

Data Management Planning:

Planning for an Interagency Climate Portal



Tom Karl

*Director, National Climatic Data Center
Chair, Global Change Research Program*

Overview

- Vision for an Interagency Climate Portal
- National Climate Assessment (NCA)
- NCA Data Management Planning
 - Considerations
 - Requirements
- Approach
 - End-to-end data management
 - Earth System Grid Federation
- Challenge

Calls for an Interagency Climate Portal

- **National Climate Adaptation Summit**, called for by John Holdren, Director of the White House Office of Science and Technology Policy, issued “seven priorities for near-term action” in 2010, including
 - “**Creating a federal climate information portal.**
This would provide single-point access to data from all relevant federal agencies and programs and would evolve over time into a more ‘national’ portal with information about relevant non-federal efforts.”
- Also, **Executive Office of the President**, Committee on Environment, Natural Resources, and Sustainability (CENRS) Roundtable on Climate Information Services called for
 - an **interagency climate information portal plan**, focusing on National Climate Assessment

Where to Begin?

- A National Climate Portal as envisioned is an immense task
- Use the National Climate Assessment (NCA) to focus the initial scope
- The NCA can be...
 - A starting point, extensible to interagency data
 - A driver of requirements for Data Management Planning: transparency, access, usability, data quality, and business rules
 - A test case for connecting robust and federated data structures with appropriate user experiences

About the National Climate Assessment

- A product AND a sustained process
- More stakeholder engagement and contributions
- More transparency, traceability, accessibility and usability
- More data, more authors, more users
- A suite of Indicators (discrete and composite, physical and societal)
- Next report due in 2013, draft due in 2012, Federal Advisory Committee just formed
- Snapshots plus frequent updates
 - More frequent interim reports, updated indicators and fresh web-content are expected... BUT
 - An assessment is necessarily a snapshot in time and should be reproducible

NCA Data Management Planning Considerations

- **External Data**

- Clear need to move beyond government data and the peer reviewed scientific literature as information source, but who guarantees access and a continuous archive?

- **Quality**

- Information Quality Act requirements (highest level)
- Peer review requirements (highest level)

NCA Data Management Planning Considerations

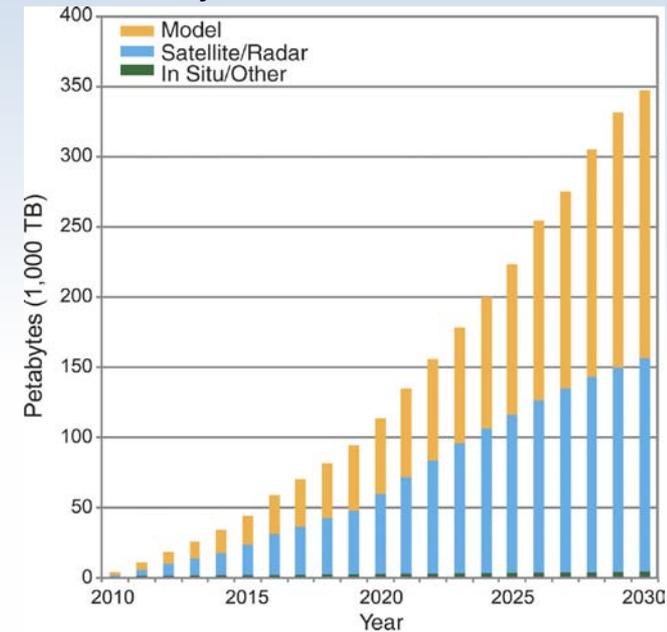
- **Volume**

- Enormous volume growth
 - E.g. physical data growth driven by satellite/radar and model data

- **Kinds of data**

- Physical climate data
 - Model data, reanalysis, satellite...
 - e.g. downscaled and right-scaled data needed and new model runs (CMIP5)
- Also socioeconomic, biophysical, ecological...
- Original data (e.g., satellite and in-situ observations)
- Synthesized data (e.g., model data, indicators, GIS application)
- Interpreted products (e.g., regional assessments)
- Also will need to consider ‘new’ types of data such as: State and local government, public utilities, traditional knowledge, operational data, insurance risk data

