pacific northwest aquatic monitoring partnership

Guidance for Implementing Successful Data Management & Sharing

March 6th, 2012

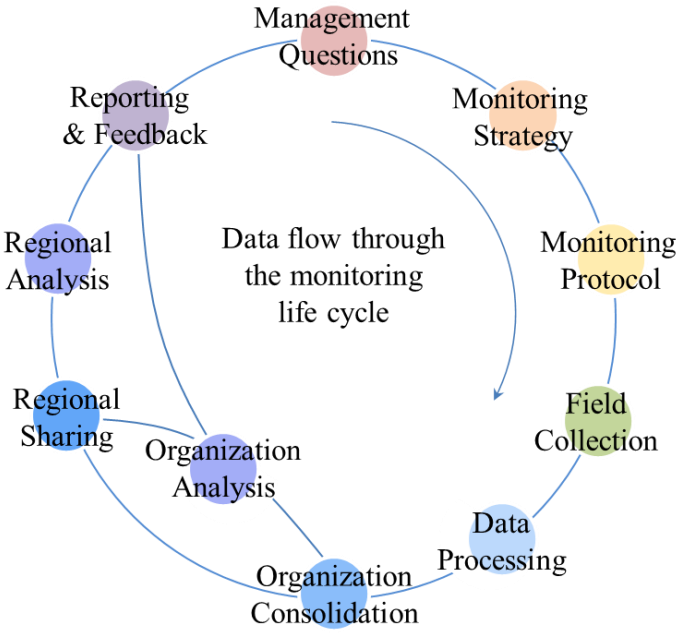
PNAMP Publication No. 2012-004

Accessible, high quality data are fundamental to a successful monitoring program. Effective use of these data requires the application of best management and data sharing practices - whether the data were gathered by direct observation, generated through remote sensing, or derived (as with a calculated metric or indicator).

This approach outlines five actions to improve the flow of data through the monitoring life cycle. Each of these recommended actions points to opportunities for organizations that monitor fish and aquatic habitats to evaluate and potentially improve their data management and sharing practices. Managing a healthy data flow throughout an aquatic monitoring program's life cycle is vital to enable the type of data management and sharing needed to support management decisions, to fulfill legal reporting obligations, and to understand emerging issues such as invasive species and climate change effects. This guidance should be useful to Pacific Northwest Aquatic Monitoring Partnership (PNAMP) signatory partners, participants, funders, and other interested aquatic monitoring practitioners in the Pacific Northwest (PNW) in identifying the structural components of a comprehensive data management and sharing approach and evaluating gaps in existing programs.

THIS APPROACH

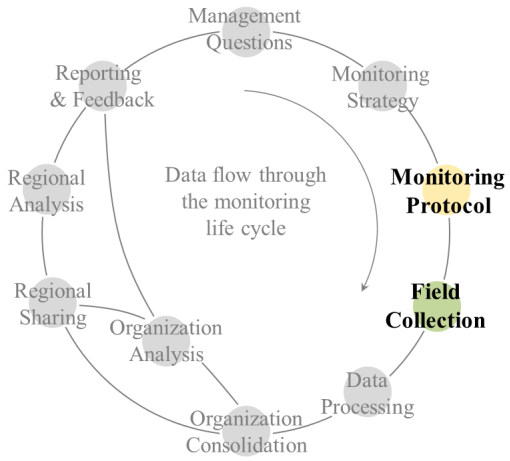
- Provides five major actions that need to be considered by data managers and their organizations
- Points to best practices and standards for data management within an organization
- Highlights the ways regional data sharing can be better implemented
- Provides examples of actions PNAMP partners are currently implementing



PNAMP promotes the adoption of best practices among its partner organizations (federal, tribal, state, and NGOs) for the management of monitoring data and, where mutual needs exist, the sharing of data. Great strides have been made in the PNW to develop strategies for the management of data across multiple programs and projects that monitor aquatic biota and their habitats. The concurrent activities of PNAMP facilitated projects and PNAMP partner projects have resulted in opportunities to share and integrate the data generated by these efforts.

