RECLAMATION
Managing Water in the West

Technical Memorandum: 86-68260-13-01

Survey of River Restoration Programs Concerning Data Stewardship Successes and Challenges

River Restoration Pilot Studies: Data Stewardship

(Hart Mine Marsh before and after construction. Photographer: Gregg Garnett, Lower Colorado River Multi-Species Conservation Program)

U.S. Department of the Interior
Bureau of Reclamation

August 2013

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
Douglas Clark, Ph.D., Physical Scientist, Bureau of Reclamation, Technical Service Center
Curtis Brown, Ph.D., Research Director, Bureau of Reclamation
David Hansen, Soil Scientist, Bureau of Reclamation
Arthur Coykendall, Environmental Specialist, Bureau of Reclamation
James Nagode, Information Technology Specialist, Bureau of Reclamation

Technical Memorandum No. 86-68260-13-01

Survey of River Restoration Programs Concerning Data Stewardship Successes and Challenges

John Carlson **15 August 2013**
Peer Reviewer Date
Physical Scientist, Bureau of Reclamation, TSC, 86-68240
Title, Agency, Group, Mail Code

Sonja Kokos **23 August 2013**
Peer Reviewer Date
Adaptive Management, Bureau of Reclamation, LC8040
Title, Agency, Group, Mail Code

Scott Durst **26 August 2013**
Peer Reviewer Date
Fish and Wildlife Biologist, USFWS, FWS R2 NM ES
Title, Agency, Group, Mail Code

Tom Czapla **29 August 2013**
Peer Reviewer Date
Propagation & Research Coordinator Upper Colorado River Endangered Fish Recovery Program U.S. Fish and Wildlife Service, Post Office Box 25486, DFC
Title, Agency, Group, Mail Code

Scott Vandercooi **5 September 2013**
Peer Reviewer Date
Deputy Chief, GCMRC (Supervisory Fishery Biologist), USGS, 2255 North Gemini Drive, Flagstaff, AZ, 86001-1637
Title, Agency, Group, Mail Code

Contents

Executive Summary

The Reclamation Leadership Team directed the Data Stewardship Core Team to conduct a pilot evaluation of data stewardship practices for Reclamation river restoration efforts. The Core Team developed a survey (Appendix A) of current data stewardship practices and submitted it to 17 ongoing river-restoration projects to determine what lessons have been learned as data stewards have managed project data assets.

The survey was based upon the Reclamation data lifecycle (Figure 1). Respondents were asked to answer open-ended questions as to the successes and challenges they had experienced for each of the eight phases of the Reclamation data lifecycle. In addition, they were asked to relate their successes and continuing challenges related to both data standards and the use of acquired data to inform Reclamation decision-making. Finally, each project contact was asked if a written data management plan had been prepared for his/her river restoration project.

The researchers contacted 17 river restoration projects and received 14 completed surveys. *Technology, collaboration, funding challenges, data standards development, expertise,* and *integrated data management planning* were the most frequently mentioned themes overall. Responses from the various programs indicated that they faced similar challenges as their scientists pursued project river restoration work.

River restoration programs are often large and complex undertakings that can generate substantial volumes of data in a variety of formats, often gathered from multiple agencies. Such large and complex volumes of data cannot be managed without the application of technology, but application of technology alone is insufficient for successful data management. Without internal and external collaboration, fully integrated data management planning, the implementation of data standards, and pertinent data management expertise, the data collection effort risks disorganization. Without these additional elements, respondents spoke of trying to manage data acquired and managed using incompatible protocols, naming conventions, formats, sampling designs, custody practices, QA/QC processes, analytic procedures, metadata documentation, and the like. Without these elements in place, what could result was a disorganized and uncoordinated assembly of ad hoc, stovepipe processes that carried the risk of allowing personnel to go off in different directions and return with data products that did not meet decision-maker requirements.

Besides the challenges listed, other problems project data stewards faced included determining and implementing data sharing best practices, developing successful archives, managing multiple databases on multiple computers over multiple years, managing legacy data and unconventional data (eg. specimen collections), overcoming field work challenges, managing data interpretability issues, making

certain that contractors deliver quality data, serving both technical and policy requirements, mission/scope creep, and managing high volumes of data requests.

With respect to data planning, of the 14 respondents, 9 reported having no written data management plan, 3 reported that they were working on a plan or having a draft plan, 1 reported having a partial plan, and 1 reported having a fully developed plan. The recent hiring of project data stewards has resulted in the sharing of plan elements for the first time.

In sum, the survey indicated some commonality in successfully meeting requirements of data stewardship as identified in the data life cycle elements. These successes were linked to:

- Involvement of management in defining the objectives for data collection and monitoring programs and in reviewing status and progress
- Involvement of principal investigators and cooperators in planning data acquisition, QA/QC requirements, and documentation (metadata) during all phases of data collection, analysis and reporting
- Agreement by all participants on data access, data formats, and data maintenance requirements
- Regular or annual reporting on data collection and monitoring efforts as they relate to requirements for the program

These data gleaned from this survey can be viewed as a collection of the combined experiences that river restoration programs have accumulated with respect to data management. In many cases, the various projects have faced quite similar challenges, such as data access and sharing, contracting issues, IT issues, collaboration challenges, etc. These problems have often been faced in isolation, i.e. most often without the benefit of input from other river restoration project input. Undoubtedly, collaboration across programs would be useful. Indeed, a consensus opinion emerging from a data stewardship workshop held in May 2013 was that communication, coordination, and collaboration amongst projects would be beneficial both to individual projects and to Reclamation as a whole. A community of practice would facilitate such collaboration and a river restoration coordinator could facilitate the implementation of such a community.

