



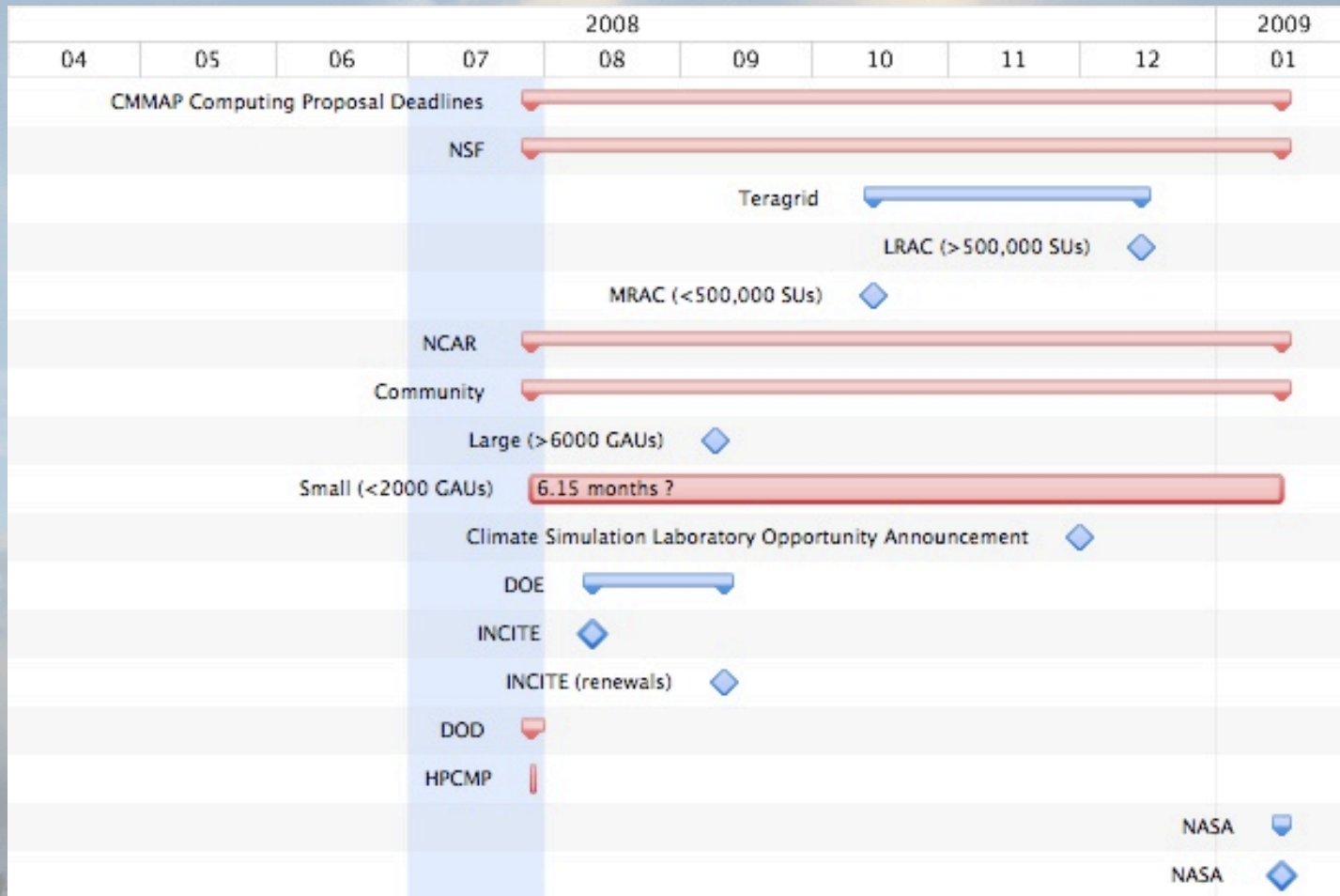
CIWG Breakout

July 2008
Semi-annual Meeting

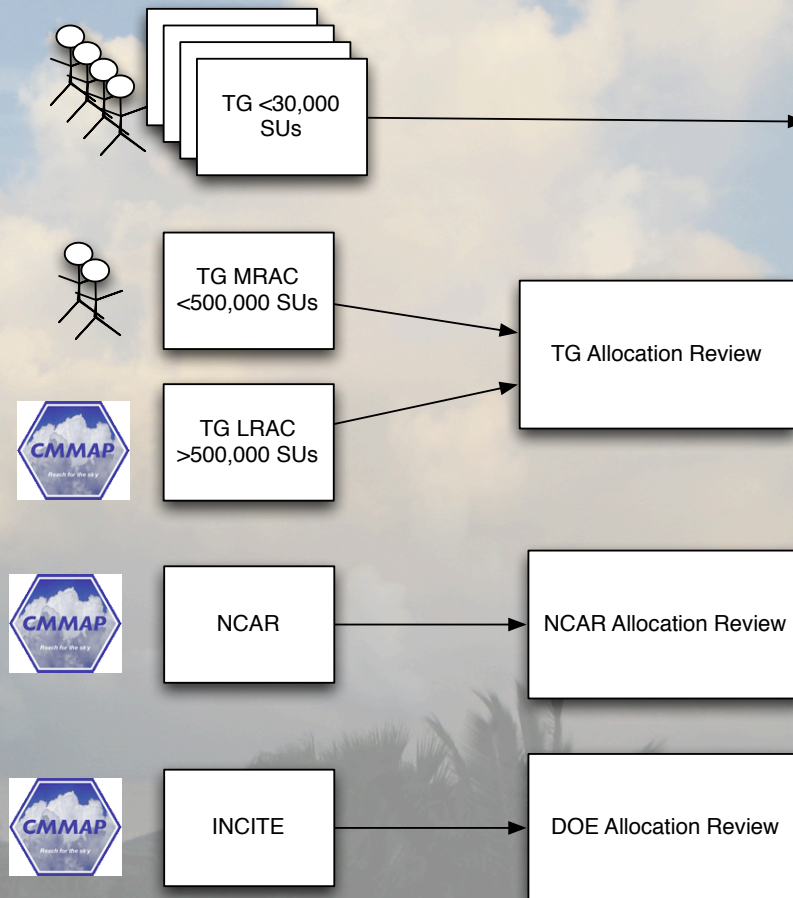
Issues

- Allocations (what, who, when)
- Data loading (for Data Management Plan/WP)
- Data products and conventions
- Digital Library (Wiki, RSS)

