

Results
of the
UCSB Faculty/Researcher Survey
on
Data Curation

Greg Janée

January 23, 2013

Outline

- *Background & motivation*
- *Survey design*
- *Questions asked*
- *Implementation*
- *Results & analysis*
- *Comments received*
- *Conclusion*

Background

- *Data Curation @ UCSB project*
 - Organized by Library, ERI, Office of Research
 - Funded by EVC's office
- *Goals*
 - Characterize state of data production and data curation on campus
 - Identify needs, solutions, gaps
 - Give Library guidance on future staffing, services
- *Strategy*
 - This survey
 - Researcher interviews; data, workflow profiles
 - In-depth case studies

Similar surveys

- *Surveys on curation & libraries*
 - CalPoly
 - <http://crl.acrl.org/content/73/4/349.abstract>
 - University of Colorado, Boulder
 - doi:10.1353/pla.2011.0049
 - Georgia Tech
 - <http://crln.acrl.org/content/73/1/10.full>
 - Oxford
 - <http://tinyurl.com/agdurc7>
- *Why another?*
 - Look at the UCSB situation
 - Ask previously-unasked questions

Survey design

- *Short and sweet*
 - Maximize response rate at expense of depth
 - 5 questions, 5 minutes
 - All multiple choice with “other” values allowed
- *Blanket (non-sampled)*
- *Anonymous*
- *Questions asked*
 - Each answer leads to a plan of action
- *Questions not asked*
 - No clear implications
 - Answered by other means
 - Factoid

Question 1

- *In the course of your research or teaching, do you produce digital data that merits curation?*
 - Yes
 - No

Question 2

- *Which parties do you believe have primary responsibility for the curation of your data, if any? (check all that apply)*
 - Myself
 - Campus IT
 - Lab manager
 - Lab research staff
 - Department
 - External research partner
 - External data repository
 - Funding agency
 - Campus library
 - UC Curation Center (UC3)
 - Other (please specify)

Question 3

- *Are you mandated to provide for (or otherwise participate in) the curation of your data, and if so, by which agencies? (check all that apply)*
 - NSF
 - NIH
 - NASA
 - DOE
 - Other (please specify)

Question 4

- *What data management activities could you use help with, if any? (check all that apply)*
 - Creating data management plans at proposal time
 - Identifying appropriate data repositories
 - Interacting with data repositories
 - Assigning permanent identifiers to data (i.e., just as DOIs are assigned to journal articles)
 - Identifying appropriate data registries and search portals
 - Creating and publishing descriptions of data (or “metadata”)
 - Day-to-day data storage
 - Backup or archival data storage
 - Capturing computational provenance (the steps used to create the data)
 - Obtaining more information about curation best practices
 - Identifying funding sources for curation support
 - Other (please specify)

Question 5

- *With which departments, programs, and ORUs are you affiliated? (check all that apply)*
 - Long list of campus entities
 - Other (please specify)

Other survey parts

- *Additional comments*
- *Signup to serve as case study*
 - Decorrelated from survey proper
- *Mailing list signup*
 - Decorrelated from survey proper

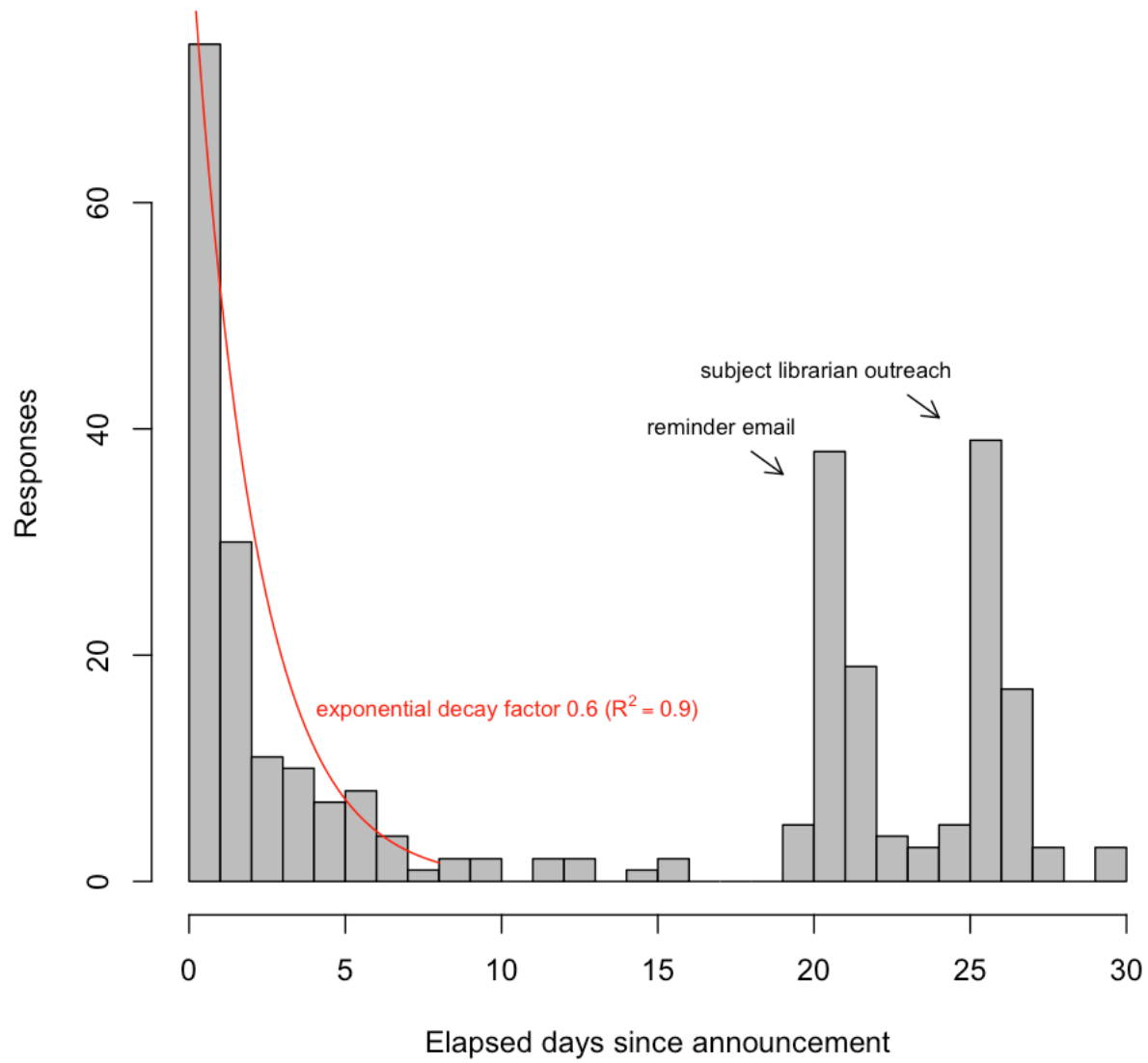
Implementation

- *Online (SurveyMonkey)*
- *Target audience contacted via email*
 - Faculty: via Academic Senate
 - Faculty and other researchers: through campus ORUs via Office of Research
 - Two follow-ups

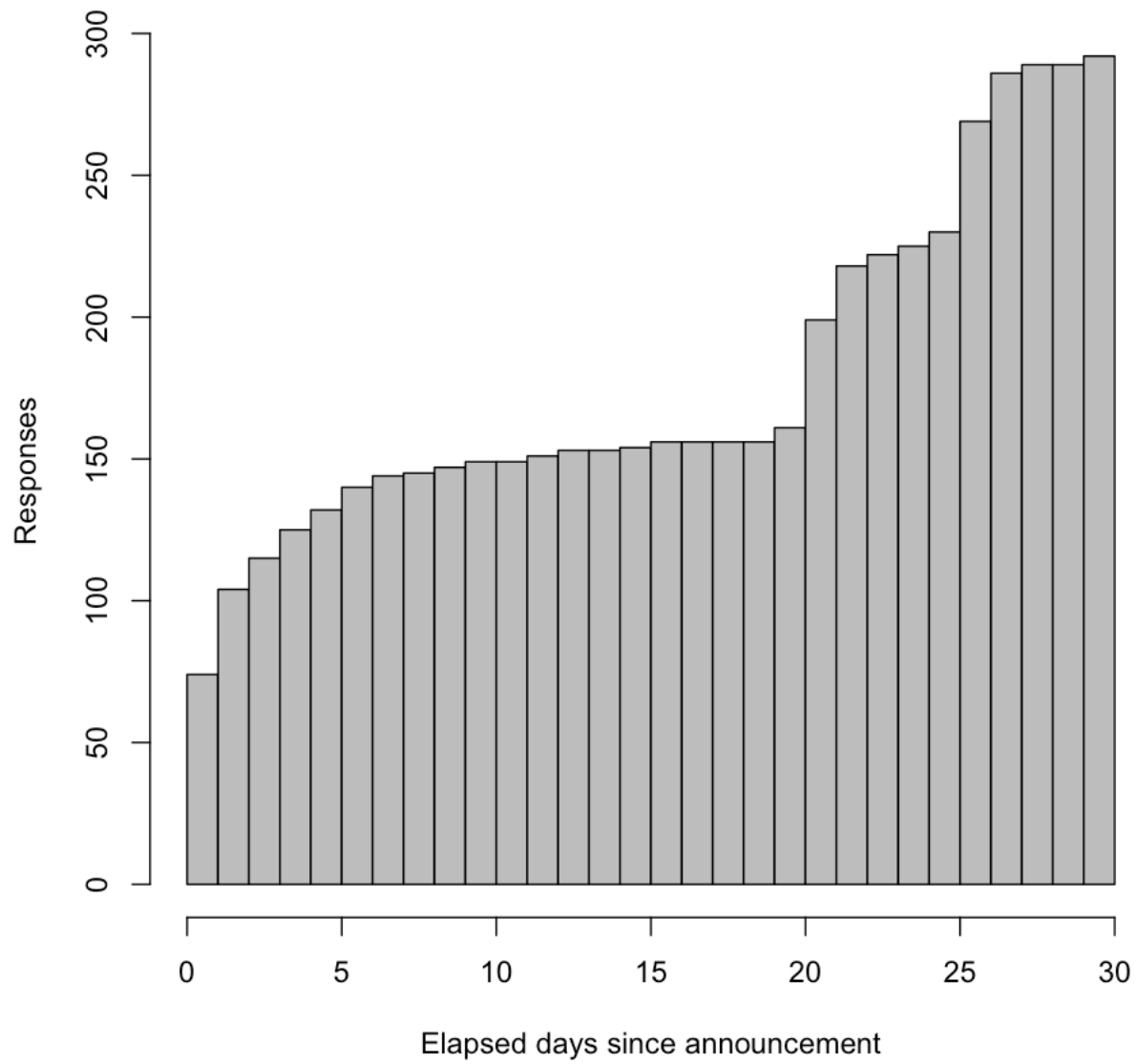
Result

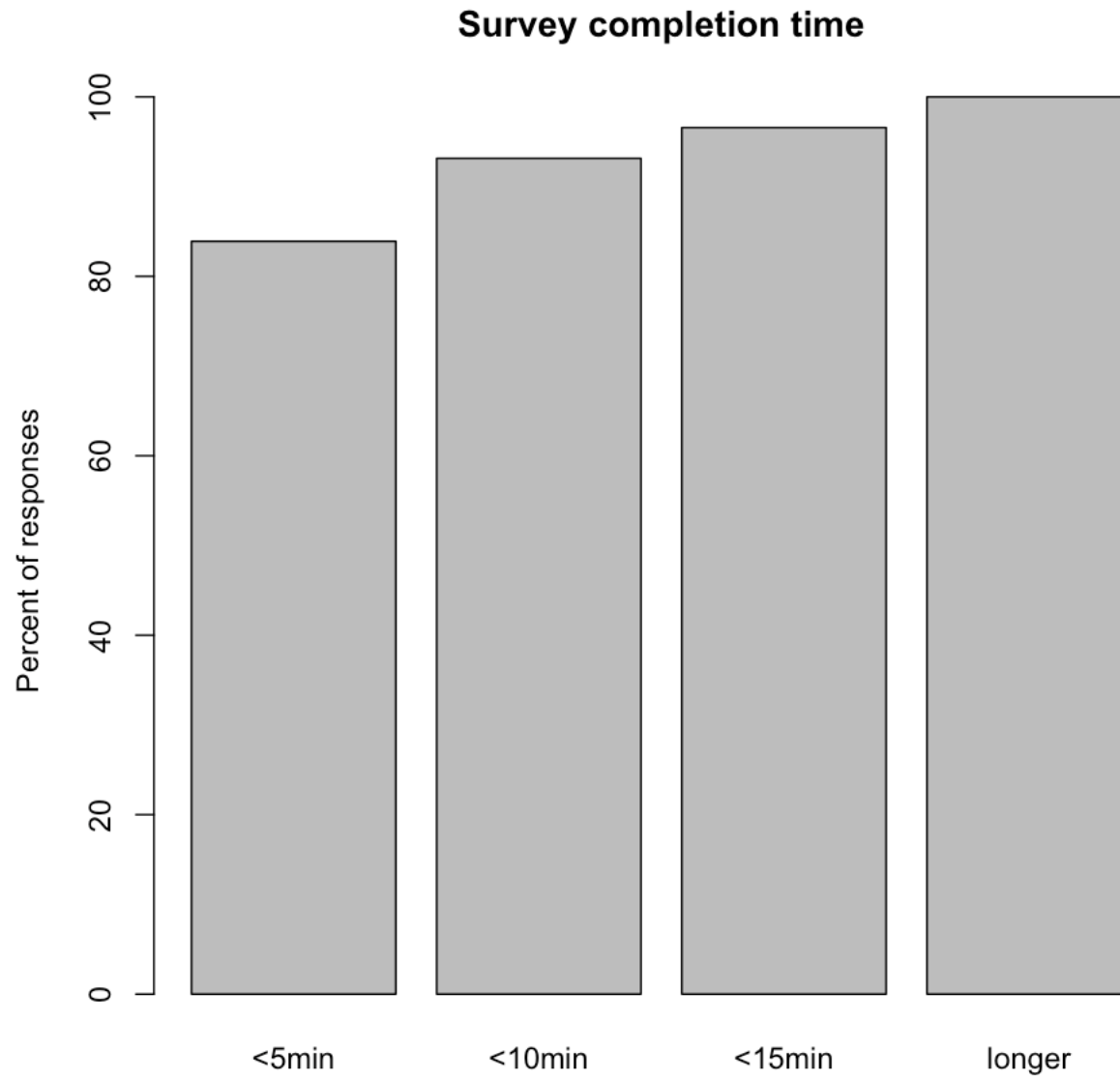
- *294 responses*
 - Represents 1/3 of estimated 900 faculty, researchers

Survey responses over time

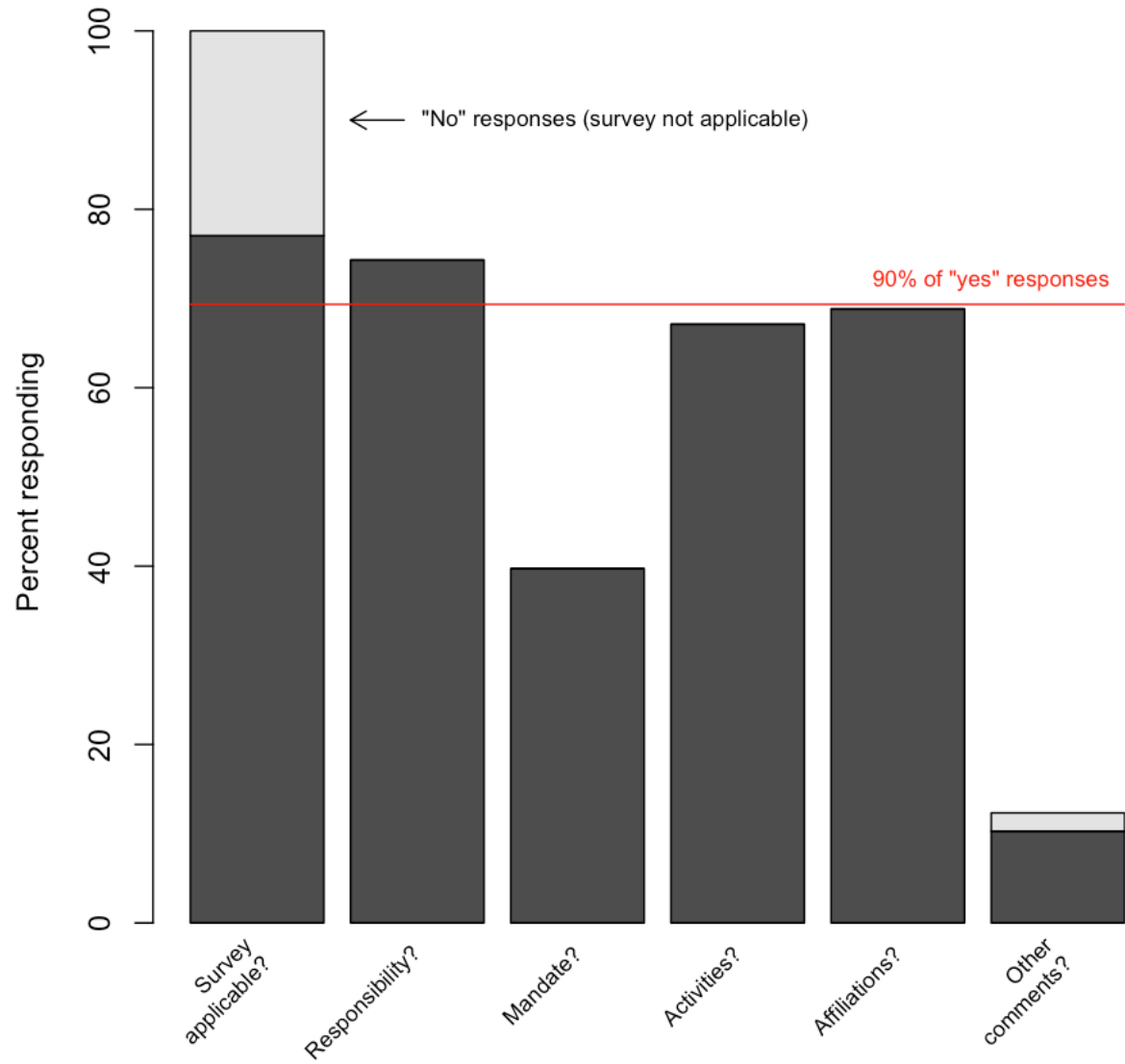


Cumulative survey responses over time





Survey completion percentage

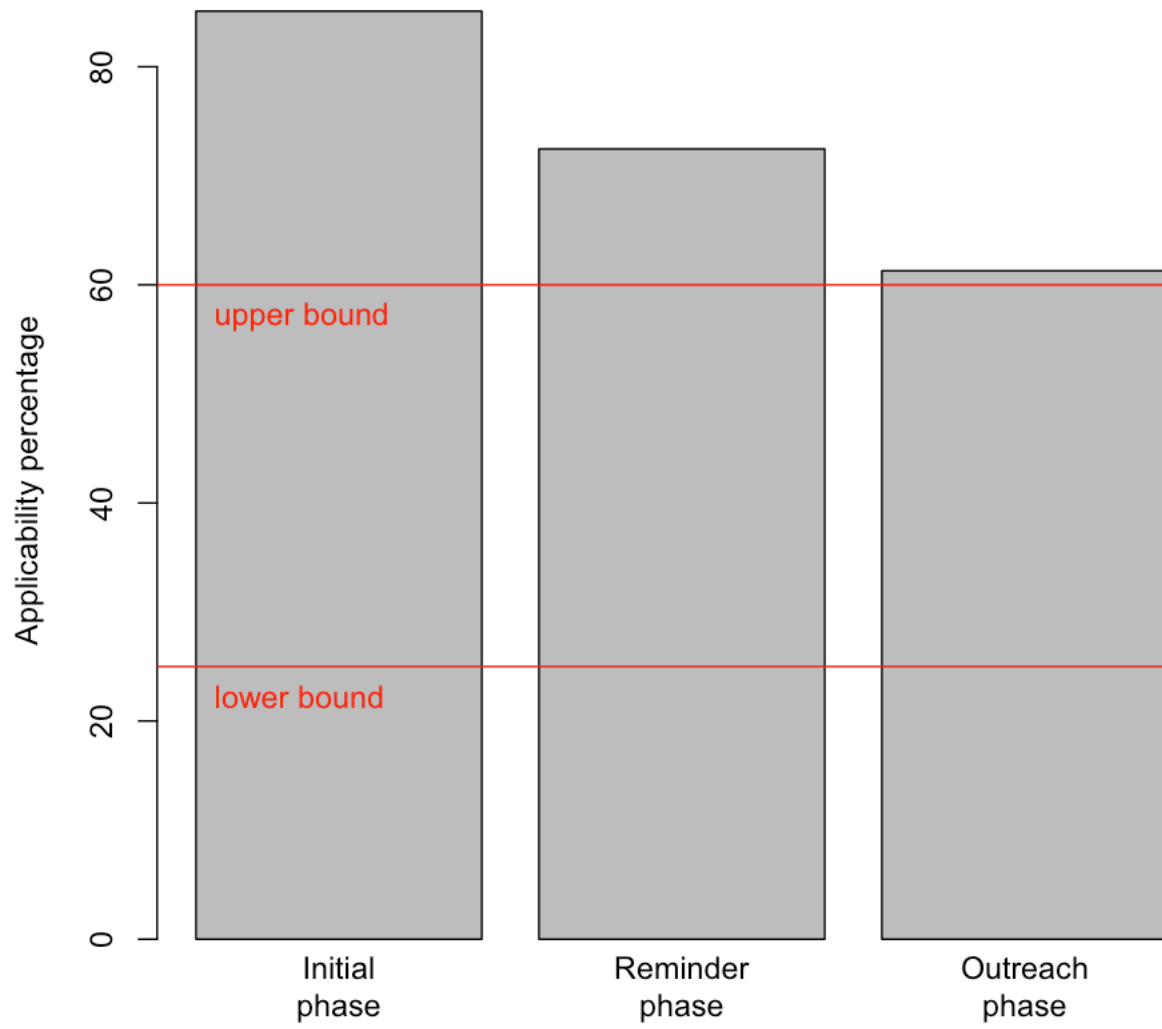


In the course of your research or teaching, do you produce digital data that merits curation?

