Data are 'Trust Resources':
a Strategy for managing
the U. S. Fish and Wildlife Service's scientific data

Developed by the Science Committee's

ad hoc Subcommittee on Data Management

Co-chairs:

Joel H. Reynolds*, Western Alaska LCC

Emily D. Silverman*, Migratory Birds, Division of Migratory Bird Management

Kelly Chadbourne, Region 5, National Wildlife Refuge System, Inventory & Monitoring

Sean Finn, Great Northern LCC

Richard Kearney, Region 8, Office of Science Applications

Kaylene Keller, Region 8, National Wildlife Refuge System, Inventory & Monitoring

Chris Lett, Information Resources & Technology Management, Data & Systems Services

Socheata Lor*, Region 6, National Wildlife Refuge System, Inventory & Monitoring

Kristin Shears, Division of Cost & Performance Management

John Swords, Ecological Services, *Environmental Conservation Online System* Program

*Science Committee member

28 March 2014

Acknowledgements

The Subcommittee thanks the Science Advisor, Gabriela Chavarria, for championing our efforts, and our respective supervisors who recognize the critical importance of data management to the Service. We also extend our sincere appreciation to Jen Jones at NCTC for facilitating the development of this strategy; to Margaret Beer (NPS), Tom Chatfield (BLM), Emily Fort (USGS), Thomas Miewald (USFWS R1), Stan Smith (USGS), and Alan Temple, Eric Kelchlin, and Mark Richardson (NCTC) for sharing their expertise and efforts; and to the members of the FWS Data Management listserv for their very helpful suggestions.

Suggested Citation

Reynolds, J. H., Silverman, E.D., Chadbourne, K., Finn, S., Kearney, R., Keller, K., Lett, C., Lor, S., Shears, K., Swords, J. 2014. Data are 'trust resources': a strategy for managing the U.S. Fish and Wildlife Service's scientific data. Report from the U.S. Fish and Wildlife Service's *ad hoc* subcommittee on data management. Pp 31 + xxvi. doi: TBD

Table of Contents

Information is a valuable national resource and a strategic asset to the Federal Government, its partners, and the public. In order to ensure that the Federal Government is taking full advantage of its information resources, executive departments and agencies must manage information as an asset throughout its life cycle to promote openness and interoperability, and properly safeguard systems and information. Managing government information as an asset will increase operational efficiencies, reduce costs, improve services, support mission needs, safeguard personal information, and increase public access to valuable government information.

Open Data Policy – Managing Information as an Asset
May 9, 2013
Director, Office of Management and Budget
Executive Office of the President

Data quality problems plague every department, in every industry, at every level, and for every type of information … [and] the costs are enormous. Studies show that knowledge workers waste up to 50% of time hunting for data, identifying and correcting errors, and seeking confirmatory sources for data they do not trust. ….

When data are unreliable, managers quickly lose faith in them and fall back on their intuition to make decisions … and implement strategy. They are, for example, much more apt to reject important, counterintuitive implications that emerge from big data analyses.

Fifty years after the expression "garbage in, garbage out" was coined, we still struggle with data quality. But … fixing the problem is not as hard as many might think. The solution is not better technology: It's better communication between the creators of data and the data users; a focus on looking forward; and, above all, a shift in responsibility for data quality away from IT folks, who don't own the business processes that create the data, and into the hands of managers, who are highly invested in getting the data right.

Data's Credibility Problem
By Thomas C. Redman, President of Navesink Consulting Group
The Harvard Business Review Dec. 2013.

Executive Summary

In order to make sound decisions and safeguard public information, the U.S. Fish and Wildlife Service (Service) must manage its scientific data effectively and transparently. Recent Executive Orders and Department of Interior (DOI) policies mandating improved open access to data extend this responsibility. The ready discovery and delivery of well-documented data are also necessary for the success of both our conservation partnerships, including the Partners for Fish and Wildlife Program, Joint Ventures, Heritage Programs, and Landscape Conservation Cooperatives, and our national action plans, such as the National Fish, Wildlife, and Plants Climate Adaptation Strategy. Further, addressing the long-term challenges of climate change and other landscape-scale threats, via efforts such as Strategic Habitat Conservation and conservation design frameworks, requires highly collaborative data-intensive conservation science and effective long-term maintenance of data. Service actions backed by rigorous science from documented and defensible data are less likely to face legal challenges.

Strong partnerships, science-based management, and public accountability demand an integrated program of data management that is understood and supported throughout the Service. Recognizing that the organizational efficiency and scientific credibility of the Service is compromised by insufficient attention to data management, the Science Advisor, Gabriela Chavarria, and the Science Committee brought together experts from across the agency to create an *ad hoc* Subcommittee on Data Management in 2012. The Science Committee and its Oversight Committee charged the Subcommittee to develop a *vision* for the management of Service's scientific data and a *high-level strategy*, articulating key challenges, necessary tasks, and recommendations addressing all aspects of data management.

The following pages outline a broad strategy to improve management of the Service's scientific information, motivated by an ambitious vision: **Scientific data are 'trust resources' and the foundation of our future conservation actions**. We describe the challenges to effective data management, focusing on the culture and organizational structure of the Service, as well as the need for commitment and resources. We recognize that a one-size-fits-all solution is neither appropriate nor feasible given the Service's diverse programs and distributed organization. A successful approach must address local, regional and national level needs with appropriate communication and coordination across these levels and with other DOI bureaus.

We detail four critical strategic directions – increasing awareness and support, implementing policies and procedures, developing the workforce, and budgeting for the long-term – and associated actions that will provide the foundation for an integrated approach to data management. The initial recommended actions address both local and national levels. The goal of the recommendations is to create an agency-wide ethos supporting consistency and excellence in data management.

The critical first step is the designation of leadership to create a home and voice for data management in the Service (Action 1.1). This entails (i) creating a national position to coordinate and lead data management in the Service, (ii) establishing a cross-Service committee to provide expertise and assist the new position, and (iii) identifying a community of early

adopters to champion new initiatives and speed development and sharing of local solutions. Following establishment of this leadership team, the Strategy recommends a comprehensive evaluation of the Service's current data management practices, including staffing, technical infrastructure, and budgeting (Action 2.1). This will guide refinement of the leadership team's near-term priorities and long-term goals.

Simultaneous with these planning steps, we propose action to address the urgent local need for technical capacity and resources 'on the ground' and to set the stage for success with the necessary Service-culture transformation. An award process will initiate one strategic three-year term investment per year in a Program or project that articulates a clear solution to pressing data management needs (Action 3.1). These awards will prototype the development of local solutions that meet regional and national needs and serve as case studies for training.

Having completed the evaluation and refined near- and long-term goals, we recommend action on the technical-capacity staffing needed to address the identified priorities (Actions 3.2-3.4). At this point, a broad, integrated effort to communicate with Service personnel about data management should also be initiated (Action 1.2). Longer-term, capacity building actions, including the creation of a robust and predictable budget model, follow.

These recommended actions represent investments in the Service's conservation mission that will quickly pay dividends, starting with improved efficiency for core scientific work and magnifying into more effective conservation decision-making. Taking action will begin reducing the high hidden costs the Service currently pays for its lack of adequate data management. Some recommended actions require re-allocation of staff time, some require re-direction of resources, and some require the creation of new positions. Central to all recommendations is the fact that the Service will only achieve the full value of its scientific data when it places as much emphasis on working with data in the office as it places on collecting data in the field.

The Service is a leader in developing collaborative conservation partnerships based on sound principles, like adaptive management and structured-decision making. These approaches have been integrated into the Service's culture and operations, and serve as a model for other organizations. It is now time for the Service to focus on leading the development of sound practices for conservation data management. Increasingly, partners and public are requesting access to the Service's scientific data. Our ability to efficiently and accurately provide data and information will define the Service's conservation impact, influence, and relevance.

Data management in an organization like the Service – with its multitude of overlapping offices, programs, stations, and partnerships – is complex. Staff needs to be skilled in the technical aspects of data management and have an understanding of the biological systems that the data represent. Collaboration must be facilitated and communication should be open; the sharing of resources and solutions needs to be seamless. If good data management were quick and easy, it would already be in place. This Strategy provides a framework to build effective data management into all Service science and to ensure we 'do the job right', from planning through management action, thus securing the long-term value of our conservation investments.

The Need

The U.S. Fish and Wildlife Service has a data management problem. There is a large disparity between the resources devoted to collecting data and those devoted to documenting, managing, and using data to inform science-based conservation decision-making. The following examples, drawn from the authors' experiences, illustrate many aspects of this disparity.

The status update for a listed species is delayed almost a decade because staff do not know how to extract monitoring data from the database in which they were stored.
Diagnosis: Lack of Standard Operating Procedures and inadequate documentation for data management; insufficient accountability for curating data for future use lead to problems when staff turn-over or offices move.

Scientific reports created by Service-funded contractors include tables and maps, but no data delivery. Data cannot be reused or repurposed.
Diagnosis: Inadequate contracting language and lack of collaborative data systems result in lost opportunities and increased costs.

Collaborations among Programs and with partners are harmed due to misunderstandings and inconsistent interpretations of policies related to 'data sharing' and 'open data.'
Diagnosis: Inadequate communication and lack of data sharing policies and procedures.

Science staff are pulled from regular duties to create awkward work-arounds in response to data requests, because servers are inadequate for the volume of data requested. Time is wasted due to network connection speeds that are too low to support the use of Regional data systems developed to ensure consistency and updating of core data.
Diagnosis: Inadequate hardware systems, software design, and/or internet connections for local data management needs, and lack of local staff with technical skills for diagnosing problems.

As these example highlight, Service Programs are not meeting the challenges of modern data management. The problems are ongoing and field staff regularly express the need for improvements: In 2000, the Refuge System's Fulfilling the Promise effort determined that the equivalent of 166 FTEs/year ($8.8 million or ~25% of all the Refuge System's biological survey effort) were spent on data collection efforts that were neither used for current decision-making nor managed in a manner that would allow future use (see Appendix A: Fulfilling the Promise WH9.1, 2002; Hal Laskowski, pers. comm. 2013). Though the reasons for this lack-of-use were not addressed directly by the Fulfilling the Promise report, the experiences of the current Subcommittee members suggest that it was largely due to bottlenecks associated with inadequate data management, which are not exclusive to the Refuge System. While the recently initiated Inventory & Monitoring Program has made advances on data management and use within the Refuge System, substantial problems remain for all Service Programs (Appendix B).

These data management problems are concerning because the quality of the Service's decision-making depends on the quality of its management of scientific data. Science-based decision making relies on ready access to scientific information. The production of scientific information – population and trend estimates for game birds, survival rates for species-at-risk, avian mortality estimates from wind turbines, predicted habitat response to controlled burns, maps documenting the occurrence of endangered species, etc. – requires high quality, well documented data. Access to high-quality data requires that Programs devote attention, expertise, and resources to managing their scientific data both during the project phase when the data are created (data stewardship) and during the transition to archiving and subsequent curation for long-term re-use (data curation). Insufficient attention to data management impairs the effectiveness and efficiency of current management decisions and severely limits the future value of data investments. A well-functioning data management program is critical to the Service's mission.

The Service's data management deficit is neither surprising nor unusual: comprehensive attention to data management, especially the curation of data for long-term use, is not standard practice in the conservation science community. Studies suggest that 80% of the data used as the basis for *peer-reviewed science publications* becomes irretrievably lost within 20 years (Gibney and Van Noorden 2013; Vines et al. 2013). With the exception of certain long-term monitoring efforts, the rate of loss is expected to be even higher for 'unpublished' data, such as those generated by many Service projects. Managing scientific data predominantly at the project level, with only limited resources committed to making sure the data are adequately curated for long-term use, practically guarantees the data won't be available for such future uses. Yet historical data are critical to investigating climate change impacts and developing adaptation strategies, for example, topics that were not of concern when the data were collected.

The Solution: A data management vision and strategy

The time has passed for debating *whether to act*; the time has come *to make it happen*. Full success demands comprehensive and integrated changes throughout the Service, as do the recent Executive Orders and DOI policies[1]. Some Programs and Regions have begun the transition,[2] but piecemeal efforts will not bring about the changes necessary to achieve our conservation mission. The Service can meet these converging data management demands, and meet them efficiently in a period of declining budgets, by taking steps to ensure that the conservation community is better able to realize a return on our data collection investments, and by implementing a shift in culture that fully embraces a new vision of the Service's scientific data:

1 Executive Office of the President, 9 May 2013, Open Data Policy – Managing Information as an Asset http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf; DOI open data plan, http://www.doi.gov/data/upload/DOIOpenDataFY14Plan.pdf

2 For example, the National Wildlife Refuge System's Inventory and Monitoring program is advancing in addressing their data management needs; Region 1 has developed an Information Management Strategy and provided training for a variety of audiences, and the Science Applications program has supported further development of the Environmental Conservation Online System (ecos.fws.gov/ecos/home.action).

Scientific data are 'trust resources' and the foundation of our future conservation actions

Service employees readily recognize and value trust species. It follows that the data we collect about these species and their habitats are also 'trust resources': data are the basis of our current and future understanding of the species we are charged to protect and are essential to our conservation actions. Just as Service employees dedicate their efforts to maintaining and sustaining the future viability of trust species and ecosystems, they must maintain and sustain our conservation data for the value the data hold for future resource managers, stakeholders, and scientists. Promoting this vision across the agency will improve our capacity to manage trust species and advance the Service's mission to work across organizations, jurisdictions, and disciplines to achieve conservation benefits. This positive message emphasizes the long-term value of every employee's current work, a value that can grow over time given appropriate data stewardship and curation.

The Strategy described in this document outlines actions that will initiate the needed cultural and behavioral shift. We identify current challenges to reaching the vision, and recommend a suite of strategic actions to overcome these challenges and achieve the vision. Given the dynamic nature of the problem, the recommendations are detailed to the degree necessary for the reader to comprehend the Subcommittee's intention but not to the point of prescriptiveness. Many of the recommendations address existing Executive and Secretarial Orders, Congressional guidance, and Departmental and Agency policies (Appendix C).