Background

The Reclamation Leadership Team directed the Bureau of Reclamation (USBR) Data Stewardship Core Team, consisting of Curtis Brown (Research and Development), James Nagode (Information Resources Management), Arthur Coykendall (Policy and Administration), David Hansen (Mid-Pacific-GIS), and Douglas Clark (Denver Technical Services Center) to conduct a pilot study of data stewardship practices in those river restoration efforts for which Reclamation had a substantial interest. As a first task, the core team developed a survey of current practices and submitted it to 17 ongoing river-restoration projects. The purpose of this exploratory survey was to determine what successes and challenges river restoration data stewards have experienced as they managed project data assets.

Methods

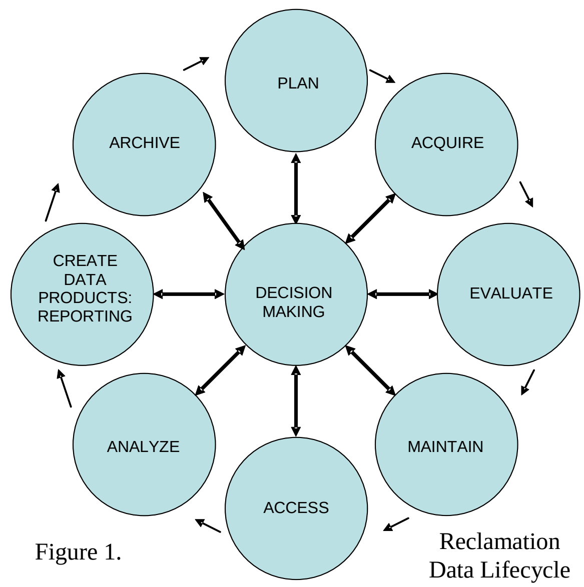

Figure 1. Reclamation Data Lifecycle

The survey was based upon the Reclamation data lifecycle (see Figure 1). Respondents were asked to list the successes and challenges they had experienced for each of the eight phases of the Reclamation data lifecycle. The questions were open-ended (see Appendix A). In addition, respondents were asked to describe the successes and challenges they have encountered related both to the use of data standards and the use of acquired data to inform Reclamation decision-making. Finally, each project contact was asked if a written data management plan had been produced.

In all, 17 river restoration projects were contacted. These were listed on the Reclamation River Restoration website www.usbr.gov/river and in the publication: *Bureau of Reclamation River Restoration Program*: *A Summary of 16 Programs and Shared Institutional Challenges (USBR R&D Office, 2011).*

Data stewards or data points of contact from the following projects were surveyed:

- Platte River Recovery Implementation Program
- Gila River Basin Native Fishes Conservation Program
- Lower Colorado Multi-Species Conservation Program
- San Joaquin River Restoration Program
- Federal Columbia River Power System
- Savage Rapids Dam Removal
- Trinity River Restoration Program
- Upper Colorado River Endangered Fish Recovery Program
- San Juan River Basin Recovery Implementation Program
- Bay Delta Conservation Project
- Middle Rio Grande Endangered Species Collaborative Program
- Clear Creek Restoration Program
- Glen Canyon Adaptive Management Program
- Battle Creek Salmon and River Restoration Project
- Missouri River Ecosystem Restoration Plan
- The Yellowstone River Intake Diversion Dam Modification
- M2 Whitefish Island and M2 WDFW

Results were collated, coded into a single database, and then tabulated across programs.

Results

In all, 14 River Restoration Projects responded to the survey. Several river restoration projects were eventually excluded from the survey for a variety of reasons. The Gila River Project data stewardship contact person indicated that the project is not a river restoration project, but a species conservation project. In addition, the project was said to contain so few data points that a database, and hence, a data management system were not justified. The Missouri River Ecosystem Restoration Plan was defunded in 2012. The Yellowstone River Intake Diversion Dam Modification restoration data collection effort is set to occur at a future date. The Red Bluff project was dropped because responders indicated that there would be no biological or physical monitoring conducted. The Battle Creek Project was contacted, but the personnel who responded indicated that an external agency was doing the project data management. This agency did not respond to the survey.

We compiled the survey responses, developed a classification scheme, and then coded content at each phases of the data lifecycle. In what follows, we give the reader an overview of the most important themes mentioned in each phase of the data lifecycle and then follow with some concrete examples of various successes and challenges which are derived from the surveys themselves.

Planning Phase

The key themes that emerged during the planning phase were involved with developing collaboration, making use of integrated project planning processes, and making use of data stewardship expertise. Examples are listed below within each thematic category.

Successes

- *Collaboration*. One respondent reported incorporating contractors, partners, and stakeholders into the planning process and seeking their feedback. Another mentioned obtaining regional support for good data management practices.
- *Integrated Planning*. Formalizing the design/planning process. For instance, aligning the data collection to the scientific management questions the project was designed to address and to NEPA compliance requirements or testing data collection processes in advance.
- *Expertise*. For instance, convening a technical advisory group of experts to assist in developing the most effective data collection methodologies.
- *Other Successes*. Building flexibility and feedback loops into planning, i.e. making adjustments to the plan based upon feedback and other project experience.

Challenges

- *Collaboration*. Examples included difficulties in convincing project managers to subscribe to good design and planning procedures, gaining internal support for the importance of developing and keeping data standards and maintaining data quality, coordinating data collections amongst multiple agencies, or convincing stakeholders and other interested parties to participate in planning meetings. Collaboration with multiple partners, many collectors from different organizations and agencies was viewed as challenging. In one case Reclamation data collectors were spread over two regions and four separate offices.
- *Integrated Planning*. One respondent had difficulty obtaining adequate pre-project cost and task documentation/justification and determining appropriate statistical analysis methods. Another talked disparagingly about the backwards process of documenting what data were actually been delivered and only then determining what should and should not be entered into the database.
- *Expertise*. Difficulties in obtaining expertise in a timely manner was listed.
- *Other Challenges*:
  - Managing multiple databases, multi-year projects, voluminous data sets, and legacy data sets.

