JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Starting the data conversation: informing data services at an academic health sciences library

Kevin B. Read MLIS, MAS; Alisa Surkis, PhD, MLS; Catherine Larson, MS; Aileen McCrillis, MS, MPH, AHIP; Alice Graff; Joey Nicholson, MLIS, MPH; Juanchan Xu, MD, MS

APPENDIX C

Interview results

Basic	Current data services	We use the servers and stuff
Basic	Current data services	Verizon building—computational department, we work with their high performance computing cluster for our genomics data
Basic	Current data services	Method for large file exchange that was introduced 5 years ago—use that
Basic	Current data services	Skirball IT, they provided external hard-drives and have an automated backup service, but when I got here, I was the second PC lab. They were Mac based, so I don't know if it [the PC computers] has been backed up on Skirball IT.
Basic	Current data services	I like the offsite server, that's really nice, I back my laptop up to that, I've been able to pull files off when stuff happens with my computer and I've had no problems, it's been great
Basic	Current data services	"I've been pleased with IT and the Library getting all these journals on line, that's been a major win"
Basic	Current data services	Pretty much everything, everything I can
Basic	Current data services	Uses Biostats Core—is on a number of my papers
Basic	Current data services	Supports half and FTE in her group—that's all grant money
Basic	Current data services	Uses X for all medical statistical analyses
Clinical	Current data services	Uses a data manager—research IT for storage
Clinical	Current data services	Statistician—primary appointment biostats, but she's dedicated to cardiology
Clinical	Current data services	We do have a shared drive—but we realized it's too small, just had a couple of databases and that's it
Clinical	Current data services	Nope—no REDCap or Velos, have been exposed to but they wouldn't use, she doesn't know everything about it but someone in lab used it for collaboration and said it was a nightmare, although that person was not so computer comfortable.
Clinical	Current data services	He has his data managers
Clinical	Current data services	CTSI helped with REDCap—helped with training
Clinical	Current data services	She doesn't use any type of data support system. The data managers do have a system, but she's not aware of what it is.

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Current data services	Internally, Damon set up everything on a server. Nothing's stored on our computers, so he's got an internal server arrangement.
Clinical	Current data services	We work with Rachel Brody and she's very good in terms of the data analysis for the bio repository samples—she's superb, really good at facilitating that
Clinical	Current data services	Running the data, organizing it, and getting it back to us
Clinical/Pop Health	Current data services	Used legal council before for all HMO agreements
Clinical/Pop Health	Current data services	Then moved to research IT so research IT signs as data custodian
Clinical/Pop Health	Current data services	Haven't used any support
Clinical/Pop Health	Current data services	Coordinator created the REDCap—used some support through help desk
Clinical/Pop Health	Current data services	Use shared drive
Clinical/Pop Health	Current data services	"I use REDCap for a survey" (Courtney)
Basic	Data catalog: opinion (-)	Scares me the most is an unguarded repository of data where creepy people can get their hands on the data
Clinical	Data catalog: opinion (-)	I have no idea
Med Ed?	Data catalog: opinion (-)	Researchers and faculty are notorious of being difficult to get to do what we're supposed to do, so what's the incentive—has to be built in or no one's going to do it
Med Ed?	Data catalog: opinion (-)	Difficulty with meta tagging will be dealing with all the different kinds of research
Med Ed?	Data catalog: opinion (-)	Med education and survey and pop health research tough to keep in
Basic	Data catalog: opinion +	NIH would have to decide how much they want to know for the raw data.
Basic	Data catalog: opinion +	Need to figure out how to communicate it—like informed consent needs to be written at a 6th grade level, that's probably what we're going to have to do. Most people don't have any concept about what storage architectures are
Basic	Data catalog: opinion +	Radiation oncology—money is all made in the simulations—that's real computational
Basic	Data catalog: opinion +	Senior physicist—probably want to contact him—responsible for all their computational designs. They would want to catalog their data—I would imagine it would be hugely cost effective if they had a database to draw on for all these
Basic	Data catalog: opinion +	Radiology also has huge data needs, and retrieval is their issue—they've had the money to do their own cataloging—use technical fees

JMLA Journal of the Medical Library Association

EC
Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data catalog: opinion +	I think that would be really useful, because when I'm relying on people in my lab, I don't worry about it, but I've had instances, like this summer, I have students come and go, so they aren't deeply engaged in the project, and they send me the data from the mass spec, so I don't know what to do with the data, and have to ask, what do these columns mean? I don't have a permanent resource to get advice and help in interpreting.
Basic	Data catalog: opinion +	Would want to limit access to her lab
Clinical	Data catalog: opinion +	Think out the workflow, cuz I'm not going to want to spend my time searching for data when someone asks for it—would need mechanism to field these requests, do the extractions
Clinical	Data catalog: opinion +	Also having a way to access these big public datasets—we're all separately paying for access to these datasets, and it's the same exact dataset
Clinical	Data catalog: opinion +	It would be interesting if there were some centralized thing. People collect data regarding smoking stating, cancer studies, cardiovascular studies, even electronic cigarettes, and tobacco research, it would be kinda cool if I could with a button, "I wonder who at NYU has data about prevalence of e-cigarette use?" How many cigarettes a day? Who's using smoking cessation? Mental illness and smoking? So let's say people are studying schizophrenia and they are asking do you smoke? Someone like me I could say, "huh, so it happens that of these patients who have these three comorbidities are being treated"
Clinical	Data catalog: opinion +	If there was some sort of way to pull the data [it would be useful]
Clinical/Pop Health	Data catalog: opinion +	Thinks it's a good idea
Clinical/Pop Health	Data catalog: opinion +	If someone's writing a grant about smoking, all they need from me is 2 numbers, I'm never going to do a smoking study, so I'd be happy to have them use it, I would want to be contacted but I would want it discoverable
Statistics	Data catalog: opinion +	"I think everyone should…it would depend again on how much people are willing to share data. I still notice the tendency of researchers to say, we'll put it on the public website, but we'll put it in some way so that it won't be usable." "Yeah its publicly available! U can go on the site! But everything is hidden somewhere, their contact, what everything means, etc., so a lot of work needs to be done in this regard"
Statistics	Data catalog: opinion +	"I suggest the requirement for sharing comes from the higher ups, that you do research in this institution, you need to share your data"
Basic	Data catalog: opinion +	It's conceivable now that basically u could have links to lots of information that could be related to the publication, make it a lot more valuable—allow people to reproduce things, but also maybe test an idea that they have
Basic	Data catalog: opinion +	If they can see the full distribution, not just a bar [graph or] plot. The theorists want this, they want the numbers. The data are packed into the paper and hyperlinked away.

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data catalog: opinion +	"I think more important than the data itself is the tools—a repository of the surveys, the tools they used, whether it's what they made themselves or is publically available"
Clinical	Data catalog: opinion +	All of work has cross-disease implications. Really rich to have access to data sets that others are using.
Med Ed?	Data catalog: opinion +	Data catalog could help—We could find some colleagues who have questions they want to answer
Med Ed?	Data catalog: opinion +	Generally you want to collaborate with people, not just give them your dataset to do with it what you want
Med Ed?	Data catalog: opinion +	There's one thing to have a list, but to actually make it useful, need mechanism to query it and have a way to connect
Med Ed?	Data catalog: opinion +	Collaboration button "someone is interested in your research!"
Med Ed?	Data catalog: opinion +	How with this data catalog we updated—how often update and get data refreshed?
Med Ed?	Data catalog: opinion +	Have to think about flexibility with categorizing data
Basic	Data collection	The software we use is available on the Internet, but it had a flaw. So my son wrote a shell to fix the software on the Internet
Basic	Data collection	For florescence from a microtiter plate, that's the company that made the instrument using their proprietary software.
Basic	Data collection	Sequence data, we use software that we bought and use in the lab—it may not be the most up-to-date—but there's probably free software out there.
Basic	Data collection	Least squares software that I wrote that was perfect for what we needed to do, that just produced numbers
Basic	Data collection	Collection of the data is done on an instrument with its own software—it comes from the company that sold you the instrument.
Basic	Data collection	Have to compress it somewhere—start to analyze the data—simplify it
Basic	Data collection	20-fold compression—that's what were generally working with—easy to move back and forth from the server
Basic	Data collection	Raw data, particular on the microscopy side, are enormous files, take a lot of processing by custom built software.
Basic	Data collection	Files that have been analyzed—from one image, they acquire 2,000 frames; in each frame, there are about 200 florescent point
Basic	Data collection	He will go into raw data if he has to, if there's a question, something looks weird, he may go in and dig into more detail, but other than that, he usually he doesn't go into the raw data
Basic	Data collection	Occasionally they will crosscheck—have files sent to 2 different people, have 2 different people measure, and then compare—there is some subjectivity, but not a lot
Basic	Data collection	Deal with info at the level of genes and different expression levels and multiple cell types and then a lot of this image data, which will correlate with some of the cells and where are they located

JMLA Journal of the Medical Library Association

EC
Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data collection	We do a lot of imaging, so a lot of what we do is collecting a lot of data and collecting all the related metadata
Basic	Data collection	Dealing with capturing those images, analyzing that data to test models on how cells are communicating
Basic	Data collection	Generate a lot of files from those programs and export to ether various image formats or Excel
Basic	Data collection	Of intermediate analysis involving various programs, and then some numerical data that would be in tabular form—cell locations or molecular locations (from images)
Basic	Data collection	Kinetic component and then the quantitative measurements
Basic	Data collection	Tracking how a given plasmid was constructed
Basic	Data collection	Generate raw file—generate flowjo analysis—then you export to Excel to move it around—then you export it to prism to make your graphs
Basic	Data collection	With programmed quality checks, data are fed in analysis pipeline. About 20% are rejected (due to, for example, it's irreproducible). 80% are accepted into database
Basic	Data collection	Data are mainly used for identification of peptides in the wide range including human, mouse, yeast, e coli, and other organism that people share common interest in research
Basic	Data collection	Then it needs to be converted to a more compatible format to for easy access it when using the cluster
Basic	Data collection	Cleaning process is very much a connected part of the process, do analysis, generate qc read out based on analysis, then might go back and correct some of it and do it again. He's never thought about it as 2 different things, for them it's together
Basic	Data collection	We modify the correlation between activity in one cell between another cell. Maybe that synapse is important in how the two cells work together.
Basic	Data collection	The state of the brain at the time
Basic	Data collection	What was being done to the cell at the time.
Basic	Data collection	(Age of the animal, was anesthesia being used. What time of the day was it.) In Vivo, we play a sound, the frequency of the sound, how long it was (1 mm or so) precisely what happened when.
Basic	Data collection	We change this from one day to the next, might totally be different sounds being played. So if I'm looking at someone's data, I need all that data, if I don't have it, it's useless.
Basic	Data collection	Do super duper hard experiments that take us a long time where we're basically beating our heads against something for 2 years, building the device, testing the device, something very on the edge of what's possible. Once it works, we get something out the device, and then we write the paper, and then we're onto the next
Basic	Data collection	Everything has to be electronic—except lab notebooks, everyone has them written. At some point, I'd love to scan them all in

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data collection	There was a whole workflow—for both cases that were in two different labs, one at NIH and one at the Scripps, analysis was done. The NIH lab, they not only did the analysis, but they continue to develop ways to analyze and come up with new algorithms.
Basic	Data collection	We ran into many complicated problems. With the data, you can't compare numbers across, because the amount of proteins are different and the biological factors are different.
Basic	Data collection	The only way to run it is to use that particular lab's software. This is the lab that is constantly developing new software
Basic	Data collection	Take LIFF files and convert to Photoshop images that we can use
Clinical	Data collection	Have procedure of piloting reviewing, checking the consistency of the data—doing double entry to check the data quality
Clinical	Data collection	Images from MRI are converted to use with analysis program
Clinical	Data collection	(Go through free surfer to calculate volumes and surface areas; could be numbers or parts of images etc.)
Clinical	Data collection	My epidemiologist then moves it into SAS when she uses the data. But it lives in SPSS because I can't use SAS very effectively
Clinical	Data collection	Form data—entered largely in raw fashion, there is some processing that goes on, from primary data we generate other measures. We collect height, weight, etc., collect ffv1. From that we can calculate a bunch of secondary things, like predicted ffv1 from age and height. Or from age height gender and observed ffv1, you can calculate lung age. Or height and weight to BMI. We keep all the processed data.
Clinical	Data collection	In his own clinic, he enters everyone's data in after every clinic—also have some of the same measurements—they're patients so they cant be merged with the research data, but you could search clinical data for case studies and the like—not actually incorporated, but there's this ever burgeoning data
Clinical	Data collection	With the big project we randomize them on discharge and get referral to state quit line or our own counselors.
Clinical	Data collection	Around smoking cessation—of patients seen by our counselor in clinic, we contact them in a month to ask questions such as are they still smoking? Have they been back to Bellevue?
Clinical	Data collection	We're always missing data with surveys. That's a data collection issue—we have people look over the survey before the person leaves and tell them to fill out the missing parts, but even with that, not 100%
Clinical	Data collection	Chart abstraction is a challenge—jump from ICIS to Epic poses difficulties, and then also the hurricane happened and some mothers delivered at Mount Sinai and we're trying to get the data from them—trying to work with medical records, but data retrieved is not as accurate as we would expected due lacking of efficient communication and understanding. Also, same data can be recorded at multiple places in EMR. Completeness of data can be an issue.

