

The following is supplemental information to the 2013 Report “Institutional Readiness for Data Stewardship: Findings and Recommendations from the Research Data Assessment.”

SURVEY INSTRUMENT

Introduction

IRB Statement

Please report on the data generated from one research project throughout the entire survey. Choose a project of significant importance to your work (either in-process or complete), regardless of your intent to share or preserve the data.

PROJECT INFORMATION

Question	Available Response	Mandatory Y/N
1. Your name:	Free text	N
2. Project name:	Free text	N
3. Briefly describe the project you are using to answer this survey:	Free text	N
4. Briefly describe the data you are using to answer this survey:	Free text	N
5. Indicate your academic status:	<ul style="list-style-type: none">• Academic Faculty• Research Faculty• Academic Professional• Postdoctoral• Graduate Student• Other (please specify) – free text	N – multiple answers possible
6. Indicate the status of this project:	<ul style="list-style-type: none">• In planning stage• In progress• Completed• Other (please specify) – free text	N
7. a. Is this project grant funded?	<ul style="list-style-type: none">• Yes• No	
b. If yes, are you the PI/Co-PI?	<ul style="list-style-type: none">• Yes• No	

c. If yes, who is the funding agency/source?	•	
8. Identify the GT schools or research centers to which you belong:	<ul style="list-style-type: none"> • Drop down menu of GT schools & research centers • Free text 	N – multiple answers possible
9. Identify the GT schools or research centers, as well as external partners, affiliated with this project:	<ul style="list-style-type: none"> • Drop down menu of GT schools & research centers • Free text 	N – multiple answers possible

Data characteristics

10. Choose all of the following formats that best describe your research data (examples of specific file extensions are included):	<ul style="list-style-type: none"> • Audio (.aif, .au, .iff, .mp3, .wav) • Computer Aided Design / CAD (.dwg, .dxf, .pln) • Data (.csv, .dat) • Data - Statistical / SAS, SPSS (.sav, .sdq, .spv) • Data - XML (.xml) • Database (.db, .mdb, .pdb, .sql) • Geographic Information Systems / GIS (.gpx, .kml, .kmz) • Image (.bmp, .gif, .jpg, .png, .ps, .psd, .svg, .tif) • MATLAB (.m, .mat, .mex) • Portable document format (.pdf) • Spreadsheet (.wks, .xls) • Text (.doc, .docx, .log, .rtf, .txt) • Video (.avi, .mov, .mp4) • Web (.html, .xhtml) • Don't know • Other (please specify)- free text 	N – multiple answers possible
11. Indicate the approximate amount of data the project is expected to generate:	<ul style="list-style-type: none"> • 1 - 500 gigabytes (GB) • 500 - 1000 GB • 1 - 500 terabytes (TB) • 500 - 1000 TB • 1 - 500 petabytes (PB) • > 500 PB • Don't know • Other (please specify)- free text 	N – multiple answers possible

Data Management

12. Identify who manages the data associated with this project (check all that apply):	<ul style="list-style-type: none"> • PI or Co-PI • IT staff within your school or research center • Other designated person on project • Collaborative responsibility • External project partners • Third party data center • No one • Don't know 	N – multiple answers possible
--	---	-------------------------------

