Data Thinking Before Data Crunching

Grace Peng
CISL Research Data Archive

SEA Conference 2016
April 5, 2016

Questions

- Can this problem be answered with data?
- What type of data can help me answer this question?
- Where can I find the best data for the job?
- Does the data match the data documentation?
- How do I cite the data for reproducibility? DOI

Meta: Can this problem be answered with data? Should it?

- Is it measurable?
- Or related to something measurable?
- Is it technically possible?
- Ethics
  - Privacy
  - Attract attention from wrong places

Define: What type of data can help me answer this question?

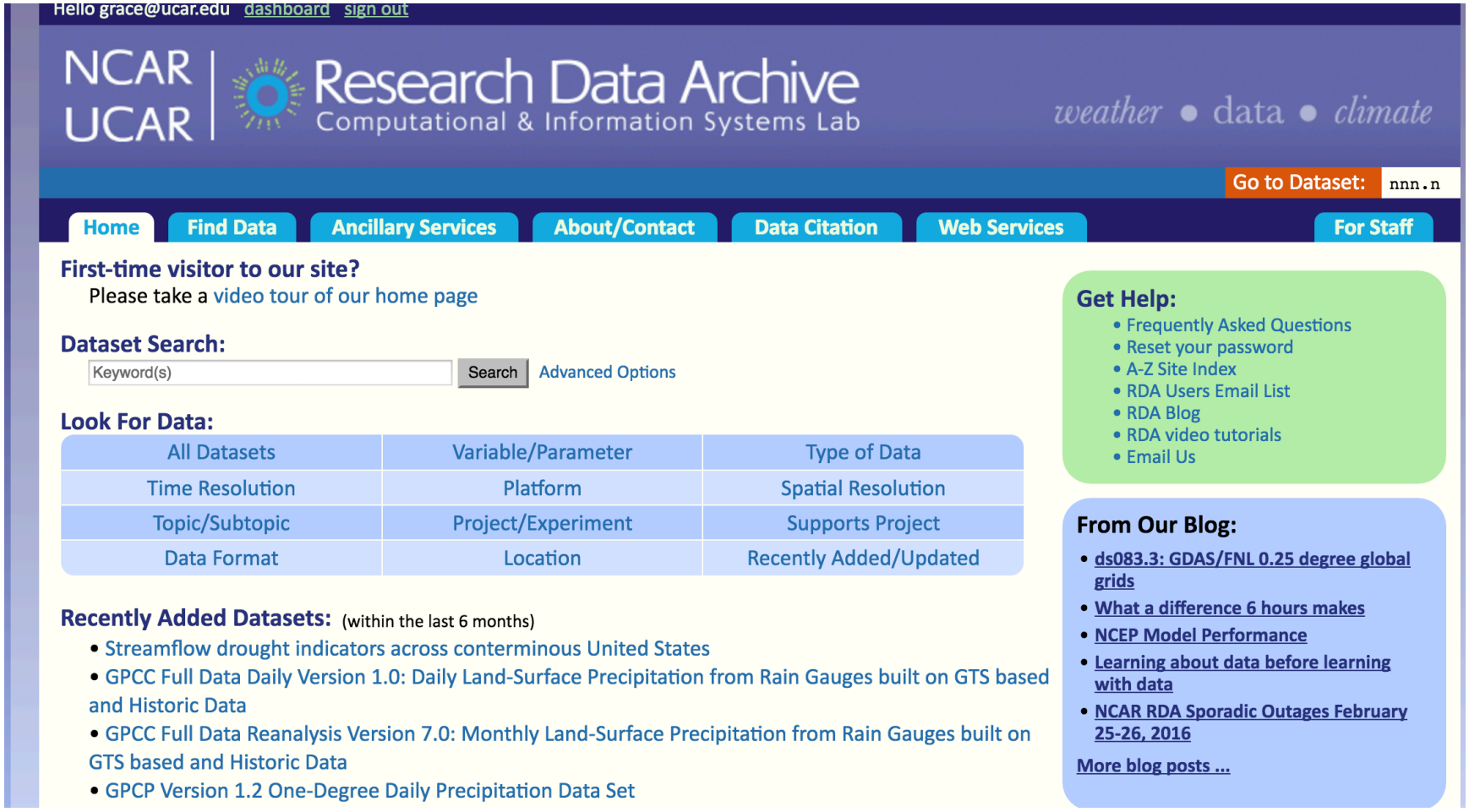
- Physical measurements/parameters
- Spatial coverage and resolution
- Temporal coverage and resolution
- Precision and accuracy
- In-situ, remotely sensed, and/or modeled?
- Does it require data fusion?
- Think outside the (physical) box

Data Discovery: Does this data exist?

