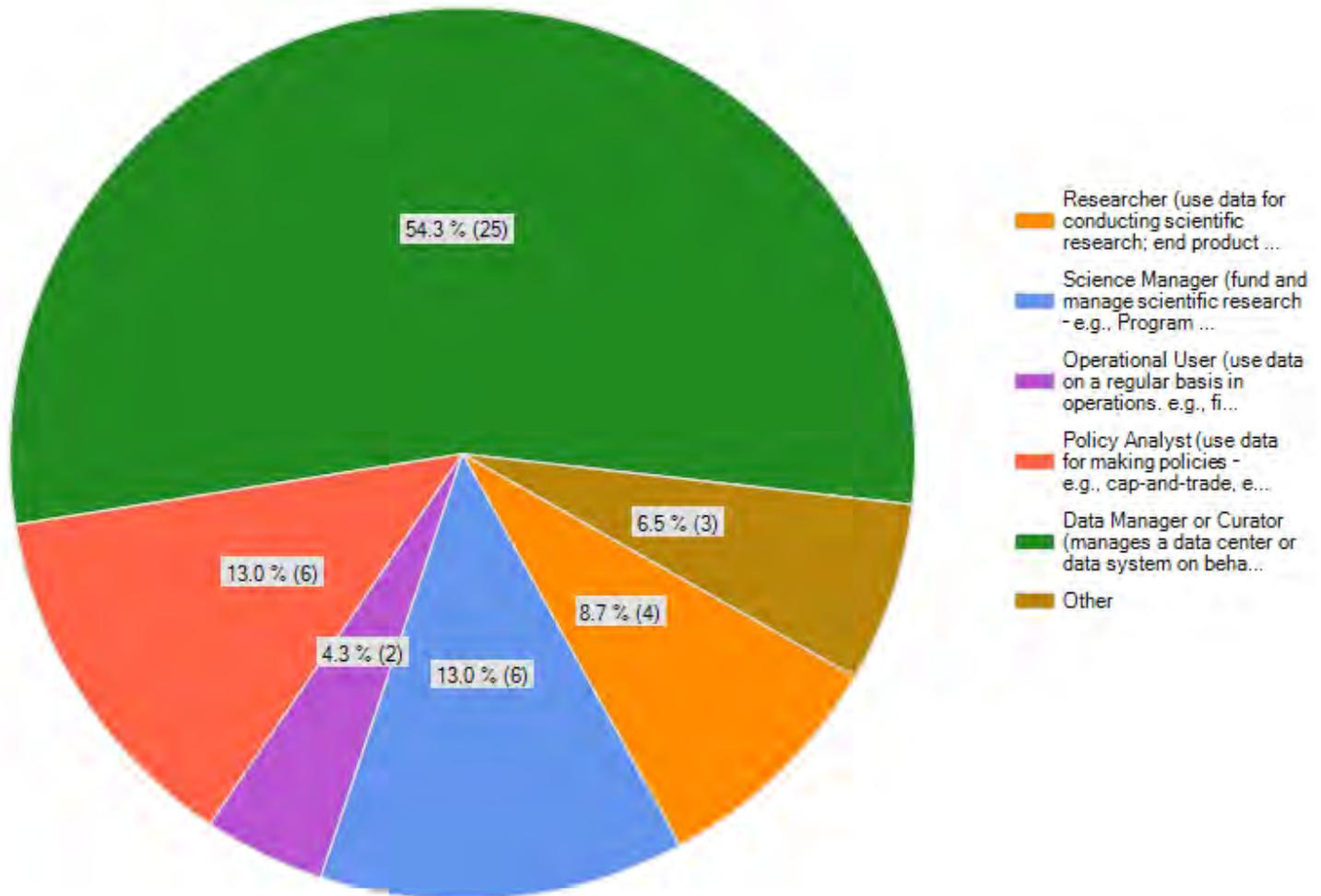
Purpose and Use of the Survey Results

- To provide common baseline of the current landscape to be used as input to break-outs
 - Everyone received all survey results
- This presentation highlights some of the findings
- Points for discussion in break-outs:
 - Developing strawman policy and plan elements (See Matrix)
 - Note the question number and refer to it in the full package each facilitator has

Survey Demographics

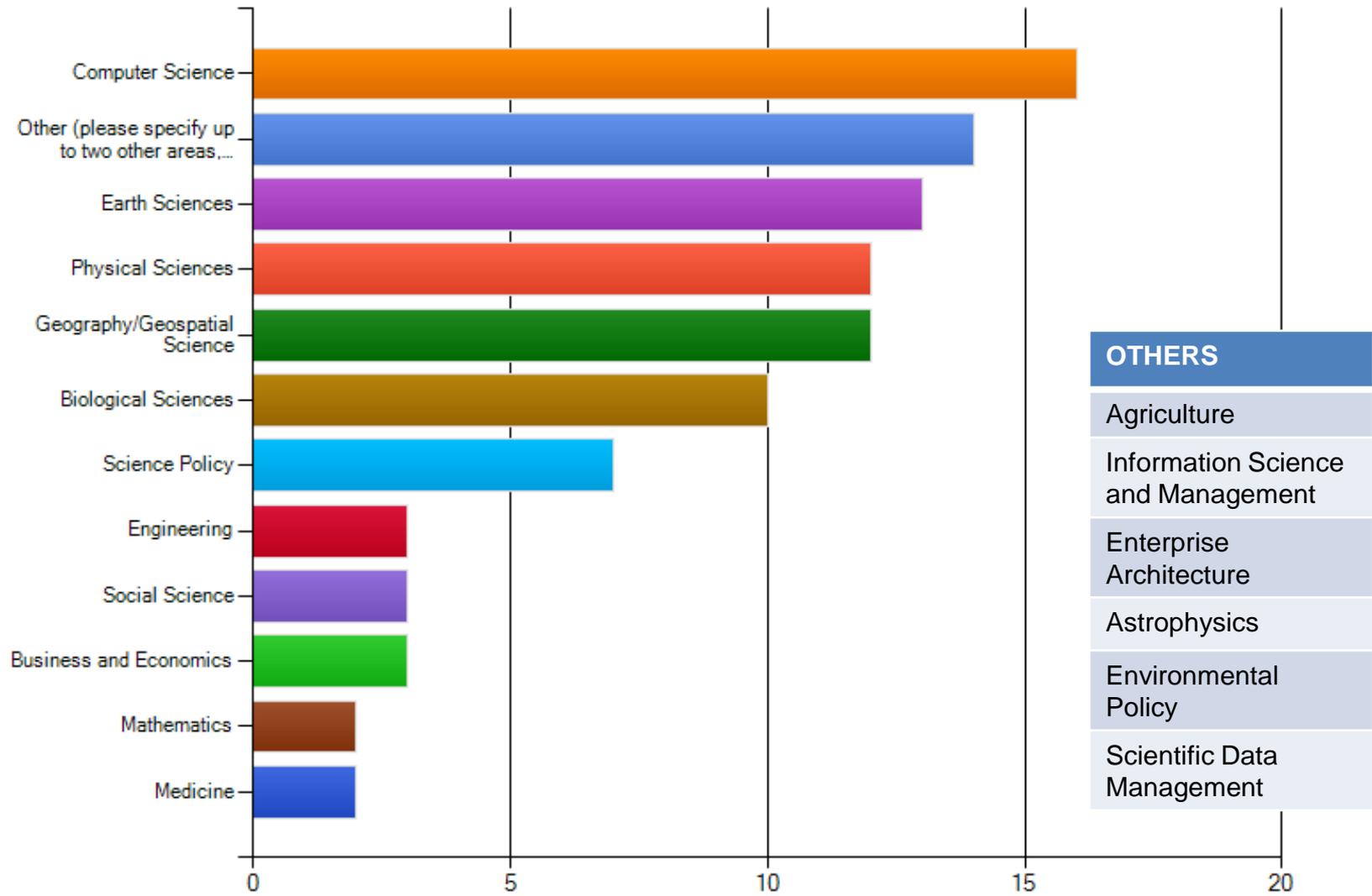
- 46 total surveys
- Mean of 35 active responses

5. To which data stakeholder group would you have your primary identification?



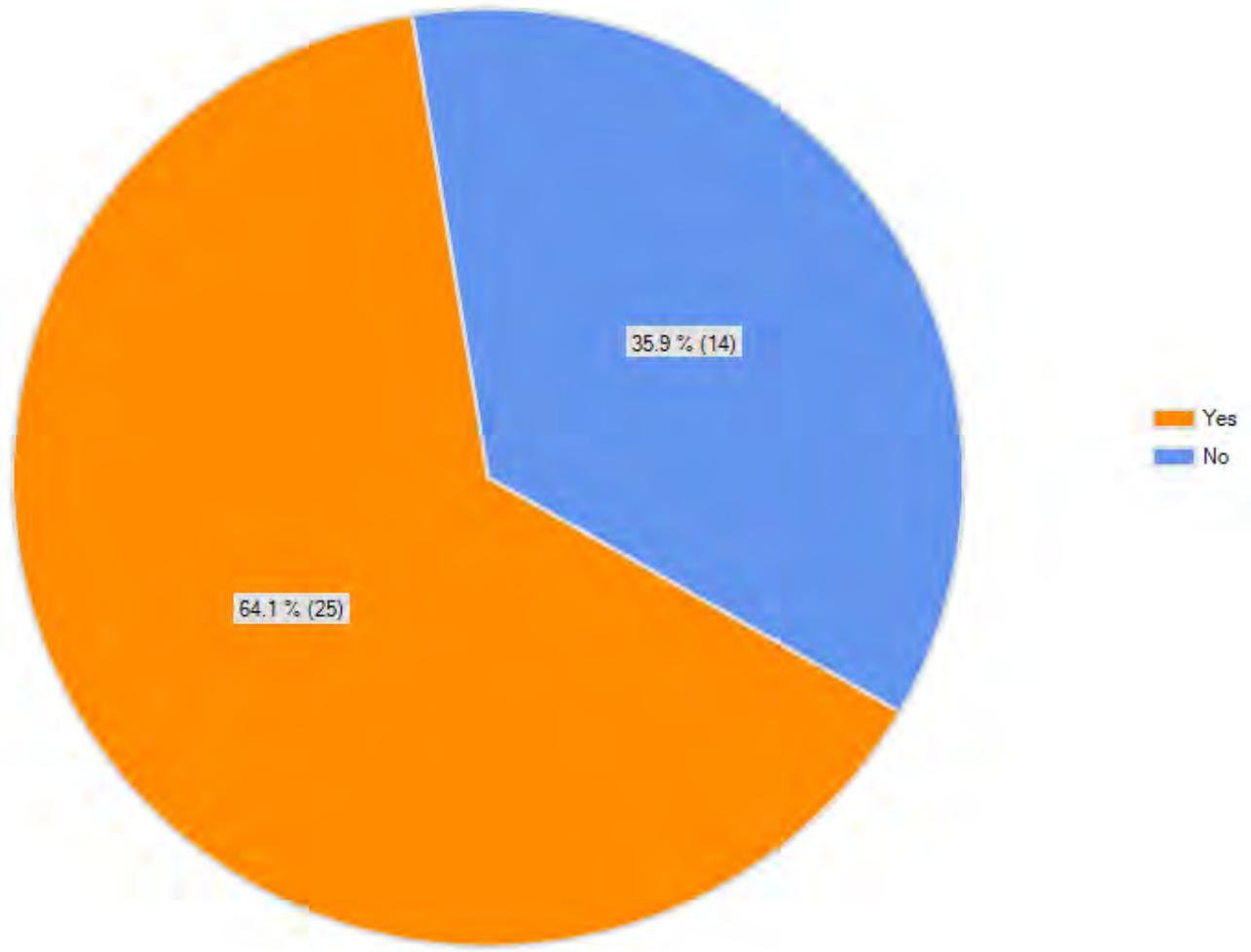
3. What disciplinary area do you work in?

N=46



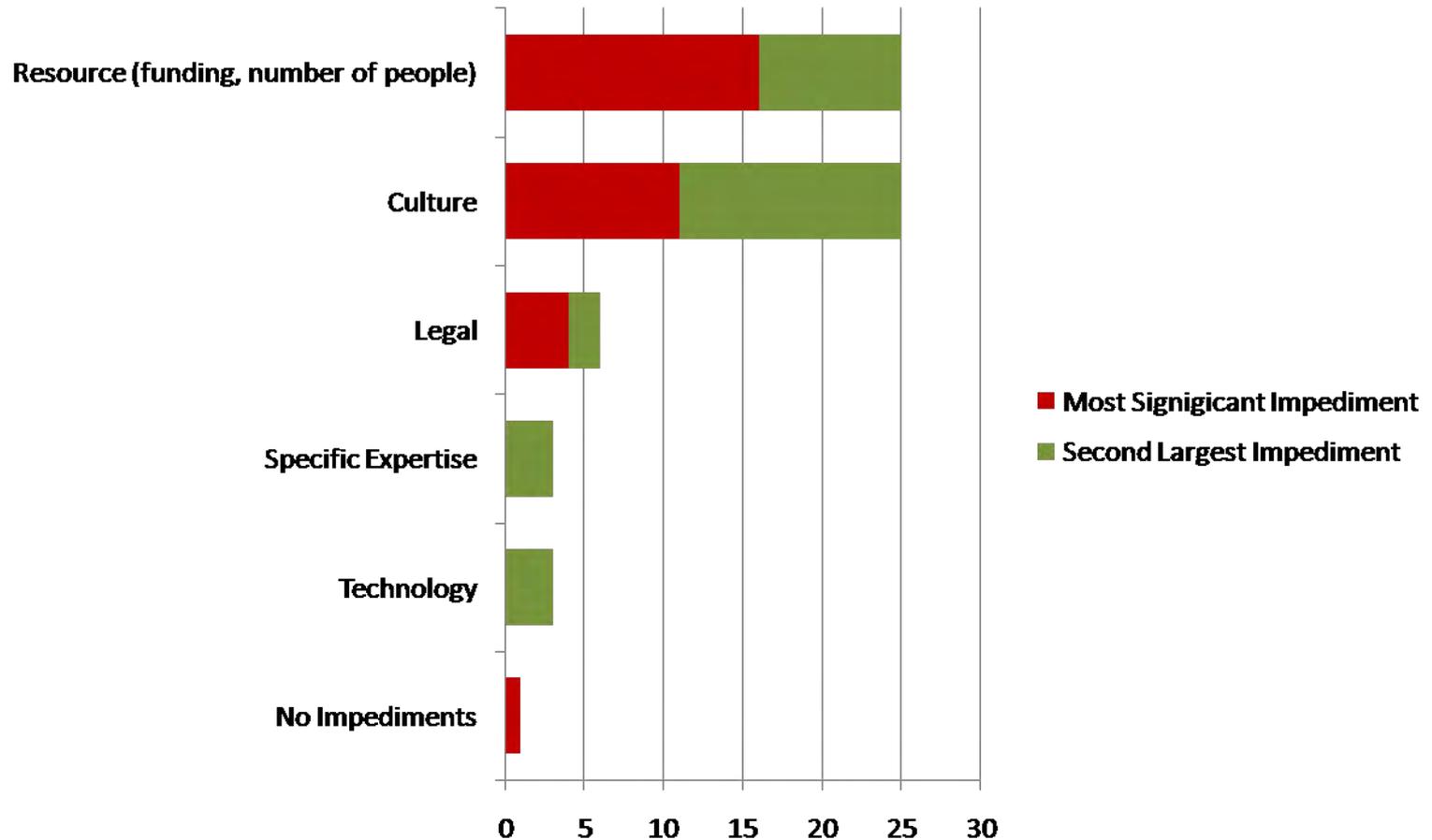
8. Does your agency manage data as an enterprise asset?

N=39

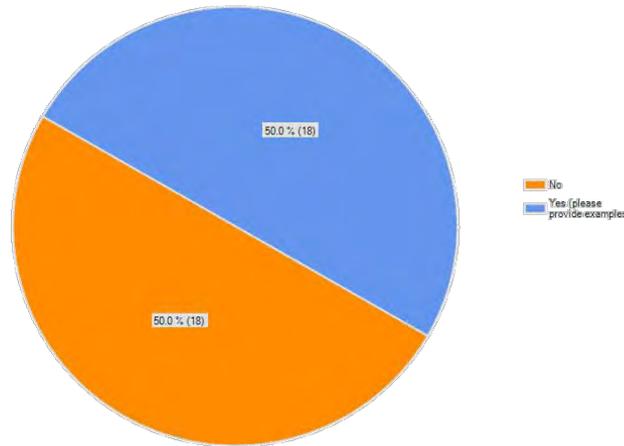


16. What are the key impediments you see, if any, to managing scientific data as an enterprise asset at your agency? Please indicate the top two barriers by labeling them chronologically, (1) & (2).

N=32



25. Are you aware of scientific data that no longer has a custodian but might need to be preserved for future use?



N=36

We refer to these as "orphans" where no programmatic sponsors exists. We have had to address several collections falling into this category. There is a group proposing that ICSU CODATA form a group to inventory data needing rescuing.

Large quantities of phenological data from individual observers. Various examples of data from professors who are nearing retirement.