	<ul style="list-style-type: none"> • Other (please specify)- free text 	
13. Indicate where the data generated by this project are currently stored (choose all that apply):	<ul style="list-style-type: none"> • Hard drive of the instrument which generates the data • PC hard drive • External hard drive • Departmental server • CD/DVD • USB flash drives • Internet-based storage (e.g., cloud or grid storage) • Don't know • Other (please specify) – free text 	N – multiple answers possible
14. Indicate how often backups are made for the data associated with this project (check all that apply):	<ul style="list-style-type: none"> • Hourly • Daily • Weekly • Monthly • Annually • Never • Don't know • Other (please specify) – free text 	N – multiple answers possible
15. a. Identify how long you plan on keeping the data associated with this project:	<ul style="list-style-type: none"> • < 1 year • 1-5 years • 5-10 years • > 10 years • Indefinitely • Don't know 	N
b. Please explain why you plan on keeping the data for this amount of time:	Free text	N
16. a. Designate if you have a data management plan or policy:	<ul style="list-style-type: none"> • Yes (proceed to question 14b) • No (proceed to question 14c) • Don't know (proceed to question 15) 	N
b. If you do have a data management plan or policy, indicate the reasons why (check all that apply):	<ul style="list-style-type: none"> • Required by IRB • Required by funding agency • Required by school or research center • Good practice • Other (please specify) – free text 	N—multiple answers possible
c. If you do not have a data management plan or policy, indicate the reasons why not (check all that apply) :	<ul style="list-style-type: none"> • Lack of information about data management plans • Not necessary • Other (please specify)- free text 	N—multiple answers possible

Data Use

17. Identify all of the following who have access to all or part of the raw data associated with	<ul style="list-style-type: none"> • Researchers, students, and staff working on the project • Other members of the affiliated 	N – multiple answers possible
---	--	-------------------------------

<p>this project (check all that apply):</p>	<ul style="list-style-type: none"> GT School or Research Center • Researchers at other institutions • Project sponsors • General public • Don't know • Other (please specify) – free text 	
<p>18. Identify all of the following who you would like to share raw project data with, if you had the ability to do so (check all that apply):</p>	<ul style="list-style-type: none"> • Researchers, students, and staff working on the project • Other members of the affiliated GT School or Research Center • Researchers at other institutions • Project sponsors • General public • Other (please specify) – free text 	<p>N – multiple answers possible</p>
<p>19. Choose all of the following reasons why you do not share the raw data associated with this project (check all that apply):</p>	<ul style="list-style-type: none"> • Confidential, proprietary or classified information • Intellectual property concerns • Possible misinterpretation of data • Time or effort required to make data available • Lack of appropriate tools for sharing or publishing data • Other (please specify) – free text 	<p>N – multiple answers possible</p>
<p>20. Choose all of the following methods you use to share all or part of the raw data associated with this project (check all that apply):</p>	<ul style="list-style-type: none"> • Collaborative web space (e.g., wiki, blog, Google Docs) • Data portal or database driven web site • Email • External storage device (e.g., USB drive, CD/DVD) • Hard copy or print • Don't share data • Don't know • Other (please specify) – free text 	<p>N – multiple answers possible</p>

