Data Management and Documentation Survey



1. Organizing the data and related files is important for current staff, and it is especially when there are staff transitions. For example, in DC, they organize all materials by source agency, then by type of file (data, programs, documentation). Do you have a standardized directory structure for storing the following file types: (check all that apply)

	Response Percent	Response Count
Data	74.3%	26
Metadata developed in-house	42.9%	15
Related documents (data use agreement, documentation from source)	57.1%	20
No standard directory structure	22.9%	8
	Comment	12
	answered question	35
	skipped question	0

2. Do you have written procedures for the following: (check all that apply)

	Response Percent	Response Count
Source data intake and storage	48.6%	17
Processing data	51.4%	18
Documenting data	45.7%	16
Backing up data	34.3%	12
No written procedures	34.3%	12
	Comments	12

35	answered question	
0	skipped question	

3. Would you be willing to share your procedures?

	Response Percent	Response Count
Yes, only with NNIP partners upon request	34.3%	12
Yes, only with NNIP partners through the NNIP website	17.1%	6
Yes, with the public through the NNIP website	22.9%	8
No	25.7%	9
	Comments	13

35	answered question
0	skipped question

4. What procedures do you use to safeguard confidential data? (check all that apply)

	Response Percent	Response Count
Drive encryption	28.6%	10
Password protection	77.1%	27
Storage in locked cabinets	45.7%	16
Signed data confidentiality forms from each staff member	48.6%	17
MOU clauses specifying minimum cell size or other standards for release	37.1%	13
No formal data protections	8.6%	3
Other (please specify)	48.6%	17
	answered question	35
	skipped question	0

5. Where are the data stored? (check all that apply)

	Response Percent	Response Count
On individual computers	31.4%	11
Shared network drive	85.7%	30
External hard drives	34.3%	12
In the cloud	25.7%	9
Other (please specify)	14.3%	5
	answered question	35
	skipped question	0

6. What is the primary computer software you use for:

	SAS	Stata	SPSS	R	Excel	ArcGIS	None	Rating Count
Data management	8.8% (3)	8.8% (3)	5.9% (2)	5.9% (2)	44.1% (15)	11.8% (4)	14.7% (5)	34
Statistical analysis	8.8% (3)	17.6% (6)	38.2% (13)	2.9% (1)	23.5% (8)	2.9% (1)	5.9% (2)	34
Mapping	0.0%	0.0%	0.0% (0)	0.0% (0)	0.0%	94.1% (32)	5.9% (2)	34

Other (please specify)

17

34	answered question	
1	skipped question	

7. How often do you store and document programs or code used to clean and prepare data?

	Response Percent	Response Count
Always	14.7%	5
Frequently	61.8%	21
Rarely	14.7%	5
Never	8.8%	3

Other (please specify)

6

34	answered question
1	skipped question

8. How often do you produce metadata for the following datasets:

	Always	Frequently	Rarely	Never	Rating Count
Source datasets	8.8% (3)	55.9% (19)	35.3% (12)	0.0% (0)	34
Derived datasets	12.1% (4)	48.5% (16)	39.4% (13)	0.0% (0)	33

Other (please specify)

4

answered question	34
skipped question	1

9. In what format is your metadata created and stored?

	Response Percent	Response Count
ArcGIS	26.5%	9
MS Access	14.7%	5
Word processing/spreadsheets	61.8%	21
Custom metadata system	14.7%	5
Do not create or store metadata	2.9%	1
Other (please specify)	23.5%	8
	answered question	34
	skipped question	1

10. Do you follow any metadata standard?			
	Response Percent	Response Count	
No	52.9%	18	
Yes, Federal Geographic Data Committee Standard (FGDC)	5.9%	2	
Yes, Data Documentation Initiative (DDI)	2.9%	1	
Yes, in-house standard	23.5%	8	
Yes, other (please specify below)	14.7%	5	
	Other (please specify)	8	
	answered question	34	
	skipped question	1	

11. Would you be willing to share your in-house metadata standard?			
		Response Percent	Response Count
Yes, only with NNIP partners upon request		46.2%	6
Yes, only with NNIP partners through the NNIP website		7.7%	1
Yes, with the public through the NNIP website		30.8%	4
No		15.4%	2
	ansv	wered question	13
	ski	ipped question	22

