

Data Management Survey Handout for Research Computing Day

2-1-2014

Cynthia Dancel

University of Central Florida, cindy@ucf.edu

Find similar works at: <http://stars.library.ucf.edu/lib-docs>

University of Central Florida Libraries <http://library.ucf.edu>

Recommended Citation

Dancel, Cynthia, "Data Management Survey Handout for Research Computing Day" (2014). *Libraries' Documents*. 144.
<http://stars.library.ucf.edu/lib-docs/144>

This Handout is brought to you for free and open access by STARS. It has been accepted for inclusion in Libraries' Documents by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.



The UCF Research Data Management Survey Report

November 2013

Research Computing Day at IST

February 11, 2014

Penny Beile, PhD

UCF Libraries

Acknowledgments

Information Technologies and Resources

Joel Hartman

Institute for Simulation and Training

Brian Goldiez

Computer Services and Telecommunications

Bob Yanckello

Jim Ennis

Tim Larson

Office of Research and Commercialization

Ivan Garibay

Josh Roney

UCF Libraries

Barry Baker

Sai Deng

Lee Dotson

Selma Jaskowski

UCF Libraries Survey Administration Team

Bobby Ciullo

Erica England

Joel Lavoie

Special thanks to University of Florida colleagues Mark Sullivan and Laurie Taylor for sharing a draft of their data management survey and allowing us to modify it for our purposes.

Report

Purpose

The Research Data Management Survey was conducted at the suggestion of members from Computer Services and Telecommunications (CS&T), the Institute for Simulation and Training (IST), and the Libraries. The purpose was to gain insight into faculty research data practices and needs to better inform decision-making about campus-wide research data management services and support.

Questionnaire

Laurie Taylor and Mark Sullivan, from the University of Florida Libraries, provided a copy of an instrument that was being used at UF to survey faculty about their research data management practices and needs. This survey was reviewed by the Vice Provost for Information Technologies and Resources and members from CS&T, IST, the Office of Research and Commercialization (ORC), and the Libraries, and was subsequently modified based on their feedback. The final survey contained 33 items.

Sample

Josh Roney (ORC) provided the names and email addresses of 524 researchers who had received research funding (listed in the ARGIS database) since January 1, 2010. At request, ORC also provided a list of faculty who attended a recent research presentation. After reconciling the names with the existing survey panel an additional 25 people were added to the distribution list.

Distribution

The survey was uploaded into Qualtrics and the initial invitation was emailed to participants on September 30. Three reminders were sent each following Monday, until the survey closed on October 30. In sum the survey was sent to 549 individuals, however thirteen emails bounced back, one person replied that she was no longer at UCF, and another person responded that she only submitted a research proposal for others and could not complete the questionnaire, resulting in 534 valid recipients. Of them, 110 (20.6%) opened the survey, however thirteen participants did not select any responses, leaving 97 (18.2%) who partially or fully completed the survey. All responses are reported.

Results

Results of the survey are summarized in the following pages and were compiled from Qualtrics reports and a raw data file. Contact Penny (pbeile@ucf.edu) for questions or additional analysis. A comma delimited Excel file, with identifying information stripped, is also available.

Several recipients contacted the survey administrator to offer insight into their data management strategies. Of particular note is the work conducted by FSEC; one project, Building America Partnership for Improved Residential Construction provides a searchable frontend to the research data. (<http://www.fsec.ucf.edu/en/research/buildings/ba-pirc.htm>).

SECTION 1: DEMOGRAPHICS

1. What is your professional status?

Of the 94 people who responded to the question, the overwhelming majority indicated that they hold *Faculty* status (n= 79, 84%). This is followed by *Administrator* (n=9, 10%), *Staff* (n=2, 2%), *Postdoc* (n=2, 2%), *Graduate student* (n=1, 1%), and *Retired faculty* (n=1, 1%). No residents or undergraduate students are represented. Three people did not answer this question.