- Constant pressures to change instituted data collection plans, especially pressures to increase both the number of data points and the reporting speed
- Managing a large volume of data requests,
- Managing poorly planned legacy collections, and
- Synthesizing large volumes of data.

Acquisition Phase

The challenges of field work, use or lack of technology, and presence or absence of collaborative processes were the predominant areas of focus related to the data acquisition phase of the lifecycle.

Successes

- *Field Collection*. One respondent mentioned successfully overcoming the challenges of conducting field sampling using a drill rig mounted on a floating platform, which owed to excellent forward planning by the drill crew.
- *Technology*. Standardizing and automating data entry with forms and templates, using automated data gathering technologies such as acoustic and laser field sensors, centralizing data storage and processing, and recording data electronically in the field were among the technological procedures that were viewed as helpful.
- *Collaboration*. Cooperation amongst and between Reclamation offices and other program participants was viewed as essential. One example was obtaining a regional commitment to share data acquisition and management processes. Another was gaining cooperation from program participants to centralize the data collection.
- *Other Successes*.
    - Developing and implementing multiple year RFPs for data collection.
    - One respondent reported success in retrieving legacy data from contractors
    - Large scale mark and recovery efforts that involved tagging over 10,000 fish with unique identifiers
    - Acquiring data under budget, which allowed for the collection of additional data.

Challenges

- *Field Collection*. One respondent wrote that restrictions on land access excluded some key monitoring points in the monitoring network. Other challenges: inclement weather, high flow events, remote locations, public interference, along with equipment installation, management, and retrieval.
- *Technology*. Data located on different hard-drives in many offices, sometimes over multiple years.

- *Collaboration*. Challenges with respect to requiring principal investigators to assume primary responsibility for data quality and management. Challenges associated with providing data access for different groups within the same office.
- *Other Challenges*.
    - Time and budgetary shortfalls,
    - Data formats and compatibility issues
    - Managing legacy data
    - Obtaining data from contractors

Evaluation Phase

Collaboration, followed by applications of technology, and making use of technical expertise were the predominant themes respondents mentioned in the surveys with respect to the evaluation stage of the data lifecycle.

Successes

- *Collaboration*. One office reported success obtaining regional commitments to develop standards. Another talked about coordinating with multiple entities to develop studies and evaluate draft reports. In addition, collaborative QA/QC efforts between Federal and contract staff were reported. Finally, another data steward reported collaboration across disciplines for improved:
    - Statistical methods
    - Query and analysis methods
    - Sampling methods such as fish marking and tracking.
- *Technology*. Automation of evaluation/data checking processes, error checking, and implementation of templates were cited as successes.
- *Expertise*. One office wrote about making use of external expert review of methods, data quality, and adaptive management practices. Another reported obtaining recommendations from experts to enhance future practices. Still another wrote about obtaining expertise on mark-recapture methods (and related statistical analyses) to review out-migrant salmon monitoring. Finally, another data steward mentioned convening and making use of a scientific advisory board or technical team to insure data quality.
- *Other Successes*.
    - Developing , agreeing upon, and using evaluation standards to insure quality

Challenges

- *Collaboration*. Some respondents, evidently reflecting on past experiences, called for early and frequent communication among evaluators, report writers, external reviewers, and project staff with respect to data quality expectations. In addition, one respondent expressed a concern that review recommendations were not being taken sufficiently seriously. Another expressed the concern that principal investigators were not always undertaking due diligence with respect to quality control and assurance. Still another noted that the participation of multiple organizations in study development "can result in a clashing of intellects and dueling scientists, who refuse to accept others' opinions as valid and worthwhile."

- *Technology*. "One of the largest challenges is trying to catch all potential errors before data are entered into the main project databases. Design changes or the addition of new types of data to be collected present challenges since they require the development and testing of new checking algorithms," one office reported.
- *Expertise*. One data steward noted that contractors and researchers in his/her project were found to lack expertise with data evaluation.
- *Other Challenges*.
    - Metadata standards not sufficiently developed or implemented, "making it difficult to evaluate the quality of the data."
    - The challenge of "gathering data using exact step-by-step protocols to ensure that all the necessary data are captured."

Maintenance Phase

Collaboration, technology, expertise, and time-budget issues were the primary themes reported with respect to the maintenance phase of the data lifecycle.

Successes

- *Collaboration*. Several respondents noted that requiring project principal investigators (and others intimately involved with the data sets) to take ownership of their data was critical for maintaining quality. Another office noted that vigorous agency management enforcement of data integrity policies was seen as contributing to the maintenance of data integrity and longevity.
- *Technology*. Several technologies were mentioned as contributing to the maintenance of data quality: for instance, establishment of a database management system such as SQL Server, developing a web-enabled database for monitoring data, or creating a common internet portal for data posting and sharing.
- *Expertise*. Successfully hiring and retaining staff with expertise in data stewardship was mentioned as being critical for success during the maintenance stage.

Challenges

- *Collaboration*. Keeping all cooperators on track producing consistent, timely data products was a challenge that received two mentions. The challenges of coordinating with multiple principal investigators received mention as well.

- *Technology*. One respondent noted that using electronic spreadsheets contributed in his/her estimation to the introduction of typographic and copy/paste errors. With another respondent, this individual also lamented the amount of time required for converting to a database management system.
- *Expertise*. Turnover of key staff at the project and program management level was noted as a challenge for the data maintenance.
- *Time and Budget*. Time and budget constraints were mentioned as adversely affecting data documentation and QA/QC reporting. Given these constraints, the question was raised as to how much QA/QC documentation is practicable? One data steward wrote that he set forth the benchmark principle that someone 50 years from now ought to be able to read the documentation and repeat the experiment. Still another respondent mentioned the challenges of obtaining QA/QC in a timely manner, "not waiting until months after the data were collected". Finally, sequestration and GSA-mandated server space limitations were mentioned as possibly constraining one project's ability to maintain its data library.
- *Other Challenges*.
    - Limitations on data storage
    - Identifying what data should be retained over time
    - Documentation of data that is sufficient for appropriate use and retention over the long term.
    - Study designs evolve over time requiring that older data be integrated into newer designs with database structures that accommodate past and present designs. The accuracy and precision of some measurements improves over time, for instance. In such cases, care needs to be taken in not assigning current precision to older data. Metadata also evolves over time, which can result in incompatible standards for data collected during different periods.
    - Acquiring and keeping iterations of data during their ongoing life stages, prior to QA/QC, during QA/QC, and final version.
    - Finally, one respondent asked a question that surely gives every project manager, data manager, or scientist nightmares: "What happens to data when it is used and errors can't be rectified?"