- Prior work
  - Check their data citations
- Data papers
  - Describe how new datasets are prepared
- Data centers
  - Use their search features
    - Metadata: data about data
  - Read their "What's new" announcements

RDA Search Matrix

rda.ucar.edu



Data and Metadata Standards

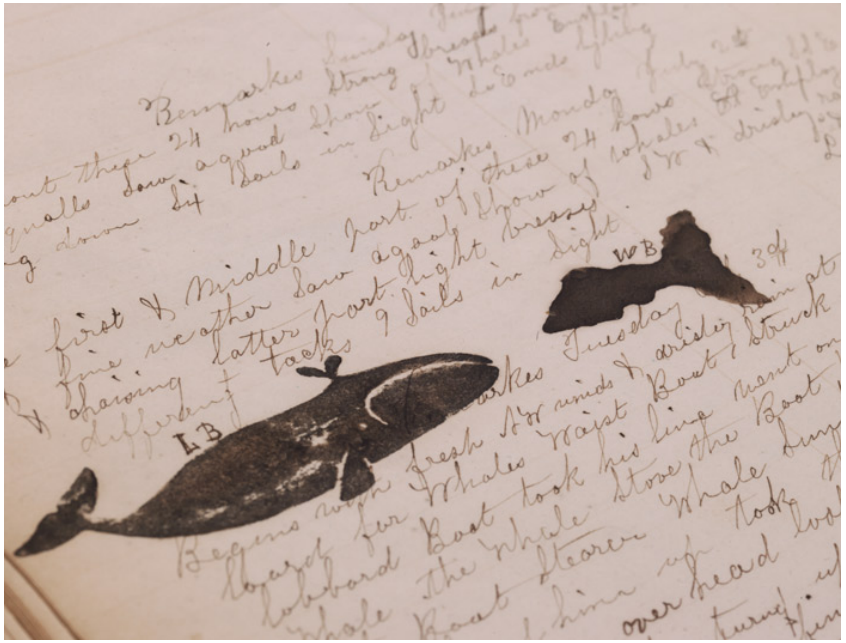
- Global Change Master Directory
  http://gcmd.nasa.gov/learn/keyword_list.html
- CF (Climate and Forecast) NetCDF
  http://cfconventions.org/index.html
- WMO GRIB1, GRIB2, BUFR
- ANSI 19115-X
- Federal Geographic Data Committee (FGDC) 1998
  http://www.usgs.gov/core_science_systems/csas/metadata/standards.html