Better management of the Service's scientific data will not be achieved overnight, nor simply through creation of more policies and guidelines. Rather, it will be achieved by an appropriate blend of top down and bottom up efforts that converge to shift the workloads, behaviors, and attitudes of the project-level Service employees who both collect the majority of the Service's scientific data and are the natural stewards for ensuring those data are adequately documented and managed in the near-term. There is desire at the field level for this transition – but while many recognize the need for this shift, changes have not occurred, largely due to lack of the necessary technical staff, IT infrastructure, procedures, and processes. This fundamental transition at the field level is feasible if Programs embrace the full costs of a project – including documenting and maintaining data – and do things right (in addition to doing the right things). More importantly, while estimates are not available, we are convinced that a large portion of the implementation costs of this transition will be offset by increased productivity and efficiency at the project level, given sufficient commitment of resources and attention from the Directorate and supervisors.

The Challenges

Successful maintenance and delivery of information is complicated and requires specialized technical expertise, sustained commitment of resources, and proper organizational priorities,

behavior and accountability. For example, access to data necessitates *infrastructure* (server space and maintenance, bandwidth capacities, machine-to-machine communication standards and systems) and *policies* (data privacy, open data requirements, metadata documentation standards), and has *behavioral components* (adequate staff know-how and time devoted to data quality assurance/control processes, roles and responsibilities for data backup and publication). A key feature for the Service, or any organization where data collection arises from mission-oriented research rather than enterprise-wide business operations, is the distinction between data stewardship during the project phase and data curation over the long-term (Smith et al. 2011): the former must meet the immediate and flexible needs of the project, while the latter must provide for ongoing preservation and wide access. These are distinct needs and require distinct skills and capacities.

Figure 1 summarizes the data management lifecycle. Inherent in the lifecycle is the need for communication among the parties responsible for the different aspects of data management, especially during the transition from the project-phase to long-term curation, stages which are often handled by different groups. The examples in the introduction demonstrate that the Service's data management practices do not move our data competently and comprehensively through the full data lifecycle. Why not? A simple answer is 'because there isn't enough importance placed on data management.' However, the Subcommittee unpacked that simple answer and identified three underlying fundamental challenges: culture, organization, and resources, each of which must be overcome to achieve the vision where *Service data are recognized as 'trust resource's*.

Challenge 1: Cultures of the Service and the Scientific Community

The cultural traditions of the Service emphasize boots-on-the-ground data collection over other stages of the information production process – managing, analyzing, documenting and reporting information derived from field data. To a degree, this reflects an earlier period when conservation science needs were more proscribed in space and time, and more easily addressed by dedicated individuals adequately skilled in all phases of information production from design, through data collection, to reporting. But, the modern Service's conservation science needs are increasingly more complex in *scale*, *scope*, the *scrutiny* they receive, and the *evidentiary standards* against which they are held. For example, teams of collaborating specialists are essential to setting North American Waterfowl Management Plan goals, conducting Strategic Habitat Conservation, addressing priority science needs identified by the Landscape Conservation Cooperatives, or implementing the National Fish, Wildlife, and Plants Climate Adaptation Strategy. The shift to more team-oriented collaborations mirrors developments in other fields (e.g., medical practice in the U.S., Gawande 2011).

Technical specialists in database programming, data management, and information science play key supporting roles in these collaborations; collectively, these skills, along with statistical skills in study design and analysis, GIS skills in mapping and data visualization, and computer science skills in network design and administration, constitute the emerging field of *data science*. Unfortunately, academic training in traditional natural resource science and

management does not emphasize data documentation and management practices as part of the standard skills and responsibilities of the research scientist. This culture is slowly changing, as witnessed by the growth of data curation and discovery portals (e.g., Genbank[3] and Dryad[4]), the increasing recognition of data publications by top-tier journals and funding agencies (e.g., *Ecology*[5], the National Science Foundation[6]), and the expanding requirement for data archiving as part of the granting and publication processes (e.g., National Science Foundation[7], The Public Library of Science[8], and the Service's own *Journal of Fish & Wildlife Management*[9]) (Burke 2014). These shifts in academic training and research have yet to be uniformly incorporated into the culture of natural resource agencies. The Service will only achieve the full conservation value of its scientific data when it places as much value and emphasis on the data-science skills needed in the office as it places on the data-collection skills needed in the field.

Challenge 2: Organizational Structure

The Service's diverse and decentralized Programs make it difficult to achieve efficiencies of scale and impede the sharing of tools, workflows, and lessons learned, let alone technical capacities and skills. Further, while most data collection occurs at field stations, most data management resources and technical staff are housed in Regional offices. This leaves field staff with, at best, limited resources for addressing the myriad technical components of project data management (Figure 1). Besides creating greater need for local technical capacity, decentralization also creates inherent barriers to communication across stations and Programs, preventing rapid distribution of local innovations and thus slowing evolution of data management within the Service.

While the Division of Information Resources and Technology Management (IRTM) plays an important role in the technical and networking components of the Service's data management, the issues under discussion are not strictly within IRTM's domain. IRTM predominantly focuses on enterprise-oriented issues, such as network services and security, and is neither staffed nor charged with providing project-level data management support. Agencies such as the Bureau of Land Management (BLM), the National Parks Service (NPS), and U.S. Forest Service (USFS) have been successful with enterprise-wide implementation of certain data management efforts and workflows (pers. comm., Tom Chatfield BLM, 4 April 2013; Margaret Beer NPS, 7 March 2013; Gretchen Moisen USFS, Aug 2010), but those successes have focused on national scale programs lending themselves to standardization due to homogenous features (e.g., the Forest Service's nationwide Forest Inventory and Analysis Program[10]). Such programs

[3] http://www.ncbi.nlm.nih.gov/genbank/,
[4] http://datadryad.org/, accessed 25 Jan 2013
[5] http://esapubs.org/archive/instruct_d.htm, accessed 25 Jan 2013
[6] http://datapub.cdlib.org/2012/10/09/data-to-receive-recognition-from-nsf/, accessed 25 Jan 2013
[7] http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp, accessed 25 Jan 2013
[8] http://www.plosone.org/static/policies#sharing, accessed 25 Jan 2013
[9] http://www.fws.gov/science/guideforauthors.html#DataArchiving, accessed 27 Jan 2013
[10] http://www.fia.fs.fed.us/, accessed 27 Jan 2013

are exceptional cases in the Service.

A one-size-fits all approach is not a feasible solution to the Service's diverse, decentralized, project-level data management needs. Technical support – the modelers, database programmers, and data managers necessary to collaboratively support the field staff in their conservation science activities – must be local and originate from the Programs, field stations, and projects. The data management tasks associated with project cataloging, data discovery and sharing (e.g., long-term data curation), however, should be standardized and centralized as much as possible. This would help make the local solutions more robust to staff turnover while serving long-term data discovery and access needs at Regional and national levels. As the introductory examples illustrate, staff turnover and isolated data systems can result in severe loss of data and thus significantly impact the Service's ability to conduct business.

Challenge 3: Commitment and Resources

Achieving the vision requires commitment from Service management for a variety of initiatives, including:

- an effort to communicate the importance of data management throughout the Service,
- strategies to incentivize data management activities,
- investments in infrastructure and capacity (it is pointless to promote the vision without providing the capabilities to achieve it), and, finally,
- implementation of systems for accountability.

A successful data management transformation will touch all levels and most Programs, and require commitment from every Service employee, and their supervisors, who engage in science. Importantly, the transformation requires a shift in priorities by project managers and supervisors to recognize that data management (including documentation and curation) is part of the project, not an addition, and that project planning must dedicate adequate resources to completing these components, as well as the design, field activities, and reporting components. The breadth and depth of the necessary transformation can only be achieved with the commitment and engagement of the Directorate.

The most critical resources for any successful data management program are appropriately trained and qualified staff. The Service's shift toward activities requiring teams of specialists increases demand for technical capacity, a demand with which the Service's staffing allocation has not kept pace. The Service employs dozens of science staff to collect data (biologists, physical scientists, etc.) for every technical staff employed to support transforming data into information and making data accessible to and usable by others (data managers, data programmers, statisticians, biometricians, etc.; see Table 1). Although some progress has been made – the first Regional Refuge Biometrician was officially hired in 2003, a century after the founding of the Refuge system, and similar positions are now staffed in at least three regions – there is a long way still to go: a 2013 review of refuge staffing in Region 8 revealed that the only substantive change in capacity since the 1940s, when many of the units were established, has been the addition of biologist positions, predominantly in the 1970s and 1980s. The refuge

biologists on the 50 Region 8 refuge stations make up roughly 14% of total Region 8 refuge staff and conduct more than 700 surveys a year. Managing the data from these surveys is 'collateral duty' for the refuge staff, and there is only limited technical assistance to help manage the growing body of data, convert it into information, and make sure the data are documented and accessible for later use. Nor is such refuge support the primary duty of any of the handful of Regional staff to which refuge personnel turn (e.g., the five geospatial analysts or data managers in the Regional office). Further, there is no statistical help available for refuge staff for survey design or data analysis (Appendix B).

As the Promise Team found in 2000 (Appendix A), most current surveys do not have written protocols, adequate survey designs, or data management appropriate for making future use of the data. The disparity in supply of, and demand for, data service capacity limits efficiency of information production, delays achievement of the information's conservation value, and, at worst, wastes resources on collecting data that are never used to further our conservation mission, nor made available for future use. Re-allocating funds that currently support the collection of data that are never transformed into usable information (~$9 million in 2000 dollars for the Refuge system alone, Inventory & Monitoring Database Team, Fulfilling the Promise WH9.1, 2002, reproduced in Appendix A; Hal Laskowski, pers. comm. 2013) to provide adequate technical support would greatly improve the efficiency and impact of the Service's conservation science. The Refuge system is working towards this as part of the Inventory & Monitoring Program, but their efforts do not resolve the Service-wide problem.

Table 1. Staffing summaries for bureaus, mainly within DOI, showing the number of science staff generally engaged in data production (Biological and Physical science job series 0400 and 1300, excluding management positions) and the number of technical staff (Data Managers, Database Programmers, Statisticians, etc., job series 2210, 1529, 1530, and 1531). These results greatly overestimate the number of technical staff primarily focused on data management (one component in second column from right), since the relevant job series (Information Technolgists, 2210 job series) includes all general IT staff, enterprise architects, and other non-data management IT positions. The results also underestimates the available design and data analysis staff (italicized component in second column from right), as it does not include positions classified as quantitative ecologists, quantitative fisheries scientists, ecological modelers, etc., that are not in the 1529 or 1530 job series. Source: OPM Fedscope Dec 2012, doi: http://www.fedscope.opm.gov/ (accessed June 2013).

Agency	**Total Number of Employees**	**Science staff generally engaged in data production** Biological & Physical Scientists	**Data Scientists and IT Support Positions** Information Technologists (*Statisticians*)	**Ratio of producers to Data Scientists and IT Support**
National Park Service	21,193	2,277	~410 (*10*)	6:1
Environmental Protection Agency	17,720	4,857	733 (*95*)	7:1
Bureau of Land Management	9,764	3,203	357 (*0*)	9:1
U.S. Geological Survey	8,887	5,953	557 (*48*)	11:1
Forest Service	30,186	15,768	674 (*41*)	23:1
Fish & Wildlife Service	9,693	4,270	188 (*9*)	23:1

The Strategy: Directions and actions

Complicating the search for solutions to our data management problems is the fact that the three core challenges to effective data management detailed above – Service culture, organization, and resources – are not independent of one another and cannot be addressed separately. The issue of inadequate connectivity at field stations is an illustrative example:

> Modern data management is supported by an information infrastructure built on machine-to-machine communication. Such architecture is central to implementation of the Executive Office's Open Data policy and must be the basis of our Regional and national data management, given the Service's organizational structure. But in many places Service IT infrastructure is inadequate: the average internet connection speed from a Region 8 refuge office, even urban offices, is ten-times slower than the average U.S. speed[11]; this is a substantial cost to Service productivity.
>
> **Challenges**: Service culture must recognize the importance to our long-term mission of a smoothly functioning information infrastructure, flagging and addressing such problems before they become crises. Diagnosing and resolving issues with the Service's distributed information infrastructure requires technical capacity be available at the station level. Maintaining this infrastructure needs to be part of standard operations.

Approaches to improving data management must address elements of all three challenges. The Subcommittee's recommended actions for addressing the Service's data management deficit are organized into four strategic directions, which can be broadly described as buy-in, policies and systems, personnel, and funding. As with the three challenges, these four categories of action are not independent of one another but highlight different aspects of our data management problems and their solutions. We detail key initial actions under each direction; additional recommended actions are listed in Table 2.

Strategic direction 1: Develop awareness and support for all stages of the data lifecycle at all levels of the Service

Successful models for intelligent, effective, and secure data management share a number of components (Mosley 2008, Brown and Henstorf 2014 – reproduced in Appendix D). First, they typically use a data lifecycle framework to structure tasks and track accomplishments within their highly complex organizations (CEOS WGISS 2011). Second, they rely on an effective leader with the authority to promote and implement the lifecycle model and oversee adherence to its tenets and requirements. And, third, they require the organization as a whole to actively engage with and support data management infrastructure and activities. The subcommittee has adopted a data lifecycle model used by the Bureau of Land Management but modified to fit the

[11]Refuge offices averaged internet connection speeds of < 1 Megabyte per second (Appendix C, Table 1), while average U.S. connection speeds are almost ten times faster (8.6 Mb/s, http://mashable.com/2013/08/22/fastest-internet-world/, accessed 5 Feb 2014).

Service's particular needs and capacities (Figure 1). This is being adopted by Region 1 in its Regional Information Management System training program.