Applicability question

- *Yes: 225 (77%)*
 - Or, 25% of target population of 900
- *Non-response bias*
- *Improve bounding of true value using late responses*
 - Hypothesis: late responders more likely to respond, even negatively, due to pestering/guilt
 - Ergo, applicability percentage closer to true value

Applicability percentage by survey phase

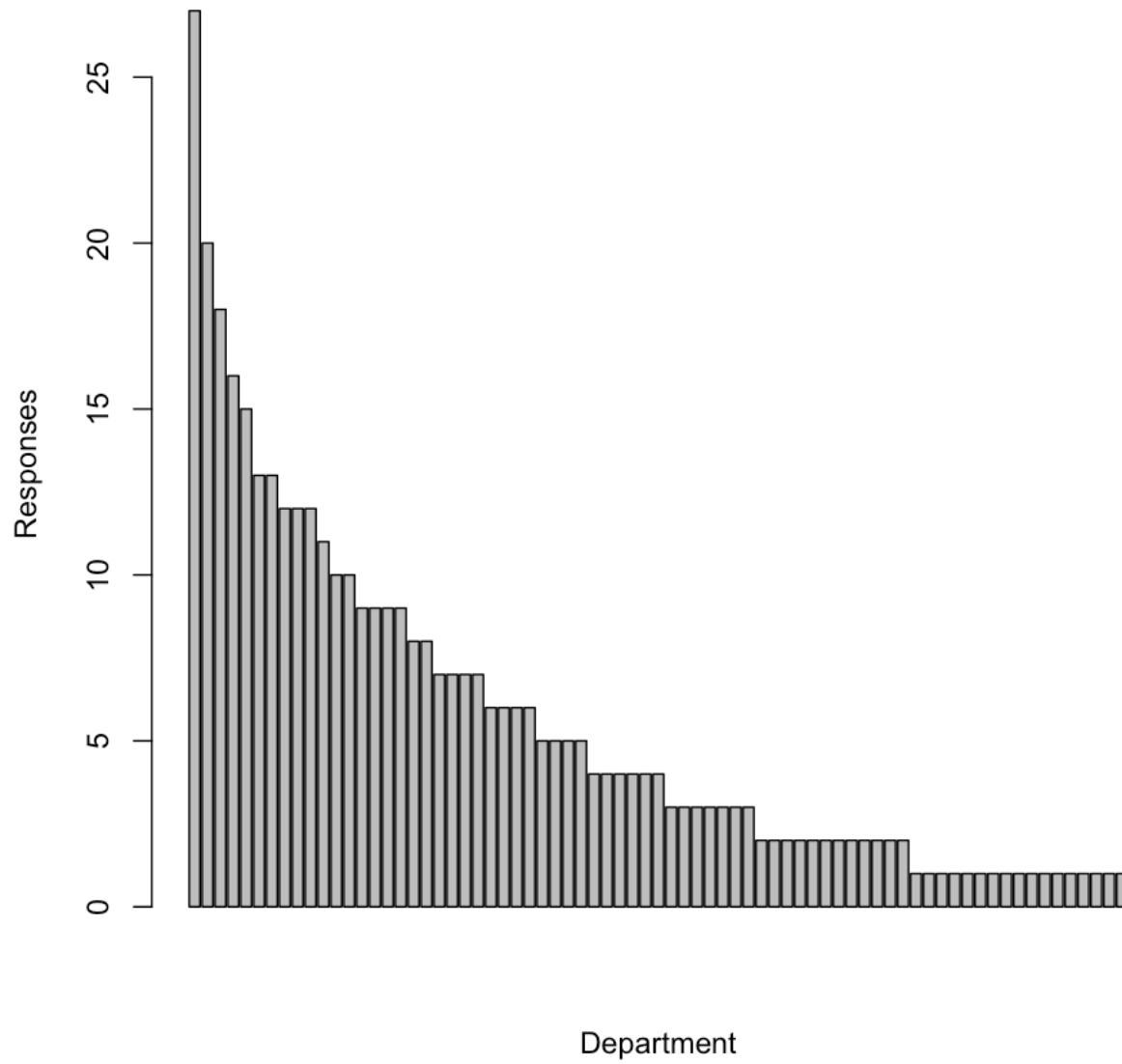


Applicability question

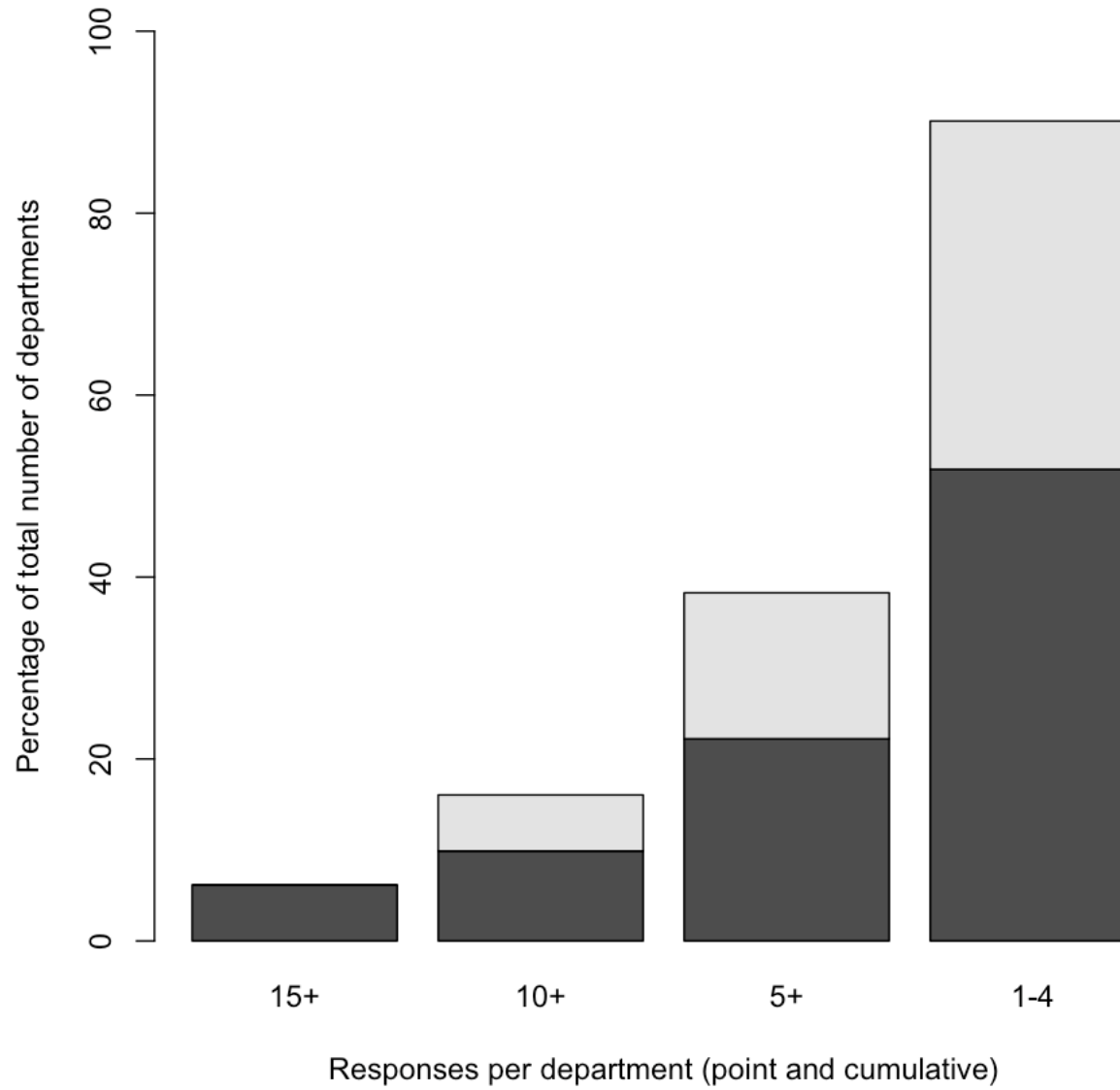
- *Observations*
 - Great interest in survey
 - Good participation rate
 - Curation applicable to significant proportion of campus researchers
- *Implications for Library*
 - Timely topic

With which departments, programs, and ORUs are you affiliated?

Yes responses by department



Yes responses per department



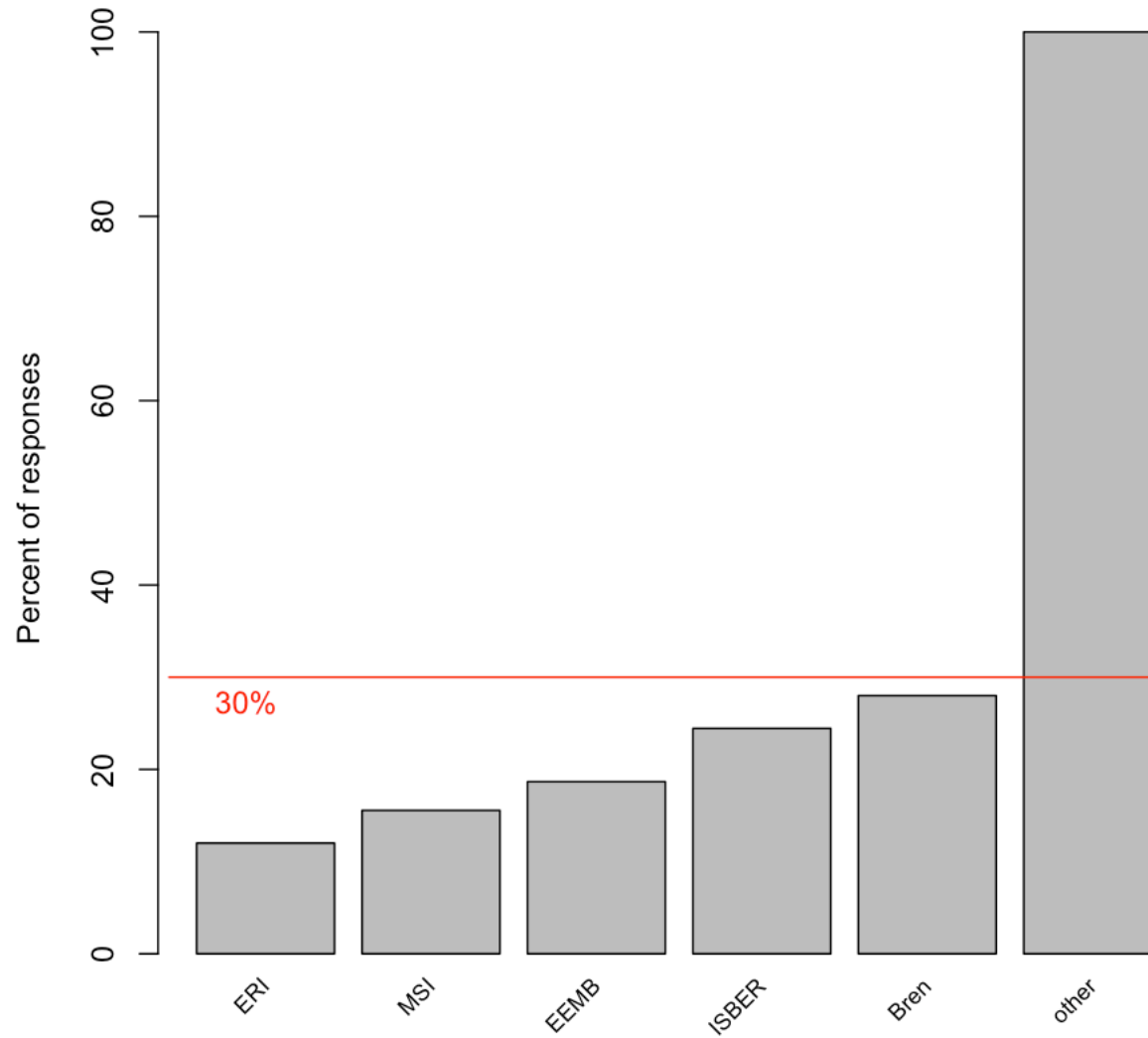
Missing departments

- *Center for Black Studies Research*
- *Counseling, Clinical, and School Psychology*
- *English as a Second Language*
- *Exercise & Sport Studies*
- *Global Peace and Security*
- *Jewish Studies*
- *Military Science*
- *Women, Culture, and Development Studies*

