System Development: Data Enrichment
Planetary Nomenclature Project

Golnaz Shapurian, Thomas Allen and the ADS Team

ADS Users Group Meeting, 9-10 Nov. 2022

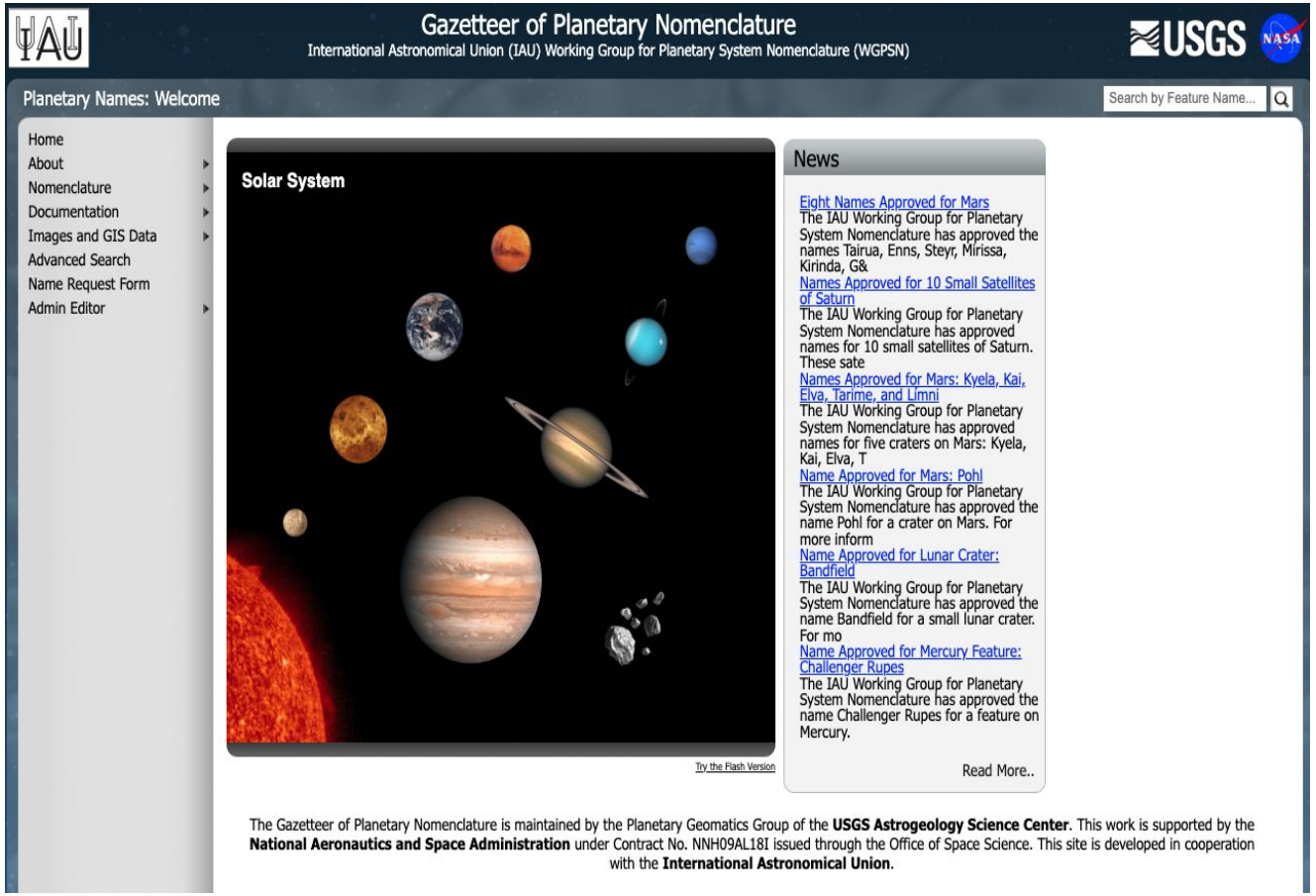
Smithsonian Institution

astrophysics data system

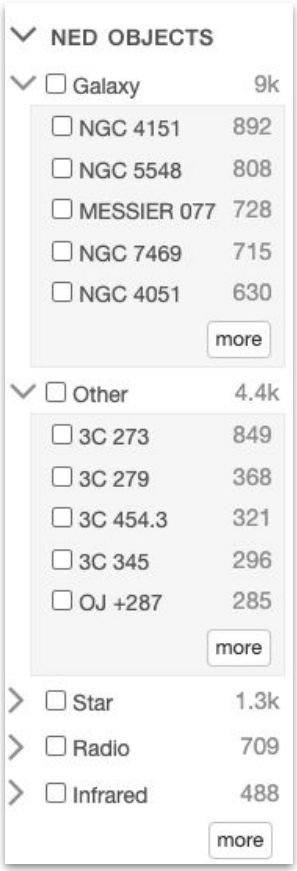
NASA

Entities of Interest to the Community Served by the Expansion

- Planetary Science Literature
  - 16,000 planetary feature names from the USGS/IAU Gazetteer of Planetary Nomenclature
  - Many names used on multiple objects - Styx
  - Many names overlap with other entities in the field - Herschel
  - Testing on Martian feature names



User Experience



Example of an ADS filter. The Planetary Names filter will be similar.

Planetary Names: Mars > Craters > Cassini

1 2020JGRE..12506104W 2020/03 cited: 12

Similarities and Differences of Global Dust Storms in MY 25, 28, and 34

Wolkenberg, P.; Giuranna, M.; Smith, M. Â. D. *and 2 more*

> *25–27°E) near the Cassini crater. 5 Seasonal variations of zonally averaged (a) dust opacities and (b)*

2 2019JGRE..124.1913D 2019/07 cited: 24

