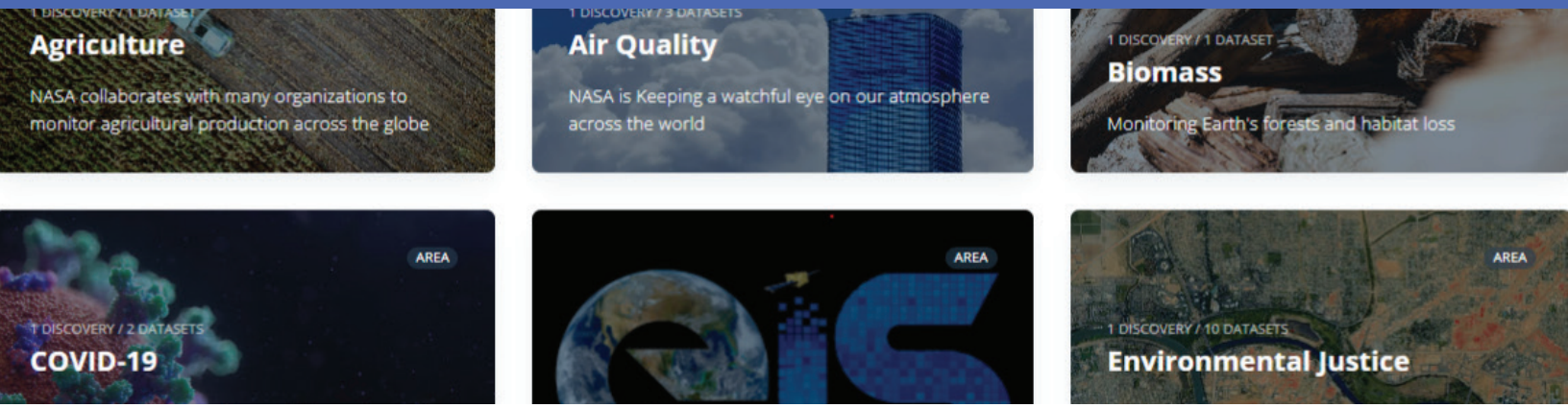
FALL 2022 | ISSUE 3

IMPACTful NEWS

Updates from the Interagency Implementation and Advanced Concepts Team



The new VEDA dashboard (VisEx) with tiled thematic areas containing datasets for users to browse, interact with, and explore.

VEDA Visualizes Data

The Visualization, Exploration, and Data Analysis (VEDA) project is NASA's unified open-source science cyberinfrastructure for data processing, visualization, exploration, and geographic information systems (GIS) capabilities. The project originated from the communal need for a standardized set of tools to analyze, explore, interact, and process NASA Earth science data in the cloud. VEDA provides users with a collaborative science environment (e.g., Jupyter notebooks) for data analysis and exploration; an interactive visual interface for storytelling; exploratory GIS and data analysis capabilities; a dynamic scaling platform for large scale processing; and a modeling and data assimilation infrastructure scaled through high performance computing and cloud.

VEDA's public-facing component, VisEx, is an interactive dashboard supporting the exploration, visualization, and analysis of NASA datasets and features such as data-driven Environmental Justice stories. The latest

ALSO IN THIS ISSUE

Assessing Federal Agency Earth Observation Needs

Smallsat Data Explorer Enhancements

ADMG Historic Data Recovery Efforts

Advancing Earth Science Through ML

Science Discovery Engine

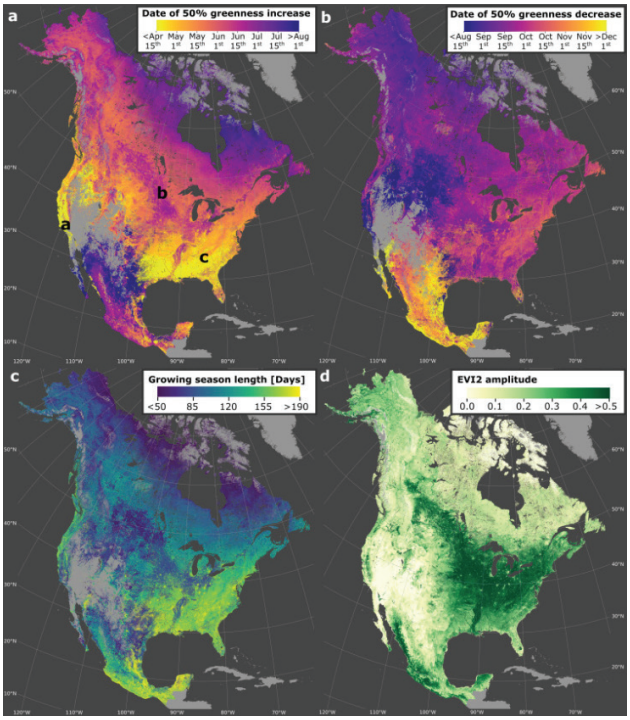
IMPACT
Interagency Implementation and Advanced Concepts Team