Leading departments

27	<i>ERI</i>	11	<i>MCDB</i>	7	<i>Materials</i>
20	<i>MSI</i>	10	<i>CompSci</i>	7	<i>MechEng</i>
18	<i>EEMB</i>	10	<i>EarthSci</i>	7	<i>MediaArts</i>
16	<i>ISBER</i>	9	<i>ChemBiochem</i>	6	<i>ECE</i>
15	<i>Bren</i>	9	<i>CompLit</i>	6	<i>LatinAmerStds</i>
13	<i>Geography</i>	9	<i>FeministStds</i>	6	<i>Linguistics</i>
13	<i>Physics</i>	9	<i>PoliSci</i>	6	<i>Psychology</i>
12	<i>Anthropology</i>	8	<i>Sociology</i>	5	<i>BMSE</i>
12	<i>History</i>	8	<i>NRI</i>	5	<i>ChemEng</i>
12	<i>MarineSci</i>	7	<i>Education</i>	5	<i>FrenchItal</i>

Cumulative responses by department

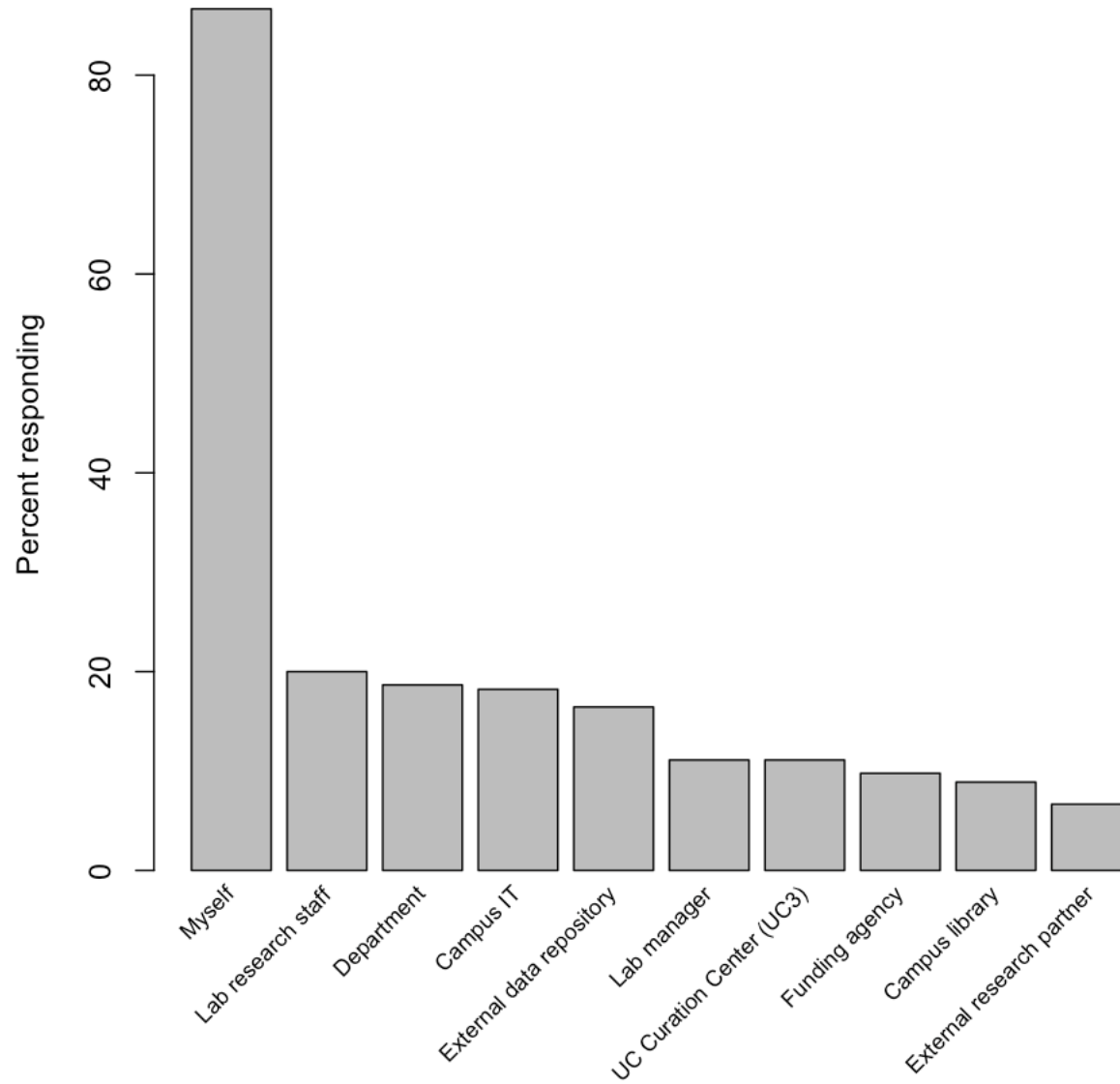


Department question

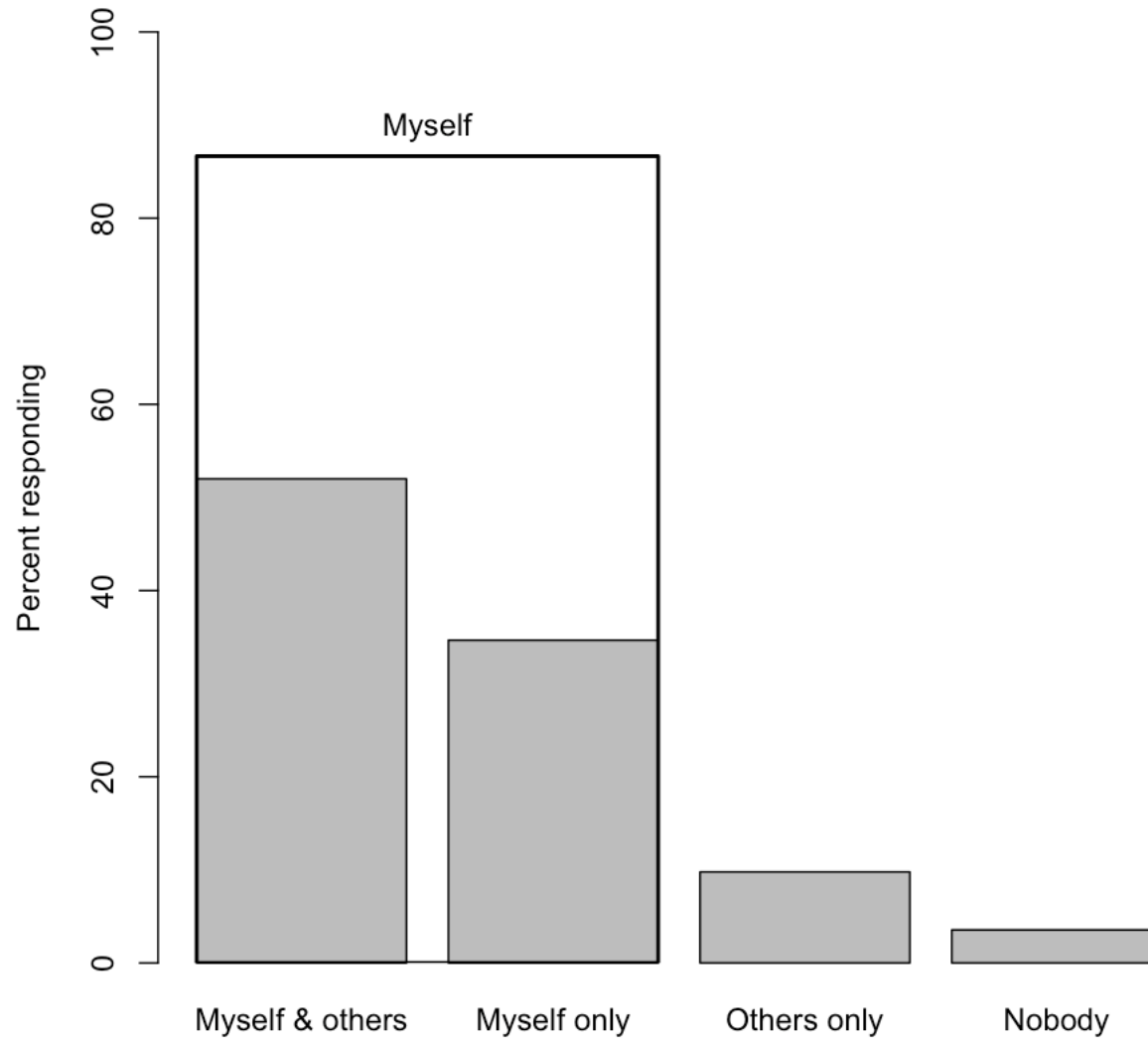
- *Observations*
 - Virtually every department responded affirmatively
 - Responses not uniform, but nevertheless broadly distributed across departments
- *Implications for Library*
 - Data curation is a campus-wide concern
 - Many researchers
 - Every department
 - Argues for campus agency to address curation

Which parties do you believe have primary responsibility for the curation of your data, if any?

