

Environmental Data Management: Challenges & Opportunities

Mohan Ramamurthy,
Unidata
University Corporation for
Atmospheric Research
Boulder CO

25 May 2010
Environmental Data Management Workshop
Silver Spring, MD

Unidata's Mission

To provide the data services, tools, and cyberinfrastructure leadership that advance Earth system science, enhance educational opportunities, and broaden participation.

We: - *Facilitate data access via push and pull systems*

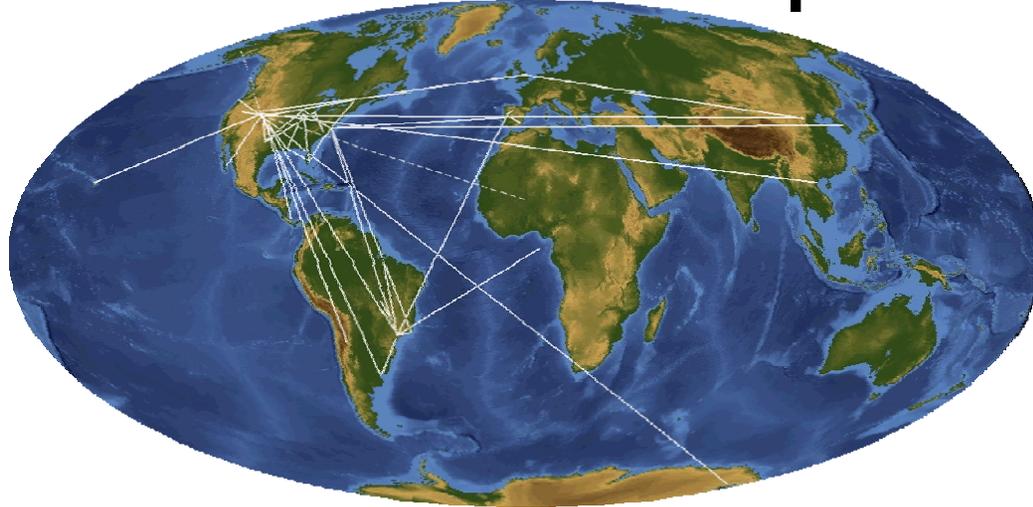
- *Develop and provide tools for data access, management, analysis, and visualization*

- *Provide support to our users, including training on our tools and technologies*

- *Help in bringing the people together to address data-related issues (e.g., data needs, agreements, facilitating discussions on standards and conventions, etc.).*



Unidata: A Snapshot



Data Access and Distribution:

Over 30 data streams
Both push and pull technologies
are used
Unidata is NOT a data archiving
center.

Software & Middleware:

Data Distribution: LDM
Remote Data Access: THREDDS,
ADDE, and RAMADDA
Data Management: netCDF and
UDUNITS
Analysis and Visualization: GEMPAK,
McIDAS and IDV

Training and users workshops

Community Seminars

Equipment Awards to universities

Intellectual commons and expertise

Unidata's engagement with NOAA

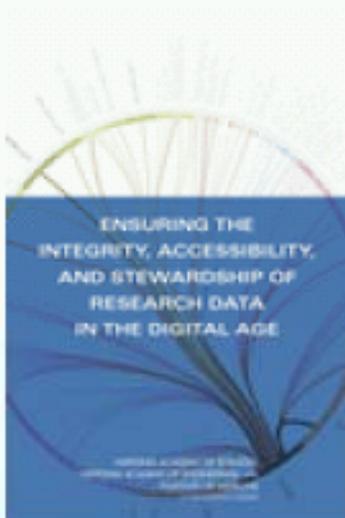
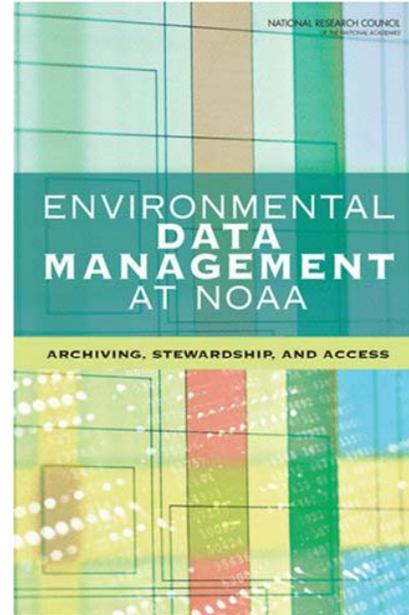
- Unidata been working with NOAA since its inception.
- In fact, Unidata was created to distribute real-time NWS weather data to universities.
- Our partnership with NOAA is a long and successful one, one that continues to this day.
 - e.g., Unidata is working with NWS and NCEP to bring AWIPS-II to universities.