A Diverse Array of Fluvial Depositional Systems in Arabia Terra: Evidence for mid-Noachian to Early Hesperian Rivers on Mars

Davis, Joel M.; Gupta, Sanjeev; Balme, Matthew *and 4 more*

> *paleolake deposit in Cassini crater. (e) HiRISE image of inverted paleolake deposit associated with the Aram*

3 2017JGRE..122.2294V 2017/11 cited: 11

Constraining the Date of the Martian Dynamo Shutdown by Means of Crater Magnetization Signatures

Vervelidou, Foteini; Lesur, Vincent; Grott, Matthias *and 2 more*

> *Huygens, Cassini, Antoniadi, Epsilon, Tikhonravov, Eta, Iota, and Herschel. In case of overlap*

Objective & Challenges

- Objective
  - to identify United States Geological Survey (USGS) terms in the manuscripts archived by ADS
- Challenges
  - 16013 terms identified for 45 targets (ie, Mars, Moon, etc) and 54 types (ie, Crater, Albedo Feature, etc)
  - Duplicates
    - Adams: crater on Mars, named after Walter S.; American astronomer (1876-1956).
    - Adams: crater on Moon named after John Couch; British astronomer (1819-1892); Charles Hitchcock; American astronomer (1868-1951); Walter Sydney; American astronomer (1876-1956).

Objective & Challenges – cont.

