DATA MANAGEMENT PLANNING FOR NASA SUPPORTED PLANETARY ANALOGUE SCIENCE.

L.C. Welzenbach[1], M. Glimoclija[2], A. Steele[3], and M.D. Fries[4] [1]Galactic Analytics (gameteorite@gmail.com), [2]Dept. of Earth and Environmental Sciences, Rutgers University (m.glamoclija@rutgers.edu) [3]Geophysical Laboratory, Carnegie Institution of Washington, (asteele@carnegiescience.edu.) [1]NASA-JSC (Marc.d.fries@NASA.gov)

Introduction: All proposals submitted to NASA ROSES are required to provide a data management plan (DMP)[5]. The reasons why are outlined in NASA's *Plan for Increasing Access to the Results of Scientific Research,* published December of 2014: "NASA is committed to following Federal guidelines that all data from federally funded research should be made as widely and freely available as possible while safeguarding the privacy of participants and protecting confidential and proprietary data…" [1]

While preliminary data or analyses, laboratory notebooks, physical objects, laboratory samples or specimens are exempted from the DMP, the information has use for other data archives and sample repositories. The repository selected and data captured should minimally conform to NASA guidelines for data archiving, e.g NASA Planetary Data System (PDS) [2]. Given that much of the data generated by planetary analogue research may not fit in to PDS, other repositories can be selected that align with science goals and types of data generated.

Planetary Analogue Research: Planetary analogue research typically includes field collection of geologic samples, and therefore should include a detailed curation and data management *plan* to collect and manage a comprehensive set of metadata. That metadata should meet the needs of a complex array of scientific goals and programmatic requirements. Defining the range of information to be captured requires a set of standards and protocols to meet the needs of different disciplines (planetary petrology, organic chemistry, biology, microbiology, engineering, etc.) which will accommodate the proposal's science goals. An additional important consideration for establishing a consistent, archival, and accessible data management system for planetary analogue samples is the eventual transferal of these collections to agencies or institutional repositories.

The goals for such a plan are outlined in *NASA ASTEP* grant *NNX14AT28G Preservation of Life Signatures in Cached Evaporites – Use of a Planetary Analogue to develop sampling, curation and a database of analogue sites and samples for use in Solar System Exploration*. From those proposed goals, we created a Planetary Analogue Management system (PAMS) to broadly outline an approach for identifying data collection and data management requirements. This current effort focuses on identifying data standards, exploring sources of data management planning and archiving, and tests an established sample data management system using Arctic Mars Analogueue Svalbard Expedition (AMASE) data [8,9].

Data standards. Sample metadata, along with generated analytical data, are recognized assets on a national scale [3]. There are many sources of information regarding data standards by discipline, tools for generating required metadata, and data archiving resources. Data captured will adhere to established metadata standards (e.g. Federal Geographic Data Committee standards) which can be generated online through tools provided by the USGS. [4] Data and metadata are selected for optimal migration to established data archives such as PetDB, MetDB and others available through the National Science Foundation (NSF) supported Interdisciplinary Earth Data Alliance (IEDA). Data standards will be established based on the data repository selected prior to data collection. Data would be managed through an established database (e.g. DREAM-Medusa) with a schema that allows data to be transferred to the appropriate data repositories (e.g. PetDB, EarthChem, etc.) Core data to be captured identified for field activities and curation, with science discipline/goal specific data will be finalized in future work. It is recommended that a dedicated data curator manage all phases of data collection and management, and educate of all parties to data collection and curation requirements.

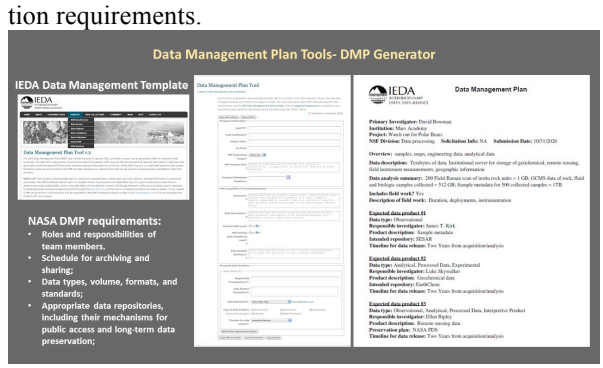

Fig. 1: IEDA DMP generator for NSF proposals

Data mangement planning tools:

National Science Foundation data management tool generator. The recommended data management plan for planetary analogue research could be adapted and generated from a modified version of the IEDA DMP tool (Fig. 1) This tool meets requirements for archiving data in most IEDA listed repositories, and should meet the NASA ROSES data management plan guidelines [5].

DREAM-Medusa. DREAM (Depository of References for Earth and Analytical Materials) is an institutional-level, archival system for managing geologic samples selected as a pilot database system for PAMS. The system and software (called Medusa) were designed and implemented at the Institute for Study of the Earth's Interior (ISEI) of Okayama University in Misasa, Japan [6]. DREAM builds on the efforts of IEDA data repositories by including architecture that manages field data, samples and analytical data in a laboratory setting, and it documents curation activities (Fig. 2).