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data collection	All chart abstraction is entered into REDCap
Clinical	Data collection	Information is recorded through being present and listening to patient/physician encounter. Access EPIC and add information that is online in EPIC (vitals) and add to paper records. Paper records are later scanned into EPIC.
Clinical	Data collection	Works off baseline clinical data. Confirms things that have been mentioned earlier, grades severity. The sheets are understandable by everyone, even a layperson. She then scans these into EPIC. The system most sponsors or investigative groups use on their end to manage trial data is Medidata Rave
Clinical	Data collection	We're using standard video equipment. Sub-professional, we're using a camera to collect the video now
Clinical	Data collection	Forms—we are handwriting it, and we store it that way and give them a paper copy
Clinical	Data collection	Suzanne, my coordinator, does the bulk of data entry—interface between paper and database—manual entry
Clinical	Data collection	I'm the clinical implementer and I analyze the data—specify what I want and go over the report that they prepare
Clinical	Data collection	We also have a core set of measures to assess. Self-confidence, depression. As new measures come across in the literature it gets added to the trial (e.g., sleep data)
Clinical	Data collection	Typically, she does her own data analysis
Clinical	Data collection	Sometimes, we just look at the images; sometimes, we have to send the images elsewhere to use advanced analysis using software. (I've used a zillion different software packages)
Clinical	Data collection	Usually put it in Excel—the data itself
Clinical	Data collection	Industry-sponsored projects, with their own databases, those have clinical managers, clinical coordinators
Clinical/Pop Health	Data collection	A lot of large secondary data analysis
Clinical/Pop Health	Data collection	She buys data—she deals with data use agreements, and if legal needs to be involved, etc.—it can be very tedious.
Clinical/Pop Health	Data collection	Data entered into REDCap, and then she creates the variables for analysis
Clinical/Pop Health	Data collection	24-hour dietary recalls with the same people over the phone
Clinical/Pop Health	Data collection	For dietary recall data—enter it into their system—"ASA24 is a freely available web-based tool that enables automated self-administered 24-hour recalls."
Clinical/Pop Health	Data collection	Data all goes through different levels of coding and cleaning
Med Ed?	Data collection	Transform data from online programs into Excel and then into SPSS

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Med Ed?	Data collection	Some data collected through the SIM center
Statistics	Data collection	We don't collect data, but rather analyze data that
Statistics	Data collection	There are different philosophies for how data should be preprocessed, so right now we have every image processed in several different ways
Statistics	Data collection	Not every subjects image has the full set of different processing
Statistics	Data collection	Download unprocessed raw imaging data, do preprocessing ourselves
Clinical	Data collection	All homegrown—all created by programmers in house, goes back to 1985
Clinical/Pop Health	Data collection	Some is just taking estimates form the literature and plugging into the model
Basic	Data collection: opinion (-)	Most problems with exchange of data and slowness of computer in analyzing but we just buy new computers for that
Clinical/Pop Health	Data collection: opinion +	Overall the data collection and data entry into REDCap is pretty organized, beyond that its organized chaos as its mostly her manipulating the data
Basic	Data collection tool	TECAN
Basic	Data collection tool	STORM
Basic	Data collection tool	TYPHON
Basic	Data collection tool	Bio-rad
Basic	Data collection tool	Build our own recording devices, we use our own software—don't buy any
Basic	Data collection tool	ImageJ
Basic	Data collection tool	Acquisition software, and first step data analysis are custom
Basic	Data collection tool	Patch-clamp—sold by axon Instruments
Basic	Data collection tool	Prism
Basic	Data collection tool	FlowJo
Basic	Data collection tool	ImageJ
Basic	Data collection tool	OpenSliceCrowd
Basic	Data collection tool	GPMDB is a database of tandem mass spectra and their assigned peptide sequences. The purpose is to aid in the difficult process of validating peptide MS/MS spectra
Basic	Data collection tool	PGx is a tool for proteogenomics mapping.
Basic	Data collection tool	GeneSpring
Clinical	Data collection tool	Enter questionnaire into data base, do some quality checks
Clinical	Data collection tool	RedCap
Clinical	Data collection tool	Done on siemens machine here at med center
Clinical	Data collection tool	Voxpro
Clinical	Data collection tool	Access databases
Clinical	Data collection tool	REDCap
Clinical	Data collection tool	REDCap

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data collection tool	Huge database of looking at the diversity of the microbiome
Clinical	Data collection tool	A SurveyMonkey
Clinical	Data collection tool	EPIC
Clinical	Data collection tool	Randomized trial, we're using Sherlock
Clinical	Data collection tool	iAnnotate, and we are handwriting it, and we store it that way and give them a paper copy.
Clinical	Data collection tool	CRFs into oracle database
Clinical	Data collection tool	Me and Leo are setting up our own database
Clinical	Data collection tool	Qualtrics
Clinical	Data collection tool	Redcap
Clinical	Data collection tool	Velos
Clinical	Data collection tool	Industry sponsored, through their databases
Clinical	Data collection tool	Medidata Rave
Clinical	Data collection tool	Flowjo (one software widely used)
Clinical	Data collection tool	Different kits or assays that are proprietary
Clinical/Pop Health	Data collection tool	Qualtrics
Clinical/Pop Health	Data collection tool	Dedoose
Clinical/Pop Health	Data collection tool	Redcap
Clinical/Pop Health	Data collection tool	Snap survey software
Clinical/Pop Health	Data collection tool	ASA24
Med Ed?	Data collection tool	Qualtrix
Med Ed?	Data collection tool	Redcap
Med Ed?	Data collection tool	SurveyMonkey
Statistics	Data collection tool	Done on the NYUMC server
Basic	Data format	Lab notebooks
Basic	Data format	CSV
Basic	Data format	All the files in their own format—typical for every single lab, if would like to have common platform, will have to figure out
Basic	Data format	ImageJ—turns into Excel type of sheets
Basic	Data format	No field specific standards
Basic	Data format	Lab notebooks
Basic	Data format	Some proprietary formats

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data format	Excel
Basic	Data format	Paper notebooks
Basic	Data format	Excel files
Basic	Data format	Word
Basic	Data format	Every instrument manufacturer has its own raw format. For example, some raw data file can only be read on Windows machines, not on Mac.
Basic	Data format	Excel spreadsheets that consist of three numbers
Basic	Data format	Proprietary formats from the recording amplifiers
Basic	Data format	Simulations in MATLAB (.m) files
Basic	Data format	Lab notebooks
Basic	Data format	Lab notebooks
Basic	Data format	Data is electronically recorded—it can be on their own computer but everything is also on shared drive
Basic	Data format	TIFF files
Basic	Data format	JPEG
Basic	Data format	Excel spreadsheets
Basic	Data format	Word docs
Basic	Data format	Spreadsheets with peptide counts.
Basic	Data format	LIFF files
Basic	Data format	Photoshop images
Clinical	Data format	Analysis and checks are done in SAS
Clinical	Data format	SAS
Clinical	Data format	Stata
Clinical	Data format	SPSS
Clinical	Data format	paper
Clinical	Data format	SPSS
Clinical	Data format	Paper
Clinical	Data format	SPSS
Clinical	Data format	DICOM Standard
Clinical	Data format	An SPSS dataset
Clinical	Data format	Use standard imaging standards, but he's creating new things where there are no standards (he's creating the standards)
Clinical	Data format	Excel spreadsheets
Clinical	Data format	SAS
Clinical	Data format	Paper

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data format	Excel
Clinical	Data format	Paper
Clinical	Data format	Spreadsheet (.csv)
Clinical	Data format	Computer-based forms
Clinical	Data format	Paper
Clinical	Data format	Spreadsheets
Clinical	Data format	SAS
Clinical	Data format	SPSS
Clinical	Data format	Excel
Clinical	Data format	Excel
Clinical/Pop Health	Data format	SAS
Clinical/Pop Health	Data format	Stata
Clinical/Pop Health	Data format	Excel
Clinical/Pop Health	Data format	Feed into an Excel spreadsheet
Clinical/Pop Health	Data format	SPSS
Clinical/Pop Health	Data format	Pencil and paper surveys
Clinical/Pop Health	Data format	Excel
Clinical/Pop Health	Data format	Stata
Clinical/Pop Health	Data format	Dat file
Med Ed?	Data format	Paper
Med Ed?	Data format	SPSS
Med Ed?	Data format	SAS
Med Ed?	Data format	STATA
Med Ed?	Data format	R
Med Ed?	Data format	Excel
Basic	Data management cost allocation	I don't. When I'm writing my NIH grant, I'm using a modular budget so I'm not specifying how…it would be a luxury to specify or ask for such a thing

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data management cost allocation	He mainly has grants where u don't need to specify where u use the money
Basic	Data management cost allocation	No, not allowed to buy computers through R01s at NYU
Basic	Data management cost allocation	Yeah, we would include money for the drives and stuff
Basic	Data management cost allocation	Budget for the storage
Basic	Data management cost allocation	Server—indirect cost—so we wouldn't pay for it so we wouldn't allocate for it
Basic	Data management cost allocation	Not very much, but to a certain extent
Basic	Data management cost allocation	Data storage and computation have been included into grants. But the cost on data management has been very seriously underestimated.
Basic	Data management cost allocation	Costs—mainly on staffing.
Basic	Data management cost allocation	No. Up until you [the library] got in touch with us, I didn't even know it was something you could budget.
Basic	Data management cost allocation	No
Basic	Data management cost allocation	I've started to do that now
Basic	Data management cost allocation	I do think it should be in your grants, and u gotta pay as you go, a lot of people think it should be free
Basic	Data management cost allocation	Avon foundation, 1.25-million-a-year grant sitting there now that hopefully they're going to approve. In that grant is hefty bioinformatics and storage costs—a lot of mutation analyses in there. Storage and analysis—that's where the costs are. Just puts down as storage costs and makes up a number—he has no idea what it should cost—we've sat around the table working this out with Jim, John Speakman, and they don't know what it should cost either
Basic	Data management cost allocation	So far no, because the scale of what we do is pretty small, and we don't do this routinely. I haven't really thought about allocating something.
Basic	Data management cost allocation	"Not specifically, I always do modular budgets"
Clinical	Data management cost allocation	Data management and statistical analysis in grant (we have a statistician too)

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data management cost allocation	Support for statistician and data entry person
Clinical	Data management cost allocation	Allocated funds to programmer on one
Clinical	Data management cost allocation	The other one is barebones, but the department of health is doing some data merging, etc., and he's paying them
Clinical	Data management cost allocation	No
Clinical	Data management cost allocation	Yes
Clinical	Data management cost allocation	Depends on the project what I put in there. An average project would have 5% for the data management for the clinical side, at least 10% of one person for data management of imaging data
Clinical	Data management cost allocation	Full day a week for image processing
Clinical	Data management cost allocation	For most of my grants, I set aside 10% effort for data management, if its bigger, it will go up—who is usually Aloq or his brother.
Clinical	Data management cost allocation	The first couple years of the study when we explicitly had an outside contractor and one was an NYU employee on salary support for a year, there was—we had a budget line item.
Clinical	Data management cost allocation	Challenges with VA—took a year and half to get laptop purchase approved, and then it was approved on next budget year, so had to wait longer—laptops for usability studies, screen capture, etc.
Clinical	Data management cost allocation	Had in my grant a data manager, and then my research coordinator does a lot of course
Clinical	Data management cost allocation	This is billed back to the trial sponsors at a measured rate, i.e., how long does the staff work on this (time focused).
Clinical	Data management cost allocation	Sponsoring groups include ECOG, Southwestern Oncology Group (SWOG), Radiation Therapy Oncology Group (RTOG), Gynecologic Oncology Group (GOG), others.
Clinical	Data management cost allocation	I can't. I wanted to, but the NIH, well the guys at SPA were telling me that it's supposed to come out of indirect funding.
Clinical	Data management cost allocation	Just piloting—and don't think he answered for other, but don't imagine he would have been the one to deal with, but for data coordinating center imagine they are must haves
Clinical	Data management cost allocation	Internal data management team funded in the grant.
Clinical	Data management cost allocation	Data management in the subcontracts in the grants

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data management cost allocation	Always puts in research coordinator in the grants
Clinical	Data management cost allocation	Allocate costs for clinical data managers sometimes
Clinical/Pop Health	Data management cost allocation	She does allocate funds for her programmer
Clinical/Pop Health	Data management cost allocation	She puts server management dollars in there in case anyone were to charge me here
Clinical/Pop Health	Data management cost allocation	No money for data management
Clinical/Pop Health	Data management cost allocation	DM's time
Clinical/Pop Health	Data management cost allocation	We always put a line for data management
Med Ed?	Data management cost allocation	Have some funds in our current grants, but we haven't found the right person with the right skillset to be our data manager
Statistics	Data management cost allocation	She doesn't. "We should be doing that but we usually get those who generate the data to do the management part to a large extent. When they give it to us, there is still management to be done, but it needs to be done by a statistician who will know how it will be used
Basic	Data management responsibility	I'm the only one responsible for all the data, since I'm the one who stays here.
Basic	Data management responsibility	Computer science student, who comes once or twice a week and helps with things, but he'll leave at some point
Basic	Data management responsibility	Everyone does their own thing
Basic	Data management responsibility	Each person is in charge of dealing with their own data (until it comes to statistical analysis)
Basic	Data management responsibility	No, this is done at the individual level
Basic	Data management responsibility	No, for after the process or responsible for long term management
Basic	Data management responsibility	I guess I'm in charge of it
Basic	Data management responsibility	No one specifically is assigned for data management. Everyone is supposed to contribute, responsible for their data

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data management responsibility	Data repository is not managed by MCIT, but by computational lab people (who?).
Basic	Data management responsibility	During the process the person, postdoc, student—they are responsible for storing the data and analyzing it,
Basic	Data management responsibility	When they leave, they turn it over to me.
Basic	Data management responsibility	Individuals have control over their own data. Other people don't usually interact with that data
Basic	Data management responsibility	She's probably responsible for it—long term
Clinical	Data management responsibility	Yelena is responsible for management data throughout
Clinical	Data management responsibility	Only a couple of people have access, and they are responsible for data entry and distributing datasets for analysis
Clinical	Data management responsibility	Person in charge of neuroimaging has her own system
Clinical	Data management responsibility	Statistician uses the data
Clinical	Data management responsibility	A full-time image programmer, part of our group
Clinical	Data management responsibility	On database side, data manager handles that
Clinical	Data management responsibility	Mathematician and full-time image programmer manage the tech side of the images
Clinical	Data management responsibility	Clinical team—nurse responsible for the overall management of all the data from clinical exams
Clinical	Data management responsibility	MRI imaging group is largely responsible for quality control
Clinical	Data management responsibility	Pet imaging—responsible for that data
Clinical	Data management responsibility	I've added them [the radiologists?] to our protocol so they can play with my data too.
Clinical	Data management responsibility	Have other investigators who are processing my data—doing network analyses and the like
Clinical	Data management responsibility	Postdocs, coordinators, volunteers, manually entering data

