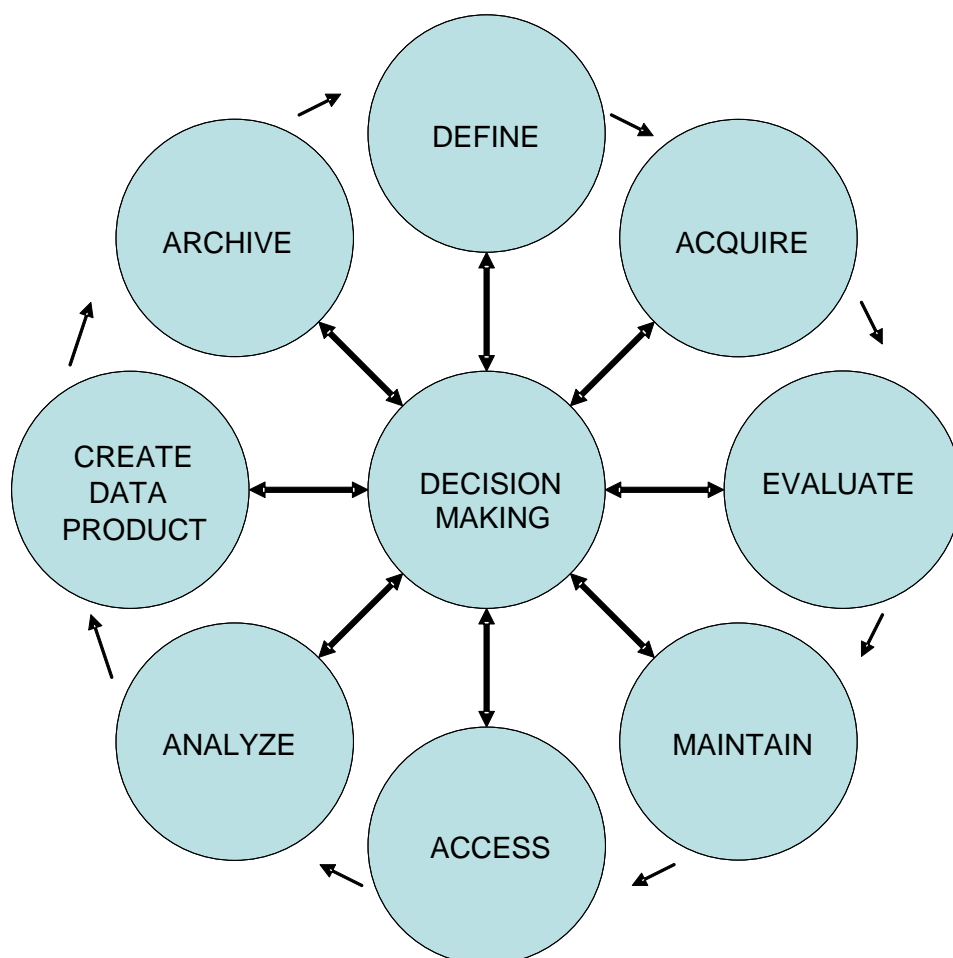


# RECLAMATION

*Managing Water in the West*

Technical Memorandum No. 86-68260-13-02

## Survey of the Reclamation Research Community Concerning Data Stewardship Practices



## **Mission Statements**

The mission of the Department of the Interior is to protect and provide access to our Nation's natural and cultural heritage and honor our trust responsibilities to Indian Tribes and our commitments to island communities.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

**BUREAU OF RECLAMATION**  
**Technical Service Center, Denver, Colorado**  
**Data Stewardship Core Team and**  
**The Emergency Management and GIS Group 86-68260**

**Technical Memorandum No. 86-68260-13-02**

**Survey of the Reclamation Research  
Community Concerning Data  
Stewardship Practices**

**Douglas Clark, Ph.D., Physical Scientist, Bureau of Reclamation, TSC**  
**Curtis Brown, Ph.D., Dir. of Research, Bureau of Reclamation**  
**James Nagode, IT Specialist, Bureau of Reclamation,**  
**Arthur Coykendall, Environmental Specialist, Bureau of Reclamation**



DATE	PEER REVIEW	CODE
NOV13 2013	Signature <i>John R. Carls</i>	85-824000
	Printed Name <i>John R. Carls</i>	
	Signature	
	Printed Name	
Author Initials		Peer Review Not Required

OFFICIAL FILE COPY		
DATE	SURNAME	CODE
Classification:		
Project:		
Control No.:		
Folder I.D.:		

