AirFlow Pipeline

Hunter Woodruff, Quinton Jasper, Chris McCorkle,
Isaiah Raspet, Richard McCormick

School of Informatics, Computing, and Cyber Systems

NAU Northern Arizona University

USGS science for a changing world

ARES

Client: Trent Hare, Cartographer, U.S. Geological Survey

Team Mentor: Vahid Nikoonejad Fard, Researcher, SICCS

Introduction

Planetary Data in Space Exploration

Every year, thousands of images from planets and moons across the solar system are collected by researchers. These images facilitate a multitude of scientific endeavors, from the Mars 2020 Mission to the upcoming Artemis Missions.

Big Data Brings Big Challenges

Each image captured from an extraterrestrial body requires processing and analysis. This processing is done with an application named ISIS (Integrated Software for Imagers and Spectrometers), which uses nearly 300 different filters.

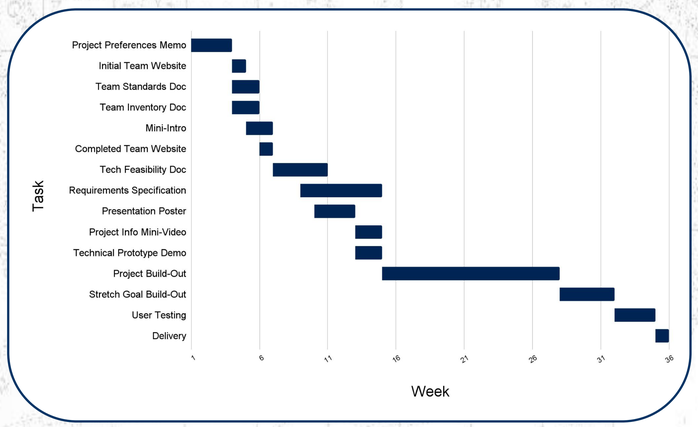
Planetary Picture Processing Pipeline

Our project will combine the hundreds of different filters used in the ISIS application into a simple, streamlined pipeline. Usable by anyone, even those with no coding experience, this solution will greatly increase image processing efficiency.

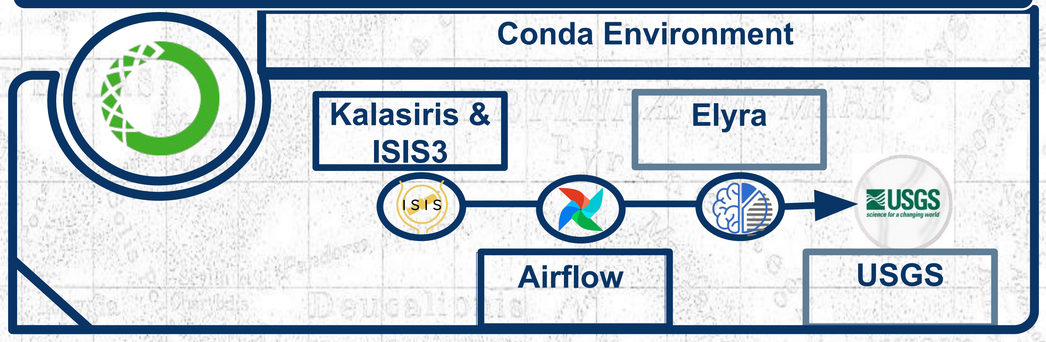
New Tools For New Worlds

With this new streamlined image processing pipeline, researchers around the world will be able to analyze planetary imagery faster and easier than ever before, opening untold possibilities for science!

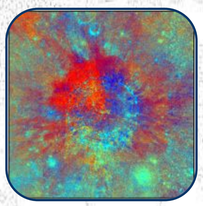
Project Timeline



Proposed Solution Overview



The path of Mark Watney from *The Martian* overlaid onto imagery processed by ISIS3



Multispectral images from Copernicus Crater

25143, Itokawa

Numisia Crater, Vesta,

Images processed with ISIS3 are captured from moons, planets, and asteroids.

Solution Explanation & Feasibility

Technical Challenges

The technical challenges we are facing are broken into three categories. The first, **pipeline software,** encapsulates the software that will hold and execute the pipeline. The second, **Graphical User Interface**, will increase usability and allow for drag-and-drop building of pipelines. Finally, **containerization** will allow the final project to be packaged & distributed.

Technical Analysis

Pipeline Software: Apache Airflow is the software we have found to fit best with our and our sponsor's needs. It has a built in interface that will show the pipeline's status as it is being executed.

Graphical User Interface: In addition to the user interface provided by Airflow, Elyra will allow researchers to build pipelines visually, with minimal need for coding, increasing efficiency and throughput

Containerization: The biggest hurdle we have run in to. We have explored many options and have finally decided on a Conda environment as the best way to contain the project.

Technologies

Python
A lightweight and portable solution for scripting and data science.

ISIS3
Powerful suite of image processing tools for planetary data science.

AirFlow
Created by Apache, this tool allows scientists to quickly and easily build data processing pipelines.

Conda
Open source distribution system for Python projects.