GCMD Sample

```
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","SURFACE WINDS","","","","10685919-bc01-43e7-901a-b62ac44627f3"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","SURFACE WINDS","WIND SPEED/WIND DIRECTION","","","a92f49f3-e2ee-4ef4-b064-39311ffb95d3"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","TURBULENCE","","","","226d05da-dd0b-4314-919a-0b259ce724b5"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","UPPER LEVEL WINDS","","","","592d49c4-e8ae-4ab4-bf24-ae4a896d0637"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","UPPER LEVEL WINDS","WIND SPEED/WIND DIRECTION","","","661591b3-6685-4de7-a2a4-9ce8ae505044"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","VERTICAL WIND MOTION","","","","841a7ac7-5981-4e93-895f-1b57c3d892a0"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","VORTICITY","","","","858a80ff-5aa4-4590-b2e2-e88a802a6ee4"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","WIND CHILL","","","","8297bc97-9531-40da-950b-ac7e9d6e1b55"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","WIND PROFILES","","","","dcc6cbbf-23a0-4ae7-bfbd-6207d35c741f"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","WIND SHEAR","","","","05cf5b56-0f86-4819-b713-1272b97b06c5"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","WIND STRESS","","","","ef034881-8bf4-403f-a4ee-c68771769c93"
"EARTH SCIENCE","ATMOSPHERE","ATMOSPHERIC WINDS","WIND TENDENCY","","","","69526601-5607-46e0-954a-251249de80fe"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","","","","","162e2243-3266-4999-b352-d8a1a9dc82ac"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DROPLET DISTRIBUTION","","","","cbb0d517-462a-46fe-a0e6-32555f7e7f23"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","","","","62019831-aaba-4d63-a5cd-73138ccfa5d0"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","HEAT INFLUX","","","49fd6f11-5682-4d27-8fc6-66bf3faadf39"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","MOISTURE FLUX","","","925f563d-908a-4671-b750-23d0f3e42310"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","MOISTURE FLUX","DOWNWARD MOISTURE FLUX","","1dc6063b-892d-4879-8551-1e346dd3f2e7"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","MOISTURE FLUX","UPWARD MOISTURE FLUX","","49cad94d-0e93-44cb-a8a2-8e83d603463b"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","RADIATIONAL COOLING","","","5bac3ef6-5e30-4f14-a5dc-8065c7fcba55"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","RADIATIONAL DIVERGENCE","","","c7259da4-18dd-4196-91ff-a68087978349"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","THETA-E ENTRAINMENT","","","cfa49843-2d36-4709-8969-b176432adf78"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","VORTEX STREET","","","ba4a9964-8323-45df-a372-b4e2f3eef9e5"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","VORTEX STREET","KARMAN VORTEX STREET","","2d00d3c4-2ef3-49f6-9261-6184f6517b4f"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD DYNAMICS","WATER VAPOR TRANSPORT","","","a997c21b-ca61-4e78-8828-aa3e144976c3"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","","","","0cfcbaa7-727b-4199-8cca-93824b427e9b"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","CLOUD CONDENSATION NUCLEI","","","ebbf8642-3da1-4401-a779-3e56550a029d"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","CLOUD DROPLET CONCENTRATION/SIZE","","","47812ef8-b64b-4988-9ae4-31f3581ae9a5"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","CLOUD LIQUID WATER/ICE","","","05ac9d3e-bc44-41fa-ace0-c41bf3ebee97"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","CLOUD MASS FLUX","","","804fb334-1c74-4070-bd5f-848014a6e220"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","CLOUD OPTICAL DEPTH/THICKNESS","","","4bc483b1-dd64-4e97-bfd3-c0e755df6308"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","CLOUD PRECIPITABLE WATER","","","b709d6fc-f0cf-47de-bdbb-1cd875b5f3ab"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","COLLISION RATE","","","76bcb8e0-1c07-4783-9d15-3a22203f7849"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","DROPLET GROWTH","","","63effad4-4323-486d-a81b-e0bf3264e5c9"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","DROPLET GROWTH","ACCRETION","","ab702934-0959-45fc-a523-81e1aa0c09c8"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","DROPLET GROWTH","ACCRETION","RIMING","889253e1-e189-4f75-bdc7-7e612b19e3ae"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","DROPLET GROWTH","AGGREGATION","","8e484ec4-50fd-4c08-9c96-6ad483e170ad"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","DROPLET GROWTH","COALESCENCE","","d5d64790-db29-451d-b022-a461dac06228"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","PARTICLE SIZE DISTRIBUTION","","","00d6fb2f-16d5-4949-afec-a1adbd600a58"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","SEDIMENTATION","","","8d66dbbe-886e-449d-bfd2-93fc8d357ccd"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD MICROPHYSICS","SEDIMENTATION","SEDIMENTATION RATE","","bee9f657-c115-4d73-a10c-7e05e00db574"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD PROPERTIES","","","","c9e429cb-eff0-4dd3-9eca-527e0081f65c"
"EARTH SCIENCE","ATMOSPHERE","CLOUDS","CLOUD PROPERTIES","CLOUD ASYMMETRY","","","4c737490-1486-418f-81f4-c50c47da117d"
```