# Contents

	Page
<b>Executive Summary .....</b>	<b>v</b>
<b>Background .....</b>	<b>1</b>
<b>Methods.....</b>	<b>1</b>
<b>Results .....</b>	<b>1</b>
Survey Response Rate.....	1
Question 1: Number of Research Projects .....	2
.....	2
Question 2: File Formats and Data Size .....	2
.....	2
Question 3: Metadata Utilization Rates .....	3
Question 4: Data Storage Depositories .....	4
Question 5: Data Back-up.....	5
Question 6: Information Management and Analysis Systems.....	7
Question 7: Hardware and Software for Data Sharing .....	8
Question 8: Sensitive Data Access .....	8
Question 9: Sufficient Speed for Data Access or Transfer .....	9
Question 10: Data Archival .....	9
Question 11: Data Discovery .....	10
Question 12: Biggest Challenges in Data Management .....	11
Question 13: Group Data Stewardship Policies.....	13
Question 14: Other Information, Suggestions, and Comments .....	14
Question 15: Reclamation Affiliation.....	16
<b>Summary .....</b>	<b>16</b>
<b>Discussion and Recommendations . .....</b>	<b>17</b>
<b>APPENDIX A: The Survey .....</b>	<b>18</b>



## Executive Summary

In April of 2013, the Reclamation Research and Development (R&D) office distributed an electronic survey to its research community to better understand types and amount of the computer resources their researchers require to manage their project data. This survey was part of a long-term effort to understand data stewardship practices in Reclamation and to subsequently improve the availability and use of related tools, practices, and standards. The survey solicited information about storage, access, computer speed requirements; metadata usage; policy guidance; management of sensitive data; and overall challenges. Several user profile questions were also asked.

In brief, this study reached the following conclusions:

- The response rate was 35.2%, i.e. 38 of 108 personnel responded to the survey.
- The majority of respondents were stationed at the Denver Technical Service Center.
- 87% of respondents were principal investigators on 2 or fewer research projects.
- The research that occurred using spreadsheets, model runs, GIS layers, imagery, documents, drawings, and databases, primarily took place in the 0-10 gigabyte range. In the >10-100 GB range, photos and other imagery predominated. In the >100 GB range model runs, imagery, and drawings predominated.
- The most frequently used data format types were spreadsheets, documents, and photos.
- Some 44.7% of respondents stored metadata with 0-10% of their data. A little more than 50% of respondents stored metadata for 0-20% of their data. About 84% stored it for from 0-50% of their data. Storing metadata with data was clearly not widely practiced amongst this sample of researchers.
- The primary storage devices were the Reclamation network drive and personal computers. A little less than half (45.7%) of responding researchers stored their data in more than one location.
- The Reclamation network and external hard drives were the back-up systems of choice. Ten (27%) of the researchers had more than 1 back-up system and 1 (2.7%) had 3.
- When asked, “Describe any information management systems (e.g. software, services, frameworks, methods, etc.) you use to maintain, organize, analyze and document your data”, surprisingly, 45.9% of researchers indicated that the question was either not applicable to their project or that they made use of no information management or analytic software. For those who did, the most frequently used types were



spreadsheets, discipline-specific software, and numerical/statistical packages.

- The most commonly used processes for data sharing were in rank order: email, physical storage devices such as thumb drives or DVDs, and Reclamation network drives.
- Only 10.5% of respondents reported managing sensitive data. Those who did reported using proprietary internet protocols, putting the data under the control of another agency, or restricting access to strictly to team members.
- 15.8% of respondents reported that sufficient speed in accessing or transferring data was a significant constraint on where they stored their data.
- 68.6% of respondents said that their data would be archived on a Reclamation network drive, a place, significantly, where they would not be readily discoverable by potential users interested in examining or using them.
- When researchers were asked where other researchers, managers, or the public could discover their data they reported that they could be discovered in rank order on a USBR website (26.3%), on PropC/R&D reports (21.1%), by contacting the researcher (21.1%), in a publication (15.8%), on an external website (10.5%), on the Reclamation network (10.5%), by contacting a research partner (10.5%), or by attending a professional meeting (5.3%). Of 23 total observations, 11 (47.8%) were for websites of one type or another, indicating that about half could readily be discovered electronically.
- The most reported overall challenges for data management among researchers were related to data access and sharing.
- 78.2% of respondents wrote that they worked in a unit that did not have any official guidance for the retention, security, and storage of data.
- When asked what other information, suggestions, and comments they had with respect to their R&D data, the most frequently mentioned remarks suggested that R&D should have a data port, that Reclamation should have a data management program, that the Reclamation FTP site should be upgraded, and that, overall, Reclamation should upgrade its data sharing capability.

## **Background**

In April of 2013, the Reclamation R&D office sent out a survey to 108 researchers who had current research projects in order to better understand the computer resources they use to manage research data. This survey was part of a long-term effort for understanding data stewardship practices in Reclamation and for subsequently improving the availability and use of related tools, practices, and standards. It also supported a larger effort to chart a course for data stewardship in Reclamation. A copy of the survey can be found in Appendix A.

## **Methods**

In April of 2013, Reclamation's researcher director, Dr. Curtis Brown, sent out an email message to 108 researchers asking them to take an electronic survey regarding their computer resource requirements.