Comparison of geochemical database systems

Database system	Target	Geochemical dataset	*In situ* visual	Curation	Open source
GEOROC[a]	igneous	○			
GeoReM[b]	RMs[b]	○			
PetDB[a]	seafloor	○			
SedDB[a]	sediment	○			
NAVDAT[a]	igneous[c]	○			
GRL[b,d]	any kind	○			
SESAR[e]	any kind			○	
MetPetDB[a,f]	metamorphic	○	○	○	
Medusa DB[g]	any kind	○	○	○	○

GEOROC, GeoReM, PetDB, SedDB, NAVDAT and GRL handle bulk and *in situ* geochemical datasets. MetPetDB handles both geochemical dataset and curation, but the target is limited to metamorphic rocks. 'Target' denotes the target rock type. 'Curation' indicates the ability to trace the current sample location and the location of any sample subsets (e.g., thin section, solution, etc). '*In situ* visual' indicates the ability to visualise locations of *in situ* analysis on images.
[a] Online database system.
[b] Any kind of rock but limited to reference materials.
[c] Sampled from North America only.
[d] Available at http://www.earthchem.org/grl/.
[e] Lehnert *et al.* (2000).
[f] Spear *et al.* (2009).
[g] Database system run by software proposed in this study.

Fig. 2: Table comparing DREAM with IEDA database schema from [6].

The DREAM system was selected because the design can be used to help develop protocols for data capture and management. The database schema and metadata are consistent with many IEDA data archives [7]. Unlike analytical data repositories such as PetDB, DREAM also captures complex sample-specific information including field information, analytical results, analytical methods, curation data, and it has the ability to link to multimedia information and bibliographic references. Developed in part by accountability, it compiles information about sample use, movement, location, and storage of scientific data with the sample. This has the added benefit of enabling decisions to be made about future use and preservation of a sample. Access to the system is through a web interface, and data entry can be portable, (used with android and iOS applications) so that sample curation and analytical collection workflow can be managed in real-time. The Medusa software is open source, modifiable, and uses data formats that allow for transfer of data and access within and among agencies and institutions.

AMASE Data. As a test of the DREAM-medusa system, we will apply data from the AMASE planetary analogue research program. AMASE data document field excursions to various localities of three years 2009, 2010 and 2011, for a total of 588 records. The primary data captured represent an inventory of samples by location, purpose, with a of physical attributes. The overall framework of the AMASE database is robust and contains many of the core fields required to document and archive each sample. There are shortcomings, however, wherein the recorded information is inconsistent across samples and users, and considerable variability exists in the quality and extent of the data captured. This is a common outcome given difficulties associated with field studies resulting from workload, physical burdens and user skill. For all of these reasons, the AMASE database is an excellent example from which to improve the process of data capture and managemet planning and education prior to the start of research.

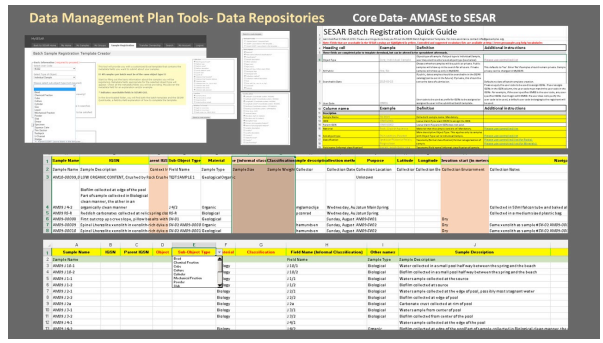

Fig. 3: First step to archive AMASE data is to create IGSN (International Geo Sample Number). Data are mapped to templates provided by IEDA.

Conclusions: Every DMP should include not only data standards, but a database system to manage and archive all data associated with NASA supported sample research. DREAM-Medusa is one option for a data system that is a powerful database system that is open source and modifiable. User interfaces are intuitive, versatile and easy to navigate. The database is available (http://dream.misasa.okayama-u.ac.jp/documentation/). Once DREAM-Medusa is fully installed and modified for PAMS schema and metadata, AMASE data will be imported for testing. AMASE data are also being mapped to templates for future importation to IEDA data archives (Fig. 3).

References:

[1]www.nasa.gov/sites/default/files/atoms/files/206985_2015_nasa_plan-for-web.pdf
[2] pds.nasa.gov/tools/standards-reference.shtml
[3] www.whitehouse.gove/files/documents/ostp/NSTC%20Reports/Revision_1-22_09_CL.pdf
[4] www2.usgs.gov/datamanagement/index.php
[5] smd-prod.s3.amazonaws.com/science-green/s3fs-public/atoms/files/Data_Mgmt_Plan_guidelines-20110111.pdf [6] Yachi Y. et al. (2014) *Geostandards & Geoanalytical Research,* 38, 95-102. [7] Lehnert K. et al. (2000) *G*^{3} ,v.1, n.1 [8] Steele A. et al. (2009) *LPSC XLI,* Abstract #2398. [9] Conrad P. et al. (2010) *LPSC XLII,* Abstract #2218.