Historic measurement data that have not been archived

Extra-mural research data is usually very poorly managed and preserved, and gets worse over time. It is usually left up to the principal researcher or a university archives that does not have the resources or skill-levels to maintain it

EPA emap data, - but no program is stepping up to manage the data

Retired field book data

Historic baseline information (4)
 -Apollo from NASA
 -Coastal aerial photos from 1970's

Data for projects terminated due to funding

People retire and data collected is not preserved

Lots of biological species inventories

Survey results which may need to be preserved for later verification of results, or for possible repurposing to new experimental uses

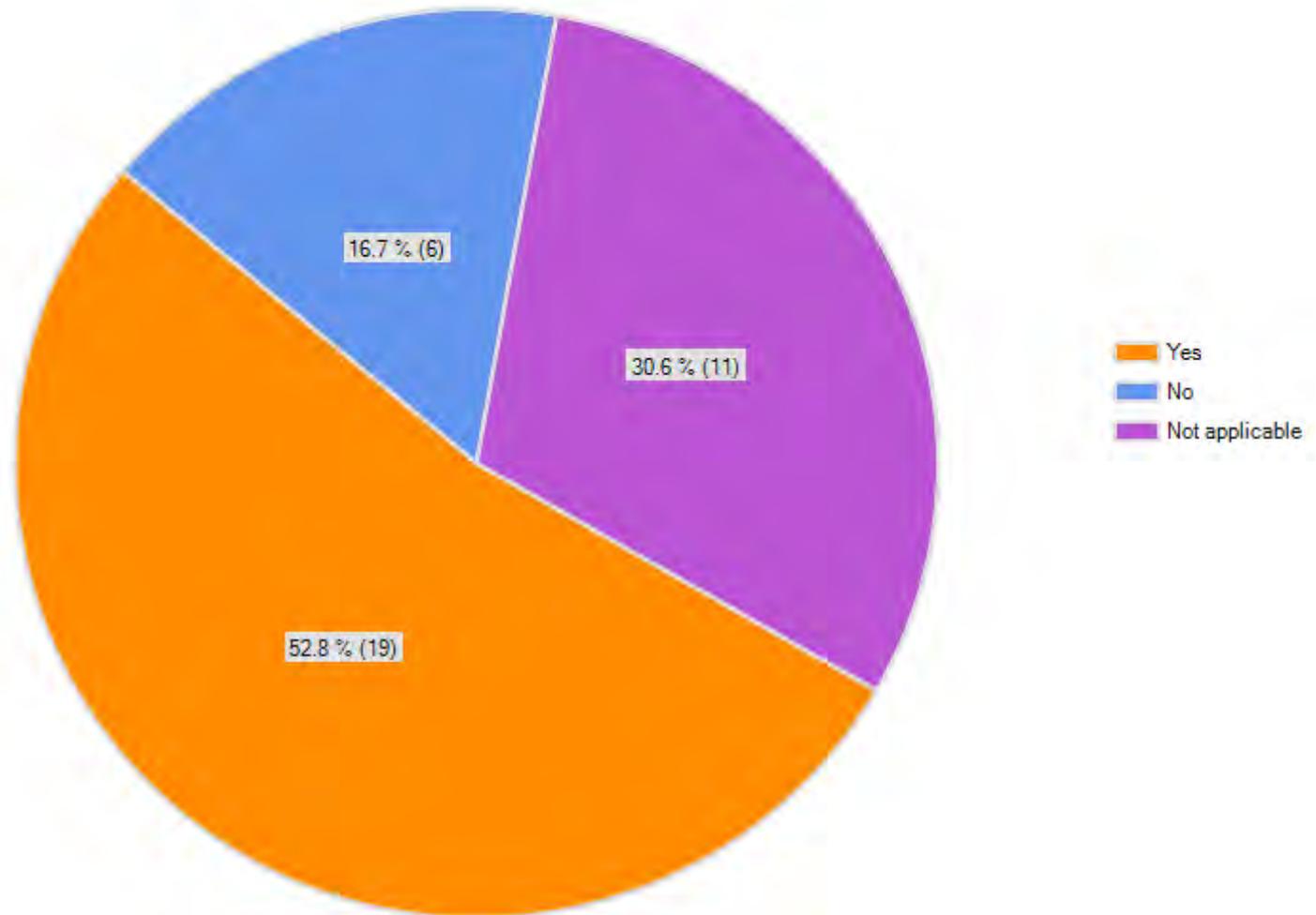
Sometimes, our data managers will come to us with data sets that should be archived at our NASA ES data centers. They are reviewed in a formal process for decisions about whether they should be included.

Technical Reports

Manual weather records at a county park that are going to be recycled.

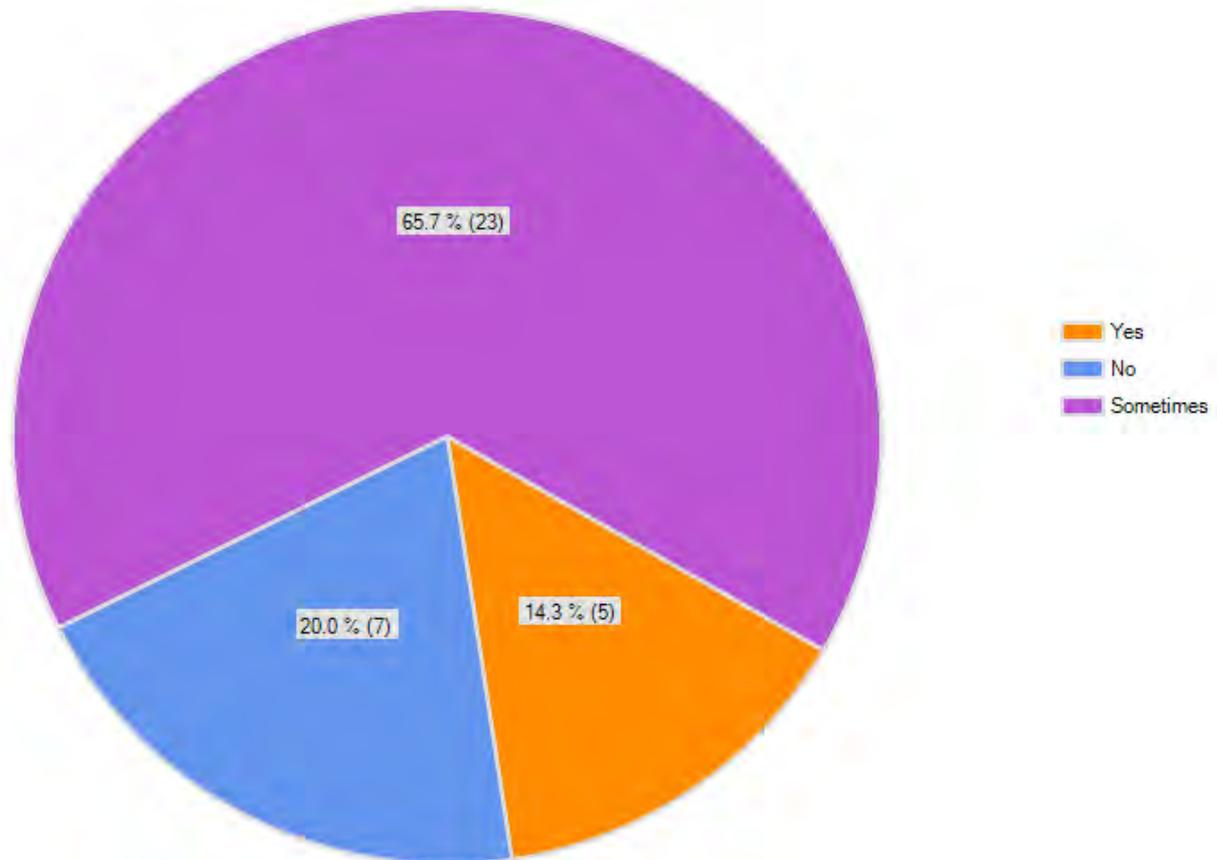
39. In the operational environment have you encountered problems in accessing data that you needed because of controls placed on the data by data producers/providers?

N=36



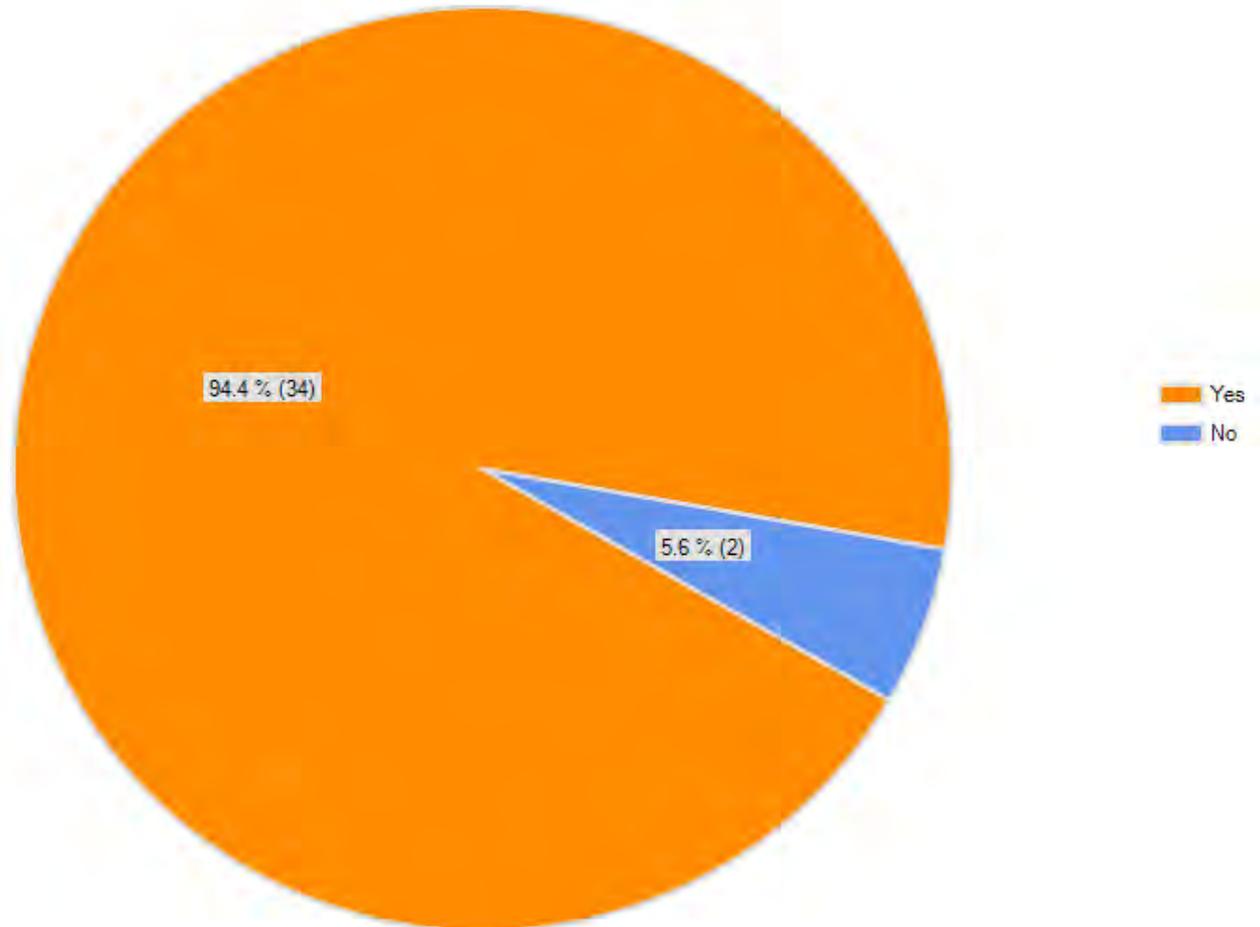
57. In planning projects or a research program do you have sufficient knowledge of data sets and other information products that are available or will be produced by other relevant efforts either in or outside your community of practice?

N=35



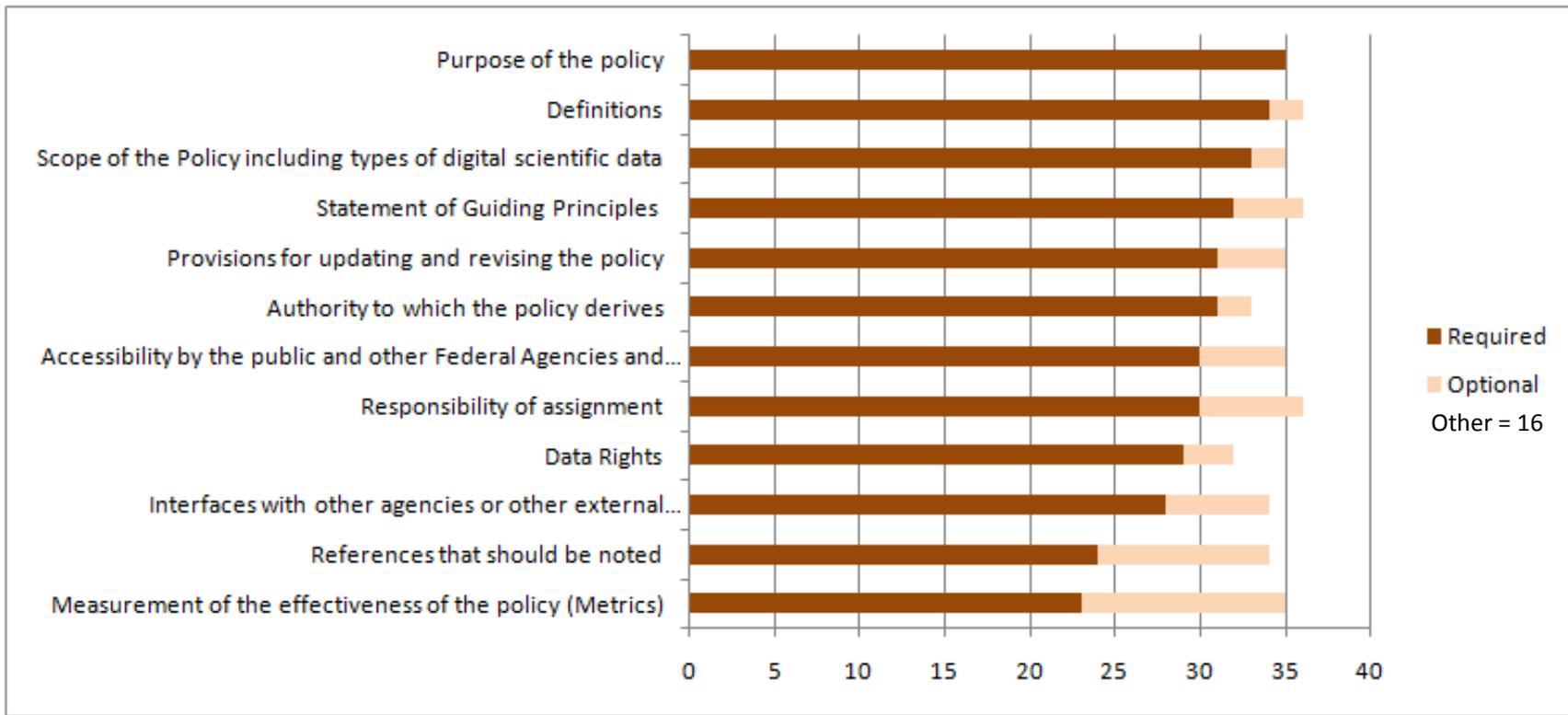
9. Considering what would work in your community, is there some level at which a policy could be crafted that you believe would work for your entire agency?