Questions were asked about electronic data storage formats; metadata usage; data backup, storage, and archive processes; information management and analysis requirements; data discovery and sharing procedures; sensitive data management; requirements for speed of access and transfer; major data management challenges overall; and group data management policies. Several respondent profile questions were also asked about the office where the researcher was stationed, how many research projects he/she managed, contact information (optional) etc. Survey responses were compiled, coded, tabulated, graphed, analyzed, and interpreted.

## **Results**

### **Survey Response Rate**

Surveys were sent out to 108 researchers and 38 were received. The overall response rate was, therefore, 35.2%.

## Question 1: Number of Research Projects

Question 1 asked respondents to report how many S&T projects they were managing as principal investigator. Table 1 summarizes the results. Nearly 87% had 2 or fewer projects.

How many S&T projects are you currently conducting as the PI?		
Answer Options	Response Percent	Response Count
1	52.6%	20
2	34.2%	13
3	5.3%	2
4	5.3%	2
5	2.6%	1
6	0.0%	0
7	0.0%	0
8	0.0%	0
<i>answered question</i>		<b>38</b>
<i>skipped question</i>		<b>0</b>

**Table 1: Number of research projects by number of respondents.**

## Question 2: File Formats and Data Size

Table 2 summarizes the data requirements of the various file formats that researchers used for their research projects. These included: spreadsheets, model runs, GIS layers, imagery, documents, drawings, and databases. Spreadsheets led the way, followed by documents, photos, models, GIS layers, drawings, and then databases (Figure 1 and Table 3). When a format was used, i.e. its data requirements were above zero, by far the most widely used size bracket was the >0-10 gigabytes (GB) range. In the >10-100 GB range, photos and other imagery predominated with 5 mentions, while models, GIS data layers, and documents tied at 3 mentions. In the >100 GB range models, imagery, and drawings tied at 2 mentions each and the rest of the categories had one mention each.

Please identify the electronic data storage formats you are making use of for all of your S&T research projects and their approximate total size in GB. <a href="#">MB/TB to GB calculator</a>					
Answer Options	0	>0-10	>10-100	>100	Response Count
Spreadsheets GB	1	31	2	1	35
Models and model runs GB	11	10	3	2	26
GIS layers GB	13	9	3	1	26
Photos/images/graphics GB	3	22	5	2	32
Documents GB	1	29	3	1	34
Drawings GB	13	10	0	2	25
Databases GB (e.g. Access, FoxPro, Oracle,	13	7	2	1	23
<i>answered question</i>					<b>38</b>
<i>skipped question</i>					<b>0</b>

**Table 2: Storage requirements by format type**

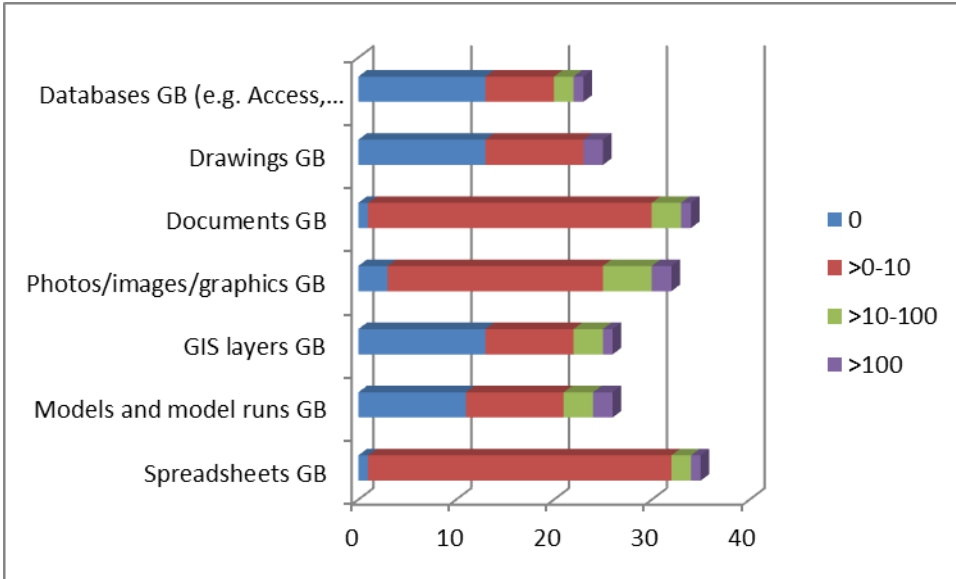


Figure 1: Usage by storage format type.

The most frequently used format types were, in rank order, spreadsheets (34 of 38), documents (33), and photos (29) (Table 3). At the lower end were models (15), GIS layers (13), drawings (12), and databases (10).

Answer Options	Use Frequency
Spreadsheets GB	34
Models and model runs GB	15
GIS layers GB	13
Photos/images/graphics GB	29
Documents GB	33
Drawings GB	12
Databases GB (e.g. Access, FoxPro, Oracle, MS/MySQL, etc.)	10

Table 3: Overall use frequency for categories greater than zero. N=38.