Current Allocation Proposal Deadlines



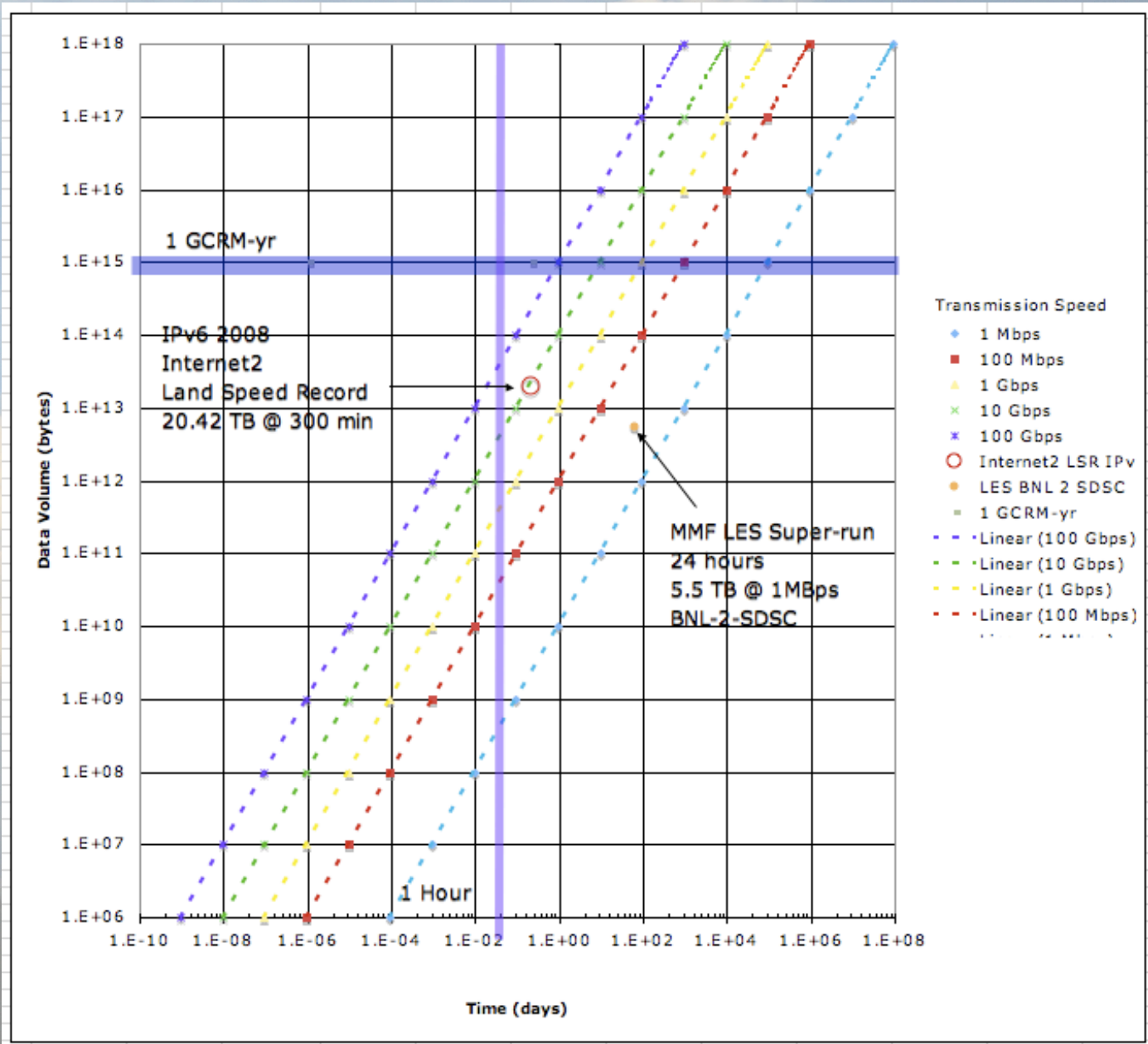
Proposal Strategies



Current Data Loading

```
2.3T      .
2.0T      ./source
114G      ./source/CMMAP_AMIP
92G       ./source/CMMAP_SUPERLES
91G       ./source/CMMAP_SUPERLES_4xCO2
3.7G      ./source/PRITCHARD
1.7T      ./source/SAM6.7_SR
14G       ./source/SAM6.7_SR/OUT_2D
1.6T      ./source/SAM6.7_SR/OUT_3D
113G      ./source/SAM6.7_SR/OUT_MOMENTS
51G       ./source/SAM6.7_SR/OUT_MOVIES
137M      ./source/SAM6.7_SR/OUT_STAT
296G      ./staging
114G      ./staging/CMMAP_AMIP
92G       ./staging/CMMAP_SUPERLES
91G       ./staging/CMMAP_SUPERLES_4xCO2
```

Internet Data Transfer Capacity





CMMAP Digital Library

Edit primary links

Active forum topics

- Allocation Proposal Schedule (as of: 2008-07)
- Computing Time
- Proposals
- Fortran90
- Data

more

hellyj

- Create content
- Login
- Search Data Catalogue
- My account
- Recent posts
- Administer
- Log out

Allocation Proposal Schedule (as of: 2008-07)