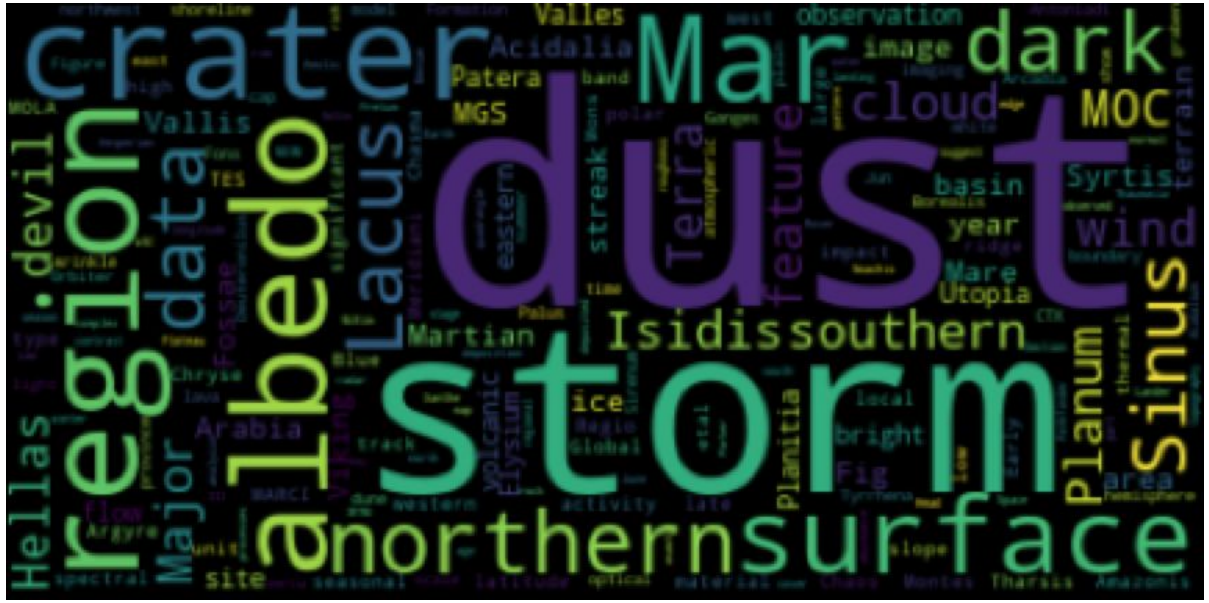
- Challenges
  - Span
    - Arabia: Albedo Feature on Mars.
    - Arabia Terra: Terra on Mars.
  - Relevance
    - Adams: Martian Crater, Lunar crater, Asteroid, President of US, City/Town, Family name.
    - Herschel: Martian Crater, Lunar Crater, Crater on Mimas, ESA Space Observatory, Asteroid, Reflecting Telescope in England, Telescope in Spain, Family name, Company name, English Science Award, Island in Canada, Mountain in Antarctica, Town.

Hypothesis & Approach

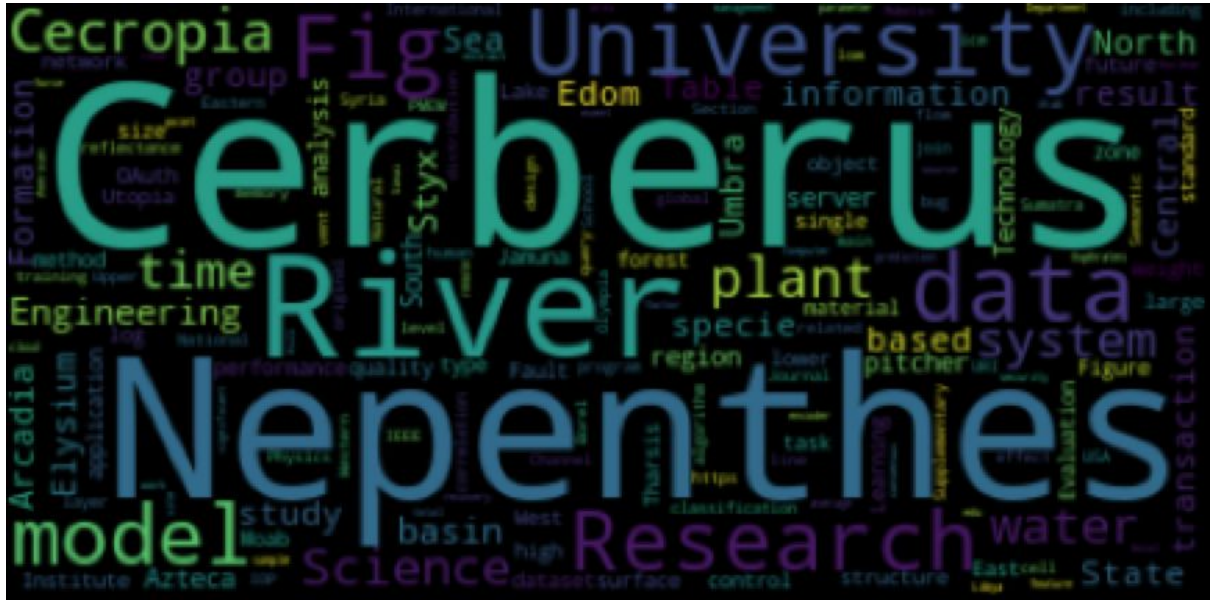
- Hypothesis
  - Capture the context of the term from the tokens around the term.
- Approach
  - Identify the term in the full text.
  - Select 128 tokens around the identified term.
  - Extract keywords, up to 10, from the selected text.
  - Decide the context of the term from the keywords.
    - How?