N=36



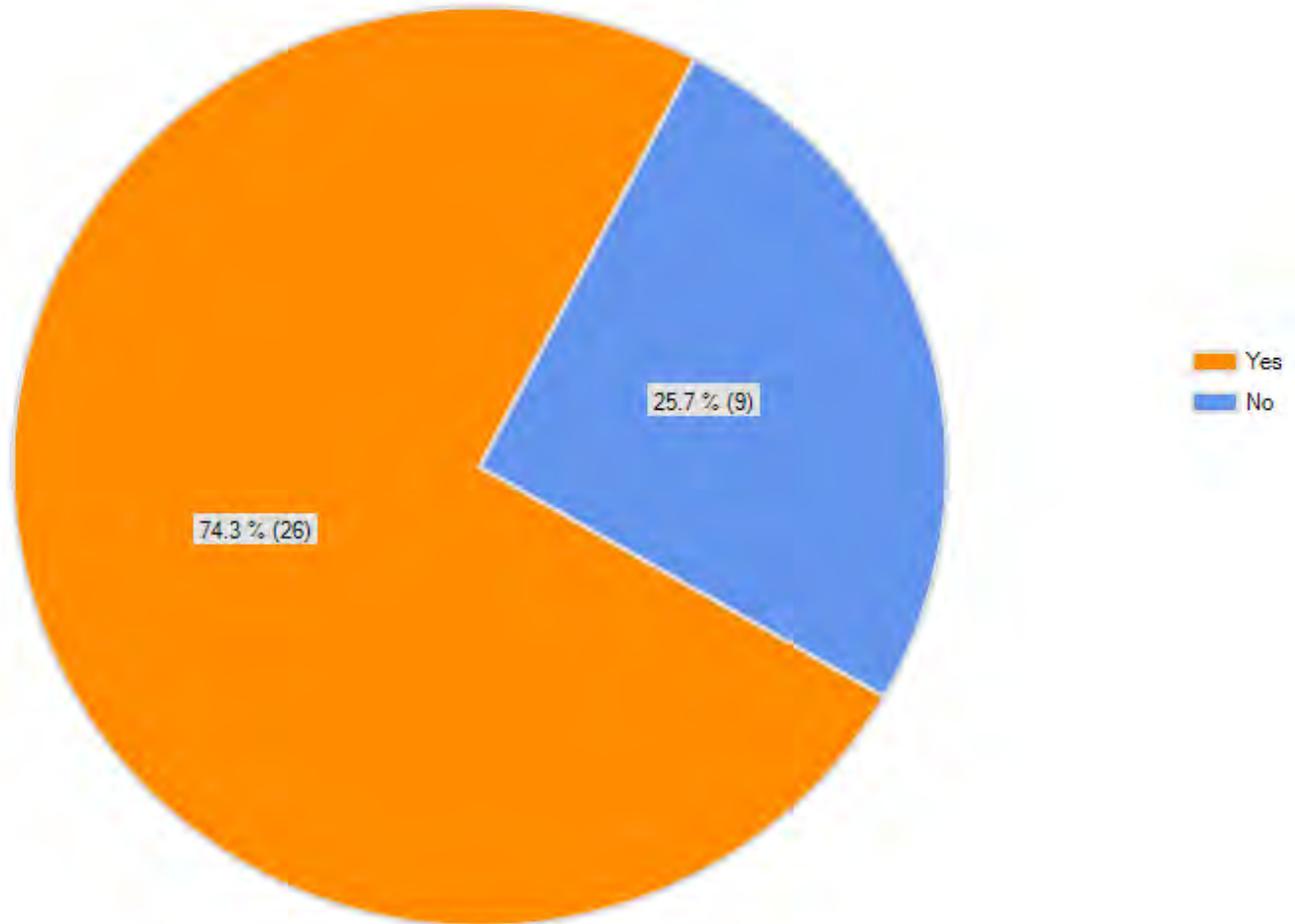
10. If you think about what elements should be in an Agency level Data Management Policy, please check all that you would include as either necessary or optional. Then add any elements not on the list.

N=37



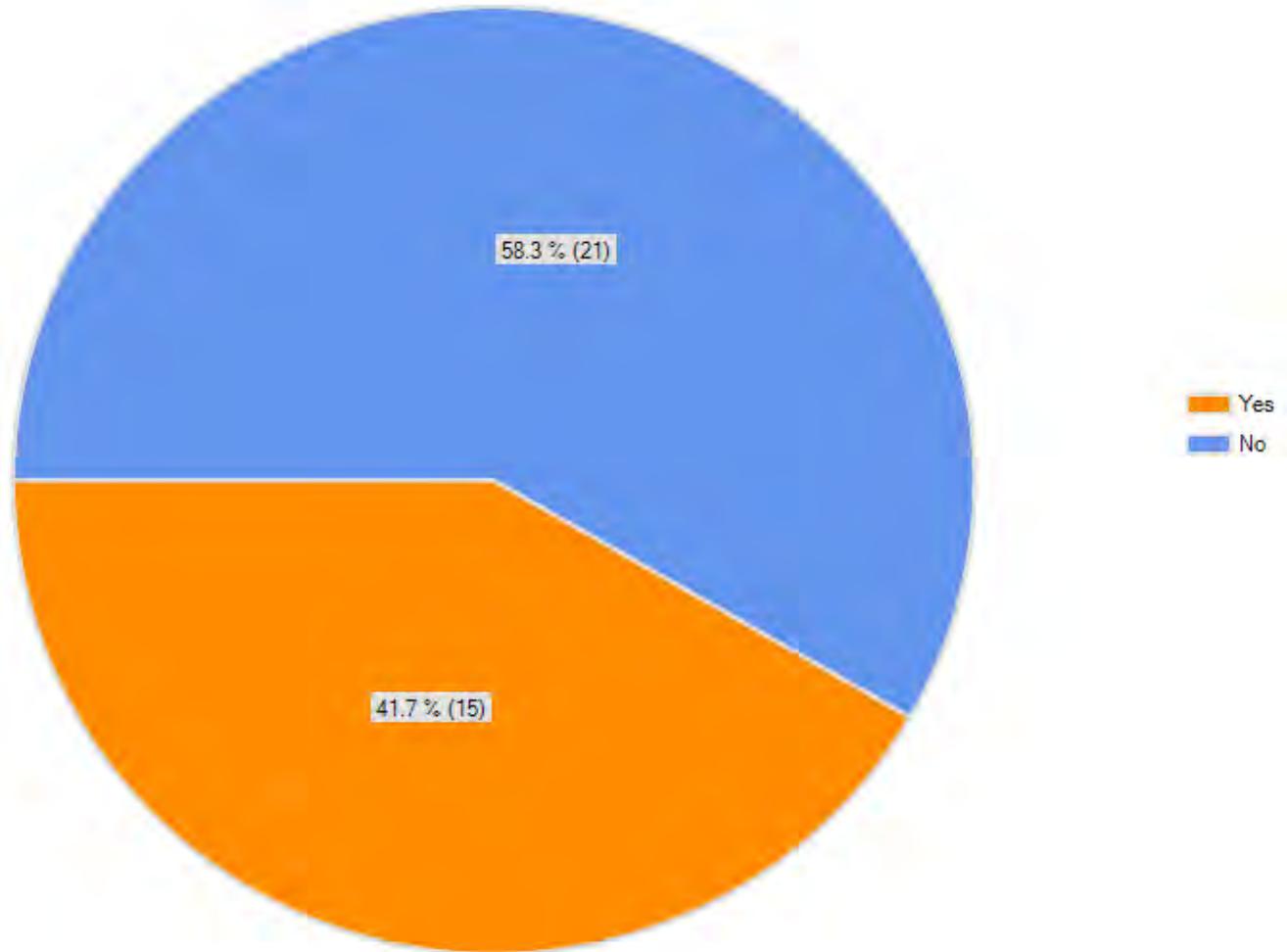
15. To be useful for operational users, should an effective agency policy ensure that datasets useful in the operational environment (e.g., those used in homeland security, environmental disasters, meteorological forecasting, etc.) be treated differently than datasets not deemed critical?

N=35



14. Should there be different expectations for a data management plan for data generated from extramural research as opposed to in-house research?

N=36



14. Should there be different expectations for a data management plan for data generated from extramural research as opposed to in-house research?

14 . A. If Yes, what would be different?

More stringent data quality, provenance, etc. evaluation; possibly reduced commitment to preservation

When data is available to other users, who is responsible for maintaining data, etc

Should be fully open and publicly available either:

1) after publication, or 2) after a set amount of time if the data is not published. Requires the creation of a data repository.

More detail on data rights since they are more complex

Potential conflicts between the agency and the extramural organization's data management policy would need to be resolved in data management plan

Data from alternative sources would require flexibility to ensure access to all relevant data sources. A "minimum" expectation may not match the host institutions plans.

I would expect greater rigor for extramural than for in-house where in some cases I would expect exploratory work to be carried out.

Managing data can be VERY EXPENSIVE. When the cost to manage the data is more than the cost to generate the data, then there is a problem. The scope of the 'plan' needs to be adaptable.

Data from Extramural research results would likely be transferred to institutional (government) organization for preservation and longer term management. So the plans should cover when and how such transitions occur.

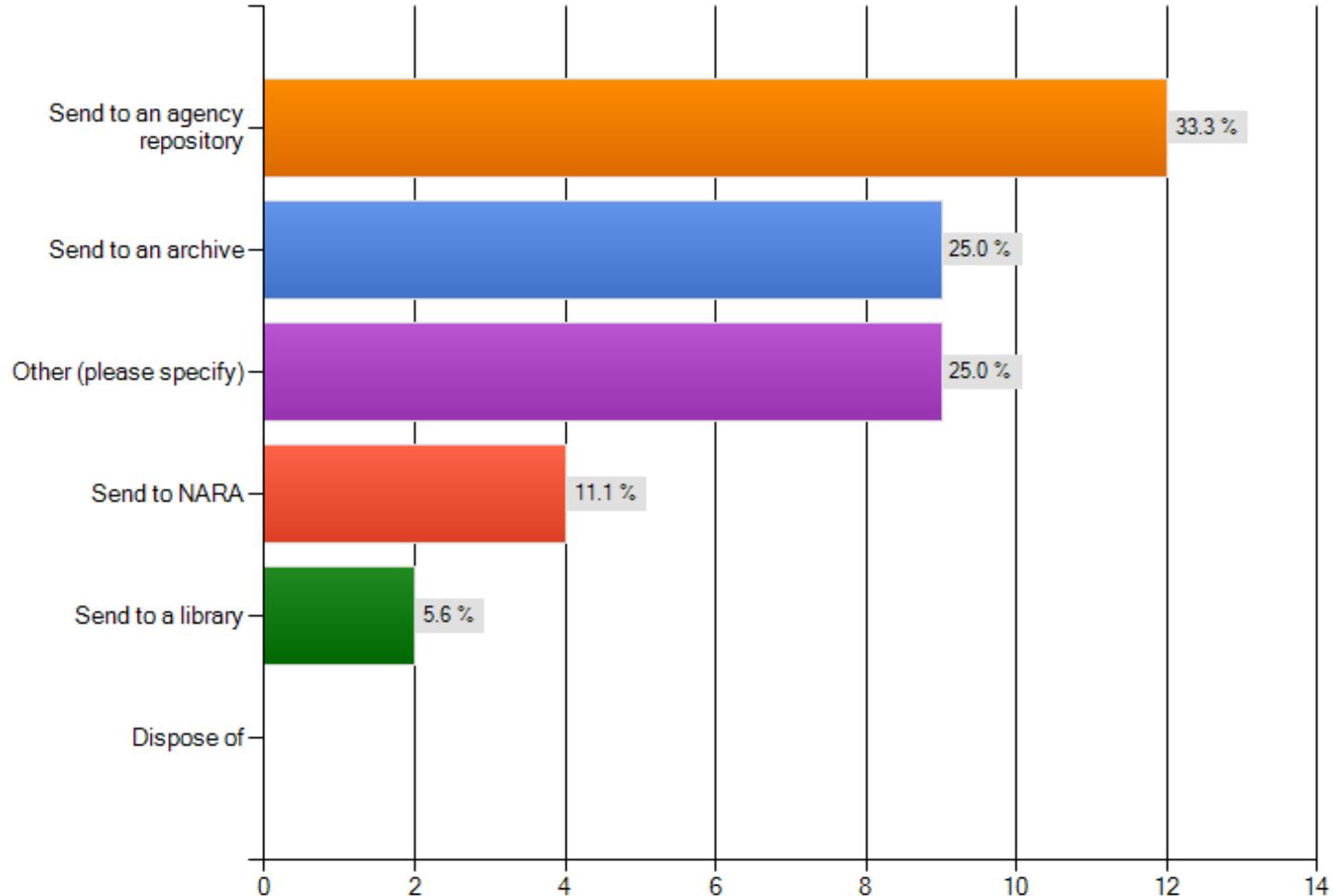
Only if it is clear that another party/cooperator/funder has the responsibility for data management.

Data management roles for managing extramural data may be best managed by the extramural researcher. Links between the in-house research management systems and the extramural system should be developed.

Data management that is external to science. For example, our data management includes descriptive information of our research facilities, and history of research performed at location. There may be stronger requirements to document methods, landscape, etc... because there would not be in-house understanding upon completion of a project. If a University performs some field tests, we may require greater documentation than if it was performed on one of our fields. Links to publications, links to biographies of scientists

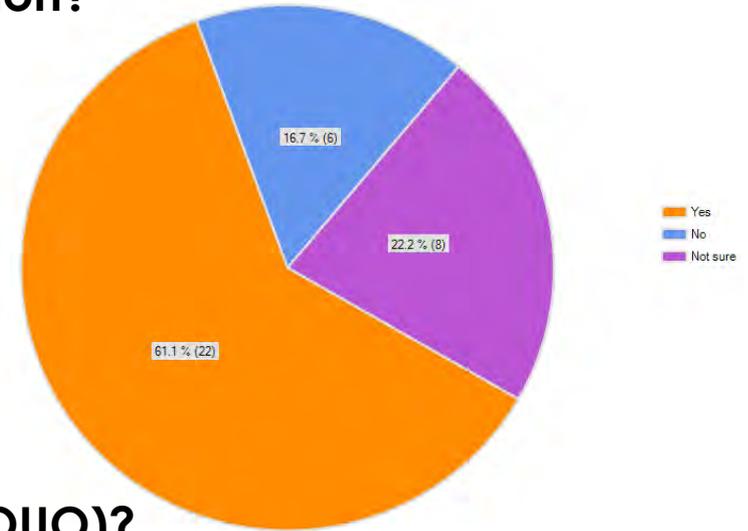
19. At the end of that time period what should be the disposition of the data and management responsibility when you no longer wish to maintain it?

N=36



37. In your community of practice, do you have controls for unclassified but restricted information?

N=35



38. What categories do you use with marking used (e.g. Official Use Only-OUO)?

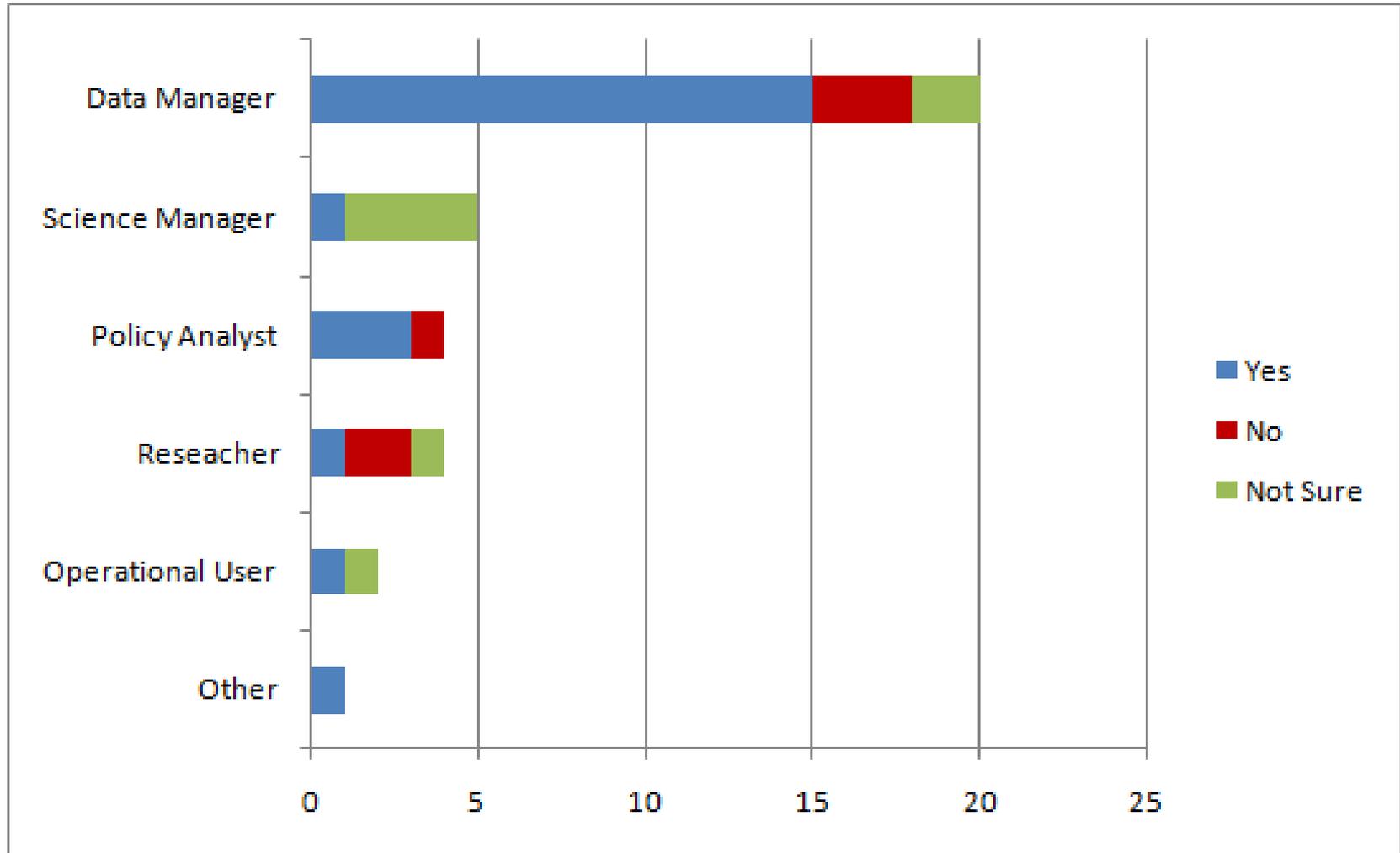
N=17

SBU (Sensitive But Unclassified) (6)
Agency Deliberative
CBI (Confidential Business Information) (3)
Contract sensitive
Deprecation
FOUO
OUO (4)
Proprietary (4)

ITAR
Secret
Agency Only
Competition Sensitive
Program Use Only
Regional Use Only
see DoD Directive 5230.24, Distribution Statements on Technical Documents