Example of projected growth in “Physical” climate data



J T Overpeck et al. Science 2011;331:700-702

NCA Data Management Requirements

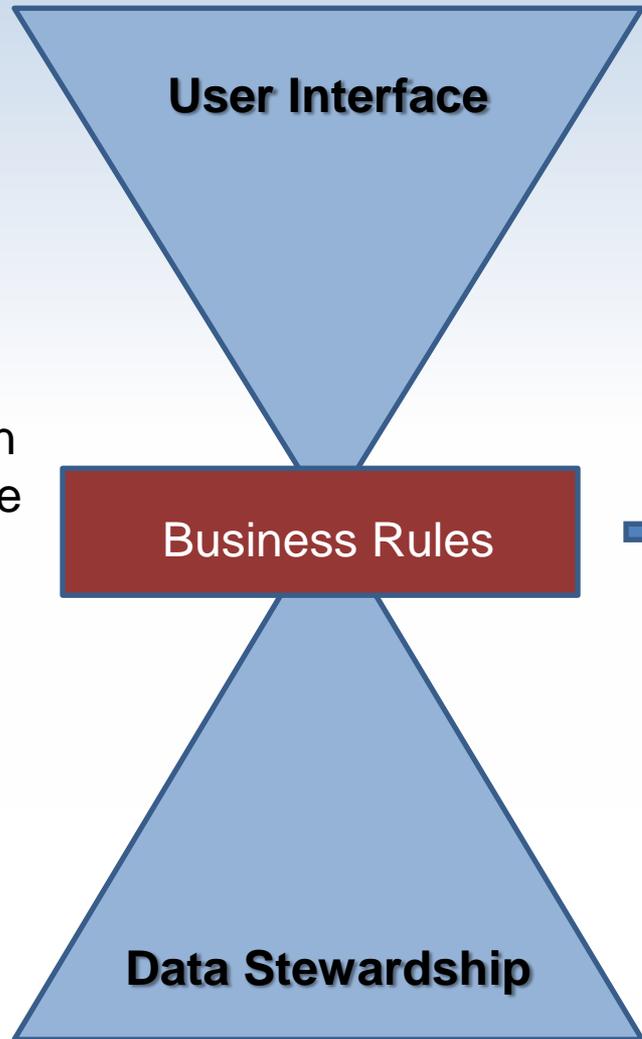
- NCA will likely not ‘host’ basic data, but will provide central access to federated system (critical questions remain for external data)
 - Highly-processed, derived data used in the report may be centrally hosted
- Must accommodate huge data sets
 - Too big to move, too big to host all data in one location
 - Implies distributed data architecture, federated storage
- Metadata must adhere to common standards
- Uncertainty must be documented
- Quality and original source of data must be clear and peer reviewed
- Data system must be interoperable
 - Use of data.gov must have clear guidelines
- Narrative assessment and graphics must link to source data and information

Approach: End-to-end Data Management

User requirements should drive business rules

Missing Layer
Making the connection between user interface and complex data system – different communities of effort

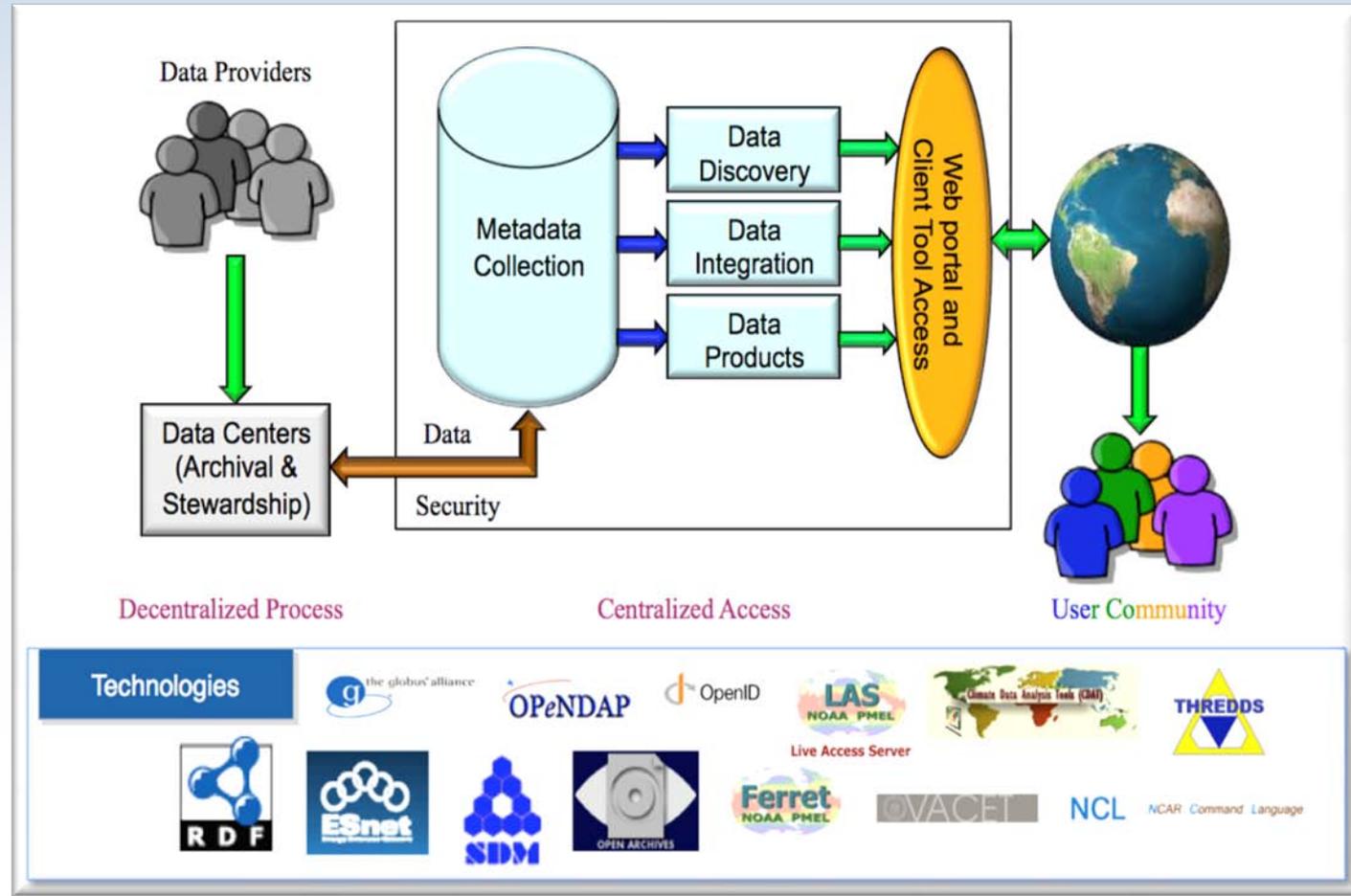
Federated data system



- What data sets?
- What formats?
- What time periods?
- How to cite data consistently?
- How to document data reviews?
- How to ensure traceability from key findings to data?
- How will we curate the data used on the assessment?
- How will we deal with non-traditional data?

Approach: The Earth System Grid Federation (ESGF)

- Too much data to move, must leave in place
- Promote sharing of knowledge, software and tools among partners
- Define APIs and protocols for interoperability among data centers
- Collaborative development of some software components
- Deployment of common software infrastructure



Data Management Planning Challenge

- Develop an implementable strategic plan for the immediate data management demands of the NCA, and the longer-term interagency/ National Climate Portal

Thank you



Tom Karl

*Director, National Climatic Data Center
Chair, Global Change Research Program*