

Research Data Life Cycle Matrix		Date: 10 January 2013						
					[1]	[2]		
Organization	Proposal	Acquisition	Analysis & Synthesis	Contribution	Discovery & Access	Use & Reuse	Publication	Preservation
Geological Survey of Alabama (comments from Denise Hills, dhills@gsa.state.al.us) State Geological Survey Firstly, a place where oil and gas legacy data exists	not always involved - companies are required to provide their data to our agency through the regulatory agency (Alabama Oil and Gas Board). Potentially, we could encourage the OGB to require companies to provide standardized information	not always involved - again, we might require data providers (oil companies) through the OGB. Difficulties might arise, though, as companies really protest at changes in requirements	often not involved directly - although we sometimes reprocess, etc., once we have the data	We ARE the repository - but our data are not always readily accessible	see previous cell - some data is searchable, but we have no standards applied to data	NEED BETTER METADATA - need standards for physical samples Additionally, there are issues with proprietary data. Can say what you "learned" from the data, but can't show the ACTUAL data	Much data is proprietary - may not be able to be published in ANY form, let alone as a repository	KEY for us - and a huge challenge, as we do not currently have funding to do this, but is very necessary. May again have to deal with the OGB rethinking how they do their permitting/regulation to include a preservation of the legally required data collected
Geological Survey of Alabama (comments from Denise Hills, dhills@gsa.state.al.us) State Geological Survey A second part is data collected for the survey - mapping, water information, etc (will be a new line entry)	depends on the proposal - we have federal, local, state level grants and programs. Not a normal process, necessarily	Data collectors - these are GSA employees. Each group may have standard procedures, but they may not (yet) be documented.	Data producers - GSA employees and/or contractors. GSA has standards (maybe not documented); contractors have standards (which we may or may not have documented) Concern: We may or may not have backups (difficult to backup physical samples)	We ARE the repository - but our data are not always readily accessible.	see previous cell - some data is searchable, but we have little to no standards applied to data	NEED BETTER METADATA - need standards for physical samples in particular Additionally, issues with proprietary data	Some data is proprietary (sometimes for a limited time, sometimes for perpetuity)	KEY for us - and a huge challenge, as we do not currently have funding to do this, but it very necessary, particularly for older data/samples. As we get new and different outside funding, we may look at ways to incorporate this cost. However, we need to address legacy data.