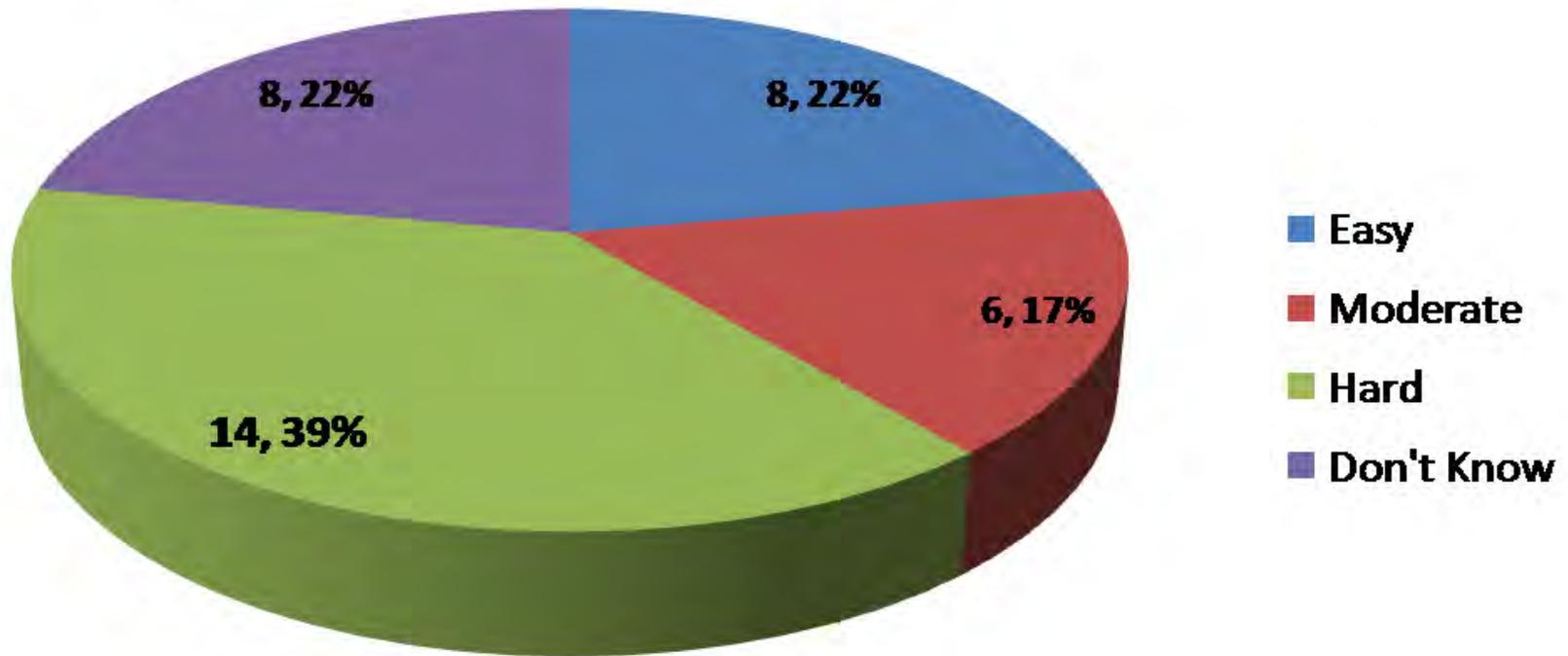
37. In your community of practice, do you have controls for unclassified but restricted information?

N=36



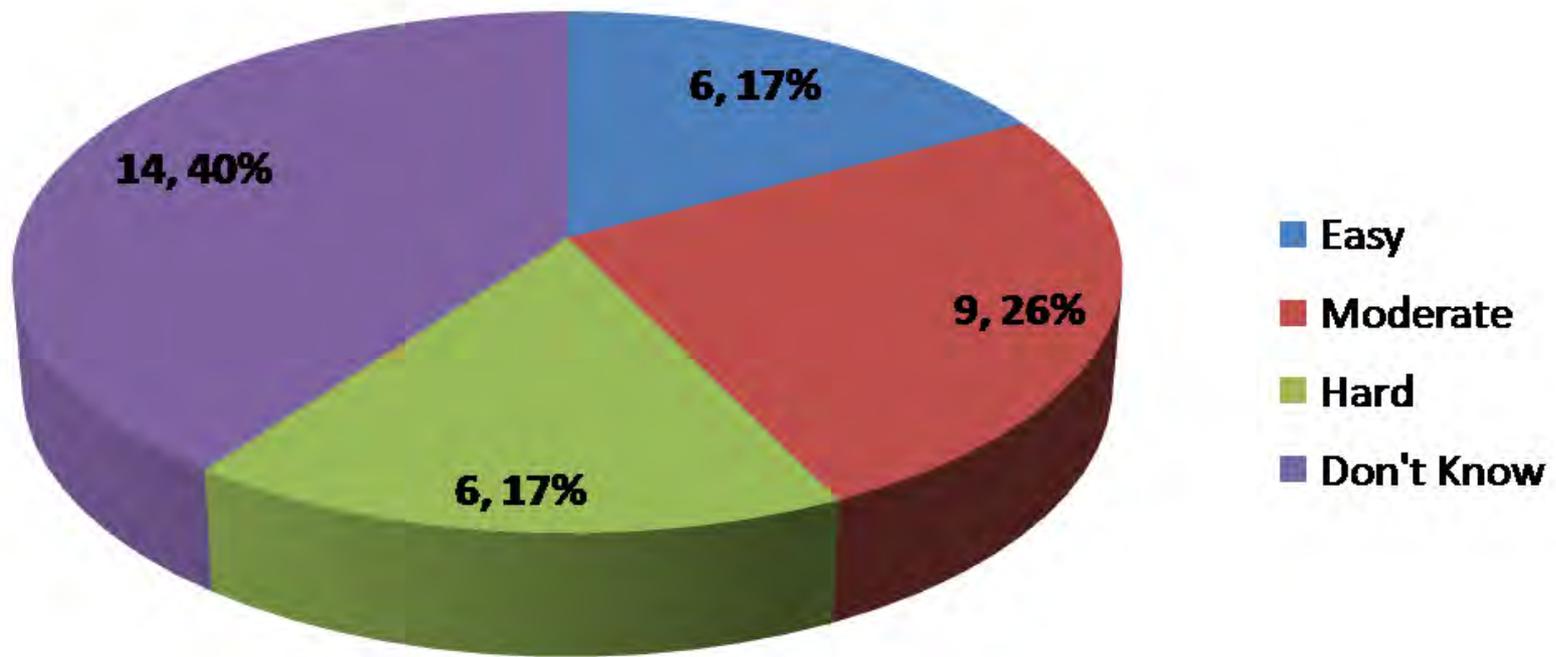
40. How easy is it to know about rights and restrictions on the use of data?

N=36



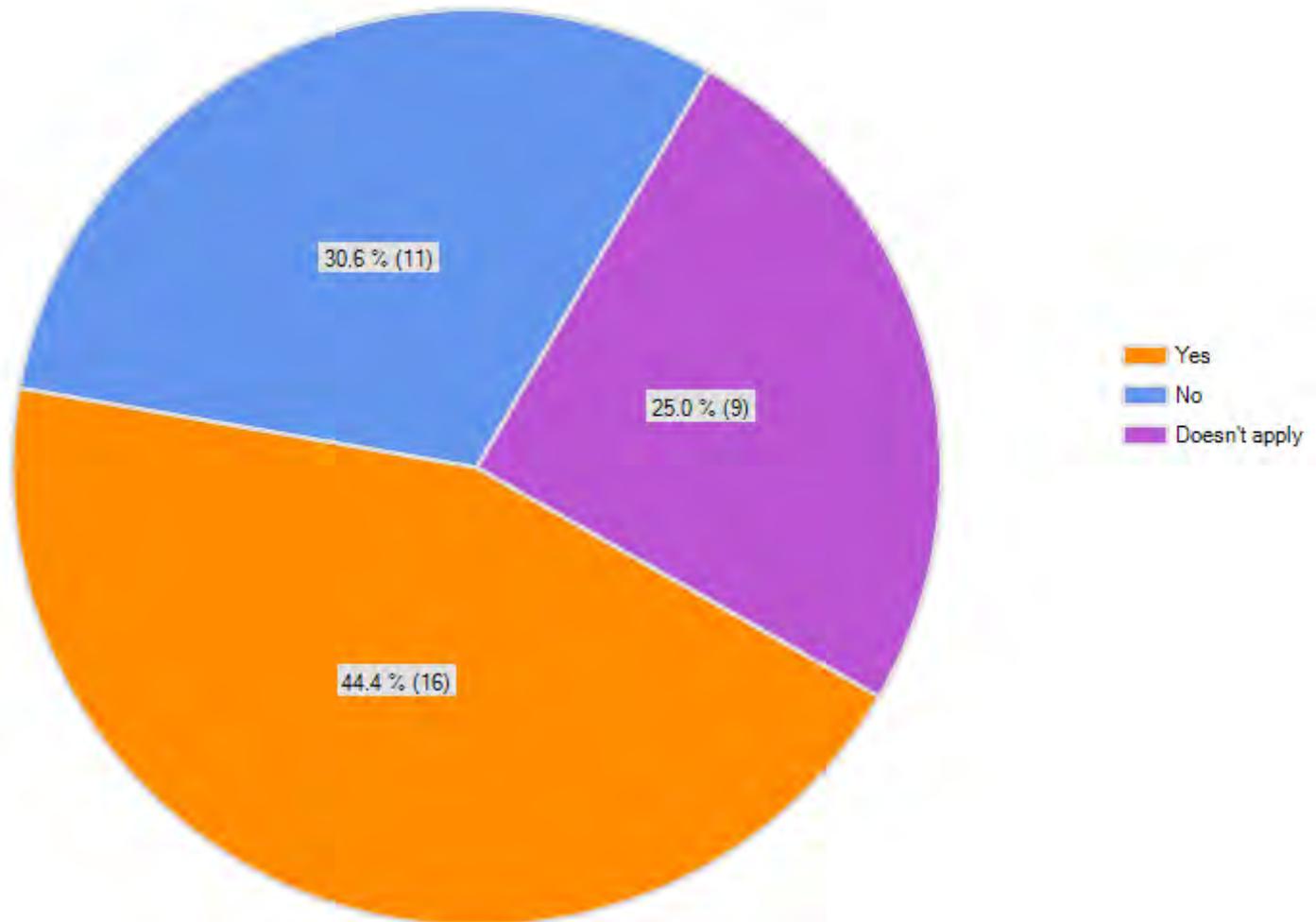
41. Taking cultural issues aside, if an Agency places formal restrictions or controls on data you need (e.g. unclassified but restricted such as Confidential business information), how hard or easy is it for you to implement the controls put on by other agencies?

N=35



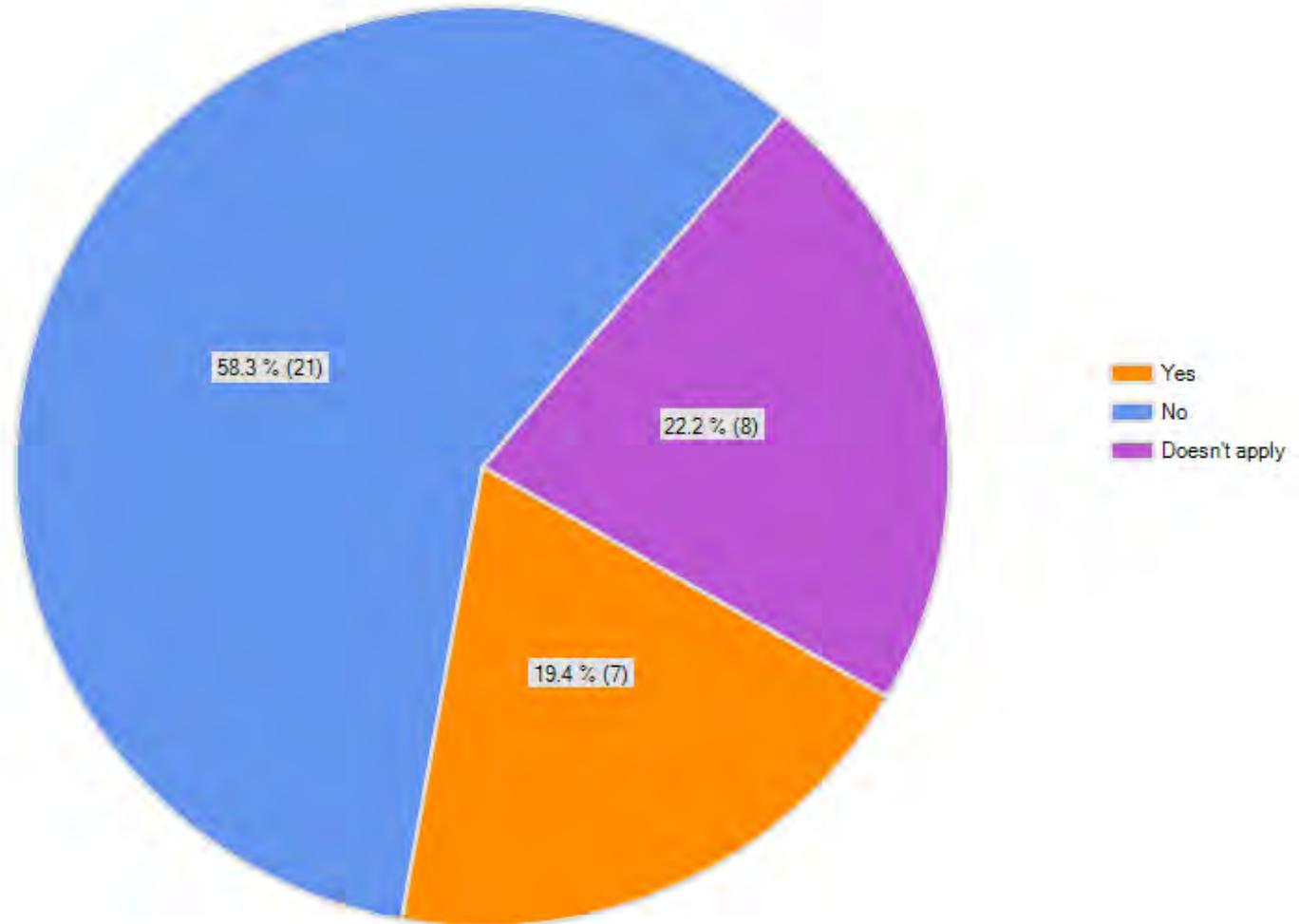
43. Do Embargo dates (Dates that a researcher can keep data for their own analysis before public release) apply to datasets that are actively managed by your community?

N=36



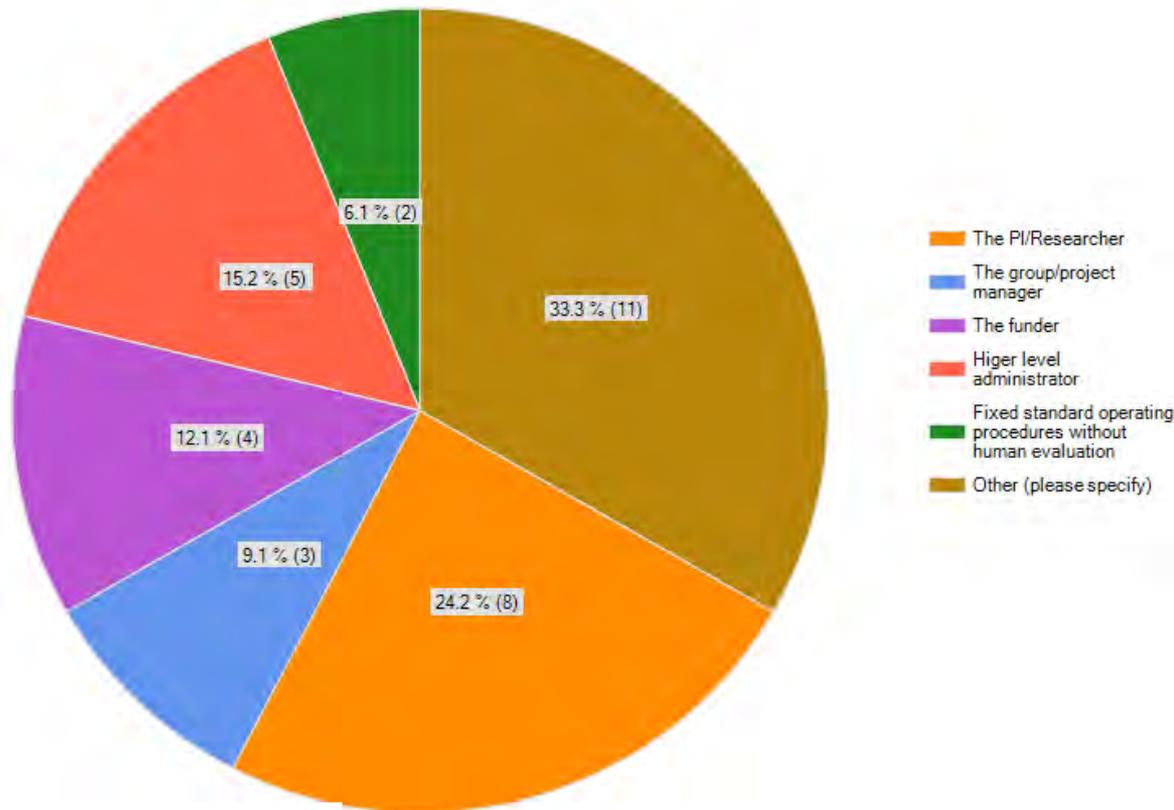
44. Do Sunset dates (Dates where data sets can be disposed of) apply to datasets that are actively managed by your community?