Access Phase

Applying technology, collaboration, data sharing, and maintaining expertise were the major themes respondents mentioned in association with the access phase of the data lifecycle.

Successes

- *Technology*. Applications of such technologies as shared drives, central data repositories, SharePoint servers, FTP servers, GIS databases, SQL servers, use of the Environmental Protection Agency (EPA) Exchange Network, use of portal technology, and the implementation of websites were said to contribute to success during the access phase of the data lifecycle. One respondent introduced the concept of the "data packages". Data were organized and documented by project managers in the Maintain phase, then rolled into zip files that could be posted for distribution.
- *Collaboration*. Collaboration was achieved internally and externally for the program by developing:
  - clear communication on common objectives within the program and between partners,
  - a common identity amongst those involved as an interagency program, i.e. people from different agencies took on a project oriented identity, and
  - regular and frequent meetings of program participants
- *Data Sharing*. Axiomatically, access was implemented by first developing processes for data sharing inside and outside the agency.
- *Expertise*. Expertise was viewed as critical-- both within the river restoration program and among collaborators, especially for technical solutions for access problems.

Challenges

- *Technology*. Lack of technology for common access was cited as one challenge. Differences in software/database platforms were also viewed as impediments to data sharing and thus, access.
- *Collaboration*. One respondent noted that schedule and time limitations sometimes made collaboration difficult. Another reported that there was no scheduled data sharing with the larger team. Still another opined that few analysts are willing to share draft data, "and the rest of the project team ends up waiting sometimes years to see any results!!!"
- *Data Sharing*. Differences in software/database platforms were said to hinder data sharing.
- *Expertise*. There was a need for expertise with respect to technological options for implementing data sharing and access between organizations. In one case, experts were sometimes said to be too busy to fully manage access issues, especially when they worked outside the agency.
- *Other Challenges*.
  - The sheer volumes of data that had to be stored and/or transferred and the related demands made on staff and IT systems.
  - Access to original data-- not just data summaries.
  - Difficulty obtaining data from contractors

Analysis Phase

The dominant themes cited for the analysis phase of the data lifecycle were, in order of frequency, technology, collaboration, expertise, and integrated data and project management.

Successes

- *Technology*. One data steward reported that the use of geospatial technology and geospatial data materially contributed to the quality of river restoration habitat analyses and also helped to "identify gaps and correlate activities with respect to critical habitat.” Among the other respondents, technology writ large was also said to facilitate query-building, automate error checking, data integration, and generally, make possible the management and analysis of voluminous data sets. Use of available technology also made it possible to undertake integration of a large variety of data types both from within the organization and from independent external sources.
- *Collaboration*. For numerous scientists across many agencies to conduct analyses, data stewards deemed it to be axiomatic that interagency collaboration must exist. Data sharing amongst these agencies was an essential part of this collaboration. Obviously, intra-agency collaboration must exist as well. For instance, one respondent noted that cooperation between modelers and the larger team insured that appropriate analyses were conducted and completed. The requirement for an annual report was mentioned by one respondent as an effective stick to enforce collaboration by making it imperative for principal investigators to take ownership of the management of their data. This measure, in turn, helped to insure analytic excellence
- *Expertise*. Respondents noted that expertise and technical competence are at a premium during the analysis and report generation phases of the data lifecycle. One respondent reported that “the hiring of a Research Statistician by GCMRC in 2012 has improved the quantity and quality of analyses conducted on data collected by many of our research and monitoring projects. The emphasis has been on biological and ecological studies, but this staff member's expertise is also available to other projects being conducted by GCMRC scientists.” Another noted that a high caliber researcher "conducted additional data analysis, above and beyond what was proposed and required, which yielded more robust analyses at no extra cost to Reclamation."
- *Integrated project and data management*. Respondents noted that it was important to take a holistic, integrated view of the project and its data management. For instance, one wrote that it was necessary to request from those that were to do the final data evaluation what exactly they wanted to evaluate so that the queries and protocols could be built towards this goal. Another respondent noted that the "information end products and associated analytic methods are important guides for structuring the data within the database." Still another wrote, "Motto: lead with planning the evaluation and reporting requirements."

Challenges

- *Technical*. Technology became difficult to apply to analyses when diverse data sets were organized and structured in different ways according to one data steward. Centralizing the project database was yet another challenge, especially when it drew upon a large variety of data sources. Another respondent noted that the time commitment to develop and test error checking software was a challenge.
- *Collaboration*. One data steward reported that where participating entities have differing priorities, opinions, and desires, difficulties can emerge with respect to the successful synthesis of the data required for analysis to occur. Inconsistent analytic results can also be generated, which can, in turn, be difficult to interpret. One respondent wrote, "With so many different groups and research efforts, it is difficult to know who is doing what and if efforts are being duplicated, and what is available to incorporate into existing models." Other challenges to collaboration during the analysis phase included:
  - Lack of collaborative tools.
  - Different priorities and opinions among participants in the analysis.
  - Coordination among investigators respecting both the analyses to be conducted, and, additionally, modeling data requirements.
- *Expertise*.
  - Lack of staff with sufficient expertise or training to undertake analysis was mentioned.
  - Lack of adequate review for the analysis that was performed
- *Other Challenges*. Other issues emerged during this phase:
  - One respondent noted the challenges associated with "staying within reasonable conclusions based on data precision/accuracy limits."
  - Another called for instituting processes to review contractor data for completeness.
  - Receiving data from different sources in a variety of data formats caused problems for some respondents.
  - Another data steward noted that the collected data were of insufficient quality to use in the required analysis, thereby reducing the robustness of the results.
  - Finally, budgetary issues were mentioned. Low or inefficient staffing levels were said to make it difficult for principal investigators to find the time for data analyses and hindered decision-makers from making use of the data.

