

Imaging Spectroscopy Processing Environment on the Cloud (ImgSPEC): A Prototype of an Open Science, Scalable Processing Environment

E. Natasha Stavros¹, Philip Townsend², Hook Hua³, Sujen Shah³, Namrata Malarout⁴,
Winston Olson-Duvall³, Marjorie Lucas⁵, William Phyto³, David Thompson³, Justin Merz⁶,
Amruta Vidwans⁷, Adam Chlus⁸, Alexander Torres⁴, Dan Yu⁴, Michelle Gierach³, and
Kevin Marlis⁴

¹University of Colorado at Boulder

²University of Wisconsin

³Jet Propulsion Laboratory, California Institute of Technology

⁴NASA Jet Propulsion Laboratory

⁵Jet Propulsion Laboratory

⁶University of California Davis

⁷University of Wisconsin Madison

⁸University of Wisconsin-Madison

November 24, 2022

Abstract

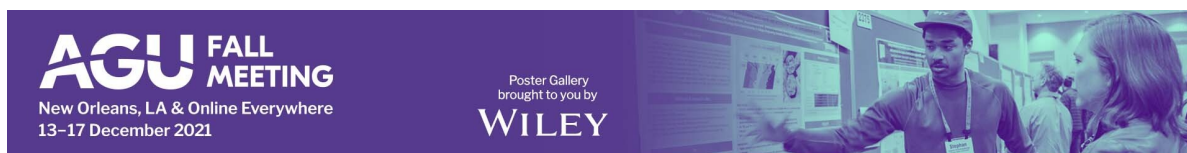
Imaging spectroscopy data is becoming more readily available from different satellite and airborne platforms. As this data becomes more prolific, there is a need for shared data tools and code for wrangling, cleaning, and analyzing it. The geospatial Imaging Spectroscopy Processing Environment on the Cloud (ImgSPEC) pioneers an on-demand science data processing platform with scalable back-end compute. It considers user experience and facilitates open science. ImgSPEC enables users to create data products in areas of interest using default workflows from registered algorithms, while also enabling users to customize scripts and workflows. ImgSPEC seamlessly interfaces with NASA Earthdata Search and tracks appropriate metadata for reproducibility when generating data products to share with others. Users can work in their preferred workspace (e.g., Rstudio, Jupyterlab, or command line) thereby facilitating use of open science software packages and collaborative coding through Git. ImgSPEC leverages existing NASA-funded information technologies such as the hybrid on-premise/cloud science data system (HySDS) and the Multi-mission Algorithm and Analysis Platform (MAAP). It also creates seamless interfaces with NASA-funded ECOSIS – a crowd-sourced spectral database, and ECOSML – a crowd-sourced model database. We demonstrate ImgSPEC on the Terrestrial Ecosystem use case processing through to foliar traits and fractional cover, thus aligning with driving thrusts for the NASA Surface Biology and Geology (SBG) Science and Applications Communities. As this technology is more widely adopted the interface with Amazon Web Services and NASA Earthdata search will enable broader use of more data (publicly available or loaded by the user) across more domains.

Imaging Spectroscopy Processing Environment on the Cloud (ImgSPEC): A Prototype of an Open Science, Scalable Processing Environment

E. Natasha Stavros (CIRES, CU Boulder)

UW Madison: Philip A Townsend, Amruta Vidwans, and Adam Chlus; JPL/Caltech: Hook Hua, Sujen Shah, Namrata Malarout, Winston Olson-Duvall, Marjorie Lucas, William Phyo, David R Thompson, Alexander Torres, Dan Yu, Michelle M Gierach, and Kevin M Marlis; University of California Davis: Justin Merz

PRESENTED AT:



IMGSPEC VALUE PROPOSITION

ImgSPEC prototypes and refines open science use cases for using open diverse open data across missions by stress testing bounding conditions on contributions by users outside of NASA missions for on-demand product generation.

INSPIRATION

Imaging Spectroscopy is high dimensionality spectral data and is rapidly emerging as a go-to data source for many applications such as characterizing biodiversity and quantifying plant traits and ecosystem function.