Action 1.1: Create a home and voice for data science[12]

Our primary recommendation is that the Service establish a **Data Science Advisor** within the office of the AD for Science Applications. The Data Science Advisor would be provided with dedicated resources and authority to oversee and promote coordination of data management within the agency's many science programs, and to integrate the Service's data assets with agencies and organizations in the Department, the Federal Government, and our State, Tribal, and non-governmental partnerships. The position would be responsible for fostering the cultural change described above, in part by promoting, incentivizing, and tracking adherence to the various data policies applicable to the Service's scientific endeavors.

This recommendation resolves two fundamental challenges to improving data management within the Service: (i) bridging the gap between enterprise-oriented IT and the reality of the Service's widely diverse scientific data collection and use efforts, and (ii) bringing sustained attention and voice to the technical complexity of modern science-based Service efforts. The position will serve as a communication link and translator between IT and the Service science programs. The Data Science Advisor would work across the Service's Programs to identify and help address the field-level needs for core technical services and guidance, as well as to aid collaboration among those creating the Service's scientific data and those charged with managing and curating them.

The Data Science Advisor would be aided by a Service-wide **Data Science Committee**. This standing committee would be composed of individuals with data management and program management experience from across the Service's Programs and Regions. The Data Science Advisor and Data Science Committee would design and help promote an institutional program of high quality, repeatable, enforceable, and rewardable management practices for the Service's scientific data in support of the Service's conservation actions. Initially, the committee would be charged with assisting the Data Science Advisor in implementing the recommendations of this Strategy. Subsequently, the committee would assist the Data Science Advisor in data management promotion, assessment, and governance.

Once established, the Data Science Committee would identify a group of data management champions, or early adopters, throughout the Service to aid in communication and outreach regarding tools and resources, development of case studies and needs assessments, and related activities. These data champions would naturally organize into a Data Science Community of Practice, a formal forum promoting communication among those actively engaged in data management, the sharing of lessons learned, and codification of best practices (Wenger et al. 2002). Organizers of the Region 1 Regional Information Management Strategy,

[12] As noted previously, *Data Science* embraces the skills, methods and tools of both managing and analyzing data. This terminology captures the broad, long-term technical capacity needs of the Service, as data management is both a major element and a fundamental first component in any process of data analysis and use.

Region 8 and national NWRS I&M data managers, the Subcommittee authoring this Strategy, and the LCC Network Data Management Working Group currently interact informally and would logically form the nucleus of this group.

Eventually, attention must be focused on effective organization of technical staff. In addition to responsiveness to local needs, other key considerations include promoting cross-program, cross-level communication among the technical staff for sharing of technical solutions; promotion of standardized data management tools, systems, and workflows; and adherence to all components of the lifecycle. Re-organization could magnify the currently isolated voices of technical staff with respect to identifying barriers to efficient and effective data management and allow for rapid evolution of improvements in our data management, thus benefiting the Service's science activities. Key technical staff might be combined into a Center for Data Science with supervision through the Data Science Advisor, a successful approach in many institutions with scarce-yet-essential data science assets (Brown and Henstorf 2014; reproduced in Appendix D). Effective organization to foster and sustain cross-program, cross-level communication among the technical staff, however, must remain responsive to local needs.

Action 1.2: Develop a communication and training plan

A critical step in changing how we manage data entails educating and informing all staff associated with the Service's scientific efforts about the central role that data management plays in achieving the Service's mission. Both data management and individual employees' interactions with data are complicated and diverse, so an effective communication plan must be multi-faceted and positive. The communication plan will need to serve several distinct audiences. The Directorate will benefit from information on shared data management among the many Service partners and conceptual approaches to information science and management. Managers will benefit from improved understanding of how and where data management efficiencies translate into staffing and workflow planning, while supervisors will benefit from assistance in overseeing the performance plan adjustments needed to prioritize and conduct data management and in tracking data management-related performance metrics. Field staff will benefit from training that demonstrates both tangible examples of how quality data management enhances their contribution to conservation, as well as the available tools, resources, workflows, and assistance for implementing data management systems. The specialized needs of technical staff were described above.

The Subcommittee has established some groundwork for effective communication with the development of a Service-wide SharePoint site[13] and ListServ[14], but does not have the specialized knowledge necessary to craft a comprehensive communications plan. The Data Science Committee should be tasked with drafting an initial communications plan (henceforth, the Strategic Communications Plan). This action should be undertaken after completion of *Action 2.1*, described below. We also suggest four sub-actions related to communication,

[13] https://connect.doi.gov/fws/Portal/fwsdm/SitePages/Home.aspx
[14] fws-data@lists.fws.gov

training, and support.

Sub-action: Develop an all-employee video message from the Director that emphasizes the vision: our scientific data assets are an important part of our conservation legacy ("Data are 'trust resources' "). The message should include an introduction to the Service-adopted data management lifecycle.

Sub-action: Incorporate key messages on the importance of management of the Service's scientific data into NCTC's core programs (i.e., Foundations, Refuge Academy, Project Leaders Academy, ALDP, SUTL) and to the scientific integrity/code-of-conduct training. The ultimate goal is for data management to be a regular component of the communications and training materials that reach employees, from very early in their employment and then throughout their Service careers. One key message is the importance of project implementation plans, including data management planning elements.

Sub-action: Develop awards specifically to reward excellence in data management. This could range from promoting targeted Performance awards to emphasizing that data management contributions qualify for the current Rachel Carson Award for Scientific Excellence and Science Leadership Awards, or considering a new Service-wide or even DOI-wide data management award.

Strategic direction 2: Create an implementation plan to advance policies, procedures, performance measures, and technical infrastructure

Following their creation and staffing, the Data Science Advisor and standing Committee should undertake the series of initial activities listed in the remainder of this report, balancing 'planning while doing.' Some actions require more formal and in-depth scoping of needs and development of solutions (e.g., balancing workforce); others are ready for action, though requiring development of thoughtful implementation plans (e.g., promoting awareness of Open Data policies and tools); still others simply require broader Service-wide communication of activities, tools and resources already developed and implemented in specific Programs or Regions (e.g., Region 8's contracting language for data products that conform to the various data policies). We describe only the first three priority actions under this Direction; others are listed in Table 2. The Data Science Advisor and Committee should keep in mind the paramount goal is increasing quality of data management across the Service through improvements in efficiencies, standardization of data management processes, and access to technical resources.

Action 2.1: Initiate an evaluation of data management practices and needs

Data management experts within and outside the Service should be brought together to conduct an assessment of the Service's current data management resources, deficits, and priorities (see NRC 2007 for an example of such an external review at NOAA). This assessment should identify critical scientific data collection activities, then evaluate:

- current policies and their implementation rate, including metadata standards and project

metadata publishing;

- data management-relevant IT infrastructure;
- staffing resources, enterprise-wide data management plans;
- existence of and adherence to formal data management plans;
- data curation; and,
- Program budget allocations to data management.

The assessment should particularly address the distribution of technical staff and expertise across Service Programs. Other land or resource management and conservation agencies (e.g., NPS, BLM, USGS, NOAA/NWS, NOAA/NMFS, EPA, US Forest Service) should be surveyed to better understand their data management staffing, and their funding of data management work. This evaluation (henceforth, the Strategic Evaluation) should be focused at the Program level. USGS is in the final stages of such as assessment, which may serve as a model (Mike Frame, personal communication to Joel Reynolds, 10 March 2014).

Action 2.2: Identify core Service data and devise plans to safeguard

In a period of scarce resources, new initiatives need to prioritize effort and attention. Data management reforms should begin with the most critical and essential Service data. Thus, the new Data Science Advisor should survey Service Programs to identify core data sets, starting with those data products identified by Programs as being *influential* or *highly influential* under the Information Quality Act[15], then moving to core long-term monitoring programs, programs collecting data on priority resources, and finally to other programs based on demonstrated importance to Service and/or partner decision making, taking into consideration each project's data management needs.

The current data management practices for each core data set should be evaluated for the features identified in *Action 2.1*, though in detail appropriate to project-level stewardship and curation. This exercise would serve as a template for the development of a data management 'tune-up' process that would aid in establishing robust and efficient data management systems. Ideally, the initial tune-up team would consist of Service or partner staff identified by the Data Science Advisor and Data Science Committee as having both the necessary skills and the willingness to volunteer on a brief review detail. Later iterations could be coordinated with training programs, thus serving as an apprenticeship program for data scientists in training. Tuned-up data systems could serve, in turn, as data management case studies for training materials.

Action 2.3: Develop a process for curating non-core data and rescuing historic data

Not all scientific data collected by the Service, now or in the past, warrants preservation and long-term curation. The *Action 2.2* prioritization will not capture a large portion of the Service's current data collection efforts, and only touches some historic data. A triage process must be

[15] Section 5 15 of the "Treasury and Consolidated Agencies Appropriations Act of 200l"

developed for data sets identified as worth the investment of long-term curation. This should be an on-going process with input from the broader conservation community, should be driven by considerations of societal benefits, and should explicitly involve coordination with other agencies. The triage development would include reviewing similar investment strategies established in sister agencies (e.g., NOAA - see Principle #7 in NRC 2007; US Forest Service FIA Program; NPS Vital Signs; BLM).

Strategic direction 3: Increase capacity and balance the workforce

A recent commencement address at Harvard Medical School described the large gap between the U.S. medical community's 'lone cowboy' model, where an individual doctor is expected to have all the skills, knowledge and abilities to provide complete health care, and the more appropriate 'pit crew of specialists' model that is actually required to address effectively the complexity of modern health care (Gawande 2011). The description applies to modern natural resource management and conservation, where our increased understanding of natural systems combines with increasingly complex management decisions, yet we continue to support the 'lone cowboy' field biologist, who is expected to know and do everything. Delivering the highest quality science and natural resource management requires we embrace a more collaborative approach built upon 'pit crews' of specialists.

Good data management requires staff skilled in the technical aspects of data science as well as staff skilled in the natural resource sciences the data represent, i.e., teams with both data collectors and data managers. Meeting these needs requires a re-balancing of the Service workforce to provide the necessary technical specialist capacity (Table 1). This balancing will require creation of new positions for future hires, modification and revision of performance measures for existing positions, and training for current staff (Actions 3.2-3.4). To initiate the rebalancing of staff and the development of related capacity, we recommend the immediate establishment of funding for ready-to-go project-level data management needs, as described in *Action 3.1*.

Action 3.1: Award three-year term funding to develop project-level capacity

To initiate solutions to the myriad local data management needs while the strategic evaluation (*Action 2.1*) and its follow-on activities are underway, we propose an annual process to award three years of funding to a project or station to that is ready to bring its data management process up to speed, but is lacking the necessary capacity. Funds would be dedicated for one new award each year. The Data Science Advisor and Committee will develop the competition, including the proposal format and evaluation criteria. The evaluation criteria should include consideration of the transferability of the project's products and lessons to other Service data management needs. The awards would be equivalent to what is needed to fund a GS-11/12 Step 6 data manager, but the competition should be structured so that recipients can apply the funds for whatever activities are most strategic for developing capacity and supporting staff: term positions, contract services, hardware, software, training, etc. One goal of the competition would be that Programs leverage

the one-time investment to transition to self-supported long-term solutions.

Action 3.2: Create new data management positions

The Service needs staff with modern data management skillsets, and must identify and develop new positions to support new technical needs. In many cases, effective stewardship of natural resource data requires both technical skills in data science and an understanding of the biological system that the data represent. Such interdisciplinary knowledge helps overcome communication barriers and will transform Service culture, while providing the core capabilities needed to produce information from data and provide for long-term access to data. Using the results of the Strategic Evaluation, the Data Science Advisor and Committee, in coordination with Human Resources, should review and modify existing job classifications, as necessary. Such a review might well recommend the creation of new job series to meet the capacity gaps identified by the evaluation. For example, the sole 400-series position 'requiring' statistics or mathematics, the research Wildlife Biologist, can be satisfied with coursework in chemistry, physics, soils, or geology, which represent distinctly different skill sets. Efforts to define new job positions should also consider employee retention and explore the construction of a career track to ensure that the technical talent the Service cultivates continues to grow and remains with the agency.

Action 3.3: Modify existing positions and/or performance standards to address data management duties

Using the results of the Strategic Evaluation, the Data Science Advisor and Committee should work with the Programs responsible for the core Service data (identified in *Action 2.2*) to ensure that all stages of the lifecycle receive adequate accountability and are clearly incorporated into the performance standards or, if necessary, the position descriptions of the relevant staff. This effort should emphasis a 'many hands make small work' approach that redirects a small but critical portion of each employee's time to specific data management tasks. These changes will only be successful if accompanied by accountability measures in the performance standards of higher level supervisors and managers.

Action 3.4: Improve training opportunities

System managers, technical, and scientific staff need access to cutting-edge best data practices so they can guide the Programs they serve and support conservation decision-making. Data management training and communication should be coordinated with other Interior bureaus, and within Service Programs. Regions 1 and 8 have recently collaborated with technical staff from the BLM to provide training in the data management life cycle and policies[16]. Additional effort should be directed to: review and promote existing training materials from sister bureaus (e.g.,

[16] All materials are available (for those within DOI) at https://drive.google.com/a/fws.gov/?tab=co#folders/0B4en1VzizsQtT1pFRFYOZVIwZOU

USGS[17], BLM[18]); create Service training modules on the data lifecycle and practical vision; developing data management plans, metadata authoring and tools, open data policies and tools, and other technical aspects of data management; schedule training webinars; adopt a user-friendly data management handbook; incorporate data management into future GIS Workshops at NCTC; and expand NCTC's data-related offerings.