Reporting/Creating Data Product Phase

For the reporting phase of the data lifecycle, respondents identified the major themes to be, in rank order: timely and quality reporting, use of technology, and collaboration.

Successes

- *Timely and Quality Reporting*. Reporting was considered successful when it was timely, of good quality, transparent, and frequent. Reading through the surveys, one finds that the requirement to write annual reports was considered to be helpful for insuring that data and information were delivered promptly to project managers and decision-makers. One respondent also viewed standardizing report formats for management as a measure that could help to insure timely reporting.
- *Technology*. Technology once again was reported as key to success. For instance, having the data stored in a centralized location was considered important for facilitating the timely production of data report products. Another respondent reported that their project's "DBMS has reporting features that allow it to generate MRGESCP (The Middle Rio Grande Endangered Species Collaborative Program) administrative, technical, and outreach-type reports in a timely manner using very few additional resources." Still another noted that the use of models helped to define the data universe, i.e. models require explicit data inputs.
- *Collaboration*. Once again collaboration was repeatedly mentioned as essential to project success. Often inducements were seen as helpful. For instance, principal investigators were given the responsibility to create scheduled reports and data products. Support from Reclamation management holding report preparers responsible for timely reporting was also viewed as contributive to success.
- *Other Successes*. One other factor mentioned as responsible for reporting success under the report phase was a highly professional and technically competent staff.

Challenges

- *Timely and Quality Reporting*. Reporting became a challenge when investigators were not held accountable to meet deadlines or when scopes of work did not require interim deliverable reports. Slow peer review processes could also impede the reporting of findings. One respondent wrote, "Timeliness is one the biggest challenges associated with reporting. As part of the USGS, our goal for products is often as publications in the peer-reviewed literature. This is rarely a quick process and can hamper our ability to produce products as quickly as we and those who rely on the information we produce would like." Another respondent wrote that when management does not hold report preparers accountable, reports may come in extremely late, or may never be delivered at all.

- *Technology*. On the challenge side, using email technology for reporting became problematic when the email systems were changed or went down. Moving to a web-based system was viewed as one workable alternative. That said, however, another respondent wrote that their project's current website and reporting processes led to the dissemination of outdated reports.
- *Collaboration*. One collaborative challenge was mentioned. Advisory committee acceptance and approval of findings was said to occasionally impede the reporting process.

- *Other Challenges*.
    - Staffing shortages could negatively affect the project's ability to manage, share, and evaluate data and report written reports in a timely manner.
    - Finally, "serving both technical and policy reporting needs" presented challenges for one respondent.

Archiving Phase

Two themes dominated the archival phase of the data lifecycle. The first was the role of technology in successful archive. The second was organizing an archival process, and part and parcel to that, finding or preparing a data repository.

Successes

- *Technology*. A variety of technologies were seen as contributing to archival success. The first was providing for a central computerized location for the project data such as a server, portal, website, etc. For instance, one respondent wrote, the "current MRGESCP website and DBMS are proving invaluable in this area, and archiving the data has not been an issue." Implementation of a data schema was also thought to show promise for contributing to a successful archive. One project reported that development of "data packages" or completed data sets that can be archived was viewed by users as contributing to successful data sharing. Finally, one project data steward reported that his office had ambitiously undertaken to digitize its entire library:

    "GCMRC library houses and archives historic GCES (Glen Canyon Environmental Studies) and current materials related to Glen Canyon Dam, Grand Canyon and Colorado River studies. The materials date back prior to the inception of our Center, and consist of research reports, maps, seismic graphs (traces), LIDAR data, video tapes, film, and Colorado River overflight photographs. We have embarked on a multi-year project to digitize and preserve our library holdings and send applicable materials to NARA (National Records and Archive Administration)."

- *Archival Process and Repository*. The second major theme for the archival phase revolved around simply having an archival process and a data repository. The forms these processes and repositories took were quite mixed. One respondent felt that publishing by itself offered some measure of archiving. Others considered that archival was the responsibility of those conducting the data collection. Presumably, these might be contractors, other agencies, or other offices within Reclamation. For some, putting the data on a centralized server was synonymous with archival. For others, the National Archive and Records Administration would be the ultimate repository.

Challenges

- *Technology*. Use of technology was viewed as necessary, but not sufficient for successful data archival. For instance, one must *have data* to archive. Thus, one respondent spoke of the need to make sure he received all final copies of finished data sets, so that he would have a basis to develop and maintain an archive. This requirement had to be communicated on a regular basis to those delivering data sets to the project. Other challenges associated with technology had to do with acquiring disk space for the increasing number of voluminous data sets. Formatting issues were also a problem, as will be discussed below.
- *Archival Process and Repository*. Process and repository challenges included (a. not having access to the actual raw data and (b. the challenges of having data of many types of media in many offices. Here is a case in point:

  "A number of disparate offices in PN and the TSC participated in this study. Individual offices were responsible for maintaining and archiving their own files. There was not a single point of archival for all the data collected in this study. There was a very diverse spectrum of data (AUTOCAD, GIS, soils lab data, geologic logs, hazardous materials testing, etc.) collected in this investigation and it is nearly impossible to concisely archive all these data into a single master archive."

