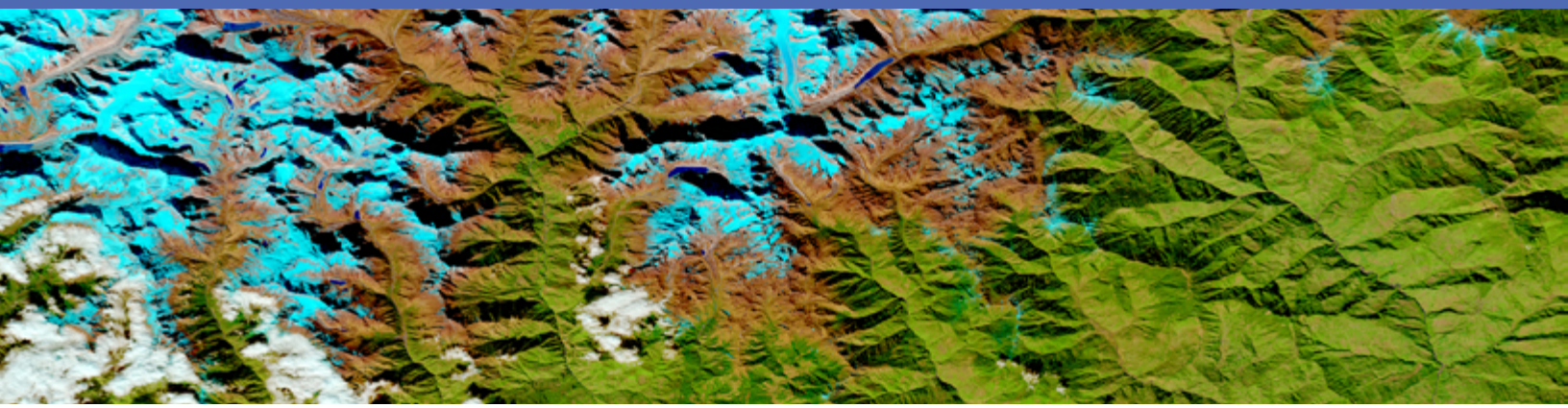
FALL 2023 | ISSUE 4

IMPACTful NEWS

Updates from the Interagency Implementation and Advanced Concepts Team



HLS image of glacier fields

IMPACT and IBM Develop AI Geospatial Foundation Model

In January 2023, a formal partnership between NASA and IBM Research was established via a Space Act Agreement. Machine learning (ML) developers within IMPACT were recruited to probe the possibilities of building artificial intelligence (AI) foundation models (FMs) with the goal of accelerating innovation in Earth science research. FMs represent a new paradigm in machine learning techniques. These models are initially pretrained on comprehensive, unlabeled datasets, then fine-tuned with domain-specific content. The adaptability of FMs is advantageous for researchers because a single model can be tailored for a variety of downstream applications. IMPACT and IBM developers also collaborated with Clark University's Center for Geospatial Analytics, the European Space Agency (ESA), USGS, and Oak Ridge National Laboratory.

At the IBM Think conference in May 2023, IMPACT director Rahul Ramachandran and IBM Research announced the

ALSO IN THIS ISSUE

Science Discovery Engine Developments

Facilitating Satellite Earth Observation Needs

VEDA Powers Open Science

Release of Additional Spire Soil Moisture Data

HLS Opens Source Code and Completes Data Archive

IMPACT
Interagency Implementation and Advanced Concepts Team

Geospatial FM (cont)

production of the first geospatial foundation model specifically trained with Earth observation satellite imagery. The model was trained on NASA's Harmonized Landsat Sentinel-2 (HLS) dataset, which provides surface reflectance data from NASA/USGS and European Union satellites. This model, named Prithvi, empowers numerous potential Earth science research pursuits, such as monitoring floods caused by natural disasters and mapping burn scars from forest fires.

IMPACT developers are keen to show Earth science researchers how to leverage FMs in their work. To that end, Manil Maskey, Brian Freitag, Iksha Gurung, and Muthukumaran Ramasubramanian representing **IMPACT's machine learning team organized a workshop** in collaboration with the Institute of Electrical and Electronics Engineers Geoscience and Remote Sensing Society (IEEE GRSS). Geoscientists who participated in the workshop learned how the FM was developed and successfully fine-tuned the FM using IBM's WatsonX.ai, an enterprise-ready AI and data platform.

Harnessing technology advances in data and information systems to expand community impact.