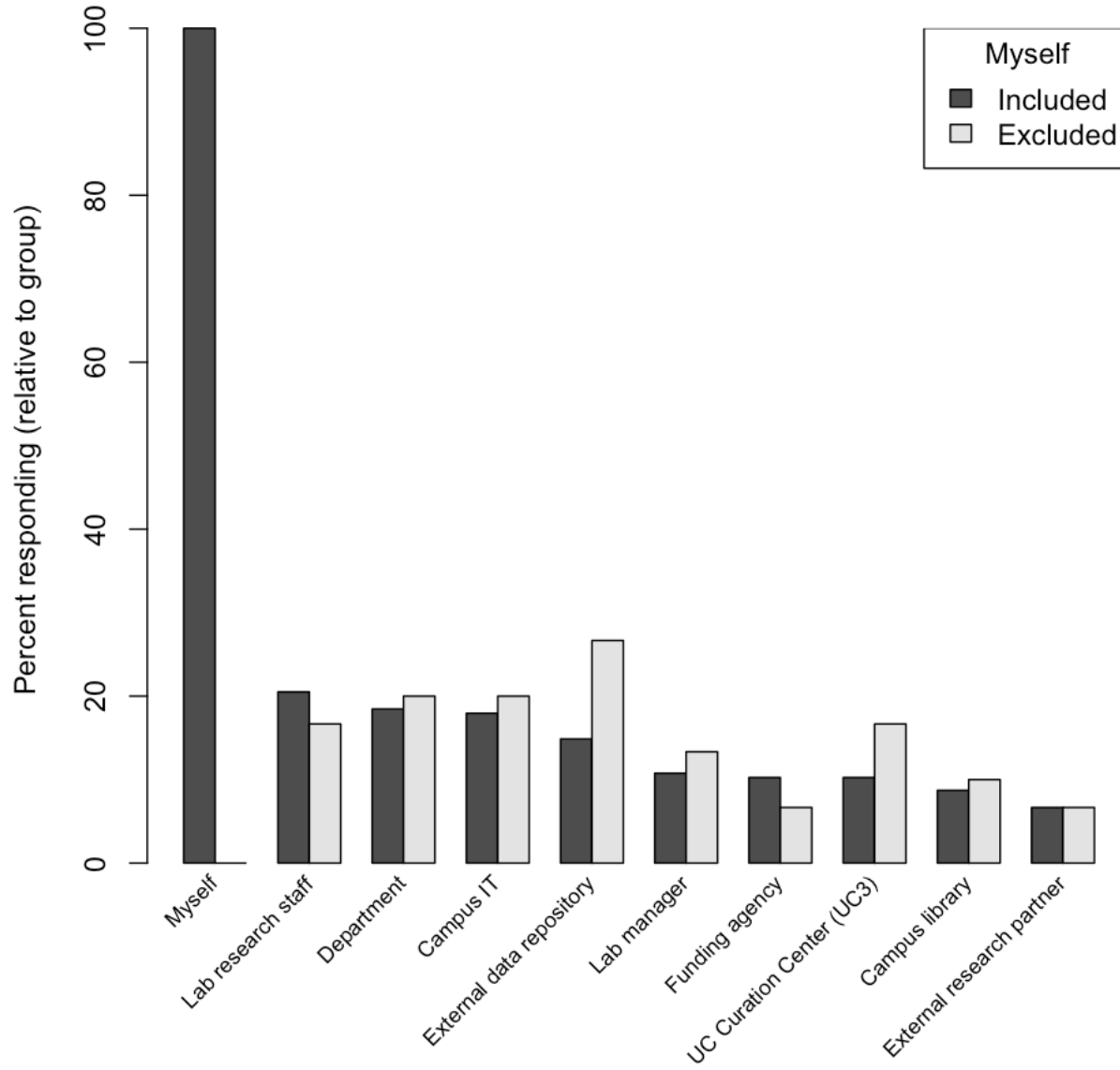
Which parties do you believe have primary responsibility for the curation of your data?



Responsibility, myself versus others



Responsibility, myself versus others



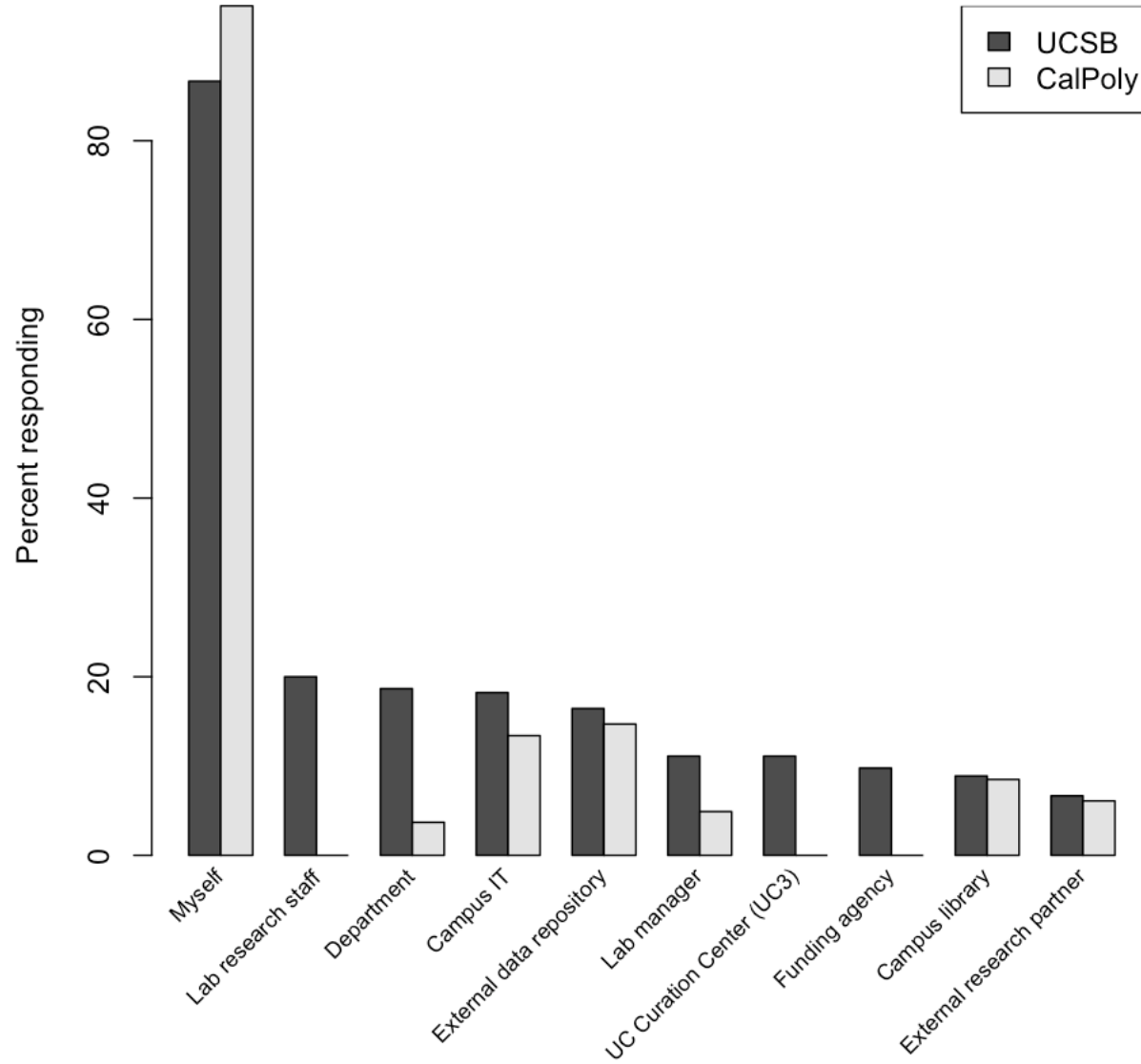
Responsibility question

- *Observations*
 - Universal personal commitment
 - Combined with reliance on/collaboration with partners
- *Implications for Library*
 - There is (some) hope for curation!
 - Individual researcher is point of engagement
 - And service consumer
 - Researchers may need help interfacing with partners

Other responses

- *Established solutions*
 - “Journals where we publish our data”
 - “Professional societies”
 - “Collections repository within my department”
 - “my e-mail group [...] has archives saved at UCSB”
 - “The research program with which the project is affiliated”
 - “publicity office”
- *Somebody else*
 - “PI / adviser”
 - “Co_PIs in my research group”