- *Other Challenges.*
  - As mentioned above, another challenge for successful archival was inconsistencies in data formatting. "Various file formats and data idiosyncrasies make it difficult for data users to find data" reported one respondent. Some data were also reported to be in outdated formats.
  - One data steward noted that many older project files were either not in electronic form or the electronic version had been lost. In such cases, when paper files were also misplaced or lost, adequate archival of the data was compromised.
  - Staffing and funding shortages presented challenges, as well. Archiving materials can be very time consuming and require meticulous attention to detail which is costly. In addition, one

respondent noted, in a work atmosphere in which personnel are urged to quickly finish the project at hand, so they can move on to the next one, data archival can be overlooked.

Informing Management Decisions

Collaborative themes, technological themes, and themes related to data interpretation predominated in the decision-maker survey category.

Successes

Respondents reported numerous successes with respect to using project data to inform decision-making. For instance, one project lead used monitoring data in design of habitat construction. Other decision-makers used data to guide dam removal, schedule flows, enhance the persistence of stocked aquatics, and control non-native fish. More generally, project data supported adaptive management programs on behalf of ecosystem restoration and endangered species recovery.

- *Collaboration*. Collaboration was essential for successful decision-maker support. Several respondents noted that the requirement for annual reporting could facilitate collaboration. One wrote:

  "Annual Technical Reports and restoration Goal Technical Feedback Group meetings, in addition to more frequent flow scheduling meetings providing updates on system status and internal briefing papers, have allowed data analysis and ongoing monitoring results to be communicated to decision makers and stakeholders, informing management decisions by Reclamation management and the Restoration Administrator."

  Another respondent noted that the process of opening flow scheduling proposals to a wide constituency fostered collaboration. After the proposals are submitted, "A collaborative team then puts it all together into final proposals."

  Periodic meetings with presentations where research results could be shared with the larger project team were also said to facilitate collaboration in behalf of decision support.
- *Technology*. With respect to technology, one respondent wrote that models communicate how data will be used to help decision-makers. "There are several examples of decision analysis software packages that have been explored in PNAMP (Pacific Northwest Aquatic Monitoring Partnership)." Another mentioned the usefulness of reporting results on a website. Still another noted that the database management system creates the ability to synthesize existing data relevant to ecosystem recovery.

Challenges

- *Collaboration*. Challenges to collaboration were mentioned. Gaining data acceptance committee consensus as to the path forward for the project created a problem in one effort. Another respondent noted that it is important to schedule meetings with stakeholders to discuss report findings, "as many stakeholders do not read reports." Still another respondent mentioned that it is incumbent upon project personnel to find suitable methods and vehicles to make project data available for management decisions "to as large an audience as possible."

- *Technology*. On the challenge side, project reviewers recommended development of a decision support system that "accounts for new results and uses modeling to select among management options" would add significant capacity.
- *Interpretability*. Making data interpretable and then interpreting them responsibly were challenges that survey respondents raised. "Often people think that the data collection itself will inform decisions without consideration of the information content of the data," one respondent noted. Another mentioned that it is essential to translate data results in a manner that optimizes management comprehension and utilization. With respect to the actual use of the data, "Agency decision makers tend to favor data that meet their particular agendas and either disregard or downplay data results that are not favorable to their agency's interests."
- *Integrated Data Management*. Fully integrated data management planning received mention in the challenge section. "Making sure Management and Research Questions have been fully defined to influence the appropriate data collection" was a case in point. Another respondent noted that, "Retroactively asking questions and then looking for the data to address them is not efficient." Still another respondent reported: "Synthesizing the existing data is a huge undertaking, and although we have a plan for doing it, we have not been able to come to agreement as to what information should be synthesized first."
- *Other Challenges*.
    - Developing data standards, presumably to facilitate data processing, was also mentioned as being essential for informing decision makers, as well as implementing QA/QC procedures for in-house and contractor data collectors.
    - Finally time and budget issues once again came into play: "Time and funding to collect, process, and report on ongoing project efforts is limited, but does exist. Time and accountability. Management must REQUIRE better and more frequent reporting to ensure data results are available to make management decisions throughout the life of a project-- not just at the end when it's too late."

Data Standards

Successes

Respondents reported considerable success both using existing and/or developing new data standards. Substantial success was also reported in the deployment of protocols, naming conventions, and metadata, for instance. Successes were also reported with respect to hierarchical naming conventions, standardized metadata processes, templates, standardized codes, standardized protocols, and standardized reporting.

- *Technology*. Technology was reported to be helpful in the pursuit of standards. Template development was mentioned prominently.
- *Collaboration*. Obtaining general agreement as to where to start and how to proceed was mentioned as critical to successful pursuit of data standards.
- *Integrated Data Management*. Advanced planning was said to contribute to successful standards development according to some respondents. More than one mentioned that standards should be developed prior to any data collection. In addition, one respondent noted that multiple reviews throughout the whole process, from planning through reporting, can help to insure "that data collected are appropriate and will address management questions."

Challenges

On the challenge side, one respondent expressed concern about an absence of specific standards across Reclamation and between agencies. In some cases, existing, tried-and-true standards were adequate. In others, standards had to be developed. In still others, project personnel went to standards sources such as: http://www.fws.gov/stand .

- *Technology*. Making certain that scientists were using the most recent version of a template presented a challenge for one project. Keeping up with evolving technology was viewed as another challenge. For instance, the change from 400 kHz to 134 kHz PIT tags brought into question whether scientists would be able to detect older tags. In another project a web-based database was being developed to plot data in a standardized format.
- *Collaboration*. Once again, respondents viewed collaboration among personnel, activities, and agencies as critical to make data standards effective. Relatedly, resistance to standards was decried as problematic. One data steward wrote: "Data requests/demands challenge understaffed agencies/programs; agencies don't *always* collaborate/communicate with others when developing criteria, protocols, methods, standards, etc."
- *Other Challenges*.
    - One data steward created the following list of challenges to data standardization: "Many data collectors, many protocols, differing

applications, separate organizations with separate reporting and storage requirements. One extra step to deal with the data steward. The data steward cannot interpret data, so although data can be shared, without proper reporting, sharing raw data runs the risk of bad and/or contradictory interpretation by unqualified personnel."