NOAA-Unidata Partnership

- Unidata technologies like LDM, netCDF, THREDDS Data Server are in widespread use by NOAA entities.
 - Virtually, all NOAA data centers use Unidata technologies. We work closely with many individuals and groups represented at this workshop.
 - LDM is in operational use by NWS to distribute WSR 88-D data and NCEP model output.
 - Unidata staff serve on several NOAA committees.
 - Unidata has hosted NOAA-sponsored workshops to address data issues.

Broad Recognition of the Importance of EDM

- “Environmental Data Management at NOAA: Archiving, Stewardship, and Access.”



Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age

Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age; National Academy of Sciences

ISBN: 978-0-309-13684-6, 188 pages, 6 x 9, paperback (2009)

Need for Partnerships

- Many organizations are trying to deal with Environmental Data Management.
 - i.e., NOAA is not alone in addressing this issue;
- All stakeholders - data collectors, providers, consumers in all sectors - have a role to play.
- Adopt, adapt and only as a last resort develop (technologies, tools, and ideas)

Linking EDM to Mission

- Data management is not the end, but only a means to an end.
- 64K question: How can NOAA's EDM best support its mission?
 - Focus on facilitating and enhancing effective data services for all consumers of environmental data, both within and outside NOAA.

Science, Education and Workforce Development

- **A Plea:** Even though NOAA is a mission agency within the Department of Commerce, please do not relegate attention to science, education, and workforce.

The Ultimate Objective of a Data Service

- Providing the right data in the right format at the right time for ready use
- Finding the right data is almost as important as providing the data
- Users don't care as much about technology or organization as they do about transparency and usability;

Source: Peter Cornillon (2003)

Overarching Challenges

- 1) Technological
- 2) Organizational
- 3) Cultural

2) and 3) are the most difficult to address

Challenges: Specifics

- Data discovery
- User community is not monolithic
- Lack of complete (or even sufficient) metadata, including information on data quality.
- Too many formats, types, scales, etc.
- Too much burden on the end user to find and effectively access and use data

Challenges: Interoperability

- Avoiding stovepipes and silos;
- Establishing abstract service interfaces;
- Supporting flexible composition and chaining of services;
- Picking and implementing the right standards and conventions;
- Contributing to development of important standards and conventions.

Challenges: Data Stewardship

- Preserving valuable and irreproducible data;
- Anticipating and providing all the metadata that will be needed for use of such data into the far future;
- Providing provenance information throughout the data lifecycle.
- Finding a way to ensure adequate funding for data preservation and data recovery.

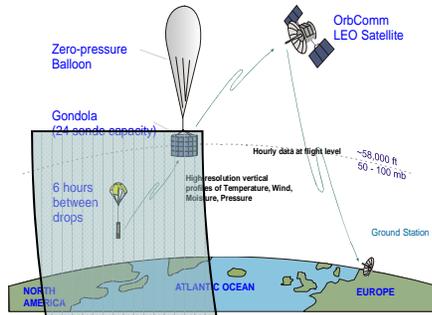
Challenges: Avoiding Technical Obsolescence (Both Hardware and Software)

- Funding some risky projects on the leading edge that use cutting edge technologies;
- Reacting quickly when technology changes or when you made a mistake;
- Evaluating innovative technologies to attract and keep enthusiastic employees;
- Encouraging professional development.