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data management responsibility	Aloc runs our servers, not actually NYUMCIT—offline
Clinical	Data management responsibility	Everyone on his project has access to all the data, all the time
Clinical	Data management responsibility	We sort of have a data manager—this is one of those things I'm learning—say the hospital smoking cessation study. I am working with PI at Pop Health so he has some project managers working with him. For this project, we've had a few project managers, one of whom has an affinity for data management. Assigning that role has been on and off again.
Clinical	Data management responsibility	Everyone who is working on it has access to the study data. The PI can't be supervising everyone all the time.
Clinical	Data management responsibility	We have dedicated staff for managing data.
Clinical	Data management responsibility	I set up the Wikis.
Clinical	Data management responsibility	He manages the data, which he likes—he knows it, he has control over it, doesn't depend on someone else, doesn't have to wait, can just do it
Clinical	Data management responsibility	Industry-sponsored projects, with their own databases, those have clinical managers, clinical coordinators
Clinical/Pop Health	Data management responsibility	The hands-on data management and manipulation the programmer and statistician manages but she's always answering questions
Clinical/Pop Health	Data management responsibility	Uses a statistician off site who is totally amazing and gets it and knows what questions to ask and is meticulous
Clinical/Pop Health	Data management responsibility	"It's for me and my coordinator"
Clinical/Pop Health	Data management responsibility	Now there's a 3rd person who's starting to do some analysis
Statistics	Data management responsibility	There are imagers, computer scientists, engineers, doing different things.
Statistics	Data management responsibility	Usually one person on one project. Avoid the conflicting versions and stuff by doing that, if its someone else's project, she doesn't touch it
Basic	Data organization	Everyone is responsible for storing their own data. Normally when they leave, they leave behind copies of their data on lab computers, which we are getting backed up. I try to encourage them to store their data in certain formats

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data organization	I get all their lab notebooks, and for their data files, I say, on the lab computer, I want a folder that has all of the experiments that were published, the data for the experiment that were published in a way that I can cross-reference to your lab notebook (with your name and date) because then the lab notebook will be the explanation
Basic	Data organization	There is a little bit standardization—we have a workgroup that has about 30 users—neuroscope—used by many labs, also used within the lab, but not perfect for everything,
Basic	Data organization	Not always documented because it was in the last paper. Bet everything needs to be really well documented
Basic	Data organization	Electrophysiology standards—came from the Stockholm-based group—nobody uses it. Hard to tell if people will use it. U set up a system and if u know it'll work for the next 5 years, you'll use it, but if it's changing year to year might not
Basic	Data organization	Use the shared drives in a disciplined matter—if someone needs to use, will tell other people in the lab, I need to use it next week
Basic	Data organization	Divided and organized by date and image number
Basic	Data organization	Make a folder each day and have different folders inside—one folder that's all the images we have acquired, then a folder of analyzed data
Basic	Data organization	Some stuff is in file name or folder name, but pretty much all in lab notebook. Every acquisition will be correlated with a lab book of what was done for that experiment—regular lab notebook, not electronic "we still write"
Basic	Data organization	When someone moves out of the lab, they have to give all the data files to someone in the lab, and give very clear instructions of how to go into the data
Basic	Data organization	Organized people will have folders of image files with Word doc explaining what it is
Basic	Data organization	Other people might just have folders with dates—again all depends on how careful people are
Basic	Data organization	I've had people come in who want to analyze data from 10 years ago, I've been able to dig out this stuff
Basic	Data organization	Info retrieval can be difficult though, can be clunky, depending on my memory, knowing something was there
Basic	Data organization	Unconsciously probably
Basic	Data organization	The upfront handling, people do what they're most comfortable with their own data. Then when we're publishing think more about it, how we're going to present it, and then different journals sometimes have different requirements
Basic	Data organization	As far as organizing it? It isn't organized" (laughs)
Basic	Data organization	Postdoc, you write it down, and it's illegible. Individual lab notebooks scattered throughout. I've got copies of their files, and analysis files, multiple backups, but everyone is in charge of their own files. It might be the same experiment, but each person might have their own way of writing it down.
Basic	Data organization	No established data workflows

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data organization	If I'm working on a manuscript, all the MATLAB figs, scripts that a relevant for a figure, I will have a folder for each figure, they will put all that information into that folder
Basic	Data organization	No overarching organizational scheme
Basic	Data organization	I will have a folder for each figure, they will put all that information into that folder
Basic	Data organization	If something becomes critical, it goes into the Dropbox—then you'll have all the information related to that paper in a folder
Basic	Data organization	Under each name (of people in lab)
Basic	Data organization	For clinical trials, yes, for lab based, everyone knows I'll wring their neck if data is not on shared drive
Basic	Data organization	"Not really standard procedures, although I should, I realize that" (laughing)
Clinical	Data organization	Data manager, and we are aware about what it takes to have complete, accurate, and well protected data
Clinical	Data organization	So for each type of procedure we set up process
Clinical	Data organization	(Re: data dictionary) For publicly available data—Don't create own, use ones they offer
Clinical	Data organization	(Re: data dictionary) For prospective—he hasn't had a big enough one that it was necessary, its small enough that he kind of knows what everything is and means
Clinical	Data organization	Files all organized—working copy, and on a regular basis copy into permanent master. All in the data folder, everything except imaging, which is a little different
Clinical	Data organization	Variables have names and labels within SPSS
Clinical	Data organization	Relational databases linked to individual subject members—not containing personal identifiers
Clinical	Data organization	Data dictionary, we sort of have one, because you can't possibly have a big dataset without having one. But we don't have one that is set up in the classic sense.
Clinical	Data organization	"I've tried to start putting together different folders for everything, tried to organize it, but there's a lot of data management stuff and you don't get any formal training in data management"
Clinical	Data organization	Have generic system for organizing data. Use this to create paper "cheat sheets" that are trial specific that collect information trial sponsor or investigative group wants to collect.
Clinical	Data organization	It's all organized by grants, my whole infrastructure is stored on the wiki.
Clinical	Data organization	Work together with data coordinating center to design—most recently worked with Cleveland
Clinical	Data organization	Now ongoing prospective studies are more organized out of Michigan
Clinical	Data organization	Truth be told in clinical trials, there is not data to look at because I'm supposed to be blinded and I'm not supposed to look at it (he's not organizing the data, data is held at the data coordinating center)
Clinical	Data organization	There is an effort to try to make sure there are common elements that would allow greater uniformity of sharing—part of it is this notion that nephrologists don't get in the sandbox and play nicely
Clinical	Data organization	Sit down, this is what we want the data to look like—and the data coordinating center exports the data, I don't really have direct access to the data—biannual investigator meeting. I can't even log in
Clinical	Data organization	Yes, have data dictionaries and data cleaning

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data organization	Data cleaning—looking for outlier variables and quality
Clinical	Data organization	Hebrew Home for the Aged and St. Johns. Documented data management.
Clinical	Data organization	Postdocs are put on the IRB, learn about data mgmt., training, meet with statistician (St. Johns), access granted through internal system.
Clinical	Data organization	Data dictionary: Usually have some sort of file describing fields
Clinical	Data organization	Sometimes if I make major changes I keep a copy so I have an audit trail, so I'll have 20 different versions, but it's not really a problem, it's just something I do
Clinical	Data organization	Sometimes can have challenges with who has what version, putting it back together—sharing data—collaborative work
Clinical	Data organization	Folders—each folder has subfolders, I'm fairly well organized—subfolders within subfolders
Clinical	Data organization	Share raw lab data using Dropbox
Clinical/Pop Health	Data organization	Programmer is supposed to annotate all her SAS programs to write what all the codes mean, why do we include this not that, all our decisions
Clinical/Pop Health	Data organization	Had a doc with all the explanations in Word, so have that as a backup
Clinical/Pop Health	Data organization	No standard methodology or process
Clinical/Pop Health	Data organization	We have a data dictionary but it's kind of a short handy one.
Clinical/Pop Health	Data organization	Data dictionary is where the main translation would have to be, or they would have to know how we named the variables
Clinical/Pop Health	Data organization	Then we export it and through the data cleansing process and SPSS, that's where the final data and final variables are, no intermediate step
Clinical/Pop Health	Data organization	Have original, code, and final on the G drive
Clinical/Pop Health	Data organization	We have files created for different iterations as we go through cleaning variables, and then usually have a final output data file, and those can be read into SAS
Clinical/Pop Health	Data organization	For secondary datasets, we have the original as it was reported to us
Med Ed?	Data organization	All the data that came before RoMEO are now in the RoMEO dataset—they've merged into 1
Statistics	Data organization	Put them in one place organized so we can go into one site, ftp, or something to download them
Statistics	Data organization	We have a system here, we organize after we have finished the project (at least to some degree)—organization of the code, keep original data, the script that will change the data, the analysis, the results—each has its own directory (*these are within each project directory, which are within each individual researcher's directory*)
Statistics	Data organization	There will be a read me file that describes what things are, where they are

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Statistics	Data organization	This is a requirement of hers—she goes and checks to make sure the researchers are doing it—you should be able to understand where and what things are if the researcher is not around
Statistics	Data organization	No standard processes—depends on type of data, depends on the question
Statistics	Data organization	Right now it's by investigator and then by project—so still need to rely on memory that oh there was a dataset that might be useful for this statistical method, go into memory, and try thinking of where it would be
Basic	Data organization: opinion (-)	Always my nightmare that someone will say, ten years from now, "I don't think I can reproduce that result, can you validate what you published?" And the person who generated it is gone
Basic	Data organization: opinion (-)	That's an example of the whole problem. You need to have certain standards for something like a lab notebook. It's hard to impose [on print]. Software forces you to follow a format.
Basic	Data organization: opinion (-)	Human images—certain journals in order publish u have to upload your data—but it's a total mess because it was organized form the top down
Basic	Data organization: opinion (-)	People want someone else to manage their data
Basic	Data organization: opinion (-)	Academia: All the knowledge stays with the postdoc or student—won't stay in my head long won't stay in their head long
Basic	Data organization: opinion (-)	When someone leaves the lab the level of organization of the data and their level of conscientiousness of how they labeled and organized their data for the next person—not always up to par
Basic	Data organization: opinion (-)	Computer files, don't know what file names are, then you have the handwritten lab notebook, bad handwriting
Basic	Data organization: opinion (-)	Cataloging items that are stored, their lineages, histories, etc., can be really challenging, probably something we're not that good at
Basic	Data organization: opinion (-)	If don't have good understanding, have to pretty much throw that away
Basic	Data organization: opinion (-)	The challenge in organizing data is the size of the data. They are just too huge.
Basic	Data organization: opinion (-)	The messiness of the data, like any data sets in any other research area.
Basic	Data organization: opinion (-)	The proprietary software that runs the machinery names everything by date. So if two people do different experiments on the same day, the files could have the same name.
Basic	Data organization: opinion (-)	A weirdness in the medical center that I've never experienced before—health care is the major focus—it has to be—but because of that, it's been hard to operate—purchasing is awkward, human resources I awkward—shipping and receiving has gotten much better, accounts payable is terrible—all these things u kind of rely on as an institution
Basic	Data organization: opinion (-)	I'd love to say I would stop and catalog—but probably wouldn't—to carefully construct all the data from the project, it's probably already easy enough to find

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data organization: opinion (-)	I don't think it's the kind of stuff that most people would really want—to have clear and really good read me docs for every script etc.—I think it would be a really big headache—if someone asks, then I can do it, I'll do it for the 4 people who really care, but to have to do it all the time for everything, I could spend my time better elsewhere
Basic	Data organization: opinion (-)	They filtered the data in their own way, and u query that data, all u can pull out is their filtered data and they got it wrong. What they thought was a mutation wasn't a mutation
Basic	Data organization: opinion (-)	Gene expression data is cataloged elsewhere—but I don't know how to find most of it anywhere, not catalogued in a way, unless u remember where the data was
Basic	Data organization: opinion (-)	Yes, that is the most challenging part—would be looking at the data. What is the most fair way to present the data? What is the most fair way to represent the data that ends up in the final collection of data.
Basic	Data organization: opinion (-)	"At the time that I started my lab, we weren't really collecting this data, it started gradually, so I didn't really set up procedures"
Clinical	Data organization: opinion (-)	Pulling in data, concerned about quality—problematic to do quality control of existing data
Clinical	Data organization: opinion (-)	Retrospective studies of EMR data—biggest data quality problem is just being confident in the data—makes it a bit scary
Clinical	Data organization: opinion (-)	Data quality problems could be a combo of putting data in wrong in the EMR or pulling data wrong
Clinical	Data organization: opinion (-)	That's the problem, every new project I have to figure out a new workflow, including who's going to pull the data, who's going to do the next steps.
Clinical	Data organization: opinion (-)	It's a big problem, we don't have an army of people, we only have one statistician, that's it—could definitely use a workflow where one team member is a database programmer who would pull the data
Clinical	Data organization: opinion (-)	But SAS and SPSS handle longitudinal datasets differently—one uses columns, one uses rows. You have to write code to flip it to deal with it. So need to be able to create a dataset that you can use it in whatever form you need it to.
Clinical	Data organization: opinion (-)	Challenges in organizing and keeping it together
Clinical	Data organization: opinion (-)	The data is not that easy for searchability, if you wanted to do data query, you can't really do that without doing select if and ___and ____, and then SPSS gets mad at you sometimes and won't give you anything and you have to go in and figure out what if and or is messing it up.
Clinical	Data organization: opinion (-)	Another issue is missing data. There are a lot of ways to deal with missing data, drop the person, or do imputation. Right now, it's just missing, we haven't imputed anything. "I was writing a grant for New York, I had magna give me a dataset. Trying to run regression models. SPSS kept saying it couldn't run the models because it said it only had 6 people in it (because there were missing fields)"
Clinical	Data organization: opinion (-)	It was a nightmare to set up the database for the hospitalized smokers. 3 failed subcontracts to set up the databases—it eventually came together but it did delay the work. We started out on paper and are catching up on the backlog now.