Word Cloud of the Mars-Albedo Feature

Planetary



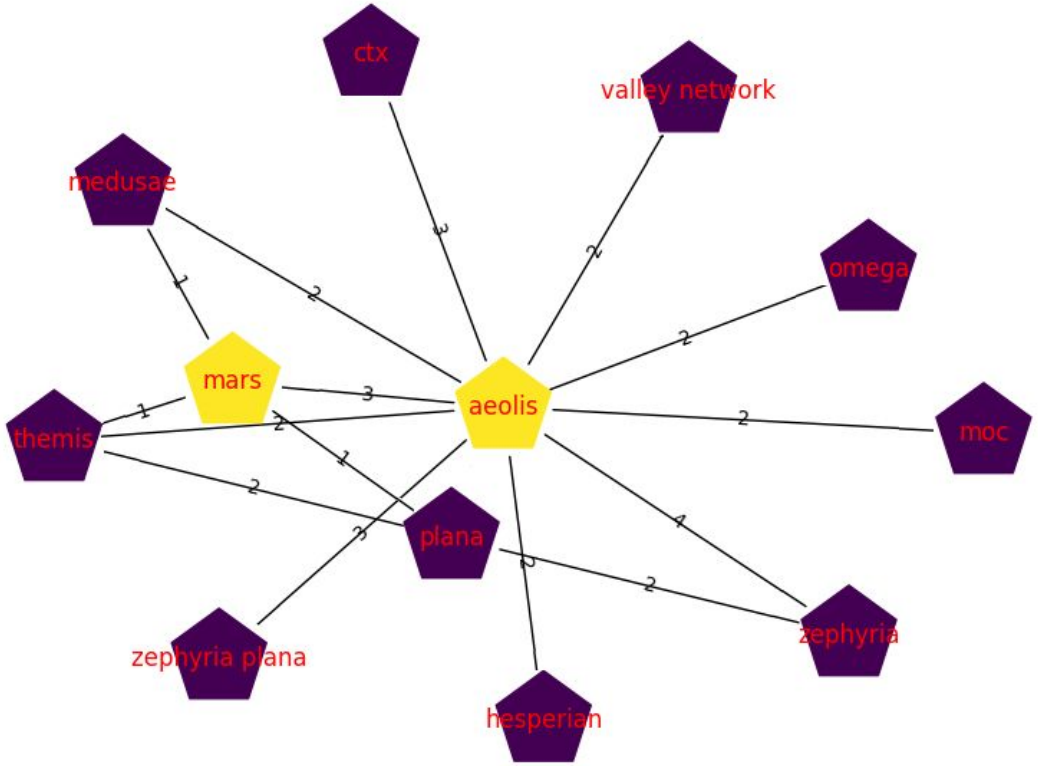
Non Planetary



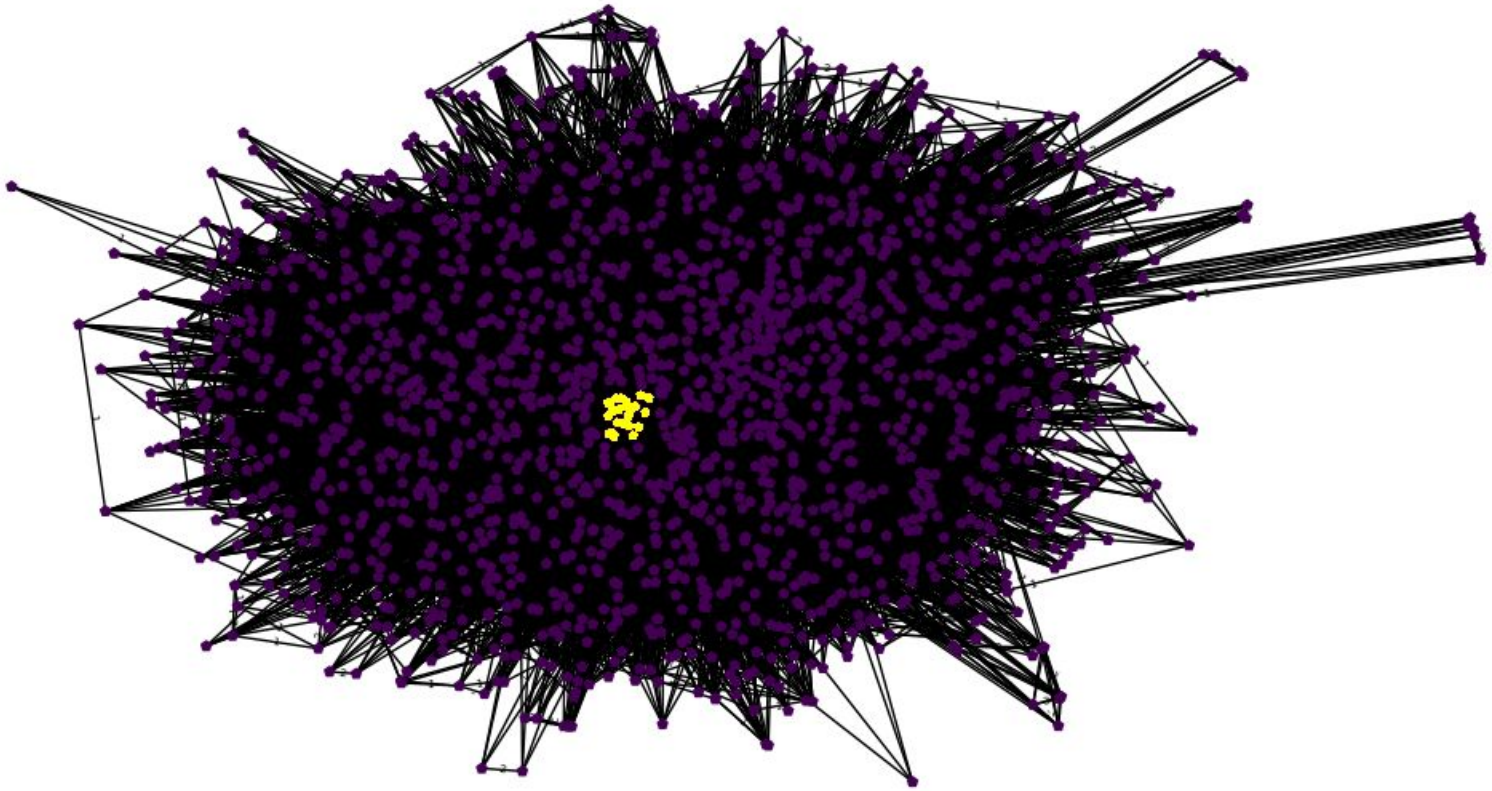
The Knowledge Base System – Training

- Query ADS
  - Extract planetary records referring to USGS terms (positive)
  - Extract records with USGS terms that do not refer to planetary terms (negative)
- For each set
  - For each record
    - Identify the term in full text
    - For each instance
      - Select the section around the term
      - Extract top keywords
  - Setup the relation
    - Count how many times a pair of keywords appear together
    - Count how many time a keyword appeared with the feature name
- Create a graph for each set

Partial Graph for Feature Name Aeolis Mars-Albedo Feature



Full Graph of the Planetary Names For Mars-Albedo Feature



The Knowledge Base System – Prediction

- Query ADS for USGS term, no other filter
- For each record
  - Identify if relevance
  - Get top keywords
    - Identify the USGS term in the full text
    - Select the section around the USGS term
    - Extract top keywords
  - Query knowledge Base Graph
    - For each top keyword for each positive and negative graph
      - If there is a path between the term and the keyword in graph, get the weight of the path
  - Sum the weight of path, if any, for all the keywords, from each graph
    - Vote for the side with highest score

Prediction Example When Relevant

Bibcode: 2020JGRE..12506330M
Collection: ['astronomy', 'physics']
Feature Term: Adams
Relevance: {'Moon':6, 'mountain':0, 'asteroid':0, 'Mars':0}
Excerpt with Keywords: GHz only lower than the visible and degraded mare units. However, more extensive regions in Region 2 postulate the higher d T B values as the cryptomare patches. Second, in Brisbane and Peirescius (67.8°E, 46.4°S) craters of Region 3, there occurs a well agreement between the cryptomare units and the higher d T B values. But, from Peirescius crater to **Adams** crater (68.4°E, 31.9°S), at least six patches indicate the similarly higher d T B values as in Brisbane and Peirescius craters, implying that they are likely the undiscovered cryptomare. Thus, the distribution of the possible cryptomare unit is likely more extensive than that obtained by the visible data. Moreover, the distribution of the patches in Region 3 is largely in
Prediction: Adams Moon Crater with confidence 0.9 (8 terms matched in KB positive, database is astronomy, journal is planetary, combined number of target and type appeared in the full text exceeds 10 times)