Harnessing technology advancements in data and information systems to expand community impact

IMPACT
Interagency Implementation and Advanced Concepts Team

VEDA (cont)

dashboard release introduced the platform to public users, allowing scientists at all levels to engage with NASA Earth science data related to agriculture, air quality, biomass, ocean/water quality, and more. In the future, VEDA aims to continue highlighting NASA Earth science data discoveries and Environmental Justice reports, ensuring the datasets, materials, and language are up to date.



The Harmonized Landsat 8 and Sentinel-2 (HLS) data is achieving unprecedented temporal resolution which is enabling exciting new research. Bolton et al. (2020)

Assessing Federal Agency Earth Observation Needs

The Satellite Needs Working Group (SNWG) launched the 2022 Satellite Needs Survey this past June. The SNWG identifies federal agency Earth observation needs, connecting agencies to current and upcoming satellite missions and identifying and implementing new activities that benefit multiple agencies. Participation in the 2020 survey increased by 50%, and the 2022 cycle acquired a similar amount of submissions. During the eight-month assessment period, the SNWG Management Office at IMPACT will coordinate interviews and solution discussions with stakeholders in NASA, USGS, and NOAA. The SNWG Management Office is also kicking off formulation for four activities from the 2020 cycle: vegetation indices derived from Harmonized Landsat Sentinel-2 (HLS) data, near real-time data products for NASA's upcoming TEMPO mission, a merged planetary boundary layer product, and expansion of the Pandora network to improve air quality forecasts. An important addition to the 2022 cycle is the Stakeholder Engagement Program (SEP). The program, led by the SNWG Management Office, will ensure that training and data products necessary to support implemented services are established and enhanced.

Smallsat Data Explorer Enhancements

The CSDA team rolled out a number of enhancements to the Smallsat Data Explorer (SDX) tool in FY2022. These enhancements included a transition from an order processing-based system to a quota-based system which, along with the introduction of improved download capabilities, puts data more directly into users' hands. The SDX interface was updated with a revised layout to better support reviewing discovered data. The team also developed spatial heat map and temporal histogram features for the search interface to improve discovery of spatially and/or temporally disparate data. In addition, new Spire and Planet data are now available in the SDX.

The Smallsat Data Explorer web interface showing the new heat map layer

APT Simplifies ATBD Production and Publication

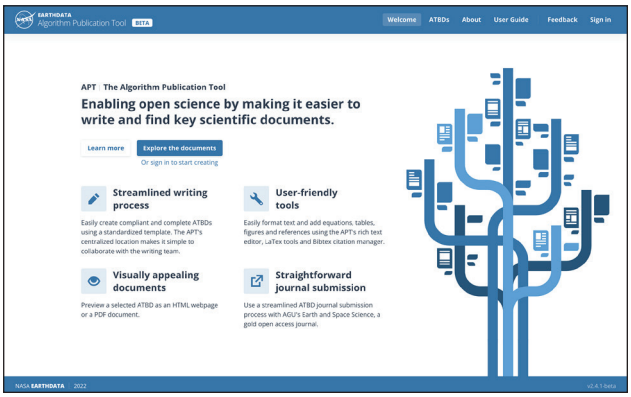
An Algorithm Theoretical Basis Document (ATBD) describes the algorithms used to convert radiances, whether from passive or active remote sensing instruments, into geophysical parameters. These documents are useful for scientific study and are an important component of open, reproducible science. Historically, producing ATBDs has been difficult because authors must carefully consider what content to include and who may be searching for the documents. Also, data users may encounter frustration when searching for unpublished ATBDs. The recent release of the Algorithm Publication Tool (APT) has the potential to significantly enhance ATBD production and accessibility. IMPACT, in partnership with developers at Development Seed, designed and created the APT to ensure ATBD contents

Leading the development of innovative open data systems to support rapidly evolving data production and management needs

IMPACT
Interagency Implementation and Advanced Concepts Team

APT Simplifies (cont)

are searchable, citable, and obtainable. Through this improved ATBD stewardship, the knowledge behind data products is more openly shared, thereby meeting NASA's Open Science directive. The APT is now available to science teams for ATBD construction and release. Soon the public will be able to access completed and released ATBDs on the APT website.