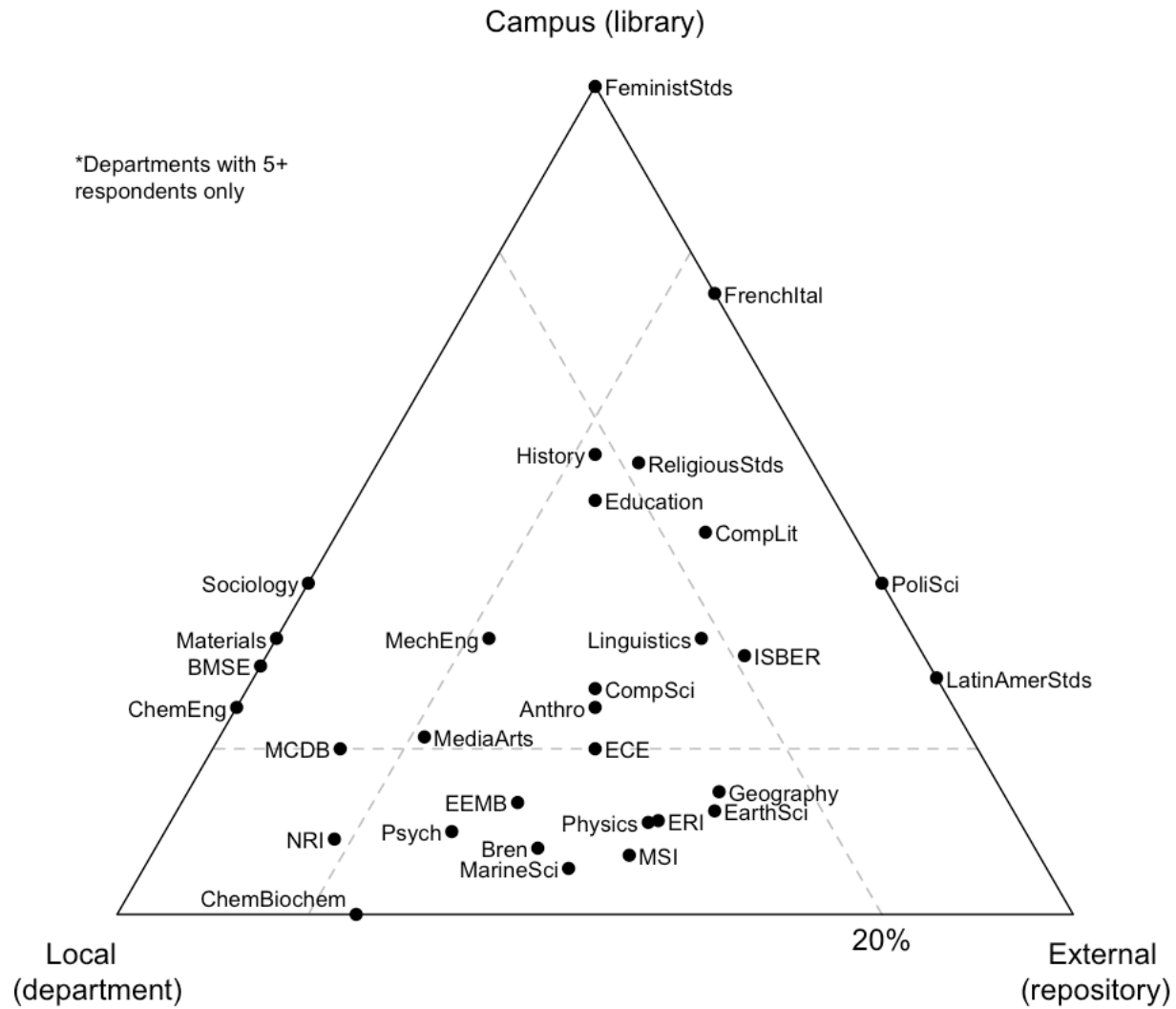
UCSB/CalPoly reponsibility comparison



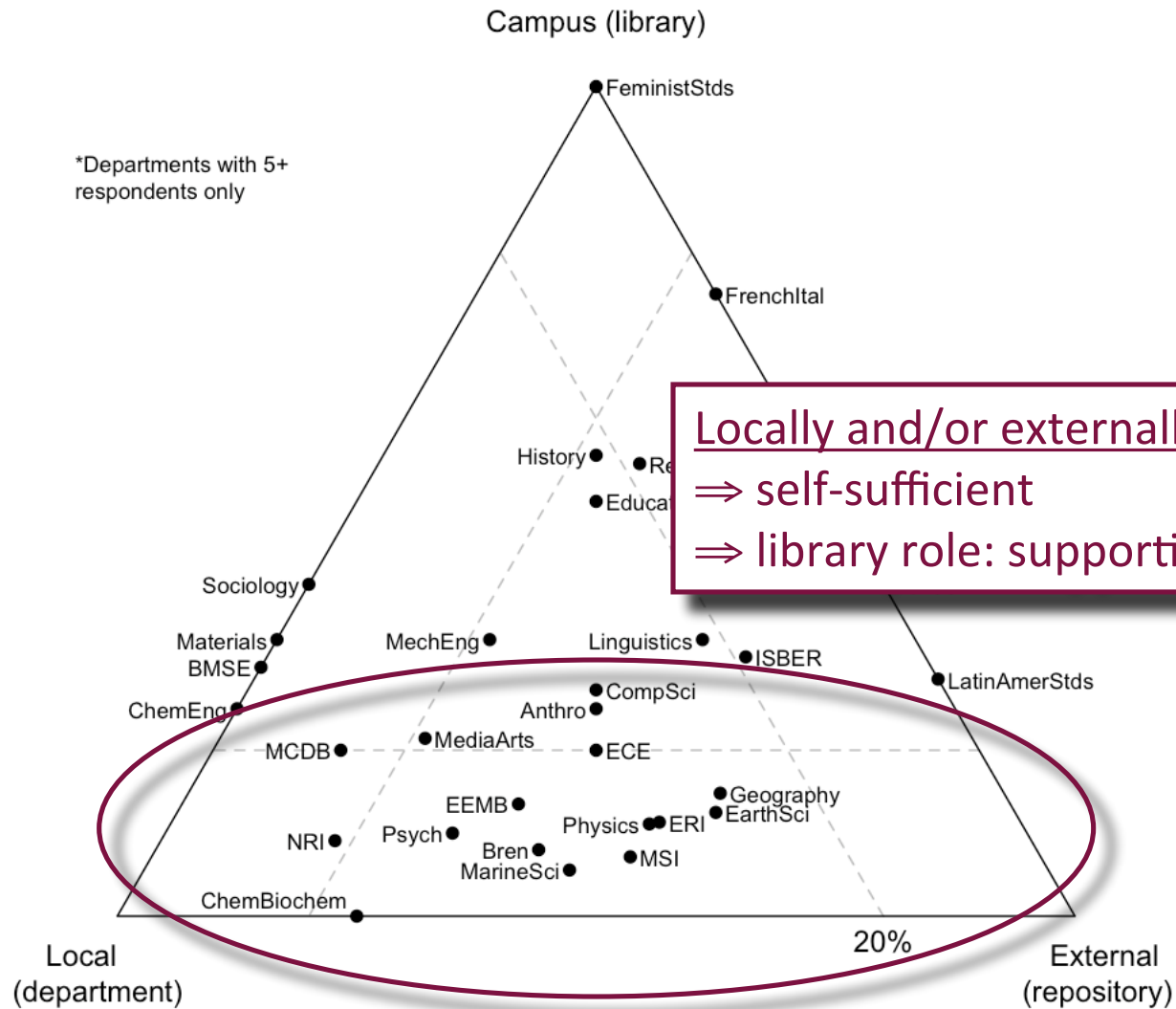
Responsibility spheres

- *“Local”*
 - Lab manager
 - Lab research staff
 - Department
- *“Campus”*
 - Campus IT
 - Campus library
- *“External”*
 - External research partner
 - External data repository
 - Funding agency
 - UC Curation Center (UC3)

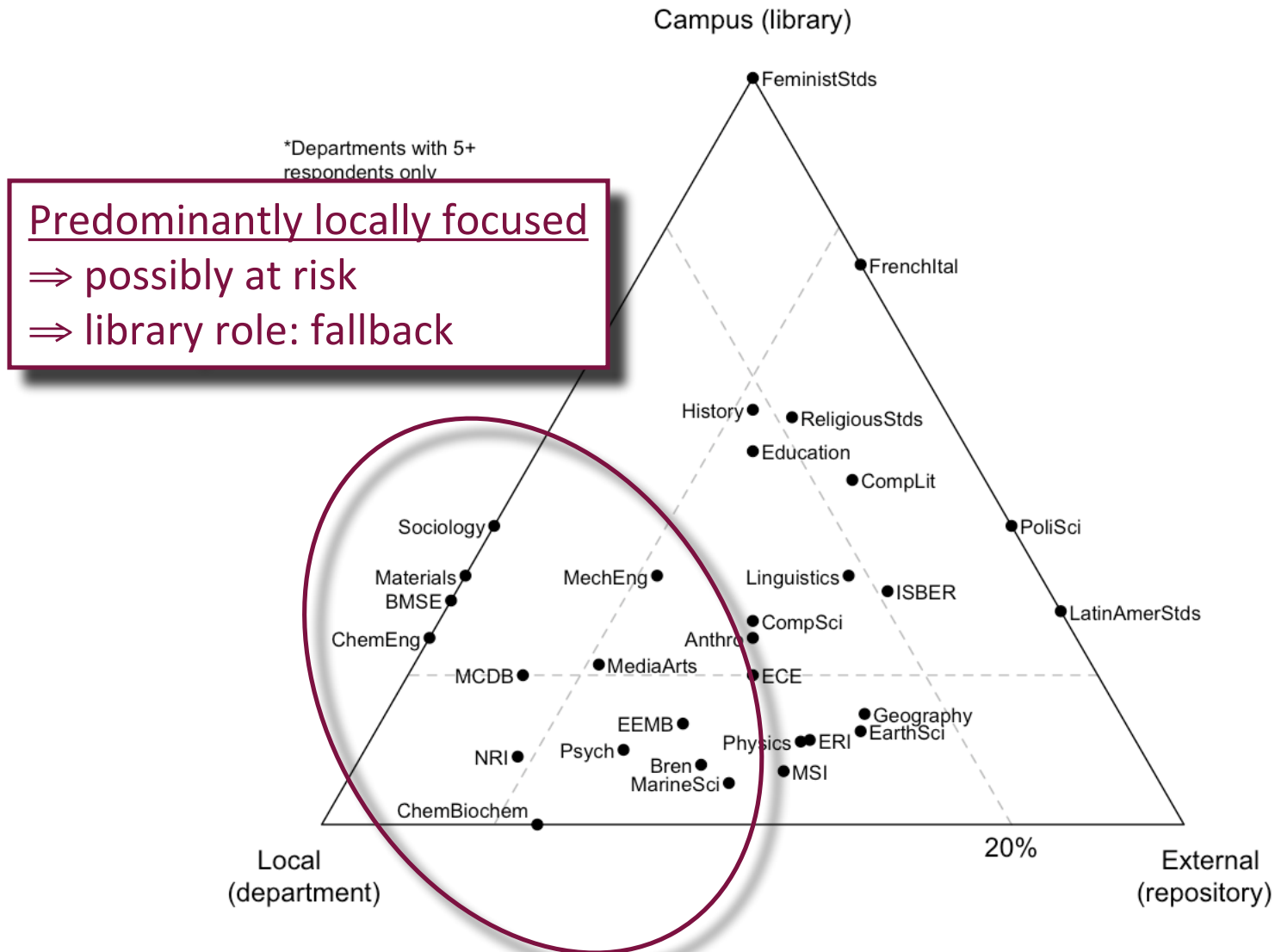
Distribution of departments* with respect to responsibility spheres



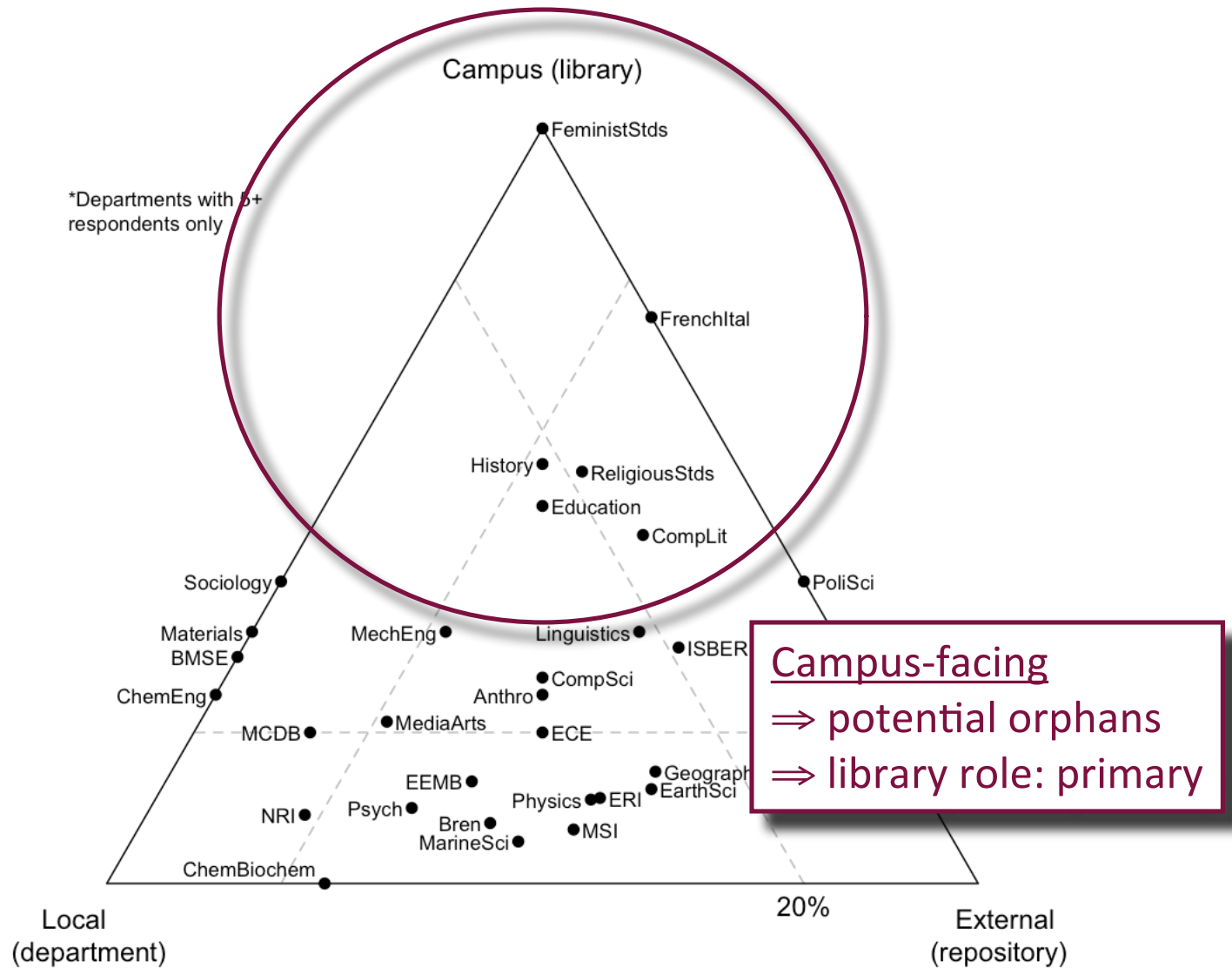
Distribution of departments* with respect to responsibility spheres