JMLA Journal of the Medical Library Association

EC
Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data organization: opinion (-)	You don't think about it until it's too late. Our data manager is in the process of doing it but I don't have a great handle on it right now. Having a data dictionary organized by topic would help
Clinical	Data organization: opinion (-)	It's tricky and messy because it's part of U01 so it's a consortium. So we have to set up a de-identified set to share and multi-site analyses and we're behind on that, with the trouble setting up the databases, the hurricane, another part is that we don't [have a dedicated manager] (because the data manager is also doing 6 other things) so it's not like I can say to the consortium, here's the name phone of the person, contact her.
Clinical/Pop Health	Data organization: opinion (-)	Keeping it clean and training programmers—it's very complex data and no one can drop in and know it
Clinical/Pop Health	Data organization: opinion (-)	Someone is starting to work with me to do the analysis and that's the main challenge, explaining the variables—it's not complicated, just translation
Statistics	Data organization: opinion (-)	"Still trying to figure out who has which data!"
Clinical	Data organization: opinion (-)	It's very hard for me to get data out of Velos—I often can't get what I need someone has to go in and pull it for me
Clinical	Data organization: opinion (-)	There hasn't been a really good handoff system. Some data were collected a long while ago but have not yet been processed due to lack of resources
Clinical	Data organization: opinion (-)	Some data were stored at a "secure" place, but now I don't have access to it so that I can just analyze myself
Clinical	Data organization: opinion (-)	What data mgmt. support is missing
Clinical	Data organization: opinion (-)	Lack of statistical support. Statistician comes in once a week. But could use him more.
Clinical/Pop Health	Data organization: opinion (-)	Strength with this research is all the different types of data, so making sure these can all line up and speak together is a real challenge—that's the number one challenge for me
Clinical/Pop Health	Data organization: opinion (-)	Can be challenge over in who's working on the project, and it's not always documented
Clinical/Pop Health	Data organization: opinion (-)	"We've been screwed, it's taken us 2 weeks to redo things, it drives me crazy"
Med Ed?	Data organization: opinion (-)	No data infrastructure—disparate datasets sit on hard drives, we have to compile when we want to use the data for research
Statistics	Data organization: opinion (-)	It probably should be done, but she hasn't figured out the best way to do it. It's really knowledge of the analysis that is most important—data management is part of data analysis at this point
Basic	Data organization: opinion +	It would be nice if was really easy for everyone to access everyone's data. 3 or 4 years ago I said—anytime someone joins the lab, I sit down with them and say how I want them to keep their lab notebook.
Basic	Data organization: opinion +	They've realized they need to have a common platform.

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data organization: opinion +	I think the library is well positioned to help with that. But any overarching scheme usually scares the hell out of me because the medical center will do what's right for the medical center and I need to do what's right for me
Basic	Data organization: opinion +	Research people need to play by their own sets of rules
Clinical	Data organization: opinion +	I collaborate with people in the school, they say it's too difficult to describe data. To me its fundamental to have meaningful straightforward data
Clinical	Data organization: opinion +	Billing data—it's as clean as it gets
Clinical	Data organization: opinion +	Not a big deal to keep it data organized. Naturally good at organization, and there is good communication across research team. This helps avoid and solve problems.
Clinical	Data organization: opinion +	He's pretty comfortable that he can just do what he needs to do in Excel
Clinical	Data organization: opinion +	I find that if I'm not organized, I have way too many balls in the air, so I am meticulous about keeping it organized on my computer
Clinical/Pop Health	Data organization: opinion +	I've learned a lot about how important data management is just from having to do it myself, so I've been talking to someone about how to do it, how to do it well, and how to write it into a grant
Statistics	Data organization: opinion +	This (rigorous) system has developed from my long and painful career of trying to pull out old data
Clinical	Data organization: opinion +	Data quality as service? With a good data dictionary and standardized procedures, almost anyone could help with data quality/cleaning.
Med Ed?	Data organization: opinion +	What we feel a huge need for is tagging our variables more consistently to help when we combine data. "I really love REDCap for that"
Basic	Data preservation	I would like to be able to access the data indefinitely. File formats can change, and that is a worry that I can't access it because the format changed.
Basic	Data preservation	I would say it's not unusual that I might go back 20 years ago. So it's not impossible that I would go back. I would rather be able to read it forever.
Basic	Data preservation	We have at least 2 copies of everything, sometimes 3
Basic	Data preservation	Just stay on those hard drives and servers, although he has put a few datasets on the CRCNS site
Basic	Data preservation	NIH requires for 10 years—that's another interesting thing, I came from university brought a truck load of papers, because we needed to keep the data but no one ever looked at it. No one really checks
Basic	Data preservation	Most of the super computers are useful for big bulk work. But in terms of data mining, every lab, in every field of science, it's still the personal computer that wins
Basic	Data preservation	Raw data does have a back up in microscopy core for at least some
Basic	Data preservation	Once you're done with data, you keep it on that hard drive
Basic	Data preservation	I definitely don't want to get rid of data if it's not published
Basic	Data preservation	Once it is published, try to keep the data for as long as he thinks people will be interested
Basic	Data preservation	Keep data for at least 5 years after its been published. But he isn't sure he always does it

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data preservation	Basically images in little database, proprietary file tells u where everything is. If u lost that file, or it got corrupted, if u really wanted to go in to get it, u could always get the images out
Basic	Data preservation	Would save at each stage—so data amplifies throughout the project
Basic	Data preservation	Often make 2 copies, I have at least one copy, and then the individual who collected the data will have their own copy. members of the lab will take a copy when they go on
Basic	Data preservation	Take them with them in case they need to revise a paper or just want a copy of it
Basic	Data preservation	Ideally my entire career (laughs) and then going forward, other peoples' careers are building here. So decades would be ideal really
Basic	Data preservation	No responsibility for data after the process
Basic	Data preservation	Ideally, data should not be thrown away
Basic	Data preservation	Usually 5 years, he would think.
Basic	Data preservation	Forever
Basic	Data preservation	Everyone in the lab had to prove that they had a daily back up plan—and Rachel would hound them until they did that, so I feel like we're at least taken care of
Basic	Data preservation	Infinite
Basic	Data preservation	Too huge to host forever
Basic	Data preservation	When we all used to use jump drives and then we had Jaz drives, when I heard it was going to be phased out, I loaded it all to hard drives, CDs, DVDs—that's how we had to do it for years.
Basic	Data preservation	Not a problem anymore, we all use the same programs now, there's been a consolidation, it wasn't the case before
Basic	Data preservation	Now at least everything is backed up, and I can find it everyone in the world
Basic	Data preservation	Backs everything up on a spinning disk hard drive, doesn't worry about it because everything is backed up elsewhere too, but he might replace it with a flash drive so he doesn't have to worry about breaking it, but flash is a little slower
Basic	Data preservation	All his data from past 10 years, 500 GBs, on a hard drive that he takes with him everywhere, backs everything up to it everyday
Basic	Data preservation	If the data's bad, not for long, but it can be useful forever
Basic	Data preservation	Probably ten years [data should be available] for data that didn't make it to publication. For data that did get published, there's a central database that everyone could put it in , [that would be good]
Basic	Data preservation	Count on the MCIT computers and servers being backed up—haven't had issues yet, haven't lost any data we can't recover
Basic	Data preservation	I haven't deliberately gotten rid of any, but it's growing exponentially, so it's getting more difficult to keep everything
Basic	Data preservation	Want to keep the data around because there are new methods created to get more info out of data.
Clinical	Data preservation	I would die if I came in and one day was my data was gone—don't back up as much as I should, I guess I trust that the shared drive is ok.

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data preservation	Mostly implicit in the process—save dataset, data dictionary, code that was used to generate results. All is organized on the server
Clinical	Data preservation	He has a copy on his external hard drive—that's the back up
Clinical	Data preservation	Forever
Clinical	Data preservation	External back up systems that are linked to them
Clinical	Data preservation	We have an integrated network that's off the NYU hub of all the comps that are involved
Clinical	Data preservation	Publication=end point—becomes a final view—all input into that publication becomes frozen
Clinical	Data preservation	Forever—I'm doing autopsies now and doing brain scans of people from 30 years ago
Clinical	Data preservation	There will always have to be someone to manage it—the default is him, but if he has money the default quickly shifts to someone that he's paying to do it.
Clinical	Data preservation	I don't know—whatever NIH wants.
Clinical	Data preservation	I haven't been backing up stuff that's on the VA server, I probably should
Clinical	Data preservation	She backs it up every week on an external hard drive
Clinical	Data preservation	Keep it forever
Clinical	Data preservation	The paper trail for clinical trials is stored a defined number of years both in original paper and in EPIC.
Clinical	Data preservation	She keeps her original documents and they are also scanned into EPIC. That's her backup.
Clinical	Data preservation	They must hold onto data until the "close out visit" with the trial sponsors. This can be many years after the trial is done. After that, the data is then under the control of NYU(?) or sponsors?
Clinical	Data preservation	Personally she feels print data should be held for a minimum of 7–10 years after a patient's death.
Clinical	Data preservation	Now that more data is electronic, it may be useful to keep data forever.
Clinical	Data preservation	Most people are signing a release form saying that I can keep it forever
Clinical	Data preservation	If it's de-identified it should be permanent. For NIH, I suppose, but for my own data, it's not as important.
Clinical	Data preservation	6 months after done—moved to NI DDK—data repository
Clinical	Data preservation	Everything—whole Cleveland database gets dumped, part of the public knowledge (NIDDK)
Clinical	Data preservation	That'll be the same for Neptune, that'll be the same for care GN, now I think every R01—at the completion of the trial mandatory to share
Clinical	Data preservation	In perpetuity
Clinical	Data preservation	After research, we still keep the data available to postdocs for secondary data analysis. Paper ideas abound from the datasets.
Clinical	Data preservation	Usually keep anonymized data—supposed to keep published for years and years
Clinical	Data preservation	But if data doesn't need to be stored, not supposed to store it. Patient identity is stripped as soon as possible
Clinical	Data preservation	Just keeps the data on the virtual drives

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data preservation	You're supposed to keep it a really long time. You can't hold on to identified though. Or my research doesn't warrant it
Clinical	Data preservation	I would be dead if all this disappeared
Clinical	Data preservation	It still sits on my drive,
Clinical	Data preservation	Infinitely, as long as I'm here
Clinical/Pop Health	Data preservation	Backed up in Philly and NJ now
Clinical/Pop Health	Data preservation	HMO data—they had a secure file transfer system so she doesn't have originals—that stuff is really not hers to share, a lot of data
Clinical/Pop Health	Data preservation	After, it just sits on their server and collects dust
Clinical/Pop Health	Data preservation	"Forever" [Laughs]
Clinical/Pop Health	Data preservation	Backed up on an external hard drive—all de-identified data so there's not a big risk in keeping it in multiple places
Clinical/Pop Health	Data preservation	"Right now everything's here—right after Sandy made copies in different places but now everything here"
Clinical/Pop Health	Data preservation	"Forever, this is important [laughing], isn't that what everyone says?"
Clinical/Pop Health	Data preservation	"I'm pretty fanatic about not saving anything on desktops, everything goes on the shared drive. The only exception is the IRI dataset"
Clinical/Pop Health	Data preservation	One small one is just living on our shared drive somewhere. Haven't thought about what I'd want to do with it when done
Clinical/Pop Health	Data preservation	"I mean forever, I would never want to get rid of it"
Med Ed?	Data preservation	We do individual backups too, portable hard drives.
Med Ed?	Data preservation	Has copies in Dropbox and box, but it's not a centralized process
Med Ed?	Data preservation	"And there's always your sent folder when you're really desperate"
Med Ed?	Data preservation	Registries—never finished—building new cohort, working on connecting all datasets
Med Ed?	Data preservation	We're trying to emulate Framingham, so we're on the long path
Med Ed?	Data preservation	Want to keep as much data for as long as possible
Med Ed?	Data preservation	No goal of dumping educational data warehouse data
Statistics	Data preservation	Data is stored in these directories "forever"
Statistics	Data preservation	Data be available forever—and usable for that long as well
Basic	Data preservation	I think 10 years was a made up number, 5 years is probably more realistic, after that technology changes