- Problems crept up with respect to legacy data. "Existing data have been collected by many different agencies, stakeholders, and consultants-- and are stored in different formats and systems".
- Finally, time and labor constraints could impede standards development: "Getting the data standard process started then facing delays due to time constraints and timeliness. This, when you only have a short time period to influence data standards prior to collection in the field."

Written Data Management Plan

Survey respondents were asked in follow-up emails if their projects had a written data management plan. Of the 14 respondents, 6 reported that their projects did not have a formal written data management plan, 3 reported that they were working on a plan or had a draft plan, and 1 reported that his program did have one. The other 4 responses are listed below

1. "Nothing for scientific data, but protocols for Congressional and Administration data calls."
2. "Has a scope of work on data management on their website."
3. "…(F)or the fishery monitoring activities, there are a number of protocols or "formal" guidance on how data are to be collected (e.g. CAMP, various IEP protocols, etc.), but I think you are perhaps looking at a broader, over-arching document that would address data management across a spectrum of disciplines, and I don't think one exists." From the same project, another respondent reported: "I don't think we have a data management plan. We do have an ecosystem monitoring plan but it doesn't address the nitty-gritties of data management."
4. "We have several written plans for data management that cover different purposes. Getting folks to actually use them, though, is another story altogether!"

Assuming that we are looking for an over-arching data management plan in which each of the phases of the data lifecycle are addressed, numbers 1, 2, and 3 above would be coded a "no". Number 4 could be said to have a partial plan. To summarize then, we would have:

No Plan	9
Plan in Progress or Draft	3
Partial Plan	1
Plan	1

One may conclude that the majority of river restoration programs do not have written data management plans.

SUMMARY

Table 1.0 is an exhaustive, rank-ordered list of the frequencies of the various themes respondents mentioned either as successes or challenges across all survey questions and across all respondents. Each will be discussed in its order.

Table 1.0	**Total**
Theme	**Mentions**
The use of technology, templates, and/or web processes to facilitate data management and processing	69
Obtaining collaboration, buy-in, and support for data stewardship practices from management, partners, researchers, principal investigators, advisors, reviewers, and/or stakeholders	56
Time, effort, turnover, and budget issues	24
Acquiring, making use of, and keeping relevant expertise	24
Using and/or developing and implementing clear data standards, protocols, naming conventions, and metadata	15
Data collection viewed, planned, and managed in an integrated, holistically planned manner. Looking at the whole and the interdependent processes and components when undertaking data management.	13
Data sharing and reporting issues including data structuring and data formatting issues	11
Having or developing an archival process and location	11
Managing multiple databases or computers	10
Locating, managing, and/or applying standards to legacy/historical data sets and metadata	9
Managing voluminous data and requests for data.	8
Making use of feedback and adjustment mechanisms	8
Access, permitting, and other difficulties such as remoteness in collecting data in the field	6
Data intepretability and data interpretation	6
Timely, transparent, quality, and successful reporting.	6
Contractor issues	5
Data loss or data available in an unusable form for what is required, eg. the availability of summarized data when original data are required.	3
Obtaining data from contractors	3
Managing data requests	2
Project scope creep	2
Change management	2
Identifying who is managing the data	2
Managing non-traditional data	2
Serving both technical and policy requirements	1
Training issues	1
Managing multiple year projects	1
What happens when data cannot be rectified?	1
Excessive demand on IT systems	1

The two most frequently mentioned themes across all questions were related to technology and collaboration. River restoration programs are often enormous undertakings. Vast amounts of data must be acquired, evaluated, analyzed, reported on, submitted to decision-makers and shared. Multiple federal, state, and local agencies, private entities, and multiple stakeholders take part. Such undertakings would be impossible without computer technology and without collaboration.

Technology, we noted, was particularly important during the access, analysis, and archival phases, though it was mentioned in nearly every phase. Application of shared drives, central data repositories, SharePoint servers, SQL servers, the use of portal technology, and the implementation of websites were all put forth as helping to foster access. During the analysis phase, the application of geospatial technology was found to be especially useful for data management and processing. Technology writ large was also said to facilitate query-building, automate error checking, expedite data integration, and, generally, make possible the management and analysis of voluminous data sets.

Technology was necessary, but not sufficient. Collaboration was mentioned at every phase of the data lifecycle, but especially at the planning and evaluation stages-- and for the development and implementation of data standards. Planning ideally engages all the participants: management, scientists, reviewers, and stakeholders. Decision-maker requirements should drive the construction of the respective scientific questions, which would then drive the data requirements, and analytic methods. Scientists and reviewers can offer insights as to what data should be gathered, how they should be evaluated, how they should be analyzed, and how they should be reported to decision-makers. Stakeholders may have substantial historical experience to contribute, as well as local knowledge. Working collaboratively, these groups of individuals can make substantial contributions to overall project and data management planning . In their absence, as we have seen, data may be gathered for disparate purposes, using incompatible formats, a variety of protocols, problematic mixtures of different documentation formats, naming conventions, and standards—and the list goes on. For all these reasons collaboration was found to be essential to the river restoration data management process.

Time, turnover, and funding constraints, or inefficient use of resources, which comprise the next most frequently mentioned thematic category, all can and do have obvious impacts on data management. With insufficiencies in these areas corners are cut, delays are compounded, evaluations lack rigor, and in general, best practices are not strictly followed. As a result, data can be compromised, and decision-makers and resources put at risk. Similar things can be said of expertise. Lack of expertise can result in faulty sampling designs, data collection protocols, laboratory protocols, analytic methods, and results interpretation. Alternatively, expertise offers constructive assessment, which can ultimately improve project and data stewardship practice.