N=36



52. In your agency, who decides if and how long data will be retained?

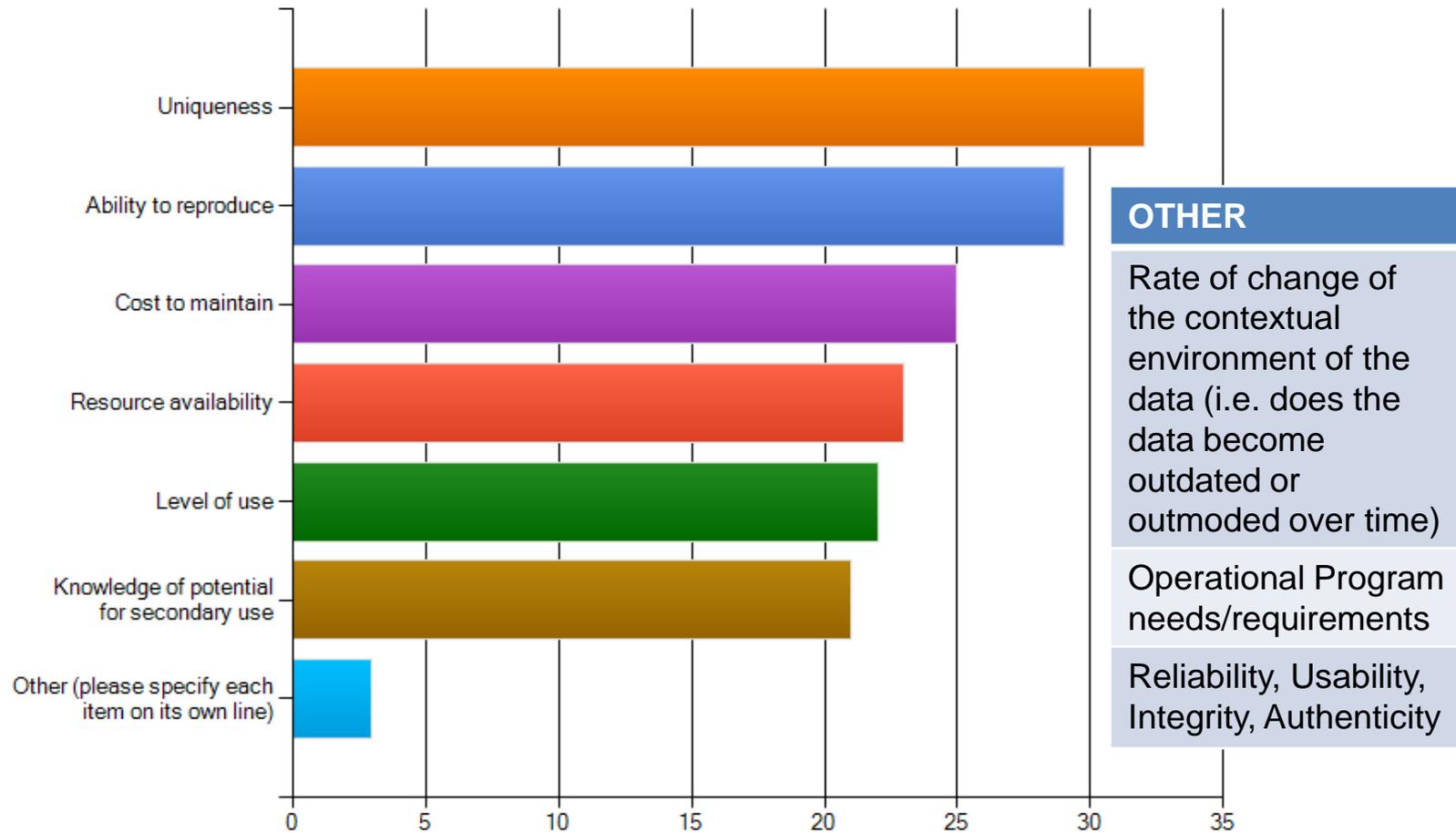
N=33



OTHER
Combination of fixed SOP, funding entity, PI, and data stewards/archivist
Peer advisory panels
Established record schedule for that data
Combination of NASA, Data Center managers, and users
Researcher decides "if"; archive decides "how long"
Appraisal archivists
Records management schedules for scientific data

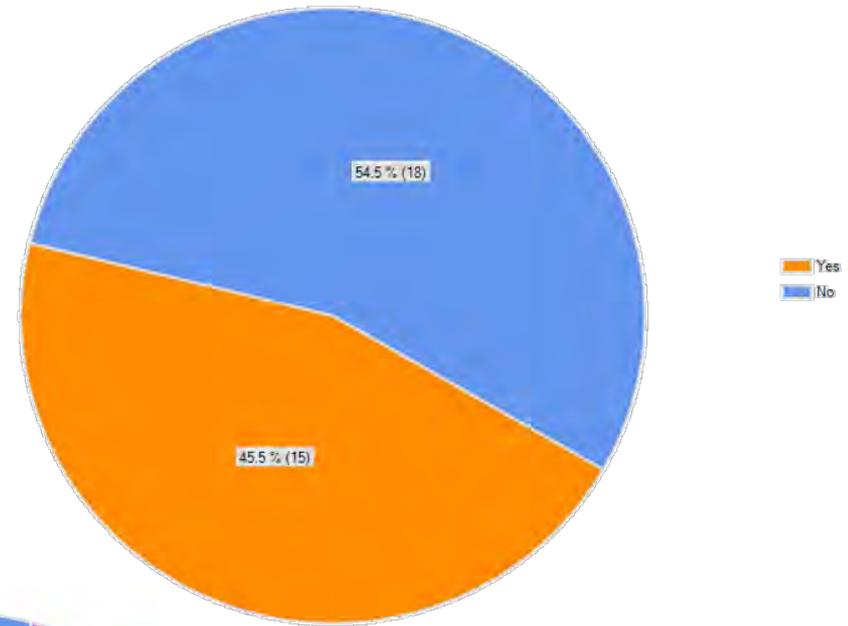
51. What attributes of data should be evaluated to determine the length of time it should be retained? Check all that apply.

N=31



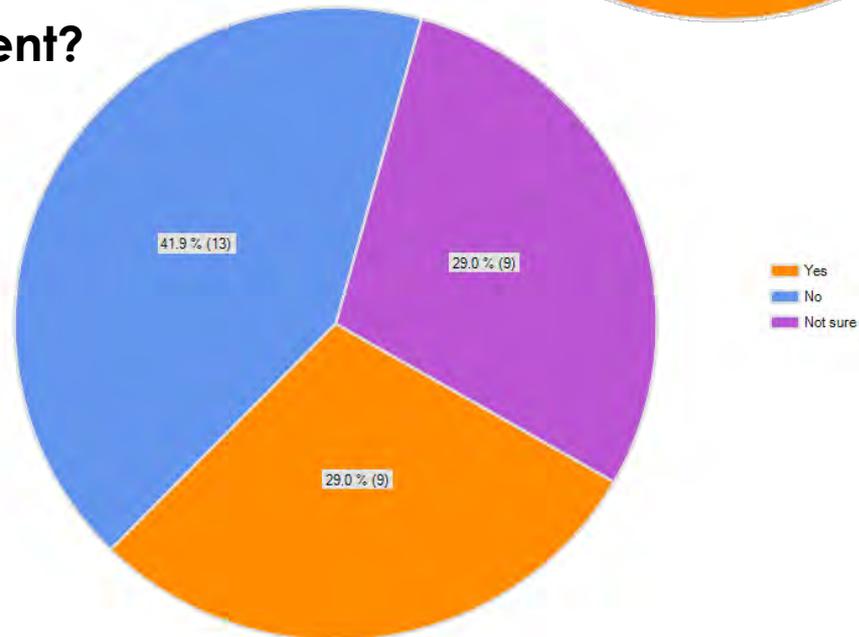
49. Within your community are there processes to evaluate how long data sets should be maintained and kept available based on their perceived value?

N=33



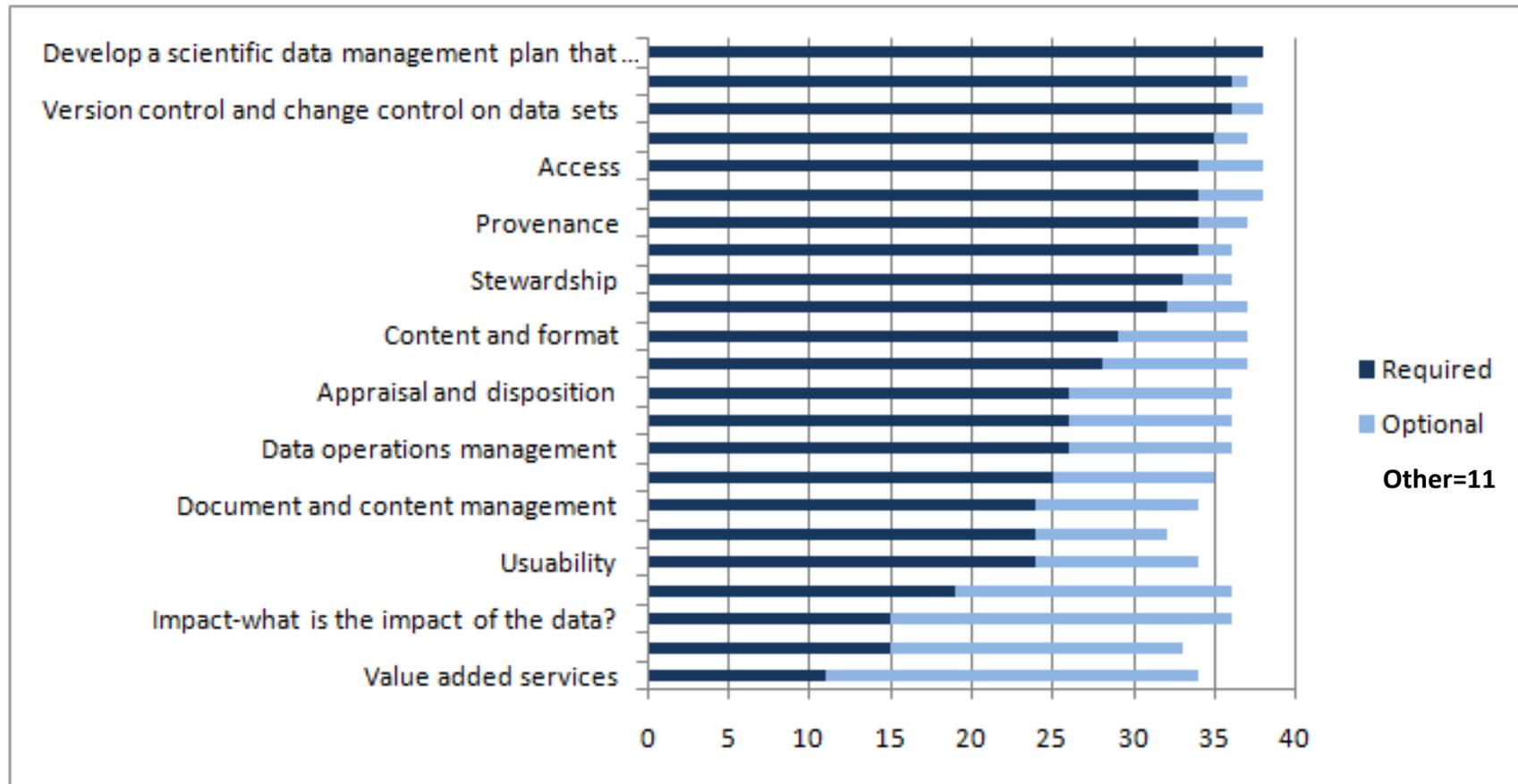
50. Are they sufficient?

N=31



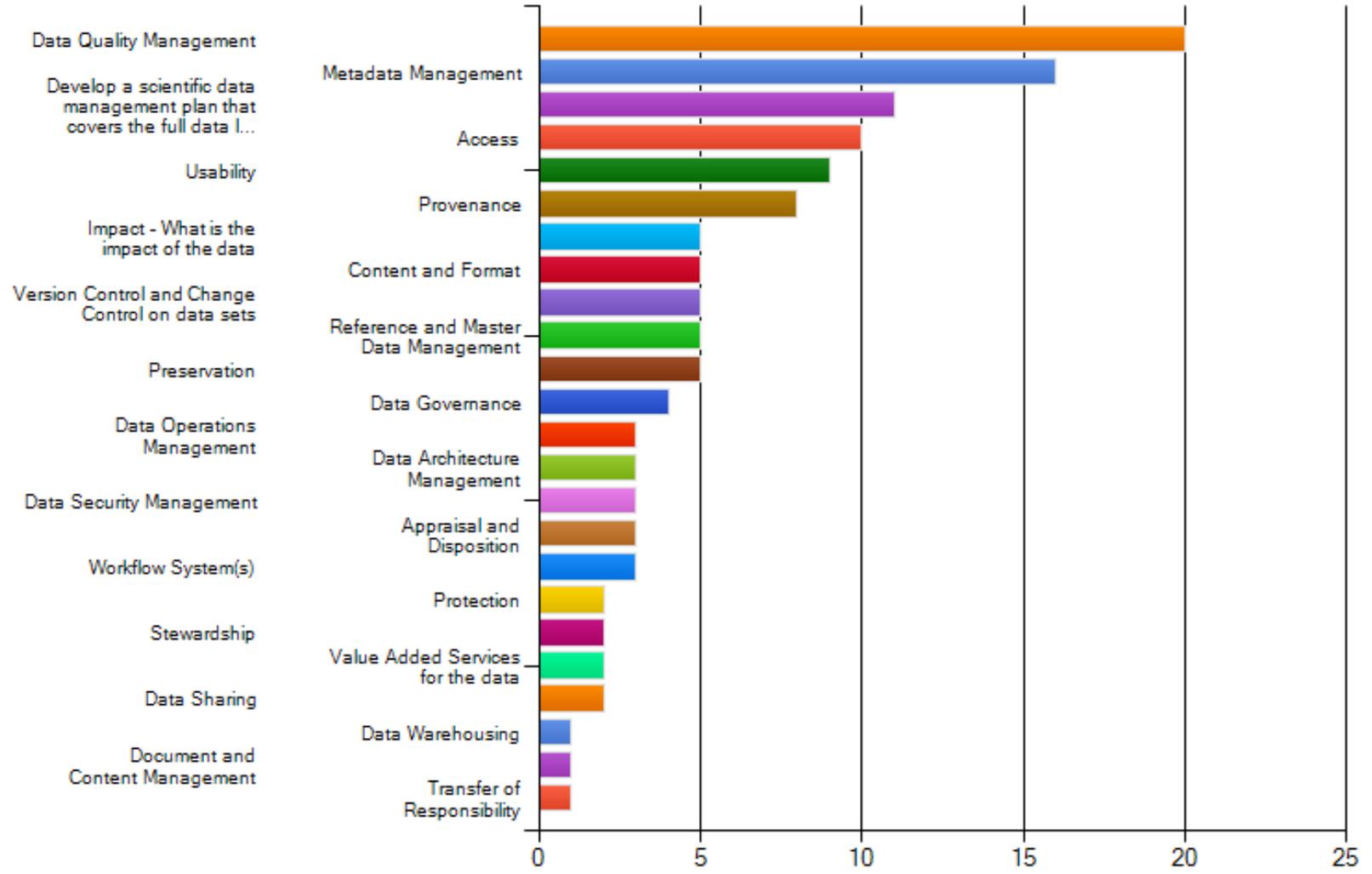
12. If you think about what elements should be in a Data Management Plan for a specific program, grant or research project, please check all that you would include as either necessary or optional. Then add any elements not on the list.

