Multi-Instrument Database (MIND): Cloud-based Database System for DUV Raman and Fluorescence Research and Flight Instrument Engineering Telemetry I. J. Doloboff[1], V. M. Paez[2], E. J. Eshelman[1], E. Hara[3], G. Wanger[1], V. J. Orphan[4], J. P. Amend[3], L. Beegle[1], and R. Bhartia[1]. [1]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109 (Ivria.Doloboff@jpl.nasa.gov). [2]Georgia Institute of Technology, North Avenue, Atlanta, GA, 30332. [3]University of Southern California, Department of Earth Sciences, 3651 Trousdale Pkwy., Los Angeles, CA 90089. [4]California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125.

Introduction: Active management of data as an asset is now required by NASA for the essential support of scientific integrity, reproducibility, timely publication, and public access [1], [2]. The Multi-INstrument Database (MIND) project, funded by the NASA Astrobiology Institute (NAI), is a novel relatable data repository. It features a web-based front-end that permits user interaction with stored raw data. The current spectral library includes data from the unique suite of deep UV fluorescence/Raman instruments built at Jet Propulsion Laboratory (JPL), as well as commercial instruments used for sample analysis (i.e. visible Raman imagers, microscopic images, SEM, etc.). MIND serves three primary functions: 1) a long term repository that can track sample collection, store data, and enable cross-lab collaboration or use from multiple instrument platforms; 2) facilitate analysis of large sets of data, taking advantage of cloud computing capabilities; 3) allow correlative analyses from multiple instrument platforms. The challenge, therefore, is development of a data storage architecture that accommodates many data formats, and can enable swift retrieval of data taken from a given sample at a number of different scales (i.e. spatial). This well-organized relational database enables data visualization features that can quickly generate maps of a single sample at all analyzed spatial scales. MIND also is intended for future storage of SHERLOC (Scanning Habitable Environments with Raman and Luminescence for Organics and Chemicals – Mars 2020 [3]) component, subsystem, and system test data; storage of this data would support mission operations and future mission concept development.

Although digital scientific data is easily shareable, it is hampered by the absence of a basic ontology at the laboratory level. Using relational database theory, a defined taxonomy of data/instrument components, and functional characteristics modeled upon user actions, MIND acts as an ontological data storage and processing system. MIND operates as a carefully curated, nimble, and secure data management system for long- and near-term data access by laboratory, science operations, and the greater scientific community.

NAI Life Underground: This NASA Astrobiology institute funded program uses a multi-scalar approach to laboratory and field experiments, and complementary analytical techniques, to perform *in situ* life detection and characterization of microbial life in the subsurface. The combination of various analytical methods including deep UV Raman and fluorescence spectroscopy, electron microscopy, and molecular biological techniques, operating over multiple spatial and spectral scales, demands efficient data archiving to compare multiple reference methods. The dynamic effort derives from research performed at diverse subsurface environments of astrobiological interest; therefore, curating and coordinating the multifarious data is of equal importance. The NAI Life Underground project that funds MIND does so because the unified effort will help to answer questions regarding the subsurface microbial distribution and potential for biosignature formation and potential for preservation through the alteration of the local geochemical environment.

The Spectral Pipeline: Data collected with the suite of deep UV Raman and fluorescence, visible Raman, and electron microscopy instruments at JPL constitute the first sequence of our spectral pipeline. This pipeline is a recurring down-selection of sample spot analysis using different techniques afforded by the robust suite of instrumentation, followed by subsequent handling of raw and derived data. Beginning with location of organics and microbes, the pipeline continues classification of the biomass and related minerals, before correlation of these features with minerals and other substrates. Generating layered spectral maps guided by principal component analysis (PCA) of both initial and down-selected data, facilitates detection of biosignatures derived from the correlations between the data sets (i.e. correlation between microbial activity and inorganic/organic features). Use of data from the sequences of this spectral pipeline as well as spectral standards, obligate a robust data management system and database for retrieving, analyzing, sharing, and saving these data sets. MIND enables research efforts that function across a diverse array of field sites and instruments to ensure access, preservation, and organization of data that will be used in future publications, proposals, and research opportunities.

