THE MARS TARGET ENCYCLOPEDIA NOW INCLUDES MARS PATHFINDER AND MARS PHOENIX TARGETS. K. L. Wagstaff[1], R. Francis[1], M. Golombek[1], S. Lu[1], E. Riloff[2], L. Tamppari[1], and T. C. Stein[3], [1]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109, USA (kiri.wagstaff@jpl.nasa.gov), [2]University of Utah, Salt Lake City, UT, 84112, USA, [3]Washington University in St. Louis, St. Louis, MO, 63130, USA.

Introduction: The Mars Target Encyclopedia (MTE) is a growing collection of information about surface targets identified by landers and rovers on Mars [1]. The MTE links targets to publications that provide information (e.g., composition, provenance, interpretation) about each target. Mission science team members, planetary science researchers, educators, students, and the general public can find relevant information about targets of interest or compare targets found by different missions. We use a combination of automated text analysis and manual review to populate the MTE with target information obtained from scientific publications. The MTE began with targets observed by the ChemCam instrument on the Mars Science Laboratory (MSL) rover. The MSL Analyst's Notebook provides access to MSL target information derived from ~6000 LPSC abstracts from 2014 to 2016 [2].

Our current goal is to extend the MTE to include targets observed in historical Mars surface missions, including Mars Pathfinder, Mars Phoenix, and the two Mars Exploration Rovers. These extensions to the MTE will enable cross-mission analyses to identify patterns in targets as well as unusual exceptions. In this abstract, we report on additions to the MTE that include targets for Pathfinder and Phoenix. We are in the process of integrating that information into the Analyst's Notebook and for delivery to the Planetary Data System to make the content available to all.

Mars Pathfinder Targets: We identified 591 abstracts from the proceedings of the Lunar and Planetary Science Conference (LPSC) between 1998 and 2020 that reported results related to the Mars Pathfinder (MPF) mission. We found that 65 documents (11%) mentioned at least one MPF target, with 467 total target name occurrences ("mentions"). The target mentions comprise 63 distinct targets; many of them occur in multiple documents. Figure 1 shows the distribution of distinct mentions of targets within LPSC abstracts over the last 23 years of the conference. As we expect, the number of mentions is highest in the year that follows each mission's landing and decreases over time. Some MPF targets are mentioned as late as 2013. The most commonly mentioned MPF targets are *Yogi* and *Barnacle Bill*. Note that there is no convention for formatting target names. In this abstract, we use italics to visually highlight target names, but some authors use quotation marks or (most commonly) no distinguished formatting for the target names.

The MTE links targets to their mentions within documents, so a search across the collection brings all information about a given target into one place. For example, target *Yogi* appears in 23 documents, with statements that include "At least one rock, Yogi, lies on a soil pedestal, which could have formed from aeolion [sic] scour" [3] and "The other three (Half Dome, Wedge, and Yogi) are higher in sulfur" [4]. By searching the MTE, users can combine disparate information from multiple sources into a more complete view of what is known about each target. The MTE also provides the URL of the source abstract so users can immediately access the full text (PDF) of the publication.

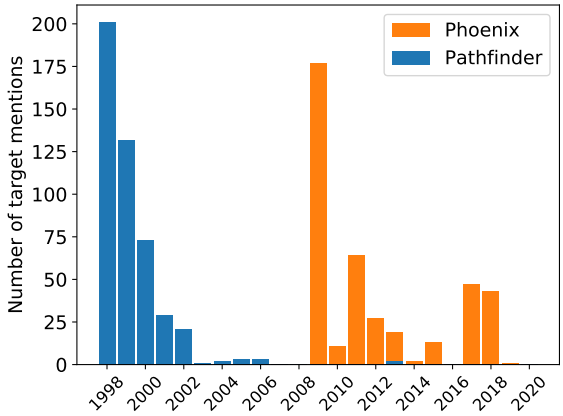

Figure 1: Number of target name occurrences (mentions) within LPSC abstracts for the Mars Pathfinder and Phoenix missions over time.

Mars Phoenix Targets: The first LPSC publications that mention Phoenix (PHX) science results occur in 2009 (see Figure 1). We found 391 relevant LPSC abstracts from the years 2009–2020. Of these documents, 33 (8%) mentioned at least one PHX target, with a total of 52 distinct targets and 406 total mentions. The most commonly mentioned PHX targets are *Wicked Witch, Rosy Red,* and *Snow White*. Browsing the MTE excerpts from the eight documents in which *Wicked Witch* is mentioned yields statements that include "A sample of sublimation lag was delivered to TEGA in the 'Wicked Witch' sample" [5] and "The Mars Phoenix Lander's TEGA instrument detected a calcium carbonate phase decomposing at high temperatures (~700°C) from the Wicked Witch soil sample" [6].

Methods: Collecting information about Mars surface targets from scientific publications for the MTE requires several steps. The first challenge is to compose a list of valid target names for each mission. We surveyed

Table 1: MTE Mars surface targets identified in this study.