Answer	Response	%
Faculty	79	84%
Administrator	9	10%
Staff	2	2%
Postdoctoral Fellow	2	2%
Resident	0	0%
Graduate Student	1	1%
Undergraduate Student	0	0%
Other (retired)	1	1%
Total	94	100%

2. What is the size of the research team that you typically work with?

Of the 93 people who responded to the question, the majority (n=57, 61%) selected *1-5 people* as the size of the research team they typically work with. This was followed by *6-12 people* (n=25, 27%) and *more than 12 people* (n=11, 12%).

Answer	Response	%
1-5 people	57	61%
6-12 people	25	27%
More than 12 people	11	12%
Total	93	100%

3. Do you collaborate with researchers from other institutions?

Of the 93 people who responded to the question, the majority (n=84, 90%) indicated that they collaborate with researchers from other institutions; only nine (10%) noted that they do not.

Answer	Response	%
Yes	84	90%
No	9	10%
Total	93	100%

4. What college and/or institute(s) or center(s) are you affiliated with? Check all that apply.

The 94 respondents selected multiple answers, for a total of 118 affiliations spread across 21 campus units. Every college except for the College of Business Administration was represented.

Participants who selected *Other* (n=10) were asked to indicate their affiliation. These areas are not noted in the table below, but include: AMPAC, Center for Humanities and Digital Research, Environmental Systems Engineering Institute, Florida Center for Nursing , Florida Space Center (2), the Libraries (2), NanoScience Technology Center, and Undergraduate Studies.

Answer	Response	%
College of Arts and Humanities	8	9%
Burnett Honors College	3	3%
College of Business Administration	0	0%
College of Education and Human Performance	3	3%
College of Engineering and Computer Science	13	14%
College of Health and Public Affairs	15	16%
College of Medicine	11	12%
College of Nursing	9	10%
College of Optics and Photonics	4	4%
Rosen College of Hospitality Management	2	2%
College of Sciences	22	23%
Florida Solar Energy Center	12	13%
Institute of Simulation and Training	6	6%
Other	10	11%

5. What department(s) are you affiliated with?

Of the 86 responses, 50 unique departments were represented. The number of respondents from each department is not provided in order to maintain anonymity.

Departments	
Advanced Materials Processing & Analysis Center	Information Literacy and Outreach
Anthropology	Institute for Simulation and Training
Biology	International Studies
Building Research	Materials Science and Engineering
Burnett School of Biomedical Sciences	Mechanical and Space Engineering
Business	Medical Education
Center for Autism and Related Disabilities	Microbiology and Molecular Biology
Chemistry	Music
Child, Family, Consumer Services	Nanoscience Technology Center
Civil, Environmental & Construction Engineering	Nicholson School of Communication
Communication Sciences and Disorders	Nursing
Computing and Information Technology	Office of Research and Commercialization
CREOL	Philosophy
Criminal Justice	Physics
Electrical Engineering and Computer Science	Psychology
English	Public Administration
Florida Center for Nursing	School of Social Work
Florida Space Institute	Sociology
Florida Solar Energy Center	Solar Systems Research Division
Graduate	Solar Technologies Research Division
Health Management and Informatics	School of Visual Arts and Design
Health Professions	Teaching, Learning, and Leadership
Health Services	UCF Police Department
History	Women's Studies
Hospitality	Writing and Rhetoric

6. If your research is or has been supported by any funding agency or agencies in the past five years, please list them.

The 84 respondents identified a total of 120 funders/funding agencies. Only 19 agencies appeared more than once. Funding agencies and the number of times listed are summarized in the following table. No number indicates the agency was mentioned only once.

Funding agencies	
AGDF	National Geographic
Air Force Office of Scientific Research	National Institute of Aging
Air Force Research Laboratory (Ball Aerospace) (2)	National Institute of General Medical Sciences
American Academy of Real Estate	National Institute of Health (8)
American Association of University Women	National Institute of Justice
American Chemical Society	National Institute of Mental Health
American Lung Association	National Institute of Neurological Disorders & Stroke
American Nurses Foundation	National Institute of Nursing Research
American Speech-Language-Hearing Foundation	National Institute on Drug Abuse
Army Research Institute (2)	Nat'l Inst on Minority Health & Health Disparities
Army Research Laboratory	National Library of Medicine