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data preservation	Keep it in the same place right now, probably would be good to be able to move it somewhere else
Basic	Data preservation: opinion (-)	We had long discussions, had plans, there were some good people, they left, and then they started to charge us excessively I would say
Basic	Data preservation: opinion (-)	The reality is software programs change, all these things change, so in reality, you don't generally go back more than 5 years
Basic	Data preservation: opinion (-)	They haven't been able to set it up on the server here at NYU, mainly due to the lack of enough resource from IT. (Hospital takes higher priority than research.)
Basic	Data preservation: opinion (-)	No procedures. I'm sure if this was a clinical dept. that would be different. It's not human data, so basic labs are little fiefdoms.
Basic	Data preservation: opinion (-)	I don't think it's happened that I couldn't get it, but the barrier was so high, it wasn't worth my time.
Basic	Data preservation: opinion (-)	I'm about to use up my shared drive space soon
Basic	Data preservation: opinion (-)	Evgeny Nudler— they built their own servers, expensive really high quality hard drive—it was stolen—it had a year's worth of work on it—crystal structures, not backed up
Basic	Data preservation: opinion (-)	When asked "Why don't u use storage available here," said, "we like to do our own thing"
Basic	Data preservation: opinion (-)	There is no readme file or data dictionary included with it. Usually, the work is done and gone through publication. So they basically store all their data and then gave their files to me, and so if I have any questions, I look at the spreadsheets
Basic	Data preservation: opinion (-)	To go back to the raw data, I'd have to ask my grad student.
Basic	Data preservation: opinion (-)	Big problem is when people leave the lab and the data is not organized in a way that she knows what's going on, what everything means.
Basic	Data preservation: opinion (-)	Some people leave read me files about where everything is saved (there is only one person she can think of who did it very well over 17 years); most don't leave anything
Basic	Data preservation: opinion (-)	They run out of time and say, "Here are all my DVDS"
Basic	Data preservation: opinion (-)	Keeping data for a long time—it can be a problem, when I came here I brought data on optical discs that you can't read anymore
Clinical	Data preservation: opinion (-)	Have specific needs that clinical trials don't have—hard for our data to fit in their database
Clinical/Pop Health	Data preservation: opinion (-)	She has some data that she's supposed to destroy when research is complete (as part of the data use agreement) but she doesn't want to do it, she thinks it's a shame
Clinical/Pop Health	Data preservation: opinion (-)	I would be happy to have anyone else manage it at any step of the way (laughing) but I would want to have some control over who uses it and how it's used
Clinical/Pop Health	Data preservation: opinion (-)	"We had bad experience with Sandy that some people lost data on the servers so we back it up elsewhere"
Basic	Data preservation: opinion (-)	But possible to lose metadata—so it depends how good your notebook entries are
Basic	Data preservation: opinion (-)	Haven't thrown anything out but moving all this stuff around is a pain
Basic	Data preservation: opinion (-)	One of the programs we used a lot, a lab program, isn't around anymore. I don't think I have any computer that could open those files directly

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data preservation: opinion (-)	We've had to go in with a little help from them [people who left the lab] and it was ok, but it was hard—looking at every version which one had the most info
Basic	Data preservation: opinion (-)	Would see figure in PowerPoint presentation, but then had to find data behind the PowerPoint
Basic	Data preservation: opinion (-)	And I usually found there were multiple copies on multiple drives, sometimes slightly different versions
Basic	Data preservation: opinion (-)	Info retrieval can be difficult though, can be clunky. It's dependent on my memory, knowing something was there
Basic	Data preservation: opinion (-)	Often when a trainee leaves, u lose the info, even if they give u a hard drive, u don't know where it is
Clinical	Data preservation: opinion (-)	Can't even plug in an external hard drive into the VA system
Clinical	Data preservation: opinion (-)	Data loss do occur, not often though, during the transfer for some unknown reason
Clinical	Data preservation: opinion (-)	Right now she has a program that backs it up continuously, but they don't want to pay for that ("we have trial version, but I think we're not gonna buy it, think we're gonna do it manually")
Clinical	Data preservation: opinion (-)	On Dropbox ("I know we're not supposed to use Dropbox")
Clinical	Data preservation: opinion (-)	Exporting the data from REDCap, Qualtrics. Translation from survey to Excel. Issue for data cleaning & quality.
Basic	Data preservation: opinion +	There is a lot data on the server that may not be needed any more. But people always think they might come back to re-work on those data someday even though nobody really does.
Clinical/Pop Health	Data preservation: opinion +	I would be happy to have anyone else manage it at any step of the way [laughing], but I would want to have some control over who uses it and how it's used
Basic	Data preservation: opinion +	Want consistent parameters between images that are supposed to be compared together
Basic	Data preservation: opinion +	Micro array data is really managed by this consortium, so u can access the raw data and the processed data—that's a very accessible interface. That's also a very recent thing. There's probably a lot of microarray data that's nowhere near as standardized as this
Basic	Data preservation: opinion +	One of the strengths of that consortium is they standardized everything
Basic	Data preservation: opinion +	There may have been 20 different groups collecting data, but they were all using the same standards
Basic	Data reuse	As though we're going back and saying we can reanalyze so and so's data, but most of our experiments are directed towards a particular answer
Basic	Data reuse	We might compare and see if the two results are compatible with each other.
Basic	Data reuse	Almost every time when you're done, you have to go back to the original experiment
Basic	Data reuse	Sometimes try, sometimes difficult—have to find the data—contact the person or figure out the formats—depends on how computer savvy someone is and how much time person is willing to spend with the data,
Basic	Data reuse	Don't remember, but it may happen—if they had better data
Basic	Data reuse	Normally, no. Perhaps the reason is, particularly the measurements that come from animal cells, the control animals change—try to stay in the animal colony

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data reuse	I'd say 10 years is probably the longest that we've gone (in using data)
Basic	Data reuse	Yes, go back and reuse old data
Basic	Data reuse	Many of his projects are connected. Data re-use do occur.
Basic	Data reuse	Oh sure. As much as you can. But you have to say when you're doing it. For example: the first few papers that I put out, changes in the electrical patterns between two cells. There's the core of the paper, then a broader look in another paper. Some of these data are hard to come by and as much as we can get out of it we do. Might be the same neuron but looking at different aspects of what the experiment yields.
Basic	Data reuse	Do you have difficulty using old data, determining what exactly was going on? All the time. A lot of this time we're running in place.
Basic	Data reuse	We do revisit old data—we're getting a paper together now where we'll have to revisit an old dataset—we're seeing something very different so we're having to compare with old dataset
Basic	Data reuse	Yes. For example: Some research started 10 years ago, but he's still mining the data and publishing on it
Basic	Data reuse	I just got a paper provisionally accepted that the research was done 6 years ago—resurrected this data, filled in the blanks that remained, and writing 2 papers on it—been sitting there, no one else had thought about doing this, so it's still ok
Basic	Data reuse	Yes we've reused our data a lot, using different materials, or different cell lines/stem cells
Basic	Data reuse	Not really
Clinical	Data reuse	It depends, most of the time u use it for some other sub-study or analysis
Clinical	Data reuse	He reuses a tremendous amount of data (his own, other peoples, other peoples in new ways)
Clinical	Data reuse	All studies are considered extensions of each other, so they will go back and look at an older dataset, so nothing is put away for good.
Clinical	Data reuse	Have never used another person's data, doesn't know that they never would but never have.
Clinical	Data reuse	Sure, oh yeah, all the time—longitudinal, knowledge is progressive, often go back to reexamine things
Clinical	Data reuse	Image—u have thousands of ways to interrogate an image
Clinical	Data reuse	If we're using the same subjects we identify that we're doing that
Clinical	Data reuse	We'd go back. One grant I had, we're doing pet MRI. One thing we found is for both our researchers and radiologists, we needed to do a cognitive battery to confirm their cognitive status (needed the gold standard)
Clinical	Data reuse	Yes, it just keeps growing!
Clinical	Data reuse	n/a [she did not talk about past projects—all projects are currently active]
Clinical	Data reuse	Yes.
Clinical	Data reuse	The study of the impact of primary care resident physician training on patient weight loss was conducted a few years back, and now we have just finished a similar study with extended scope.

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data reuse	She is aware of situations where data (both biological samples and accompanying data) has been reused. Reexamination of blood samples from colon cancer patients led to discoveries about mutations associate with the KRAS protein allows patients to be screened in advance of certain therapies.
Clinical	Data reuse	I'll go back to data I've collected and analyze it for results. I might stratify it
Clinical	Data reuse	[Don't think so]
Clinical	Data reuse	"Don't generally use other peoples data mainly due to my limitations"
Clinical	Data reuse	Data sets lifespan is long. Almost 12 years and still using the same sets.
Clinical	Data reuse	He has on occasion
Clinical	Data reuse	I'll go back and use things, may take figure out of something I wrote 5 years ago to use as a background figure in a current paper
Clinical/Pop Health	Data reuse	When I reuse my own data that I've already bought, I have to write a new proposal, I can't just willy nilly do what I want with the data
Clinical/Pop Health	Data reuse	Have reused SAS programs as well
Clinical/Pop Health	Data reuse	Used some open source SAS programs from the web ("here's the SAS program for how to make a comorbidity index," for example)
Clinical/Pop Health	Data reuse	Nothing is ever really done, everything's a work in progress—so I keep protocols open for a while
Clinical/Pop Health	Data reuse	In the future certainly, I haven't done it just because I haven't had time
Clinical/Pop Health	Data reuse	Whole bunch of external data from city and state, some we buy from commercial companies
Med Ed?	Data reuse	Yes absolutely, I think that's a core idea of the registry—this is data that u could answer a bunch of questions are—so one of our missions is to collaborate with other researchers
Statistics	Data reuse	Yes, there is sometimes more uses for the data—it's always a possibility, and frequently it happens. The one thing I have not gotten around to having a system for and work out is actually cataloging all of those already cataloged things (*in the researchers personal directories*).
Basic	Data sharing	Most often happens if we were collaborating. It happens reasonably often. Usually when collaborating, people don't need to see the raw data. The data gets processed into numbers, and usually they are satisfied to see the final numbers and images, and they are not asking to get down in the weeds and see all those data and reanalyze them. There's more trust in that.
Basic	Data sharing	We've had 14 papers published by people we've never heard of—using our data
Basic	Data sharing	If somebody asked for our datasets, the most efficient way is to buy hard drives and send the hard drives—downloading through the Internet would be very, very long. Nobody knows any other solutions that would be faster that I know about

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data sharing	Mail hard drives
Basic	Data sharing	Only share raw data with people on projects
Basic	Data sharing	Outside the group, would only share information (preliminary results)
Basic	Data sharing	The files are analyzed, the results are there, there's a certain level of trust in the field that I don't need to go to the raw data, I believe that you did it right
Basic	Data sharing	The only time I have seen raw data being reviewed is during investigation of scientific fraud
Basic	Data sharing	Outside the group, I have never seen that, never inside my group or in anyone else's (if it's not part of a collaborative project)
Basic	Data sharing	I've had people come in who want to analyze data from 10 years ago, I've been able to dig out this stuff
Basic	Data sharing	No (other's people data)
Basic	Data sharing	Already have an obligation to upload the genome sequence data. I think that's what the data share goal is more, large sets that you could look at in another way
Basic	Data sharing	We definitely do share in collaborative projects—project we do in collaboration with a group in Switzerland
Basic	Data sharing	Share by have common access to a common server—share the platform using the same passcode
Basic	Data sharing	We occasionally look at publicly available data, rare, but probably going to get more common
Basic	Data sharing	Researcher from any organization can go into database and see what has been observed for certain proteins
Basic	Data sharing	Data is open source, so is the data repository.
Basic	Data sharing	Every project that he's involved in now, both the data and the software are open source. They encourage people in the field with publication to shall the data. They search literatures for data in interest and they contact the author for data sharing.
Basic	Data sharing	Yes. They have been using open source data, as mentioned before.
Basic	Data sharing	Yes, by email. These Excel files. There have been times when people wanted the raw files. So I've mailed a DVD. 95% it's pretty tight [the organization and structure of the data] and analyzed.
Basic	Data sharing	Want to know only medical center people could have access to the metadata
Basic	Data sharing	Yeah, all the time
Basic	Data sharing	FDA—"What they want is pharmaceuticals to actually post failed clinical trials—so we could data mine enormously important data"
Basic	Data sharing	Drug that was shown to not work. "All I know is the paper that was published," it would be great to get access to the data

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data sharing	She (a collaborator) gave us a list of data (not the raw data) but then we compare that to the previous data, so we've done comparisons that way. I just got an email from the foundation that supported the project—CHDI—they are in the process of trying to figure out the network of proteins that's related to this disease, so they asked me for this data so that they compare what they've done and put it in their network.
Basic	Data sharing	I think my student and postdoc probably went through other people's data to compare with our data—that would be important to do for the publication
Basic	Data sharing	But someone who wanted to do a rigorous comparison between our data and another group's data that did something similar. So her interest was to do a rigorous comparison to see what's in common and what's not
Basic	Data sharing	GEO for gene expression data, not really anywhere else
Clinical	Data sharing	Collaborations with other cohorts like ours, a lot of our work up to know, we have collaborated with a Swedish and an Italian cohort
Clinical	Data sharing	Website of the cohort consortium. Share description of data, don't share actual data
Clinical	Data sharing	IRB data: they have done that many times—sometimes it's anonymized data, sometimes it's de-identified, it's specific to each project
Clinical	Data sharing	Once the primary aim is done, and since NIH wants it, I think sharing is ok.
Clinical	Data sharing	People have requested theirs, and she provided de-identified dataset, which was easy for her to do—easy to find, de-identify, share, was transparent.
Clinical	Data sharing	Yes sure. Typically its done through NIH-related mechanisms, data are de-identified, haven't put in repository yet but we've talked about it
Clinical	Data sharing	Typically what happens, someone writes an email, I have this project I'd like to do it with u, do u have these materials, this is our objective—it's not a blind sign up and just take it, we use our judgment
Clinical	Data sharing	Some of our investigators here are discussing having limited datasets available on a website
Clinical	Data sharing	Each of the grants that we have at NIH has some sort of data sharing agreement
Clinical	Data sharing	Use other people data all the time—use FTP sites
Clinical	Data sharing	Got internal funding to add LPs to a subset—going to call the people who already participated in stuff and try to get LPs from them, until the money runs out.
Clinical	Data sharing	Wants to really do it in a big way (informationist project—want to create anonymized data set) (I would like to take a subset of that data for those people who have a complete dataset, and anonymize it and create a publicly accessed database that you would need a password to enter—want to only have people who have complete data because other stuff won't be particularly valuable to the dataset—replica of his dataset
Clinical	Data sharing	"Not too much, I guess in theory I could if we wanted to do a multi-site kind of thing."
Clinical	Data sharing	No, not yet [related response in question 11]

