IMPACTful NEWS

Updates from the Interagency Implementation and Advanced Concepts Team

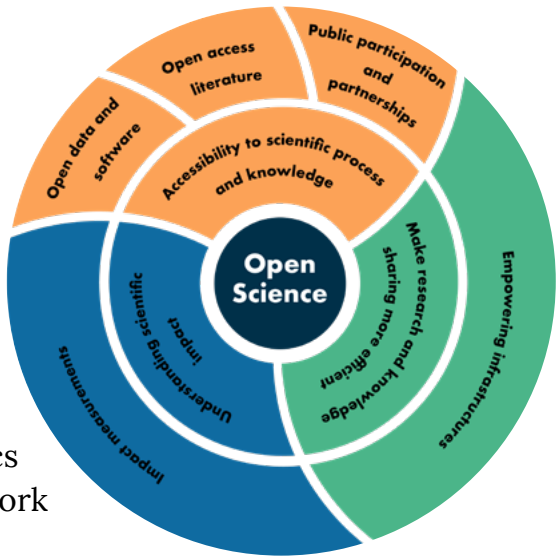
ADMG Launches CASEI

The new CASEI user portal home page

The Airborne Data Management Group (ADMG) has been busy developing, designing, building and curating content for the new airborne and field campaign discovery portal called CASEI - the Catalog of Archived Suborbital Earth Science Investigations. This user portal was released in June 2021 and serves as a knowledge center for exploring information pertaining to NASA airborne and field campaigns and finding the data products available. CASEI is a unique inventory of curated information about the campaign context, research motivation, funding, and platform, instrument and measurement details. Data collected by non-satellite instruments are important to many areas of study. CASEI currently has 60 curated campaigns of the 152 identified NASA-funded or NASA-participated campaigns. Many more will be curated in the coming fiscal year. The release of CASEI demanded a coordinated effort from the ADMG team, other IMPACT team members, and Development Seed team members. Explore CASEI to learn more about NASA's airborne and field campaigns.

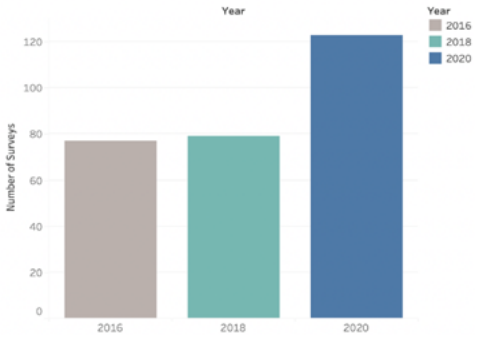
Open-Source Science

IMPACT focuses on open science and open data, a cross-cutting theme across all NASA projects, especially how it relates to data and metadata. As a part of this effort, Dr. Rahul Ramachandran presented "The Role of Data Systems to Enable Open Science" for the Earth Science Informatics Technical Committee Webinar for the Geoscience and Remote Sensing Society, a technical Society of the Institute of Electrical and Electronics Engineers. For FY 2022, each project in IMPACT has aligned work with NASA open science policies.



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

Connecting Federal Agencies to NASA Data



Number of responses returned for the past three SNWG survey cycles

There is a growing interest in the Satellite Needs Working Group (SNWG) biennial survey of Federal agencies. Survey completion showed a 50% increase in 2020. Due to this increased response, NASA has established the SNWG Management Office at IMPACT to manage the assessment of agency needs for satellite Earth observations and the implementation and operations of solutions. The SNWG Management Office coordinates all SNWG funded projects, including those from previous surveys. Also new this year is a SNWG Stakeholder Engagement Program (SEP) which provides training and community engagement data products. The Management Office also produced an automated survey evaluation of reported needs and tools used by all assessment participants. IMPACT is using this evaluation to inform new tool development and make updates prior to the next assessment survey in 2022.

Hi-Res Look at Wildfires in the West

The Harmonized Landsat and Sentinel-2 (HLS) project generates analysis-ready data products using Landsat-8 and Sentinel-2 that improve upon the temporal resolution of the data products derived from these sensors when used independently. Harmonization requires bringing consistency to the data regardless of the source. Over the last year, the IMPACT HLS project has refactored the HLS v1.4 algorithm for cloud optimization, expanded the algorithm to provide global coverage, and converted the data to cloud-optimized geotiffs for the HLS v2.0 data products. In addition, the project created browse imagery from these products formatted for use in NASA's Global Imagery Browse Service and Worldview applications. The data are available to users from the HLSL30 and HLSS30 data set landing pages.