Army Research Office	National Oceanic & Atmospheric Admin (3)
Army Reserve Education Assistance Program	National Renewable Energy Lab (2)
Atlantic Housing Partners	National Science & Technology Council
Austin Tsutsumi ATA Honolulu, HI	National Science Foundation (24)
Autism Speaks	Nemours Hospital
Bauer Foundation Corp.	North Atlantic Treaty Organization
BlueCross BlueShield of Florida	NYSTAR
Brown and Caldwell	Office of Juvenile Justice & Delinquency Prevention
Carollo Engineers, Inc.	Office of Naval Research
City of Edgewater, FL	Office of the Attorney General
City of Orlando, FL (2)	Orange County Government
City of Palmetto, FL	Orange County Health Department
City of Sarasota, FL	Orange County Utilities
County of Maui Department of Water Supply, HI	Owens Corning
Defense Advanced Research Projects Agency	Pacific Northwest National Laboratory
Department of Children & Families	Philips
Department of Defense (4)	Plasmonics
Department of Education	Polk County Utilities, FL
Department of the Navy	Research Corporation for Science Advancement
Electric Power Research Institute	Robert Wood Johnson Foundation
Environmental Protection Agency	RosTek Associates, Inc.
Federal Emergency Management Agency	Sandia National Labs (2)
Florida Alliance for Assistive Services & Technology	Siemens
Florida Blue Foundation	Solar Rating & Certification Corporation (2)
Florida Council on Compulsive Gambling	Southwest Florida Water Management District
Florida Department of Education	Scientific Research Corporation
Florida Department of Health (3)	St. John's River Water Management District (2)
Florida Energy Systems Consortium	State of Florida (3)
Florida Hospital (3)	The Nature Conservancy
Florida Northwest Health Foundation	The Substance Abuse and Mental Health Services Administration
Florida Sea Grant	
Florida Space Institute (2)	UCF College of Medicine
Harn R/P Systems, FL	UCF College of Nursing (2)
Hilton Orlando	UCF Libraries' Professional Development Award
Institute of International Education	UCF Office of Research & Commercialization (4)
Interactive Management Group	UCF School of Public Administration
International Research & Exchanges	University of Oregon
Intertek	US Air Force
Kennedy Space Center	US Army Corps of Engineers
Kimley Horn, FL	US Department of Agriculture
King of Fans, Inc.	US Department of Energy (18)
Library of Congress	US Department of Health & Human Services
LIFE Institute	US Department of Justice
MacArthur Foundation	US Department of Transportation
Magruder Foundation	US National Park Service
NASA (8)	US Navy Naval Air Systems Command
National Art Education Foundation	Visit Orlando

National Endowment for the Arts	Water Management Districts of Florida
National Endowment for the Humanities	Winter Park Health Foundation

7. Do any of your funding agencies require you to manage, store, or share research data in a particular way?

Of the 83 people who responded, 51 (61%) replied that they are required to manage their data while 32 (39%) indicated that they are not.

Answer	Response	%
Yes	51	61%
No	32	39%
Total	83	100%

SECTION 2: DATA COLLECTION

8. What type(s) of data do you generate? Please indicate an approximate percentage.

The 84 respondents indicated that they generate a wide variety of data with the approximate percentage (Average Value) of their data dedicated to that type. The following chart illustrates the range of types of data generated.

Participants who selected *Other* (n=10) were asked to indicate the type of data. The types are not noted in the table below, but include: experimental, focus group transcript data, human performance data, metadata, online survey data files, qualitative, sensor data, simulation data, software program, and survey data.

Answer	Average Value	Responses
Numerical data, e.g. ocean temperatures (%)	62.03	73
Text, e.g. historical records and literature (%)	28.35	48
Still images (%)	24.19	37
Audio files (%)	27.37	19
Video files (%)	22.95	21
Medical data, e.g. patient health information (%)	49.47	17
Biochemical data, e.g. raw and processed "omic" data (%)	18.17	12
Tabulated data (%)	34.79	39
Other (%)	53.58	10

9. What format(s) are your data in? (file extension, etc.) Please list all that apply.

The 75 people who responded have data in a wide variety of formats. The following chart illustrates the range of formats used to identify participants' research data.*

Three other types of data were entered that did not fall into a particular category. Those include: origin, test results, and website.

