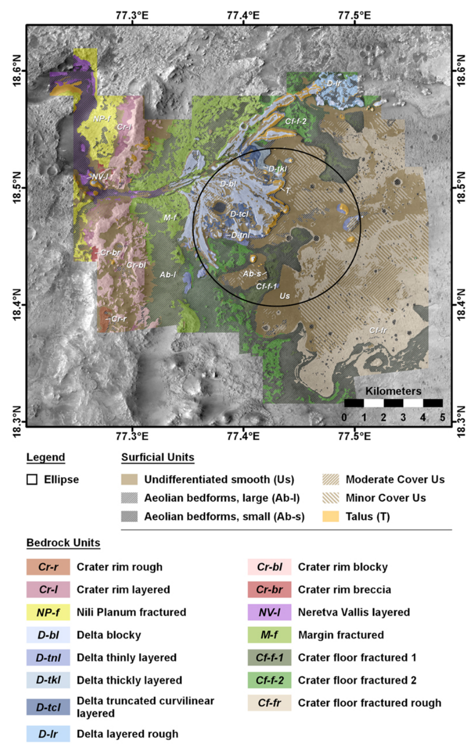
CS486C – Senior Capstone Design in Computer Science

Project Description

Project Title: Planetary Maps: One Tool to Find Them All	
Sponsor Information: USGS science for a changing world	Sarah Black, Research Physical Scientist - sblack@usgs.gov Marc Hunter, IT Specialist - mahunter@usgs.gov Planetary Geologic Mapping Program USGS Astrogeology Science Center

Project Overview:

Space exploration relies on cutting edge technologies and an enormous body of knowledge amassed by the planetary science community over the past several decades. One facet of planetary science is planetary geologic mapping. Planetary geologic maps are used to inform space missions during both planning and exploration phases and are also used by researchers conducting their own science investigations beyond mission operations. For example, on February 18, 2021, the Mars 2020 Perseverance rover landed in Jezero crater to begin its exploration of the Martian surface. Before landing, the Mars 2020 science team mapped the geology of the landing site and surrounding region, and is now using this map (shown here) to identify regions of scientific interest for mission planning. Mars scientists will also use this map (and others) to explore the local and regional geology and to extrapolate in situ (in place) observations to areas that the rover can not reach.



The NASA-USGS Planetary Geologic Mapping (PGM) Program is based out of the USGS Astrogeology Science Center in Flagstaff, Arizona. Members of the PGM group work to support the planetary science community, while also conducting research of their own. One primary focus of the PGM group is creating tools and resources that can be used by the planetary science community to make better use of the large volume of data that is available, especially in finding and using published geologic maps.

One challenge scientists face when beginning a project is finding all geologic (or other related thematic) maps that have been made for a particular region of interest. There are two main avenues for publishing planetary maps: 1) through the USGS as standardized map products, and 2) through journal articles and conference proceedings. All USGS-published maps are discoverable through the USGS Planetary Geologic Mapping Program website. However, because there are specific requirements and standards for a USGS geologic map, only a portion of planetary maps are published through the USGS. Thus, a large percentage of the mapping work that is conducted by the planetary science community is ultimately shared through non-USGS publications. These maps are also extremely valuable informational products. Currently, there is no singular, comprehensive, searchable database for planetary maps that have been published through journal articles or conference proceedings. The lack of this tool is notable in the planetary science community, with many scientists voicing frustration over the difficulties of finding and using non-USGS maps. Without a comprehensive database of planetary maps, researchers and mission scientists lack an efficient way to search for existing resources. Ultimately, they may never find existing map products which would be helpful, and potentially impactful, to their work. This may also result in repeated efforts and certainly a reduction in efficiency and investment. The inability for scientists to find all available map products can also lead to and perpetuate citation bias – where the most often cited maps continue to be used, and others are "lost."

Over the years, members of the planetary science community have tried to address this issue. However, due to various limitations on funding and time, none of these projects have progressed very far. For example, Dr. Kiri Wagstaff (JPL) used machine learning to create a local data dictionary and image content classifiers for the Planetary Data System (complete publications list here), and Dr. Henrik Hargitai (Eotvos Lorand University) worked to build the International Catalog of Planetary Maps (related paper here). While these are both impressive works, neither are fully automated nor comprehensive. However, these serve as excellent examples, resources, and inspiration for this project.