Database Architecture: MIND is built on a website front-end (HTML/CSS/JQuery) and a database back-end (PHP/MySQL) hosted on Amazon gov-cloud servers. The MIND ontology includes all data products generated by an instrument or experiment—it is sample-centric. Uploading data to MIND can be performed manually via the web interface or automatically via PHP scripts and within LabVIEW™ programs operating on various instruments and test assemblies. Individual experiments are associated with a particular sample, and that sample may have an arbitrary number of experiments from multiple instruments associated with it. The centralized database allows access to, and production of, raw, working, derived (versioned), and analyzed data in one place that is accessible to all authorized users. An integrated QR code-based physical sample storage system enables detailed tracking of samples throughout workflow. Linking of samples via parent-child-sibling relationships is a non-trivial feature that provides a large proportion of the flexibility offered by MIND. Data visualization in the form of quicklooks generated by PHP scripts provide rapid visualization of individual data channels within an experiment. An anticipated future update will include spectral processing capabilities. MIND is in the early stages of performing large-scale analyses that will permit a searchable spectral library capable of peak searching and identification. A PHP layer performs the extraction of large sets of data based on specified criteria provided by a LabVIEW™ program. In the absence of software capable of spectral processing with large datasets, cloud-based computing for spectral data processing and visualization is desirable. Using MIND to perform computing that a local computer cannot is one of the benefits of using a cloud-based database.

SHERLOC (Mars 2020): As SHERLOC begins testing of components, sub-assemblies, and full system tests as part of the instrument development process, the volume of data, and the diversity of data formats generated, will be immense. Laser test data, for example, is collected for the ultimate purpose of measuring the stability of M^2—a calculation of the divergence of a pure Gaussian unfocused beam propagating through space—and the average output energy of the laser, as well as evaluating spot size as a function of ambient temperature. Laser lifetime testing is conducted in a Martian simulation chamber to evaluate performance in Martian atmospheric conditions. Laser beam images as well as energy, temperature, and pressure readings are collected as individual data channels. These data demonstrate individual laser performance as well as comparisons between multiple lasers or assemblies. Demonstrating that a laser or a given sub-assembly fulfills verify and validate project requirements necessitates organization, correlation, and analysis of all associated data.

MIND will function as the data management system for SHERLOC engineering telemetry, dividing data into two streams: derived data/analysis; and hardware records associated with individual test procedures, assemblies, and models. In this instance, data visualization using quicklooks enabled by a JavaScript graphing library provide easy access to thermal, spectral, and electrical plots for each experiment. To address the needs of SHERLOC test data, organization of individual components and levels of a project/assembly can be mapped onto the existing MIND data topology. For example, metadata standards established in the MIND ontology can accommodate verify and validate project requirements by dividing Program, Project, System, and Sub-System sessions. The sample-centric nature of the relational database and the linking of samples via parent-child-sibling relationships the structure affords, are what make MIND universally flexible; a singular data ontology can support both scientific research data and flight instrument engineering telemetry.

Conclusion: One of the most compelling features of deep UV spectroscopic methods is that they are non-contact and non-destructive; unaltered samples are available for future downstream analysis by complementary data acquisition methods [4]. This feature is augmented when the reusable assets (i.e. data) are captured, curated with distinct attributes and associations, organized carefully, and made easily accessible to researchers. With the data generated by laboratory-based and *in situ* downhole instruments in one place, the reusable assets stored in MIND can support not only NAI Life Underground's spectral pipeline, but also future research efforts, and mission instrument development.

References: [1] OSTP. (2013) *Memorandum for the Heads of Executive Departments and Agencies*. [2] NASA, (2014) *Memorandum: Plan for Increasing Access*. [3] Beegle, L. W. et al. (2015) *IEEE,* 90, 1-11 [4] Bhartia et al., (2008) *AS*, 62(10):1070-7.