Support for Research Data Management

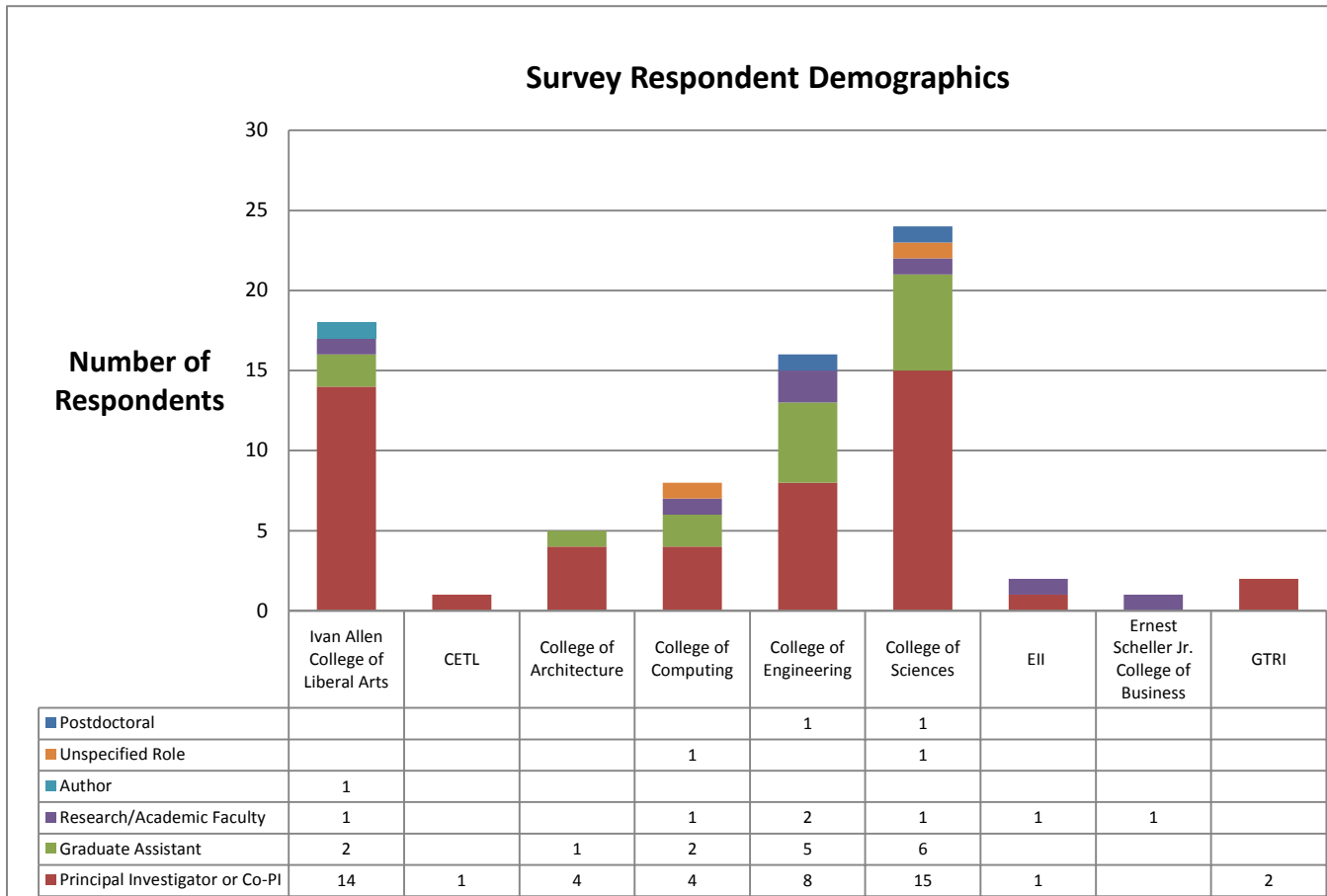
<p>21. Identify which of the following services might be useful in regards to the management of research data (check all that apply):</p>	<ul style="list-style-type: none"> • Assistance meeting data sharing or data management requirements of funding agencies • Information about developing a data management plan • Information regarding data management best practices • Assistance selecting data to preserve for the long-term • Tools for sharing research data as research is on-going • Data storage & backup while research is on-going 	<p>N – multiple answers possible</p>
--	--	--------------------------------------

	<ul style="list-style-type: none"> • Long-term data access and preservation • Assistance applying metadata to research data • Assistance finding and accessing data resources • Information about citing data resources • None of the above • Other (please specify) 	
22. Would you be willing to participate in a follow-up interview regarding research data management?	<ul style="list-style-type: none"> • Yes • No 	N
23. Please provide any additional comments regarding research data management, potential data curation services, or this survey:	Free text	N

FULL SURVEY RESULTS

Questions that asked for personally identifying information have been removed

5. (Indicate your academic status) and 7. (Identify the GT schools or research centers, as well as external partners, affiliated with this project):



