From Sensor to Archive: Observational Data Services at NCAR’s Earth Observing Laboratory

Mike Daniels
Computing, Data and Software Facility
NCAR/Earth Observing Laboratory
NSF Aircraft operated by EOL

C-130 near Kiribati (Christmas Island)

G-V over the Colorado Rockies

www.eol.ucar.edu
NSF Radars and Lidars operated by EOL

ELDORA airborne radar in Guam

S-PolKa radar in the Maldive Islands

www.eol.ucar.edu
NSF Surface Instrumentation, Profilers and Sondes operated by EOL

Tower setup in Owens Valley, California

Wind profiler in Miami (mobile and fixed)

Upsondes (mobile and fixed)

Dropsondes (from various airborne platforms)

www.eol.ucar.edu
Other supplementary field project data (domestic and foreign)

Radar networks

Satellite Data

Model Output

Mesonets

www.eol.ucar.edu
Expanding the Internet to New and Unusual Places

Kiribati (aka Christmas Island)

Aircraft

Balloons

www.eol.ucar.edu
Through comprehensive data services, our mission is to improve the quality of data files and datasets and maintain the integrity of the associated metadata, software, and infrastructure.

CatB: “Given enough eyeballs, all bugs are shallow.”
Data Citation and Persistent Locator Issues
Difficulty in Tracking Publications

It’s important to show the results of and types of research accomplished with our facilities and data.

- Process: manually search Web of Science and other online citation indexes for text such as “C-130”, “S-Pol”, “INDOEX”, etc.
- Very labor intensive
- In 2007, this took approximately 4 months of work

Updated metrics 1997 to 2007
DOI as a service? The Field Catalog: NSF’s Internet Hub for Field Projects

Mission Summaries and Status Reports

Model and Forecast Products

Research Products

Operational Products

…with links to Google Earth visualizations, data files, chat, forums, etc.

www.eol.ucar.edu
DOIs for Software? EOL’s Diverse Platform Support

Diverse Software Engineering:

- S-Pol Radar
- Integrated Sounding System
- G-V
- C-130
- Scanning Atmospheric Backscatter Lidar
- Electra Doppler Radar
- Python Environment for Radar Processing
- Reorder and translators
- Raman-Shifted Eyesafe Atmospheric Lidar
- Scientific Data Processing
- CP-2 signal processor
- NIDAS
- Aeros
- FORAY
- AVAPS
- Integrated Surface Flux System
- NIMBUS
- Multiple Antenna Profiling Radar
- Solo Radar Editor
- Driftsonde
- Ncplot, ncpp, xpms2d
- Rapid Dow
- Zebra
- NEXRAD
- Ka Band Radar
- TDLS
- GAUS
- Field-Programmable Gate Arrays

Data acquisition software, real-time displays, analysis software, communications (satcom/WAN) and QC tools for data EOL manages

*Engineered software and data systems that are at the heart of state-of-the-art observing systems*
DOIs for Data: CDS Data Management

FTP
Email
Web
Ingest Streams
Media
Field Catalog
Snail Mail

Various Dataset Sources

QA/QC Processing and Tracking

“value-added” Composites and Format Conversions

Archive and Distribution
DOIs for data acquired from non-EOL sources: e.g. supported arctic projects

- Surface Heat Budget of the Arctic Ocean (SHEBA)
- Western Arctic Shelf-Basin Interactions (SBI)
- Arctic System Science (ARCSS)
- Arctic Transitions in the Land-Atmosphere System (ATLAS)
- International Tundra Experiment (ITEX)
- Cooperative Arctic Data Information System (CADIS, A-CADIS)
- Bering Sea Projects
  - Bering Sea Integrated Ecosystem Research Program (BSIERP)
  - Bering Sea Ecosystem Study (BEST)

We also support hydrology, oceanography and atmospheric chemistry data via outside funding
What granularity for EOL DOIs and when are they issued?

• Given a large project with aircraft, soundings, radars, model output and satellite data do we:
  – Assign a DOI for each data file?
  – Assign one DOI for all datasets for the project?
  – Assign separate DOIs for datasets from each major platform?
  – What about ancillary data? Do we assign DOIs or does the providing institution?

• We are thinking to assign DOIs for each major platform data associated with the project (e.g. C-130, S-Pol), outside datasets that we have “value-added”, and data for which no DOI exists

• It may be beneficial to only issue DOIs when processed data are released so as to prevent pubs from referencing preliminary data
What text (esp authors and contributors) do we use for data citations?

- Main Project PIs have spent years in forming this research project (authors)
- Others spend much time acquiring data in the field project phase (contributors, maybe list EOL itself?)
- Our group (CDS) will be responsible for maintaining the integrity of the DOI itself (publisher)
- It might be beneficial for data quality and distribution experts to be listed so they can be consulted for dataset questions (contributors)
- This decision is expected to require Lab Management approval
How far back do we go? EOL Legacy Data Projects

- Data and Metadata Rescue
- Collect and Digitize Project Documentation
- Digitize items such as Aircraft Videos
- Media and Format Migration (e.g. GENPro to NetCDF)
- NCAR Iron Mountain Archives
- Sounding Legacy Project (w/CSU)
- T-28 Aircraft Facility Archive
- Long-term Archive (e.g. MSS to HPSS transition)
Thank you!

www.eol.ucar.edu

Mike Daniels, EOL/CDS
daniels@ucar.edu