UNIVERSITY OF WASHINGTON

SCHOOL OF MARINE AFFAIRS

Dear VADM Conrad C Lautenbacher
National Oceanic and Atmospheric Administration
Room 6811
14th St & Constitution Avenue, NW
Washington, DC 20230

December 18, 2007

I am writing you with respect to NOAA Science Advisory Board's (SAB) recommendations on Data Archiving and Access Requirements across NOAA. These recommendations are based on the presentation of findings of the SAB's Data Archiving and Access Requirements Working Group which we established in 2006 to provide advice to NOAA on maintaining comprehensive data archives and providing efficient access to them for research and management.

With this set of recommendations NOAA SAB addresses two current archive and access issues at NOAA. The National Research Council report on archiving and access is soon to be released. NOAA SAB awaits the results of that report before addressing the additional issues of: 1) coping with data volume and data stewardship for earth-science data and for non-NOAA datasets; 2) unified interdisciplinary NOAA standards and protocols; and 3) the role of Centers of Data in archiving and access.

Issues Identified

First, multiple versions of datasets exist because NOAA archives data and data products for many reasons. However, a policy on retention of multiple versions is needed.

Second, NOAA will benefit by clarifying the roles, responsibilities, and requirements of the participating elements of the archive in which the Comprehensive Large-Array Data Stewardship System (CLASS) will be a major element.

Issue 1. Archiving Multiple Versions of Datasets

NOAA is required by law to archive its central mission data for weather, climate, oceans, solid earth and space. Data and products are archived if they require extensive processing to reproduce, e.g., output from computer models, satellite-derived products, and radar. Data are archived that support NOAA's regulatory responsibilities as in

3707 Brooklyn Avenue N.E. Seattle, Washington 98105-6715 (206) 543-7004 FAX: (206) 543-1417
uwsma@u.washington.edu www.sma.washington.edu Campus mail: Box 355685

fisheries and climate. Similarly data that are used for producing scientific assessment are archived as in the case of State of the Climate reports or data used by the Intergovernmental Panel on Climate Change (IPCC). Finally, non-NOAA data that support NOAA's missions are also archived. These archival needs occupy a significant amount of NOAA's data storage capacity and may be difficult to access.

Different versions of datasets proliferate in NOAA because they are modified as additional data become available, new scientific discoveries occur and more thorough scrutiny is applied as datasets are used. Given these facts, NOAA data archives are expected to grow exponentially. With this increase in cost there are increased costs of storage and data stewardship. Can costs be reduced if archives minimize the number of dataset versions that are retained?

If so, the question is which version of a dataset should be kept and which should be abandoned? In current practice, NOAA sometimes finds that the "needed" version is not available as was the case when a Freedom of Information Act request could not be fulfilled because the stations and data for a classic 1990 paper [Jones et al. 1990] on climate effects on urbanization were not available. Should NOAA have a retention policy for data cited in papers by NOAA authors? Is there a time limit on what data to archive?

NOAA SAB recommends that NOAA should develop a retention policy for multiple versions of databases and that policy should take into account: the benefits and costs, user needs, NOAA mission requirements, legal and regulatory constraints and National Research Council (2007) recommendations in Environmental Data Management At NOAA: Archiving, Stewardship and Access. [This document was not available at the time the Data Archiving and Access Reliability Working Group met].

A useful first step in this direction would be to hold a workshop including NOAA and its data user communities for the express purpose of developing recommendations on a retention policy.

Issue 2. CLASS and the NOAA archive

CLASS is being developed as the storage element in a NOAA archive. However, there is not a shared vision across NOAA of what elements the archive would include. This is due to the fact that CLASS has evolved over time and there is a widespread misunderstanding of the CLASS role and purpose within NOAA. This lack of understanding is handicapping development. In addition, there is a need to develop trust across NOAA among those elements whose active participation in the archive is essential.

NOAA SAB believes that a NOAA-wide archive should incorporate data-originating and data-managing elements. Such an archive allows better use of data and information to meet all of NOAA mission objectives. In order to do this, the interacting roles of CLASS, data centers, centers of data, and legacy systems must be clarified because the

are essential to the effectiveness of the NOAA archive. This requires a fresh look at the data-system architecture that best meets NOAA's needs.

The NOAA SAB recommends that NOAA should define its archive requirements and, based on these requirements, clarify the roles and responsibilities of CLASS, data centers, centers of data and legacy systems. The SAB recommends that NOAA establish an archive architecture group to define the elements and to track progress toward achieving them.

The SAB appreciates the leadership and work on-going in NOAA to resolve data archiving and access issues and to anticipate future needs as new systems come on line and new technologies develop. The SAB's Data Archiving and Access Requirements Working Group continues to assess other issues noted above.

Sincerely,

David Fluharty, Chair NOAA SAB
Wakefield Professor of Ocean and Fishery Sciences
College of Ocean and Fishery Sciences
School of Marine Affairs
University of Washington
3707 Brooklyn Ave. NE
Seattle, WA 98105

206/ 685-2518
206/ 543-1417 fax

cc: Mary Glackin
Richard Spinrad
Jack Hays
Cynthia Decker
Jim Beck