Distribution of departments* with respect to responsibility spheres



Distribution of departments* with respect to responsibility spheres

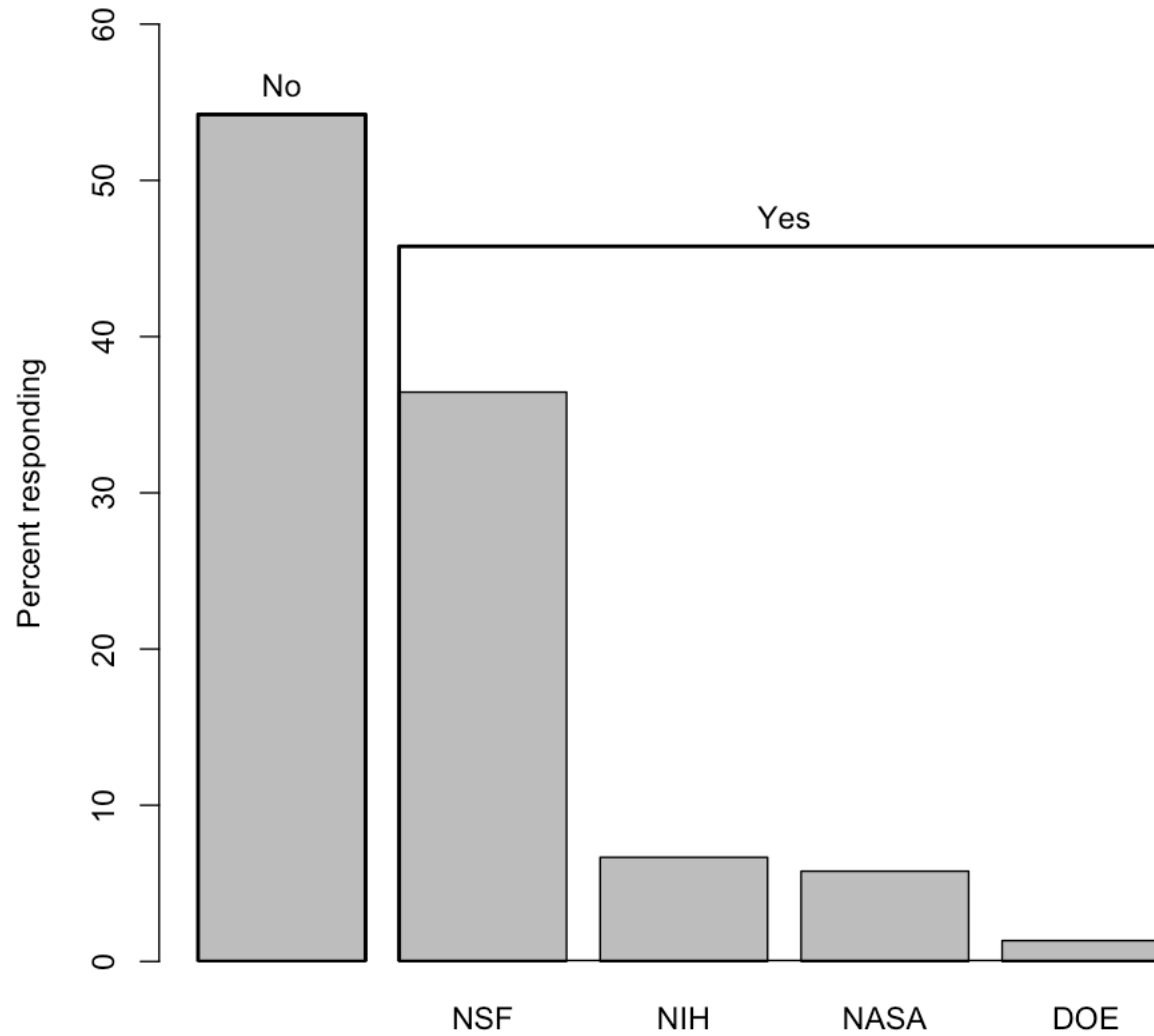


Responsibility vs department

- *Observations*
 - Departments differ with respect to placement of responsibility
 - Excluding “myself”, that is
- *Implications for Library*
 - Departments may need differing levels, types of support
 - Different roles, relationships to departments
 - Supporter
 - Fallback custodian
 - Primary custodian

Are you mandated to provide for (or otherwise participate in) the curation of your data, and if so, by which agencies?

Are you mandated to provide for the curation of your data, and if so, by which agencies?



Other responses

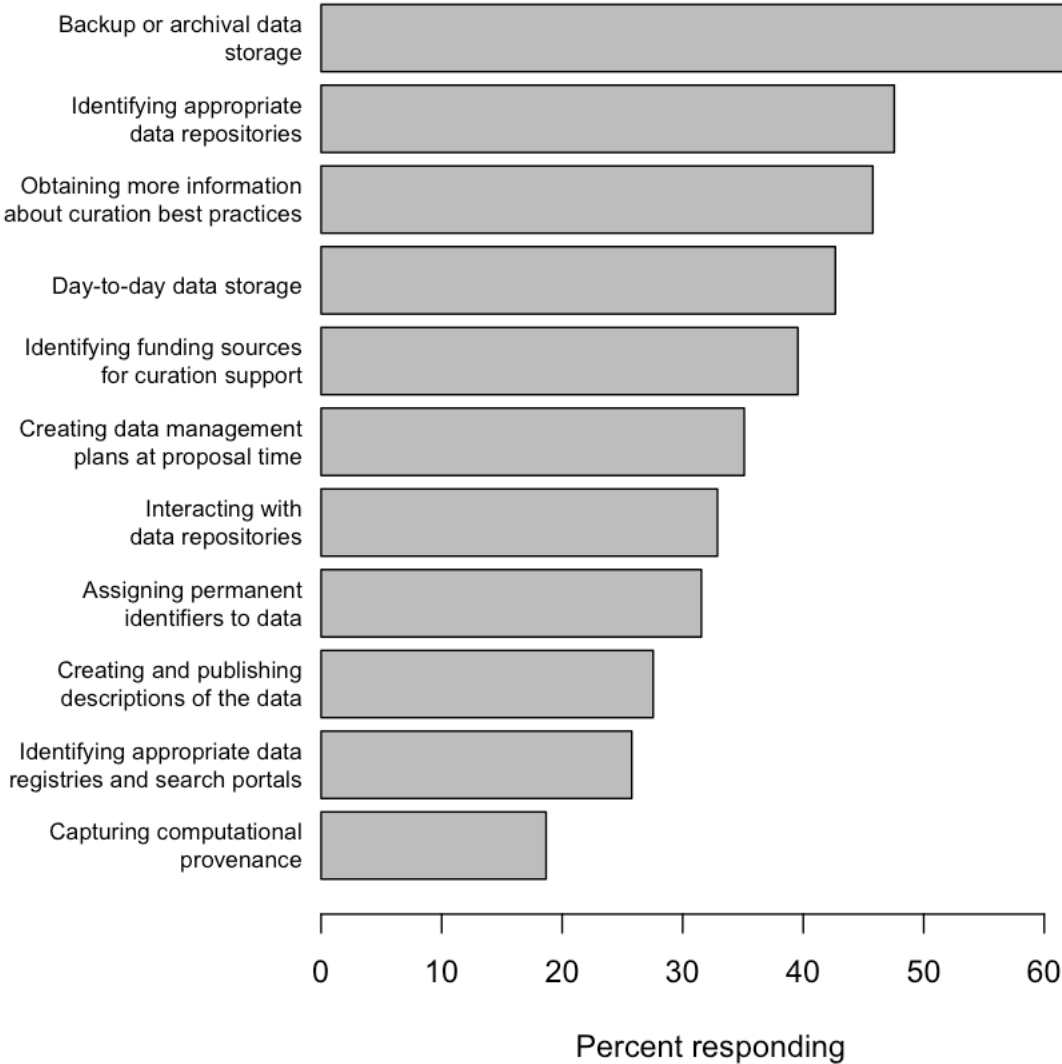
- *Additional agencies*
 - DOD, EPA, NEH, USGS, USGS-NEHRP
 - Department of the Interior
 - NARSAD
 - Social Science Data Archive, Moscow, Russia
 - ACLS
 - Instructional Development
 - Chancellor’s Outreach Advisory Board
 - “PHARMA COMPANIES WHO PROVIDE GRANTS”
- *Comments*
 - “NSF doesn’t have any enforcement policy”
 - “[agencies] expect universities to contribute to long term data storage”

Mandate question

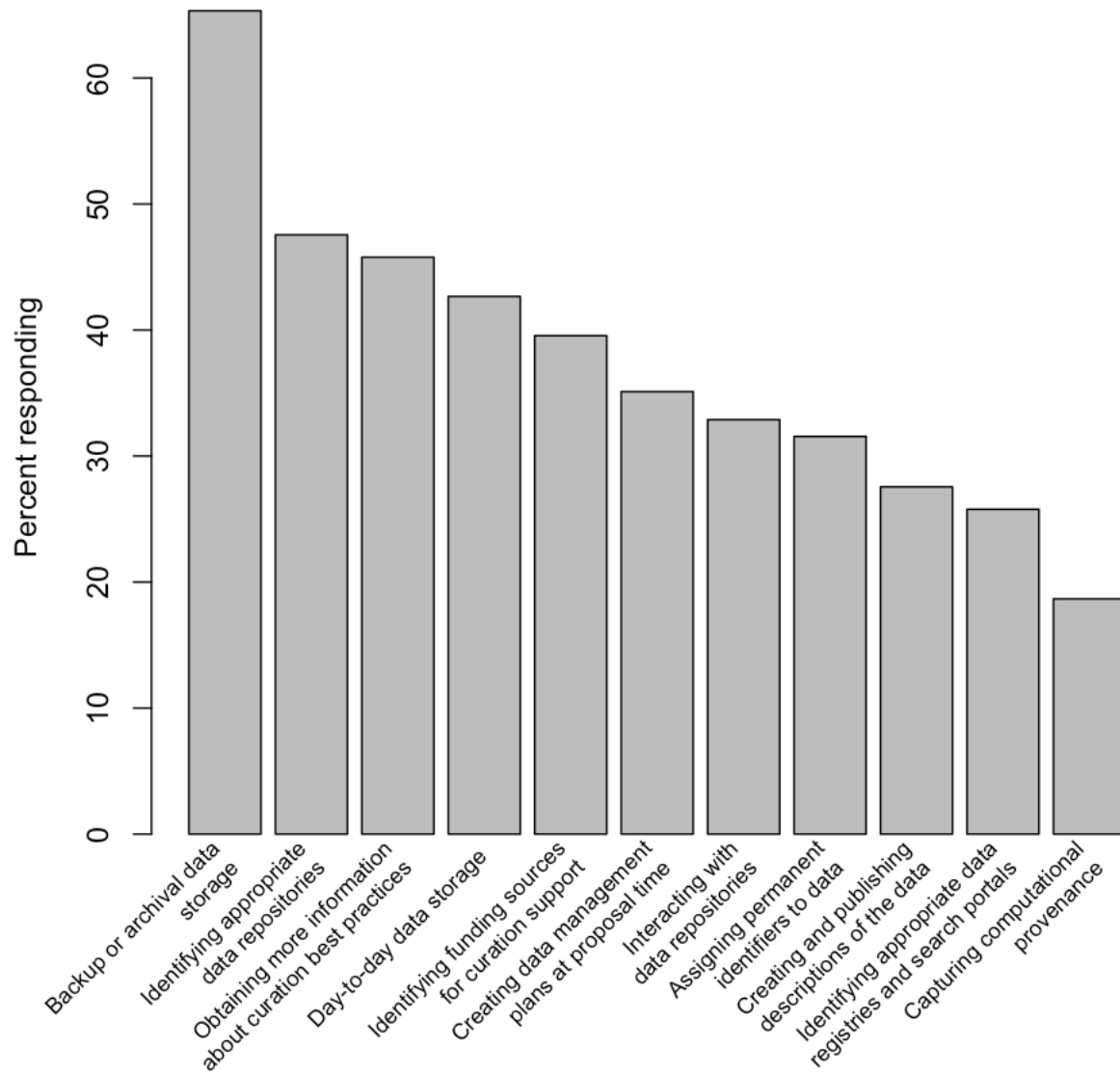
- *Observations*
 - Significant proportion of researchers are already mandated to address curation
 - Many sources of mandates
- *Implications for Library*
 - Researchers may need help...
 - Understanding mandates
 - Addressing mandates at proposal time
 - Identifying appropriate solutions, services, partners

What data management activities could you use help with, if any?