12. How do you disseminate data through your website? (check all that apply)

	Response Percent	Response Count
Downloadable files of aggregate neighborhood indicators	64.7%	22
Record-level non-confidential data	20.6%	7
Static images of charts, maps or other graphics	79.4%	27
Interactive maps	64.7%	22
None of the above	5.9%	2
Other (please specify)	32.4%	11
	answered question	34
	skipped question	1

13. In what format do you disseminate aggregate neighborhood indicators? (check all that apply)

	Response Percent	Response Count
Excel	56.3%	18
csv	59.4%	19
SAS/SPSS/Stata	6.3%	2
Shapefile	28.1%	9
Other (please specify)	34.4%	11
	answered question	32
	skipped question	3

14. If you share data online through a static website or interactive portal, how do you handle data updates?

	Response Percent	Response Count
The system or static displays accommodate multiple years of data.	62.5%	20
We update the pages with the latest year, and make the archives available to users.	40.6%	13
We update the pages with the latest year, and no longer share the older data through the website.	21.9%	7
	Comments	10
	answered question	32
	skipped question	3

15. If you post your downloadable data to a consolidated page on your web site or an external data portal, please list the URL(s) where the data are posted.

	Response Percent	Response Count
Hosted by our organization	100.0%	15
Hosted by external organization	40.0%	6
	answered question	15
	skipped question	20

16. Thank you for participating in the survey! Please share ideas about other data management and documentation topics you'd like to learn about from other partners.

management and documentation topics you'd like to learn about from other partners.	
	Response Count
	6
answered question	ı 6
skipped question	29

Page 1, Q1. Organizing the data and related files is important for current staff, and it is especially when there are staff transitions. For example, in DC, they organize all materials by source agency, then by type of file (data, programs, documentation).

Do you have a standardized directory structure fo...

1	I try to be consistent in how I store data- organize by type of data then source. This lack of standards will change as we move into open data.	Aug 26, 2014 8:35 AM
2	we reviewed and updated our structure at the beginning of the year; point being that we have structure but it needs to evolve periodically	Aug 25, 2014 12:36 PM
3	Metadata needs additional development - data limitations - esp. how data is collected and should be interpreted is often not written down in enough detail	Aug 22, 2014 7:48 PM
4	Right now everything is in topical files awaiting decision about the data platform.	Aug 22, 2014 4:02 PM
5	GiS data stored on computer by source agency and then by year.	Aug 22, 2014 12:05 PM
6	Raw data on the LAN and in internal database; proposals, signed MOUs filed physically and electronically	Aug 19, 2014 11:45 AM
7	The quality varies by dataset. We're working on more standardized approaches to better document them all.	Aug 15, 2014 3:23 PM
8	Metadata varies in quality across datasets	Aug 15, 2014 12:27 PM
9	It's always a work in progress, but we have standards that we try to follow.	Aug 11, 2014 6:46 PM
10	We have one, but aren't always good at using it.	Aug 7, 2014 4:14 PM
11	we haven't done much with the metadata items developed in house but it is there (historical metadata)	Aug 7, 2014 2:58 PM
12	We are working on a more standardized structure for cross-state indicator data, but it would probably not apply to "project specific" data.	Aug 7, 2014 2:37 PM

Page 1	Page 1, Q2. Do you have written procedures for the following: (check all that apply)		
1	Our procedures vary from staff to staff, they need to be standardized	Aug 28, 2014 11:40 AM	
2	We are working with Chicago (City & Chapin Hall) on adopting their metadata structures and tools	Aug 26, 2014 8:35 AM	
3	We have standard procedures for the above, but I am not sure that all of them are written.	Aug 25, 2014 12:36 PM	
4	Documentation is written into the programs written for the purpose	Aug 22, 2014 7:48 PM	
5	Ad hoc approach in the past will have to be more systematic with transition to a data platform.	Aug 22, 2014 4:02 PM	
6	From a data management standpoint, I am essentially a one-person shop, so I (loosely) follow procedures, but they are not written down. I enjoy taking NNIP surveys because they remind me of things I probably should do/have!	Aug 22, 2014 3:27 PM	
7	Some general notes, not as formally written up as we would like	Aug 22, 2014 11:28 AM	
8	We need to conver these procedures into more technical documentation/SOPs	Aug 19, 2014 11:45 AM	
9	The quality varies by dataset. We're working on more standardized approaches to better document them all.	Aug 15, 2014 3:23 PM	
10	Only for some data sets. We don't have written procedures for everything.	Aug 15, 2014 12:27 PM	
11	We document everything by project, so we know where to look when we need it. But that is just understood, not necessarily in writing	Aug 13, 2014 11:30 AM	
12	yes we have some written procedures but they are not formalized	Aug 7, 2014 2:58 PM	

