![StreamNet logo - Fish Data for the Pacific Northwest]

Data Sharing Top Ten List
StreamNet*
2010/06/15

These are the 'top ten' data management and sharing recommendations condensed from 'Considerations for Regional Data Collection, Sharing and Exchange' by StreamNet (www.streamnet.org). This list, in no particular order, represents aspects to consider and address in developing data management and sharing approaches. They are not specific instructions.

1. **Make full use of information technology in the field**

   Better and cheaper ruggedized and waterproofed tools (e.g., digital pens, PDAs, tablet PCs, etc.), when appropriately matched to field conditions, provide the capability to enter data directly in the field and avoid the workload and risk of errors from later data entry. Quality controls like range checks, specified formats, pull down lists, etc. can be built in to simplify data entry and improve data entry speed and data quality.

2. **Manage data on an agency-wide basis**

   Agency-wide data systems can leave responsibility for and control of data at the field level yet at the same time consolidate data for the entire agency. This allows for consistent quality control, helps assure data consistency, supports automated summarization or analysis at local and agency levels, enables canned reports at local and agency levels, simplifies translation of data to regional standard formats, and enables posting to data delivery systems as 'nodes'. Such systems can also facilitate development of descriptive information (see #6)

3. **Assure data quality**

   Quality assurance procedures should be defined and applied by the agencies collecting the data. These procedures should be part of initial data collection and address each step in collection, maintenance, analysis and publication. QA can be significantly enhanced and simplified when built into agency-wide data management systems and data capture tools in the field (range checks, required formats, drop down lists, etc.).

4. **Standardize data collection to the degree practicable**

   Data management and sharing are greatly simplified when like kinds of data are collected using the same methods. Although not always possible, the greatest benefit occurs when data methods are standardized within an agency, and sharing data is simplified if methods can be standardized among agencies. When this is not possible or desirable, then it is critically important to describe the methods, data definitions and coding used (see # 6).

5. **Standardize data management and sharing to the degree practicable**

   Basic standards for managing and sharing data are available and should be used. The Northwest Environmental Data-network (NED) published several best practices documents related to recording location, creating a data dictionary and developing a data management plan, which are available through the Pacific Northwest Aquatic Monitoring Partnership (PNAMP, www.pnamp.org). Consistent use of agency code lists is helpful, and sharing could be enhanced if agencies collaborate to create common coding schemes.

6. **Describe data sets so they can be located, understood and used by others**

   Although local data collectors already understand their data, it is important to fully describe the content of data sets so that new local staff members, other agency staff and users outside the agency can understand them and use them properly. Descriptive information (metadata) should accompany all data sets made available for sharing and include information on what the data are, collecton methods, data definitions, codes, locations, and contact information. Metadata contents are defined in formally adopted standards like the FGDC Biological Extension, with help available from regional groups like PNAMP, regional database projects and NBII . Agency data management systems can facilitate much of the metadata creation.

7. **Publish metadata and data on the Internet**

   Metadata for data sets that are made available for sharing should be published on the Internet as web services, a prerequisite for being able to locate he data through online clearinghouses, data portals or exchange networks. Ultimately, the data should also be published as web services. Agencies and/or projects can choose to publish the metadata and data themselves or use regional database projects to publish them, at least until agency database systems are in place. Publishing metadata will preclude the need for independent data inventory efforts.

8. **Write a data management plan**

   A written data management plan for data collection efforts will help to clarify the steps needed and to assure that the ultimate treatment of the data is considered early in project planning and not end up as an afterthought. This plan should not be long and complicated, but should provide a general overview of how the data will be managed, stored, analyzed and shared. Assistance is available from PNAMP and regional data management projects.

9. **Commit to sharing data collected using public funds**

   In general, data collected using public funding should be made available for use by others. Agencies should establish data sharing procedures, and funders may require data sharing as a contractual requirement. Legitimate limitations on sharing data should be addressed in the data management plan and in the metadata. Such limitations can be enforced through data sharing agreements when subsequent use of the data needs to be restricted, with data being released only to entities that sign the agreement. Concerns over releasing data prematurely or treatment of sensitive data should be addressed through regional policies (see # 10).

10. **Resolve issues by developing regional data sharing policy**

   A number of concerns over sharing data are already known, such as establishing timely sharing standards, restricting use of data by others before the originator has time to analyze and publish results or recommendations, or restricting release of sensitive information. There are various ways to resolve these concerns, and these and other issues should be addressed through regional policy developed collaboratively among data collecting agencies, funders, and entities that rely on the data. These can also be addressed in data sharing agreements.