Action 1. Describe protocols

Protocols set the context for collection, management, analysis, and reporting of data within a monitoring program. Without planning and documentation of protocols, data flow can be inefficient and inappropriate to the needs of the monitoring program. A fully described protocol includes description of the monitoring objectives, study design including methods and resulting metrics and indicators, and programmatic information. Protocols allow an outside data user to understand if two or more datasets are compatible for combination or if a crosswalk is necessary.



TOOLS TO DOCUMENT PROTOCOLS

In 2010, PNAMP sponsored a web-based tool, MonitoringMethods.org, to document protocols and constituent methods. This tool has the capacity to accept protocol descriptions across a range of aquatic monitoring programs. As monitoring practitioners across the region use the tool, it will not only provide a ready-made mechanism for the individual programs to document their respective protocols but will allow refinement of protocols across programs in support of data sharing.

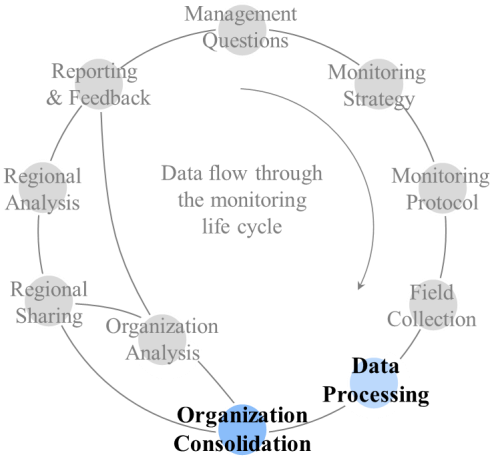
One important aspect of a protocol is the study design, which includes the spatial and temporal aspects of a project. PNAMP is currently sponsoring the development of the Master Sample tool. This tool uses the generalized random tessellation stratified (GRTS) technique to create a random subsample and spatially balanced list of sites for an area of interest. In 2009, PNAMP sponsored development of a prototype for the Lower Columbia River. In 2012, the tool will be developed as a broadly-available resource and it will facilitate documentation of monitoring locations across the region.

MONITORINGMETHODS.ORG: monitoringmethods.org/

FOR THE LATEST PNAMP INFORMATION:
Monitoringmethods.org: pnamp.org/project/3134
Master Sample Tool Development: www.pnamp.org/project/3263

Action 2. Standardize data processing

As organizations move from the early data-management model of individuals entering their own data on spreadsheets and maintaining them on personal computers, practices that are more sophisticated are becoming standard procedure. These practices include using standardized data entry forms, performing quality control checks, adhering to quality assurance guidelines, consolidating data on an agency-wide basis, prioritizing long-term maintenance of multi-user databases, and using automated and secure backup. As an organization's data management modernizes, often data professionals – such as data stewards, coordinators, modelers, and programmers – may be involved in designing and maintaining the technical features of the entity's data flow.



WAYS TO KEEP CURRENT WITH BEST PRACTICES

Data management is evolving from an individual-centered recording activity to coordinated consolidation of agency data and automated data handling and management, requiring new expertise, infrastructure, planning and standardized practices.

PNAMP recognizes that professional standards for the management of natural resource data are still evolving. A cadre of data professionals is emerging who have the training to guide organizations in identifying and applying best practices within their organization.

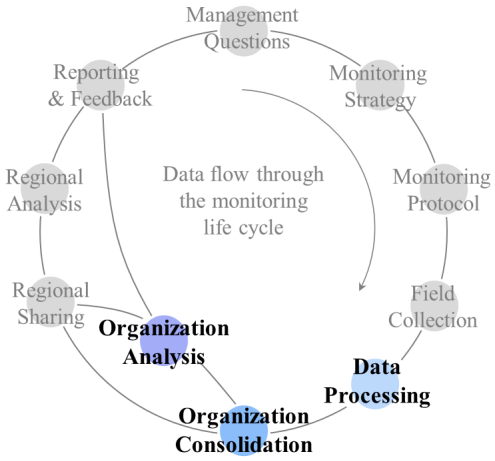
In order to support best data management practices and the emerging community of data professionals in the PNW, PNAMP has several activities planned to support a Community of Practice:

- Maintain an online database of technical articles and written resources on data management and sharing topics
- Develop an easily, accessible way for the data practitioner community to share information on current data management topics and discuss issues of interest
- Sponsor webinars and/or mini-workshops that focus on technical issues in data management and sharing

FOR THE LATEST PNAMP INFORMATION:
Data Best Practices: pnamp.org/project/3265

Action 3. Develop metadata - don't leave your dataset naked

Without metadata, anyone using a dataset will have a multitude of questions and uncertainties about the content, quality, and context of the data. National standards exist for developing a metadata report for a dataset with spatial attributes. A biological extension is available for describing the non-geographic aspects of data, including sampling methods and data creation. A new International Organization for Standardization (ISO) metadata standard is currently being released. Metadata should be applied at all levels of data consolidation from creation through consolidation and analysis.



Metadata documentation is not a commonly embraced activity throughout the region. While the need for various levels of metadata is recognized, the evolving national standards, the complexity of the standards that exist, and lack of time, expertise, and resources have discouraged widespread adaptation of this business practice.

REGIONAL SUPPORT FOR METADATA

PNAMP sponsored the 'Regional Guidance on Metadata for Environmental Data' report ,which provides an overview of the various formats available for developing metadata and a decision tree to aid in selecting a format. A 1994 Presidential executive order resulted in the creation of the Federal Geographic Data Committee (FGDC). FGDC established the Content Standard for Digital Geospatial Metadata (CSDGM) that is currently required for all datasets developed by federal agencies or through federal funding. While the word 'geospatial' often invokes the idea of a map, any measurement that has locational information associated with it has geospatial attributes. The CSDGM standard provides guidance to document a wide range of complexity within a dataset, but the required metadata elements are only a subset of all the available metadata elements within the CSDGM. The ISO standard will replace CSGDM in upcoming years. PNAMP will facilitate technical exchange to help with regional adaptation of these new standards.

To continue promoting widespread development of metadata, PNAMP is sponsoring the prototype development of a tool, the "Metadata Builder". The intent of this tool is to develop the core of a regional standard metadata report using existing online databases that document aspects of a monitoring project. The tool should provide an easier pathway for documenting a monitoring dataset including needed project level information.

PNAMP will also pursue other opportunities to promote the widespread development of appropriate and accessible metadata.

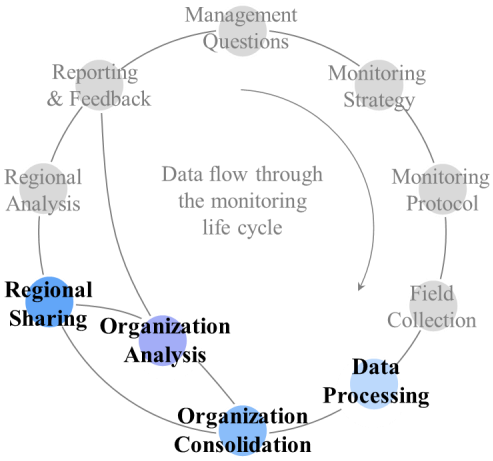
REGIONAL GUIDANCE ON METADATA:
Regional Guidance on Metadata: pnamp.org/document/2771

FOR THE LATEST PNAMP INFORMATION: pnamp.org/project/3139

Action 4. Use common terminology

Even the words we use to describe data content and types are important. Unless we agree on their meaning in context of a monitoring program, data users can make mistakes in the handling and interpretation of another program's data.

Using common glossaries is a start to lessen confusion over terminology. Another mechanism is the development of crosswalks and data dictionaries that indicate where different terms have the same meaning. Ontologies, which define the words that describe an area of knowledge and the relationships within that area, are a more formal type of controlled vocabulary. Ontologies are important to building the ability to access and discover data.