Science Discovery Engine Developments

In late 2022, the Science Discovery Engine (SDE) became the first capability of its kind to establish an infrastructure for searching vast quantities of NASA Science Mission Directorate (SMD) data and information. Users can perform free-text searches and apply text-based facets to refine results based on SMD topical area, data repository, source name, and document type. The **beta version of the SDE went live on the SMD website** in December 2022, and Kaylin Bugbee, SDE team lead, unveiled SDE's capabilities to attendees at the American Geophysical Union (AGU) 2022 Fall Meeting.

The **SDE leverages advanced insight engine technology** to provide enhanced search capabilities. Insight engines apply relevancy methods to describe, discover, organize, and analyze data. Since launching SDE, IMPACT developers and SDE collaborators have continued to add customizations and workflow enhancements that have improved user search results. Recently they built a web application called the SDE Indexing Helper to simplify and optimize the data curation process. The software allows data curators to easily review and select particular URLs to index from diverse science sources. The SDE team has also continued to refine both the back- and front-end features of the SDE. In the coming months, curators will continue to add substantial new content to the SDE, and the team will also integrate additional natural language processing techniques to fine-tune the scope and specificity of user search results.

ADMG Achieves Full Release of CASEI for Improved Suborbital Discovery

On July 11 of this year, the Airborne Data Management Group **announced the full release of the Catalog of Archived Suborbital Earth Science Investigations** (**CASEI**). CASEI includes a web-based, openly accessible search and discovery interface for learning about NASA suborbital research. After years of development, the catalog now contains extensive, searchable metadata for more than 100 campaigns (65% of all currently known NASA campaigns), more than 500 instruments, and more than 120 airborne and surface platforms. Metadata curation efforts will continue next year, and several major user interface enhancements will be implemented, including visual maps of platform tracks and locations for all campaigns. To promote enhanced use and understanding of CASEI, **ADMG team members presented a NASA webinar** on August 30th to over 75 attendees interested in learning more about CASEI's capabilities. ADMG team members have been meeting with DAAC and NASA Earth Science and Data Information System (ESDIS) representatives to define the transition process that will shift operations of CASEI from IMPACT to ESDIS in the future. Continued CASEI maintenance will then be in the care of the DAACs.

SNWG MO Facilitates Meeting Federal Civilian Agency Needs

The Satellite Needs Working Group Management Office (SNWG MO) was established within IMPACT three years ago to oversee NASA's response to federal agency needs for Earth observations. Every two years, the Satellite Needs Working Group (SNWG) administers a survey to assess the Earth observation measurement and data product needs of 25-30 federal civilian agencies. Through a careful review process by NASA and partner agencies, responses to these surveys ultimately become the inspiration for new NASA projects and data products that provide novel Earth science data. Some prime examples of solutions previously identified and now managed by the SNWG MO include the Harmonized Landsat and Sentinel-2 (HLS) project, the Airborne Data Management Group inventory (CASEI), and Observational Products for End-Users from Remote Sensing Analysis (OPERA). Large-scale projects such as these are made possible due to SNWG MO's leadership in managing swift and comprehensive efforts to identify critical needs and facilitate the development of each solution.

SNWG MO (cont)

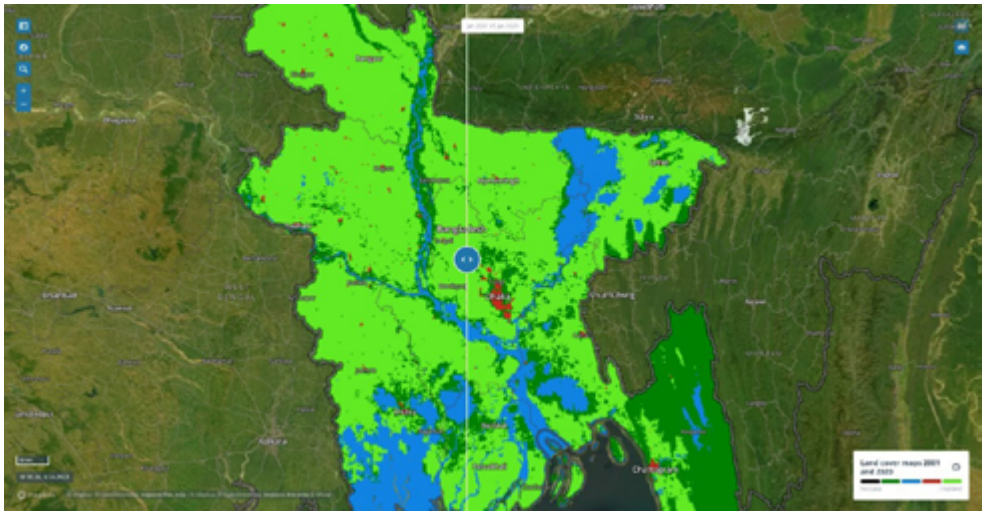
In the 2022 assessment cycle (concluded June 2023), the **SNWG MO led NASA's review of 115 surveys from more than 25 agencies**. All surveys were sorted into nine Earth science thematic areas, and NASA assessment leads conducted follow-up interviews with 99 respondents and drafted thorough assessment reports. Several new workflow management practices were implemented during this cycle to enable assessment teams to collaborate more efficiently and to communicate better with agency representatives. The results of these efforts will lead to selection of several projects to begin in 2025 that directly meet the needs of NASA data users.