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data sharing	Most people ask me for the questionnaire, not many ask for the actual data, would have to get IRB and stuff
Clinical	Data sharing	Did share data with PhD student
Clinical	Data sharing	I haven't used a lot of other peoples data
Clinical	Data sharing	She has never personally shared data
Clinical	Data sharing	Physicians certainly share information verbally about their clinical experiences with different diseases and patient types.
Clinical	Data sharing	Yeah, it was for a paper, this guy did a journal club for our papers, and I sent the raw de-identified data.
Clinical	Data sharing	There is an effort to try to make sure there are common elements that would allow greater uniformity of sharing—part of it is this notion that nephrologists don't get in the sandbox and play nicely
Clinical	Data sharing	MEMO—team of investigators here and other institutions. Used a shared data set.
Clinical	Data sharing	New center for stroke disparity. Common core.
Clinical	Data sharing	Creating registry. To keep measure across studies. This is in development that the field is working on as a whole
Clinical	Data sharing	Will check with fellow PIs to share data quality/data mgmt. procedures and training manuals.
Clinical	Data sharing	Usually just he has access to data, collaborators via email
Clinical	Data sharing	On rare occasion people have asked for raw data—it's usually not people distant, people I know
Clinical	Data sharing	People have asked me for grant drafts, for help with their own grants
Clinical	Data sharing	People have asked for methods sections, I've sent them those
Clinical	Data sharing	Might use data from a collaborator, not from some random person I don't know
Clinical/Pop Health	Data sharing	Don't share right now
Clinical/Pop Health	Data sharing	It's her data, she's not averse to sharing it, there's a lot that can be looked at that I'm not looking at, but I don't have any current plans to share
Clinical/Pop Health	Data sharing	Would have to get the IRB to approve sharing, but then it would probably be just giving them a file, it's a small enough dataset that we could do that
Clinical/Pop Health	Data sharing	There's a similar investigator doing a similar study, she's using that dataset. She contacted him, he just sent the file to her
Clinical/Pop Health	Data sharing	She's done other research using large *institutional* datasets
Clinical/Pop Health	Data sharing	Get an email from a student who wants to use it for x project
Clinical/Pop Health	Data sharing	Share via email or upload onto something NYU had down at Washington, files 2.0 or something

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical/Pop Health	Data sharing	More of the sharing we do is of instruments and data collection tools
Clinical/Pop Health	Data sharing	Most recent R01—using Admin data from DOE
Med Ed?	Data sharing	Challenge has always been how to describe the data for other people who might be interested in it
Med Ed?	Data sharing	Sample size is generally small so would like to eventually collaborate with other institutions
Med Ed?	Data sharing	Explicitly want to make this data available
Med Ed?	Data sharing	There are rules with the registry
Statistics	Data sharing	We don't generate data, we analyze data, it belongs to the person who generated it—they need permission to share. It's not too common that it's shared, often people say no, I don't want you to use my data
Basic	Data sharing: opinion (-)	Would never consider using someone else's data
Basic	Data sharing: opinion (-)	Often someone will ask the data—and so I say yes, I have the data, do u have someone to come get it? And usually it stops there
Basic	Data sharing: opinion (-)	The thing with imaging is some parts of it are so specialized in terms of experimental design, it may be much less frequent that people will want to dig into dataset
Basic	Data sharing: opinion (-)	Biggest issue may be how good were the antibodies, how reliable were the cells—[may not just trust other people's data]
Basic	Data sharing: opinion (-)	Ethical issue—genomic data can be used for people identifying purpose, so there are restrictions there.
Basic	Data sharing: opinion (-)	The bigger problem is not sharing data. It is where to put the shared data. There was a place called Tranche. With the funding dried out, it slowly died.
Basic	Data sharing: opinion (-)	Just this year (2014) people at Duke used it, published really fast in some crappy place, and then I was hurt by that
Basic	Data sharing: opinion (-)	Copied my experiment from what they saw in a poster, they had some engineers reengineer our device, and then beat us to the punch
Basic	Data sharing: opinion (-)	"Who does it benefit publishing after it's published? People can read the paper if they care"
Basic	Data sharing: opinion (-)	If people could get their hands on it, they didn't have to do all that work, they can just get the data and profit from that, so why am I motivated to do that, put in all the hard work and then someone else will just profit from it
Basic	Data sharing: opinion (-)	May not want people to have direct access to that data, I want people to know it exists so they know to contact me if they're interested
Basic	Data sharing: opinion (-)	The spreadsheet format is, the raw data can be depending on the algorithm used, I'm guessing it can generate different—be interpreted differently—so the spreadsheet might be different depending on what you use, so maybe people need the raw data, but most people can't use it because they don't have the software.

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data sharing: opinion (-)	They'd have to negotiate with the lab where we interpreted the data. It's not proprietary but they [the lab] did create it, so strictly speaking, for the raw data, they'd have to go to the material that came out of the machine. I suspect people don't want that, so we provide spreadsheets and I think we did provide it in the supplementary material [of the article].
Basic	Data sharing: opinion (-)	Raw data—a lot of it doesn't look very pretty, so you don't really want to put it out there
Clinical	Data sharing: opinion (-)	It's challenging outside the institution—issues of who's using it, data quality issues
Clinical	Data sharing: opinion (-)	We asked the VA for our set of 700 veterans in our study, this is in our consent form. We asked, "Can you give us utilization, inpatient, outpatient over a year after enrollment in the study?" So far all we've gotten is a very long text file.
Clinical	Data sharing: opinion (-)	There may be legality/ownership issues with sharing data as it is "owned" by the sponsor.
Clinical	Data sharing: opinion (-)	Qualitative data. I would hope the NIH wouldn't mandate that
Clinical	Data sharing: opinion (-)	Here's the thing, you'll have data overload. How many people, if my data were available, someone could run a meta-analysis. I see a lot of utility to it. For this to work though, people need standardized instruments like NIH Toolbox. If people make their own tools it's not useful. I just don't think all data should be available forever unless it's linked and useful.
Clinical	Data sharing: opinion (-)	Double-edged sword of the shared dataset—High quality statistician at employ but could not get access to the raw data
Clinical	Data sharing: opinion (-)	But no, I won't share my data. My data is my proprietary data. I will share it with my collaborators, absolutely would not put it up somewhere
Clinical/Pop Health	Data sharing: opinion (-)	Medicare data, there's no sharing protocol, u have to go and ask for it yourself.
Basic	Data sharing: opinion +	A lot of data the published is under-analyzed, only outsiders can realize that but it's true
Basic	Data sharing: opinion +	Also quality checks. Now it's almost impossible to communicate through figures in science—need images etc. U can go beyond that and provide compressed but organized datasets that u need to look into to decide if I'm right or wrong
Basic	Data sharing: opinion +	People want control of their data—but there are benefits to sharing—get cited, like if you have a popular paper
Basic	Data sharing: opinion +	It's a typical excuse that only the people who made the data can understand it, yes, true with cutting-edge science, but different levels, certain level can be used by others
Basic	Data sharing: opinion +	Everybody can share data—u have to document it—this is what people are afraid of, including us—like I know this animal was a little sick—that's something that I just know when I look at the data, but this is also documentable
Basic	Data sharing: opinion +	Ur scared someone analyzes your dataset and come to a different conclusion. But that will happen. And that's also free advertisement—people will reanalyze because it's controversial, and then add new data

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data sharing: opinion +	Proteomic data so far does not have this issue since it is not patient identifying. In the future if the proteomic data becomes better and finer,
Basic	Data sharing: opinion +	I understand the more u share, the more u have, it only makes fiscal and community sense
Basic	Data sharing: opinion +	Would be great to share if there was a movie on how to use this device or a doc file describing, but I would want some sort of check in—like people would say I'd like this info, and then I could give them a password, and then they could get it
Basic	Data sharing: opinion +	I'm all in on sharing, we have empty desks in the lab where we will train people for free on our technology, we don't wanna be on their papers, they can just come in a learn what they want to learn
Basic	Data sharing: opinion +	I want strict regulations to make sure there isn't misconduct.
Basic	Data sharing: opinion +	So do I feel threatened that people are using my data? Not really, we see so many different things, and make sense of heterogeneity.
Basic	Data sharing: opinion +	I could see how published data might be useful. Like if we publish one image, but we have others related, that might be useful. But data that hasn't been published at all, I'm not sure if it'll be useful, and also u don't know if you'll ever come back to it, might want to use that data in the future
Clinical	Data sharing: opinion +	One major issue is women are volunteers and we're supposed to protect their identity—we are part of the CI cancer consortium—she's the chair of the secretariat—we meet once a year—this issue came up (making data public)—and it's a big concern for a lot of cohorts—need to guarantee the confidentiality of the data
Clinical	Data sharing: opinion +	With more sharing, data dictionaries will become more important
Clinical	Data sharing: opinion +	There are a number of large datasets that intrigue me, I was surprised a couple years ago, student came to me, said they had access to insurance dataset, wanted to look at mental health status and economic status—100s of thousands of records with all this financial data, wealth every year
Clinical	Data sharing: opinion +	I wouldn't see in a problem sharing my data, I'm pretty junior, usually it's a fellow or a medical student who wants to do a project, and she advises the project, then publish together—it's a win win
Clinical	Data sharing: opinion +	These two projects aren't about their health, so I'm in an area that is more lenient. Re: sharing
Clinical	Data sharing: opinion +	For my own data, it doesn't matter to me, I can see where some people are opening themselves up to if they made any mistakes, but if I made a mistake someone should call me out on it.
Clinical/Pop Health	Data sharing: opinion +	Nobody has ever asked
Clinical/Pop Health	Data sharing: opinion +	If it's federally funded, it seems like the right thing to do to share
Statistics	Data sharing: opinion +	They've spent a lot of time generating their data, they don't want other people publishing on it
Statistics	Data sharing: opinion +	But now the thinking of young researchers has changed, and people think data should be shared and there's all these places to share it

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data size	I would guess at this point, we could take all of our data and we would be talking tens of 100s of gigabytes
Basic	Data size	10 datasets all together
Basic	Data size	30 images, each images is 1GB—generate 30GB in one day. For 1 paper=500 images and each image is 1.5GB ~ half a terabyte in one project
Basic	Data size	Raw is 1GB, and then analyzed is 2GB
Basic	Data size	Images form the microscopy core—those are about 100MB per file, about 10–20 per project
Basic	Data size	Patch-clamp data somewhere around the same (100MB per file)
Basic	Data size	They estimate probably 10 terabytes overall per year
Basic	Data size	Each experiment might have several thousand, 10s of thousands of images
Basic	Data size	10s of GBs per week in terms of just image data for people who are making a lot of 3D movies of lymph nodes or tissues
Basic	Data size	Few terabytes a year
Basic	Data size	Mass spectrometry—2–3GB
Basic	Data size	In a project, analyze 95 tumors, 15 mass spec per tumor=very large datasets
Basic	Data size	DNA and RNA sequencing—even larger
Basic	Data size	The size of the spreadsheets is small: in kilobytes
Basic	Data size	1–10 terabytes per person per year. Safe estimate of total stored now: 100 terabytes
Basic	Data size	16 terabytes
Basic	Data size	Probably in the gigabytes range
Basic	Data size	I don't think it was that much (TB)
Clinical	Data size	Health study—close to 100GB data (their 100GB drive is almost full)
Clinical	Data size	MRI study—issue of space. Worked with IT—they provided us with a site or something that we could upload it
Clinical	Data size	Just HCUP is close to 100GB
Clinical	Data size	Run 1–2 subjects/wk, SPSS datasets are small
Clinical	Data size	One folder w/300 gig and only 110 have been used so far
Clinical	Data size	Our data is probably in the 1–2 terabyte range
Clinical	Data size	But the actual size of files I have no idea
Clinical	Data size	Currently have 277 unique individuals, going to have 500.
Clinical	Data size	Audio and video of screen capture are probably the biggest files (focus group is 1 hour 25 mins long /each)
Clinical	Data size	25 interviews with providers and staff are 45 mins each
Clinical	Data size	Numbers are up in the 1,000s so nothing too big
Clinical	Data size	Eventually, we're going to have 500 gigabytes of videos
Clinical	Data size	75 videos over 5 years…maybe 200 videos

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data size	1 study 140, another 130, so databases are not huge
Clinical	Data size	Some are very small, up to a few MB
Clinical/Pop Health	Data size	Size, doesn't think she has the terabyte of data
Clinical/Pop Health	Data size	75,000 patients and every single Medicare claim for a year for them, so not small—and that's just one dataset
Clinical/Pop Health	Data size	About 400 individuals
Clinical/Pop Health	Data size	Ballpark on data size
Clinical/Pop Health	Data size	10,000 consumers and receipts—soda project is 8,000 so probable 12–15,000 over the last 5 years
Clinical/Pop Health	Data size	IRI data—in neighborhood of 1–2 terabytes
Med Ed?	Data size	Most of our datasets aren't so big, we can just sent them electronically
Statistics	Data size	Let's say maybe 2,000 variables for ultimately 400 subjects, right now we have 200 and something
Basic	Data storage	Large data (shared data): stored and processed in the data repository mentioned before.
Basic	Data storage	The lab notebooks get stored in the lab. In the high shelves [because they are not often accessed?].
Basic	Data storage	Leave the data on the lab computers.
Basic	Data storage	We have our cluster that we created, which was very good because we weren't affected by the flooding
Basic	Data storage	We have various servers, and we have various hard drives in people's drawers
Basic	Data storage	Put them in external hard drives—don't use cloud or anything like that
Basic	Data storage	Have a couple hard drives already not backed up that haven't been published
Basic	Data storage	Patch-clamp—everything is contained within the computers that are generating it and on an external hard drive to make room for acquisition
Basic	Data storage	Isn't on a central server because we don't have enough space to hold everything at one time. So we'll use the MCIT servers for what we're currently working on, and then when done, move data elsewhere
Basic	Data storage	Kept on the CDs, DVDs, hard drives
Basic	Data storage	Stored on the super cluster
Basic	Data storage	Genomics data—stored on the cluster
Basic	Data storage	Various Macintosh hard drives
Basic	Data storage	Portable back up (external hard drive) that they pass around occasionally
Basic	Data storage	Use Dropbox—"Dropbox is better than the server because you have that backup"
Basic	Data storage	Use Google drive and Dropbox, although he knows you're not supposed to