Type	Response	Annotation
Audio	9	Audio (2), .mpeg, .mp3 (2), .mp4, .wav (2), .wma
Databases	11	Filemaker, Online survey database, .dat (3), .sql (2), .mat (4)
Geographic information data organizers	4	.gis, .lyr, .prj, .shp
Graphics	18	.gif (3), .jpg (7), .png, .tif (7)
Presentation	2	.ppt (2)
Remote sensing	1	LiDAR
Scientific data	1	.fits
Simulation engines	2	.bpp BEopt, .enb
Source code	5	.cpp, .stk, hyperRESEARCH files, HDF5, VTK
Spreadsheets	59	.xls and .xlsx (45), .csv (12), .jnb (2)
Statistical analysis software	32	.dta (3), .jmp, minitab, SAS (8), SPSS (17), STATISTA, statistical files
Text	55	.pdf (8), .doc and .docx (31), .asc (2), .txt (14)
Video	2	.mov, .wmv
Virtual machines(?)	4	.sav (4)

**Note that some files may not be categorized correctly. This was a best guesstimate.*

10. How is your data labeled or annotated? Please check all that apply.

The 84 respondents selected multiple answers, with *Manually, by myself or a member of my research team* being chosen 65 (77%) times. This was followed by *Automatically, through a data collection tool* (n=37, 44%) and *Referentially, with an associated codebook* (n=22, 26%).

Answer	Response	%
Automatically, through a data collection tool	37	44%
Manually, by myself or a member of my research team	65	77%
Referentially, with an associated codebook	22	26%

11. Please estimate the volume of research data for your most data-intensive project of a typical project in your field:

Approximately two-thirds (n=63, 64%) of the 83 respondents indicated that the volume of data produced for a typical data-intensive project was under 50 GB. Another 14% (n=12) selected the 50-100 GB range, and 21% (n=18) exceed 100 GB.

Answer	Response	%
< 1 GB	18	22%
1 - 50 GB	35	42%
50 - 100 GB	12	14%
100 - 500 GB	6	7%
500 GB - 1 TB	5	6%
1 - 50 TB	6	7%
50 - 100 TB	1	1%
> 100 TB	0	0%
Total	83	100%

SECTION 3: DATA STORAGE

12. How do you store your data? Please check all that apply.

The 84 respondents selected multiple answers, with *Personal laptop/desktop* (n=55, 65%), *External Hard drive/CDs/DVDs* (n=52, 62%), and *College or departmental computer network* (n=51, 61%) the most highly selected ways to store research data.

Online solutions, such as Dropbox, Google Docs, and/or Amazon Cloud, generated another 32 responses (38%). Much less popular were *Discipline-specific databases* (n=7, 8%) and *Professional association storage* (n=5, 6%).

Twenty-two (22) respondents selected *Other institutional storage* or *Other* as an option. These responses are not noted in the following table, but include: IST server (2), FSEC (2), locked file drawer (2), other institutions (2), Sharepoint (1), ORC (1), webpage (1), Sandia (1), graduate student computers (1), own TB mini network (1), document management system (1), and networked RAID backup system (1).

Answer		Response	%
Personal laptop/desktop		55	65%
External hard drive/CDs/DVDs		52	62%
Online (e.g. Dropbox/Google Docs/Amazon Cloud)		32	38%
College or departmental computer network		51	61%
Other institutional storage (please note where)		14	17%
Professional organization/association storage (e.g. ICPSR, available with published findings)		5	6%
Discipline-specific databases (eg, National Center for Biotechnology Information / NCBI)		7	8%
Other		8	10%

13. How long do you need your data stored?

Eighty-three (83) respondents answered the question based on three types of data: raw, intermediate/working, and processed/ready for publication. Five options were provided, ranging from *Less than a year* to *Indefinitely*. For each type of data, responses gravitated toward *1-5 years* and *Indefinitely*.