Strategic direction 4: Budget for data management

Data management is part of every science endeavor; it is not an activity that one engages in as time and resources allow. An often stated rule-of-thumb is that up to 30% of a project's budget should be allocated to data management and analysis. The Strategic Evaluation will provide additional insight relative to this rule, by surveying data management budget allocations at other natural resources agencies. Until data management and analysis support is part of Service culture, the Programs need to target funds to bring their data management systems and capacities to needed levels (*Actions 3.1-3.4*). Once *Actions 4.1* and *4.2,* described below, have advanced far enough, they will provide the basis for developing a comprehensive data management budget initiative. Ultimately, long-term base funding will need to be identified for (i) Service Programs that host long-term data curation sites, such as ECOS - the Environmental Conservation Online System, as well as (ii) maintenance and upkeep of fundamental data management resources, such as hardware upgrades and network connectivity sufficient to support local access and use of Regional, national or Program-specific integrated data systems, etc. These costs must be recognized as Operation & Maintenance costs, just like any other aspect of the Service's extensive physical infrastructure assets.

Action 4.1: Establish a funding policy for national, Regional and field station support of project data management and analysis

The results of the Strategic Evaluation will provide a basis for developing a funding strategy that meets the needs for personnel capacity, technology, and operating expenses. The strategy should consider incorporating infrastructure and hardware costs into station operating costs, reflecting their fundamental role in our 'conservation science' business operations.

Action 4.2: Establish a Service-wide funding strategy for long-term curation of core data

In addition to addressing staffing and technical capacity needs (*Strategic Direction 3*), the Data Science Advisor and Committee will, in discussion with all Service Programs, develop a long-term funding strategy for supporting curation of the Service's core data sets (identified in *Action 2.2*). The strategy should aim for dedicated funding, if the solutions are to serve as the backbone of the Service's long-term data management structure. The strategy should review existing non-

[17] http://www.usgs.gov/datamanagement/index.php

[18] http://www.blm.gov/pgdata/etc/medialib/blm/wo/Information_Resources_Management/policy/blm_handbook.Par.17422.File.dat/h1283-1.pdf

enterprise-wide efforts (e.g., ECOS: the Environmental Conservation Online System; ServCat: the Service Catalog), as well as those systems developed or developing as part of the recent Open Data policies (www.data.gov, dataone.org), the LCCs (LCMAP[19]), or partners (databasin.org, the Avian Knowledge Network, etc.). The strategy should identify a limited menu of approaches covering the majority of the Service's core data curation needs. The strategy should also summarize current Service efforts and resources allocated to curation and utilization, if known, of the core data sets.

The Future

The data crisis is here. Business as usual is not working and we are wasting money on data collection with insufficient planning and resources directed at managing and using our data, let alone making it available for other users. In an ideal world, we would stop counting, measuring, and recording long enough to get our data house in order: clean-up, document, evaluate what we have, set up infrastructure, and plan for what we need to move forward. In the world in which we operate, we must dedicate ourselves to finding time at the margins. Across the Service, supervisors must commit to freeing up time and resources for staff to focus on data recovery and management, infrastructure assessment, and workflow planning. Ignoring these tasks today creates more crises and inefficiencies tomorrow. The Region 8 Inventory and Monitoring team recently demonstrated the value of setting aside such time by holding its first "Data Management Strategic Pause" on Celebrate GIS Day. Seven staff spent the day addressing issues of server file management with the task of identifying project folders for archiving, deletion, or restructuring. This one day pause reduced duplication (49 file directories from 66), archived completed projects, reorganized and consolidated project folders, removed outdated material, and synchronized isolated project files back to the appropriate server folders. These day-to-day actions are pushed aside by regular crises and urgent deadlines, but with regular attention, these activities make for more efficient operations, including more efficient project management and thus more effective decision-making, conservation delivery, and beneficial conservation outcomes. Instituting *data management plans* is a pro-active step that would eliminate many of these issues in the future. Such activities, distributed across Service Programs, improve the quality and defensibility of the information we use to make management decisions, consequently strengthening our conservation actions and improving future operations. Strategic pauses are also a means of tackling the enormous triage problem associated with rescuing and curating the backlog of Service-collected data that will otherwise disappear. (*Action 2.3: Develop a process for curating non-core data and rescuing historic data*).

All the actions recommended in this Strategy (Table 2) aim at the same goal: transforming Service culture and capacity so that management of our scientific data becomes a

[19] https://www.sciencebase.gov/catalog/?community=LC+MAP+-+Landscape+Conservation+Management+and+Analysis+Portal

standard part of everyday operations. Implementing these actions requires resources in a time of shrinking budgets and overtaxed staff. Yet so does ‘no action’ – in the myriad inefficiencies that reduce our conservation effectiveness and ultimately stem from inadequate attention to, and capacity for, data management. These long-term costs to the Service and our conservation partners greatly exceed the time and resource costs of achieving our data management vision. Such a cultural transformation will ensure that we automatically ‘do the job right,’ from planning through management action, and prevent current crises from undermining our ability to secure the long-term value of our conservation science investments.

Initial Actions

1) **Action 1.1**: Initiate Creation of Data Science Advisor and USFWS Data Science Committee; identify data management champions at all levels.

   A) Gain Directorate's commitment to position; determine location and level.

   B) Draft role and responsibilities, job description, and initial tasks.

   C) Draft charter for Committee; establish Committee membership.

   D) Poll sub-committee members, OSA, and Science Committee to identify at least one data management champion from each Region and Program; Ask Regional GIS, Data management teams, and Refuge I&M for names.

OUTCOME. Oversight for management of scientific data in the USFWS. Improved cross-program implementation of data management policies, communication and coordination regarding solutions and resources, and mechanisms for identification and prioritization of shared Service-wide needs regarding data management life. Core community to advance initiatives.

2) **Action 2.1**: Strategic Evaluation. Initiate assessment of Service's data management practices as a joint effort of new Data Science Committee, OSA, IRTM, and external experts.

   A) Develop proposal outlining assessment process, specific questions, information needed and available.

   B) Analyze existing data management landscape and produce report.

   C) Basis for guidelines on effective implementation of the data lifecycle.

OUTCOME. A knowledge-based plan for change: Assessment of the scope of data management in the Service. Recommendations on resources needed, included equipment and staffing; baseline for assessing progress.

3) **Action 3.1**: Award process to build data management capacity.

   A) Data Science Advisor and Committee develop a process to award funds; evaluation to include consideration of the impact of proposed activities and the transferability of products and lessons learned.

   B) Awards run for three years, one award initiated each year.

   C) Programs are encouraged to extend positions and activities beyond the three-year award period, leveraging this initial assistance into long-term solutions.

   D) Projects form basis for case studies and transferable solutions.

OUTCOME. Immediate action on pressing, well-defined data management needs; balance 'planning' activities (*Action 2.1*) with 'doing' activities aimed at the battle lines where the data are being collected; seed local capacity and solutions for project-level data management.

Table 2: A comprehensive list of recommended Key Actions with ratings of each action's potential impact and feasibility. The recommended initial actions are indicated by bold font.

Definitions: **Potential Impact** (Will the action lead to desired changes in scientific data management?). High–The action is likely to help mitigate a barrier or improve scientific data management; Medium–The action could possibly help mitigate a barrier or improve scientific data management; Low–The action will probably not contribute to meaningful barrier mitigation or scientific data management improvement.
Feasibility (Is it possible to implement the action within reasonable time, financial, staffing, ethical, and other constraints?) Very High–The action is ethically, technically, AND financially feasible; High–The action is ethically and technically feasible, but may require some additional financial resources; Medium–The action is ethically feasible, but either technically OR financially difficult without substantial additional resources; Low–The action is not ethically, technically, OR financially feasible.

Key Action	**Impact**	**Feasibility**
Strategic Direction 1: Develop awareness and support for all stages of the data lifecycle at all levels of the Service		
Action 1.1: *Create a home for data science*		
Create Data Science Advisor.	**High**	**High**
Staff a Data Science Committee.	**High**	**Very High**
Identify data management champions to carry the DSA message, e.g., NWRS I&M Data Managers.	**High**	**Very High**
Develop the DSA Program. Reassign select staff to report directly to DSA, while continuing to have them located in Regional/Program offices to support local needs.	High	Medium
Action 1.2: *Develop a communication and training plan*		
Data Science Committee contracts with communications specialist and develops a Strategic Communications Plan.	High	High
Develop website for access to data management guidance and resources, as well as a forum for communication and collaboration.	High	High
Create and distribute an all-employee message focused on data management.	Medium	Very High
Incorporate data management messages into core Service Training programs (Foundations, Stepping Up To Leadership, ALDP) and annual science integrity training.	High	Very High
Establish incentives program, including data management performance awards and possible Service-wide Excellence in Data Management recognition.	Medium	Very High
Developing metrics for success relating to effective communication (e.g., quantity of data shared via data.gov, number Service employees participating in data management training, number of Request for Proposal processes with explicit data management requirements, use of curated data sets).	High	Very High
Promoting avenues for data sharing through data portals, data publications.	High	High

Strategic Direction 2: Create an implementation strategy to advance policies, procedures, performance measures, and technical infrastructure		
Action 2.1: *Initiate an evaluation of data management practices and needs*		
Convene sub-committee of Data Science Committee to conduct evaluation, include representatives from OSA, IRTM, and external experts.	**High**	**Very High**
Survey science/resource agencies and Service Programs to determine (i) staffing numbers for data producers, managers, and analysts [refine Table 1], (ii) project and Program level budget allocations for data management (iii) performance measures for quality of data management activities, and to (iv) evaluate associated quality of data management strategies, and (v) summarize lessons learned. (Evaluation should distinguish stewardship activities from curation.)	**Medium**	**High**
From survey, develop recommendations on strategies, staffing and budget allocations, and performance metrics for a high quality, effective data management approaches.	**High**	**High**
Using the information from the survey, set targets for Regions and Programs for hiring teams of specialists.	High	Medium
Evaluate existing Program organizational structure and budget allocations using the recommendations developed above; identify positions that should be refocused, or added, to address data management and analysis needs.	High	Medium
Action 2.2: *Identify core Service data and devise plans to safeguard*		
Compile list of Service's influential information.	Medium	Very high
Survey Programs to identify core datasets.	High	High
Evaluate management of each core dataset, standardize and update as needed.	High	Medium
Develop data management "tune-up" process.	High	Medium
Action 2.3: *Develop a process for curating non-core data and rescuing historic data*		
Compile and review approaches employed by sister bureaus and Service Programs to determine which data to manage for long-term use, include evaluations of these approaches by interested user groups.	High	Medium
Review 'open data' policy guidance, including requirements for use of specific data cataloging and/or long-term curation sites (e.g., data.gov).	Medium	Very High
Convene a Service working group to share approaches to triage and data rescue; develop common criteria, strategies, and best practices; identify means for sharing resources, processes, and workflows.	High	High

Strategic Direction 3: Balance the Workforce		
Action 3.1: *Institute an award process to speed development of local capacity.*		
Define competition guidelines (schedule, evaluation criteria and process, administration procedures)	**High**	**Very High**
In-reach and promotion of competition	**High**	**Very High**
Hold first award	**High**	**Very High**
Develop case study & transferable lessons learned	**High**	**Very High**
Action 3.2: *Create new data management positions that combine technical and biological skills*		
DSA and Committee, jointly with HR, create new data management and data analyst positions descriptions that cross job series and families.	High	High
Develop a career track within the Service science and technical experts in order to promote retention.	High	High
Provide Regions with assistance in developing technical specialist job descriptions and support in interviews and review of applications.	High	Med/Low
Action 3.3: *Modify existing positions and/or performance standards to address data management duties*		
Incorporate data management tasks into existing job series as appropriate (e.g., a field biologist would be required to complete data and metadata entry, follow standard protocols etc. but would not be required to build a data management system for storage and analysis of monitoring data).	High	High
Incorporate feasible, meaningful, and SMART data management objectives into relevant EPAPs and performance plans from field staff to SES staff.	High	High
Action 3.4: *Improve training opportunities for current positions*		
Review training materials from sister agencies.	Medium	Very High
Develop data management case studies.	Medium	Very High
Create training modules for data lifecycle and vision.	High	High
Create data management handbook.	High	High
Develop web and NCTC training on data management.	High	Medium

Strategic Direction 4: Budget for data management		
Action 4.1: *Establish a funding policy to assist national, Region and field station managers support project data management and analysis*		
Develop cost guidance for project-level and long-term maintenance costs.	High	High
Coordinate with DOI's IT Transformation and look for efficiencies.	High	Medium
Establish baseline funding for bandwidth	High	Medium
Direct that project/Program budgets include data management costs (short- & long-term); When data systems are developed, budgets must include long-term maintenance in addition to funds to build the system.	High	Medium
Make project funding contingent on acceptable data management plan with a long-term maintenance strategy (requires a number of significant sub-actions).	Medium	Med/Low
Action 4.2: *Establish Service-wide funding strategy for long-term curation of core data*		
Review current Service strategies and associated costs, as well as overall resource allocations to long-term curation.	High	Medium
Review existing non-enterprise-wide curation systems, as well as systems developed or developing as part of the recent Open Data policies. Summarize their associated benefits and costs and any barriers to use for long-term curation of Service core data.	High	High
Identify a limited menu of approaches and systems addressing the majority of the Service's core data curation needs and develop communication plan to promote use of these sites.	High	High