Research Data Life Cycle Matrix		Date: 10 January 2013							
					[1]	[2]			
Organization	Proposal	Acquisition	Analysis & Synthesis	Contribution	Discovery & Access	Use & Reuse	Publication	Preservation	
NOAA National Oceanographic Data Center (NODC). Long-term national archive for oceanographic data. (inputs from Ken Casey, Don Collins, John Relph, maybe others?)	We rely on high level policies from NSF/NOAA etc. to require data management (& plans) for submission of (well-documented) data to the national archival holdings. NODC also works with data producers as they develop their data management plans. We have relationships with some programs, like NOAA's Ocean Acidification Program and the NOAA Office of Exploration and Research, where we are funded to participate very actively in the early stages of a PI's efforts... developing a data management plan, working with them to define the data formats to be used, etc.	In NODC context, acquisition is acquiring data from a data producer/provider for archival stewardship, rather than acquiring data during a field experiment, mission, etc. However, NODC can help inform planning of acquisition especially with regard to documenting the data as they are acquired.	In some cases, NODC does conduct analysis and synthesis activities. We typically call this sort of activity "building products"... in our case, those products are typically some form of aggregated, quality-controlled product that is composed of data from multiple producers and brought together in a common format. An example is the World Ocean Database.	NODC works with intermediate data assembly centers to acquire data, and to make sure that data management is in line with policy and Archive needs	Project-specific discovery and access methods may support submission of data sets to a national Archive. However, data management practice needs to be well considered, especially with regard to consistency to support machine-to-machine interactions, and documentation (descriptive information) to support long-term preservation and understanding of the data. // (Ken:) Discovery and access are at the heart of NODC operations... starting from the inside, and working outward... First, we work to build at least minimal metadata for everything in the NODC Archive. Then, we expose that through human web interfaces like the Geoportal Server, make it searchable via OpenSearch and CSW, and publish it so that Google and other search engines can harvest it. We then work actively with federally mandated "uber portals" like Data.gov and Geo.Data.Gov to provide them with our metadata. We are also now actively working with "data networks" to provide federated discovery and access to our archive holdings. Two in particular are the WMO Information System (WIS) and NSF's DataONE.	NODC provides free and open access to data, and supports standard data access protocols. In addition, NODC actively works to help data producers to manage their data using interoperable formats and standards	NODC publishes data sets for long-term access, with a persistent unique identifier (NODC accession number). We are also working with the rest of NOAA to support DOI publication using EZID.	Long-term preservation can support many of the other phases. NODC has routine redundant backup procedures with backup storage in multiple geographic locations.	
NASA ORNL DAAC (Sue Heinz, Technical Project Manager)	Field Campaigns or Datasets are discussed with User Working Group. Communicated to Sponsor (NASA). Provide initial scope of work and cost to incorporate. Individual Data providers occasionally contact the DAAC or NASA to request archive.	Various Field Campaigns - generally NASA Sponsored. We do have a MODIS web services tool now - Data is pulled from another NASA Data Center. Significant ingest process is initiated to add to database and cite. Future missions will be planned to communicate with data providers specific formats to streamline ingest process	Reviewed by DAAC Scientist and Quality control by technical staff (GIS, Metadata, General Science)	Published	ORNL Website, ESDIS/EOSDIS Websites, Mercury, ECHO, GCMD		DOI assigned	Archived and backed up on site	