Tue, 07/29/2008 - 2:46pm — hellyj

Latest summary schedule

Add new comment 1 attachment

CIWG

2008 Annual Meeting

Tue, 07/29/2008 - 1:02pm — hellyj

First blog entry

hellyj's blog 2 comments

Developer Tools

- Migrating code to Future NSF resources (<http://www.sdsc.edu/us/consulting/migration.html>)
- Script automation (<http://www.ibm.com/developerworks/aix/library/au-satbuildscript.html>)
- Compiler quirks (<http://en.wikipedia.org/wiki/Compiler>)

Read more

Recent blog posts

- 2008 Annual Meeting

more

Recent comments

- Model Taxonomy has been updated
16 hours 21 min ago
- ECMWF Multi-grid modeling idea
19 hours 3 min ago

Who's new

- mikepritchard
- kelley
- mbranson
- leoners56
- jasonc

Who's online

There are currently 0 users and

Publication of Scientific Data

CONTROLLED PUBLICATION OF DIGITAL SCIENTIFIC DATA

How to balance free and open access to scientific data with privileged access to new results by authors while protecting them from being scooped by competing interpretations of their own data.

Although the principle of equal access to data is a key aspect of U.S. government-funded science policy [10], there are strong, institutionalized, though sometimes contradictory, incentives for investigators to maintain proprietary control over that data; there is also increasing commercial and in some cases federal pressure to treat data as a commodity [4]. Efforts by the scientific community to prevent potentially deleterious international commercialization of scientific data through the World Intellectual Property Organization (WIPO) have had some success, thanks to support from the U.S. State Department. A recent example is the Anti-Piracy Bill (H.R. 2652) passed by the U.S. House of Representatives in 1998 but never approved by the Senate; it was similar in some ways to the WIPO proposal. Related pressures continue to build, including from within the U.S. private sector. It seems the commercialization of scientific data and treating it as a commodity represent an increasingly important aspect of how scientific data is published today; further complicating this scenario is the growth of the Internet-based business sector and the increasing commercial value of the data itself, especially biomedically significant data. These changes have influenced many aspects of scientific research, including the published content of professional journals, both online and on paper. The special role of research data in the advancement of science and its distinctly non-commodity character were identified as threatened by efforts to put a price on data [5].

Efforts by some scientists and policymakers to prevent the commercialization of scientific data reflect a certain irony and tension attending the purpose and politics of such data. On one hand is vigorous support for free and open access to the data consistent with the scientific

John J. Helly, T. Todd Elvins, Don Sutton, David Martinez, Scott E. Miller, Steward Pickett, and Aaron M. Ellison

STD-DOI Publication and Citation of Scientific Primary Data

http://www.std-doi.de/front_content.php

Camino Info News Mac News Tabs Google Links to calculators (d... SENSIAC - Technical ... Development of a Nov... ALOS - ADEN DQPC a... Operationally Respon... http://www.nbp.usap... ADCP Figures (with fr... The Project Gutenber...

Publication and Citation of Scientific Primary Data

- Home
- About STD-DOI
- News
- Documents
- Contact
- Internal

News

- DSJ Special Issue: Open Data for Global Science
- TIB completes pilot of project to install DOIs for scientific datasets
- Data publication in the open access initiative
- Open Access als Publikationsalternative unter Wissenschaftlern kaum bekannt
- New DOI Registration Agency for scientific data appointed: 1,500,000 datasets to be registered by end of 2005
- Creative Commons zum Geburtstag