N=38



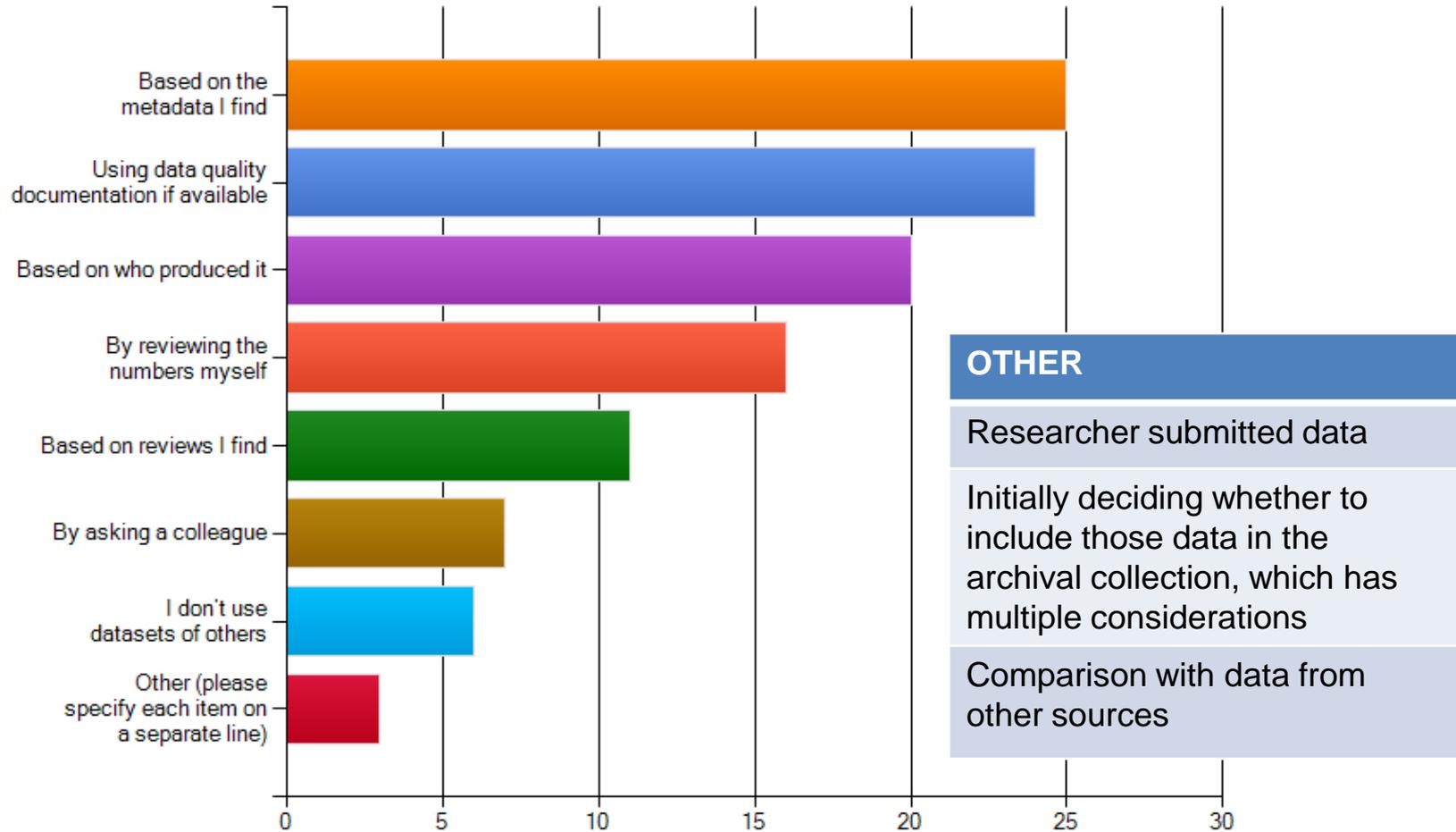
29. What three (3) elements of a data management plan would be most helpful to you in evaluating data for use in making and defending policy recommendations?

N=36



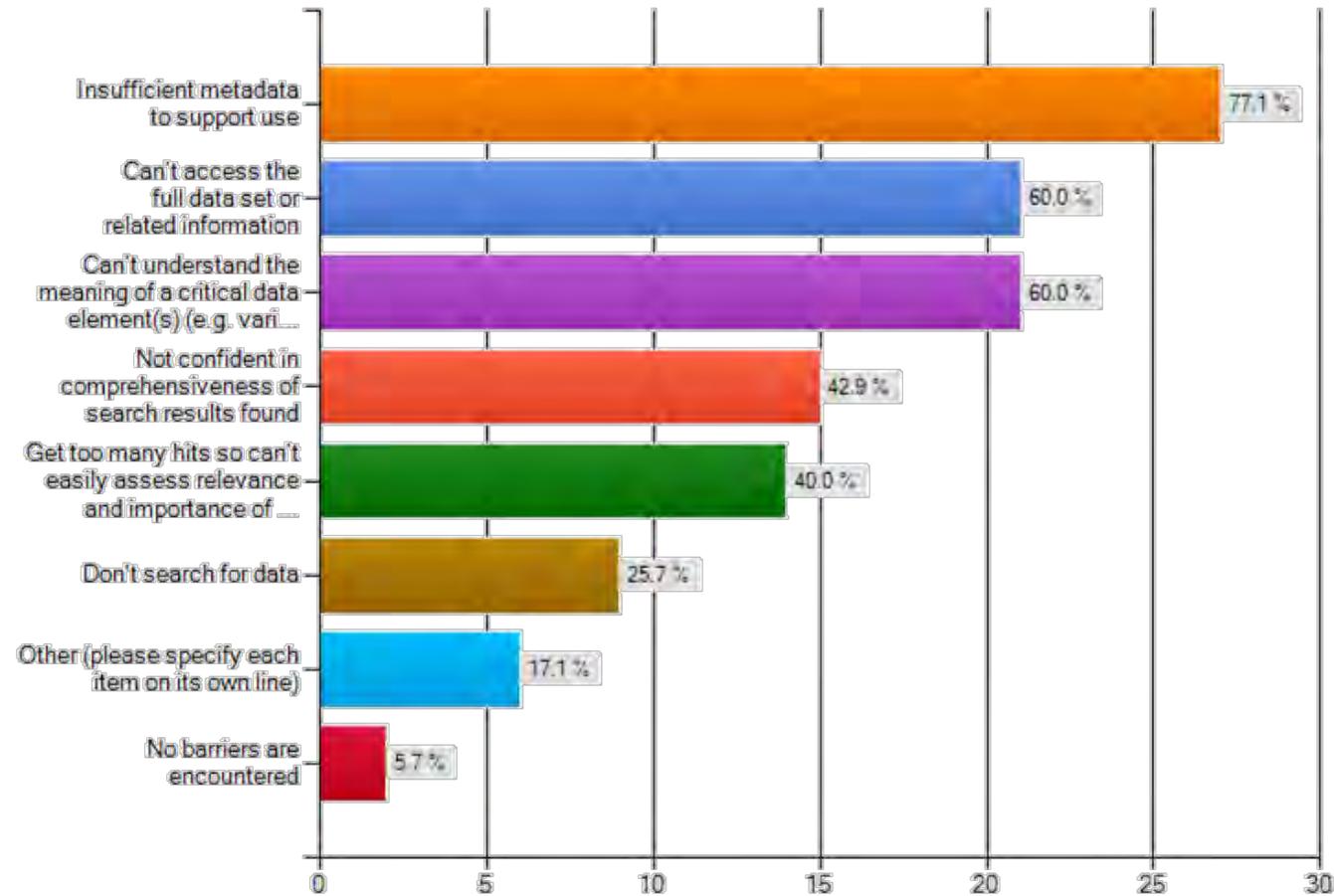
24. In using data sets of others, how do you assess whether the data are of sufficient quality for your use?

N=36



23. What barriers do you encounter when searching for or re-using “found” data generated by others? Check all that apply.

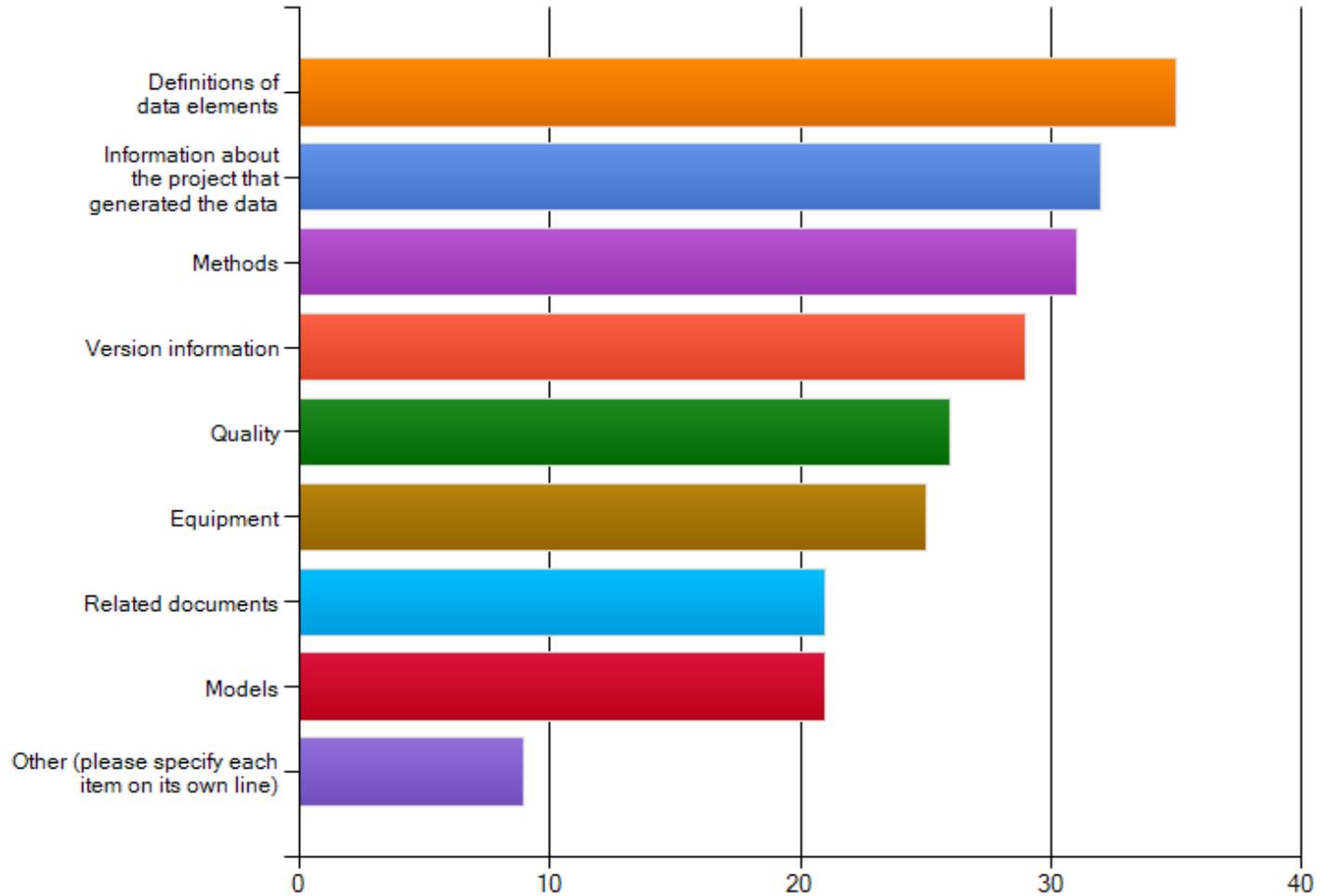
N=35



- OTHER**
- Linking of publication with data source
 - Federated search across information systems
 - Standardization of metadata across information systems
 - Getting permission to access this data
 - Need to derive metadata from existing metadata
 - All of the above are cited as reasons for not finding data in our data systems
 - Find multiple versions with uncertain lineage
 - Incorrect or misleading metadata
 - Insufficient information on QA/QC procedure details
 - Dead links in metadata

22. In your research community what level of documentation do you typically need to support use of “found” or other people’s produced data? Check all that apply.

N=36



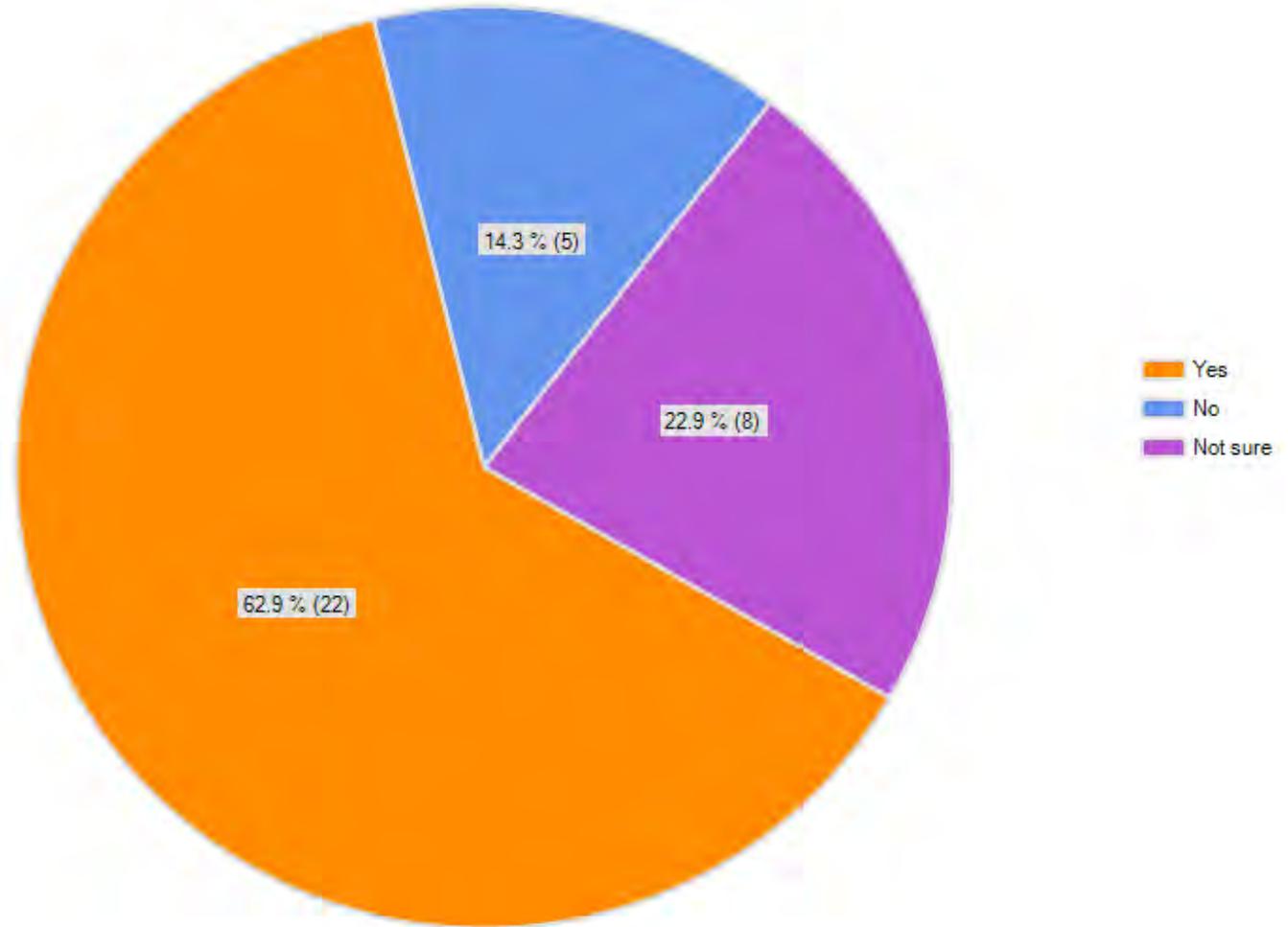
22. In your research community what level of documentation do you typically need to support use of “found” or other people’s produced data? Check all that apply.

N=36

OTHER	OTHER (cont.)
<p>Author</p> <p>Taxonomies and other metadata necessary to link the “found” data with other datasets</p>	<p>Software toolkits</p>
<p>A statement describing the original intended use of the data.</p>	<p>Sensor calibration data in easy to use (computer readable and easy to parse) format</p>
<p>"Fitness for use.</p>	<p>Geospatial information (location and datum)(2)</p>
<p>Known limitations of the data</p>	<p>Calibration references</p>
<p>Data Format definition</p>	<p>Space</p>
<p>Code libraries and Code snippets (as examples) and sometime code itself</p>	<p>Time format (m.d.yh.m.s)</p>
<p>Ancillary information including calibration, validation, etc.</p>	<p>Time as it relates to other events (e.g., two hours after a heavy rain event, two weeks after plowing, etc.)</p>

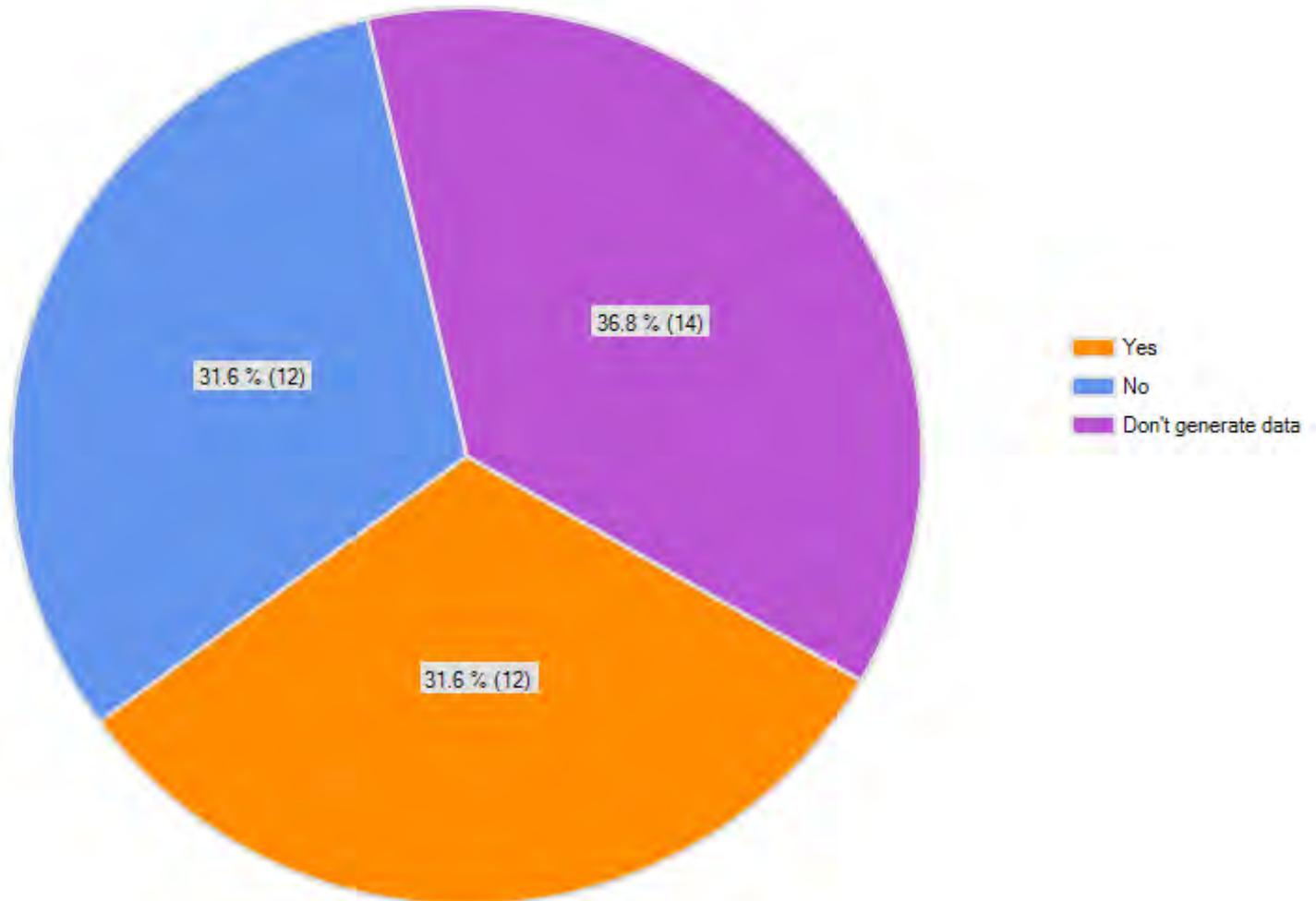
46. Do problems exist in your environment associated with accessing the appropriate version of data?

