PLANETARY GEOSCIENCE ONTOLOGY TESTBED: USING SEMANTIC TECHNOLOGY TO ENHANCE THE DISCOVERY OF GEOLOGIC MAPS. M. A. Hunter[1], J. A. Skinner[1], C. M. Fortezzo[1], and C. Okubo[1], [1]USGS Astrogeology Science Center (2255 N. Gemini Dr., Flagstaff, AZ 86004)

Introduction: Over the past fifteen years the World Wide Web has become increasingly more accessible, despite the voluminous growth in content. This is due in large part to the development of web semantic technologies, which allow for more accurate and relevant query results through explicitly defined relationships between conceptual objects [1], rather than relying on string matching alone. The Resource Description Framework (RDF) language is a commonly used, eXtensible Markup Language (XML)-encoded format for describing such relationships as subject-predicate-object triples, in which each element references a Universal Resource Identifier (URI) [2].

As part of a larger effort to test and advocate for open solutions within planetary spatial data infrastructures (PSDIs), the U.S. Geological Survey (USGS) Astrogeology Science Center is testing an interdisciplinary planetary geoscience ontology as a means to enhance data discovery [3]. This ontology is being authored in World Wide Web Consortium standard Web Ontology Language (OWL2), and managed via Stanford's Protégé v5.5 application [4]. A major component of this testbed is to assess the semantic potential of geologic maps as released in a Geographic Information System (GIS) form.

Current Work: For a user to determine if a given geologic map is potentially of interest they must synthesize the description, interpretation, and understand the components like the correlation of map units, and symbols used for topography, structures, and mapped features. All of these map elements must work for the scientific information to be communicated a coherent manner. This ontology is being designed to model such relationships between map elements while concurrently testing an expanded schema for geologic maps in GIS formats, and International Organization for Standardization (ISO)-19139 metadata (the XML schema for describing geospatial datasets). This effort is not intended to 'read' maps as people do, but rather, expose standard map information alongside related planetary geoscience data in a machine-readable format that supports semantic queries (i.e., SPARQL Protocol and RDF Query Language). When combined with other geospatial datasets such as International Astronautical Union-approved nomenclature, quadrangle boundaries, dune and cave catalogs also represented in the ontology, they become much more accessible to the entire community than in isolation.

To demonstrate this capability, two overlapping geologic maps of different scales have been migrated into an extended schema which captures the full name, description, and interpretation of units, as well as their geologic age as epoch and major group (Figure 1). Each unit description was assigned a geomorphic property and each interpretation a geologic process, both enumerated in the Planetary Geoscience ontology. Subclasses in different branches of the ontology are connected by custom predicates such as "kindOf", "partOf", "representationOf", "constrainedTo", "mappedOn", "ofBody", and "publishedBy". These relationships, coupled with geographic properties of individual features and enhanced metadata, provide multiple avenues through which a user may discover a given dataset.

The development of this ontology has largely followed guidance from Noy and McGuiness' 2011 *Ontology 101* publication, where classes, their properties (slots) and values (faces) are introduced only where necessary to make a distinction necessary to determine if a dataset may be of interest [5]. Core competency questions were developed to test ideal query criteria, and have been a guiding force of the ontology structure. These benchmark questions have already highlighted some basic limitations of the test data, which are due to inconsistencies between maps, across authors, scales and basemaps used.

An early finding of this work is that legacy maps from pre-Viking Orbiter imagery may not be translated into standard geologic process and properties as reliably as recent maps, requiring significant deviation from the authors' original interpretations. The ontology continues to evolve in an iterative manner as new ideas are introduced, tested, and discussed, and can be expected to change significantly from its current state.

Future Work: In order to determine the effectiveness of this methodology, and provide findings to inform future work, the full set of testbed data and metadata must be validated and migrated into a Postgres database, and then mapped to the OWL2 ontology. This has typically been accomplished through a series of steps, as documented in USGS OFR 2011-1142, in which data are converted from a database to Geographic Markup Language (GML), and then to RDF triples [6]. Recent works have utilized applications such as the Protégé's OnTop extension, which is capable of mapping class properties in an RDF ontology directly to columns in a spatially-enabled Postgres database, and we expect to use this approach in the testbed [7]. Once this has been completed, a set of basic SPARQL queries will be used to demonstrate basic functionality and the potential for implementing this approach across a greater variety and volume of data.

Results of this research will be included with findings from nomenclature, cave and dune databases as a recommendation for geospatial data access as part of a realized PSDI. This includes serving maps as an Open Geospatial Consortium (OGC) standard Web Feature Service (WFS), exposing service metadata through an OGC Catalog Service for the Web (CSW), as well as developing open tools, schemas and policies to support collaboration across institutions. Future work may bring this subject into the broader planetary science community so that all data custodians have the opportunity to participate in an open data community while empowering end users to find useful data from documented sources.