Page 1,	Q3. Would you be willing to share your procedures?	
1	It depends on what the material is but we would be willing to share different documents at different levels.	Aug 28, 2014 11:40 AM
2	Would be willing to discuss with other partners, not sure there is great value in sharing, since I think these tend to vary according to scale/org capacity	Aug 25, 2014 12:36 PM
3	Procedures often in Foxpro with some Visual Basic to Excel processes. We will be retooling to more commonly used software.	Aug 22, 2014 7:48 PM
4	The final decision will rest with the new board for the reconfigured Memphis partnership.	Aug 22, 2014 4:02 PM
5	No = N/A in my case	Aug 22, 2014 3:27 PM
6	At least not at this time.	Aug 22, 2014 1:31 PM
7	We have none at this point, but know that these are needed.	Aug 22, 2014 12:05 PM
8	Yes, After we have time clean up the documentation	Aug 22, 2014 11:28 AM
9	We would love to be in a position that the SOPs are in a state that the public would understand!	Aug 19, 2014 11:45 AM
10	Unfortunately, we don't really have established procedures to share.	Aug 15, 2014 6:41 PM
11	Our systems are pretty antiquated	Aug 7, 2014 3:21 PM
12	potentially we would be willing to share some of these procedures as long as it didn't add delay to our already busy schedules	Aug 7, 2014 2:58 PM
13	Once the structure is more developed we hope to make it "open data"	Aug 7, 2014 2:37 PM

age 1,	Q4. What procedures do you use to safeguard confidential data? (check all that a	pply)
1	We tend to avoid working with confidential data. We have processes to handle confidential info (usually survey-based) within our center.	Aug 26, 2014 8:35 A
2	HIPAA and FERPA systems security compliance and use of best practices	Aug 25, 2014 12:36 F
3	Intermediate files with most critical security issues addressed are created quickly and the source material is locked separately	Aug 22, 2014 7:48 P
4	Just to clarify - I sign a confidentiality form, as I am the only one who has access to confidential data, but not all staff members do. We've had MOU clauses in the past, but not frequently. (We only have a handful of truly confidential datasets, though.)	Aug 22, 2014 3:27 P
5	Security Groups so only certain people can access data on a secure server	Aug 22, 2014 11:28 A
6	checking to see if our LAN drive is encryted by our IT dept. We do have a confidentiality agreement that staff will be signing this year	Aug 19, 2014 11:45 A
7	All hard drives are encrypted and large data sets are stored on secure servers that are encrypted and striped. Our servers are located in a secure facility with 24 hour surveillance, maintenance, and extremely limited access.	Aug 15, 2014 6:41 P
8	Data storage on protected server drives with rights restrictions	Aug 15, 2014 3:23 P
9	D3 has a password protected directory for proprietary data.	Aug 15, 2014 11:44 A
10	We adhere to any and all procedures that the source of the data deems necessary. This is usually captured in the MOU, or the general contract. Everyone is a little different so procedures vary.	Aug 13, 2014 11:30 A
11	We've done everything! Our preferred method is to restrict access on our shared drives to certain users, but many data providers require other arrangements. The most restrictive agreement we have requires data to only be used on a locked computer, not connected to the internet, in a locked office.	Aug 11, 2014 6:46 P
12	Electronic files are stored in secure server	Aug 8, 2014 8:52 Al
13	Health Department and other HIPAA data kept on removeable media and processed on laptop that is kept locked and not connected to network, no one in room not signed confidentiality forms.	Aug 7, 2014 3:21 PN
14	We don't use confidential data yet.	Aug 7, 2014 3:04 PM
15	IT safeguards (firewalls, acl's etc)	Aug 7, 2014 2:58 PM
16	We try not to work with it directly, when possible (e.g., requesting deidentified data sets).	Aug 7, 2014 2:37 PM
17	We have an off-site partner who handles confidential information. They then aggregate the data and send it over to us for dissemination. Thus there is no direct link between what is public facing to the confidential information.	Aug 7, 2014 2:10 Pf