### Question 3: Metadata Utilization Rates

Question 3 sought to learn how often metadata were stored with data. Table 4 indicates that on average Reclamation R&D researchers stored metadata with 26.9% of their data. The standard deviation was 4.9 and the median was 20%. Table 5 shows the cumulative distribution of metadata utilization. Some 44.7% stored metadata with 0-10% of their data. A little more than 50% of respondents stored metadata for 0-20% of their data. About 84% stored it for from 0-50% of

their data. Storing metadata with data was clearly not widely practiced amongst this sample of researchers.

Roughly what percent of these data have documentation (metadata) stored with the data? (Metadata are information describing the who, what, where, when, why and how of the data collection.)			
Answer Options	Response Average	Response Total	Response Count
%	26.92	1,023	37
<i>answered question</i>			<b>38</b>
<i>skipped question</i>			<b>0</b>

**Table 4: Metadata utilization rates.**

<i>% of Data</i>	<i>Frequency</i>	<i>Cumulative % of Researchers</i>
10	17	44.74%
20	3	52.63%
30	7	71.05%
40	2	76.32%
50	3	84.21%
60	0	84.21%
70	0	84.21%
80	3	92.11%
90	2	97.37%
100	1	100.00%

**Table 5: Cumulative frequency of metadata storage with data. N=38**

## Question 4: Data Storage Depositories

Question 4 asked where researchers principally stored their electronic data, for example: their PC, external hard drive, Reclamation network, external cloud storage, partners' data storage, etc. Respondents reported the following categories, listed in rank order (See Table 6):

- The Reclamation network drive (62.2%)
- Their personal computer (PC) (56.8%)
- Partner storage (21.6%).
- External hard drive (5.4%)
- A flash drive (2.7%).
- External server (2.7%).
- Cloud storage (2.7%).

Of 37 respondents 17 (45.9%) reported using 2 or more storage systems, and 2 (5.4%) reported using 3 (Table 6).

Respondent	PC	USB Server/Network	Flash Drive Backup	Partner Data Storage	External Hard Drive	External Server	Cloud Storage
1	X	X					
2	X						
3	X		X				
4		X					
5				X			
6		X					
7	X	X					
8		X					
9	X			X			
10		X					
11		X					
12		X					
13		X			X		
14		X					
15	X				X	X	
16	X	X					
17	X						
18	X						
19	X	X					
20	X	X					
21	X						
22		X		X			
23		X					
24	X	X					
25				X			
26	X			X			
27	X	X					
28		X					X
29		X					
30	X	X					
31				X			
32	X	X		X			
33	X						
34	X						
35		X		X			
36	X						
37	X	X					
<b>Total</b>	<b>21</b>	<b>23</b>	<b>1</b>	<b>8</b>	<b>2</b>	<b>1</b>	<b>1</b>
<b>Percent</b>	<b>56.8</b>	<b>62.2</b>	<b>2.7</b>	<b>21.6</b>	<b>5.4</b>	<b>2.7</b>	<b>2.7</b>

Table 6: Data storage facilities. N=37.

### Question 5: Data Back-up

In Question 5, researchers were asked how they backed up their data. In rank order (Table 7) 67.6% used the Reclamation network, 27% an external hard drive,

Reclamation Researchers' Data Stewardship Practices

13.5% indicated that they used their PC for back-up, 5.4% used a flash drive, 8.1% partner storage, and 2.7% an external server, 2.7% cloud storage, and 2.7% did not know. The Reclamation network and external hard drives were the back-up systems of choice. Ten (27%) of the researchers had more than 1 back-up system and 1 (2.7%) had 3 (Table 7). Most researchers depended upon a single back-up system for their storage.

Respondent	PC	USBR Network	Flash Drive Backup	Partner Data Storage	External Hard Drive	External Server	Cloud Storage	Don't know
1	X	X						
2					X			
3			X					
4		X						
5		X						
6	X			X				
7		X						
8		X						
9								X
10		X						
11		X						
12		X						
13		X			X			
14		X						
15	X						X	
16						X		
17					X			
18			X					
19	X	X			X			
20		X						
21		X						
22		X		X				
23		X						
24		X			X			
25				X				
26		X						
27	X	X						
28		X						
29		X						
30					X			
31		X						
32		X						
33					X			
34					X			
35		X			X			
36		X			X			
37		X						
Total	5	25	2	3	10	1	1	1
Percent	13.5	67.6	5.4	8.1	27.0	2.7	2.7	2.7

Table 7: Data back-up processes. N=37

## Question 6: Information Management and Analysis Systems

Question 6 asked the researchers to report on the information management systems they used to maintain, organize, and analyze their data. Examples might be various software packages, services, frameworks, methods, etc.

