NASA Data Recovery / Operation ZETA

Zebulon Mcknight
UTK EECS
Knoxville, TN, USA
zmcknigh@vols.utk.edu

Eric Yang
UTK EECS
Knoxville, TN, USA
eyang7@vols.utk.edu

Trevor Eisenbacher
UTK EECS
Knoxville, TN, USA
teisenba@vols.utk.edu

Ahmed Ghazi
UTK EECS
Knoxville, TN, USA
aghazi2@vols.utk.edu

***Abstract*—NASA's extensive data repository contains thousands of datasets spanning over six decades of scientific observations, but many remain inaccessible due to obsolete data formatting practices. This project aims to cleanse and parse the datasets into a usable format. We intend to use a JavaScript framework, namely React, to visualize this data. This project will demonstrate how NASA's invaluable data can still be used to enable new research opportunities across disciplines while establishing a replicable framework for modernizing other legacy scientific data repositories. We plan to focus on finding habitable planets among the data and highlight how NASA's extensive dataset can be curated for broad public use.**

Keywords—Data modernization, educational, general public, habitable, habitability, human, humanity, legacy format conversion, scientific data visualization, NASA datasets.

I. Objective

The objective of this project is to make NASA's vast array of public datasets more accessible and usable by parsing, labeling, cleaning, and presenting the data with modern data processing and visualization techniques. Our project will transform complex raw data into easily digestible and interactive formats promoting broader usage in education, research, and public engagement. We will focus specifically on identifying habitable planets among the chosen datasets by analyzing conditions such as gravity, air quality, and other necessary elements for human survival. We aim to present this data in an accessible and digestible format for the public, and we hope to inspire them to dig deeper into the stars of NASA's untapped databases.

II. Motivation

NASA has been recording and publishing data across various domains since the 1960s, resulting in a vast database of scientific discoveries spanning decades of research efforts. While these datasets are technically public, their age affects their usability. Many datasets, including data collected from the Apollo and Voyager missions, are stored in outdated and proprietary .TAB and .LBL formats with little use to those outside of niche research communities.

Our project is motivated to bridge those gaps between the data and the general public. By converting legacy formats into modern, standardized, and easy-to-use structures, and then layering interactive visualization tools, we aim to increase the practical reach and educational value of NASA's open data.

For the scope of this project, we want to highlight the utility of NASA's datasets by identifying habitable planets. We believe this information to be valuable for humanity, not only for the sciences but also for the arts & humanities and the inspired stories that may result from our findings. We believe the extensive databases on several planets hold information that can be broken down and analyzed, but there currently doesn't exist a widespread public effort to dive into that information nor does that information exist in a form easily accessible to the public. By synthesizing the data into an accessible form, we strive to pioneer a renewed focus on the potential of space for humanity and skyrocket humanity's general knowledge of the cosmos.

III. Data

We will source data from across NASA's databases, focusing on planetary data such as average temperature, atmospheric composition, magnetic fields, presence of water in any state, gravity, air quality, and other quantifiable data that could be utilized to assess habitability. Our initial datasets will be sourced from NASA's open data repositories, specifically from their collections on the Apollo and Voyager missions. We have already investigated datasets containing information on solar wind recorded by Apollo 12, and data on Jupiter's atmosphere collected by Voyager 1. Such data would be valuable variables necessary to determine whether a planet would be habitable for humanity.

We have established the following guidelines for selecting data:

1. The data represents significant scientific milestones with high educational value, such as planetary statistics for planets that are already recognized as

habitable or near-habitable. We want to ensure the information becomes easily quantifiable to the general public.
2. The data can reasonably be processed in a semester-long time frame. We want to process data for multiple planets to uncover their habitability, and we cannot afford to be stuck on data that cannot be reasonably processed within the project timeline.
3. The data has formatting and / or accessibility issues. One of our main goals is to demonstrate how NASA's large datasets can become more accessible for the general public, so we want to ensure we can process difficult or inaccessible data for public consumption.

For this project our data will need to pass through a multi stage processing pipeline, including parsing, storing, and visualization.

We anticipate encountering proprietary formats (e.g., legacy binary structures), incomplete metadata, and inconsistent labeling schemes. These issues will necessitate writing custom parsers, data normalization scripts, and robust cleaning pipelines.

Our parsed data will be stored in modern, easily readable formats, such as CSVs or JSON. We intend to produce clean, modern storage formats containing the processed NASA data for future use by ourselves and others through methods such as a specialized front-end display and simpler, text-based formats/files. By modernizing these file formats and maintaining metadata, we can provide these parsed files as a resource for other data science projects.

Finally, we aim to present a subset of the available data using modern visualization techniques in Python and React, providing an educational resource sourced directly from NASA.

IV. Team Responsibilities

Zebulon Mcknight will take the lead on data parsing and be responsible for providing other members of the team with workable data in convenient formats, such as Pandas Dataframes. In addition he will take the lead on data requisition, and be in charge of selecting viable and relevant datasets to parse.

Eric Yang will assist Zebulon with data collection and research the best methods to quantify what exactly "habitable" means for the scope of this project, and he will also utilize React to create an intuitive front-end design that will make the processed data and results both easy-to-access and interesting/aesthetically-pleasing to view.

Trevor Eisenbacher will implement Python-based visualizations of the collected data using libraries such as Plotly, ensuring that key habitability indicators are represented in clear, informative graphical formats. He will structure these outputs for seamless integration into the React frontend, preserving full interactivity and functionality. He will lead the documentation of workflows and processes.

Ahmed Ghazi will lead the development of the web application. He will work with Eric to develop an application using a JavaScript framework as well as designing the application that will be developed.

V. Timeline

October 1st: Dataset(s) chosen
October 15th: Begin development of visualization tool
October 31st: Cleanse data that will be visualized
November 15th: Finish visualization tool
November 30th: Finalize tool, finished product deadline

VI. Expected Outcome

We expect our project to result in a fully functional frontend interface for our visualization product. Using React, we will create a framework to display interactive visualizations generated in Python and exported in a format suitable for integration. Additionally, we expect to parse the processed data into convenient formats for the general public to access, view, and utilize for any future projects.

We expect to provide an accessible and digestible educational resource to students, educators, and researchers through our visualization of the data, and a technical resource to other developers interested in investigating and parsing the large, untapped potential of NASA datasets that hold the future of humanity within their 1s and 0s.