GLOSSARIES, DATA DICTIONARIES & CROSSWALKS

PNW aquatic monitoring practitioners are increasingly recognizing the need for common terminology. Several ongoing PNAMP projects are working to establish common vocabularies. MonitoringMethods.org provides a glossary of terms common to practitioners active in PNAMP. The Integrated Status and Trends Monitoring project includes a crosswalk component that compares the content and names of common measurements and metrics as collected and developed by different monitoring organizations. The PNAMP partner project Columbia River Basin Population Crosswalk Geodatabase and Online Interactive Mapping Application (implemented by Columbia River Inter-Tribal Fish Commission) is an example of another crosswalk where different salmonid population names applied to the same geographic areas are being cross referenced.

PNAMP expects additional crosswalks and common data dictionaries will be developed in the future as data sharing among agencies increases. As data management practices continue to refine, user communities may require ontologies specific to topical areas of mutual monitoring and management concern.

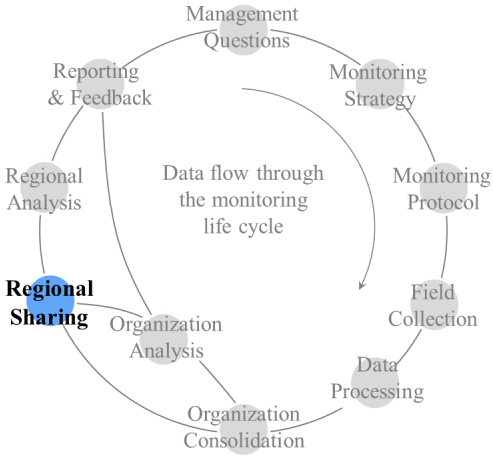
MONITORING METHODS.ORG GLOSSARY: http://monitoringmethods.org/Glossary/Index

FOR THE LATEST PNAMP INFORMATION:
Data Best Practices: http://pnamp.org/project/3265

Action 5. Plan for data sharing

The need for aquatic resource information often extends over large landscapes and involves the perspective of multiple monitoring entities. Regulatory requirements such as Endangered Species Act related Biological Opinions may require the sharing of data from different monitoring efforts. In addition, organizations often voluntarily share data with other organizations for mutual management and decision-making needs.

One mechanism of data sharing is an exchange network. Within the network, data exchange templates (DET) describe the successful 'packaging' of datasets for direct transport from one data managing entity to another. As described by the EPA "a data exchange template is a standardized format that identifies the types of information required or allowed in a particular document or data exchange." These transactions must be governed with data sharing agreements describing a commonly agreed structure of exchange, whether the network has many partners or consists of a small set of sharing organizations.



Another form of data sharing is the online presentation of spatial data supported by GIS systems. Underlying this sort of display is a geodatabase comprised of the geospatial and other attributes of the data. GIS technology has established protocols for sharing of spatial data through web services, giving this technology a current advantage for wide sharing of mapped data.

EXCHANGE NETWORKS AND INTERACTIVE APPLICATIONS

In recent years, several projects have emerged within the PNW aquatic monitoring community that promote regional data sharing best practices. The Washington Department of Fish and Wildlife and Northwest Indian Fisheries Commission is completing a Juvenile Migrant Data Exchange (JMX) for salmon in early 2012. The JMX project is an early example within the regional aquatic community of working with the Environmental Information Exchange Network. The Exchange Network guides groups in applying current technology to develop an efficient network for the exchange of high quality data among participating agencies. The Coordinated Assessments project is an ongoing PNAMP-supported effort that will result in a Columbia Basin-wide exchange network for Viable Salmonid Population (VSP) indicators and supporting metrics. Another PNAMP project, initiated in late 2011, is examining how to best exchange habitat-related indicators and metrics.

In addition to these efforts for developing exchange networks, agencies are acknowledging that there is a need for online, interactive sharing of data. The Columbia Basin Salmonid Population Crosswalk Geodatabase and Online Interactive Mapping Application, initiated in early 2012 by CRITFC, is one example of this kind data-sharing tool. PNAMP plans additional online tools for data query and access such as a monitoring site management tool. Each of these tools will provide examples of how data managers can apply geospatial technologies to meet current resource management needs for data exchange.

FOR THE LATEST PNAMP INFORMATION:
Coordinated Assessments: http://pnamp.org/project/3129
Habitat Data Sharing: http://pnamp.org/project/3266
Columbia Basin Salmonid Population Crosswalk: http://pnamp.org/project/3155