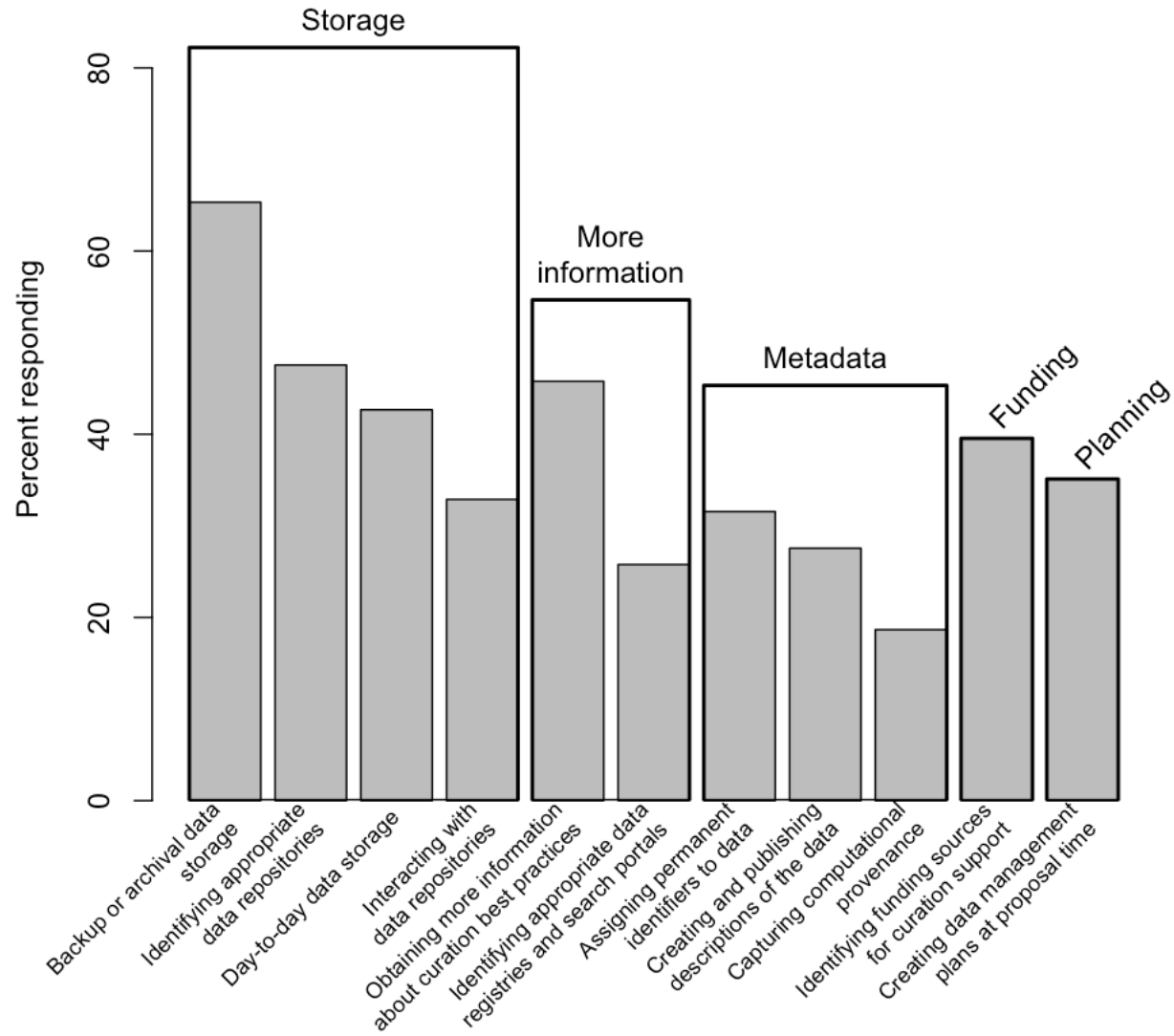
What data management activities could you use help with?



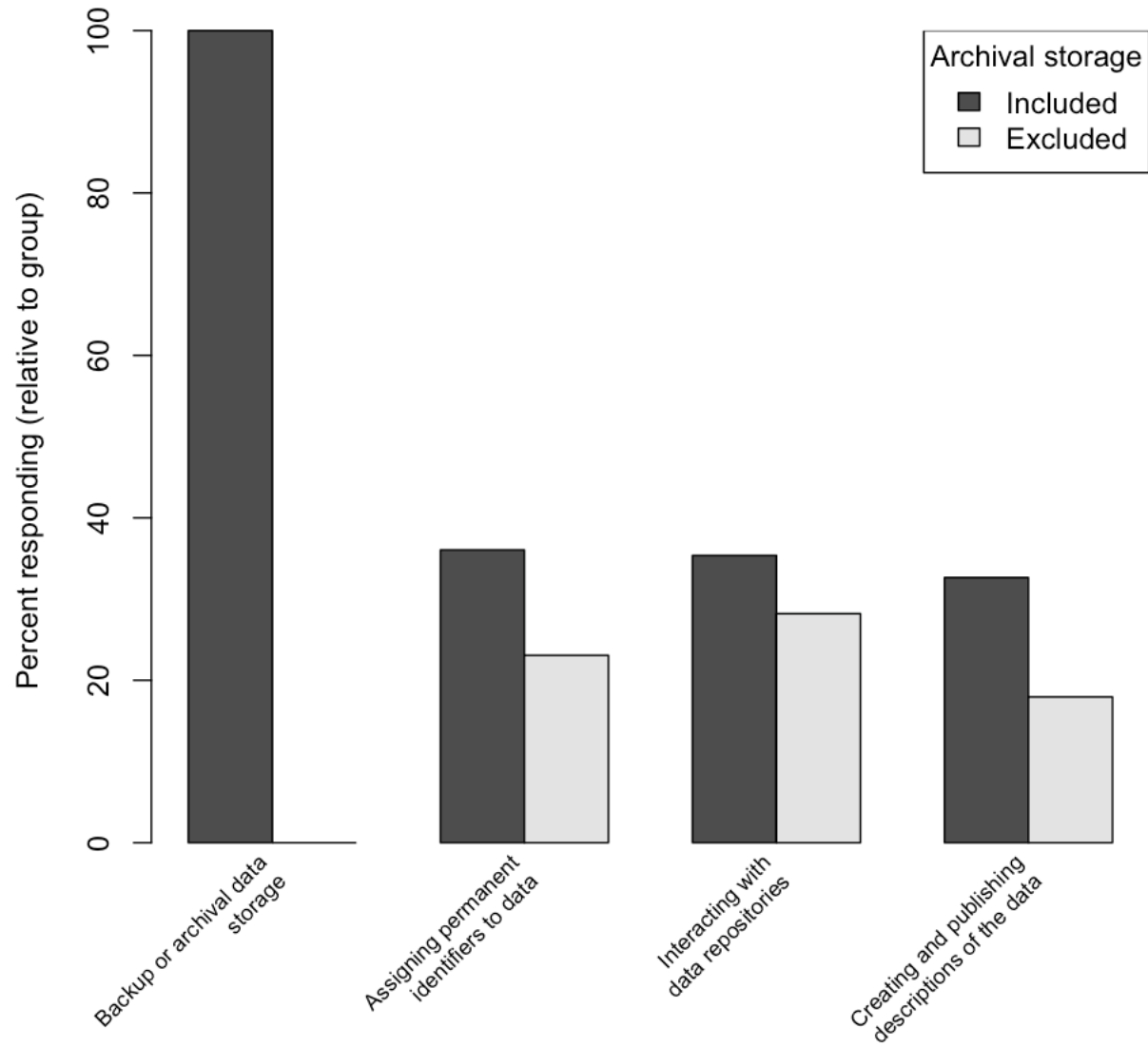
What data management activities could you use help with?



What data management activities could you use help with?



Metadata/repository help versus archival storage help



Help question

- *Observations*
 - All categories of help are desired
 - Storage dominates
 - Researchers may be unaware of the difficulties of archival storage versus day-to-day storage
- *Implications for Library*
 - Additional help needed related to repository use, metadata preparation

Other responses

- *Digitization*
 - “interviews still on cassette tape”
 - “floppy to flashdrive”
- *Education*
 - “most people will not understand what all of the above are. [...] There has to be a simpler way to describe.”
 - “info about best practices and standards for metadata”
 - “tech info on suitable data storage”
 - “managing ethical issues”
- *Access*
 - “Develop data access tools for the users (search portals,...)”
- *Storage*
 - “We need a modern cross-platform file repository”
- *Management*
 - “Maintenance of data connected with publications - articles”

Comments received

- *Kudos & interest*
 - “I welcome this initiative and for me it is very timely.”
 - “Very important. One can only speculate at the lost research insights and lost innovations that have occurred.”
 - “I got some help on my last grant from the office of research on a data management plan. Much appreciated.”
 - “Thank you for this study. I, and probably other faculty and staff, would benefit from learning more about data storage/dissemination options.”
 - “I [...] want to know what I can do to help preserve and migrate the digital data being produce by our organization.”

Comments received

- *Criticisms*
 - “Where is this survey coming from?”
 - “Some descriptions of what ‘curation’ is now being undertaken would help.”
- *Who the respondent is*
 - “I am a data manager on campus...”
 - “I am a librarian and administrator of several unit-wide databases...”
 - “I am a visiting scholar at UCSB.”

Comments received

- *Strategies*
 - “I save all my de-identified data in Gmail. I simply send anything I want to save to myself...”
 - “I use datadryad.org to publish my data with my papers and I am quite happy with their service.”
- *Content types*
 - “...digital data and physical collections...”
 - “...microfilmed archival material...”
 - “...video...”
 - “...e-mail...”
 - “...VHS collection...”
 - “...some rare [...] cartoons in french...”
 - “...‘data bank’ type data...”

Comments received

- *Problems*

- “For the most part, [...] computers and storage are not a problem, however, software tools are in short supply. One of the largest hurdles is finding tools to efficiently gather and assemble metadata into prescribed models. Libraries have been engaged with these issues (for print resources) for many years, whereas many fields of research are just starting this process. It would benefit those fields greatly to have libraries apply their experience and resources to the curation of digital resources.”
- “...almost everyone in my department needs lots of support with irreplaceable digital data that will be of great value to future generations. And we cannot take care of the task by ourselves.”
- “...[need] mechanisms that might help simplify the curation of my own laboratories data for the support of our individual research publications.”

Comments received

- *Suggestions*
 - “You may wish to include the campus Museum in your list of programs...”
 - “...strategy should complement the Keck Observatory Archive (of which UCSB is a part)...”
 - “I have never heard of UC3.”
 - “...Davidson should be recognized as central to digital research on our campus and its budget should be increased to foster faculty research...”

Conclusion

- *Survey largely confirmational*
 - Anecdotal observations representative
 - Value in having numbers
- *Takeaway points*
 - Data curation is problem for entire campus
 - Common model: researcher takes responsibility, but works with partner(s)
 - Many sources of curation mandates
 - Researchers want help with everything
 - But might underestimate need for help using archival storage
- *Implications for Library*
 - Take lead in identifying/formulating curation solutions
 - Interface researchers with those solutions
 - Host (parts of) solutions as called for, as resources permit
 - Provide education

Questions?