Data and Metadata Standards

- Infrastructure of the data world



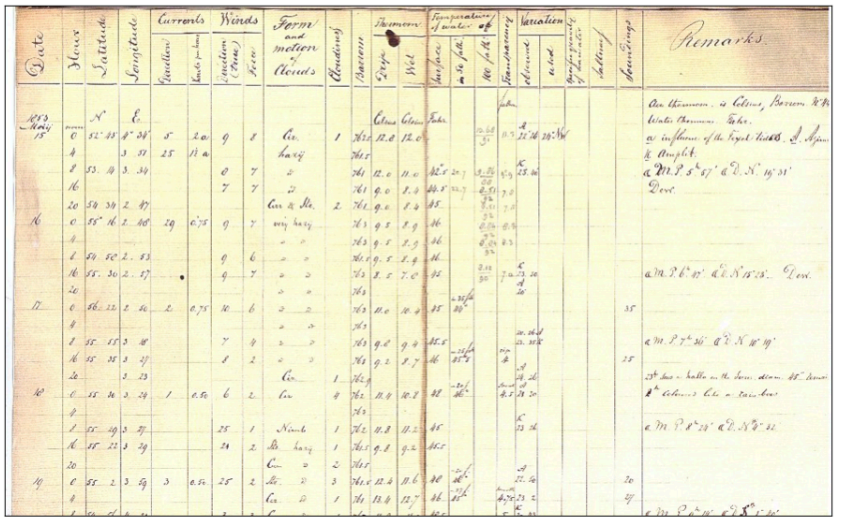