VEDA Powers Open Science

VEDA is an open platform with services that support in-browser visualization and statistical analysis, collaborative development environments for data processing and analysis, and an interactive communications medium. The platform is designed to support scientists and their teams in performing research with collaborative compute environments while also bridging the gap between the research community and the general public through visual exploration and data stories. Moreover VEDA's data services are built to community standards supporting interoperability with similar cloud-based platforms and other applications including NASA's Enterprise GIS. The open platform is designed to be modular where adopters can reuse the platform in part or in whole.

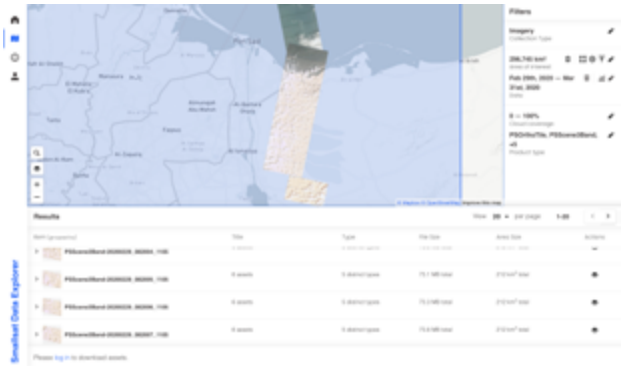
The VEDA platform offers analysis-ready Earth observation datasets collocated with open-source tools for interactive visualization, collaboration, and analysis, and processing. The rich set of data services extend beyond single interfaces, supporting user dashboards and integrating with GIS platforms.

Reuse of the VEDA platform is ongoing and supports several important NASA initiatives. For example, the **new U.S. Greenhouse Gas Center** was developed by redeploying the entire VEDA platform and customizing it to meet the needs of GHG Center stakeholders. Also, the joint NASA-ESA Multi-mission Algorithm and Analysis Platform (MAAP) data team deployed the data services components of VEDA to more efficiently serve data to MAAP biomass users. Finally, the visualization component of VEDA, powered by a dynamic tiling service, was deployed within the collaborative **NASA-USFS Fire Information Resource Management System (FIRMS)** to visualize Harmonized Landsat Sentinel-2 (HLS) data to support wildfire response.



Comparison slider for land cover in Bangladesh visualized in the VEDA dashboard

CSDA: Release of Additional Spire Soil Moisture Data



The CSDA Smallsat Data Explorer (SDX) website

IMPACT supports the Commercial Smallsat Data Acquisition (CSDA) program by operating and improving on previously developed processes and infrastructure. **CSDA has established sustainable data management processes** that make data from commercial vendors available to the broader research community and supports long-term preservation of the data to ensure scientific reproducibility. Over the past year, the CSDA team worked diligently to incorporate new vendors. They completed science evaluations of Airbus and Blacksky and initiated evaluations for GHGSat, GeoOptics, Capella Space, and ICEYE. Reports from these assessments will be available in the coming year, providing crucial insights into the capabilities and potential contributions of data from these vendors.

CSDA has recently **unveiled a suite of exciting new data products, including Spire's Soil Moisture, Planet's 8band, and Maxar's data products**, promising to revolutionize our understanding of Earth's dynamic systems. Moreover, CSDA has demonstrated its commitment to seamless data access and collaboration by mirroring critical data from the National Center for Climate Simulation (NCCS) and taking over data fulfillment support previously handled by CAD4NASA, reinforcing its pivotal role in advancing data-driven scientific research and innovation.

Open Source Tool Development

Developing and strengthening partnerships for knowledge exchange and infusing new technology to solve hard problems.