6. Indicate the status of this project:

Status	Response	%
Planning Stages	2	3%
Completed	14	18%
In progress	57	75%

8. Choose all of the following formats that best describe your research data (Choose all that apply):

Data Formats	Response	%
Text (.doc, .docx, .log, .rtf, .txt)	52	68%
Spreadsheet (.wks, .xls)	40	52%
Data (.csv, .dat)	37	48%
Image (.bmp, .gif, .jpg, .png, .ps, .psd, .svg, .tif)	34	44%
Scanned documents (.pdf)	30	39%
Video (.avi, .mov, .mp4)	23	30%
Data - Statistical / SAS, SPSS, STATA (.sav, .sdq, .spv, .dta)	21	28%
Web (.html, .xhtml)	19	25%
Database (.db, .mdb, .pdb, .sql)	17	22%
Audio (.aif, .iff, .mp3, .wav)	14	18%
Data - XML (.xml)	12	16%
Geographic Information Systems / GIS (.gpx, .kml)	6	8%
MATLAB (.m, .mat, .fig)	4	5%
Instrument-specific data format	3	4%
Computer Aided Design / CAD (.dwg, .dxf, .pln)	2	3%
Homegrown format	2	3%

The following formats were mentioned in the survey by only one participant: netCDF(.ncf), Abaqus outputs (.odb), Binary format, X-ray diffratograms (.xrdml), physical papers or materials, Vicon Motion Analysis (.c3d), genetic sequence reads, DSP code library, wave form, Atlas.ti (.hpr6, .hpr7, .atlc), scholarly materials in Lexis.

9. Indicate the approximate amount of data the project is expected to generate:

Volume of data	Response	%
1 – 500 gigabytes (GB)	39	51%
500-1000 gigabytes (GB)	15	19%
1 – 500 terabytes (TB)	13	17%
Don't know	9	12%
Smaller than 1 GB	1	

10. Identify who manages the data associated with this project (Choose all that apply):

Party responsible for data management	Response	%
PI or co-PI	56	73%
Collaborative Responsibility	21	28%
Designated person on project (not PI)	20	26%
IT staff within the school or research center	11	14%
External project partners	8	10%
Graduate Assistants	6	8%
Third party data center	2	3%
No one	1	1%
Don't know	1	1%
Research Faculty	1	1%

11. Indicate where the data generated by this project are currently stored (Choose all that apply):

Location where data are stored	Response	%
PC hard drive	55	71%
Departmental or Lab server	35	45%
External hard drive	33	43%
Hard drive of the instrument which generates the data	32	42%
Internet-based storage (e.g., cloud or grid storage)	18	23%
USB flash drives	17	22%
CD/DVD	7	9%
Physical materials stored in GT office or lab	5	6%
Shared Drive Hosted/Maintained by an outside collaborator	4	5%
Storage Area Network	2	3%
HPC Cluster	2	3%

12. Indicate how often backups are made for the data associated with this project (Choose all that apply):

Data Backups	Response	%
Daily	23	30%
Weekly	20	26%
Don't Know	14	18%
Monthly	11	14%
Annually	5	6%
Whenever IT/Storage provides backs up	4	5%
Irregularly	4	5%
Never	4	5%
Hourly	3	4%
Varies depending on data or project	3	4%

13a. Identify how long you plan on keeping the data associated with this project:

Length of Data Retention	Response	%
1-5 years	27	35%
Indefinitely	18	23%
5-10 years	15	19%
>10 years	9	12%
Don't know	7	9%
< 1 year	1	1%

13b. Please explain why you plan on keeping the data for this amount of time (these responses have been broken up based on the length of time identified in the first part of the question):

Reason for Retaining Data < 1Year	Response	%
Length of retention was established in terms of use for gaining access to data	1	1%

Reason for Retaining Data for 1-5 Years	Response	%
Length of retention is consistent with length or goal of the project	7	9%
This length will allow for re-use or future use of the data, by the original researcher or by others	3	4%
Don't expect to be able to use or to want to use data after this point	3	4%
Don't expect to be able to use or to want to use data after this point	3	4%
This is the length of time needed to complete data analysis and publish	2	3%
Data will ultimately be replaced by newer, higher quality data in the future	2	3%
To allow for audits or sponsor inquiries	2	3%

Reason for Retaining Data for 5-10 Years	Response	%
This length will allow for re-use or future use of the data, by the original researcher or by others	5	6%
Length of retention is consistent with length or goal of the project	2	2%
This is the length of time needed to complete data analysis and publish	2	2%
Data will used in future academic courses	1	1%
Professional obligation	1	1%