The next two thematic categories, namely the development of data standards, followed by integrated project and data design, fit hand in glove with the previous discussions. Our respondents noted that standardization of naming conventions, collection protocols, reporting templates, data documentation, and the like materially advanced the efficiency and the quality of their operations. Technology could sometimes help, for instance, by enforcing coding protocols. Integrated project and data design had the potential to put the project emphasis where it belongs, on the scientific questions that arise from the management decision requirements. With this approach, the study questions direct what data at what accuracy and precision are to be gathered and how they should be analyzed. In addition, collection protocols, data formats, naming conventions, and data standards are spelled out in advance. Obviously, collaboration furthers these processes, as does expertise.

Data sharing is, of course, part and parcel of data maintenance, access, and archive. Data sharing is, again, made possible by collaboration and facilitated by technology, funding, and expertise. Data archival provides one of the avenues for data sharing, but, as our respondents have noted, archival processes must be in place and a repository must be developed or acquired.

The remaining data themes listed in Table 1 primarily represent a catalogue of the challenges river restoration data stewards face. They must contend with data on many computer systems in many locations—and over multiple years. They must face the problems of legacy data, which include obsolete media and formats, poor documentation, and levels of precision and accuracy that are different from the current ones. They must find ways to handle the sheer volumes of data that are now gathered and the best practices to synthesize and integrate them. In addition, they must contend with questions about who can view what data. IT systems can be overtaxed as the result of these large volumes of data.

There are still other elements in the catalogue. For instance, accessibility issues during field investigations may result in data gaps. Users may make unfounded inferences from the data, given its precision and accuracy, or, as one respondent reported, they may choose to ignore data that contradict their own frame of reference or value system. Worse yet, what happens when the data one possesses are known to be wrong and cannot be rectified? What happens, hypothetically, if a researcher was monitoring an historical event like a flood and his/her data collection is found to have significant errors? In most cases another flood event cannot be dialed up.

Data stewards must contend with contracting issues ranging from poor data quality, to delays in delivery, to data loss, to poor documentation. In addition, more than once, respondents claimed to have difficulties obtaining the data Reclamation had paid to acquire.

Data demands can grow with project scope creep and that change must be managed. Non-traditional data such as specimen collections must be managed alongside data in databases.

The Table 1 rank order list represents significant institutional experience with the successes and challenges of data management, one from which newcomers can benefit. The implications of this are discussed below in the recommendations section.

There is one final issue to be revisited, one that is not on the frequency list. Only one river restoration program can currently be said to have a written data management plan that addresses each and every phase of the data lifecycle. At a data stewardship conference held in Denver in May of 2013, representatives expressed the need for such a document. Given the potential pitfalls of not having one, it appears that the routine production of such a document, perhaps using a template, would be advisable.

So, the challenges of river restoration data stewardship within Reclamation are quite extraordinary. The good news is that Reclamation has a cadre of dedicated managers, scientists, and engineers who have now confronted these challenges and they are finding collaborative processes, technologies, and other tools to manage them. In short, Reclamation river restoration programs now have a substantial body of data stewardship experience to draw upon.

RECOMMENDATIONS

This analysis shows that at each phase of the data lifecycle Reclamation river restoration subject matter experts, project managers, and data stewards have built up a substantial body of experience. They have learned that failure to plan a data collection can result in data inaccuracies, data inconsistencies, and data loss. They are learning how to plan for a data collection. They are also gaining experience with respect to how to collect many types of data in standardized formats, evaluate them to ensure their integrity, and maintain them in a ways that protect that integrity. They are learning how to make river restoration data assets accessible to those with a need to know, analyze them in ways that answer questions specific to our decision-maker needs, generate reports, and then successfully archive them for future use.

Given this growing body of experience it appears to us that that it would be beneficial to Reclamation if the various river restoration projects could cooperate with one another to share their experiences and their lessons learned. At present, each restoration effort is or has encountered many of the same problems. Collaboration amongst projects has the potential confront these problems as a community and share common solutions. A community of interest and practice could accomplish this goal, but such an endeavor would benefit by being facilitated by a river restoration coordination function.

Therefore, it is the recommendation of this report that Reclamation should consider the establishment of a river restoration coordination position that will have as one of its responsibilities the creation of a community of practice to address, among other things, the stewardship of data assets.

Appendix A: The Survey Questions

LIST OF SURVEY QUESTIONS FOR

RIVER RESTORATION PILOT STUDIES

I. Considering the range of data that your river restoration project collects or manages, please highlight a couple examples of successes and continuing challenges in the following areas:

a. *Planning* for the collection and management of these data.

i. Success:

____

ii. Challenge:

____

b. *Acquiring* the data.

i. Success:

____

ii. Challenge:

____

c. *Evaluating* the data quality.

i. Success:

____

ii. Challenge:

____

d. *Maintaining* the data integrity during the project's life.

i. Success:

____

ii. Challenge:

____

e. *Accessing and sharing* the data.

i. Success:

____

ii. **Challenge:**

____

f. ***Analyzing*** **the data.**

i. **Success:**

____

ii. **Challenge:**

____

g. ***Creating reports and other data products*****.**

i. **Success:**

____

ii. **Challenge:**

____

h. ***Archiving*** **the data after the project is completed.**

i. **Success:**

____

ii. **Challenge:**

____

i. **Using the data to inform** ***management decisions.***

i. **Success:**

____

ii. **Challenge:**

____

j. **Development or use of** ***data standards***

i. **Success:**

____

ii. **Challenge:**

____

II. **Please provide the name of the river restoration project and contact information for the data steward or the person whose duties most closely match those of a data steward:**

a. **Name of River Restoration Project:**

b. **Data Steward:**

   i. **Email:**

   ii. **Phone:**

The Reclamation Data Life Cycle