The Algorithm Publication Tool beta version web interface

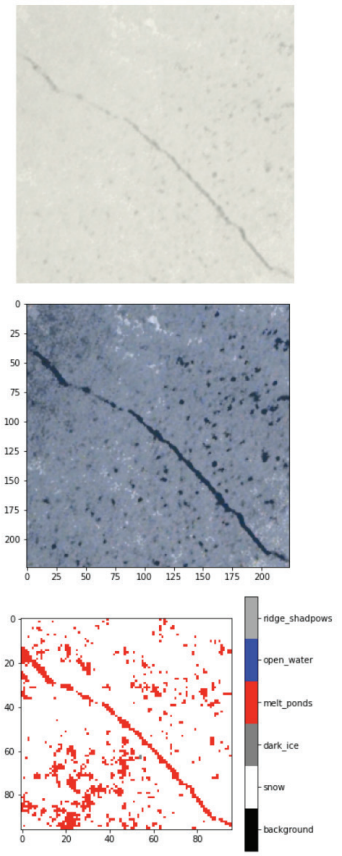
ADMG Historic Data Recovery Efforts

For more than 50 years, NASA has flown airborne sensors to carry out research, validate satellite sensors, and test newly developed instruments. Prior to 2000 there was no centralized, reliable way to publicly access all NASA airborne data. Research scientists involved in field campaigns retained instrument data for their own use and that of collaborators. Since 2018, the Airborne Data Management Group (ADMG) has been working to locate, recover, and facilitate the transition of historical data to NASA Distributed Active Archive Centers (DAACs) for proper data stewardship. This ongoing effort opens field data to newer users and brings collections out of the closets or warehouses where stored, sometimes for decades. This year, ADMG arranged for the rescue of historical high resolution air photos taken from the NASA ER-2 and other aircraft. The camera film was stored in metal cans for decades on shelves in a warehouse at NASA Ames Research Center. ADMG successfully arranged for transfer of these air photos to the NASA Johnson Space Center for digitization. ADMG handles more than just analog data recovery. One of ADMG's primary tasks is to gather the contextual information needed to reuse this older data and provide the detailed information in the Catalog of Archived Suborbital Earth Science Investigations (CASEI). This fall, two ADMG team members, Shelby Bagwell and Ashlyn Shirey, traveled to Wallops Flight Facility to gather P-3 aircraft flight information from paper records stored at the base. This information will contribute to CASEI curation efforts and create a detailed history of past campaigns and flight activities involving the P-3.

Shelby Bagwell and Ashlyn Shirey in front of NASA's P-3 aircraft

Advancing Earth Science Through Machince Learning

The IMPACT machine learning team launched a similarity search tool that enables the discovery of images in Earth science observation data that match a selected feature or event. The tool utilizes MODIS data from NASA's GIBS service. Users can click on events in the displayed map to initiate searches for similar images across 197 million square miles of satellite imagery. They can also search according to date.



Machine learning predictions of supraglacial melt

Supraglacial lakes (SGLs), also known as melt ponds, result from meltwater accumulation in topographic depressions on the surface of glaciers. SGLs affect ice sheet dynamics through a positive feedback loop in which the albedo-lowering effect of SGLs can escalate surface melt leading to increases in lake extent and depth. Small satellite imagery with high temporal and relatively high spatial resolution offers the potential to closely monitor dynamic changes in SGL formation, drainage and associated impacts on ice shelves. The machine learning team used supervised semantic segmentation leveraging a convolutional neural network architecture to train a model with imagery and labels from NASA Operation IceBridge and predict SGL occurrences in high temporal resolution PlanetScope small satellite imagery. This approach was chosen to enable rapid prototyping of a model without the need for manually creating labels ourselves, and was accomplished by downsampling the NASA Operation IceBridge data to simulate PlanetScope spatial resolution before model training. The result is a model equipped for segmenting SGL occurrences in PlanetScope daily imagery at a spatial resolution of 3 meters.

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

IMPACT
Interagency Implementation and Advanced Concepts Team

The machine learning team has also implemented OpenAI application programming interface (API) endpoints which access the GPT-3 model to enable text classification. The team used the FineTune API, which allows users to provide examples of the text and their respective ground truth, and then builds a model based on it. Once the model is trained, the text is sent as a prompt to the API, along with the model ID that

Machine Learning (cont)

represents the trained model, to get the predictions. Below is an example.

Source paragraph:

"Surface flow velocities in river channels: Remote sensing of surface flow velocities could become one component of a non-contact approach to measuring river discharge, which is an important objective of the USGS that helps the agency to fulfill our mission to provide reliable streamflow information for the Nation."