Research Data Life Cycle Matrix		Date: 10 January 2013						
Organization	Proposal	Acquisition	Analysis & Synthesis	Contribution	Discovery & Access	Use & Reuse	Publication	Preservation
	Our field programs are of short duration (e.g. 1-4 months) and can occur anywhere in the world. We operate research platforms such as radars, profilers, aircraft and sounding systems that are requestable by the NSF scientific community, typically Universities. After science proposal is written for a given field program, the PI teams work with the data managers to decide on what other data sources are required for the region and area of interest. For example, if a research project is to take place in the US, the PI team works with us to also acquire satellite, radar, mesonet and other operational data as part of the data collection for the experiment.	Often, our teams of software engineers and data managers will need to integrate user instruments, e.g. an airborne instrument measuring a specific parameter into the data acquisition systems onboard the aircraft. These instruments can be simple instruments that send their data via a serial feed of data to highly complex standalone instruments that only need access to a common time synchronization information (e.g. NTP or other) and position information via an something like an ARINC bus. We also have an extensive system (called the Field Catalog, see http://catalog.eol.ucar.edu) which is used during the field program to acquire and view reports, imagery and preliminary datasets in the field. Finally, we build real-time displays of integrated datasets of satellite, radar, model and aircraft data which, along with a customized IRC chat system, are used to provide situational awareness used in real-time to direct flight missions.	This step requires a Quality Control (QC) and/or Quality Assurance (QA) step for us. For example, QC might involve despiking of data, filtering, deciding on the best sensor to use in the final dataset among a suite of redundant sensors. In QA, this could mean comparing several similar measurements with others and perhaps then creating a "composite" dataset of similar measurements for the PI team in a common format. In our world, these composites are common with balloon sounding measurements, e.g. comparing soundings we do with soundings occurring daily by the National Weather Service. Other Quality Control procedures, for radars, for example, could involve complex calibrations, clutter removal, creation of 3-D wind fields from dual-doppler radar measurements, etc. We also meticulously track revisions to datasets, when and why they were made, etc. Of course, occasionally we may do this step again if the community finds errors at a later date and the data need to be re-processed.	Contribution could involve working with the providers on "ancillary" datasets to deliver their data in a timely manner and in a suitable format. The scheduling of data workshops (mentioned in a subsequent column) can often prompt our PI teams to submit their data just before the data workshops actually take place because part of our presentations at the workshop will show who has submitted their data and who hasn't... :) We also have some projects, e.g. A-CADIS in which we do not build the data acquisition systems, but begin the data management lifecycle once the data are collected and not necessarily during the "ground level" data acquisition process itself.	Like many other organizations, we store a very comprehensive set of metadata within a metadata database. We have tools that can export this metadata to a variety of metadata formats, e.g. ISO, THREDDS, etc. to facilitate access through several methods. NCAR has attempted to try to build a single data portal for the organization (e.g. cdp.ucar.edu) and then the various individual labs connect to that portal. The NCAR CDP intern connects to other portals via OAI, etc. However, not all groups are using the CDP and users can also go directly to our data distribution system at http://data.eol.ucar.edu/codiac/	3-12 months after a field program ends, the PI teams often get together to present their preliminary results at a "Data Workshop" meeting. Having our data folks attend these meetings can be one of the best feedback mechanisms in terms of understanding quality, formatting, access or usage problems that need to be addressed.	We also will be working with our library group to assign DOIs to datasets that are ready to be used. The most difficult parts of DOIs/citations have been 1) Deciding the granularity of a dataset appropriate for a DOI and 2) Deciding who should be on the actual citation itself.	We do not have a formal policy for the number of years a dataset must be preserved. Our archive is relatively young, say, since the 1970s, and we do have these older data online. Of course, the codes to read these legacy data are often obsolete, we have very little digital documentation on the data themselves and so they are difficult to use. Climate science is driving the need to "rescue" these legacy datasets, convert them to modern formats, create web pages for the datasets, digitize documentation and in somecases digitize the datasets themselves.
Rolling Deck to Repository (R2R) Program - underway sensor data from the U.S. academic oceanographic research vessel fleet (Bob Arko) NSF funded	Each field program (cruise) is linked to 1 or more funding awards (NSF, ONR, etc)	Research vessel technician routinely submits a package of underway sensor data (and associated documentation) directly to R2R after cruise ends - should be "an exact copy of the data handed to chief scientist	R2R performs routine/programmatic assessment (not control) of data quality for selected device types	Each chief scientist indicates proprietary data holds (if any), according to their funding agency policy, and is invited to contribute a Cruise Report to be archived along with the underway data	R2R publishes all catalog content as Linked Data (RDF resources) with an associated SPARQL endpoint, as well as traditional Web catalog browser	Ultimately original field data flow downstream into final products, global syntheses, journal articles, etc	R2R publishes a DOI for each dataset, and also plans to publish a ARK(?) for each granule for selected data types (eg. multibeam swaths) where a community use case exists	R2R submits each dataset with documentation (IETF/ISO- standard metadata records, and quality assessment report) to NOAA Data Centers for long-term archiving