Respondent	None	Spread sheet	Collaborative software, eg. MindJet, SharePoint, etc.	CAD and Drawing Software	Numeric Analysis and Statistical Software	Linux specific software	Discipline specific software	Geospatial software	MicroSoft Office and MS Addons	Carbonite cloud	Data Acquisition and Instrument Control	REDS	Endnote Bibliographic Software
1	X												
2	X												
3	X												
4								X					
5	X												
6	X												
7			X										
8	X												
9		X											
10		X		X			X						
11	X						X						
12	X												
13	X												
14	X												
15										X			
16	X												
17					XX	X	XX		X				
18		X											
19		X			X						X		
20	X												
21	X												
22			XX						X			X	
23	X												
24	X												
25									X				
26					X		XX						
27			X										
28													X
29			X										
30	X												
31								X					
32								X	X				
33							X						
34		X											
35		X			XX				X				
36		X			X						X		
37	X												
<b>Total</b>	<b>17</b>	<b>7</b>	<b>5</b>	<b>1</b>	<b>7</b>	<b>1</b>	<b>7</b>	<b>3</b>	<b>5</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>
<b>Percent</b>	<b>45.9</b>	<b>18.9</b>	<b>13.5</b>	<b>2.7</b>	<b>18.9</b>	<b>2.7</b>	<b>18.9</b>	<b>8.1</b>	<b>13.5</b>	<b>2.7</b>	<b>5.4</b>	<b>2.7</b>	<b>2.7</b>

Table 8: Data management, organizational, and analysis software utilization. N=37

Surprisingly, 45.9% of researchers indicated that the question was either not applicable to their project or that they made use of no information management or analytic software (Table 8). For those who did, 18.9% used spreadsheets, 18.9% used specialized discipline-specific software (e.g. a meteorological analysis package), 18.9% used numerical or statistical analysis software, 13.5% used collaborative software or applications, 13.5% used Microsoft Office software (exclusive of Excel which was coded to “Spreadsheets”), 8.1% used geospatial applications, 5.4% reported using software for data acquisition and instrument control, 2.7% use CAD or drawing software, 2.7% used Linux, 2.7% used Carbonite cloud, 2.7% used REDS, and 2.7% used Endnote bibliographic software. Clearly, the most heavily used packages were spreadsheets, numeric analysis, discipline-specific, and collaborative software packages. Interestingly, database management software such as ACCESS was only mentioned one time and was cited as one part of the Microsoft Office suite.



## Question 7: Hardware and Software for Data Sharing

Question 7 asked how researchers share their project data. As Table 9 indicates, email (76.3%), followed by physical storage devices (63.2%), and Reclamation Networked Shared Drives (55.3%) received the most responses. FTP (39.5%), Google Drive (23.7%), external cloud services (15.8%), and internal SharePoint sites (13.2%) shared the middle range. The least used methods were external SharePoint sites (7.9%) and websites (5.3%). One respondent listed “N/A”.

<b>If, during your research project, it is necessary to share your data with others, how is this accomplished? (SELECT ALL THAT APPLY)</b>		
<b>Answer Options</b>	<b>Response Percent</b>	<b>Response Count</b>
Physical storage device (e.g. thumb drive, DVD, external hard drive, etc.)	63.2%	24
Reclamation Networked Shared Drives	55.3%	21
Internal SharePoint site(s)	13.2%	5
External SharePoint site(s)	7.9%	3
Google Drive	23.7%	9
Non-DOI/Reclamation (external) cloud services (e.g.	15.8%	6
Email	76.3%	29
FTP	39.5%	15
Website	5.3%	2
N/A	2.6%	1
	<i>answered question</i>	<b>38</b>
	<i>skipped question</i>	<b>0</b>

**Table 9: Data Sharing Processes. N=38**

## Question 8: Sensitive Data Access

Question 8 asked respondents if and how they managed sensitive data. Table 10 shows that 89.5% of the respondents did not manage sensitive data and 10.5% did. Of those 4 respondents who did manage sensitive data, one said that he/she used proprietary internet protocol addresses. Another responded that the data were sent exclusively to USFWS offices. A third said that access was restricted to team members only. And a fourth reported that management of sensitive data would only be required in the future. Presumably, methods for their management would be worked out at that time as well.

Do you have data that are sensitive? (e.g. proprietary, ESA location data, etc.)		
Answer Options	Response Percent	Response Count
No	89.5%	34
Yes. Please explain how access to the sensitive data is controlled.	10.5%	4
<i>answered question</i>		<b>38</b>
<i>skipped question</i>		<b>0</b>

Table 10: Access to sensitive data.

### Question 9: Sufficient Speed for Data Access or Transfer

Question 9 asked whether sufficient speed in accessing and/or transferring the respondent’s data was a significant constraint on where the researcher could store his/her data. The answer was ‘yes’ for 15.8% of respondent users (Table 11).

Is sufficient speed in accessing and/or transferring your data a significant constraint on where you can store your data?		
Answer Options	Response Percent	Response Count
No	84.2%	32
Yes. Please explain.	15.8%	6
<i>answered question</i>		<b>38</b>
<i>skipped question</i>		<b>0</b>

Table 11: Sufficient speed for access or transfer

The explanations as to why sufficient speed was important included (a. the challenges of managing multiple files, (b. the slow speed of the Reclamation network compared to the personal computer (said to be 50-70% slower), (c. the impediment that slow download and processing speeds created for large model runs, (d. the challenges of communicating with remote computers, and (e. the demands that large point cloud processing made on older software and computer storage. One respondent offered no explanation. The fact that nearly 16% of respondents encountered problems in this area, suggests that a wider survey of Reclamation users’ speed challenges should be undertaken. If this percentage is found to hold bureau-wide, a serious threat to agency productivity may exist.