N=34



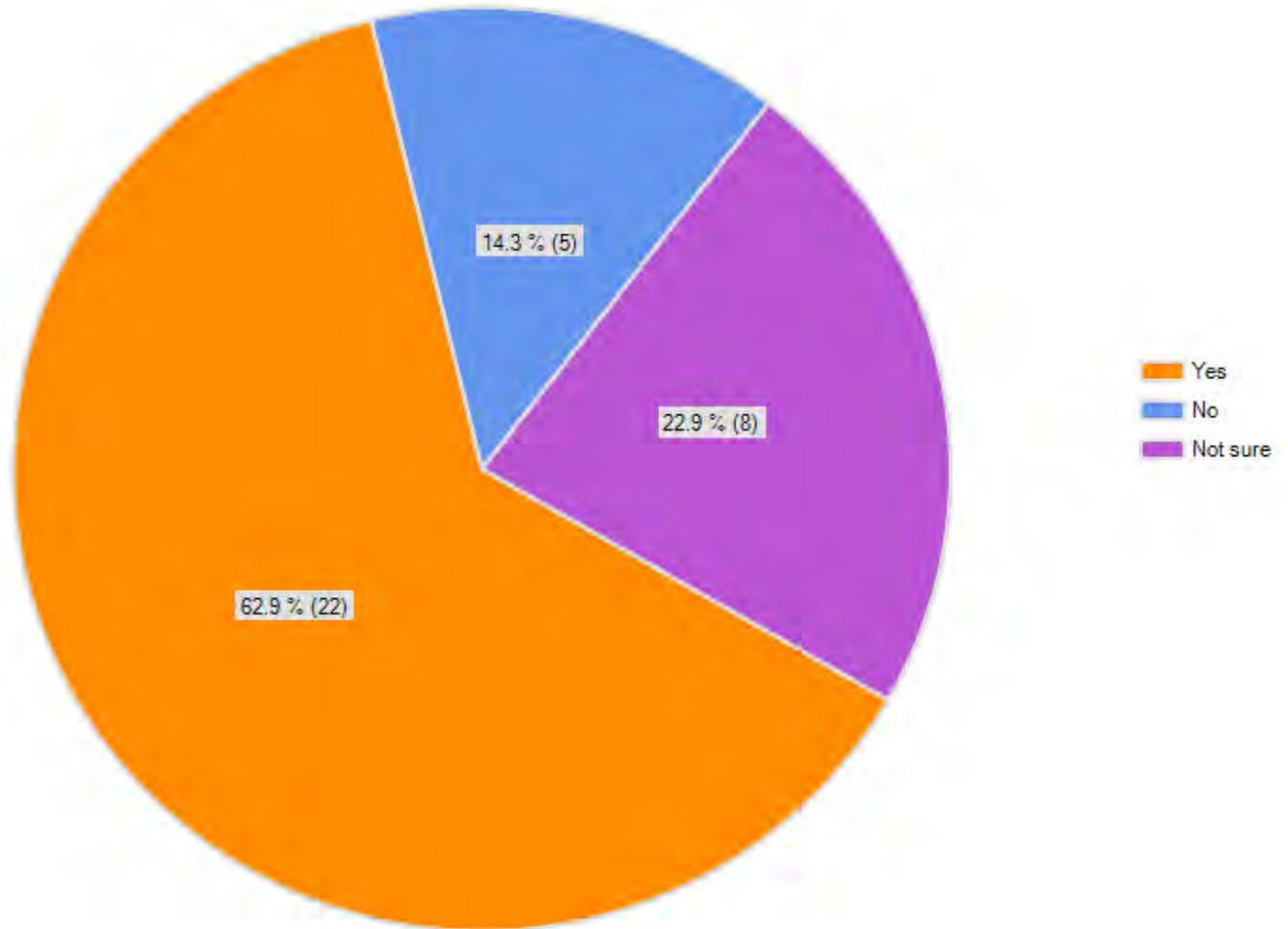
20. Generally, do you document or describe data sets that you have planned to collect or generate in a data management plan during your research planning efforts?

N=38



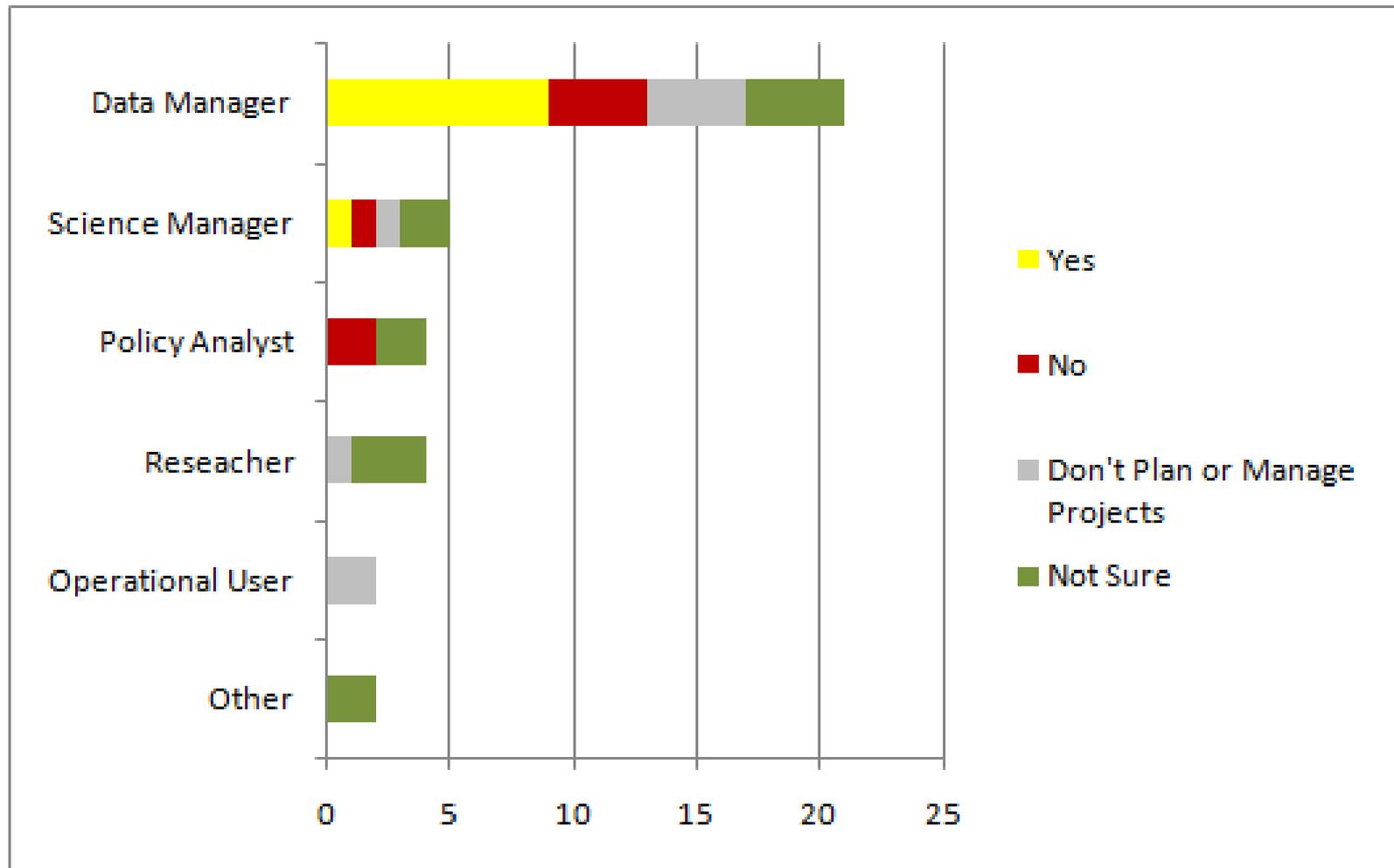
46. Do problems exist in your environment associated with accessing the appropriate version of data?

N=34



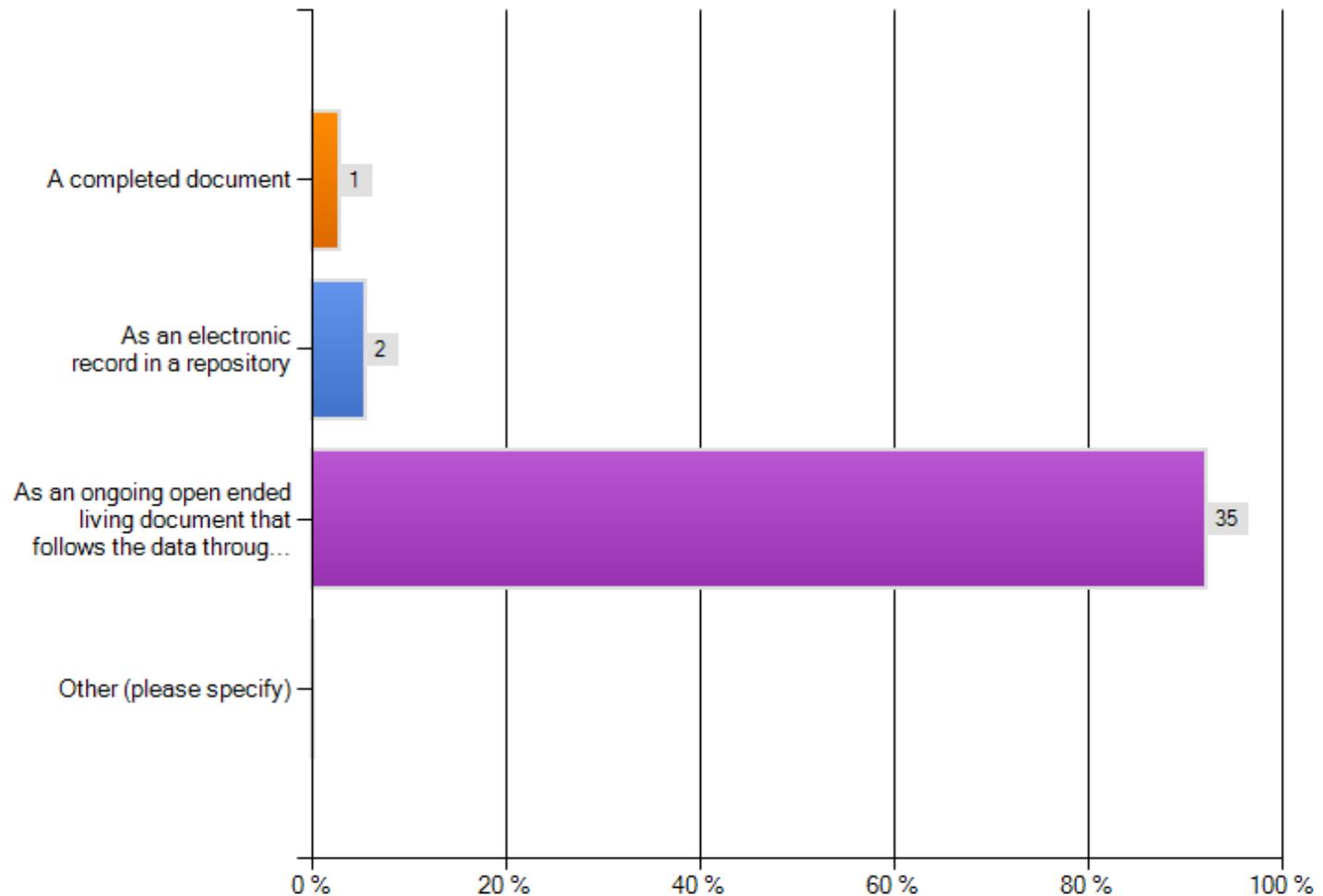
48. In planning and managing science projects do agency or program procedures exist for change control on data during the project?

N=38



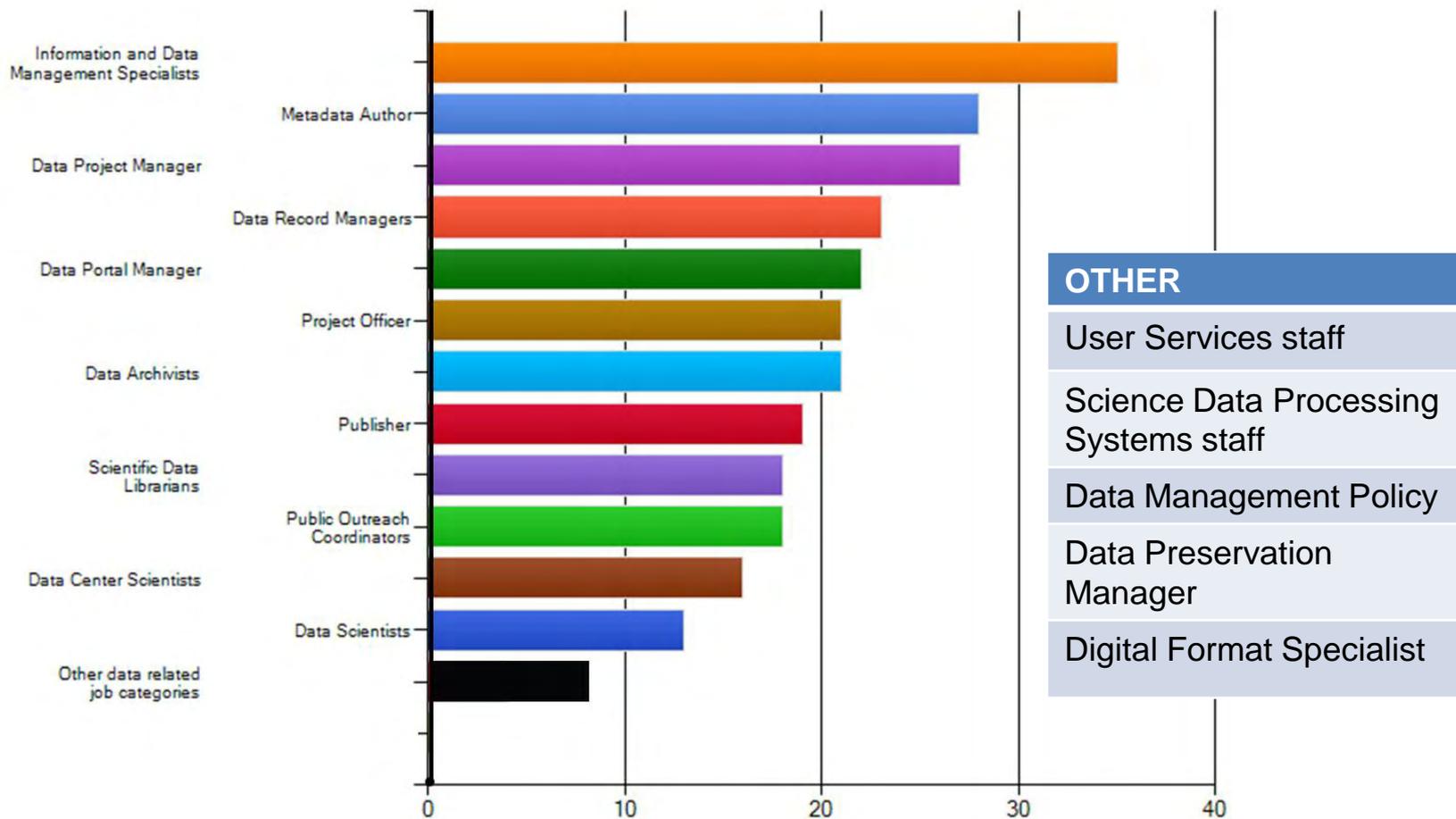
21. After a project begins, check the description of what you think a Data Management Plan should be?