Specific to planetary geologic mapping, this work will inform decisions in the near future on how to best structure mapping projects in order to make them more useful to geologists and the larger planetary science community. Multiple terrestrial geologic data models such as the Geoscience Markup Language (GeoSciML) and USGS Geologic Map Schema (GeMS) are available to adopt/extend, or serve as examples to emulate. Any future solution will be made publicly available with the ultimate goal of integrating USGS-published maps with geologic maps from other sources.

References: [1] Kamel, M. et al. (2007). Ontology-aided vs. keyword-based web searches: a comparative user study. *Large Scale Semantic Access to Content*, 639-652. [2] Berners-Lee, T. et al. (2001). The semantic web. *Scientific American*, *284*(5), 34-43. [3] Laura, J. R. et al. (2017). Towards a Planetary Spatial Data Infrastructure. *ISPRS*, *6*(6), 181. [4] Musen, M. A. (2015). The Protégé project: A look back and a look forward. *Association of Computing Machinery Specific Interest Group in Artificial Intelligence*, 1(4). [5] Noy, N. F., and McGuinness, D. L.. (2001). "Ontology development 101: A guide to creating your first ontology". [6] Bulen, A. et al. (2011). A program for the conversion of the national map data from proprietary format to resource description framework (RDF) *USGS OFR 2011-1142*. [7] Calvanese, D. et al. (2017). Ontop: Answering SPARQL queries over relational databases. *Semantic Web*, *8*(3), 471-487.

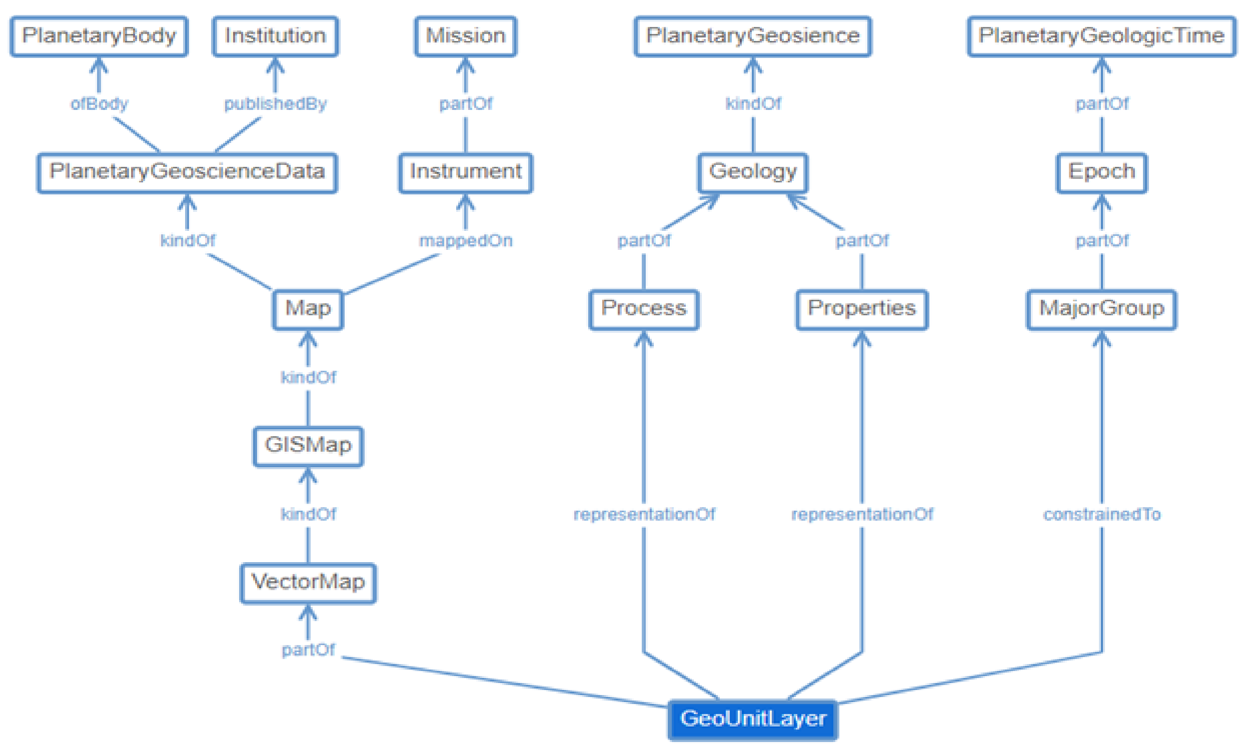

Figure 1. Graphical view of the geologic unit subclass in Web Protégé.