### Question 10: Data Archival

Question 10 asked researchers, “When your current research project is complete, where will your electronic data be stored?” By far the location of choice was the USBR network at 68.6% (Table 12). In the next range, the locations of choice were: the personal computer (20.0%), partner storage (14.3%), external hard drive (14.3%), and the R&D website/PropC (14.3%). At the lowest end, were

flash drives (8.6%), CD/DVD (8.6%), SharePoint or other Collaborative sites (5.7%), hardcopy (5.7%), and website other than R&D (2.9%). Interestingly, the only three sites that lent themselves to wide-spread data sharing were the PropC, non-R&D websites, and SharePoint (and other Collaborative sites)-- among the least used methods.

The majority of users (54.2%) employed a single archival system, 34.3% had 2, 8.6% had 3, and 2.9% had 4. System failure could, it appears, put a substantial percentage of researchers' data at risk (Table 12).

Respondent	PC	USB Server/ Network	Flash Drive Backup	Partner Data Storage	External Hard Drive	Website other than R&D	R&D Website /PropC	SharePoint or Collaborative Site	Hardcopy	CD/DVD
1	X	X								
2	X				X					
3		X	X		X					
4		X								
5		X		X						
6		X								
7		X								
8		X								
9		X								
10		X					X			
11		X								
12		X					X			
13				X	X			X		
14		X					X			
15	X	X								
16	X									
17	X	X			X					
18					X					
19			X							
20									X	
21	X	X								
22		X							X	
23		X								
24	X	X								
25				X						
26		X								
27		X	X				X			X
28		X								
29								X		
30						X	X			
31				X						
32		X								X
33										X
34		X		X						
35		X								
Total	7	24	3	5	5	1	5	2	2	3
Percent	20.0	68.6	8.6	14.3	14.3	2.9	14.3	5.7	5.7	8.6

Table 12: Archival sites and processes. N=35

### Question 11: Data Discovery

Question 11 asked researchers to explain where other scientists, managers, or the public could readily find the research data or discover its existence (e.g. via web search) once the research project was completed. (See Table 13). Only 19 of 38 respondents answered this question indicating perhaps, that data discovery was not of significant importance for many researchers. In rank order, the 19 researchers who did respond reported that their data could be discovered on a

USBR website (26.3%), on PropC/R&D reports (21.1%), by contacting the researcher (21.1%), in a publication (15.8%), on an external website (10.5%), on the Reclamation network (10.5%), by contacting a research partner (10.5%), or by attending a professional meeting (5.3%). Of 23 total observations, 11 (47.8%) were for websites, indicating that about half could readily be discovered electronically. Six of 23 (26.0%) of the observations required the individual searching for the data to contact either the researcher or his/her partner, not necessarily an effective way to make data widely discoverable.

Sixteen of the 19 researchers answering this question (84.2%) archived their data in 1 location, 2 stored it in 2 locations (10.5%), and 1 (5.2%) stored in 3. This indicates that there was little redundancy in discovery locations (Table 13).

Respondent	USBR Website	External Website	Contact Researcher	R&D Reports/ PropC	Publication	Reclamation Network	Contact Partner	Professional Meetings
1	X							
2		X						
3	X							
4			X					
5		X			X			
6				X				
7			X	X				
8			X					
9	X							
10					X			
11					X			
12	X							
13						X		
14							X	
15				X				
16						X		
17	X			X				X
18							X	
19			X					
Total	5	2	4	4	3	2	2	1
Percentages	26.3	10.5	21.1	21.1	15.8	10.5	10.5	5.3

Table 13: Data discovery locations. N=19

## Question 12: Biggest Challenges in Data Management

Question 12 was an open-ended question that asked researchers to list their biggest challenges with respect to data management. A large percentage of respondents reported that they had none (38.7%) (Table 14). Among those who did, data access and sharing were listed most often (22.6%), followed by time and budget issues (16.1%). Other issues mentioned were storage requirements (6.5%), archive and long term storage needs (6.5%), multiple users (3.2%), challenges associated with collecting the data (3.2%), data formatting issues (3.2%), changing storage media over time (3.2%), lack of a central repository (3.2%), retrieval speed (3.2%), converting data to knowledge (3.2%), and data stewardship awareness (3.2%).