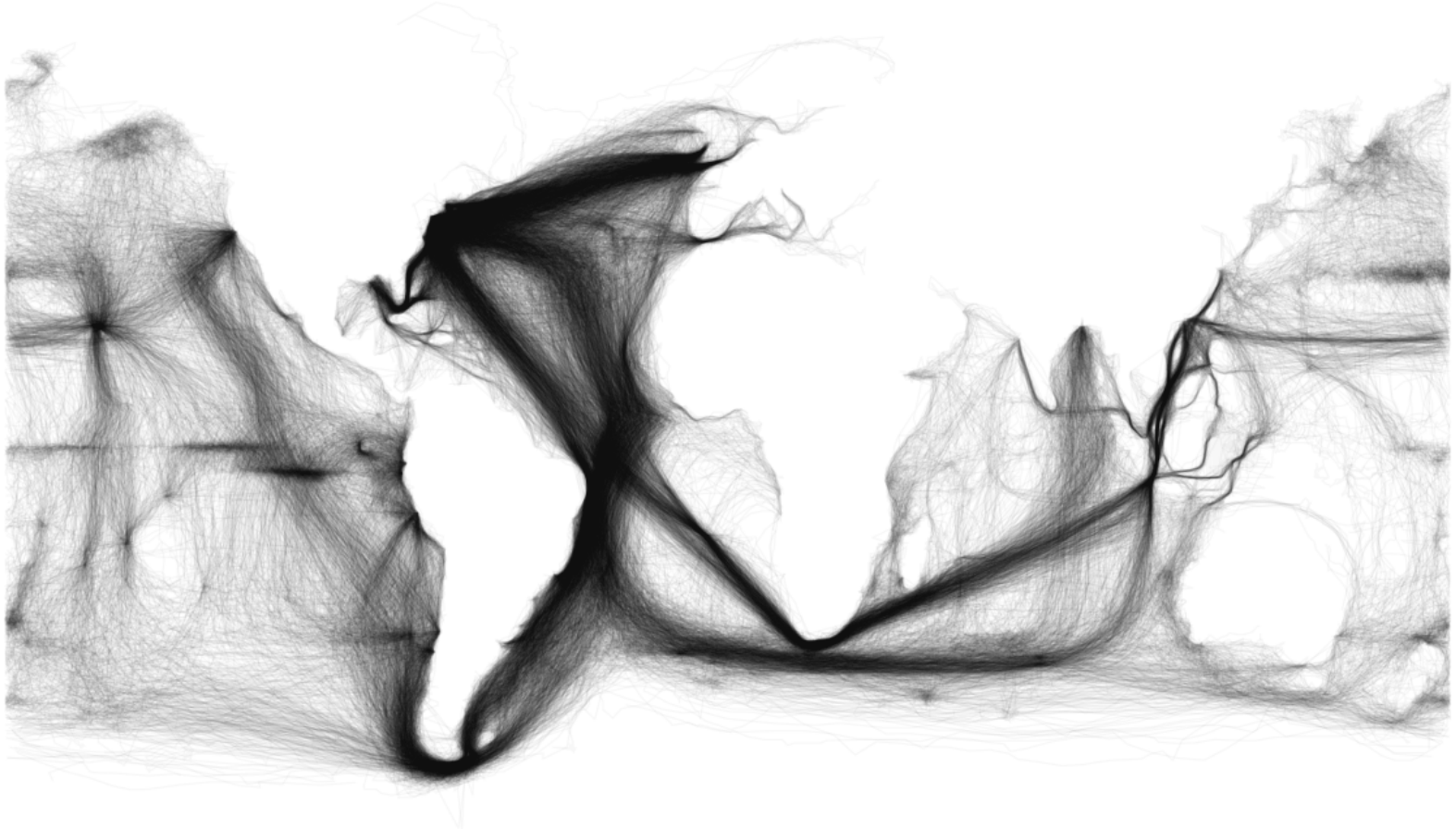
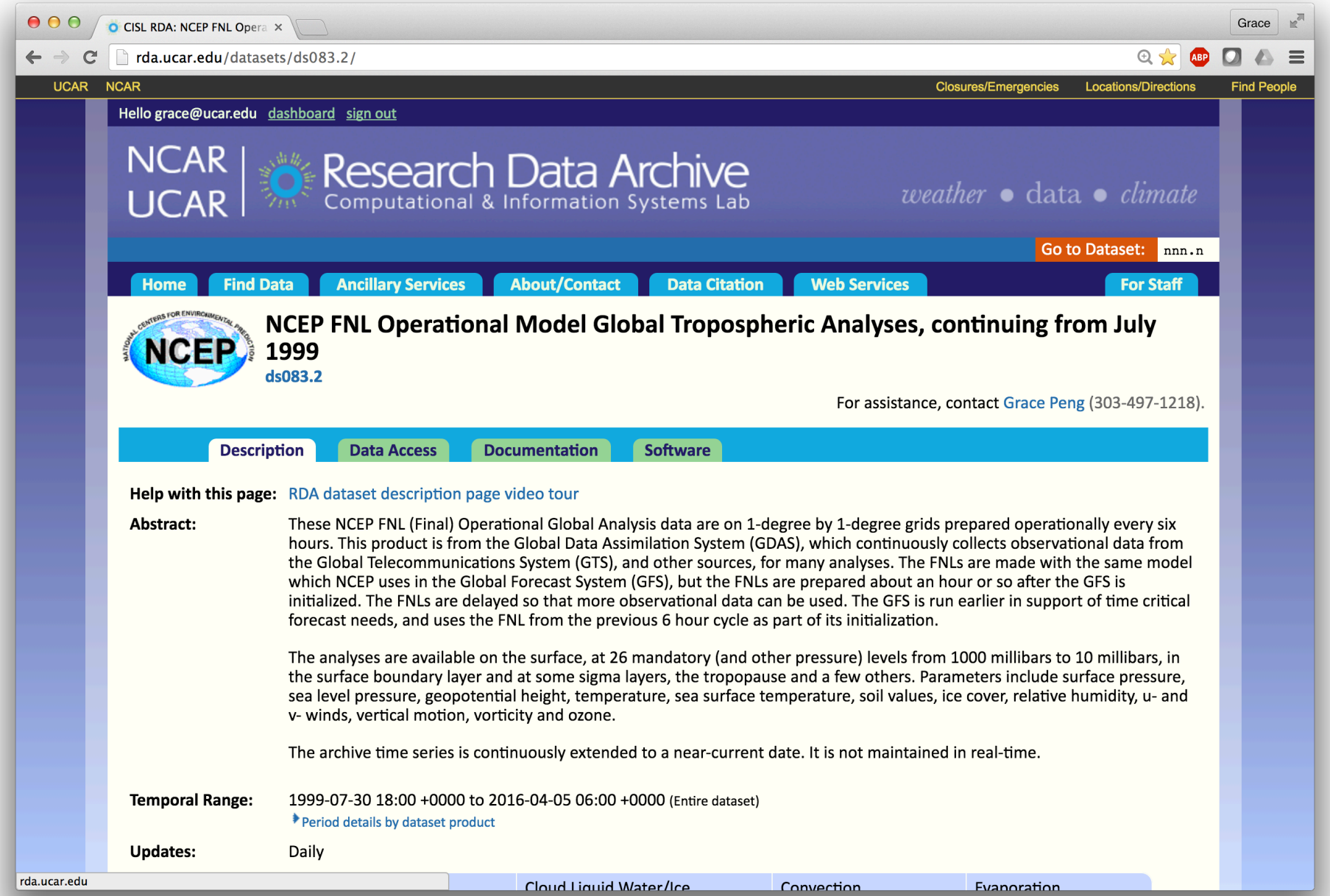
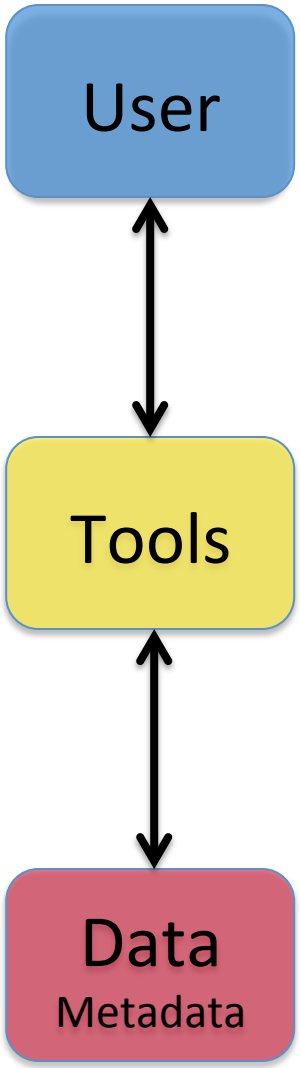


