Plans for Data Support, Oct 2000, Part 2
(2000 – 2002 and Road Maps)

- These are slides about data for the computing division (SCD) meetings in Sep and Oct 2000.

- Part 1 of these slides is in document RJ0093, 50 pages
  - This has nearly all of the slides

- This Part 2 has just a few new slides plus a few repeats

- Ready Dec 19, 2003 (12 pages), document RJ0315

Roy Jenne
Dec 19, 2003
List of Some Slides in Part 1, But Not Given Here In Part 2

Roy Jenne
Dec 18, 2003

- Work done in DSS under contracts
- Accomplishments during Sep 1999 – Oct 2000
- More DSS accomplishments FY2000
- A 35-year history of DSS projects
- History chart
- Working style for DSS
- We use computer services developed by others
- DSS is a baseline data system
- The defining work for the baseline archive
- Problems to solve to deliver good data support for research
- Necessary for data success
- Important data tasks for next 1-5 years
- Some new data happenings
  - Africa, CLIVAR, etc.
- NCEP 5-day 500 mb forecast scores, 1948 – 1998
- Examples of good data work around the USA
- What separates good data systems and ones that are prone to problems
- Why has our US – USSR data exchange been a success, 1972 – 2003?
- Should the mesoscale data job come to NCAR?
- Use tech gains to help users
- Cope with errors to permit a 50-year archive
- Not enough data backups
- Progress in scanning documents (in Oct 2000)
Road Maps for Data Support

1. Update the archives and obtain new datasets
   - Keep the main data flows coming from NCEP
   - Input data from three reanalysis projects
   - Add about 20 new datasets per year
   - Handle the mesoscale GCIP data

2. Keep providing the main data services

3. Provide more observations for reanalysis
   - Give better access to the observations
   - Also help the CLIVAR program

4. Expand COADS and other ocean datasets
   - COADS (1854 – 1997) is the world’s surface ocean dataset
   - We have about 70 other ocean datasets; add more

5. Provide data services for CEDAR (data 70 – 1000 km)

6. Assemble more metadata: Make it available
   - Write more information about reanalysis observations
   - Prepare more information on other datasets
   - Carry out our document management project

7. Technical developments to improve data delivery
   - Improve our Web delivery of data and information to users
   - Improve bulk delivery of data

8. Maintain mechanisms to obtain more data
   - Keep NCEP data going
   - Use exchanges such as US-Russia, and NCAR-China
   - Use scientific collaborations

9. Do some work on data rescue
   - Do our share of the work (limit this)

Roy Jenne
Apr 2000
Roadmaps for Research Data

What's New by 2006?

1. More data
   - Many updates; new datasets
   - Keep data feeds going
   - Open up ISCCP data

2. New reanalysis
   - Okay for ocean work
   - Better clouds/radiation

3. Some data packages
   - For African research
   - For Latin America
     - 70% done by 2006
     - Precip better, but poor

4. New technology
   - Use this
     - Tapes
     - CD-ROMs
     - Disks
     - Nets
     - Lots better

5. Types data delivery
   - Bulk delivery
   - Mid-scale access
   - Small-scale access
   - Web services
   - Better features on all of these.

6. Bulk data delivery
   - Costs down a bit.

7. Long-term archives
   - Data backups
   - Data grouping
   - Error control
     - Better procedures
       - Small software use
       - Reduced complexity
       - Better error control

Roy Jenne
23 Oct 2000
Aspects of Data Management

A. Some people tell us:
   - We can handle data retail
   - But we need help with data wholesale

B. Our DSS methods have been the best:
   - To make large scale research possible
   - To make mid-scale research work
   - To avoid timing and cost problems

C. So we use a balance
   - Do the data work to make any data use possible
   - Handle mid-scale and bulk access well
     - Do not use methods that hurt this
   - Offer some micro-scale data access
     - But be more selective about this

Roy Jenne
9 Oct 2000
A Success Story: Distributing Lots of Data

- NCAR DSS has been a leader in distributing lots of data
  - How does it work?

- We put ease-of-use high in priority
  - Do not force all users to learn a big software system
  - Do not force all users to learn a complicated format system
  - Use simple data access routines to hide unpack complexity
  - Give smaller datasets in easy ASCII formats

- We work hard to drop the costs for users
  - And also drop the real costs

- And we note
  - If we don’t have the right data—then we can’t distribute it.

See RS0093 from data use plots

Roy Jenne
6 Oct 2000
Data Distribution for Next 10 Years

- The main struggle is to:
  - Have the right data
    - And knowledge about data
  - Gather the right data products
    - Subsets, statistics, etc.
    - We will do part of this work

- Prepare the good macro data structures...
  - That permit good data access

- Keep improving mid-scale and bulk-access
  - Technology and ideas give opportunities

- Special help for some parts of the world
  - Help research in Africa
  - Help research in Latin America

Roy Jenne
13 Oct 2000
* Maintain the main flows of input data

and

* Obtain new datasets

and

* Do appropriate data checks

Roy Jenne
April 2000
Aircraft Reports Each Day, NCEP, USAF, Navy
(for 1973 - 1991)

Figure 1. Number of world aircraft reports per day for NCEP alone, and for the combination of NCEP, USAF, and Navy reports with duplicates eliminated.
Do More Data Work

- Add more data for reanalysis
- And get more data in place for CLIVAR
  - Needs reanalysis output
  - Needs data products from observations
- Plus much more

Roy Jenne
13 Oct 2000
There Is Trouble for Data Systems
(When they do this)

- Set up for micro data access, and...

- Don’t handle mid-scale access well

- Don’t handle bulk access well

Roy Jenne
9 Oct 2000
How is DSS Unique?

1. DSS has a large archive with many data types
   — Like some big federal archives
2. But a small staff
3. Have fast methods to handle large datasets
4. Support very large user projects (plus small ones)
5. Coupled tightly with research users
   — Staff has discipline expertise
   — Give personalized service
6. Some groups prepare products for us, ISCCP, climate models, vegetation, etc.
7. Best set of global and regional analyses anyplace
8. Efforts to locate and describe world archives
   — Wrote texts; national and WMO inventories

Roy Jenne
Feb 1995