Atlantic Marine Bird Survey Protocol Development:

Modeling Effective Data Management

Project description

* We propose to develop a data management protocol for the Atlantic coast sea duck and seabird surveys that the Divisions of Migratory Bird Management and Bird Habitat Conservation have conducted since 2008.
* The protocol will serve as a template for the development of protocols for all DMBM/DBHC data management efforts.
* This project will employ a collaborative, team-based approach that will ensure reproducibility and continuity, and will support staff training in the principles of effective data management.
* In addition to the protocol, we will document the development process, and provide an assessment of lessons learned and necessary resources.

**The Need**

High quality data are essential for sound scientific research, and monitoring data, in particular, constitute an irreplaceable historical record. Effective management, in turn, is critical to ensure data can be found, interpreted, and used for decision-making. And the opportunity for discovery beyond the original intent of the data collection advances research and conservation management. The Division of Migratory Bird Management (DMBM) has an impressive record of managing long-term monitoring data, and a well-established commitment to making data publicly available.

Data management is typically viewed as the process of collecting and storing observations (i.e., the data) that are error free. The Data Management Association (DAMA) defines data management (or data resource management) more broadly as the the development and execution of architectures, policies, practices, and procedures that properly organize the full data lifecycle needs of an enterprise (Fig 1). This extends beyond the expectation that data are simply error free observations. The DAMA definition implies that data management is a formal process that results high quality data, and that high quality data include associated documentation and code detailing the study design, known problems in the data, variable descriptions, procedures for data acquisition and Quality Assurance and Control (QA/QC), policies for data sharing and reuse, and detailed mechanisms for short-term and long-term storage to guard against data loss. Furthermore, all this information should be easily available, ideally in a centralized clearinghouse with oversight, open to discovery and scrutiny by others.

Considering these components of data quality and modern technological standards, it is apparent that the DMBM must update our data management processes: Our data are not easily discoverable, well documented, nor consistent in their structure, organization, or location. We need to improve the quality control of our data collection, the quality assurance of our data processing, the organization of our data storage, and the ease of data access. We lack documentation and clear procedures for all stages of the data management lifecycle and individuals lack authority and accountability for data and information management. If we fail to refine our current data management processes, our data will cease to be an asset, and will increasingly become an impediment with considerable staff time necessary to accomplish tasks related to data maintenance, access, and use. The Division needs to develop new data management procedures to improve data quality and discoverability. Done well, changes to our current operating procedures will simultaneously reduce staff time devoted to inefficient and uncoordinated data management. We propose to develop a framework for DMBM data management by drafting a comprehensive data management protocol for one survey effort, and document this effort so as to model the process and products needed for all the data sets and monitoring programs that fall under DMBM authority. This effort will inform, not only DMBM projects that generate data, but also how we handle data and information from collaborators, partners, and third-parties.

**Objectives**

DMBM will achieve effective data management when every project with data needs has a documented protocol covering the relevant stages of the data management lifecycle and the connections between these stages. The objectives of the proposed project are to:

1. Pilot an integrated, collaborative process for protocol development;
2. Draft a template that can be used to develop protocols for all DMBM data projects;
3. Evaluate DMBM data management capacity and needs.



**Figure 1.** The data management lifecycle illustrates the various activities that require coordination and documentation to ensure successful data management. Programs within DMBM are engaged in all aspects of the lifecycle

**The Approach**

We will use the National Wildlife Refuge System’s *How to Develop Survey Protocols: A Handbook* as the guide for our protocol development. This handbook provides a template for documenting all aspects of data collection and management with a general narrative and SOPs for each specific component of the data lifecycle. We will also document the process of creating the protocol to provide a road map for other efforts with lessons learned. We plan to evaluate a formal project management philosophy and collaborative project software solutions as part of the protocol development (see attached description of that process).

Deliverables:

1. A survey protocol that addresses all aspects of the data management life cycle and serves as a prototype for additional data management efforts (our application is described in the next section, see the *Whooping Crane Winter Abundance Survey Protocol* from Aransas NWR as an example of a completed protocol).
2. A formal populated relational database that is available for discovery and analysis (i.e., centralized SQL).
3. A living workspace for continued project maintenance and new project development.
4. An evaluation of the process that outlines steps, timelines, and caveats for future efforts, and identifies and estimates DMBM needs for full data management effectiveness.

Application: For this effort, we plan to use the Atlantic Coast Sea Duck and Atlantic Marine Assessment Program for Protected Species surveys. This suite of surveys was chosen for several reasons. First, the surveys were designed and implemented by current DMBM/DBHC staff. Second, many QA/QC components have been developed for these data, along with a well-constructed database that includes edited effort information, i.e., the data and database are already organized and properly formatted, and represent “the latest and greatest” of our database efforts. Finally, these surveys have a recent and relatively short history, making the effort needed for their documentation manageable; since the surveys were initiated as an experimental effort in 2008, the data also contain a certain amount of complexity, in terms of survey changes, that will be useful for developing general guidance for other surveys. For these reasons, we believe the process of protocol development can occur successfully within a reasonable timeframe, and that prototyping the process using these surveys will provide useful guidance for larger and more complex efforts.

**Project Leads:** Emily Silverman, Nathan Zimpfer (1 day/wk)

**Project Collaborators:** Tony Roberts, Kammie Kruse, Sarah Yates (2 day/month)

**Feedback and Review:** Mark Koneff, Kathy Fleming, Tim Jones, Becky Rau, Doug Tucci (1 day/month)

**Resource needs**

* This project requires a dedicated time commitment from the team members to succeed.
* To succeed in a timely fashion, the project also requires access to collaborative software. This software is especially critical if we are to build deliverable #3, the living workspace. Investment in this workspace will benefit DMBM going forward, because it will ensure that the team’s work is continued and maintained in an efficient way, in addition to creating workspaces for other data assets.

**Milestones and Timelines**

The following are the minimal milestones identified by the project leads. Once the project has been initiated, these milestones, and the necessary timelines for completion, will be discussed and set by the project team, and collapsed or expanded as necessary.

* Develop documentation detailing background and objective of the survey(s) - Silverman, Roberts, Koneff, Jones
* Develop documentation of survey design and survey protocols/training requirements - Silverman, Yates, Jones, Fleming
* Develop documentation of initial data processing and visualization methods, code (R code, R/ArcGIS interface), and analyses - Silverman, Zimpfer, Kruse, Leirness
* Create process and documentation for data storage (i.e., database design schema) and coding system and archiving - Zimpfer, Silverman, Tucci, Leirness
* Investigate requirements and/or API creation for plugging in to data.gov initiative or DMBM datacenter - Zimpfer, Rau, Tucci
* Develop data use and sharing policy - Silverman, Zimpfer, Kruse, Rau