Page 1, Q5. Where are the data stored? (check all that apply)		
1	Our cloud solution is the most secure and continuously monitored. We maintain static maps and reports on a shared network drive.	Aug 25, 2014 12:36 PM
2	Data storage issues are being re-examined	Aug 22, 2014 7:48 PM
3	ISC data are stored on a completely isolated virtual machine.	Aug 22, 2014 11:27 AM
4	only occasionally do we use external hard drives, usually for very large datasets so they don't clog up our shared drive.	Aug 11, 2014 6:46 PM
5	database	Aug 7, 2014 2:58 PM

Page 2, 0	Q6. What is the primary computer software you use for:	
1	We use multiple tools for data management, especially ACCESS	Aug 25, 2014 12:41 PM
2	Not None, but Foxpro for data management. It is still the best choice for scripting from data cleanup to report generation. Will be replacing, however. SPSS used when statistical analysis of a sophisticated sort is required. Foxpro programs address Census confidence interval issues	Aug 22, 2014 7:55 PM
3	We also use PostgreSQL and QGIS. We do conduct quite a bit of data management in ArcGIS, though, because we often receive data from our partners in ESRI formats.	Aug 22, 2014 3:27 PM
4	Access for data management	Aug 22, 2014 1:34 PM
5	We use R and SQL Server 2012 to manage the ISC database. We use excel and access for most other data sets.	Aug 22, 2014 11:34 AM
6	Excel is the most common for analysis but we are working more with R this year	Aug 22, 2014 11:32 AM
7	For data management we use PostgreSQL with a PostGIS stack.	Aug 22, 2014 11:15 AM
8	Maptitude by Caliper for mapping Weave for Data Visualization and mapping on web	Aug 22, 2014 10:27 AM
9	for data management: Access, SQL, Postgres, Text editors! Statistical Analysis: Growing use of R by staff	Aug 19, 2014 11:52 AM
10	We use SQL server for the management of large data sources.	Aug 15, 2014 6:44 PM
11	We also use SPSS for statistical analysis and Stata for mapping.	Aug 15, 2014 3:24 PM
12	We use MS Access equally with ArcGIS to do Data Management.	Aug 15, 2014 11:49 AM
13	Data management = Access, Oracle, and SQL Server	Aug 7, 2014 4:18 PM
14	Data Management and Stats performed using custom FoxPro programs	Aug 7, 2014 3:23 PM
15	we use various software applications we aren't stuck to just one i think that is what makes us more versatile. We pick tools that work well and do the jobs we need them to do. Always looking for ways to improve processes too so we aren't stuck in the same platform all of the time.	Aug 7, 2014 3:01 PM
16	Depends on data. In general, Excel is heavily used for everything too. QGIS also used.	Aug 7, 2014 2:39 PM
17	we are also moving to the open data portal offered from ESRI	Aug 7, 2014 2:11 PM

Page 2	Page 2, Q7. How often do you store and document programs or code used to clean and prepare data?		
1	Regularly, we prepare a data walk-through summary document with new data sets when we initiate a project with them.	Aug 25, 2014 12:41 PM	
2	Not frequently enough	Aug 22, 2014 7:55 PM	
3	The goal has been always, but the practice has occasionally slipped.	Aug 22, 2014 1:34 PM	
4	The quality varies by dataset. We're working on more standardized approaches to better document them all.	Aug 15, 2014 3:24 PM	
5	We are also working on formalizing version control procedures.	Aug 11, 2014 6:48 PM	
6	I am actually not the data guru but I do know we have some documentation in place for some of the procedures	Aug 7, 2014 3:01 PM	

Page 2, Q8. How often do you produce metadata for the following datasets:		
1	Need another choice - "not frequently enough"	Aug 22, 2014 7:55 PM
2	If anyone says "always," be skeptical! :)	Aug 22, 2014 3:27 PM
3	We ask for source data metadata, but don't always get it. And often the metadata is in a .pdf document which makes connecting the metadata to the data electronically difficult. we do store the metadata in a binder!	Aug 19, 2014 11:52 AM
4	not sure I am qualified to answer this	Aug 7, 2014 3:01 PM