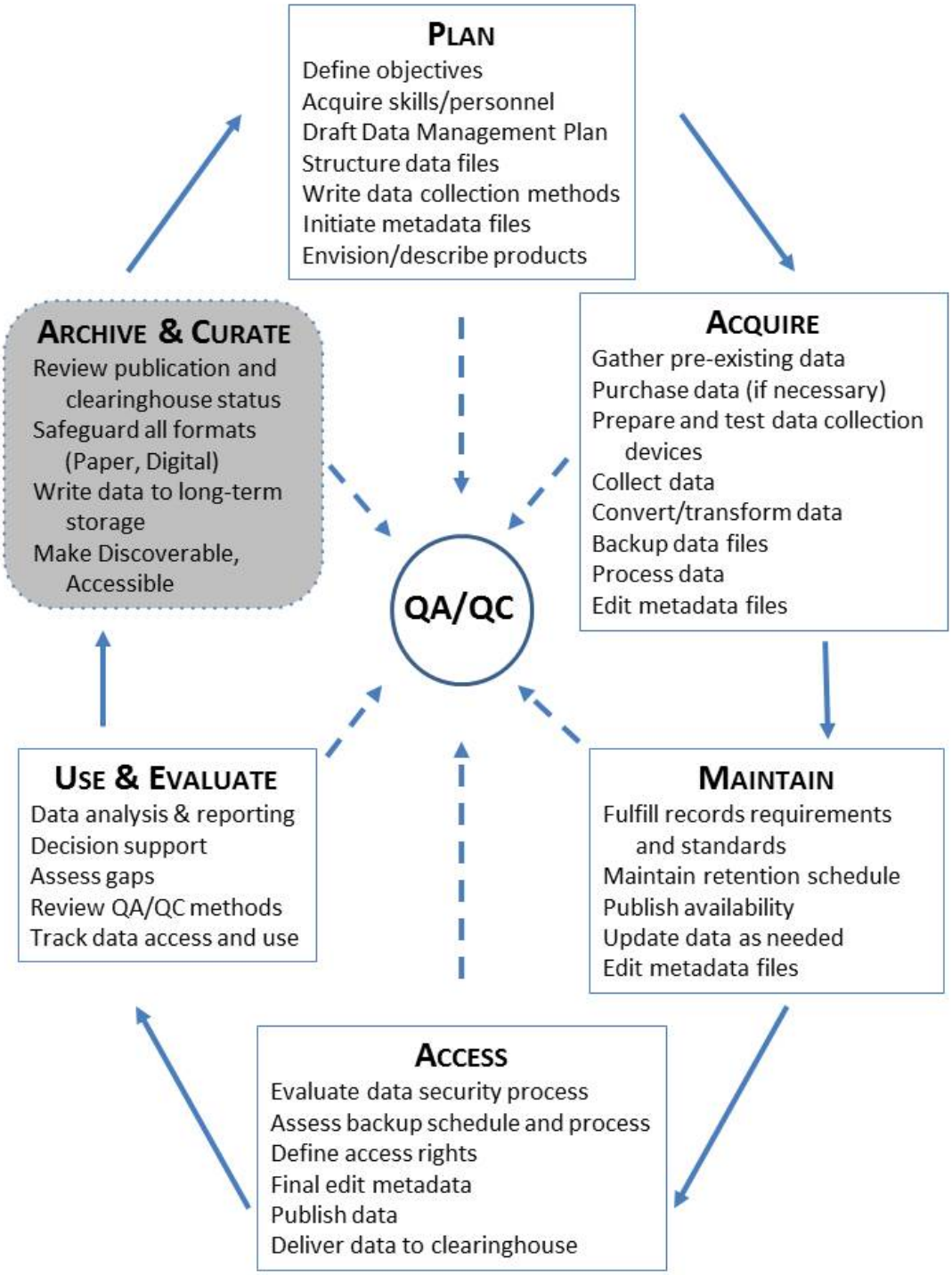


Figure 1. The Data Lifecycle (adapted from Bureau of Land Management 2006). All major stages of the lifecycle involve quality assurance and control activities (QA/QC). *Data stewardship* encompasses all unshaded components and is usually the responsibility of those who collect the data. *Data curation* encompasses the shaded component and addresses the demands of long-term preservation and access, likely by staff that did not design or collected the data. *Metadata* are the information accompanying a dataset that document and describe the data so that others can use and interpret the data. (Definitions adapted from NRC 2007.)

References

Brown, B., Henstorf, B. 17 Jan 2014. Making the most of scarce data-mining talent. Harvard Business Review Blog Network. http://blogs.hbr.org/2014/01/make-the-most-of-scarce-data-mining-talent/, accessed 19 Jan 2014. Reproduced in Appendix D.

Bureau of Land Management (2006). Data Administration and Management Handbook, H-1238. Bureau of Land Management handbook to accompany manual section 1238. http://www.blm.gov/wo/st/en/info/regulations/Instruction_Memos_and_Bulletins/blm_handbooks.html (accessed 14 Feb 2014).

Burke KL (2014) A Safety Net for Scientific Data. American Scientist 102(1): 6 doi:10.1511/2014.106.6

CEOS WGISS 2011. Data Life Cycle Models and Concepts. Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems and Services (WGISS) Data Stewardship Interest Group Technical Note 01, 26 Sept 2011. http://www.ceos.org/index.php?option=com_content&view=category&layout=blog&id=137&Itemid=218, accessed Jan 2014.

Gawande, A (2011) Cowboys and Pit Crews. The New Yorker, 26 May 2011. http://www.newyorker.com/online/blogs/newsdesk/2011/05/atul-gawande-harvard-medical-school-commencement-address.html

Gibney, E, Van Noorden, R (2013) Scientists losing data at a rapid rate. Nature, 19 Dec 2013, doi:10.1038/nature.2013.14416.

Mosley, M. 2008. The Data Management Association- Data Management Book of Knowledge, functional framework, Version 3.02. DAMA International, http://www.dama.org/i4a/pages/index.cfm?pageid=3345, accessed 22 Jan 2014.

NRC 2007. Environmental Data Management at NOAA: Archiving, Stewardship, and Access. Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA, National Research Council. http://www.nap.edu/catalog/12017.html

Smith, S., Tessler, S., McHale, M. 2011. A Data Management Plan Framework. USGS Community for Data Integration, Data Management Working Group. https://my.usgs.gov/confluence/display/cdi/DMP+Framework%2C+Part+II%2C+Data+Model+Standards, accessed 17 Jan 2014.

Vines, TH, Albert, AYK, Andrew, RL, De'barre, F, Bock, DG, Franklin, MT, Gilbert, KJ, Moore, J-S, Renaut, S, Rennison, DJ. 2013. The Availability of Research Data Declines Rapidly with Article Age. Current Biology, Available online 19 December 2013. http://dx.doi.org/10.1016/j.cub.2013.11.014

Wenger, E., McDermott, R., Snyder, W. 2002. Cultivating communities of practice: a guide to managing knowledge. Harvard Business Press, Boston, MA.

Appendix A: Report of Fulfilling the Promise, Inventory and Monitoring Database Team (WH 9.1)

Findings of a Questionnaire Regarding Monitoring Activities on National Wildlife Refuges Throughout the National Wildlife Refuge System

Prepared by:

Inventory and Monitoring Database Team
Fulfilling the Promise (WH 9.1)

January 2002

Introduction

A critical element of adaptive management on national wildlife refuges is biotic and abiotic monitoring to determine whether wildlife and/or habitat management actions have met refuge-priority goals and objectives. In order for monitoring data to be analyzed and assessed relative to refuge objectives, the data first must be stored and managed in a way to facilitate its use and interpretation. Although many refuges collect monitoring data, there is no standardized approach for storage/management of these data. Consequently, there is disparity regarding the application and utilization of these data relative to management decisions on refuges throughout the National Wildlife Refuge System. In some cases, monitoring data are collected and then simply archived in file drawers without any analyses or interpretation of results. In contrast, some refuges store the information in databases, but there is no post collection data processing. Other refuges utilize monitoring by storing/managing data in database software that facilitates data analyses and subsequent interpretation of results to assess whether management actions meet refuge goals and objectives. In accordance with 701 FW 2.1, monitoring data are to be stored and managed in a database to facilitate their effective use with regard to management decisions on refuges.

Questionnaire

A 3-page questionnaire (Appendix A) was distributed to all project leaders/refuge managers within the National Wildlife Refuge System during fall 2000 regarding biotic and abiotic monitoring activities (surveys) conducted on refuges to guide adaptive management. One questionnaire was sent to each complex, but the staffs at each satellite refuge were requested to complete it for their respective unit(s). The questionnaire included a series of 7 questions regarding software and hardware used to store data associated with monitoring activities on refuges. Specific information was requested regarding storage, management, analyses, and use of data for each monitoring activity conducted on a refuge. Respondents also were requested to prioritize the top 5 monitoring activities for their refuge(s).

Information regarding the management/storage of data for 1977 monitoring activities was reported on questionnaires received from 219 refuges throughout the National Wildlife Refuge System (Table 1). Monitoring activities on refuges involved birds (54% [1063/1977]), habitat (15% [310/1977]), mammals (11% [217/1977]), amphibians and reptiles (7% [134/1977]), abiotic (environmental) conditions (6% [112/1977]), and invertebrates and fish (3% [69/1977]). The following are the 15 most frequently conducted monitoring activities on refuges reported on the questionnaire: birds - waterfowl counts, breeding landbirds, bald eagles, waterbirds, shorebirds, marsh and wading birds, waterfowl production, Christmas bird count, and grassland birds; habitat - vegetative community, invasive plants, and moist-soil vegetation; mammals - big game; amphibians and reptiles - breeding frogs and toads; and abiotic monitoring - water gauges. The 15 highest priority (ranked) biotic and abiotic monitoring activities on refuges were the following: waterfowl counts, vegetative community, breeding landbirds, waterfowl production, waterbirds, shorebirds, marsh and wading birds, bald eagles, water gauge readings, grassland birds, invasive plants, deer, brown bears, moist-soil vegetation, and raptors (Table 2). Of the 15 highest ranked monitoring activities on refuges, 9 involved birds. Waterfowl counts was the highest ranked monitoring activity on refuges throughout the National Wildlife Refuge System.

Most of the monitoring activities on refuges throughout the National Wildlife Refuge System involved a limited number of survey techniques. For the 1855 monitoring activities on refuges for which methods were reported by respondents, there were only 46 different survey techniques (Table 3). Area counts, point counts, transects, and aerial surveys were used for 73% of all monitoring activities reported by questionnaire respondents. The most frequently reported survey technique was area counting which was used for a variety of biotic monitoring activities.

Of the 219 respondents to the questionnaire, 189 (86%) electronically stored their refuge monitoring data on a computer hard drive (Table 4). In addition to computer hard drives, some respondents also back-up monitoring data using other electronic media such as a floppy disk, Zip disk, CD-ROM, back-up tape, and/or server/network/LAN. Sixteen respondents (7%) stored data as data-entry sheets filed at the refuge without any electronic storage. Another 13 respondents (4%) had refuge monitoring data stored by another agency.

Questionnaire respondents used a variety of software for electronic storage/management of monitoring data collected on refuges. Of the 613 questionnaire responses regarding data management/storage software, 42% (259/613) identified relational databases (Table 5): Rbase (83), Access (68), FileMaker Pro (61), Dbase (33), Paradox (13), and Fox Pro (1). Forty-one percent (250/613) of the responses reported monitoring activities on refuges were electronically stored using non-relational (flat file) database software: Excel (141), Quattro Pro (81), and Lotus 123 (28). Seventy (11%) responses identified GIS programs (primarily ArcView) as data management/storage software. There were 14 (2%) responses that identified statistical programs (SAS, SPSS, SigmaStat, SysStat) as management/storage software. Seventeen responses (3%; 6 of 7 regions with $1) identified word processing programs (primarily WordPerfect) as management/storage software for monitoring data. Two responses (<1%) identified graphical software (Harvard Graphics and Power Point) as management/storage software for monitoring data. In accordance with the preference to use of database software for storage/management of monitoring data, many respondents recommended the use of Excel, FileMaker Pro, and Access for future development of specific applications to store/manage data for refuge monitoring (Table 6). Respondents also wanted databases to be user friendly and easily interface with GIS and statistical software.

Although questionnaire respondents indicated that most monitoring activities on refuges were used to make management decisions, computer programs were not frequently employed to analyze these data. Respondents indicated 74% (1286/1737) of the monitoring activities on refuges were used to make management decisions (Table 7); however, only 30% (514/1701) of the data for monitoring activities reported on the questionnaire were analyzed statistically, ranging from summary statistics (e.g., range, mean, standard deviation and error) to comparative tests (e.g., multivariate and factorial analyses of variance, t-test, regression). Respondents indicated 41% (692/1674) of the data for monitoring activities could be imported into GIS for data analyses. Similarly, 38% (662/1732) of the data for monitoring activities on refuges was geo-referenced through the use of GPS.

The questionnaire also inquired about software and hardware training needs for refuge staff responsible for collecting, storing, and analyzing data for biotic and abiotic monitoring on refuges. Most respondents (131/289 [45%]) identified the need for GIS training (Table 8). Associated with GIS, some respondents (31/289 [11%]) indicated the need for GPS training. Another 27% (78/289) and 15% (43/289) identified the need for training in the use of databases (particularly relational programs such as FileMaker Pro and Access) or statistical software,

respectively. A small number of respondents wanted training for metadata documentation and management (3/289) or population modeling (3/289).

Summary and Recommendations

Refuges throughout the National Wildlife Refuge System collect monitoring data to evaluate wildlife and habitat management actions. However, only a minority of refuges actually analyze these data on computers using database, statistical, and GIS software. Inappropriate or uneffective methods for storage/management of these data likely results in infrequent computer data analyses. For example, some refuges use graphical and word processing software for management/storage of monitoring data. Some refuges store/manage monitoring data in non-relational databases that do not facilitate effective assessment of relationships between management actions and wildlife and/or habitat responses. In contrast, relational databases can effectively store/manage monitoring data with linkages to facilitate assessment of these relationships**.** Although only a minority of refuges throughout the National Wildlife Refuge System use relational databases to store/manage their monitoring data, most of them successfully evaluate wildlife and habitat responses relative to management actions. In addition, questionnaire results indicate many monitoring activities are similar throughout the National Wildlife Refuge System and they utilize a limited number of survey techniques. Consequently, we recommend development of a Refuge Monitoring Data Management System to accomplish the following:

- Improve the use of monitoring data to make wildlife and habitat management decisions;
- Improve efficiency of the National Wildlife Refuge System by eliminating the need for each refuge to develop its own data management system;
- Facilitate analyses and summarization of biological data at the regional and national levels to aid in identifying management priorities, setting refuge system biological objectives, and improving the contribution of the National Wildlife Refuge System to biological resources;
- Facilitate exchange biological data with other organizations.