JMLA Journal of the Medical Library Association

EC
Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data storage	Google drive to share with himself and Dropbox to share with other people in the lab
Basic	Data storage	Long-term management of the data? We throw it on external backup.
Basic	Data storage	Everyone has their own PC, laptop and external drive. [It's not clear to him if those PCs are backed up since Skirball IT is mostly Mac and he's PC]
Basic	Data storage	2 big servers in our lab
Basic	Data storage	Raw data is on big server in lab
Basic	Data storage	Everyone has their own hard drive and that's backed up on the server there
Basic	Data storage	Dropbox
Basic	Data storage	Given 300GB as a lab on our remote server
Basic	Data storage	Mirror server—2 in different rooms, continuously backing up to each other, not in another location (on 6th floor, would have to melt the polar icecaps pretty bad to get up here)
Basic	Data storage	Raw data is on big server in lab
Basic	Data storage	Going to host it locally
Basic	Data storage	Shared drive
Basic	Data storage	Raw data is at Scripps
Basic	Data storage	Jeff has all the raw data.
Basic	Data storage	It may be somewhere on our server, but that data wasn't a great set of data so we haven't gone back to look at it. It's on our lab computer, but backed up using the R or D drive.
Basic	Data storage	Everybody keeps their own images on their own computer or on the server, or some of it is on CDs and DVDs
Basic	Data storage	When we're putting together a paper or her team wants to show something to her, they'll put it on the lab server. Otherwise, usually people in her lab cant access other peoples data.
Basic	Data storage	On the MCIT servers
Basic	Data storage	On CDs and DVDs
Basic	Data storage	GEO
Clinical	Data storage	We have a very extensive database that Yelena has organized that hold epidemiology data, tracked specimen use, and so on
Clinical	Data storage	In-house database for NYU health study—complex database—keep all sorts of data there
Clinical	Data storage	Data in REDCap—talking to research IT to see if they have some software suitable for migration to their system
Clinical	Data storage	Don't store anything on our own computers here, we share it on research IT
Clinical	Data storage	Shared drive has 100GBs now, and we only have a couple left, not enough at this point
Clinical	Data storage	RedCap
Clinical	Data storage	EDC that's maintained by Duke, or it's a commercial thing, they might just have a license, but it's at least housed at Duke.
Clinical	Data storage	CTMS database for some of the site information

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data storage	Prospective studies on MCIT sanctioned and supported servers
Clinical	Data storage	Recently bought an external hard drive for statistician
Clinical	Data storage	Data is on shared drive P drive, automatically backed up, only certain people have access. Also have hard copy backup paper records, stored in a cabinet in the lab
Clinical	Data storage	Have 3 servers
Clinical	Data storage	2 out of the 5 (servers) are NYU, 3 are not because MCIT we've had issues with quality control and stability—it's resulted in losses, delays, misconceptions, so we ended up maintaining our own systems
Clinical	Data storage	Right now, the data sits largely resides on our servers.
Clinical	Data storage	Imaging data is on both our servers and also on the radiology servers.
Clinical	Data storage	The data is on a shared drive with his team, under his folder, and there's the data
Clinical	Data storage	Because it's the VA and they are particular about their data, it all lives on the VA drives so we operate by VPN from here.
Clinical	Data storage	Paper survey data is kept in a locked filed cabinet in her office
Clinical	Data storage	The data on VA-related project has to be stored on the VA secure server
Clinical	Data storage	Anything that's not HIPAA compliance issue is on Dropbox
Clinical	Data storage	Medical educational research data are in EduData Warehose
Clinical	Data storage	EPIC—Medidata Rave (see above)
Clinical	Data storage	Stored remotely. Because I have NIH grants, I was able to set up our own terabytes of NYU storage—through research IT servers
Clinical	Data storage	Use PBWorks wiki
Clinical	Data storage	All biospecimens are also stored by NI DDK, they contract out where it's actually stored
Clinical	Data storage	NYU shared drive to keep the data sets.
Clinical	Data storage	Center for Health Behavior Change—shared drive
Clinical	Data storage	He doesn't store the images. Images are on the PAC system
Clinical	Data storage	Stores on hospital supported options
Clinical	Data storage	Encrypted iron key provided by MCIT
Clinical	Data storage	On the H drive
Clinical	Data storage	On R drive
Clinical	Data storage	Data in Velos,
Clinical	Data storage	Raw data in lab
Clinical	Data storage	Industry sponsored studies sit somewhere else, on their own databases
Clinical	Data storage	I have a lot of it on my shared drive (on the MCIT Servers)
Clinical	Data storage	I also have a cloud (I use Dropbox—My husband got drop box so then I got it too)
Clinical/Pop Health	Data storage	Research IT now has it

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical/Pop Health	Data storage	Hard drives or DVDs—for most of her past data
Clinical/Pop Health	Data storage	Keep the raw data in REDCap
Clinical/Pop Health	Data storage	All data is stored on MCIT server
Clinical/Pop Health	Data storage	These external hard drives are stored in locked drawer in office
Clinical/Pop Health	Data storage	IRI is on external hard drive—with 2 back ups
Clinical/Pop Health	Data storage	Everything else on shared drive
Clinical/Pop Health	Data storage	The NY BMI data—That data lives with some collaborators down at the square
Med Ed?	Data storage	Trying to use Redcap more consistently to store our datasets—love that we can create the data dictionaries, behind the firewall and all that
Med Ed?	Data storage	Have it all on multiple hard drives
Med Ed?	Data storage	Have MCIT shared drive, try to make sure everything makes it to the shared drive
Med Ed?	Data storage	Keep raw data in REDCap
Statistics	Data storage	Store the data on our personal computers
Statistics	Data storage	Data in different places—one in Pittsburgh, one at Stonybrook, another is at Columbia
Statistics	Data storage	Then analyze there because those data objects are quite large, too large for computer, and then results are stored there too (so raw, analysis, and storage on server)
Statistics	Data storage	Directory for junk—every so often they empty “junk” cuz no one knows what it is, very messy
Statistics	Data storage	Maybe 70% of the time the data is small enough to fit on our laptops and that’s what we do (store the data on our personal computers)
Statistics	Data storage	We use the server—we work with MCIT directly to get access to that server. It’s not personally her, she doesn’t lately have time to do data analysis
Basic	Data storage: opinion (-)	Not enough space [laughed]—can’t even transfer files through it
Basic	Data storage: opinion (-)	Don’t use MCIT server at all
Basic	Data storage: opinion (-)	“A little bit awkward to work strictly from the server”
Basic	Data storage: opinion (-)	“Not the cloud storage generation yet”
Basic	Data storage: opinion (-)	Concern about having the power go out in New Jersey someplace and not being able to access the data
Basic	Data storage: opinion (-)	One of the main reason they don’t use server is they don’t trust the network, they want to have it locally

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data storage: opinion (-)	Problem with Dropbox is the confidentiality
Basic	Data storage: opinion (-)	Need more storing space in accommodating more data. The current data repository is outgrowing its size.
Basic	Data storage: opinion (-)	Know researchers have found the 300GB kinda constrictive
Clinical	Data storage: opinion (-)	Problem is getting access to the HPC—running a simple stat on his datasets with 6 million patients, crashes normal computers
Clinical	Data storage: opinion (-)	HCC data—We have this database but he doesn't even know where its sitting (it's all de-identified)
Clinical	Data storage: opinion (-)	Haven't been able to get enough space
Clinical	Data storage: opinion (-)	"Because they don't give us enough space to actually do anything. They give you about the size of that laptop."
Clinical	Data storage: opinion (-)	The VPN is not ideal. If we had it on an NYU server—I'm not a tech person, so there is probably something you can do to put it in cloud, but there are definitely days when the VPN is slow or the login isn't working and it's an issue.
Clinical	Data storage: opinion (-)	Can one come up with the ideal equal access for all but secure? The VA has its own issues about data. Bellevue is very particular about—can you download software on to your computer? Oh, this computer lets you do it, but this one doesn't. Halfway through our study, Bellevue upgraded our computers and suddenly nothing works.
Basic	Data storage: opinion (-)	Real secure server storage is more expensive that these external things [hard drives]. Buying a terabyte server would be 100 dollars a week vs. just 100 dollars total for the external hard drive
Clinical	Data storage: opinion (-)	3 are not because MCIT we've had issues with quality control and stability—it's resulted in losses, delays, misconceptions, so we ended up maintaining our own systems
Clinical	Data storage: opinion (-)	Have to request access—and can get access to tools like SAS, etc., because VA has license
Clinical	Data storage: opinion (-)	Hard for VA to install software
Clinical	Data storage: opinion (-)	Takes an hour, hour and a half to upload each transcript to NVivo
Clinical	Data storage: opinion (-)	Some files stored on local driver are backed up every 10 minutes. But files cannot be edited at local server because the required software is not available here.
Clinical/Pop Health	Data storage: opinion (-)	Price was absurd to order more storage from MCIT
Basic	Data storage: opinion +	"I think cloud would be a good solution. Having data on external hard drives, we would need a second hard drive to back up the first hard drive. Sometimes we back it up on the second hard drive, but they will both be stored in the same physical location"
Basic	Data storage: opinion +	They try to avoid commercial software as much as possible. In the field, there's so much open source that they can use, usually they can avoid being licensed.
Basic	Data storage: opinion +	"If it's free, I'd like to have a hundred terabytes"
Clinical	Data storage: opinion +	It would be nice if we had our own, it's secure, but I do all my storage
Basic	Data tyoe	Visual data

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data type	Gel electrophoresis
Basic	Data type	Quantitative images of the gel
Basic	Data type	Quantified the amount of information intensity on the gel
Basic	Data type	Measurements from florescence or luminescence
Basic	Data type	Monitoring a reaction rate (as a function of time) or making a single measurement.
Basic	Data type	Quantitative PCR
Basic	Data type	A series of numbers after every cycle that we then reduce to a single number
Basic	Data type	Analyze them to get ratios; normalize the numbers; to see how mutations affect the function
Basic	Data type	Rate measurement and have lots of data points to fit to a line to get a reaction rate.
Basic	Data type	Collections of strains and plasmids and sequences that we keep
Basic	Data type	Image information
Basic	Data type	250, 500 channels, sometimes 1 hour sometimes 5 hours records
Basic	Data type	Video information
Basic	Data type	Put fluorescent markers on proteins and find where they are in the cell
Basic	Data type	Electrophysiology—measure electrical currents in the cell
Basic	Data type	Super resolution microscopy
Basic	Data type	Scanning ion-conductance microscopy
Basic	Data type	Electron microscopy
Basic	Data type	Flouro and electro—get molecular and structural
Basic	Data type	Gene expression
Basic	Data type	Micro arrays
Basic	Data type	RNA sequencing
Basic	Data type	3-dimensional image files
Basic	Data type	Data on which mouse came from which parents, and how long ago—there's a data tracking that goes along there
Basic	Data type	Catalog cytokine levels
Basic	Data type	Flow plots from harvested animals that tell you about cell types found
Basic	Data type	Microscopy data
Basic	Data type	Genomics data
Basic	Data type	Identification of proteins and characterization of their post-translational modifications
Basic	Data type	Quantitation of proteins and peptides.
Basic	Data type	Modeling proteomics experimental design
Basic	Data type	Biomarker discovery and verification
Basic	Data type	Mass spectrometry
Basic	Data type	Electrical recordings

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Data type	Behavioral experiments
Basic	Data type	Electrophysiological data
Basic	Data type	EPSP and spike train traces
Basic	Data type	Movies
Basic	Data type	Manuscript
Basic	Data type	Matlab figures
Basic	Data type	Scripts
Basic	Data type	Host locally the tumor genome atlas
Basic	Data type	Gels
Basic	Data type	In vitro assays
Basic	Data type	Test tubes
Basic	Data type	Protein interactions
Basic	Data type	Staining protein measurements
Basic	Data type	Gene transcription
Basic	Data type	Proteomics
Basic	Data type	Mass spectrometry
Basic	Data type	Imaging
Basic	Data type	Microscopy
Basic	Data type	RNA sequencing
Basic	Data type	Western blots
Basic	Data type	Reverse transcription polymerase chain reaction
Clinical	Data type	Questionnaires
Clinical	Data type	MRI data
Clinical	Data type	Radiology data
Clinical	Data type	Baseline data (demographic data, data on history, dietary data, data on main cancer incidence, cardiovascular, data on medication use)
Clinical	Data type	Patient data
Clinical	Data type	Uses national, international data from previous trials—gets data from networking
Clinical	Data type	NHANES
Clinical	Data type	HCUP
Clinical	Data type	HHC data
Clinical	Data type	State registries
Clinical	Data type	MRI studies
Clinical	Data type	Blood draws
Clinical	Data type	Cognitive data