STD-DOI - Creating access to scientific Data

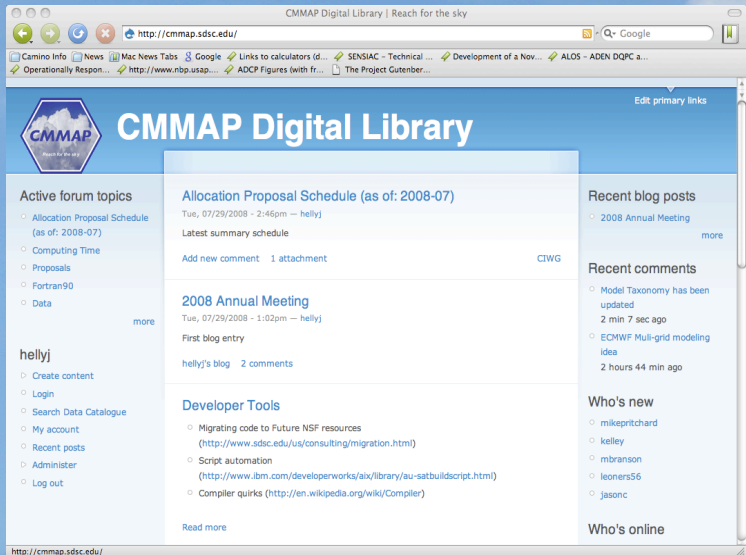
"Publication and Citation of Scientific Primary Data" (STD-DOI) is a project funded by the German Science Foundation. Its aim is to make primary scientific data citeable as publications. In this system, a data set would be attributed to its investigators as authors like it would be done for a work in the conventional scientific literature. Thus, scientific primary data should not exclusively understood as part of a scientific publication, but may have its own identity. [Download and view a short MS Powerpoint presentation on the idea behind STD-DOI: [Deutsch](#) / [English](#)]

How can I publish data and make them citeable? Follow this link to find out more about [publication of data](#).

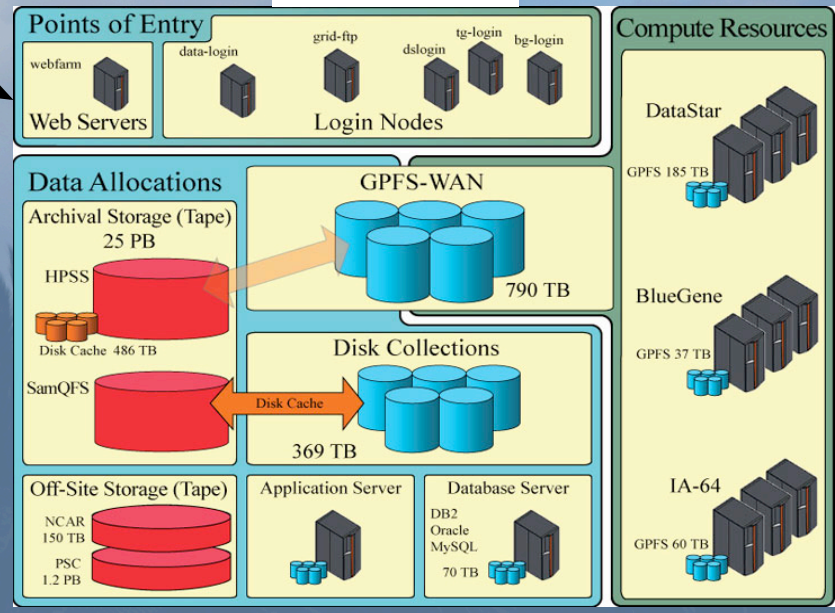
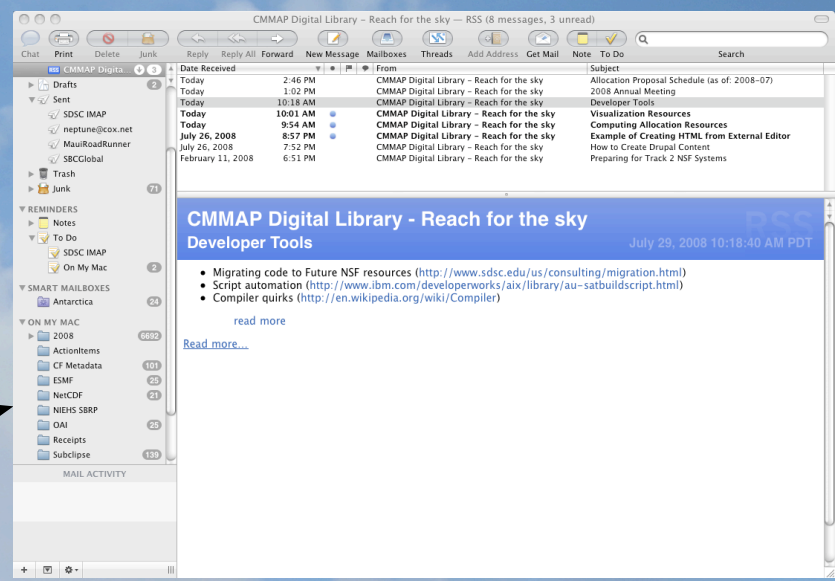
How do I cite data publications? Follow this link to find out more about [citation of data](#).