Question	Less than a year		1-5 years		6-10 years		10+ years		Indefinitely	
	n	%	n	%	n	%	n	%	n	%
Raw data	7	8	30	36	14	17	6	7	26	31
Intermediate/Working data	12	15	33	40	14	17	4	5	19	23
Processed data (ready for publication)	2	2	29	35	20	24	6	7	25	31

14. Does your research data contain personally identifiable information (PII), protected health information (PHI/HIPAA), or other types of sensitive information?

Of the 81 people who responded to the question, 60 (74%) indicated that they do not collect sensitive data while 21 (26%) noted that they do.

Answer		Response	%
Yes		21	26%
No		60	74%
Total		81	100%

15. How do you protect your data? Please check all that apply.

The 83 respondents selected multiple answers, with *Data are password protected* (n=55, 66%), *Data are regularly backed up* (n=53, 64%), and *Only certain people can access my data* (n=52, 63%) as the most popular choices. *Data are de-identified* was selected 31 times (37%), followed by *Data are encrypted* (n=12, 14%), *Data are destroyed after use* (n=6, 7%) and *Other* (n=4, 5%).

I do not protect my data was selected five times (6%). Participants who selected *Other* (n=4) were asked to elaborate on their response. These activities are not noted in the following table, but include: project ID's used in filenames, tabulated data, etc, with very limited access to ID key; locked file cabinet (2); and, it depends on the contract.

Answer		Response	%
Data are password protected		55	66%
Data are de-identified		31	37%
Only certain people can access my data		52	63%
Data are regularly backed up		53	64%
Data are encrypted		12	14%
Data are destroyed after use		6	7%
I do not protect my data		5	6%
Other		4	5%

16. Do you take measures to preserve your data? If yes, how?

Of the 80 people who responded to survey, 54 (68%) replied that they take measures to preserve their data while 26 (33%) indicated that they do not. Participants who replied to the affirmative were asked how they preserve their data. Responses follow.

Answer		Response	%
Yes, by...		54	68%
No		26	33%
Total		80	100%

Of the 68% of respondents who replied to the affirmative, most indicated that they preserved their data by backing it up. Generic responses included: making multiple backups (n=11), making multiple copies (n=12), or having multiple copies in various storage locations (n=9). A smaller number of respondents noted where they back up their data; this included on campus servers or networks (STOKES was mentioned once, n=10), external hard drives (n=5), non-specified hard drives (n=2), USB (n=1), CDs (n=1), and hard copies (n=1). Off-site storage was also mentioned, and included off-site backups (n=2), cloud (n=1) or third party agency (n=1).

Migration of file formats was mentioned as a preservation technique only twice. Other responses included file transfer, multiple media formats, raw data, research file with personal identifiers, and version control, which could also be referring to file format preservation techniques. Only one respondent noted an attempt to deposit in a preservation-type facility.

SECTION 4: DATA RECORDING AND ANALYSIS

17. Provide any technical details about the tools that you use or would like to be able to easily use for your work or research. These can be name or vendor of the software product, technical requirements of the software, special accelerators like graphical processor units (GPU), etc.

Thirty-nine (39) respondents listed a variety of technical tools used or needed to perform their research. The responses were loosely categorized into: processing, analysis and writing software or databases; processing, backup and storage network, server or cloud space; and hardware. A summary of responses follows.

Processing, analysis, and writing software and databases	Processing, backup, and storage network, server and cloud space
AMOS	Automated backup internal to UCF system (2)
Ansys/Fluent (2)	Black Armor RAID backup system
ArcGIS/GIS ((2)	Cloud storage/backup (Dropbox and HIPAA-compliant cloudspace specifically mentioned) (4)
AspenTech	DSPACE
CST Microwave Studio	Personal drives
Database with graphical viewing capabilities, basic statistics, filtering, custom output of datasets	Replication
DTreg	STOKES
EndNote	
FACTSAGE	
GPower	Hardware
Gephi	EPSON Workforce Pro GT-550 scanner
Git/GitHub (2)	Tablets
Interactive Data Language	
LimeSurvey	
Lumerical FDTD	
MathCad (Vensim) (2)	
MatLab (5)	
MS Office (2)	
NVivo (3)	
Origin	
RedCap	
REMARK'S OMR software	
R-project programs (4)	
SAS/SAS Enterprise version (6)	
SciFinder Scholar	
SigmaPlot (3)	
SPSS (5)	
SQL	
Stata (2)	
Video performance analysis software	

18. If applicable, how are you recording lab data? Please check all that apply.

The 49 respondents selected multiple answers, with *Excel (or other) files on computers in the lab* the most popular choice with 48 responses (98%). This was followed by *Lab notebooks in paper* (n=29, 59%) and *Electronic lab notebook tool* (n=3, 6%).