Research Data Life Cycle Matrix		Date: 10 January 2013						
Organization	Proposal	Acquisition	Analysis & Synthesis	Contribution	Discovery & Access	Use & Reuse	Publication	Preservation
NASA ESDIS Project (Ramapriyan)	Not directly involved. However, NASA's EOSDIS DAACs managed by the ESDIS project provide data to users who may use the data in their proposals. Some of us are involved in providing inputs to calls for proposals and/or peer review evaluation process.	We support acquisition of data from satellites, aircraft and in situ campaigns	Data products derived from satellite data are generated by instrument teams funded by NASA into higher level products and archived in the DAACs for general distribution and supporting science research and applications.	Data products are archived at the DAACs. DAACs can be considered publishers of data. There are established policies and procedures for accepting data for archiving at the DAACs. NASA mission generated data are assigned to specific DAACs and follow interface control procedures to be sent to DAACs as they are generated. DAACs can accept data from research projects after review by DAAC User Working Groups and cost assessment and approval by NASA.	There are many tools provided by EOSDIS and the DAACs to facilitate discovery of and access to data. On-line services are available for subsetting, visualization and analysis.	The data held at the DAACs are freely and openly accessible to all users on a non-discriminatory basis, and "no period of exclusive access" according to NASA's Data and Information Policy.	When data are archived at the DAACs, they are "published" - i.e., they are available to the community. We have started assigning DOI's to the datasets held at the DAACs to ensure they can be cited.	NASA has recently developed a specification titled "NASA Earth Science Data Preservation Content Specification". These are applied as appropriate and practical to current and future missions. The specification provides a checklist of contents that need to be preserved along with data. Provenance and context are covered. This specification resulted from the work with ESIP Data Stewardship Committee where we started an "Emerging Provenance and Context Content Standard (PCCS)" matrix.
Michael Camponovo, Limited involvement in some of these topic categories, Earth Data Analysis Center (EDAC)	EDAC is not directly involved in the proposal process (as far as I know) of the researcher teams around the state.	EDAC acquires most data from other sources like government agencies within the state and researchers as part of the EPSCoR program. This results in us spending a large amount of time contacting the researchers to acquire the data and the documentation.	Limited involvement		EDAC hosts data on both the RGIS website (http://rgis.unm.edu) and the NMEPSCoR Data Portal (http://nmepscor.org/dataportal).			

Research Data Life Cycle Matrix		Date: 10 January 2013						
Organization	Proposal	Acquisition	Analysis & Synthesis	Contribution	[1] Discovery & Access	Use & Reuse	[2] Publication	Preservation
Tyler Stevens, NASA Global Change Master Directory (GCMD) Tyler.B. Stevens@nasa.gov	not involved, only related to proposals that assist users with discovery of data within data catalogues	not involved,	not involved	not involved	Our primary mission: http://gcmd.nasa.gov The GCMD holds more than 28,000 Earth science data set and service descriptions, which cover subject areas within the Earth and environmental sciences. The project mission is to assist researchers, policy makers, and the public in the discovery of and access to data, related services, and ancillary information (which includes descriptions of instruments and platforms) relevant to global change and Earth science research. Within this mission, the directory also offers online authoring tools to providers of data and services, facilitating the capability to make their products available to the Earth science community. In addition, citation information to properly credit data set contributions is offered, along with direct links to data and services. As an integral part of the project, keyword vocabularies have been developed and are constantly being refined and expanded. These vocabularies are also used in other applications within the broader scientific community. Users may perform searches through the Directory's website using controlled keywords, free-text searches, map/date searches or any combination of these. Users may also search or refine a search by data center, location, instrument, platform, project, or temporal/spatial resolution.	n/a	n/a	n/a

Research Data Life Cycle Matrix		Date: 10 January 2013						
					[1]	[2]		
Organization	Proposal	Acquisition	Analysis & Synthesis	Contribution	Discovery & Access	Use & Reuse	Publication	Preservation
NASA GSFC (GES-DISC) Mo Khayat	Typically not involved, some missions however may use our data center as an archive center or via EOSDIS	not involved	Typically not involved, although we work with some missions in production related processing or algorithms	Archive and distribution center for atmospheric data; tools development to enhance external users of our data (like Mirador for search, Wizards for data subsetting, and Giovanni for exploration). Typically our contribution is of an enabling nature to help scientists or general public users find data of interest and provide them tools that could help them arrive at some information they want to extract. We also provide user services for issues with access, download, or usage of our web tools.	Develop tools for search and discovery of archived data and assist users with issues related to access or questions related to processing of data archived at our data center. Most data is publicly available to domestic and international users which in turn use these data to arrive at science of interest.	Production of data at various levels of refinement (ie, L2 to L3)	Publications related to promotion of data archived at GES-DISC, new tools developed, or extracting new information from data archived at our center.	Major involvement in preservation of data from EOSDIS assigned missions. Preservation of related documents and meta data.

1. cchandler:
enter the names / acronyms of the main communities you work with / represent
2. cchandler: data manager, administrator, librarian, PI/researcher, etc.