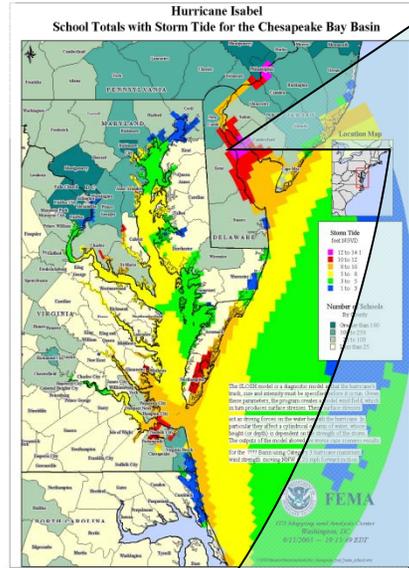
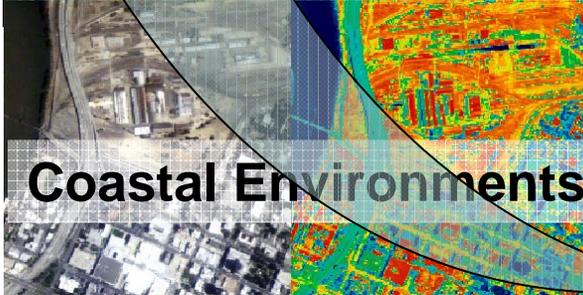
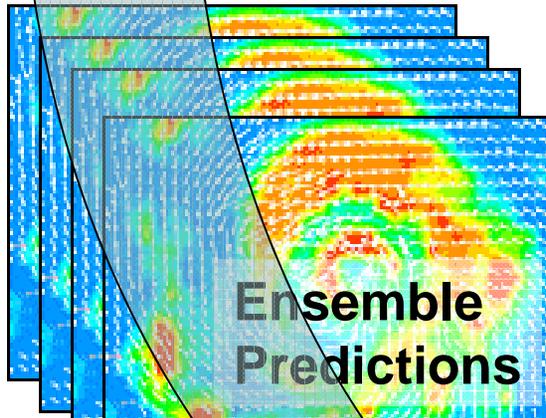
Significant Opportunities

- Helping to meet the environmental data needs of society (and every sector)
- Enabling Climate Science & Education and Climate Services
- A society and workforce that is ever more IT literate
 - There are nearly 5 billion mobile phones in the world; mostly will be powerful, roving multimedia computers

End to End Data Services



**GIS
Integration**



Need integrated services

Another Significant Opportunity

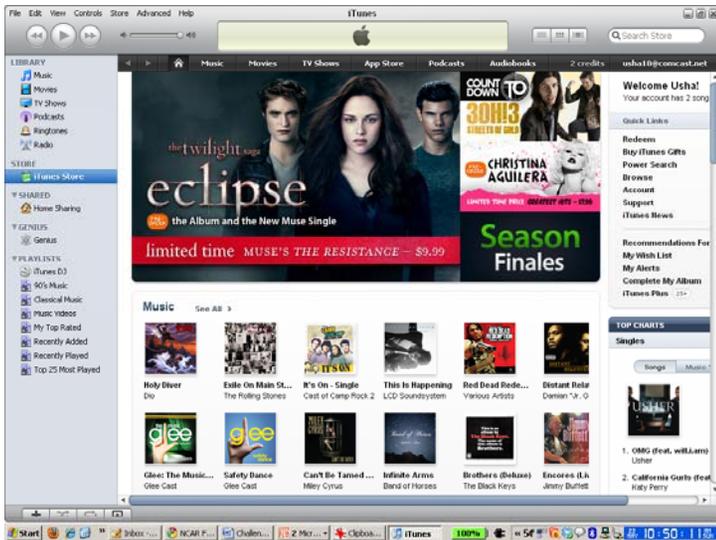
- NSF is going to ask every proposal for a Data Management Plan;
- This is a watershed event, even though it has been long time coming;
- This is a significant opportunity for NOAA

What initiatives are important for NOAA to be aware of and engaged in?

- Cloud Computing
- Real-time and Semantic Web and related services (Web 3.0)
- Data Mining – we have just scratched surface
- GPS-enabled devices (i.e., smartphones)
- Scientific Workflows
- Conventions (e.g., CF) and Standards

Seamless Access to and use of Data

We can learn from these successful implementations



- Anytime
- Any place
- Anywhere
- To Anyone



Thank you!

- Questions?
- Contact information: mohan@ucar.edu
- www.unidata.ucar.edu

Why EDM is Important?

- Data is the lifeblood of science.
- Almost every sector of society depends on access to and use of environmental data.
- Advances in science and technology are resulting in increasing volumes of data generation from diverse sources and consumption by a broad spectrum of users and applications.
- While there are many challenges, there are also significant opportunities as the world ever-increasingly relies on environmental data.

Global Participation

- Our science and operations are most certainly global in nature, so global participation and networking is crucial.
- When practical, avoid parochial data management efforts that only focus on the U. S.