If respondents indicated that they used an Electronic lab notebook they were asked to specify which one. The two ELNs identified were Google Docs and Word with embedded images storing NMR and other equipment data in a digital format.

Answer	Response	%
Lab notebooks in paper	29	59%
Excel (or other) files on computers in the lab	48	98%
Electronic lab notebook (ELN) tool. Please specify which one.	3	6%

19. Do you document or record any metadata for your data or dataset?

Of the 62 people who responded, 41 (66%) indicated that they do not add metadata to their datasets while 21 (34%) noted that they do. If respondents replied to the affirmative, they were asked about specific standards or guidelines. Those responses are reported in question 20.

Answer	Response	%
Yes	21	34%
No	41	66%
Total	62	100%

20. If you record metadata for your dataset, do you use any local, agency-specific, or national standards or guidelines?

Twenty-one (21) respondents indicated that they assigned metadata to their data or dataset in question 19. Each of the respondents also answered the follow up question as to the type of standard or guideline applied. Of the responses, 15 (71%) do not use any specific standards or guidelines, five (24%) use identified standards, and one (5%) was not sure.

The five who use standards or guidelines provided the following types: HIPAA/FERPA, FITS standard, program specific, librarians are helping us with this, and all of the above.

Answer	Response	%
Yes (please specify)	5	24%
No	15	71%
I'm not sure	1	5%
Total	21	100%

DATA SHARING

21. Do you share your data?

Of the 82 people who responded to the question, 33 (40%) selected *It depends on the project*, 25 (30%) replied *No*, and 24 (29%) indicated *Yes*.

Answer	Response	%
Yes	24	29%
No	25	30%
It depends on the project	33	40%
Total	82	100%

22. If Yes or *It depends on the project*, do you have a data use agreement (that stipulates the conditions by which someone can access and/or reuse your data)?

Of the 57 people who share or potentially share datasets, 31 (54%) indicated that they do not have a data use agreement and 26 (46%) noted that they do.

Answer	Response	%
Yes	26	46%
No	31	54%
Total	57	100%

23. If you are sharing or planning to share your data, what approach is or will be used? Please check all that apply.

The 57 people who share or plan to share their data selected multiple answers, with *Making them available informally to peers upon request* the most popular (n=40, 70%). This was followed by *Making them available online via a project or institutional website* (n=29, 51%), *Submitting them to a journal to support a publication* (n=25, 44%), and *Depositing them in a discipline-specific data center or repository* (n=18, 32%).

Answer	Response	%
Depositing them in a discipline-specific data center or repository	18	32%
Submitting them to a journal to support a publication	25	44%
Making them available online via a project or institutional website	29	51%
Making them available informally to peers on request	40	70%

24. What restrictions limit your ability to share data? Please check all that apply.

The 78 people who responded to the question selected multiple answers, with *Intellectual property* (n=36, 46%) being the largest barrier to sharing data. This was followed by *Personal*

preference/philosophy (n=23, 29%), *Self-embargo* (n=22, 28%), *Legal* (n=17, 22%), *National security* (n=6, 8%), and *Imposed embargo* (n=2, 3%). *No restrictions limit my ability to share data* was selected 15 times (19%). The option *Other restrictions* was selected five times (6%).

Participants who selected *Other* (n=5) were asked to elaborate on their response. These barriers are not noted in the following table, but include: licensed with Creative Commons, privacy is protected by using passwords for data access, contract requirements, and identifiable information (2).

