

Follow up to panel discussion at the Data Management Workshop  
Bruce Schmidt

Moderator Nancy Tosta requested that panel members identify actions to achieve goals, not just discuss obstacles. Following are ideas on specific actions that would make regional data sharing feasible and relatively easy.

The question is: What capabilities would be needed in a regional data sharing system? My premise in the panel discussion was that we need to consider "the system" at all steps in the process, from data collection to regional dissemination, and I suggest that there are specific solutions at each level, all of which are essential:

1. Local / data collection level:
  - a. Design data capture tools and data input interfaces for consistent use across agency field offices. This would speed data capture, and improve data input efficiency, consistency and quality.
  - b. Assist agencies and local offices develop and maintain local databases for data management and upkeep at field offices. This would help keep responsibility for data management and maintenance with those who collected the data.
2. Agency level:
  - a. Assist agencies develop and maintain functional, integrated internal data system(s), ideally modular by type of data. This would increase data management capability for internal uses as well as make regional sharing easy.
  - b. OR, provide support for intermediary database projects to capture agency data, by type of data. This would be cheaper and require less infrastructure.
3. Regional level:
  - a. Define data collection standards across agencies so similar data are equivalent, regardless of agency. This is the responsibility of biologists and regional administrators, not data specialists. Getting independent agencies to change current sampling to adopt regional standards has serious implications and will be a challenge. PNAMP, CSMEP are working toward this.
  - b. Define data standards (data coding, formats) across agencies so that individual data records mean the same thing. This is the responsibility of data specialists AND biologists and will require collaborative involvement.
  - c. Make databases visible on the Internet, WITH metadata. This can be done in several ways, either by each agency or intermediary database projects. This will require fully functional data management systems with web servers in the agencies and their offices, or sharing data with regional scale database projects in an agreed upon format. A common regional format, by data type, will be required of either approach.
    - i. Ideally, post data in XML so they are "findable" by portals, search engines, distributed database systems, etc.
    - ii. If data are on the web, then many approaches to dissemination are feasible (DDBMS, metadata portals, search engines, etc.) This is the "easy" part if the other steps have been satisfied.
  - d. Provide a mechanism for coordinating adoption of data related standards across agencies. There are various approaches that could include the Council, a coordinating body like NED, a regional "data center", coordination of regional scale database projects, a neutral agency with significant data expertise like PSMFC, etc.) Many approaches could work if the preceding recommendations are implemented.

As pointed out in my paper, if anything prevents data flowing from the field to the final regional application, the data won't get there, so all steps must be functional. Who should do these things

needs regional discussion, with roles for the Council, NED, PNAMP, CSMEP, agencies, and the various database projects. I should mention that StreamNet is already doing several of these items, including tool and database development, hosting of data on the web for agencies, and developing regional data standards, but only for specific types of data. Actually, much of the needed effort is in place, but in an incomplete and uncoordinated way. The key thing missing is coordination and region-wide agreement on approach.