PROBLEM DEFINITION

A broader community wants to make use of these data, but may not have the expertise or interest in processing the data themselves.

Likewise, agencies or other product developers may not have the capacity to generate all possible products and distribute them.

Finally, to make use of these products, we need more than Imaging Spectroscopy to accelerate scientific understand of the Earth system.

Datacubes, stacks of remote sensing data pertinent to a single domain of study or application, over ecological national observatories (NEON, LTER, and LTAR) would require >590 TB of data storage!

We cannot do what we have always done!

We cannot rely on a single project to create information products that can be downloaded and processed locally.

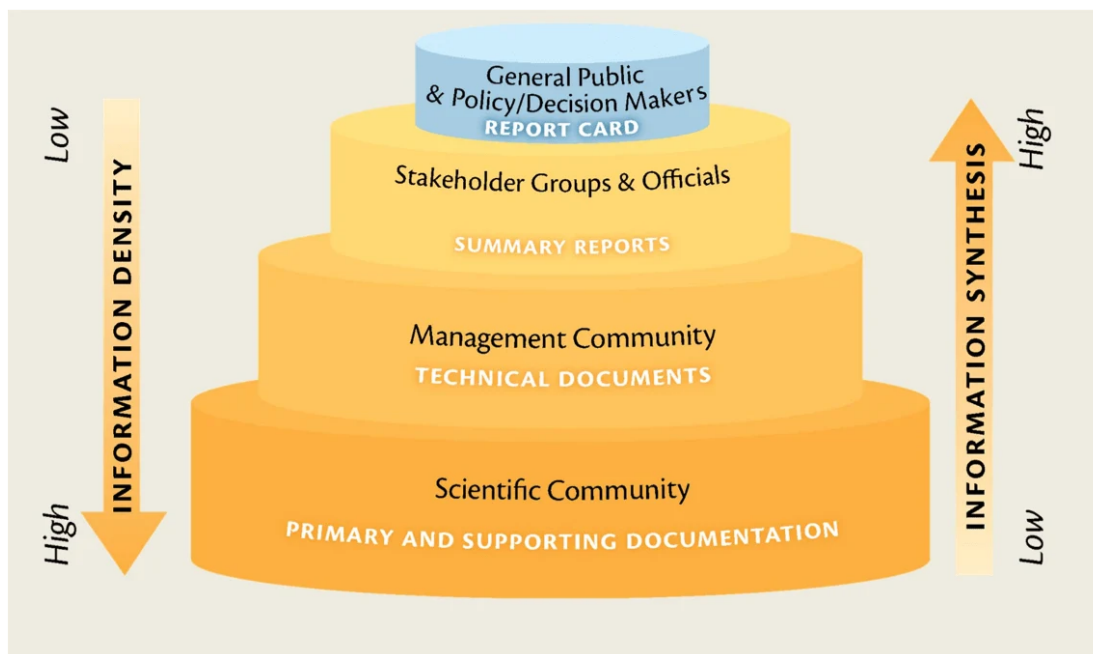
VISION AND OBJECTIVE

ImgSPEC **vision** is to provide an open source science platform that supports analysis by *diverse* perspectives through *inclusive* collaboration, that increases *access* to data, data tools, and compute resources. This Enables us to take advantage of the information age to provide decision support* with scientific advancement using *analytics by many* to create relevant information from the wealth of big data

ImgSPEC **objective** is to prototype and demonstrate an end-to-end, on-demand, processing platform on the cloud for imaging spectroscopy Level 1 calibrated radiance data through Level 3+ information products for the terrestrial biodiversity use case.

DESIGN THINKING FOR DIFFERENT USERS AND CORE FUNCTIONALITY

There are many potential users of ImgSPEC:



Manager using data visualizations to inform decisions

Field scientist interested in point locations

Educator/Student working on limited scope class project

Researcher conducting statistical and modeling experiments

Remote Sensing algorithm developer

Software developer generating tools and services

Technician providing data processing support

This results in providing core functions:

