MATTHEW MAYERNIK
PROJECT SCIENTIST / RESEARCH DATA SERVICES SPECIALIST
NCAR LIBRARY / UCAR INTEGRATED INFORMATION SERVICES
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH (NCAR)
UNIVERSITY CORPORATION FOR ATMOSPHERIC RESEARCH (UCAR)
MAYERNIK@UCAR.EDU

PRIMARY RESEARCH OR PRACTICE AREA(S):
• DATA CURATION
• DATA PUBLICATION & CITATION
• METADATA

PREVIOUS EXPERIENCE
• PH.D. UCLA – INFORMATION STUDIES

RELATED WORK (PROJECTS SPECIFIC TO WORKSHOP WITH WEB-SITES)
• Data citations within NCAR/UCP (http://dx.doi.org/10.5065/D6ZC80VN)
• **P**eer **RE**view for **P**ublication & **A**ccreditation of **R**esearch **D**ata in the **E**arth sciences
(PREPARDE, http://www2.le.ac.uk/projects/preparde)

DAS**P**OS Data and Software Preservation
for Open Science

CONTACT INFORMATION:
P.O. Box 3000
Boulder, CO 80307-3000
SURVEY OF COMMONALITY WITH OTHER DISCIPLINES
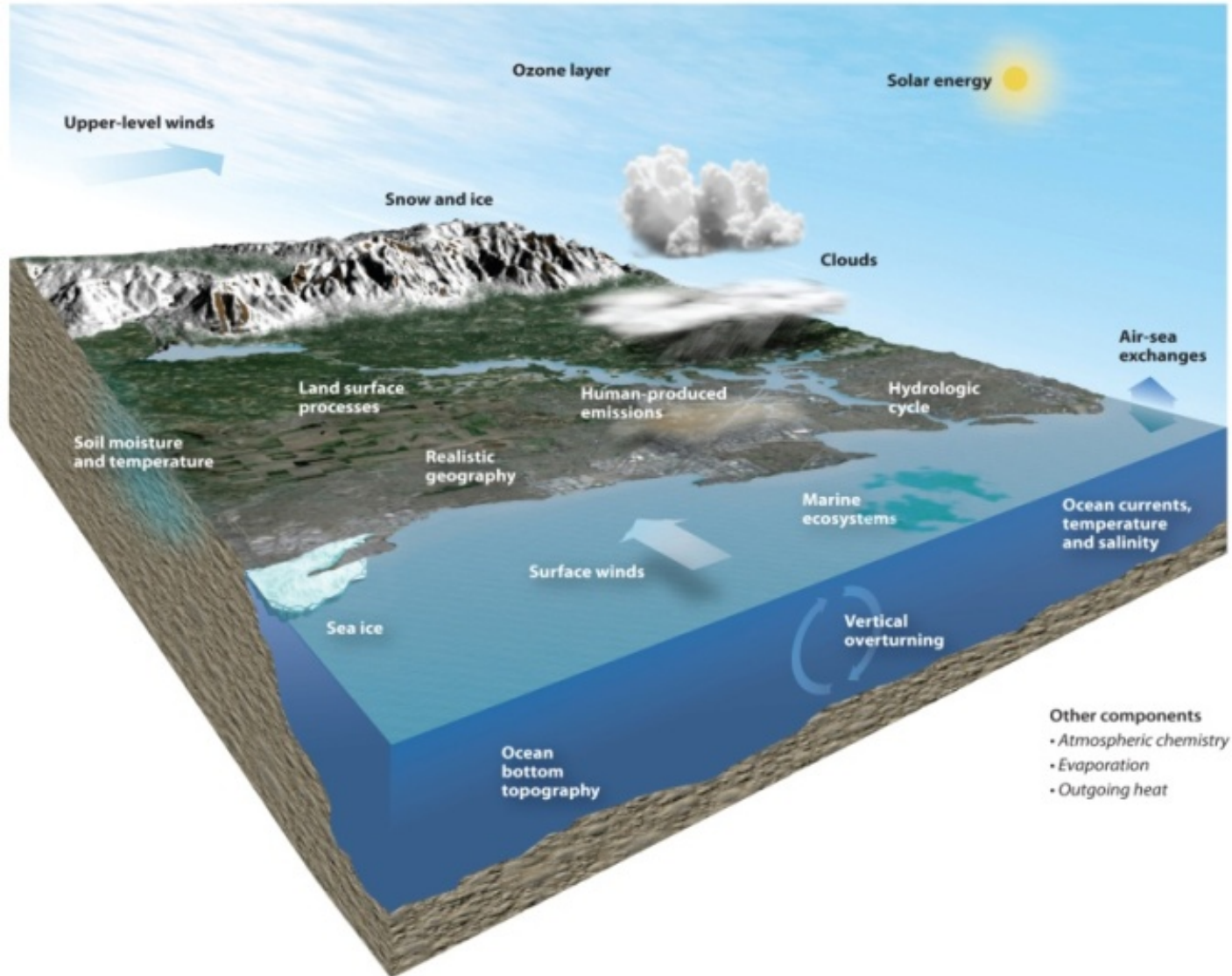**WORKSHOP 2 – JULY** 25, 2013
INDIANAPOLIS, INDIANA

Geosciences

- Some very robust and high visibility data collections
    - Atmospheric/Oceanic: NOAA, NCAR
    - Geophysical: USGS, IRIS
    - NASA: Many DAACs
    - International: DKRZ (Germany), ECMWF (UK), NERC Environmental Data Centres (UK)
- Some established and widely used standards
    - Data – NetCDF, HDF, GRIB, BUFR, SEED, SAC…
    - Metadata – ISO (19139, 19115, 19119), FGDC, GCMD…
- Most importantly: tremendous diversity!

NCAR/UCAR Scope

NCAR/UCAR Computing Facilities

NCAR/UCAR Data Services

http://www2.ucar.edu/research-resources/data-archive-services

1. **Guidelines – Near term**

   As a service, create and maintain community-developed data and software management guidelines.

   ➢ Too many *ad hoc* systems

   ➢ Complete systems: standards-based, sustainable, cost effective

   ➢ Understand Principles:

      ➢ Preservation

      ➢ Data lifecycle

      ➢ Data and metadata standards

      ➢ Data management planning

      ➢ Provenance

      ➢ Data citations and DOIs

2. Archiving and Access – Long term

Create, adapt or identify an archiving and access system for research data and software that need, but do not currently have, sufficient, secure and publicly accessible repositories.

- ➤ Too many orphan datasets – not in managed repositories
- ➤ Scholarly publications and data, rightfully so, are becoming more tightly linked, e.g. DOI's
- ➤ Cost is a major consideration
- ➤ Flexible systems; what is the appropriate level of service?
- ➤ Assess what we have, then build or adapt to meet additional needs

Near term: quantitative and comparative assessment of existing systems, and their capacity for expansion

3. Discovery – Long term

Create and maintain a unified and flexible system for discovery of UCAR publications, data, software, and services.

- Past, did well with centralized method for <u>data</u> (Community Data Portal), 10-yr old effort
- Need a new approach:
  - Expand, more data, publications, more software, and services
  - Convert to a distributed method
  - Add richness to the metadata standard
- Sustainable!

Near term: Pilot projects that connect pairs of systems

4. Preservation – Long term

Develop appropriate digital data preservation solution(s).

➢ Responsibility to preserve digital assets related to science

➢ Easily overlooked

➢ Cost is not always considered

➢ Need a suite of approaches

   ➢ Dedicated UCAR archives

   ➢ Cloud storage solutions (at UCAR or commercial)

➢ Challenge: sustaining management personnel and/or developing reliable self management tools

5. **External Integration – Long term**

   Prepare UCAR systems for greater integration with external distributed and federated systems.

   ➤ Federated? Data is mutually <u>discoverable</u> and <u>accessible</u> from multiple data systems.

      ➤ From a single point a user can reach into multiple systems, either interactively (GUI) or interoperably (scripted, web service)

   ➤ Future, standards are key, participation is a must

   ➤ Challenges – many

      ➤ Sharing storage access and computing

      ➤ Authorization, authentication for users: security concerns

      ➤ Managing large numbers of datasets

   Near term: increase people involved with developing federations – progressively track and report on trends, etc.

6. Make Data Open and Machine Readable the New Default at UCAR - Long term

Make data accessible and machine readable by publishing an API(s) and providing a data service to the general public, entrepreneurs, policy & decision makers, and as supplementary data for our scientists in the field.

➤ Exploit data for mobile and web applications

➤ Establish and maintain robust API

➤ Collaborate, extensively with public and private developers

➤ Supports societal needs, improves UCAR name recognition

Near term: Conferences, workshops, or visitor programs which bring together data providers and application developers.

Conclusions:

➢ Many improvements to demonstrate our leadership in data services!

➢ Challenges:

    ➢ Finding resources, largely human

    ➢ Setting priorities – cannot do it all

    ➢ Developing the most fruitful implementation plans

    ➢ Creating sustainable methods and procedures

    ➢ Things will change – are we planning in a flexible/adaptable manner?

➢ These data service issues are not exclusively at UCAR, they are in all organizations to some degree.

Thanks

- UCAR Information Technology Council, Data Services Working Group
  - Steve Worley – Lead
  - Matt Mayernik – co-lead
  - Mike Wright
  - Steve Williams
  - Gary Strand
  - Peter Schmitt
  - Marcos Hermida
  - Eric Nienhouse
  - Kelly Keene

Email:

mayernik@ucar.edu