Page 2, Q9. In what format is your metadata created and stored?		
1	Often built into custom report generators	Aug 22, 2014 7:55 PM
2	PostgreSQL, text files	Aug 22, 2014 3:27 PM
3	PostGreSQL database	Aug 22, 2014 11:32 AM
4	Visio for workflow; Microsoft Excel	Aug 15, 2014 3:24 PM
5	MS SharePoint	Aug 15, 2014 2:34 PM
6	HTML	Aug 8, 2014 9:39 AM
7	SAS datasets	Aug 8, 2014 9:01 AM
8	last that I knew we stored some metadata in our database	Aug 7, 2014 3:01 PM

Page 2, Q10. Do you follow any metadata standard?		
1	We have developed our process in conjunction with our Scientific Advisory Council, which reviews our work	Aug 25, 2014 12:41 PM
2	We follow FGDC when required/appropriate. We can't claim to fully follow any one standard, though.	Aug 22, 2014 3:27 PM
3	American National Standards Institute (ANSI) Standards	Aug 22, 2014 11:34 AM
4	Starting to use aspects from the Dublin Core for our tabular data	Aug 22, 2014 11:32 AM
5	WHAT!!	Aug 19, 2014 11:52 AM
6	Dublin Core	Aug 15, 2014 2:34 PM
7	We try to keep it very simple.	Aug 11, 2014 6:48 PM
8	unsure rather than No	Aug 7, 2014 3:01 PM

Page 4	Q12. How do you disseminate data through your website? (check all that apply)	
1	Reports with static maps and charts which include relevant lit review and local expert interpretation of findings. at neighborhood level.	Aug 25, 2014 12:42 PM
2	Will be totally rethinking web delivery. Most everything distributed through inquiry/ service system to this point.	Aug 22, 2014 7:56 PM
3	The platform we are aiming for will include downloadable files of aggregate indictors and interactive maps. Not sure yet about record level data.	Aug 22, 2014 4:05 PM
4	Downloadable files coming VERY soon!	Aug 22, 2014 3:27 PM
5	Interactive charts, graphs, tables	Aug 22, 2014 11:46 AM
6	All that are selected are soon to be updated on a new website.	Aug 22, 2014 11:21 AM
7	We are in the process of changing our website. Our plan is to be able to disseminate community indicator data on the new site.	Aug 15, 2014 2:38 PM
8	There is a lot of variation in what form data is released on the D3 website. It depends on the client, the dataset, and the product produced.	Aug 15, 2014 11:55 AM
9	Online database system (NEO CANDO)	Aug 8, 2014 9:10 AM
10	Interactive tables and charts	Aug 7, 2014 4:18 PM
11	interactive charts and tables too	Aug 7, 2014 3:01 PM

Page 5, Q13. In what format do you disseminate aggregate neighborhood indicators? (check all that apply)		
1	Community profiles website (fall 2014 launch)	Aug 26, 2014 8:37 AM
2	It really depends, but usually static maps, with relevant drill downs	Aug 25, 2014 12:44 PM
3	Final maps/ tables only at this point - in PDF and occassionaly XLS format	Aug 22, 2014 8:02 PM
4	Excel files of indicators have been project-related; aiming for more general accessability.	Aug 22, 2014 4:08 PM
5	geojson	Aug 22, 2014 3:27 PM
6	PDF reports	Aug 22, 2014 11:47 AM
7	HTML	Aug 15, 2014 3:26 PM
8	XML formatted spreadsheets readable in Excel	Aug 8, 2014 9:43 AM
9	Mainly just online but will do other formats upon request.	Aug 7, 2014 4:18 PM
10	Reports, Maps and graphics	Aug 7, 2014 3:24 PM
11	PDFs - We develop downloadable profiles	Aug 7, 2014 2:13 PM

