MoonDB is a data system that synthesizes analytical data acquired on lunar samples and makes them easily accessible online as a single data product with human and machine-readable interfaces that integrate the data into digital research data infrastructure. MoonDB preserves and restores data that have been scattered across the scientific literature, in online PDF documents, and in private files, having made discovery and reuse of these data difficult at best. MoonDB also serves as a home for new data generated on lunar samples, integrating new measurements seamlessly with the historical data. MoonDB helps researchers comply with US federal mandates for open access to research results1.

MoonDB started as a collaboration between the Geoinformatics Research Group (GRG) at the Lamont-Doherty Earth Observatory of Columbia University, the Astromaterials Acquisition and Curation Office (AACO) of NASA's Johnson Space Center (hereafter JSC Curation), and senior members of the lunar science community. Over the past two years, the MoonDB project focused on (1) developing the MoonDB data system consisting of a relational database to store data and metadata, and a web application that provides users with web-based software tools to browse, search, and retrieve data; (2) summarizing references for lunar geochemistry and petrology from all relevant published sources and the Apollo Sample Compendium into the MoonDB Reference Catalog; (3) compiling geochemical and petrological data and metadata from these references and from datasets contributed by researchers into data templates and ingest these into the MoonDB database; and (4) supporting lunar scientists with preparing their unpublished geochemical data for ingestion into MoonDB. To encourage contribution of new data, the MoonDB web site features a service for investigators to contribute their data. Submitted data are deposited and published in the EarthChem Library, the geochemistry data repository recommended by publishers (e.g. Nature Scientific Data), where data are assigned a persistent unique identifier (DOI) so that they can be formally cited to give credit to authors, and linked to publications.

Figure 1 provides an overview of MoonDB's architecture and software stack. The system consists of the storage database (PostgreSQL), reporting database (ElasticSearch), the MoonDB search application (Python & JavaScript), and the MoonDB API, which serves as the bridge between the ElasticSearch index and the search application and can also be used by external systems to access MoonDB data. Data are loaded into the SQL database via the MoonDBLoader web application.

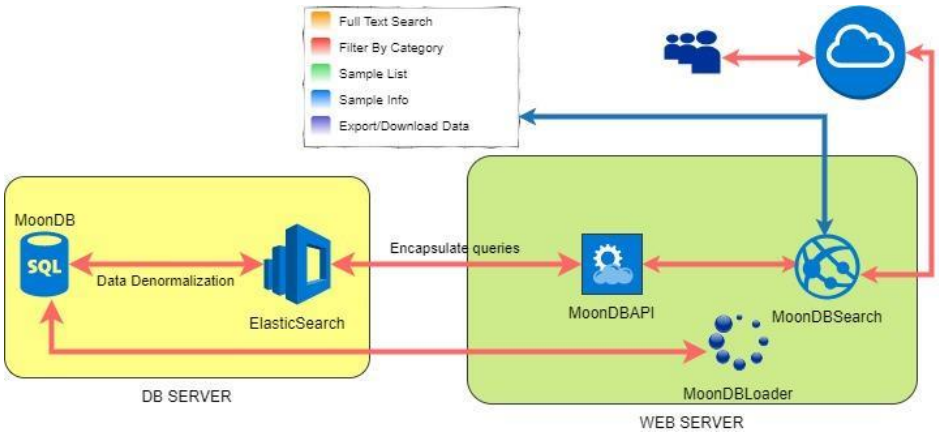

Figure 1 Architecture of the MoonDB Data System

The MoonDB database uses a customized version of the Observation Data Model (ODM2) relational database schema to store data and metadata of included datasets. ODM2 is a communi-

ty information model designed to support a wide variety of feature-based Earth observations derived from sensors and samples, and to improve the capture, sharing, and preservation of these data (Horsburgh et al. 2016; Hsu et al. 2017). ODM2 implements concepts of the Observations and Measurements (O&M) standard (Cox 2007). O&M is one of the core standards in the OGC (Open Geospatial Consortium) Sensor Web Enablement suite. The use of ODM2 gives MoonDB a flexible and scalable structure to store data and metadata for a diverse range of samples and analytical data and makes it compatible with the new EarthChem synthesis database, which is also built on ODM2.

The MoonDB search application provides the tools for users to explore the content of the MoonDB database, select the samples and data they are interested in (e.g., specific chemical parameters, analyzed material such as whole rock or minerals), view and browse them, and download them in a useful format. Users first choose samples based on attributes such as sample type (lithology), geographic feature, data availability, and reference. They can then select chemical parameters to create a customized dataset containing a set of analytical data for a specific sample or group of samples, retrieved from separately published datasets. The custom dataset can be viewed online and downloaded in different formats. The MoonDB search interface features free-text and structured, faceted searches to find, filter, and explore data stored in MoonDB, and a download option for .csv file.

The MoonDB API (Application Programming Interface) is the bridge between the database and web applications. It defines a set of Hypertext Transfer Protocol (HTTP) request messages and the structure of response messages. The API feeds the HTML5 web component of the MoonDB search application. The API supports any third party uses or applications to retrieve data from MoonDB, exposing queries that return data for samples, features, datasets, and authors. A first set of HTML5 web components that interested data facilities can embed in their web pages have been developed. These web components allow access to the sample metadata and analytical results in MoonDB from data systems that scientists may use to look for lunar data, e.g. the Lunar Sample Database, MoonTrek, the Analyst's Notebook, and other planetary data systems.

During its next development phase, the MoonDB project will add geochronological data and lunar meteorite data to the synthesis and advance integration with the Planetary Data System by developing PDS4-compliant versions of the MoonDB data as a lunar sample bundle with table, context image and document collections with XML labels and archive these with the PDS Cartography and Imaging Sciences Node (IMG).

Cox, S.J.D., Ed., (2007), Observations and Measurements – Part 1 – Observation Schema: Open Geospatial Consortium. Wayland, MA., OGC doc. no. 07-002r1, 73 p.

Horsburgh, J. S., Aufdenkampe, A. K., Mayorga, E., Lehnert, K. A., Hsu, L., Song, L., Spackman Jones, A., Damiano, S. G., Tarboton, D. G., Valentine, D., Zaslavsky, I., Whitenack, T., (2016), Observations Data Model 2: A community information model for spatially discrete Earth observations. Environmental Modelling & Software, 79, 55-74, doi: 10.1016/j.envsoft.2016.01.010.

Hsu, L., Mayorga, E., Horsburgh, J. S., Carter, M. R., Lehnert, K. A., Brantley, S. L., (2017), Enhancing Interoperability and Capabilities of Earth Science Data using the Observations Data Model 2 (ODM2). Data Science Journal, 16(4), 1-16, doi:10.5334/dsj-2017-004.