N=38



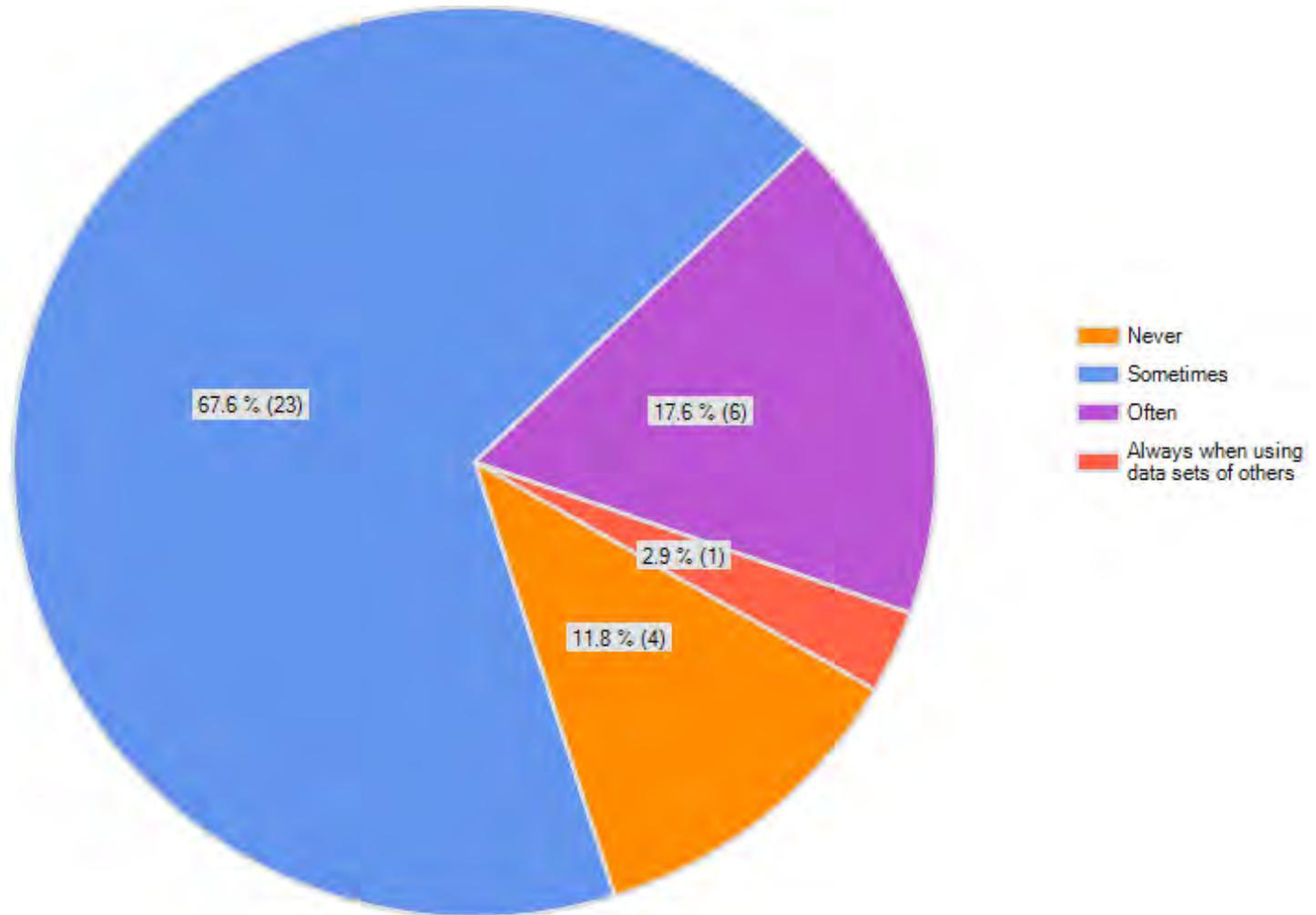
7. Please check the types of jobs listed below that your program or agency recognizes with regard to scientific data management.

N=43



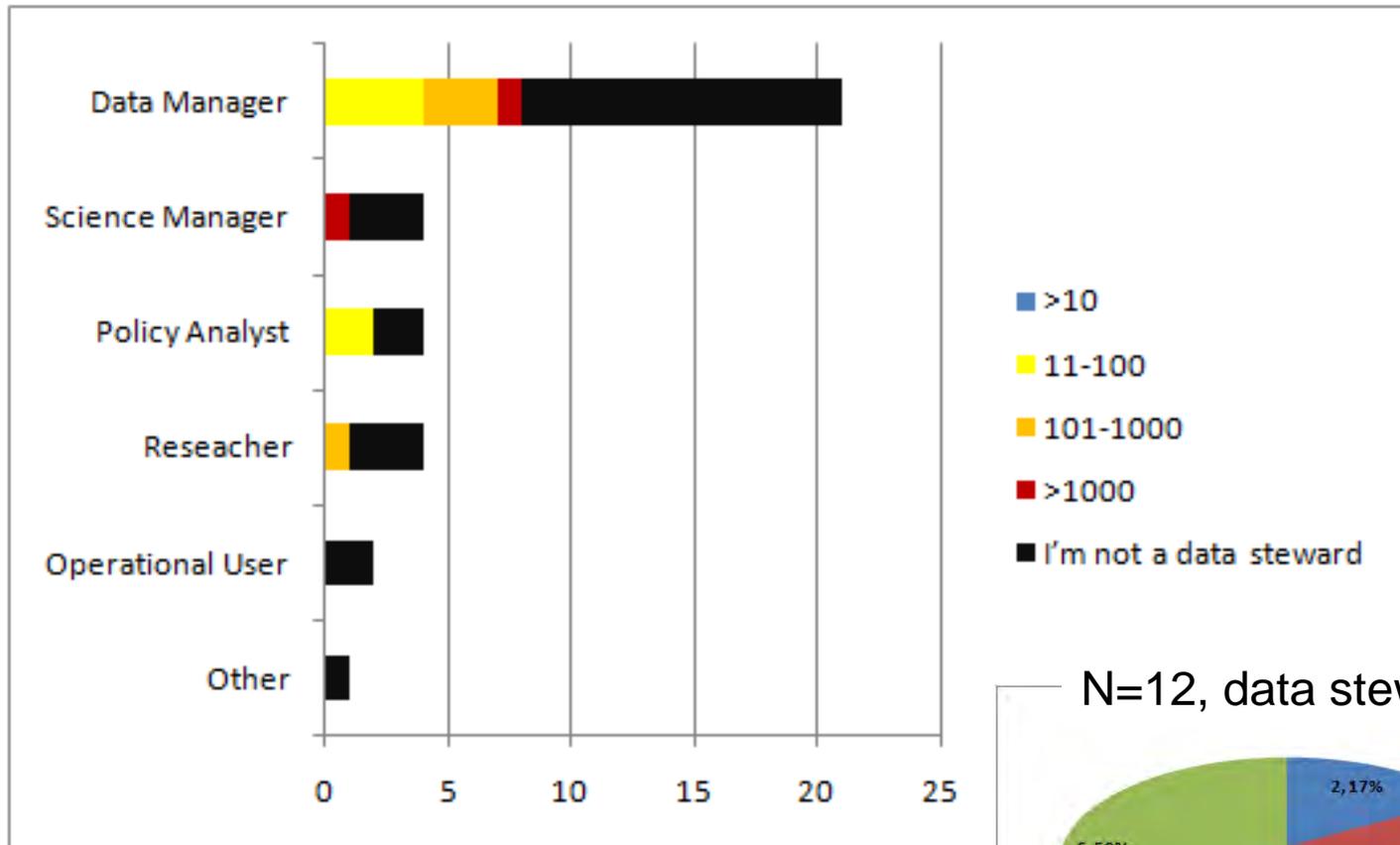
31. How often do you need a data steward who can answer your questions about an agency data set that you are using?

N=34

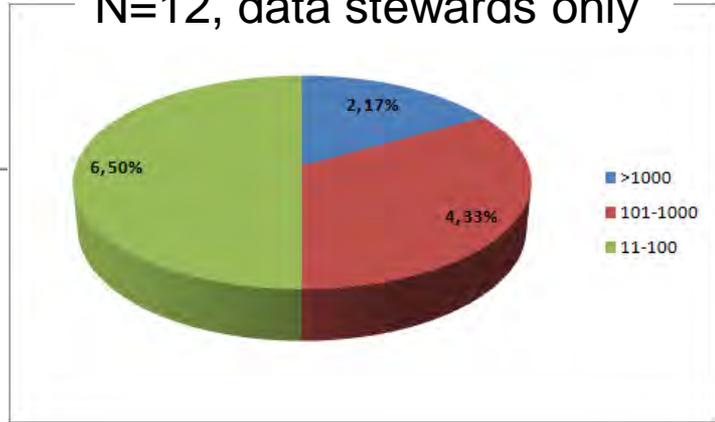


32. If you are a data steward, how many inquires do you get per year?

N=36

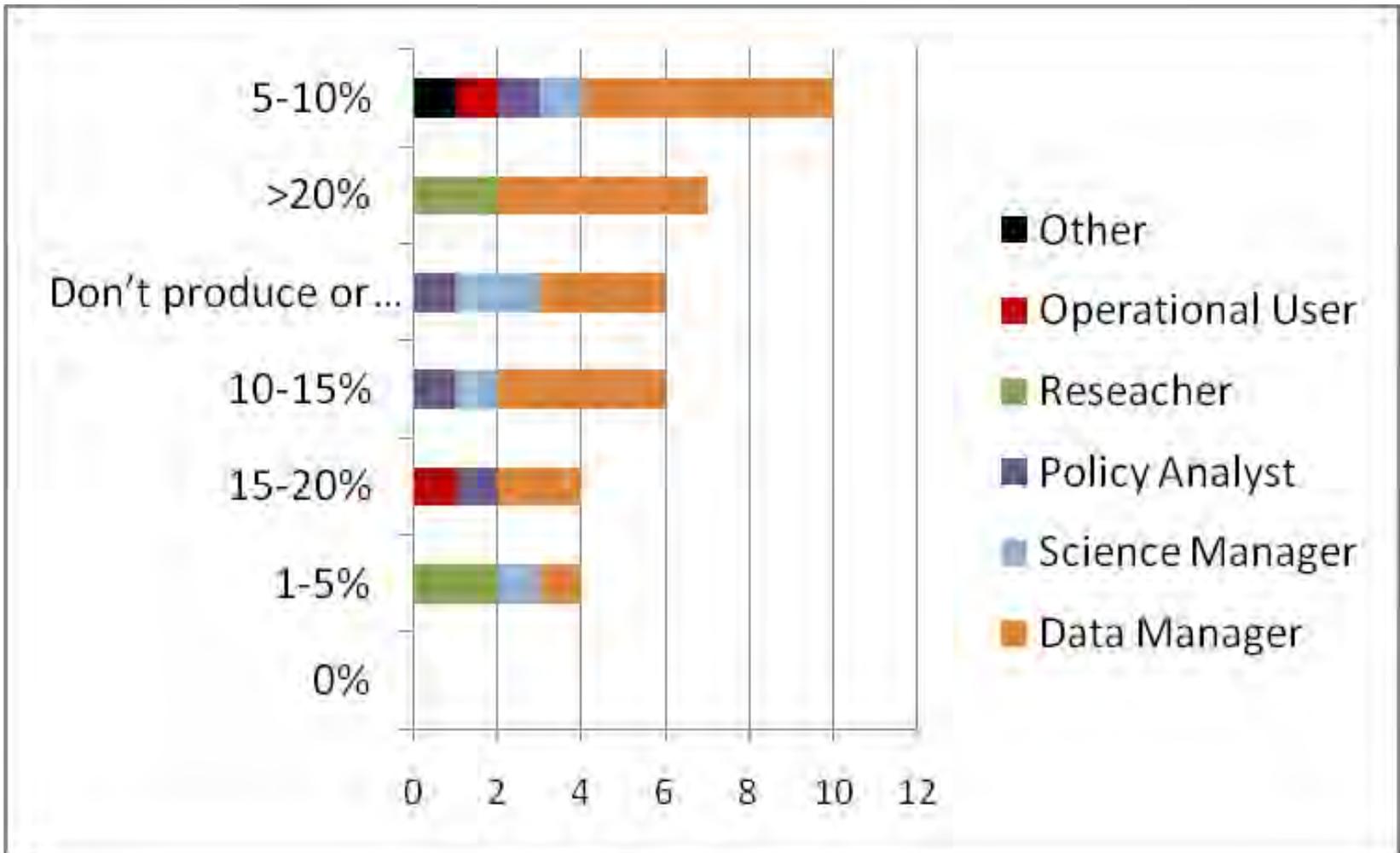


N=12, data stewards only

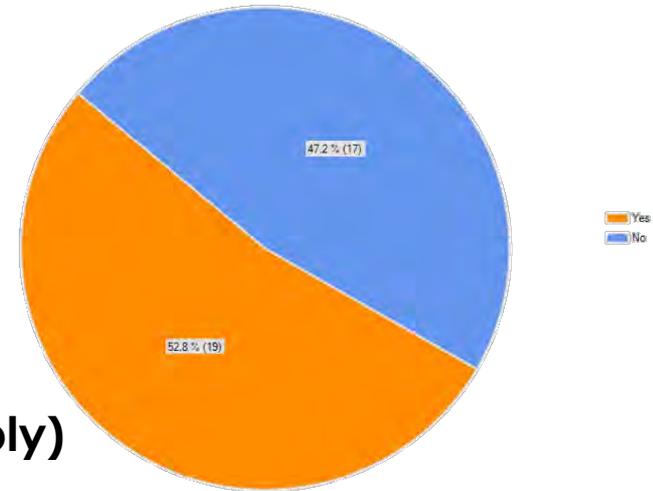


17. What percent of your efforts/resources are you willing to allocate to documenting, maintaining, and making your data available for reuse?

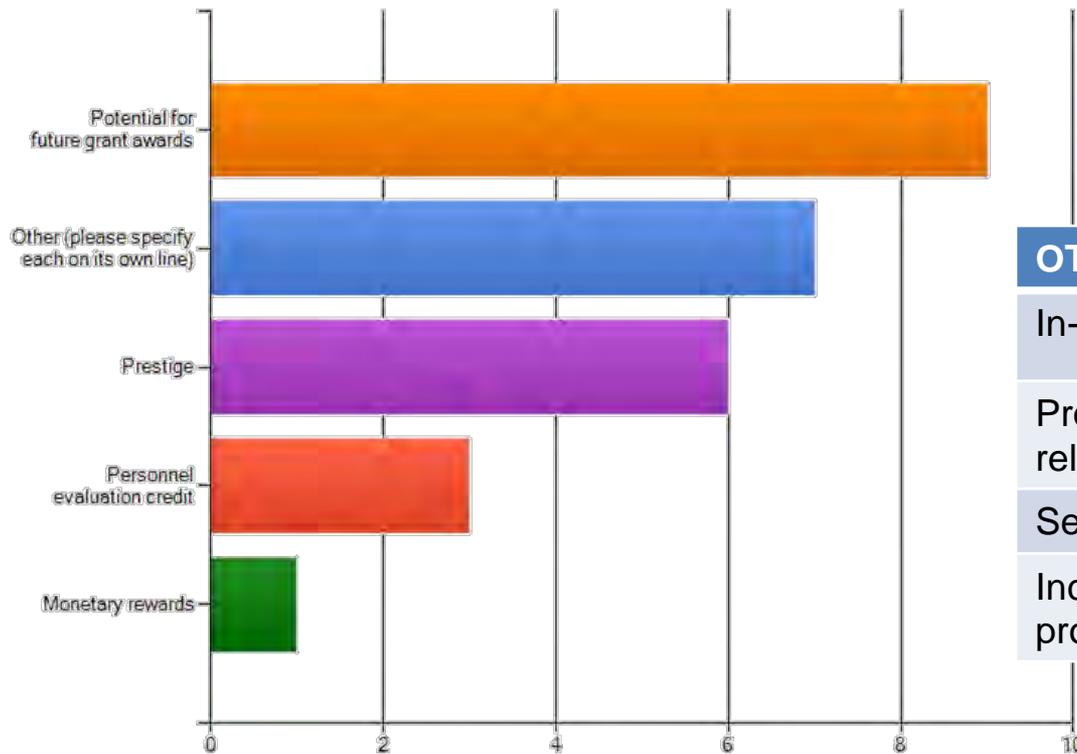
N=37



33. In your community are there incentives that promote data sharing?



34. If yes, what are they? (check all that apply)



OTHER

- In-kind sharing
- Promoting data sharing relationships (2)
- Senior management directive
- Incentive to develop a better product (2)

N=36

N=20



QUESTIONS?

Bonnie C. Carroll
CENDI Executive Director
c/o Information International Associates
1055 Commerce Park Dr. PO 4219
Oak Ridge, TN 37831
(865) 298-1220
bcarroll@iiaweb.com