Page 5, Q14. If you share data online through a static website or interactive portal, how do you handle data updates?		
1	This question requires an answer but there is no option for those who do not share data online	Aug 25, 2014 12:44 PM
2	History and timelines are critical to how we deliver data. The city makes standard property and crime information available, but with no archive or history at all.	Aug 22, 2014 8:02 PM
3	This is the goal that is reflected in the project-based approach we took in the past.	Aug 22, 2014 4:08 PM
4	Currently working on interface for this, but I think the plan is for archives to have the year embedded within the url	Aug 22, 2014 3:27 PM
5	We update data such as ACS to only show the most recent years, but any maps or visualizations will still work with older data	Aug 22, 2014 11:35 AM
6	The system accommodates multiple years, but it's manually uploaded every year	Aug 19, 2014 11:56 AM
7	Depends on the indicator set.	Aug 15, 2014 3:26 PM
8	Depends on the portal.	Aug 12, 2014 5:29 PM
9	Neighborhood profiles accommodate multiple years of data. The static charts and graphics are archived.	Aug 8, 2014 9:43 AM
10	We are actually waiting for the 2011-2015 ACS to show "real" change at neighborhood levels.	Aug 7, 2014 2:13 PM

Page 5, Q15. If you post your downloadable data to a consolidated page on your web site or an external data portal, please list the URL(s) where the data are posted.

	Hosted by our organization	
1	Site down this year during transition	Aug 22, 2014 8:02 PM
2	Historical project data still at http://cbana.memphis.edu	Aug 22, 2014 4:08 PM
3	www.datagateway.org - new design launching September 2014	Aug 22, 2014 3:27 PM
4	http://mpip.temple.edu/exploredata/default.aspx	Aug 22, 2014 1:37 PM
5	in process	Aug 21, 2014 3:09 PM
6	Various websites but here is the main one: http://bniajfi.org/vital_signs/data_downloads/	Aug 19, 2014 11:56 AM
7	https://data.kingcounty.gov/	Aug 15, 2014 3:26 PM
8	www,pinellasindicators.org	Aug 15, 2014 2:40 PM
9	viewer.infoalameda.org	Aug 15, 2014 12:31 PM
10	http://datadrivendetroit.org/data-mapping/interactive-tools/	Aug 15, 2014 12:01 PM
11	http://communitylinkcr.org/services/research-data-report-center/	Aug 13, 2014 11:35 AM
12	http://www.neighborhoodinfodc.org/comparisontables/comparisontables.html	Aug 8, 2014 9:43 AM
13	http://neocando.case.edu	Aug 8, 2014 9:10 AM
14	www.ctdatahaven.org	Aug 7, 2014 2:40 PM
15	www.neighborhoodnexus.org	Aug 7, 2014 2:13 PM
	Hosted by external organization	
1	City website is an important current portal - http://city.milwaukee.gov/mapmilwaukee#.U_fZrEi0bGM and http://itmdapps.milwaukee.gov/MyMHome/SearchDB2_prod.jsp	Aug 22, 2014 8:02 PM
_	www.stlouisdata.org, www.opendatastl.org - working on getting more of our data	Aug 22, 2014 3:27 PM
3	on these external sites	
6	on these external sites http://data.baltimorecity.gov	Aug 19, 2014 11:56 AM
		Aug 19, 2014 11:56 AM Aug 15, 2014 12:31 PM
6	http://data.baltimorecity.gov	

Page 6, Q16. Thank you for participating in the survey!

Please share ideas about other data management and documentation topics you'd like to learn about from other partners.

P		
1	We need to improve metadata and documentation processes. We are now starting to automate data cleanup for key datasets in python, and hope to share the code. Also talking to Chicago (City & Chapin Hall) on metadata standard & software catalog. This will come with expanded (and more-skilled staff).	Aug 26, 2014 8:39 AM
2	Agreements that specify rules for data sharing when numbers are low and confidentiality concerns are high. Model metadata elements that focus on data critique, limitations and "back stories" that inform well-considered analysis. Better tools for working with ACS data Better tools for statistical mapping given data uncertainty	Aug 22, 2014 8:06 PM
3	The reconfigured Memphis partnership will be interested in integrating a public neighborhood level data portal (probably NCCS) with selected data from an emerging IDS.	Aug 22, 2014 4:09 PM
4	We are about to hire staff for focus on this. this survey was useful to help us understand the skills we need to ask for, so thanks!!	Aug 19, 2014 11:56 AM
5	I'd love to hear more about data quality as part of data management.	Aug 15, 2014 3:27 PM
6	Metadata standards, file storage schemas, data and open data delivery systems	Aug 15, 2014 12:06 PM