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data type	Intravenous glucose tests
Clinical	Data type	Wide range of types of clinical data
Clinical	Data type	Physical
Clinical	Data type	Psych
Clinical	Data type	Neuro
Clinical	Data type	Medical exams
Clinical	Data type	Clinical labs
Clinical	Data type	Archive of spinal fluid and blood drawn longitudinally on hundreds of people
Clinical	Data type	CT scans
Clinical	Data type	MRI scans
Clinical	Data type	PET scans
Clinical	Data type	Blood diagnostics
Clinical	Data type	Radioligands for pet imaging
Clinical	Data type	Inflammation and tao imaging
Clinical	Data type	Numbers (age, test performances)
Clinical	Data type	Raw data (And our imaging data all sits as mostly raw)
Clinical	Data type	Processed data
Clinical	Data type	Cross-sectional dataset, eventually it will probably be a bit more longitudinal (quasi longitudinal)
Clinical	Data type	NY state—fall prevention demonstration project—get core measurement, but then also getting whole host of measurements that have to do with falls, that aren't in other projects
Clinical	Data type	Patient questionnaires
Clinical	Data type	Chart review data
Clinical	Data type	Saliva cotinine analysis
Clinical	Data type	Hospitalization data
Clinical	Data type	Exit interviews on 160 obese patients
Clinical	Data type	Survey data
Clinical	Data type	Chart abstraction
Clinical	Data type	100s of pages of transcripts, from 25 patients
Clinical	Data type	Staff surveys
Clinical	Data type	Biochemistry/biological analysis
Clinical	Data type	Bio samples, blood stool, and vaginal and rectal swabs
Clinical	Data type	Blood stool
Clinical	Data type	Vaginal and rectal swabs
Clinical	Data type	Pre-trial medical history
Clinical	Data type	Every prescription and non-prescription drug the patient is taking

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Data type	Create a baseline profile of their health that can used to measure against as patient participates in trial
Clinical	Data type	Adverse events
Clinical	Data type	Lab results
Clinical	Data type	Tumor data
Clinical	Data type	Randomized trial data
Clinical	Data type	Qualitative data
Clinical	Data type	Video
Clinical	Data type	Ethnography data
Clinical	Data type	Rich text
Clinical	Data type	Oh yeah, when I was in public health school, there was a big nexus data set that I used.
Clinical	Data type	Case report forms
Clinical	Data type	Psychosocial
Clinical	Data type	Clinical
Clinical	Data type	Demographic
Clinical	Data type	Audiotape
Clinical	Data type	Survey
Clinical	Data type	Mostly MRIs—imaging data
Clinical	Data type	Clinical data form clinical trials
Clinical	Data type	Flow cytometry
Clinical/Pop Health	Data type	Medicare claims data
Clinical/Pop Health	Data type	Health insurance claims data
Clinical/Pop Health	Data type	HMO data (integrated health care system data)
Clinical/Pop Health	Data type	Audio recording or transcription
Clinical/Pop Health	Data type	Survey
Clinical/Pop Health	Data type	Paper
Clinical/Pop Health	Data type	Data on prevalence
Clinical/Pop Health	Data type	Data on how screening tools perform

JMLA Journal of the Medical Library Association

EC
Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical/Pop Health	Data type	IRI data
Clinical/Pop Health	Data type	NYC fitness data
Clinical/Pop Health	Data type	Data is at point of purchase
Clinical/Pop Health	Data type	Influence of food env on BMI—that's all secondary data
Med Ed?	Data type	Medical school registry: 85–90% medical student provide consent to include their educational and performance data in a research registry. That includes baseline OSCE (Objective Structured Clinical Examination) performance, course grades, encounter logs, clinical skill OSCE, USMLE step clinical skills exam, graduation questionnaire.
Med Ed?	Data type	Residency research registry: >200 residents across 6 residency programs with 80–90% consent rates. Data includes clinical evaluation (faculty rating), annual OSCE performance, 360 degree data (patients, staff, peers assessments).
Med Ed?	Data type	Quality data
Med Ed?	Data type	Chart data
Med Ed?	Data type	Survey data
Med Ed?	Data type	OSCE data
Statistics	Data type	Brain imaging data
Basic	Desired data services	If the school was researching what's out there, finding the best things and making us aware of them, licensing and releasing them.
Basic	Desired data services	Would be very useful if we had some support and we had a resident curator or software guy—get all the old data together
Basic	Desired data services	We'd like to come up with solutions to have access within this community (his specialty of research)—that alone is very difficult
Basic	Desired data services	Would be good to have a cloud, Amazon or something, and someone else will manage it. I don't want it to cost us to do it
Basic	Desired data services	It would be very useful if we had some ,
Basic	Desired data services	Uploading data would require someone to be resident here, learn about our datasets, also know things we don't know. New resources could help, such as programmers—it's hard to communicate across levels, tell the postdocs, for the benefit of human kind, do these things
Basic	Desired data services	Could benefit from a few terabytes of storage
Basic	Desired data services	There's nothing that provides any sort of real structure to the information that's being stored
Basic	Desired data services	Always been an issue—where does data collected to the datacore go?
Basic	Desired data services	Something that would be interesting would be archiving all this stuff

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Desired data services	Imaging dataset for microscopy—meta data in the files might contain a lot of data on parameters of images—a lot of little details—if that could be captured in a little table or a Word file or something and it didn't have to be structured, could be just a block of info, if that were available and entered by the people who were entering into database, maybe would not have to be so structured—field for a metadata table, up to the individual
Basic	Desired data services	And then maybe a movie or something that captures one output in terms of how u process the data and things
Basic	Desired data services	For published data but then could also be useful for us to be able to access the organize data that wasn't published
Basic	Desired data services	If there was a standard for that (organizing data), that would be really helpful when someone leaves the lab
Basic	Desired data services	If there was a way to automatically upload from hard drive to server, have duplicate—kind of a gigantic Dropbox type system, that would be great
Basic	Desired data services	Would be useful internally—if there were systems that could integrate data from any given platform, that would be fantastic, it's a fantasy
Basic	Desired data services	Interested in some sort of service to help him estimate costs when putting together grant budget (and he brought it up again later)
Basic	Desired data services	Would like to be able to get a lot of public data and download it and reanalyze it, don't have the space right now.
Basic	Desired data services	I'd love something else, such as what Theodora is talking about [re: MCIT storage]
Basic	Desired data services	If there was a formal format, archival [procedure], data transparency, it would help a lot.
Basic	Desired data services	The other data management service I used here (library)
Basic	Desired data services	What I want is that as a portal for my lab, for outsiders, to get a hold of the data. The ability to interface with that would be lovely. In principle, there are other things that are more scientific—computer speed, the ability to put simulations online or larger scale simulations run on the desktops.
Basic	Desired data services	Keeping track of the old data. And the peripheral files, what's in the raw data, and some way to access or distribute the summary plots. My dream is that you can click on a figure and get the data
Basic	Desired data services	It's cyclical, once someone reaches their full stride, someone comes and takes them from you—there's a short window where they're actually really productive—because of that, it'd be nice to know here's a library of all the data—tell future generations where to go—I see the utility there
Basic	Desired data services	Data management is a really good one
Basic	Desired data services	My sense is that the technology evolves quickly. So this summer when they got more data, it was analyzed in a different way that it's a whole other story. It can't be a static entity that someone can go back and forth with. Partly b/c we work with this group that's at the cutting edge of analysis, who are constantly reinventing, it makes it hard to keep up

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Basic	Desired data services	It would be useful for people in my lab if they were forced to label their data in a certain way and keep it in a certain place where it would be easier for others in the lab to get to and understand
Clinical	Desired data services	Would be nice if IT had a centralized database we could use
Clinical	Desired data services	What would make it better is a dedicated staff—consistent people who have the expertise—if there were just some statisticians, database programmers, in the hospital who u know have worked on these things
Clinical	Desired data services	The most organized organizations he knows of have a panel of statisticians and database programmers—it makes it easier for them to get NIH grants
Clinical	Desired data services	Institution wise—PubMed type thing for data (data catalog) would be extremely useful, I have no idea what the person next door is doing
Clinical	Desired data services	Having a central place to figure out who's doing what would be helpful, he doesn't know who's using HCUP
Clinical	Desired data services	If they were told there were products that could do more for them, they'd be interested but what they have is working fine for them, and they can do things very rapidly.
Clinical	Desired data services	I'd like to improve the pipeline of getting clinical data streamed into medical records—would like better interface between data and records—as we're trying to increase patient flow, always come up against that
Clinical	Desired data services	I'd like to have a pac system here, so we could directly connect to the imaging facility—very soon to start PET imaging at NYU (been off campus until now)—we'd like to have direct conduits that way
Clinical	Desired data services	Don't like the remoteness of his location—not great for collaboration—should be surrounded by buzz so you run into people—far away so that never happens
Clinical	Desired data services	I would love to learn about available databases
Clinical	Desired data services	PET scans—different types of scans in brain, different age groups, different procedures to create that scan, collect continuous data—difficult to get, very expensive—but the data do exist
Clinical	Desired data services	Large population studies—New York area, family histories
Clinical	Desired data services	Anonymized data, one database doesn't talk to another database—if there was a neutral 3rd party that could connect the datasets into one dataset—state of New York—big problem with that is HIPAA
Clinical	Desired data services	Would love to be able to capture people in other ways, like spending habits—if could link voter registration database with database of who bought a car in the last year, that would be cool
Clinical	Desired data services	My vision is to have the master dataset that's mainly for internal purposes, that's for me and my direct collaborators. Secondarily, I would like to take a subset of that data for those people who have a complete dataset, and anonymize it, and create a publicly accessed database that you would need a password to enter—want to only have people who have complete data because other stuff won't be particularly valuable to the dataset—replica of his dataset (not full replica)

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical	Desired data services	Thinking from the beginning, these are the data sources you're going to be collecting [to do the data model], what's the likely size of this dataset, what software to put this in? Where should this live? Thinking from the beginning.
Clinical	Desired data services	Not having enough people who understand the data enough to do the analysis
Clinical	Desired data services	Better communication between everything and everyone (labs, pharmacies, etc.) would make her life easier. Politics and hierarchy can get in the way of this.
Clinical	Desired data services	It would be very useful to have all the data managed electronically. Each patient generates a big stack of paper, and the sicker the patient is, the bigger the stack—more tests, hospitalizations, etc.
Clinical	Desired data services	I have ATLAS.ti so I can login offsite—I don't know if it works going through onsite health? That's what we really want. That way, you can collaborate and code and merge everything. We need to establish something like that. Like any stats software. Like STATA, SPSS. But I'd like to promote STATA.
Clinical	Desired data services	I'd like to be able to run my data remotely
Clinical	Desired data services	Majority of studies less than 100 sample size—so there is an effort to dump the data into a larger database that would have more power
Clinical	Desired data services	Small pilot—have some money from CTSI, Leo has money—trying to put together cohort of children to show that we can do it—this will be my entre into the world of NYU as an independent data analysis, repository
Clinical	Desired data services	More external training. Using SPSS, data mgmt. training for staff. Privacy & confidentiality. Standard trainings for staff.
Clinical	Desired data services	Data repository. Applied scientists with basic scientists, there's benefit for sharing datasets but the language and understanding needs to be understood.
Clinical	Desired data services	Making remote access more seamless [would be helpful]
Clinical/Pop Health	Desired data services	Having programmers we could call on who have experience in this type of data—that's departmental infrastructure more than anything
Clinical/Pop Health	Desired data services	New tools—like Dedoose—she heard about it at a conference
Clinical/Pop Health	Desired data services	We spend a lot of time cleaning the data—this is this huge amount of work between collection and analysis—because I'm junior, I've always done that myself. In this project, it's completely bogging me down, and it would be done better and more efficiently if it was done by someone else
Clinical/Pop Health	Desired data services	It would be great to have someone from the library have some help in thinking about evaluate a workflow for data and data management—I know u have a lot of experience in this
Clinical/Pop Health	Desired data services	Coordinating a storage plan—this comes up a lot

JMLA Journal of the Medical Library Association

Supplemental electronic content to
J Med Lib Assoc. Jul;103(3):DOI: http://dx.doi.org/10.3163/1536-5050.103.3.005
www.mlanet.org

Clinical/Pop Health	Desired data services	To have some thoughts and guidance and best practices for what would work best with the systems in place at NYU—in context of NYU what would be the best way to have these datasets managed and stored
Clinical/Pop Health	Desired data services	Can never have too many programmers
Clinical/Pop Health	Desired data services	Just having access to a super computer, maybe through citrix or something so that I don't need to have this amazing computer, but I have access to processing power
Med Ed?	Desired data services	We've often thought that what we need is a data manager. We don't really have one.
Statistics	Desired data services	Cataloging of their data, which is now organized in folders by researcher and then by project.
Statistics	Desired data services	For example, if my colleagues' data was put somewhere in an organized way with minimal description as opposed to everyone keeping it on their laptops somewhere, it would be much better
Statistics	Desired data services	Would include description of study, description of data files, description of how they're organized
Basic	Perception of data usefulness	People keep a lot of samples that they've prepared, but nobody's going to see these samples, there's not just an identity assigned to it, but also a concentration assigned to it in their lab notebook.
Basic	Perception of data usefulness	The molecular bio stuff could probably be useful for the long time
Basic	Perception of data usefulness	Genome sequence data, people like to see the raw files, because you can come out with very different results—that's where raw data becomes useful.
Basic	Perception of data usefulness	Data covers thousands of experiments performed around the world. Protein and peptide quantitation can be done to analyze/compare consistency.
Basic	Perception of data usefulness	But in terms of keeping data in a box that people could get into, I think that fits poorly for the type of data that we're using—we've been burned by that sharing
Basic	Perception of data usefulness	My data is a little different because this is on the edge of technology, right on the edge of what's possible—we have these traces that are hard to follow—each comes with a story—there are a lot of shades of grey, it's not finite pieces of data—it's got flavors and nuances that may not be apparent to people looking at data
Basic	Perception of data usefulness	So this disease is a genetic disease, and there are communities that want to develop treatment, so there's an effort to share this data and help investigators make sense and with people who are publishing. So there is that sense of wanting to share.
Clinical	Perception of data usefulness	Biospecimens are very valuable because they were collected before the disease, so they're good for looking at developing disease—we have a lot of them—I think it could be used for many years, I don't know—I don't think its limited—it's a unique source of data—we're one of the oldest cohorts
Clinical	Perception of data usefulness	From a scientific perspective, it's always useful for something, at least to generate a hypothesis
Clinical	Perception of data usefulness	Even transcript data can be used again, code again around a different question
Clinical	Perception of data usefulness	The datasets have a lot of use over time.
Clinical/Pop Health	Perception of data usefulness	"This type of data, people do tend to go back to, even decades later"