Bibcode: 2011E&PSL.312..140P
Collection: ['astronomy']
Feature Term: Adams
Relevance: {'Mars':15, 'Moon’:1, 'mountain':0, 'asteroid':0}
Excerpt with Keywords: The timing of eruptions producing the studied lava flows appears heterogeneous throughout the province (Fig.1), there being no clear indicators that volcanic episodes are related to specific sectors or edifices. However, it is observed that nine of the ten lava flows younger than 500 Ma are located at the distal reaches at the lower flanks of the Elysium rise and in **Adams** crater to the NE (Fig.1). This observation could be regarded as a possible time-space relation of volcanic activity. In addition, two caldera segments of Hecates Tholus show formation ages of 150 Ma and 440 Ma. We are aware that some of the mapped and dated lava flows of similar age but emplaced on opposing flanks could have been erupted during the same
Prediction: Adams Mars Crater with confidence 0.7 (6 terms matched in KB positive, database is astronomy, journal is planetary, combined number of target and type appeared in the full text exceeds 10 times)

Prediction Example When Not Relevant

Bibcode: 2001M&PS...36.1617C
Collection: ['astronomy']
Feature Term: Adams
Relevance: {'asteroid':34, 'Moon':8, 'mountain':0, 'Mars': 0}
Prediction: Not Relevant
Excerpt: FIG. 1 7. Intimate mixture models for the spectra of Eros bright and dark materials. The Eros spectra are shown in symbols and the models are shown in solid lines. the bright material spectrum into the spectrum of the dark material using model variations. Grain Size The first plausible explanation for the albedo contrasts observed in Psyche crater is grain size. As shown by Adams and Fiice (1976), Johnson and Fanale (1973), Clark et al. (1992) and Clark(1995), the reflectance of meteorite and mineral samples increases with decrease in the average grain size. We began with our nominal Eros bright material mixture model spectrum (grain size 63 ~tm) and produced three comparison spectra at 88, 107, and 126am grain size. Figure 18 shows the spectral consequences of this variation

Bibcode: 2022Icar..38415090W
Collection: ['astronomy']
Feature Term: Adams
Relevance: {'Mars':18, 'Moon':0, 'mountain':0, 'asteroid’:0}
Prediction: Not Relevant
Excerpt: band near 1.9m, and a steep spectral-slope from ~12; features that match well with the spectra of some deposits in Ladon (see Fig. 7 a,d). Interestingly, a recent study investigating the character of the phyllosilicates at Gale crater compared with disordered clays found that glauconitic clays are a reasonable analog for the phyllosilicates in the Murray formation at Gale crater (Losa-Adams et al., 2021). Furthermore, these smectite-glauconite mixtures are indicators of long-term quiescent conditions in lakes (Losa-Adams et al., 2021). Some of the variations observed in the spectral properties of Mg-rich smectites could be due to glauconitization of the smectite to form mixed smectite/glauconite clays or even glauconite in regions of Ladon that experienced long periods of extremely low sedimentation

Prediction Results - Albedo Features

Feature Term	Correct Martian Feature	Other Martian Feature	Non-Martian Feature
Arabia	253 (75%)	77 (23%)	9 (3%)
Syria	48 (58%)	32 (39%)	3 (4%)
Tempe	22 (79%)	5 (18%)	1 (4%)

Prediction Results - Crater Features

Feature Term	Correct Martian Feature	Other Martian Feature	Non-Martian Feature
Gale	200 (95%)	11* (5%)	0 (0%)
Herschel	40 (85%)	7 (14%)	0 (0%)
Cassini	10 (66%)	0 (0%)	5 (33%)
Houston	0 (0%)	0 (0%)	101 (100%) where 80% are identified with low confidence
Basin	0 (0%)	0 (0%)	28 (100%) where 74% are identified with low confidence

* All features within Gale Crater

Next Steps & Recap

- Next Steps
  - Refine type of planetary feature identification
  - Expand to other planetary objects
- Recap
  - Building a model to identify planetary feature names in the literature
  - Provides text enrichment to a new segment of our expanding community