This data management system would address the storage/management of numerous types of data collected for monitoring activities on refuges. It would group biotic and abiotic monitoring data and information regarding management actions into modules with relational structure to create linkages amongst them in order to evaluate the response of wildlife and/or habitat to refuge management activities.

Table 1. Frequencies of biotic and abiotic monitoring activities (n = 186) conducted on refuges throughout the National Wildlife Refuge System.
NOTE – this table has been edited to remove taxa-specific entries. The full table is available online at:
https://drive.google.com/a/doi.gov/file/d/0B7xwLWClOEJxRkZpQWVyUXdjb09qel9MZ29VZ1p5RWNla0k4/edit?usp=sharing

	Region							
Monitoring Activity	1	2	3	4	5	6	7	Total
Bird Subtotal	148	96	167	209	215	162	66	1063
Mammal Subtotal	31	36	20	30	25	36	39	217
Amphibian Subtotal	22	11	18	37	41	4	1	134
Invertebrates Subtotal	5	2	11	4	12	5	1	40
Fish Subtotal	4	4	6	7	3	2	3	29
Habitat Subtotal	42	37	53	39	46	74	19	310
Abiotic Factors Subtotal	10	9	15	30	21	19	8	112
Other Subtotal	0	14	23	16	6	9	4	72
Grand total	262	209	313	372	369	311	141	1977
Number of refuges responding	29	38	35	43	37	25	12	219

Table 2. Rankings[a] for biotic and abiotic monitoring activities conducted on refuges throughout the National Wildlife Refuge System.
NOTE – this table has been removed. The full table is available online at:
https://drive.google.com/a/doi.gov/file/d/0B7xwLWClOEJxRkZpQWVyUXdjb09qel9MZ29VZ1p5RWNla0k4/edit?usp=sharing

[a]Rankings are calculated by dividing the sum of ranks for a monitoring activity (questionnaire #5) by the total number of refuges responding to the questionnaire.

Table 3. Frequencies of survey types (n = 46) used for monitoring activities conducted on refuges throughout the National Wildlife Refuge System.
NOTE – this table has been almost completely removed. The full table is available online at:
https://drive.google.com/a/doi.gov/file/d/0B7xwLWClOEJxRkZpQWVyUXdjb09qel9MZ29VZ1p5RWNla0k4/edit?usp=sharing

	Region							
Survey Method	1	2	3	4	5	6	7	Total

Total	264	251	137	383	372	234	214	1855

Table 4. Storage and management of data from monitoring activities on refuges throughout the National Wildlife Refuge System.

	Region							
Storage/management								
	1	2	3	4	5	6	7	Total
Electronic								
Hard drive	25	23	26	34	43	30	8	189
Floppy disk	3	7	0	15	4	6	7	42
Zip disk	8	3	1	6	8	3	5	34
CD-ROM	3	1	0	4	4	2	5	19
Server/network/LAN	9	3	0	2	0	0	5	19
Tape	2	3	0	0	2	3	3	13
Non-Electronic								
Paper files	0	14	8	20	0	4	3	49
No electronic storage	0	4	0	7	0	5	0	16
Stored by other agency	2	2	0	7	0	2	0	13
Grand Total	52	60	35	95	61	55	36	394

Table 5. Software for storage and management of data from monitoring activities on refuges throughout the National Wildlife Refuge System.

NOTE – this table has been edited. The full table is available online at:
https://drive.google.com/a/doi.gov/file/d/0B7xwLWClOEJxRkZpQWVyUXdjb09qel9MZ29VZ1p5RWNla0k4/edit?usp=sharing

	Region							
Software	1	2	3	4	5	6	7	Total
Relational database								
Rbase	15	2	22	1	42	0	1	83
Access	32	6	3	7	8	6	6	68
FileMaker	2	8	26	8	9	5	3	61
Dbase	10	4	2	3	1	8	5	33
Paradox	0	1	1	2	1	3	5	13
FoxPro	0	0	0	1	0	0	0	1
Subtotal	59	21	54	22	61	22	20	259
Non-Relational Databases								
Excel	53	13	12	18	19	16	10	141
Quattro Pro	43	8	6	7	4	9	4	81
Lotus 123	3	6	2	1	3	8	5	28
Subtotal	99	27	20	26	26	33	19	250
GIS								
ArcView	13	3	8	11	13	9	6	63
MapInfo	0	0	0	3	0	0	0	3
ArcInfo	0	0	0	0	0	1	1	2
Topo Scout	0	0	0	0	1	0	0	1
MIPS	0	0	0	0	0	1	0	1
MAPS	0	0	0	1	0	0	0	1
Subtotal	13	3	8	15	14	11	7	71

Table 6. Questionnaire comments received regarding the storage/management of data from monitoring activities on refuges throughout the National Wildlife Refuge System.
NOTE – this table has been edited. The full table is available online at:
https://drive.google.com/a/doi.gov/file/d/0B7xwLWClOEJxRkZpQWVyUXdjb09qel9MZ29VZ1p5RWNla0k4/edit?usp=sharing

Comment	Total
Create database which is simple, easy to use for new staff, volunteers, and interns	20
Create database which integrates with GIS and statistics	18
Develop generic /dynamic application to address numerous procedures	13
The core issue still is that data entry, data analysis, and report writing are lacking at the individual biologist level due to excessive work loads, shifting priorities, and employee turnover	6
Provide technical support for developed databases	6
Develop application in widely used software.	5
No Corel products	4
If you choose to focus on one desktop database management system, we recommend that you select Microsoft Access or Corel Paradox. Please do not select FileMaker Pro; it is a poor software package with many disadvantages	3
Create a windows based census program	2
Use ESRI software	2
Use SAS or other satistics software	2
Develop some standardized database applications and data structures to be used on data sets that are generally similar among several refuges.	2
At a minimum, the Service should develop a data vault@system accessible via the Intranet to store refuge biological data in a centralized permanent file archival	2
Many field stations, especially those with 20+ years of data, would benefit from having a database manager at the field station	2
Database flexibility to meet individual refuge needs	2
Provide field with software for databases	2
The Service should serve a centralized ProCite bibliographic database over the Intranet to document all published and unpublished refuge reports and data sets, including contact information for viewers to request the data from the original source	1
Some surveys are already standardized (e.g. BBS, MAPS, goose molting counts)	1
Create application which addresses non-traditional wildlife	1
All databases provided Service-wide or to the public should be well documented with metadata viewable by the user	1
Don't go overboard trying to standardize too many surveys. Databases should be user friendly and compatible with major statistical programs (such as SPSS, SAS, S+) and GIS software. More emphasis should go to getting biologists on refuges than spending to standardize software	1

which is constantly changing	
Use MicroSoft software	1
Window applications	1
Standardize database at least at ecosystem level and preferably at NWRS level	1
Standardize vegetation monitoring to evaluate mgmt.	1

Table 7. Analyses and use of data from monitoring activities on refuges throughout the National Wildlife Refuge System.

Region	Is data used for mgmt. decisions? Y/N	Are statistics used to analyze data? Y/N	Is data available to import into GIS? Y/N	Is GPS used in data collection efforts? Y/N
1	202 / 46	60 / 177	83 / 148	69 / 176
2	196 / 42	102 / 131	44 / 157	66 / 160
3	23 / 11	6 / 16	11 / 23	10 / 24
4	205/123	68/253	109/220	128/205
5	245 / 127	65 / 307	148 / 224	157 / 215
6	254 / 57	104 / 196	133 / 165	99 / 210
7	161 / 45	109 / 107	164 / 45	133 / 80
Total	1286/451 (74%)[a]	514/1187 (30%)	692/982 (41%)	662/1070 (38%)

[a]Value in parentheses represents the percentage of positive responses to a question.

Table 8. Training needs of refuge staffs associated with biotic and abiotic monitoring activities on refuges throughout the National Wildlife Refuge System.

	Region							Total
	1	2	3	4	5	6	7	
GIS	22	21	15	23	27	15	8	131
Databases	20	5	4	20	18	7	4	78
Statistics	12	5	0	10	10	5	1	43
GPS	3	9	0	12	3	4	0	31
Population modeling	0	0	0	2	0	0	1	3
Metadata documentation & mgt	0	0	0	1	0	1	1	3
Total	57	40	19	68	58	32	15	289

Appendix A

NWRS Biological Data Questionnaire

Fulfilling the Promises - Recommendation WH 9.1

Orgcode: _____Date: _____

Refuge: _____ (Please complete 1 questionnaire per refuge)

1.Please identify each survey procedure you conduct on the refuge which is biological or habitat oriented. Include procedures to measure numbers of wildlife, habitat, or abiotic factors which influence habitat. Record your responses on the attached questionnaire.

2.Please identify all data management software being used at the Refuge for management or storage of your biological data (examples of software are Microsoft Access, Foxpro, FileMaker Pro, etc):

3.Are you aware of biological database Applications which the *Promises* WH 9.1 Implementation Team should consider for use by the NWRS? An Application is a data management system that is created using software such as Microsoft Access. The Application is a set of data entry screens, output, menu system, etc). Please list any Applications you feel the Team should evaluate for NWRS use:

4.Please describe the types of training or technical support you may require to improve your use and management of biological data at the refuge (such as relational database use, Geographic Information Systems (GIS), and/or others):

5.Please identify your 5 highest priority biological data sets and list how many past years data is available (Example, Winter Waterfowl Census, 20 years):

6.If the refuge is using Electronic Data Storage for biological data, where and how is that data being stored at the present time? For example, does this data reside on a computer hard drive within the refuge office, is it permanently archived using some other media? Also, is the data accessible to others?

7.Please provide any comments/feedback you wish the Promises WH9.1 implementation team to consider in evaluating various database applications to manage refuge biological information:
Orgcode:
Refuge:

Group/Species/Habitat [A]	Method [B]	Season [C]	Data Storage [D]	Application / Software [E]	Data Used ? [F]	Statistics [G]	GIS [H]

[A] Identify survey procedure such as waterfowl, wintering waterfowl, waterfowl brood survey, or single species such as bald eagle. Include habitat surveys.
[B] Identify method of survey: Point Count, Transect, Area Survey (count of birds within impoundment, etc), quadrat, point intercept, etc.
[C] Identify all seasons of the year the survey is conducted: 1 = Winter, 2 = Spring, 3 = Summer, 4 = Fall, 12 = Winter and Spring, 123 = Winter, Spring, and Summer, etc.
[D] Identify data storage: P = Paper Files, E = Electronic, X = Other.
[E] If electronic storage, identify either database or spreadsheet software used (e.g., Microsoft Access) or application used (e.g., Census Database).
[F] Is the data used to improve refuge management decisions? Yes / No
[G] Is the data statistically analyzed to improve reliability of decisions ? Yes / No
[H] Is the data available for importing into a GIS ? Yes/ No

Appendix B: Region 8 Report: Supporting Refuge Data Management.

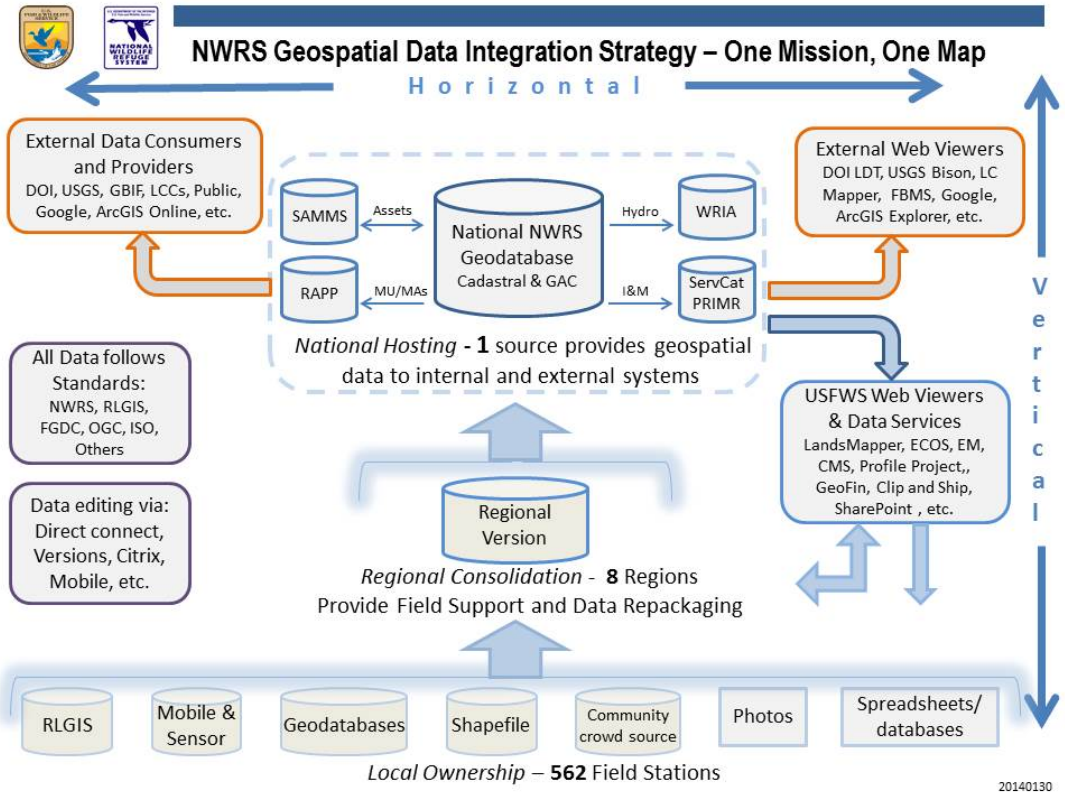
Kaylene E. Keller, Ph.D. Pacific Southwest Region 8 Inventory & Monitoring Ecologist/GIS/Data Manager
Karen Laing, Pacific Southwest Region 8 Inventory & Monitoring Coordinator
1/31/2014

Data to Support the National Wildlife Refuge System

Recently the Refuges National GIS Coordinator outlined a vision for data integration (Figure 1). The vision focuses on the national level of integration and how national databases can be used to support external data consumers. It does not articulate how high quality data will become part of the system except to indicate that Refuges and Regional staff will supply the data.