IMPACT is dedicated to open source tool development. DevSeed is a central collaborator with IMPACT in this area. One example is the published MAAP Boreal Biomass dataset that uses a combination of ICESat2, HLS, and Copernicus DEM data. This is a Level 4 product created from cloud optimized data, on cloud infrastructure, and delivered in a cloud-optimized format. **A new tool called stac_ipyleaflet** allows users to mix STAC data with local user data in interactive maps for interesting analysis.

Working with DevSeed, **IMPACT launched eoAPI**, a set of open-source, cloud-native data infrastructure tools for building Earth applications. eoAPI captures the libraries and architectures used across DevSeed/IMPACT work (most recently VEDA) and from supporting projects like the Microsoft Planetary Computer and the Amazon Sustainability Data Initiative.

Open Source Tools (cont)

IMPACT continued to push technological boundaries of what is possible with cloud native data formats. In collaboration with CarbonPlan, DevSeed/IMPACT have delivered performance benchmarks and developed guidance for new methods for Zarr visualization.

Along with DevSeed, IMPACT worked with NSIDC and ASF at ICESat-2 Hackweek to develop benchmarks for using ATL03 HDF5 via various cloud-optimized formats and methods.

HLS Opens Source Code and Completes Data Archive

The Harmonized Landsat and Sentinel-2 (HLS) project generates analysis-ready data products using Landsat-8 & 9 and Sentinel-2 that improve upon the temporal resolution of the data products derived from these sensors when used independently. As part of IMPACT's commitment to open science, the **HLS team made their codebase publicly available** through the IMPACT Github repository in May and a full release to NASA's official software catalog is in progress. This release gives HLS data product users greater visibility into the project's image processing pipeline. Users can now see how the team implements cloud processing methodologies and can apply their own image processing strategies to create custom data products. This code release fulfills Science Mission Directorate (SMD) policy SPD-41A, which requires that all products created using SMD funding be shared with the public.

Using the newly created **Blaze data transfer tool**, the team successfully transferred 11 million files, totaling almost six petabytes in data volume, from the European Space Agency (ESA) in May. In June, the team **completed historical processing** of the HLS Sentinel-2-derived (HLSS30) data, adding 1.5 petabytes of data to the existing archive. In combination with the HLSL30 archive (which extends back to 2013), the HLSS30 archive increases the temporal resolution of observations from 2015 to present allowing for more thorough scientific investigation. HLS data can be accessed at the Land Processes (LP) DAAC and on NASA Worldview.

IMPACT Projects by the Numbers

1371 CSDA Users

6 CSDA Datasets

27 APT Users

15 Active ATBDs

Create forward-looking data curation policies, tools, services and documentation by envisioning new ways to lower barriers to data and information

IMPACT
Interagency Implementation and Advanced Concepts Team

US GHG Center IMPACT Contributions

Using VEDA, a new U.S. Greenhouse Gas Center (US GHG Center) was rapidly designed by IMPACT and DevSeed team members during the summer of 2023. The **US GHG Center is a multi-agency (EPA, NOAA, NIST and NASA) effort** to integrate greenhouse gas data from observations and models into one open and accessible environment for visualizing and exploring trusted GHG datasets. The beta release contains more than a dozen datasets and includes articles, known as Data Insights, that introduce greenhouse gases and their effect on climate. The **project focus is to accelerate production and delivery of actionable GHG datasets and information**, workflows and visualizations from the U.S. federal government and the non-public sector. The IMPACT/DevSeed team will continue development of the Center during the coming year to release enhanced features and more data and data insights.

APT Updates the ATBD Creation Process

The APT website interface

The Algorithm Publication Tool (APT) is a web-based application designed to simplify Algorithm Theoretical Basis Document (ATBD) authoring, publishing, and discovery. APT is more than a tool; it is a process for supporting scientists in ATBD creation. In the past year, the APT team has **added several new updates** that improve the tool's usability, flexibility, and accessibility. Users can now choose to either author their ATBD in the tailored user interface or to download a template for use in Google Docs, Microsoft Word, or LaTeX. This new option provides more flexibility, allowing users to author offline in familiar tools or alternative online environments. Once an ATBD is completed, it is uploaded to APT for review, metadata addition, and public release. To improve document security, the team added multi factor authentication to the tool. Additionally, **six new user help videos** demonstrate specific tasks such as adding collaborators, requesting document review, or importing and adding references. **Journal-ready PDFs** are available, easing document preparation for ATBD submission to AGU's Earth and Space Science or other AGU journals. Over the next year, the APT team will shift focus from development to user support and feedback collection. Currently APT has 27 active users from projects including TEMPO, OPERA, OMPS, CSDA, and PACE. Their valuable feedback and experiences will enable the team to further improve the tool.

IMPACT Projects by the Numbers

46 VEDA Datasets

893 SDE Monthly Sessions

16 VEDA Data Stories

500+ CASEI Users

Community Recognition of IMPACT

In March of this year, several IMPACT team members were honored with prestigious NASA Silver Achievement Awards. **Manil Maskey**, who serves as the deputy manager of IMPACT, was awarded the NASA Silver Achievement Medal, and the entire **Airborne Data Management Group** (ADMG) was awarded the NASA Silver Group Achievement Award. **Rahul Ramachandran**, IMPACT's project manager, was named the 2023 AGU Earth and Space Science Informatics Greg Leptoukh Lecture recipient for his meritorious work and service that advances and promotes science discovery and solutions. **Iksha Gurung** was recognized by the UAH Earth Systems Science Center (ESSC) with the ESSC Meritorious Service Award for 2022. The award emphasized Iksha's ML-based approaches to meeting "one of the greatest challenges in science — to discover something new and interesting from an ocean-sized bucket of data." **IMPACT's VEDA team** was proud to be significant contributors to the NASA - ESA - JAXA Earth Observing Dashboard which was honored with the International Astronautical Federation Special Award on Space for Climate Change.

IMPACT @AGU2023

Monday, December 11, 2023

H11A-04 (ML), IN11A-02 (VEDA)

Tuesday, December 12, 2023

IN23A-03 (SDE), IN23A-04 (ML), IN24A-01 (ML), A23C-03 (ML), ED21A-02 (VEDA), IN21A-06 (VEDA)

Wednesday, December 13, 2023

IN31A-01 (ADMG), IN32A-08 (SDE), IN34B-05 (CSDA), IN33A-09 (MAAP)

Thursday, December 14, 2023

IN42B-05 (HLS)

Friday, December 15, 2024

IN53A-01 (FM), IN54A-02 (SDE), IN53A-05 (ML), INV51D-13 (VEDA)

"The Art of Scientific Curation" an AGU *Eos* Article

IMPACT team members Kaylin Bugbee, Deborah Smith, Stephanie Wingo, and Emily Foshee collaboratively wrote and published an article in *Eos*, AGU's science news magazine. Their article entitled "The Art of Scientific Curation" explored the benefits and nuances of ideal data curation practices and first appeared online in May. Look for it in the upcoming December paper issue of *Eos*.

Publication Spotlight

"Pixel level smoke detection model with deep neural network" led by Muthukumaran Ramasubramanian, Aaron Kaulfus, and others, was published in *Image and Signal Processing for Remote Sensing XXV*.

Pontus Olofsson co-authored a paper titled "Near real-time monitoring of tropical forest disturbance: New algorithms and assessment framework" was published in *Remote Sensing of Environment*.

Manil Maskey, Rahul Ramachandran, Brian Freitag, Aaron Kaulfus, Aimee Barciauskas, Olaf Veerman, Leo Thomas, Iksha Gurung, Muthukumaran Ramasubramanian co-authored the chapter "Dashboard for Earth Observation," in *Advances in Scalable and Intelligent Geospatial Analytics*.

Pontus Olofsson co-authored a paper that was recently published in *Remote Sensing of Environment*: "Continuous mapping of aboveground biomass using Landsat time series."

Rahul Ramachandran and Manil Maskey co-authored "Investigating Different Data-Traceability Approaches to Prevent Data Swamps" published in the September 2022 issue of the *IEEE Geoscience and Remote Sensing Magazine*.

Congratulations Corner

Congratulations to **Sujit Roy** on his marriage and **Slesa Adhikari** and **Caden Helbling** on their marriage.

Congratulations to IMPACT members **Georgios Priftis** and **Siddharth Chaudhary** on their respective marriage proposals and upcoming weddings.

IMPACT team members **Iksha Gurung**, **Muthukumaran R**, **Georgios Priftis**, and **Slesa Adhikari & Caden Helbling** purchased their first homes.

Rajat Shinde successfully completed a Ph.D. in computer science applications. **Christopher Phillips** was awarded a Ph.D. in atmospheric science. **Madhu Sridhar** completed a Ph.D. in mechanical engineering. **Nishan Pantha** recieved a MS in computer science.

Anabelle Brodsky, **Noam Mayerfeld**, **Shrey Gupta**, and **Tuong Phung** completed successful internships on the ADMG, CSDA, machine learning project teams.

IMPACT 2023 by the Numbers

14 Active Projects

123 Team Members

20 New Team Members

39 Publications

10 Outside Partner Collaborations

16 Student Research Assistants

38 Blog Posts