Figure 3. Page of the experimental universal logbook of the *Prince of Orange*, covering the period 15-19 May 1853. The logbook, designed by Maury and Jansen, was tested in practice during this trip of this ship. This happened in the preparation of the Brussels Conference in November 1853.

Time-travel through data




CISL RDA: NCEP FNL Opera
Grace
rda.ucar.edu/datasets/ds083.2/
UCAR NCAR
Closures/Emergencies
Locations/Directions
Find People
Hello grace@ucar.edu dashboard sign out
NCAR UCAR
Research Data Archive
Computational & Information Systems Lab
weather • data • climate
Go to Dataset: nnn.n
Home
Find Data
Ancillary Services
About/Contact
Data Citation
Web Services
For Staff
NCEP
NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999
ds083.2
For assistance, contact Grace Peng (303-497-1218).
Description
Data Access
Documentation
Software
Help with this page: RDA dataset description page video tour
Abstract: These NCEP FNL (Final) Operational Global Analysis data are on 1-degree by 1-degree grids prepared operationally every six hours. This product is from the Global Data Assimilation System (GDAS), which continuously collects observational data from the Global Telecommunications System (GTS), and other sources, for many analyses. The FNLs are made with the same model which NCEP uses in the Global Forecast System (GFS), but the FNLs are prepared about an hour or so after the GFS is initialized. The FNLs are delayed so that more observational data can be used. The GFS is run earlier in support of time critical forecast needs, and uses the FNL from the previous 6 hour cycle as part of its initialization.
The analyses are available on the surface, at 26 mandatory (and other pressure) levels from 1000 millibars to 10 millibars, in the surface boundary layer and at some sigma layers, the tropopause and a few others. Parameters include surface pressure, sea level pressure, geopotential height, temperature, sea surface temperature, soil values, ice cover, relative humidity, u- and v- winds, vertical motion, vorticity and ozone.
The archive time series is continuously extended to a near-current date. It is not maintained in real-time.
Temporal Range: 1999-07-30 18:00 +0000 to 2016-04-05 06:00 +0000 (Entire dataset)
Period details by dataset product
Updates: Daily
rda.ucar.edu
Cloud Liquid Water/Ice
Convection
Evaporation



- Access data with tools
- Standardized data
- Reusable tools
- Reusable data
- Happy users
- Data infrastructure
- Data that remains in use remains alive

Veracity

- Is this data what you expect?
- Is it as described in the documentation?
  - Format
  - Field contents
- Are the values physical?
- If any of these answers are no, consult with the data provider and/or data specialist. rdahelp@ucar.edu

Reproducible Science: Data Citation

- Provider/creator
- Dataset name
- DOI if it has one
- Source
- Revision, access date
- RDA citation widget

How to Cite This Dataset:

RIS

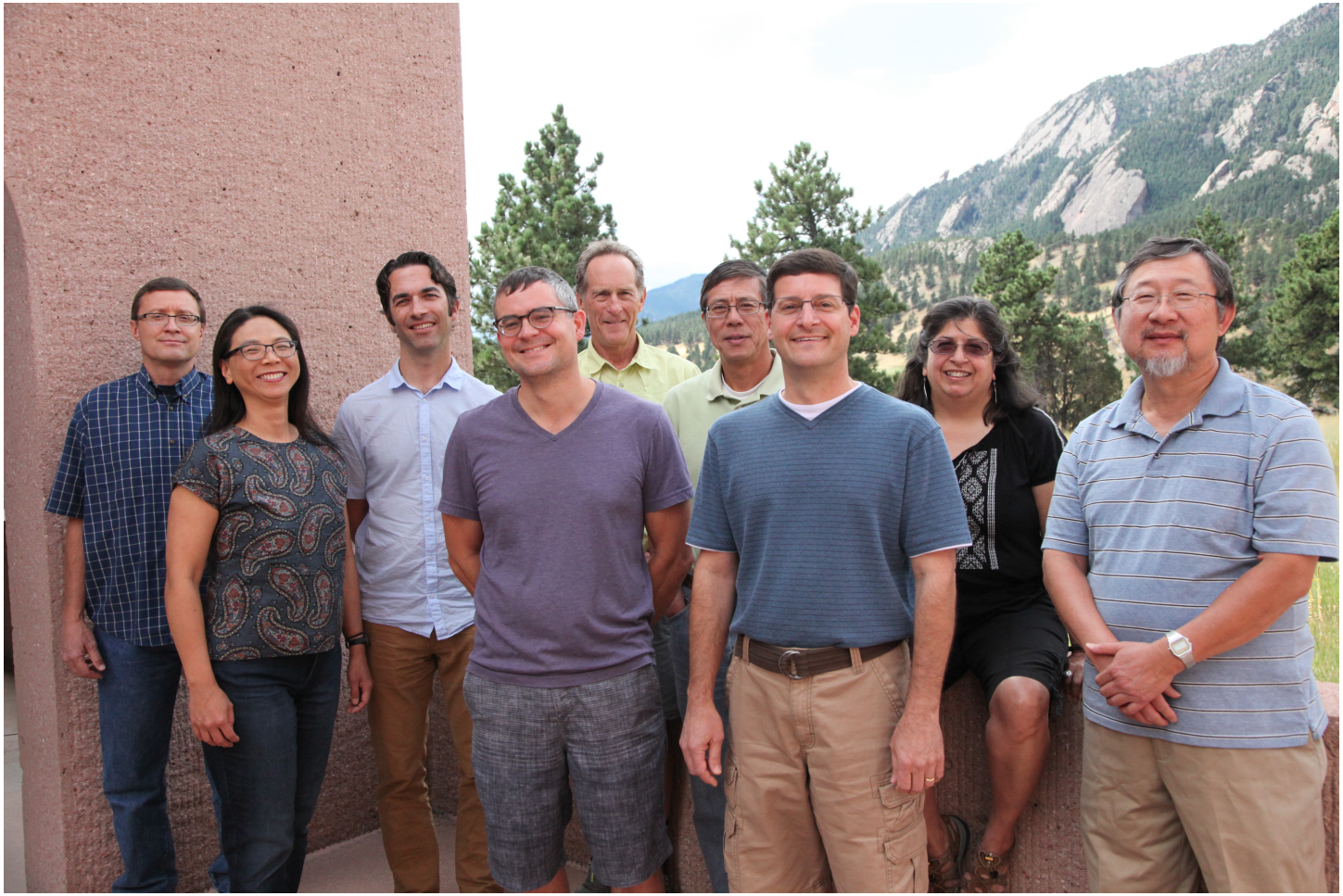
BibTeX

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce. 2000, updated daily. *NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. http://dx.doi.org/10.5065/D6M043C6. Accessed† dd mmm yyyy.

†Please fill in the "Accessed" date with the day, month, and year (e.g. - 5 Aug 2011) you last accessed the data from the RDA.

Bibliographic citation shown in Federation of Earth Science Information Partners (ESIP) style

Get a customized data citation

Questions? rdahelp@ucar.edu

http://ncarrda.blogspot.com/p/recommended-data-reading.html