Individual refuge stations operate independently of each other, without significant technical support from each other or their Regional Offices. The Pacific Southwest Region conducted a Needs Assessment of all 50 stations in the region in 2012, providing an example of the current ability of refuges to provide data imagined in the vision articulated in Figure 1. Here we use the Needs Assessment and other information to summarize data management needs for the region.

Figure 1: National Vision for Data Integration (Provided by Dan Craver and Ron Salz)



Current Status of Pacific Southwest Region Refuge Data Support

The Pacific Southwest Region Inventory and Monitoring program (I&M) conducted a Needs Assessment at all stations in the region in 2012. I&M staff visited each stations, implementing a standardized interview with refuge staff, one section of which focused on data management needs. In addition, I&M conducted a data organization pilot project at a single refuge, and an informal survey of refuge internet connections.

Refuge GIS/Data Management Needs:

During the Needs Assessment refuge staff members were asked "What are your biggest GIS/Data Management needs?"

- 44% of stations indicated that getting **additional staff to do GIS** work was a top need; 42% of stations indicated that **organizing data** was a top need; and 32% of stations indicated that getting **database technical field support** was a top need (Figure 2).
    - Creating, maintaining and visualizing GIS data takes staff time and specialized skills. To address Refuge needs we need to hire staff with the necessary technical skills. Without technical support refuge staff can take hours to figure something out that would take someone with the technical background minutes.
    - Data organization is rarely addressed in a class, but is something that everyone struggles with. Supporting staff with training in data organization and supporting staff taking the time to organize and document data is an important component in data management.
    - More information needs to be gathered to understand the specific needs that stations have for database technical support and how that support can be provided.

Refuge Data Users

The level of GIS technical support currently available to refuges in the Region is low. 32 stations have no staff members who have the primary responsibility of GIS / database management. At these stations GIS / database management is handled by multiple people and generally there is no individual to coordinate the data. 17 other stations have someone whose primary role is GIS / database management. Users with the primary role of GIS/database management support multiple stations and frequently are not classified as GIS specialists.

The following are the results for GIS users from the Needs Assessment

- 46 stations responded GIS is used at the Refuge; 4 responded that GIS was not used at the refuge.
- An average of 2.38 staff use GIS per station. 50 individuals report using GIS at stations. Often one individual provides GIS support for multiple stations.
- 72% of station staff that use GIS reported **viewing, creating and managing** GIS data; 20% reported **viewing and creating** GIS data; and 7% reported just **viewing** data.

    - The majority of station GIS users are creating data, even though their ability to produce high quality data is limited. This represents the foundation for many of the data integration tools envisioned at the national level. The region and individual stations need to provide station staff with the support to create the high-quality data that will be required to support station management, surrogate species, strategic habitat conservation and other landscape-scale planning efforts.
- Approximately 62% of these staff gained their GIS skills through **short courses or on-the-job training**, while 35% gained their GIS skills through **university courses or while working in a GIS-focused job**.

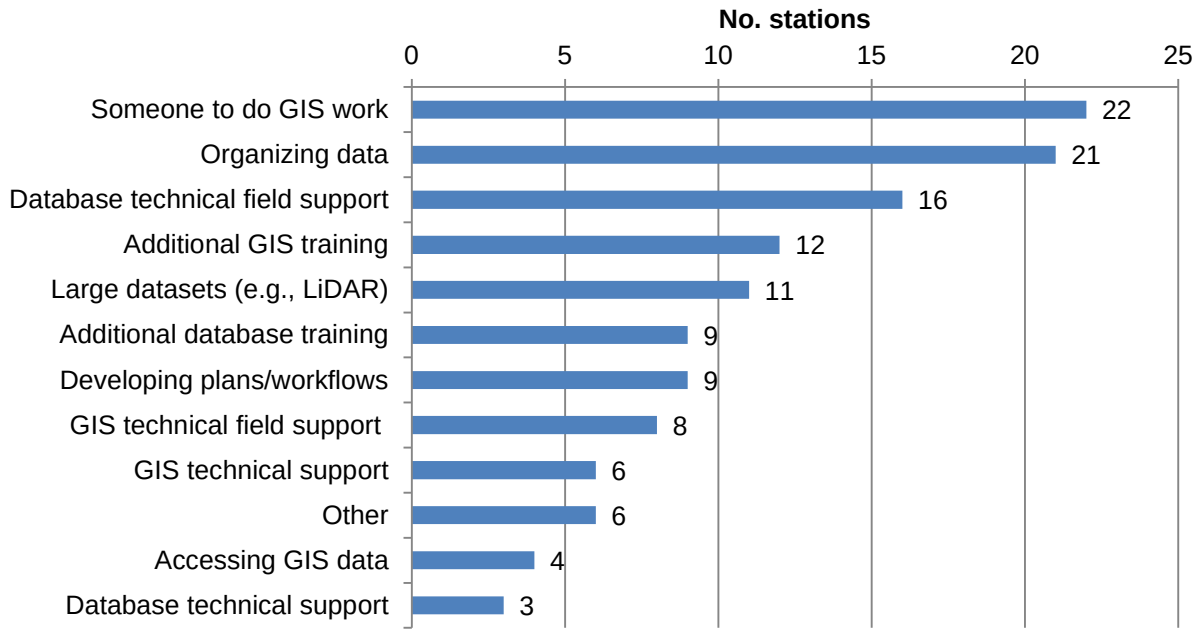
The following are the results for database users from the Needs Assessment

- 42 stations responded that they maintain databases; 7 responded that they do not maintain databases.
- 54 individuals reported maintaining databases at stations (mean = 1.6 staff/station). Often the same people supported multiple stations and also were the same people supporting GIS at the station.
    - The support of databases maintained by station staff needs to be investigated further. There may be opportunities for databases to be developed that meet multiple station needs rather than each station having to develop a database independently.
    - The long-term viability of the databases also needs to be evaluated. Is it a database that only one person understands? If that person leaves, can the database be supported? If databases are not supported/documented, we are at risk of losing the ability to use the data in the future; i.e., if we don't know what the codes in the database represent, the data is useless.

The results for non-spatial database support were similar to what was reported for GIS support. In addition, there is currently no regional technical support that station staff can access to ask questions or to get guidance on how to organize data, complete datasets or QA/QC data. It is not feasible to have a full-time data manager (spatial or non-spatial) at every station, but there needs to be more technical support for the field. The type of technical support needed by stations will vary depending on the existing level of support and overall data needs.

To understand refuge conditions better, I&M staff members assisted an individual refuge to organize refuge data to support a habitat management database. The pilot project refuge has 3 staff that have been using GIS to create, view and manage the refuge data. A fourth staff person has begun learning GIS in the last year. Each individual develops and manages their own data and nobody coordinates data for the refuge. When I&M started working with the refuge there

Figure 2. GIS/data management needs of R8 stations.



were 1500 GIS associated files found on the server and 3 laptop computers. 698 of the files were shapefiles and the majority of those were duplicates. After a first round of standardization and clean-up there are approximately 43 GIS datasets. In addition, the refuge has limited internet connectivity and a server they inherited from the Regional office. The server does not have enough space to handle the Regional DataSpace[20] library and the connectivity is not sufficient to utilize the distribution tools available with DataSpace. The refuge staff has taken advantage of the Regional GIS training in 2012 but still struggle with data structures and data organization. Additional training is needed to address data organization and how to decide how to structure data.

Needs

Data Organization.

Refuges identify Data Organization as a top data management need. The average rating of data organization on a scale of 1-10 was a 4. It was not recorded during the interview but it appeared that refuges with recent or high staff turnover rated data organization lower. We suspect that it is a result of the difficulty new staff have finding past data and documents and data that is available lacks documentation. Data organization is not taught in classes and it requires effort to set up a functioning structure. Refuge staff will need support to walk through a process to get files and

[20] DataSpace: Regional tool for distribution of regional data library. The system allows the regional office to "push" updated datasets to regional offices through direct connections with office servers.

workflows set up. When new staff inherit a computer that has 1500 GIS and data related files with no documentation the task is overwhelming.
In order for a refuge to address the issues associated with data management, the refuge staff must decide that it is a priority and allocate staff time to addressing the issue. In addition the region and HQ need to make it a priority and provide the technical support necessary to support refuge staff.
As described above refuge staff have limited time to devote to moving from their existing situation to an organized data system and structured data. Organization of refuge data and development of an efficient system that can be managed by users with limited skills and in staff turnover situations will take time and technical support. We as an agency need to re-evaluate how we approach training and support. It is not enough to have staff attend a multi-day training and then return to a refuge where they have no support.

Technical Support.
In Region 8 there is limited GIS and data support for 50 stations/refuges. All of the regional support (1 in planning, 2 in I&M and a couple of ES offices) have other primary duties and refuge support is a collateral duty. If we expect refuges to provide high quality, standardized data for national centralized systems we need to provide better support to refuge staff.
In determining how to provide technical support we need to examine how much time staff spend using GIS and conducting data management tasks. We also have to look at how data is used and the frequency different datasets are used.
During the Needs Assessment refuges where asked to estimate the time staff spent on GIS-related work.

- Staff using GIS reported spending an average of 9.9% of their time on **GIS-related work** per station. However, this may be an overestimate because it appears that a few station staff inflated their estimated time spent on GIS-related work (e.g., their total time spent on GIS added up to more than 100%).
- Approximately 60% of station GIS users spent 10% of their time or less using GIS.
    - This translates to 4 hours or less a week; most likely that time is not evenly spread across the year and comes in short intensive periods such as producing numbers for RAPP once a year. If this is the case, then we will need to work with stations to understand how GIS is used and how best to support the users. For example we may need to develop workflows to follow and provide additional technical support because the frequency of use makes it difficult to remember what was done the year before. In the GIS data that I&M has reviewed, we often see users starting from scratch each year.

One of the uses of GIS is to create acre calculations for the RAPP system. This is a once a year event. When evaluating the pilot project refuge GIS files it became apparent that every year the

staff started from scratch and the data was not consistent from year to year. Time is wasted starting over each time they have to do the task. In addition, if a user is only using GIS for short periods of time or infrequent but high intensity periods of time they may not have enough time to become familiar with new versions of the software. The limited amount of time staff can dedicate to GIS and data management requires us to re-evaluate our approach to training. If someone spends a couple of days in training but cannot apply what they learn immediately the new skills are quickly forgotten.

The following are some potential methods for providing support for data organization and technical support for refuges:

1. Set up data organization and technical coaching arrangements with refuge staff. Currently Region 8 I&M is conducting a pilot project with 4 refuge staff representing 1 refuge and 2 refuge complexes. The participants are all tasked with GIS and data management as a collateral duty. We are providing a workflow and tools for the participants to work their way through organizing their data. So far we are finding that we cannot just provide the tools; we need to provide the tools, teach them how to run them and then help trouble shoot when problems arise. In addition to working with refuge staff in addressing data organization issues we will also be providing technical support for standardization of existing data and creation of new data. After working with the staff for 6 months the I&M staff realized we needed to meet individually with refuge staff. We needed to make site visits to understand each refuges unique situation and help the refuge staff handle the backlog of data management issues. Conference calls and web-ex can help but one-on-one support is critical for success.
2. Refuge staff will also need to set priorities to help identify high priority datasets that will help support refuge management. It will take time to migrate from the existing data organization and existing structures into a standardized system. We can use existing CCPs, annual workplans and other planning documents to help identify the data needs as well and set priorities. The refuge system needs to move data from a reactionary system to developing datasets in a planned and structured manner.
3. Existing I&M staff will be able to provide organizational and technical support during the pilot project but for long-term success we will need GIS and data management support at a zone level. GIS and data managers would be able to provide refuges with technical support, data development support and assistance with maintaining centralized data systems such as ServCat and future centralized tools.

Internet Connection.

A major constraint for refuges to provide data into centralized data systems or provide data to be rolled up into national datasets is the lack of high speed internet connections at refuges. This is not a problem confined to remote refuges (Table 1). Stone Lakes NWR has the lowest internet speed of the refuges evaluated and they are on the edge of a high density urban area. Hopper

Mountain Complex offices are in an office park. Simply directing refuge managers to find a solution will not work without also providing IT support that can think creatively about solutions.

Table 1: Internet speed calculated using Speedtest.net during December 2012

Station	Download (Mb/s)	Upload(Mb/s)
ANAHO ISLAND NATIONAL WILDLIFE REFUGE (Stillwater Complex Office – Fallon NV)	1.01	0.46
ASH MEADOWS NATIONAL WILDLIFE REFUGE (Desert Complex)	1.13	1.28
DON EDWARDS SAN FRANCISCO BAY NATIONAL WILDLIFE REFUGE (San Francisco Bay Complex - Don Edwards Office 12/14/2012)	1.1	1.23
DON EDWARDS SAN FRANCISCO BAY NATIONAL WILDLIFE REFUGE (San Francisco Bay Complex – Don Edwards Office 12/4/2012)	.77	1.27
ELLICOTT SLOUGH NATIONAL WILDLIFE REFUGE (San Francisco Bay Complex – Monterey Bay Office)	5.25	0.64
HOPPER MOUNTAIN NATIONAL WILDLIFE REFUGE (Hopper Mountain Complex)	1.35	1.29
HUMBOLDT BAY NATIONAL WILDLIFE REFUGE (Humboldt Bay Complex)	3.17	0.2
RUBY LAKE NATIONAL WILDLIFE REFUGE	1.34	0.15
SACRAMENTO NATIONAL WILDLIFE REFUGE (Sacramento Complex)	1.22	1.39
SAN LUIS NATIONAL WILDLIFE REFUGE (San Luis Complex)	1.21	1.16
SAN PABLO BAY NATIONAL WILDLIFE REFUGE (San Francisco Bay Complex – San Pablo Bay Office)	1.45	0.23
STONE LAKES NATIONAL WILDLIFE REFUGE	0.34	0.11

It is possible that IT is addressing this problem but it is not known by the regional staff if they are addressing it. The following are some first steps that can be taken to begin to address the issue at a regional scale.

