

The Truth about Data Management

Sigurd W. Christensen and Les A. Hook
NARSTO Quality Systems Science Center

Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN, U.S.A. 37831-6335

NARSTO 2000
Tropospheric Aerosols: Science and Decisions in
an International Community
Queretaro, Mexico

October 25, 2000



Broad-Brush Thoughts...

***Primarily for Project Planners
and Project Funders***

- **DATA MANAGEMENT NEEDS ARE USUALLY HARD TO DETERMINE**

- *What does “data management” mean? (Not well understood)*
- *What are the overarching requirements? (e.g., OMB Circular A-110; NARSTO)*
- *What data management will be needed for the program to meet its objectives, when, and why?*

- ***IMPORTANT IF DATA NEED TO BE SHARED, INTEGRATED, AND ARCHIVED FOR FUTURE USE***

- *Accomplished by*

- *Funding agency establishes requirements and framework*
 - *Include in RFPs, SOWs, Contracts' Terms and Conditions*
 - *Coordination needed among data managers and Pls to determine specifics*
 - *Data managers support Pls*

- **DOCUMENTED DATA NEEDED LATER IN PROJECT, BUT WILL ONLY BE THERE IF PLANNED FOR VERY EARLY IN PROJECT AND COLLECTED WITH SUSTAINED EFFORT**

- *Pls can generally work with undocumented data just fine*
- *Integrated analyses using data from multiple Pls need “managed” data*
 - *Metadata to describe to external user*
 - *Common syntax for sites, variables, units, flags, and methods aids integration - enables focus on analyses rather than on fixing data and making them consistent*

- **SEVERAL LEVELS OF DATA MANAGEMENT**

- *Investigator-managed files*
- *Web-based search and file access system*
 - *Minimally managed FTP site*
- *Central integrated project data base*
 - *Facilitates QA over larger spatial and temporal scales*
- *Permanent archival*
 - *Complete metadata, good documentation, consistent data format*
- *Costs?*
 - *Reasonable: 10-20 % of project cost*
 - *Typical : < 5 % (therefore, compromises)*

- **DATA MANAGEMENT HAS SUBTLETIES AND REAL CHALLENGES**
 - *Must document what, where, when, how, who, and why*
 - *Even simple “when” often has issues of time zone and daylight vs. standard time*
 - *“Format wars”*
 - *Pls use table/spreadsheet (“wide”) format*
 - *Most fully integrated data base systems use “narrow” format*
 - *Requires interfaces, views, ongoing support*
 - *Facilitates assembling data in arbitrary ways*
 - *Going back and forth requires transposition - harder than it sounds*

- ***INVOLVE EXPERTS EARLY AND OFTEN FOR PLANNING***

- *Subject matter experts*

- *PIs*

- *Data integrators/analysts/modelers*

- *Data managers*

- *Field technicians*

- *Systems Analysts*

- *Data Coordinators*

- **ENSURING SUCCESS**

- *Initially provide strong programmatic and adequate financial support*
- *Maintain strong support, involvement, and oversight through to implementation*
- *Leverage resources across similar projects and programs*
- *Broaden the base of users*
- *Accept some compromises*
 - *An affordable data management strategy cannot be all things to all people*

- ***BOTTOM LINE: PLAN AHEAD***
 - *You make the rules*
 - *Make sure your expectations are clearly stated and understood*
 - *Involve experts to help you make decisions*

REFERENCES

- ***NARSTO Quality Systems Science Center Web Site:***
<http://cdiac.esd.ornl.gov/programs/NARSTO/>
 - *NARSTO Data Exchange Standard*
 - *Data and Information Sharing Tool (DIST)*
 - *Quality Assurance Guidance*
- ***NARSTO Web Site:*** ***<http://www.cgenv.com/Narsto/>***
- ***Permanent Data Archive Web Site, Langley DAAC:***
<http://eosweb.larc.nasa.gov/>