TOWARDS A TERRESTRIAL ANALOGS DATA PORTAL: USE CASES AND REQUIREMENTS. D.M.H Baker[1], M.E. Rumpf[2], M.A. Hunter[2], L.P. Keszthelyi[2], J.A. Richardson[1,3], P.L. Whelley[1,3], N.M. Schmerr[3], K.E. Young[1]. [1]NASA Goddard Space Flight Center, Greenbelt, MD 20771 (david.m.hollibaughbaker@nasa.gov), [2]U.S. Geological Survey Astrogeology Science Center, Flagstaff, AZ 86001, [3]Univ. of Maryland, College Park, MD 20742.

Introduction: Terrestrial analog field data are important for ground-truthing and testing hypotheses on a range of scientific and technical topics in planetary science and exploration. Due to their interdisciplinary nature and intersection with both the Earth and planetary sciences, analog field data do not fit easily into traditional Earth science or planetary science archives or repositories, causing most data to reside with individual researchers or in disparate locations, hindering both accessibility and discoverability. A recent survey conducted by the USGS Astrogeology Science Center [1] and the final report of the Planetary Data Ecosystem (PDE) Independent Review Board (IRB) [2] emphasize a need to establish and maintain the ability to serve, preserve, and link to the diverse and extensive terrestrial analog field data produced by NASA programs.

To this end, the USGS Astrogeology Science Center has been working to establish a Terrestrial Analog Data Portal (TADP), with the goal to open it as a data repository of terrestrial analog field data for all NASA-funded researchers [3]. Recent collaboration with NASA-funded research teams have provided valuable use cases to test ingestion and refine requirements for TADP. We summarize the technical foundation for such a portal and list findings from use cases that can feed into future requirements and best practices.

Terrestrial Analog Data Portal (TADP): The data portal is founded on the USGS ScienceBase Catalog, which is built on open standards for metadata, with an item core based on the Dublin Core Metadata Element [4]. ScienceBase shares geospatial data via Open Geospatial Consortium (OGC) standards and common commercial formats, and can also link to data hosted by other institutions. Key elements of ScienceBase include: 1) A data cataloging and collaborative data management platform; 2) Central search and discovery application; 3) Web services facilitating other applications; and 4) Research community catalogs.

ScienceBase items use persistent and unique URLs and are accessible through an application programming interface (API). The core functionality is also enhanced with custom extensions (e.g., ArcGIS REST Service). Further, the USGS ScienceBase Catalog serves as a Trusted Digital Repository, with long-term preservation provided by the USGS Core Science Systems and meets all the legal and functional requirements for NASA terrestrial analog data management.

Use Cases: The TADP is in its early stages of implementation. To further refine the requirements for TDAP and to fulfill recognized needs for preservation and public access to data products, the USGS is actively working with two groups of field analog researchers, the NASA Goddard Instrument Field Team (GIFT) and the SSERVI GEODES team.

The NASA GIFT team has conducted a range of field investigations and collected a range of data types over the past few years, including LiDAR, hXRF, UAS, GPR, and dGPS, as well as field samples with follow-on laboratory analyses. GEODES is specifically focused on the acquisition of geophysical field data, including seismic, GPR, magnetics, and gravity.

The GIFT and GEODES teams are currently preparing a number of data collections for archiving and working with the USGS TADP to ingest and release these data to the public. In doing so, the teams continue to identify a number of considerations for further development of the TADP and for community best practices in the management of field analog data:

1) Metadata and Documentation: Quality metadata and documentation is essential for making any data archive accessible, discoverable, and usable. TADP uses open metadata standards and provides a set of minimum requirements for ingestion of data into the repository. ScienceBase also provides a metadata wizard that facilitates formatting and validation of metadata files.

While the minimum metadata requirements facilitates ingestion into TADP, in practice, making the data useable to a researcher unfamiliar with the field environment and data acquisition requires more thorough documentation and metadata definition. For example, GEODES and GIFT teams have developed ReadMe templates that seek to capture a number of common and critical elements to describe both the field expedition, data archive structure, data collection methods, and format, uage considerations, and other important information. Expanding the set of recommended metadata elements and documentation in the TADP data repository would enhance user understanding and usability.