Mission	Initial targets	Basilisk additions	NER additions	Manual additions	Final targets
MPF	219	+3	+1	+18	241
PHX	50	+5	+0	+22	77

full-length peer-reviewed mission publications to construct an initial list for Pathfinder [7][8][9][10][11][12] and Phoenix [13]. The number of targets in the initial list is shown in the first column of Table 1.

Since there might be additional targets that were not mentioned in these sources, we used the Basilisk [14] vocabulary generation system to analyze the relevant LPSC document collection for each mission. Basilisk uses a "seed" list of initial terms and identifies frequently occurring text patterns in which those terms appear. For Mars surface targets, example patterns are [composition of T], [the rock "T"], and [T was imaged by]. When these same patterns occur with a previously unknown name in T's position, they are proposed for human review as a possible new target name. Manual review of the Basilisk candidates extended the list of target names, adding three targets for MPF (*Half Anvil, Ovoid*, and *Scooby*) and five for PHX (*Dodo-Goldilocks Trench, Mancha, Pet Donkey, Pet Donkey Trench*, and *Snow White Trench*).

We also trained an automated Named Entity Recognition (NER) model [15] to find additional targets. Because Mars target names often re-use common nouns (*Lamb, Runaway, Wedge*) or names with other meanings based on names of people or Earth locations (*Dagwood, Geordi, Half Dome*), context matters and a simple keyword search is not reliable. The NER model incorporates syntactic context when identifying targets within text. Unlike Basilisk, which requires only a list of seed terms, the NER model must be trained with annotated documents that show where the terms appear. This approach uses features such as word "shape" (pattern of capitalization) and character patterns before and after a term to decide whether the term is a target name. We trained an NER model using 117 hand-annotated LPSC documents as examples [16]. This step yielded one additional target for MPF (Soufflé) and no additional targets for PHX.

Finally, to ensure high reliability, we performed a manual review of all documents with at least one proposed target to check for additional target names. We added 18 MPF and 22 PHX targets by hand. The automated steps had already annotated 93% (MPF) and 71% (PHX) of the targets in advance. Manual additions included abbreviations invented within a document, such as *B. Bill* for *Barnacle Bill* or *WW* for *Wicked Witch*, alternate spellings or typos such as *Fat Top* for *Flat Top* or *Mamma Bear for Mama Bear*, alternate hyphenations such as *Humpty-Dumpty* for *Humpty Dumpty*, etc. There were also some genuinely new target names found, such as *Abu, Gerbil*, and *Peesh* for MPF and *Runaway* for PHX. We speculate that if we had a larger set of hand-annotated documents for training, the NER model would likely have found more of these targets automatically.

Conclusions and Next Steps: The Mars Target Encyclopedia (MTE) collects knowledge from diverse publications about Mars surface targets in a central, searchable database. The MTE now spans targets from the Mars Pathfinder, Mars Phoenix, and Mars Science Laboratory missions over more than two decades. The MTE is more than just a literature search tool. By connecting targets to publications, the MTE makes it easy to answer questions like "What are the existing published results about *Yogi?*" which can reveal consensus as well as disagreements and also highlight gaps in our knowledge, inspiring new research investigations.

Our next steps will include expanding the MTE to include targets from the Mars Exploration Rover missions and adding the ability to analyze peer-reviewed journal publications. In addition, we will include target properties (shape, appearance, provenance, geochemistry) that are stated within the document collection to further enrich each target's description.

Acknowledgments: We thank the NASA Planetary Data Archiving, Restoration, and Tools (PDART) program for funding this project. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References: [1] Wagstaff, K.L. *et al.* (2018) *Innovative Applications of AI*. [2] Stein, T.C. *et al.* (2019) *50th LPSC #1820*. [3] Bridges, N.T. *et al.* (1999) *30th LPSC #1907*. [4] Crisp, J.A. (1998) *29th LPSC #1962*. [5] Smith, P.H. & Phoenix Science Team (2009) *40th LPSC #1329*. [6] Cannon, K.M. *et al.* (2012) *43rd LPSC #2008*. [7] Smith, P.H. *et al.* (1997) *Science 273*. [8] Bridges, N.T. *et al.* (1999) *JGR 104*. [9] Golombek, M.P. *et al.* (1999) *JGR 104*. [10] McSween Jr., H.Y. *et al.* (1999) *JGR 104*. [11] Bell, J.F. III *et al.* (2000) *JGR 105*. [12] Stooke, P. (2012) *International Atlas of Mars Exploration*. [13] Arvidson, R.E. *et al.* (2009) *JGR 114*. [14] Thelen, M. & Riloff, E. (2002) *Conf. on Empirical Methods in Natural Language Processing*. [15] Finkel, J.R. *et al.* (2005) *43rd Meeting of the Assoc. for Computational Linguistics*. [16] Francis, R. & Wagstaff, K. (2017) *Zenodo data set, http://doi.org/10.5281/zenodo.1048419*.