Answer		Response	%
Intellectual property		36	46%
Legal (e.g. HIPAA)		17	22%
National security		6	8%
Self-embargo (I want a period of first access to my data)		22	28%
Imposed embargo		2	3%
Personal preference/philosophy		23	29%
No restrictions limit my ability to share data		15	19%
Other		5	6%

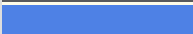

25. In general, with whom are you willing to share your data? Please check all that apply.

The 81 people who responded to the question selected multiple answers, with *Immediate collaborators* garnering the largest response (n=64, 79%). This was followed by *Others in my field* (n=33, 41%), *Others in my department or institute* (n=23, 28%), *Anyone* (n=15, 19%), and *Others outside of my field* (n=9, 11%). Only four people (5%) selected *No one*.

Answer		Response	%
No one		4	5%
Immediate collaborators		64	79%
Others in my department or institute		23	28%
Others in my field		33	41%
Others outside of my field		9	11%
Anyone		15	19%

26. Would your answer be different if mechanisms were in place to make sure that only people you authorize can get access to your data?

Of the 57 people who responded, 34 (60%) indicated that their answer would not be different while 23 (40%) noted that it would be different.

Answer		Response	%
Yes		23	40%
No		34	60%
Total		57	100%

27. If you are sharing your data by depositing data in one or more discipline-specific data repository(ies), please provide the name of the repository.

The 17 people who responded to the question listed 14 unique repositories or locations. A summary of responses follows with the number of times the repository was identified. No number indicates that the repository was mentioned only once.





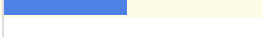

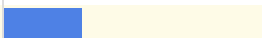

Data repositories being used	
Ameriflux	Other universities' libraries + Library of Congress
ArXiv (4)	NASA Planetary Data system
Campbell	NIH
Cochrane	Online survey site
EDBMS	Open source
Google	SOPHIA
ICPSR (3)	StartTeam

SECTION 5: CONCLUSION

28. What resources outside of your department do you need to best manage and analyze your data? Please check all that apply.

The 74 people who responded to the question selected multiple answers, with *Storage capacity* (n=46, 62%) selected most frequently. This was followed by *Computing expertise or software* (n=35, 47%), *Training on data management* (n=33, 45%), *Data/digital management system for organizing data* (n=25, 34%), *Computing capacity for analysis* (n=23, 31%), *Other external expertise/statistician* (n=22, 30%), *Data management service to outsource some of the work* (n=13, 18%), and *Other* (n=3, 4%).







Participants who selected *Other* (n=3) were asked to elaborate on their response. These other resources are not noted in the following table, but include: simplifying backups (2) and more advanced data management system.

Answer		Response	%
Training on data management (including formulating a data management plan, identifying appropriate data repositories, providing Digital Object Identifiers (DOIs), and/or documenting and describing your data)		33	45%
Storage capacity		46	62%
Data/digital management system for organizing data		25	34%
Computing capacity for analysis		23	31%
Computing expertise or software		35	47%
Data management service to outsource some of the work		13	18%
Other external expertise (e.g. statistician, informatician)		22	30%
Other		3	4%

29. Where do you get assistance now for data concerns? Please check all that apply.

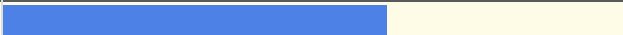


The 81 people who responded to the question selected multiple answers, with *Department or College IT* the leading response (n=46, 57%). *ORC* (n=13, 16%), *IST* (n=7, 9%), and *UCF Libraries* (n=4, 5%) were selected to a much lesser degree. Fifteen (15) respondents indicated that they seek assistance elsewhere and 21 noted that they do not get assistance.

Participants who selected *Other* (n=15) were asked to elaborate on their response. These other areas that were consulted for assistance are not noted in the following table, but include: colleagues (4); colleagues at other institutions (2); computer science experts (2); industry experts; Institute for Simulation & Training (2); economist, methodology consultant, or statistician (4); and University of North Carolina's ODUM Institute.

Answer		Response	%
Dept or College IT		46	57%
ORC		13	16%
IST STOKES computing		7	9%
UCF Libraries		4	5%
Other		15	19%
I do not get assistance		21	26%

30. Are you satisfied with the current level of assistance you receive for data?

Of the 58 people who responded to the question, 35 (61%) noted that they are satisfied with the current level of data assistance offered while 21 (36%) indicated that they are not satisfied. Two people were neutral on the question. Respondents were given the option of elaborating on their answers. Comments were loosely categorized by those who appear satisfied with current services, those appearing neutral, and those appearing dissatisfied. A summary of comments follows.