1. Identify communication person in FWS. Who handles network/internet connections for the Service.
2. Identify someone to investigate and report on the existing network structure

    a. What type of connection does each refuge have ie satellite, DSL, Cable, Comcast etc.

    b. Are the refuges getting the speed they are paying for?

c. What type of set up does the refuge have with the FWS network ie VPN, LAN to LAN something else? If they are on the LAN to LAN which consolidator are they going through (Denver or Reston)?

d. Who does the IT support for the refuge?

Once we have this information we can start looking at what needs to be done. It might be that the refuge is not getting the internet speed they are paying for so a technical person needs to work with the provider to make sure we get what we are supposed to get.

Using Existing Processes to Improve Data Development

Currently the refuge system wastes time and money by developing the same data/information to support different projects such as surrogate species, landscape conservation design, CCPs, HMPs, and WRIAs. To address this issue of redundant data development we need data managers who can coordinate data development and sharing between projects. These would be new positions at a regional level available to support regional projects and provide technical assistance to refuges. Effective coordination and management of data require specialized technical skills and cannot be accomplished as a collateral duty. Leaders at the national, regional and refuge level need to prioritize data management to maximize the return on FWS data.

Conclusions

1. If USFWS Refuges program wants to develop national data integration tools refuges must have technical support for data development.
2. The current level of technical support available to refuge staff is variable across the regions and within regions.
3. We need to look at alternatives to the traditional methods of training and look at how to best support casual users.
4. Internet connections must be improved if we expect refuges to contribute to centralized systems and use “the cloud”.
5. Regions need to evaluate current processes to find opportunities to leverage data collection efforts to support additional data uses.

Acknowledgements

Thank you to the Region 8 I&M team and Refuge staff who have participated in the Needs Assessment and the I&M pilot projects. The information and feedback is invaluable as we attempt to develop solutions to the data management challenges faced by the USFWS.

Appendix C: Key Federal, DOI and FWS Data Management Policies

Policy Type	Title	Brief Description
DOI	Department of the Interior Departmental Manual Information Resources Management, Part 375 IRM Program Management, Chapter 12: Information Resources Standards Program	The Information Resources Standards Program coordinates the development, adoption, implementation, and review of information management, automated data processing, and telecommunications standards.
DOI	Department of the Interior Departmental Manual Information Resources Management, Part 378 Data Resource Management, Chapter 1: Program Description and Objectives	The Data Resource Management Program encompasses the process of planning managing, controlling and protecting DOI data assets while supporting DOI business functions and goals.
DOI	Agency-wide Internet Accessibility Policy	The CIO web page does not restate Section 508, but does provide guidelines for 'Best Practices', responsibilities, and tools to assist in complying with section 508 requirements.
DOI	Transition to IP6	This CIO Memorandum states the Department's commitment to the operational deployment and use of Internet Protocol version 6 (IPv6).
DOI	Limiting of Administrative Rights	All DOI agencies are required to limit administrative rights to servers and computers to those staff that are required for operational use.
DOI	National Institute of Standards and Technology – (NIST) Special Publication 800-53	All agencies are required to manage the security of servers, networks (LAN, WAN), and computers must conform to security guidelines, and administrative rights to servers and computers should be limited to those staff that are required for operational use. All users with administrative rights must have an approved request, with justification, on file with the agency's Office of Chief Information Office.
Executive Office of the President	Data Reference Model, Version 2.0	The DRM is a framework whose primary purpose is to enable information sharing and reuse across the federal government via the standard description and discovery of common data and the promotion of uniform data management practices.
Executive Office of the President	National Strategy for Information Sharing and Safeguarding	Provides guidance for effective development, integration, and implementation of policies, processes, standards, and technologies to promote secure and responsible information sharing
Executive Office of the President	Digital Government: Building a 21st Century Platform to Better Serve the American People	Provides strategies and guidance for making government information more accessible to its citizens

Executive Office of the President	Memorandum for the Heads of Executive Departments and Agencies, Increasing Access to the Results of Federally Funded Scientific Research	Provides guidance to agencies to release the results of scientific research funded by the federal government publically
Executive Office of the President	Transparency and Open Government	Directs the Chief Technology Officer, in coordination with the Director of the OMB and the Administrator of General Services, to coordinate the development by appropriate executive departments and agencies, within 120 days, of recommendations for an Open Government Directive, to be issued by the Director of OMB, that instructs executive departments and agencies to take specific actions implementing the principles set forth in this memorandum.
Executive Office of the President	Building a 21st Century Digital Government	The Strategy will enable more efficient and coordinated digital service delivery by requiring agencies to establish specific, measurable goals for delivering better digital services; encouraging agencies to deliver information in new ways that fully utilize the power and potential of mobile and web-based technologies; and more.
Executive Order	Content Standard for Digital Content Standard for Digital Order 12906	CSDGM is the federal standard for the documentation of geospatial data. The standard was defined by the Federal Geospatial Data Committee
Executive Order;#OMB Circular	Project Open Data	Memorandum for the Heads of Executive Departments and Agencies regarding the Open Data Policy - Managing Information as an Asset
Federal Law or Directive	Freedom of Information Act	The FOIA grants anyone the right to access any DOI records unless DOI reasonably foresees that the release of the information would harm a protected interest or falls under one of thenine exemptions.
Federal Law or Directive	Clinger-Cohen Act of 1966	Rules for information technology acquisition and management (primarily for hardware and software)
Federal Law or Directive	E-Government Act of 2002	The purpose of the E-Gov Act is to improve the methods by which Government information, including information on the Internet, is organized, preserved, and made accessible to the public.
Federal Law or Directive	Confidential Information Protection and Statistical Efficiency Act of 2002	CIPSEA provides strong confidentiality protections for statistical information collections and statistical activities such as data analysis and sample design that are sponsored or conducted by Federal agencies.
Federal Law or Directive	Section 508 of the Rehabilitation Act of 2000	Provide open access to information products, with provisions for serving people with disabilities
Federal Law or Directive	Paperwork Reduction Act of 1995	This Act is a subpart of the Coordination of Federal Information Policy, and is intended to streamline and standardize government IT planning, purchasing, operations,

		and activities.
Federal Law or Directive	Government Paperwork Elimination Act of 1998	GPEA requires federal agencies to allow individuals or entities that deal with the agencies the option to submit information or transact with the agency electronically. The Act specifically states that electronic records and their related electronic signatures are not to be denied legal effect.
Federal Law or Directive	Patent and Trademark Law Patent and Trademark Law Act) of 1980	In 1980, the Bayh-Dole Act (PL 96-517, Patent and Trademark Act Amendments of 1980) created a uniform patent policy among the many federal agencies funding research. As a result of this law, universities retain ownership to inventions made under federally funded research. In return, universities are expected to file for patent protection and to ensure commercialization upon licensing.
Federal Law or Directive	Federal Information Processing Standards	Under the Information Technology Management Reform Act (Public Law 104-106), the Secretary of Commerce approves standards and guidelines that are developed by the National Institute of Standards and Technology (NIST) for Federal computer systems.
Federal Law or Directive	Federal Information Security Management Act	Enacted in 2002 as Title III of the E-Government Act of 2002, FISMA recognized the importance of information security to the economic and national security interests of the United States.
OMB Circular	Coordination of Geographic Information and Related Spatial Data Activities - OMB Circular A-16 (1990,2003, 2010)	This Circular provides direction for federal agencies that produce, maintain or use spatial data either directly or indirectly in the fulfillment of their mission.
OMB Circular	Requirements for Grants and Agreements - OMB Circular A-110 (1999, 2003)a.k.a. “Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations”	This Circular sets forth standards for obtaining consistency and uniformity among Federal agencies in the administration of grants to and agreements with institutions of higher education, hospitals, and other non-profit organizations.
OMB Circular	Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities – OMB Circular A-119 (1998)	The revised Circular establishes policies on Federal use and development of voluntary consensus standards and on conformity assessment activities; and authorized the National Institute of Standards and Technology to coordinate conformity assessment activities of the agencies.
OMB Circular	Management of Federal Information Resources – OMB Circular A-130	General policies that apply to the information activities of all agencies of the Executive Branch of the Federal government.
OMB Circular	Information Quality Act	Directs the OMB to issue government-wide guidelines that "provide policy and procedural guidance to Federal agencies

		for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies.
OMB Circular	Open Data Policy	Pursuant to Executive Order of May 9, 2013, Making Open and Machine Readable the New Default for Government Information, this Memorandum establishes a framework to help institutionalize the principles of effective information management at each stage of the information's life cycle to promote interoperability and openness
OMB Circular	Improving Public Access to and Dissemination of Government Information and using the Federal Enterprise Architecture Data Reference Model	Identifies proceudres to organize and categorize information and make it searchable across agencies to improve publicv access and dissemination.
USFWS	FWS Manual Chapter 274 FW 2 – Establishing Service Data Standards	FWS Manual chapter describing how to establish data standards.
USFWS	FWS Manual Chapter 274 FW 1 – Data Resource Management	FWS Manual chapter providing the authorities for and definitions of terms used in chapters Part 274, Data Management.
USFWS	FWS Information Quality Guidelines	Guidelines issued by the FWS for ensuring the quality, objectivity, utility, and integrity of information disseminated by FWS.

Appendix D: Brown and Henstorf 2014 - Making the most of scarce data-mining talent.

This is a post from the Harvard Business Review blog network. Brad Brown, the author, is director of the New York office of McKinsey & Company management consultants.

Make the Most of Scarce Data-Mining Talent
by Brad Brown and Brian Henstorf | 10:00 AM January 17, 2014

The immense promise of big data to reveal new opportunities and deliver practical business results has so far been focused on technologies and models, and less on the human challenges of staffing roles and processes to take advantage of big data's promise. The technology may be abundant, but developing, recruiting and hiring the people to use it is becoming an acute challenge for Fortune 1000 companies. Defining the roles, recruiting talented practitioners, setting up center of competence structures, establishing data governance across business units, and tying advanced data and analytics (AD&A) to the results of those businesses is lagging the deployment of tools and the collection of the data.

The Talent Gap in Big Data
By 2018, the United States alone could face a shortage of as many as 190,000 people with deep analytical skills according to a study by the McKinsey Global Institute. The study also found a looming need for over 1.5 million managers and analysts who understand big data and how to apply it to business operations. More than 70% of the Fortune 1000 companies surveyed by New Vantage Partners said it was "very difficult" to source analytical skills, with more than a third of the respondents saying their current level of AD&A skills are less than adequate.

To some extent, higher education is stepping into the gap. North Carolina State University worked with locally-based SAS — a leader in business analytic software — to offer a Master of Science in Analytics (MSA). Other universities are establishing similar programs, but until the supply of qualified candidates catches up with demand, most organizations are focusing on the internal development of big data skills by creating data literacy programs to establish a baseline level of knowledge for all their employees and setting professional tracks to build data skills and nurture career paths to retain their existing specialists.

Human resources departments are also looking globally at traditionally analytically-intensive sectors ranging from meteorology to medicine and finance to find talented candidates. This talent isn't cheap: a top data scientist can command a $300,000 salary in the current market.

The Nature of the Roles in an AD&A Organization

There are clear categories of staff roles required to drive a successful advanced data and analytics agenda. Companies that are far along in advanced data and analytics have adopted a roughly similar model organized around a center of excellence with three types of talent:

1. *Technical and Data Specialists*: These positions range from data quality managers who ensure the collected data are clean and accurate to business solution architects who assemble the data and organize it so it is ready for analysis.

2. *Analysts and Data Scientists*: This includes the foot soldiers of the function who sift through mountains of data seeking insights and the "ninja" scientists who create sophisticated models to predict customer behavior and allow advanced customer segmentation and pricing optimization.

3. *Business Analytics and Solutions specialists*: These people are aligned by domain and sometimes sit within the business units they serve. This category would include insights analysts who turn models into actions and are the primary interface between the center of excellence and the business units.

Form a Center of Excellence to Extend Scarce Resources

Advanced organizations are making the most of their scarce resources by centralizing their analytics organizations into centers of analytic excellence (COEs) to act as a hub to serve business units and departments. Success in companies that have adopted a COE model depends on strong leadership by a leader who can break down silos and foster a strong culture of customer service for internal customers who may lack confidence in the value and trustworthiness of data models. The best COEs measure performance not on volume or speed but by their impact on business success. COEs need clear governance on how advanced data and analytical decisions are made, laying out how impact is measured and constantly reviewing the COE's agenda to ensure the business units are using the team in a way that leads to practical results.

The lessons we have learned from our work in big data is that success comes when companies make strategic hires from hotbeds of data-driven cultures such as Silicon Valley to identify and recruit the new breed of leaders who understand the technology, science and data behind advanced data and analytics. Successful big data practitioners leverage strategic partnerships to obtain scarce talent across geographies while sourcing globally with the understanding that the analyst and scientist roles are especially dispersed. Finally, big data success requires a change in culture to be driven from the top to retain the new breed of data-driven managers while

convincing the rest of the organization to be more oriented and aligned around big data and its potential.

Achieving Impact by Changing the Business Culture

Most important, COE leaders are acting as change leaders to drive data-driven approaches into the business, and shift the culture from art to science. Some leaders are achieving this by focusing on the business units that can become reference cases across the entire corporation. In addition, leaders are praising and rewarding the individuals and groups in their company who have made the transition to advanced analytics and big data. With techniques like these, COE leaders are creating a wave of front line change.