Predicted classification from the ML model:

{ "GPT_prediction": "water and energy cycle"}

IMPACT Projects by the Numbers

5
Country Biomass Maps (MAAP)

5
Biomass Dashboard Products (MAAP)

75
Campaigns Curated (CASEI)

pyQuARC Becomes a Service

pyQuARC is an automated tool for performing metadata quality assessments developed by IMPACT's Analysis and Review of the CMR (ARC) team. pyQuARC builds off of QuARC which was developed in coordination with the Commercial Smallsat Data Acquisition team. Running QuARC ensures that metadata is complete, consistent, and correct prior to publication or for routine checks. The release of pyQuARC as a service makes it easier for users to implement metadata review processes. They can interact with a simple API interface, so they don't require any technical expertise to perform the metadata quality checks. QuARC is currently in beta release as pyQuARC is not yet finalized, but the most recent version can be accessed from the pyQuARC github.

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

IMPACT
Interagency Implementation and Advanced Concepts Team

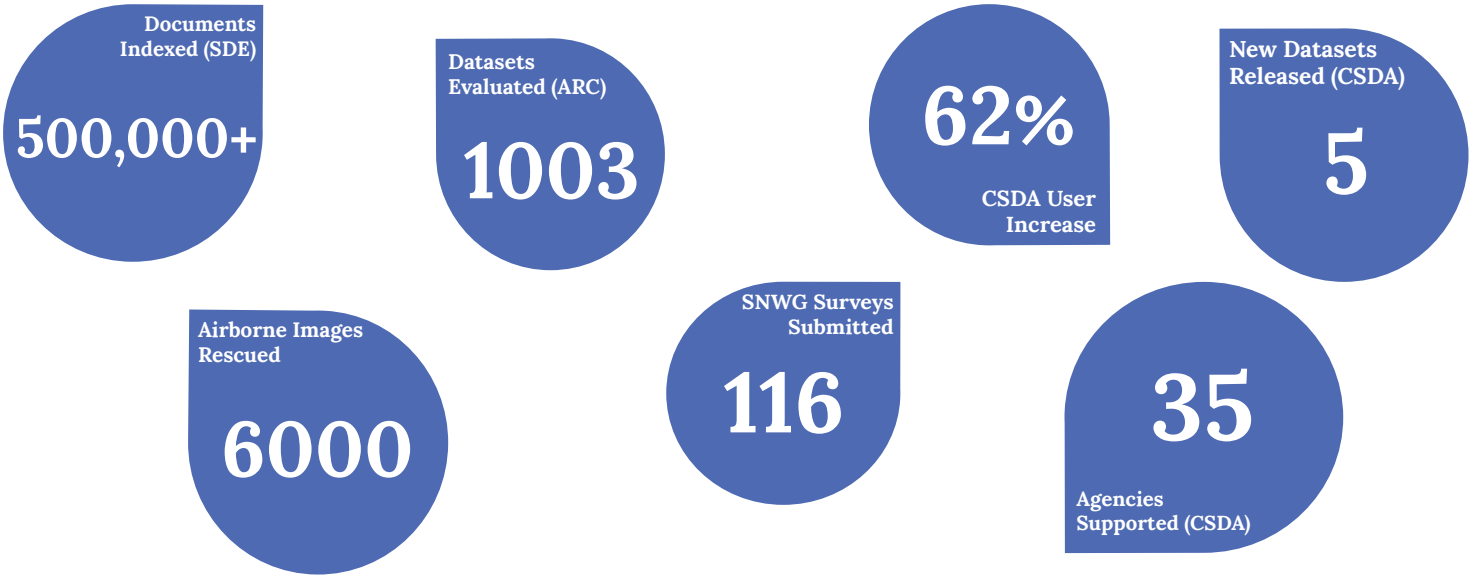
Science Discovery Engine Releases Beta Version

The Science Discovery Engine (SDE) project within IMPACT is a key component in NASA's effort to construct an open-source science infrastructure. SDE is building an integrated Science Mission Directorate (SMD) search capability to enable discovery of open data and information across SMD's five divisions: Astrophysics, Biological and Physical Sciences, Earth Science, Heliophysics and Planetary Science.

In May 2022, the team released the beta version of SDE to NASA users. This version included data from Biological and Physical Sciences, Earth Science and Heliophysics. Recently, the SDE team completed the indexing of data and information from all five of the SMD divisions. SDE also supported the Frontier Development Lab (FDL) SMD Knowledge Graph Discovery challenge this year. The FDL investigated whether NLP techniques could be used to develop more effective data discovery methods by embedding modern scientific language models into search tools. In September 2022, the SDE team presented at the Open Source Science Data Repositories workshop. The purpose of this workshop was to engage with NASA data repositories to plan the next steps in NASA's open science initiative.

Throughout the year, the SDE team developed an SMD vocabulary extraction workflow that leveraged over 50 glossaries, thesauri, and keywords across the SMD to generate term lists such as platforms, instruments, and missions. The first iteration of the term lists have been integrated into the SDE, and a primary goal of FY23 will be to refine the existing lists and develop new lists.

IMPACT Projects by the Numbers



IMPACT hosts a seminar series, called Tech Talks, aimed at sharing technical expertise across various topics relevant to IMPACT efforts. Some of the tech talks featured IMPACT team members and others highlighted external guest speakers. All tech talks are recorded and available online from the IMPACT website after the event.

This year's speakers were:

- Dr. Kelly Rose and Vic Baker: "Developing a Smarter Way to Search - Parsing the Online "Forest" To Find Data for Your Research Needs Using a Scalable Integrated Data Discovery System"
- Dr. Surya Durbha: "Five Geo-Computational Challenges for Rapid Assessment of Post-Disaster Areas Using Earth Observation (EO) Data"
- Dr. Marquita Ellis and Dr. Davis Wertheimer: "Training Foundation Models Anywhere with Hybrid Cloud"

IMPACT @AGU2022

Monday, December 12, 2022

IN12A-07 (CASEI), ED16B-07 (SDE), IN16A-08 (ML), SY12B-0393 (SNWG), SY15B-0412 (SNWG), ED11B-01 (ML), ED12C-0368 (Partnerships)

Tuesday, December 13, 2022

H21C-11 (ML), IN22D-0328 (ML), IN22D-0332 (ML)

Wednesday, December 14, 2022

IN32C-0396 (ADMG), IN35A-01 (CASEI), IN35B-0406 (QuARC), IN34A-05 (CSDA)

Thursday, December 15, 2022

IN43A-04 (CSDA), IN41A-05 (SDE), C44A-08 (ML), IN45A-02 (VEDA), IN45A-07 (ML), IN42B-0332 (FAIR), IN42C-0342 (MAAP), SY45D-0661 (Open Science), H46D-07 (ML), N42A-09 (ML), IN45B-0356 (ML), B42C-07 (HLS)

Friday, December 15, 2022

IN56A-08 (ML)

Student Spotlight

IMPACT is a close collaboration between NASA and the University of Alabama in Huntsville (UAH). A key component of this partnership is the opportunity IMPACT provides for graduate and undergraduate students at UAH. As part of the IMPACT team, students in the fields of Earth science and computer science gain direct hands-on research experience that enhance their classroom experience

Machine Learning: **George Priftis**, **Satkar Dhakal**, **Rajesh Pandey**, **Meghana Paramesh**

MAAP: **Anish Bhusal**

SNWG: **Binita Gyawali**

ADMG: **Jillian Ethridge**, **Pottimuthi Alekhya**

HLS: **Eric Rice**, **Udaykumar Bommala**, **Nishan Pantha**

CSDA: **Ankur Kumar**, **Suraj Regmi**, **Dan Pham**, **Panday**, **Lance Linder**, **Girdharilal Pandey**, **Rajashree Dahal**, **Meilia Tecson**

Science Discovery Engine: **Saurav Upadhyaya**

VEDA: **Paridhi Parajuli, Prakash Chaudhary**

Interns: **Shankar Bahadur**, **Kyle Song**, **William Diana**, **Michael Tsai**, **Petru Zubek** – Ridgefield High School

Topu Saha – New Jersey City University

Shreya Shrestha – Vanderbilt University

Melanie Shariff – University of Colorado, Boulder

Ryan Aspenleiter – University of Texas at Dallas

Deepak Menon – Poolesville High School

Hadeel Farhan – Northeastern University

Bryanna Gutierrez-Coatney – Arizona State University

Congratulations Corner

Congratulations to IMPACT member **Camille Woods** and family who welcomed a baby boy, Cayden King, in February 2022.

Congratulations to IMPACT member **Brian Freitag** and family who welcomed a baby boy, Ender Freitag, in June 2022.

Aaron Kaulfus successfully completed a Ph.D. in atmospheric science from UAH.

Madhu Sridhar completed a Ph.D. in mechanical engineering from UAH.

Ashlyn Shirey received her M.S. in atmospheric science from UAH.

Jillian Ethridge was selected for the NASA Student Airborne Research Program and completed this internship over the summer. Her SARP research project applied machine learning to ocean science, and she was chosen to represent the SARP program by presenting her work at AGU in December.

IMPACT 2022 by the Numbers