2) Processing Levels: Human collection of field measurements using diverse field instruments and post-processing procedures ensures that each dataset will be unique to each field team and site. We have found it important to clearly structure and document the archive levels to elucidate understanding of the types of data reduction methods. We follow basic PDS4-type processing levels: Raw, Processed, and Derived. Having both raw and processed data ensure that potential data user needs and understanding are met. Proper documentation is critical for a user's understanding of the data types.

3) Linkages to external data sources: For a number of reasons, projects may find it preferable to host data at an institutional repository or other community repository. There are also data already ingested in individual repositories. To enable a comprehensive search and discovery of these external sources, a primary repository must be able to discover and register these data sources and bring them into the search system.

TADP allows linkage to these datasets. GEODES and GIFT teams are testing both direct archiving and external linkage scenarios. An immediate issue linking to external archives is the variability in metadata and standards (or absence thereof). A community agreed-upon set of best archiving practices would help to align other repositories with TADP.

4) Geospatial search and discovery: The GEODES team found great value in enabling GIS-based search through selectable shapefiles of observation locations that link to the archive landing pages. TADP enables a geospatial search and retrieval through its GIS web extensions. This is enabled by complete capture of metadata and supplemental data (e.g., shapefiles of observation locations). Making geospatial search operable with external data sources where the metadata are insufficient is a challenge that could be rectified by community adoption of best practices.

5) Common dictionaries: Analog data do not have a common dictionary of metadata attributes and syntax for use of the variety of field environments and measurements. As a result, while data may be internally consistent within a field team, different teams often use different nomenclature for the same features, processes, and measurements, etc. This greatly complicates search and discovery across the repository. The community should consider how a common dictionary would be developed and maintained at a reasonable level of effort.

6) Linkages to field samples and laboratory data. Field measurements are often accompanied by collected field samples and follow-on laboratory measurements of these samples. Linkages to digital measurements or published results is more tractable than with physical samples. Proper planning prior to field acquisition is required to capture common metadata threads to enable linkages between data types. Most commonly, these can be geospatial or temporal stamps or naming conventions and metadata tags can also assist. Again, a set of best practices is needed to guide future field teams.

7) Linkages to the broader Planetary Data Ecosystem (PDE): A future goal of the TADP should be to be plugged into the larger PDE, including planetary mission data and sample/lab repositories and services.

8) Low barrier to data management: Keeping the barriers to data management low increases the likelihood that more researchers will add to the repository. Indeed, one of the findings of the PDE IRB was the significant cost and time spent on archiving in formal archives such as the PDS. A balance must be made between data quality and the effort involved in data preparation and archiving. The ScienceBase Catalog is being developed and enhanced to facilitate depositing of data by individual researchers and teams.

9) Data format and size: The diversity of field measurement types and volumes requires the ability to accommodate a range of formats and sizes. For example, data can range from kilobyte csv files to LiDAR point clouds tens of gigabytes in size.

10) Citability and data accumulation: Citable data sources are now required for a number of journals prior to publication. TADP provides DOIs to data collections that satisfies this requirement. Further, as field expeditions often span multiple years, the ability to add to update collections of data are needed.

State of TADP and Future Development: TADP provides a solid foundation for growing a much needed repository for terrestrial analog field data. Our use cases show that there are a multitude of considerations that should be translated to requirements for maximizing the data portal's potential and ability to meet the needs of the community. A field analog community group that develops and promotes best practices for metadata, standards, data formats, and documentation is essential to further the goals of TADP.

Further, as a terrestrial analogs data repository grows and is recognized as an important service, sustaining it will require a commitment and support from NASA. We hope that TADP can serve well into the future to ensure that the full scientific return on investment from terrestrial analog field data is realized.

Acknowledgments: This work was supported by the SSERVI GEODES, NASA GIFT ISFM, and USGS Astrogeology Science Center under an interagency agreement with NASA.

References: [1] Rumpf, M.E. et al. (2019) *AGU Fall Meeting* 2019. Abstract #EP21E-2211. [2] NASA (2021) Planetary Data Ecosystem Independent Review Board Final Report, https://science.nasa.gov/researchers/science-data. [3] Keszthelyi, L. et al. (2019) *4th Planetary Data Workshop* 2019. Abstract #7023. [4] ScienceBase Wiki, https://my.usgs.gov/confluence/display/sciencebase/ScienceBase, Accessed April 2021.