

## Data Quality: Yet Another Chapter in the Metadata Story

Lynn Yarmey, CCE-LTER & PAL-LTER

As our LTER site datasets flood in from the field, we Information Managers, as the 'humans-in-the-loop,' remain continually challenged by the basic yet broad issue of standardization for data integration. Equally important for integration efforts as attribute-level metadata is standardization of data quality assurance (QA) and control (QC), necessary as similar datasets are brought together at sites and across the community. While data quality has been coming into its own as an issue independent of, although related to, metadata and data access, it remains largely unexplored and underdeveloped.

At an international level, quality standardization issues have been discussed for years by organizations like the International Organization for Standards (ISO) and the Organization for Economic Cooperation and Development (OECD). Recently, the Digital Library movement has begun development of sensor network data assessment tools (Wallis et al.). However, there remain many unique data quality 'quirks' associated with dynamic, local science that remain undocumented. To address these, a few grassroots communities have formed with the intent of creating community best practices and data quality standards, e.g. the LTER IM Data Quality Working Group.

One more formal example of such an effort is Quality Assurance of Real-Time Oceanographic Data (QARTOD, <http://qartod.org>), which has been organized and funded by NOAA. QARTOD is a cross-agency effort driven to address issues of data quality description and standardization in the context of large-scale physical oceanography with, however, secure grounding in a local, hands-on perspective. Bringing together instrument manufacturers, data collectors, metadata specialists, funding agencies and data center representatives, QARTOD looks at the quality assurance and quality control endeavor from a variety of angles and perspectives. A QARTOD meeting report states: 'One of the primary challenges facing the oceanographic community will be the fast and accurate assessment of the quality of data streaming from the [Integrated Ocean Observing System] IOOS partner systems. Operational data aggregation and assembly from distributed data sources will be essential to the ability to adequately describe and predict the physical, chemical and biological state of the coastal ocean. These activities demand a trustworthy and consistent quality description for every observation distributed as part of IOOS.'

([http://nautilus.baruch.sc.edu/twiki/pub/Main/WebHome/QARTOD2006\\_v9.pdf](http://nautilus.baruch.sc.edu/twiki/pub/Main/WebHome/QARTOD2006_v9.pdf))

QARTOD is divided into measurement groups where initial active groups include CTD, in situ currents, waves and dissolved oxygen. Each group uses the creation of a common system of data flags as a mechanism for standardizing practices across disparate instrument types, manufacturers, sample analysis methods, etc. Meetings began in 2003, followed by two in 2005 and one in 2006. No plans have been made yet for the next QARTOD meeting, the community is awaiting funding decisions.

As the LTER sites continue to work towards improving data description and interoperability, the issue of data quality is likely to become more recognized as central to

integration. Community efforts like QARTOD provide a valuable forum for needed discussion, providing lessons learned and a realistic groundwork for others to build upon.

References:

Wallis, J.C., Borgman, C.L., Mayernik, M.S., Pepe, A. (in press). Know thy sensor: Trust, data quality, and data integrity in scientific digital libraries. European Conference on Research and Advanced Technology for Digital Libraries 2007, Budapest, Hungary.  
(<http://www.ecdl2007.org/>) (see also  
[http://polaris.gseis.ucla.edu/cborgman/pubs/wallisetal\\_2007.pdf](http://polaris.gseis.ucla.edu/cborgman/pubs/wallisetal_2007.pdf))