How can I participate as a data publication agent? Follow this link to find out more about becoming a [data publication agent](#).

The diagram above gives a schematic overview of the existing network of registration agents for scientific and technical data, the registration agency and the International DOI Foundation.



RSS



URL data retrieval
OAI metadata

- Basic User Menu**
- SSDBviewer
 - SSDBquery
-
- Quick Links**
- Login
 - Register
 - Information for Proponents

Home > SSDBquery

SSDBquery

Instructions:

Press the "Search" button to search the Site Survey Data Bank. All search crit

Search ->

Proposal Number: Latitude North:
 - All -
 A001
 Exp303-306
 Exp306
 Exp313
 Exp314
 Leg125-126
 Leg127-128
 Leg170
 Leg194
Longitude West: Longitude East:
Latitude South:

Site Names: [View Site List](#)
 Expedition Name:
 Data Type:
 Data Object Type:
 Format:
 Access Control:
 Submission Date: Start: End: (YYYY-MM-DD)



Portal to SIO Expedition Discoveries

- Home
- User Interfaces
- Project Summary
- Tools
 - Search Tools
 - Metadata Tools
 - Grid Tools
- Resources
 - Shipboard Data
 - Historical Material
 - Geophysical Data
 - Educational Data
 - Metadata Resources
- Education
 - Information
 - Master Documents
- Support
 - Help
 - Contact Information
- About Us
 - Collaborators
 - People

Welcome to the SIOExplorer Digital Library Project

User Interfaces



[Advanced interface](#) - requires java webstart



[Basic interface](#) - text search

Project Summary

Data, documents and images from 822 expeditions by the Scripps Institution of Oceanography (SIO) since 1903 are becoming [web-accessible](#) under a [usage policy](#) through the new SIOExplorer project, and will become part of the overall NSF-funded National Science Digital Library (NSDL). The effort is a collaboration among researchers at Scripps, computer scientists from the San Diego Supercomputer Center (SDSC), and archivists and librarians from the UCSD Library.

An overview of the project is available from recent [news releases](#) or a more detailed recent [presentation\(ppt\)](#). Funding is provided by the NSF NSDL and ITR programs and by the Scripps Institution of Oceanography.

Predecessor Digital Libraries

Conventions for Naming Datasets

- `spcam_climo_-3C-12C_1e-5_re7-10_vt01---` → CMMAP_SUPERLES
- Standardized names aid communication and data management
 - current convention is adapted from oceanographic collections done for SIO and IODP
 - none apparent for climate modeling (e.g., AMIP, CF)
 - DOI (digital object identifiers) are used in the publishing world
 - this is similar to a DOI and we may add a DOI if a DOI authority emerges for atmospheric data
 - In the meantime we should have something equivalent

CMMAP_SUPERLES_20080725014445402_20080725014445402_spcam3.clm2.r.0003-01-01-00000.mif

Standardized prefix is added
to facilitate search and identification

Original filename is untouched

Conventions for Data Formats

- NetCDF is default storage format
- Models cannot(?) afford to generate this
- Implies that some post-processing capability must be provided
- No current provision for this
- Sub-group will form to develop an initial approach