Answer		Response	%
Yes		35	61%
No		21	36%
Other		2	3%
Total		58	100%

Respondents who appeared satisfied offered the following comments:

- I do not intend to give up the control I have over my data to anyone else, within or outside the university.
- Just don't get access frequently enough due to the busy schedule of the epidemiologist.
- Yes, but there is room for improvement.
- Yes, everyone in department, college and university has always been very helpful.

Respondents who appeared neutral offered the following comments:

- In an absolute sense, the answer is no, but then again I've never expected that a unit within UCF would have the resources to be able to help faculty with such issues, so I've done fine taking care of all this myself. If this changes, that's great, but there are many things at UCF that require more resources so if these don't bubble up to a high priority, I'm not going to be surprised. In any case, I know there are researchers on campus with far more data (in terms of GB); I don't have any projects I can't handle.
- I was not aware that there was assistance with data here at UCF.
- We receive hardware and software support, not data support.
- I get none from UCF. I am very happy with my current external sources.

Respondents who appeared dissatisfied offered the following comments:

- No, limited infrastructure is available for conducting sponsored research that requires handling large data sets.
- Absolutely NOT! <edited to maintain anonymity> I have had to give away one of my UCF websites... due to minimal resources or interest at UCF. Another website... has also been given to the ... library for archiving. <edit> I am now talking with major national organizations to help with one project's management as it is too big for me to handle with the resources I now have available.
- My requirements have been limited with respect to data collection and storage.
- No. Departmental/College staff are focused on efforts to reduce their workload, rather than efforts to reduce faculty/student workload. One example is a current move to eliminate College-level servers/IT-support in favor of UCF centralized support. This creates further barriers by making service less accessible, less person-to-person contact for problem solving, and less accountability. It is a cost- and time-saving measure that will not increase research productivity.
- No, UCF should have a central repository for social science data such as the American Community Survey, U.S. Census, National Incident Based Reporting System, Uniform

Crime Reporting, education data, and electoral data used across the various colleges that are preprocessed for use by faculty, staff and students for research purposes.

- No. Would like to have an institutional mechanism in place.
- Since I'm not getting much help from UCF, I guess no.
- No, the capacity is not available to store the data.
- No!!! I really wish we had an option that would allow organization of patient data that is HIPPA compliant and accessible via cloud.
- No. I would appreciate help with 1) Storage space 2) Backup systems 3) Long term sharing platform 4) Preparing data management plans for proposals 5) Computing power for data processing.
- I just moved to the Bioannex in January 2013 and I have been waiting for many months for additional internet hookups in my lab. I think the level of service needs improvement in this regard.

31. What concerns do you have? and 32. Any additional comments? were combined for space.

Of the 41 people who responded to the question, 17 (42%) noted that they have no concerns and the remaining 24 (58%) indicated a variety of concerns. Responses were summarized into the following categories: general; data analysis support; data management and processing; data curation (storage and preservation); technical; and other. Tabulated comments follow.

Concerns
GENERAL: Training and professional development (3), access to assistance, IRB protection of data, lack of support from UCF, centralized UCF research server with limited access, institutional data repository system & storage space needed, university-provided cloud storage like DropBox
DATA ANALYSIS SUPPORT: Methodological assistance, skill and expertise in data collection, interdisciplinary intramural grants for secondary data analysis, lack of available analytic tools, university-wide licensing of software (3), better software overall
DATA MANAGEMENT AND PROCESSING: Data management and processing (8), large network service (10GB or higher), sharing of data without personal maintenance
DATA CURATION (STORAGE AND PRESERVATION): Storage (2), loss of important data (2), HIPAA-compliant research data storage, long term backup, not sharing all research data (only data sets pertaining to publications)
TECHNICAL: Support for network access issues, support for software/hardware assistance, massive, high-speed scanner for scannable paper surveys
OTHER: Faculty comprised committees