Reason for Retaining Data for >10 Years	Response	%
Length of time will allow for re-use or future use of the data, by the original researcher or by others	6	8%
The data are rare or irreplaceable	2	2%
Length of retention is consistent with length or goal of the project	1	1%

Reason for Retaining Data Indefinitely	Response	%
This length will allow for re-use or future use of the data, by the original researcher or by others	12	
Cost for Storage is negligible	3	
Ability to reproduce results or support claims	2	
Professional norm or obligation	1	1%
Length of time needed for analysis and publication	1	

14a. Designate if you have a data management plan or policy:

Data Management Plan	Response	%
No	48	62%
Yes	18	23%
Don't know	10	13%

14b. If you do have a data management plan or policy, indicate the reasons why:

Reasons to have a data management plan	Response	%
Required by funding agency	5	6%
Good practice and/or helps research	5	6%
Someone in lab or collaboration wants it	4	5%
Community expectation	1	1%
IRB	1	1%
Preparation for sharing data	1	1%

14c. If you do not have a data management plan or policy, indicate the reasons why not:

Reasons not to have a data management plan	Response	%
Not necessary	27	35%
Lack of information about data management plans	26	34%
Other, unspecified reasons	5	6%

15. Identify all of the following who have access to all or part of the raw data associated with this project:

Current Access to Research Data	Response	%
Researchers, staff, and students working on the project	72	94%
Project Sponsors	13	17%
Researchers at other institutions	11	14%
General Public	7	5%
Co-author	2	3%
Web Designer	1	1%

16. Identify all of the following who you would like to share raw project data with, if you had the ability to do so:

Who should have access to research data?	Response	%
Researchers, students, and staff working on the project	58	75%
Other members of the affiliated GT School or Research Center	20	26%
Researchers at other institutions	32	42%
Project Sponsors	16	21%
General Public	9	12%
No one	2	2%
Future Co-Authors	1	1%

17. Choose all of the following reasons why you do not share the raw data associated with this project:

Reasons to Not Share Data	Response	%
Confidential, proprietary, or classified information	34	44%
Time or effort required to make data available	30	39%
Intellectual Property concerns	23	30%
Lack of appropriate tools for sharing or publishing data	22	29%
Possible Misinterpretation of data	21	27%
Data will not be understood or useful to people not originally on the project	3	4%
Data used in research were already public data	2	2%
Not common practice in community or discipline	1	1%
Not necessary	1	1%
Want to ensure that original researcher gets credit	1	1%
No funding to preserve or share data	1	1%

18. Choose all of the following methods you use to share all or part of the raw data associated with this project:

Methods for Sharing Data	Response	%
Email	45	58%
Collaborative web space (e.g., wiki, blog, Google Docs)	29	38%
External storage device (e.g., USB drive, CD/DVD)	24	31%
Hard copy or print	19	25%
Data portal or database driven web site	17	22%
Don't share data	9	12%
Networked computers or server access	6	8%
Dropbox	3	4%
Don't know	2	2%

19. Identify which of the following services might be useful in regards to the management of research data:

Desired Service	Response	%
Data storage and preservation	59	77%
Tools for sharing research data	49	64%
Information regarding data management best practices	40	52%
Information about developing a formal data management plan or other data management policies	34	44%
Assistance regarding data management best practices	29	38%
Assistance meeting data sharing and/or data management requirements of funding agencies	28	36%
Assistance finding and accessing data resources	25	32%
Information about citing data resources	15	19%

APPENDIX D. INTERVIEW RESULTS

Have a Data Management Plan?	Number of Interviewees
Yes	16
No	10

Conditions on Sharing Data/Reasons not to Share Data	Number of Interviewees
Data are sensitive (human subjects, intellectual property, planned commercialization of data, etc.)	11
Data are not of use to anyone else (too idiosyncratic, student work, etc.)	6
Will not share data until work is published or they are done with data	6
Person using data must give attribution to original data creator, or include as collaborator or co-author	6
Sharing data is too time and resource intensive	3
Concern about someone misusing/misinterpreting data	2
Asked by someone directly to share	2
Data were expensive to generate/create/collect	2