Respondent	None	Storage requirements	Time/Budget	Multiple Users	Collecting the Data	Data Access and Sharing	Data Format Issues	Changing storage media: DAT, zip, JAZ, CD, DVD, etc.	Archive/Long Term Storage	Lack of a Central Repository	Retrieval Speed	Existing Data Not Organized	Converting Data to Knowledge. Data Analysis and Interpretation.	Data Stewardship Awareness
1	X													
2		X												
3			X											
4				X										
5					X									
6	X													
7														
8		X				X								
9						X	X							
10	X						X							
11														
12	X			X										
13								X						
14						X								
15						X			X					
16				X										
17						X				X				
18	X													
19	X													
20	X								X					X
21	X													
22	X					X								
23	X									X				
24														
25						X						X		
26	X													
27	X											X		
28	X													
29														
30			X										X	
31	X													
<b>Total</b>	<b>12</b>	<b>2</b>	<b>5</b>	<b>1</b>	<b>1</b>	<b>7</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>
<b>Percent</b>	<b>38.7</b>	<b>6.5</b>	<b>16.1</b>	<b>3.2</b>	<b>3.2</b>	<b>22.6</b>	<b>3.2</b>	<b>3.2</b>	<b>6.5</b>	<b>3.2</b>	<b>3.2</b>	<b>6.5</b>	<b>3.2</b>	<b>3.2</b>

**Table 14: Challenges of data management.**

### Question 13: Group Data Stewardship Policies

Question 13 asked, “Does your group have any official guidance for retention, security and storage of data?” 77.8% answered ‘no’ and 22.2% answered ‘yes’ (Table 15). So, overall, among Reclamation researchers there appears to be a lack of policy with respect to data retention, security, and storage.

Does your group have any official guidance for retention, security and storage of data?		
Answer Options	Response Percent	Response Count
No	77.8%	28
Yes. Please provide a citation below.	22.2%	8
<i>answered question</i>		<b>36</b>
<i>skipped question</i>		<b>2</b>

**Table 15: Existence of group policy for retention, security, and storage. N=36**

Among those who responded that they had policies, 38% (Table 16) said they used the Reclamation network drive, which, they presumably thought was a Reclamation policy requirement. Another 25% noted that they complied with Reclamation records policies. The rest of the categories received a single response:

- The researcher complied with the rules of other Federal agencies
- The respondent used a database for final reports
- The respondent adhered to regional policies and procedures
- The respondent faced time and budget constraints but retained important records and adhered to records policies
- The respondent was not aware of any policy

Respondent	Comply with other Fed Agency Rules	Database for Final Reports	Reclamation Network Drive	Regional Policies and Procedures	Time and Budget Constraints	Retain Most Important Records Only	Complies with Reclamation Records Policies	Not Aware of Any Policy
1	X							
2		X	X					
3								X
4			X					
5			X					
6				X				
7					X	X	X	
8							X	
Total	1	1	3	1	1	1	2	1
Percent	13	13	38	13	13	13	25	13

**Table 16: Citation of data stewardship policies. N=8.**

## Question 14: Other Information, Suggestions, and Comments

Question 14 asked researchers to list any other information, suggestions, or comments they wanted to report with respect to their R&D/S&T data. Fifteen researchers responded to this question and proffered up quite a variety of responses. Of these 15, 2 indicated that they had no comment. The following categories also had 2 (13%) responses (Table 17):

- S&T should have a data port dedicated to searching and accessing data and reports.
- Reclamation needs a bureau-wide program and/or bureau-wide information about data management.
- The Reclamation FTP site is unreliable and possibly responsible for data losses.
- There is a need for an effective data sharing capability.

The following categories received a single mention each (7%):

- There is a need for adequate storage space.
- There is a need for faster access to network drives.
- One researcher called for common Reclamation and R&D protocols for various data structures.
- Final documents must be made web-accessible to meet disability requirements.
- There is a need for a central data repository.
- Data management best practices and/or examples of best practices should be posted on the R&D website.
- Each person should have his/her own SharePoint site
- PropC should provide data and metadata storage.
- One researcher reported leaving data management to his/her partners.
- Finally, one respondent said the publication is more important than data management.

Respondent	None	S&T Should have a Data Port for Searching for and Accessing Data/Reports	Adequate File Storage Space	Need Faster Access to Network Drives	Common USBR or R&D Protocols for Various Data Structures	Making Final Documents Accessible to Meet Requirements	Need a Bureau-wide Program and/or Information for Data Management	USBR FTP Site Unreliable: Possible Data Losses	Need Effective Data Sharing Capability.	A Central Repository for Data Management	Post Data Best Practices and/or Examples on the R&D Web	Each Employee Should have Own SharePoint Site	Provc Should Provide Data and Metadata to Partners	Leaves Data Management more important than data management	Thinks publication is more important than data management
1	X														
2		X													
3			X	X	X	X									
4							X								
5	X							X							
6								X							
7									X						
8										X					
9											X				
10									X			X			
11													X		
12							X								
13														X	
14															X
15		X													
Total	2	2	1	1	1	1	2	2	2	1	1	1	1	1	1
Percentage	13	13	7	7	7	7	13	13	13	7	7	7	7	7	7

Table 17: Final comments on data management. N=15.