What we envision to address the above shortcomings is a singular, comprehensive, and searchable database which contains references to all planetary maps published through non-USGS venues. The final product should be capable of automatically searching for, identifying, and adding new map publications as database items, and be available to the community through a simple web interface. Some of the key functions should include:

Minimum Viable Product – A Python-based tool that can automatically search for, identify, and add published map products to a simple database. Abilities would include:

- Recognize when new publications are added to online journal records and other formats of interest (e.g., conference abstracts). *Note: If needed, we can provide available journals and other sources for searching*
- Using machine learning techniques and a training dataset, train an algorithm to automatically scan new publication records and determine if they contain map products (flagged with a defined level of uncertainty)
- For publications that have been flagged as containing map products:
    - Identify and extract relevant metadata, including a link to the publication itself
    - Check for duplicates against existing database records
      Note: maps sometimes result in multiple publications or have progress presented over multiple years of conferences; for example, this in-progress map was presented over three years of the Planetary Geologic Mappers Meeting: 2015, 2016, 2017. This is a common occurrence and is a requirement of some NASA funding programs.
    - If this is a new map publication – add publication to the database as a new item, including relevant metadata
    - If this is a duplicate of an existing database record – add to existing record as additional information
    - Include a method to manually assess found records, sorted by uncertainty, such that they can be easily removed if they are not applicable. Perhaps use these false positives to refine and re-train the algorithm.
- The resulting database will be hosted by USGS Astrogeology and made available through the USGS Planetary Geologic Mapping Program website (or similar), and automatically kept up to date with any new additions.
- End users can download and view the database in a simple excel-style table format, and sort/search various attributes (e.g., planetary body, geographic location, publication date, scale) to locate relevant maps.

A complete system – this product would enable all the above functionality, in addition to:

- End users interact with and search the database through a web-based GUI, hosted through the USGS Planetary Geologic Mapping Program website. Functions supported by the GUI might include:
    - Perform a simple search using text and dropdown menus (e.g., search for a feature name, author)
    - View, sort (e.g., by date), and explore search results, including links to the original publications
    - Export search results (and all associated metadata) to a local .csv file
    - User accounts and profiles.
    - Users can create "searches" based on targeted criteria, and can name and save these, to re-run at any time with a simple click.
    - Users can configure periodic searches to run automatically, and be notified of new additions.

The end goal (stretch goals) – this product would enable all the above functionality, in addition to:

- The web-based GUI has all the functions outlined above, plus:
    - The ability for users to search on a map by drawing a region of interest.

Note: Though not required, developers may want to use the National Geologic Map Database as inspiration (or people to talk to for lessons learned, etc.) for this deliverable. Additionally, developers may want to use CartoCosmos as a starting point for the map tool. This is an excellent, fully-functional map interface that was created by a previous NAU CS capstone team.

- User-created author profiles (similar to Google Scholar), which would allow the following functions:
  - Community input and validation – Users can add database entries for maps that did not get detected by the automatic search. Users may also validate entries and correct errors.
  - Authors may add supplemental information to database entries. For example, map summaries, attachments (images, PDFs, GIS files), or links. This should only be allowed for entries on which they are listed as an author.
  - Authors may enable email notifications to alert them when a new database entry is added in which they are an author. This would allow them to view the new entry and validate it/make any necessary corrections and add supplemental information.

Successful completion of this project would be hugely impactful to the planetary science community around the world. This tool has been widely requested by the planetary science community, with scientists around the world voicing their support for its development. The current lack thereof causes valuable map products to be lost or go unused, which ultimately hinders scientists and prevents them from making use of beneficial and unique previous works. The impact of this tool would be on a global (some would say even interplanetary) scale. By creating a better system for search and discovery, planetary scientists can take advantage of all previous works, which will ultimately result in improved science and space exploration in the future.

Knowledge, skills, and expertise required for this project:

- Familiarity with Machine learning, and the various frameworks and tools available (Pytorch, Keras, Pandas, etc.)
- Containerization
- GUI/web development, for creating the web-application
- Cloud computing and data storage (AWS S3, Microsoft Azure, etc.)

Equipment Requirements:

- There should be no equipment or software required other than a development platform and software/tools freely available online. Cloud environments can be emulated through docker and S3 buckets using MinIO (https://docs.min.io/).

Software and other Deliverables:

- A strong as-built report detailing the design and implementation of the product in a complete, clear and professional manner. This document should provide a strong basis for future development of the product.
- A complete professionally-documented codebase, delivered both as a repository in GitHub, BitBucket, or some other version control repository; and as a physical archive on a USB drive.
- A docker container, ready to be deployed