Motivations to Share Data	Number of Participants
Re-use of data in the future (Ask new questions, supplement research, etc.)	10
Required by funding agency	5
Altruism/Desire to contribute to science or the public	4
Reproducibility/Reduce Fraud	3
Required by or asked for by a journal	2
Data were expensive	2

INTERVIEW INSTRUMENT

1. What is the story of the data?
 - a. Did anyone fund your research & the data you hold?
 - b. Why was this data created or collected?
 - c. How is the data processed and/or analyzed?
 - d. Are others involved in these processes (e.g., graduate students, lab assistants)
2. What tools (hardware or software) are used in collecting, generating, or analyzing your data?
For example – remote sensor, data logger
 - a. What tools are required to utilize the data? (e.g., GIS, my own software, Matlab)
3. Do you gather and manage data from others (secondary data)?
 - a. Are there data sets or collections that you know exist, but cannot access due to subscription or other barriers? If yes, what are these?
4. How do you share data with your research partners?
 - a. Are any of your collaborators in another lab or department?
 - b. Are any collaborators outside of Tech?
 - c. How adequate do you consider your methods for sharing data with your research partners to be?
5. How do you organize the data produced from your research?
 - a. Do you document how the data is generated or collected?
 - b. Do you document when files are modified or who modified the files?
 - c. Do you use keywords to describe the data?
 - d. Do you apply any specific metadata standards to your data?
 - e. How is the metadata generated?
 - f. Do you follow & document file-naming standards?
 - g. Do you maintain research-related data in non-digital forms such as paper, photographs, video or audio tapes, slides, etc.?
 - h. How adequate do you consider your methods for organizing data to be?
6. How should your data be shared with others – i.e., researchers not involved with the project or the general public?
 - a. At which stage of your research would you be willing to share your data?
 - b. How do you provide access to your data? (see survey results)
 - c. Do you place any conditions on sharing your data? (e.g., keep data confidential)
 - d. Have you been asked to make data openly available?
 - e. How do you imagine that people would find your data?
 - f. Do you need security measures due to sensitive or proprietary information?
 - g. How adequate do you consider your methods for sharing data?
7. How important is it for you or others to have access to your data over the long-term?
 - a. Are you interested in providing long-term access to all or part of your data? (e.g., Raw data? Processed data? Analyzed data?)
 - b. Who are the potential audiences for your data?
 - c. How could your data be used, reused, or repurposed?
 - d. Do you deposit your research data with a data service or repository?
 - e. IF YES: what made you choose this particular repository?
 - f. Would you be willing to deposit your data in a repository? At what stage in its lifecycle?
 - g. What preparations or actions would need to take place before your data could be ingested into a repository?

- h. Have you heard about SMARTech, the GT institutional repository? Have you considered using SMARTech to provide access to your data?
 - i. Are you aware of any discipline or subject-specific data repositories in your field?
 - j. How adequate do you consider your methods for providing long-term access to your research data?
- 8. How important is it for you to archive/preserve your data?
 - a. Are you interested in preserving all or part of your data?
 - b. Is this the same data that you are willing to share with others?
 - c. Have you ever lost data – how did it happen (e.g., old files formats, size of files, poor file naming conventions, departure of team member)?
 - d. How important would it be for you to have the ability to migrate the datasets into new formats over time?
 - e. To audit the datasets to ensure structural integrity over time?
 - f. How adequate do you consider your methods for preserving?
- 9. Does a funding agency mandate that you preserve or provide access to your research data?
 - a. Are you aware of the NSF Data Management Plan requirement or the NIH data sharing requirement?
 - b. How do you think these mandates will affect your data management practices?
 - c. What makes data management a low/medium/high priority for you?
- 10. If there were a magic wand that would make your work with research data easier, what would it allow you to do?
 - a. What is the biggest need you have with regards to managing your research data?