## Question 15: Reclamation Affiliation

Question 15 sought to learn the Reclamation group or region for each of the respondents. The overwhelming majority of the respondents came from the Denver Technical Service Center (69.4%), followed by the Pacific Northwest Region (16.7%), the Upper Colorado Region (5.6%), and, finally, the Great Plains, Mid-Pacific, and Lower Colorado Regions -- each at 2.8% (Table 18).

Please tell us your primary Reclamation Group (TSC) or Region.		
Answer Options	Response Percent	Response Count
TSC (Denver)	69.4%	25
UC	5.6%	2
GP	2.8%	1
MP	2.8%	1
PN	16.7%	6
LC	2.8%	1
<i>answered question</i>		<b>36</b>
<i>skipped question</i>		<b>2</b>

Table 18: Group or regional affiliation. N=36.

## Summary

A little over a third of the Reclamation R&D community responded to an electronic survey conducted to determine data management needs and practices related to research projects. The majority of the respondents came from the Denver Technical Service Center and most of the respondents were principal investigators for 2 or fewer research projects. Most of the work conducted using spreadsheets, model runs, GIS layers, imagery, documents, drawings, databases was in the 0 to 10 gigabyte range. Models, imagery, and drawings often required much higher amounts of space.

The most frequently used formats were spreadsheets, documents, and photos. Metadata were rarely stored with original data. The Reclamation network drive was the primary mode of storage, back-up, and archive. The most widely used information management systems (e.g. software, services, frameworks, methods, etc.) for organizing and managing data were spreadsheets, numeric/statistical software, and discipline-specific software. Email was the primary method used to share data, along with flash drives and DVDs.

Only 10.5% of researchers managed sensitive data. Security precautions varied from using proprietary internet protocols to putting the data under the control of another agency to restricting access to selected individuals. About 16% of users reported that speed of access and transfer was a significant constraint on where

they stored their data. About 70% of respondents said that their data would be archived on a Reclamation network drive. The most prominent place where data might be discovered was on websites (47.8%).

The most frequently reported challenges for data management were access and sharing. A large majority of respondents noted that their group had no written data stewardship policy for retention, security, or storage. When asked for suggestions, the researchers called for an R&D data port, an overall Reclamation data management program, and an upgraded FTP site.

## **Discussion and Recommendations**

While this sample cannot be said to be representative of the research community as a whole or of Reclamation generally, it may, nonetheless, be suggestive. The findings did not paint a favorable picture of current data stewardship practices. Metadata stored with original data shown to be much the exception, rather than the rule. Data were generally not readily discoverable. Redundancy not always practiced with respect to storage. Speed of access may be a significant drag not only on storage, but on productivity overall. Data stewardship policy was rare, if it existed at all.

Given these findings it is recommended that a Reclamation-wide assessment be undertaken of user needs related to data stewardship. This will provide a foundation for developing future guidance, training, tools, and system capabilities.

## APPENDIX A: The Survey

1. How many S&T projects are you currently conducting as the PI?
2. Please identify the electronic data storage formats you are making use of for all of your S&T research projects and their approximate total size in GB. MB/TB to GB calculator.

<b>Answer options</b>	<b>0</b>	<b>&gt;0-10</b>	<b>&gt;10-100</b>	<b>&gt;100</b>
Spreadsheets GB				
Models and model runs GB				
GIS layers GB				
Photos/images/graphics GB				
Documents GB				
Drawings GB				
Databases GB (e.g. Access, (FoxPro, Oracle)				

3. Roughly what percent of these data have documentation (metadata) stored with the data? (Metadata are information describing the who, what, where, when, why and how of the data collection.)
4. Where are your electronic data principally stored? (e.g. your PC, external hard drive, Reclamation network, external cloud storage, partners' data storage, etc.)
5. How are these backed up?
6. Describe any information management systems (e.g. software, services, frameworks, methods, etc.) you use to maintain, organize, analyze and document your data.
7. If, during your research project, it is necessary to share your data with others, how is this accomplished? (SELECT ALL THAT APPLY)

### **Answer options**

Physical storage device (e.g. thumb drive, DVD, external hard drive, etc.)  
Reclamation Networked Shared Drives  
Internal SharePoint site(s)  
External SharePoint site(s)  
Google Drive Google Drive  
Non-DOI/Reclamation (external) cloud services (e.g. DropBox, iDrive, Amazon, Mozy, etc.)  
Email  
FTP  
Website  
N/A  
Other (please specify)

8. Do you have data that are sensitive? (e.g. proprietary, ESA location data. etc.)

9. Is sufficient speed in accessing and/or transferring your data a significant constraint on where you can store your data?
10. When your current research project is complete, where will your electronic data be stored?
11. Once your research is completed, can other researchers, managers, or the public readily find your data or discover its existence (e.g. via web search)?
12. What are your biggest challenges for data management?
13. Does your group have any official guidance for retention, security and storage of data?
14. Please use this space to provide any other information, suggestions, comments, etc. that you would like to report with respect to the management of your R&D/S&T data.
15. Please tell us your primary Reclamation Group (TSC) or Region.

**Answer options**

TSC (Denver)

UC

GP

MP

PN

LC