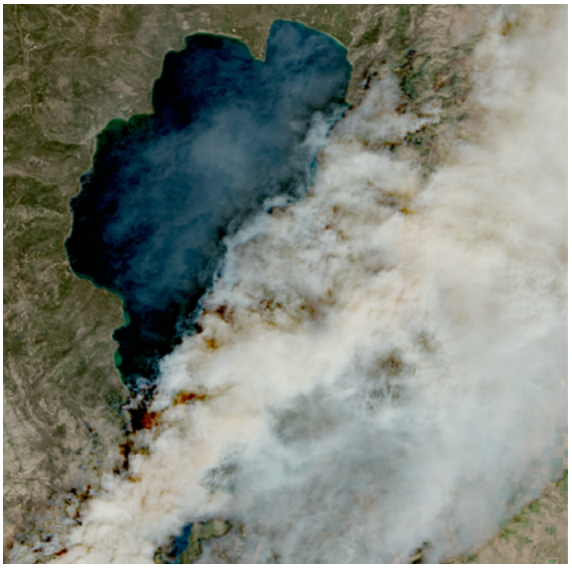


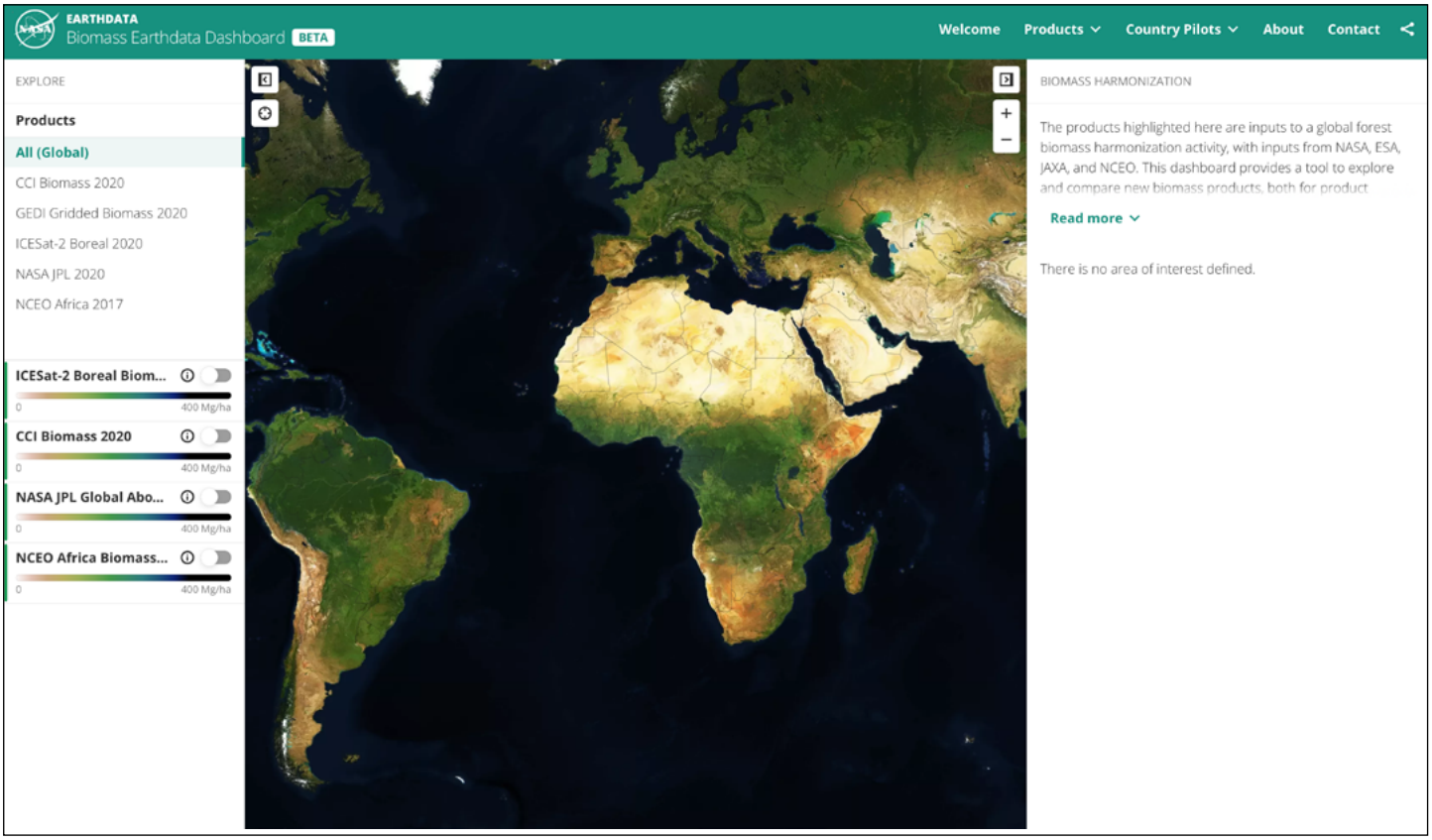
Image created using HLS data showing the Caldor Fire in South Lake Tahoe, California, USA, on September 1, 2021 at 18:49 UTC

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

MAAP Releases the Biomass Dashboard

The Multi-Mission Algorithm and Analysis Platform (MAAP) is a collaborative, open-source science tool jointly managed by NASA and the European Space Agency (ESA) that provides seamless access to NASA and ESA data and enables scientists to jointly analyze large volumes of data at scale. The NASA MAAP data team, led by IMPACT, provides many capabilities to help users more effectively find and use data within the MAAP. This initial application focuses on above-ground biomass: the size and carbon content of Earth's forests. Understanding biomass distribution and change is critical to understanding the global carbon cycle and global climate change.

The prototype MAAP dashboard was released in July 2021. The MAAP dashboard leverages capabilities created for the Environmental Monitoring dashboard initially developed in support of the COVID-19 partnership between NASA, ESA and the Japan Aerospace Exploration Agency (JAXA) in 2020. The MAAP dashboard offers a new way for users to visualize biomass products within the platform and makes it possible to curate stories around regions or topics of interest relevant to biomass change. Explore the MAAP biomass dashboard to visualize the datasets.



The MAAP dashboard user Interface

SAA Success: Petabyte-scale Imagery Search with New SpaceML Tool

A key IMPACT objective is to build strategic connections through public-private partnerships established via the NASA Space Act Agreements (SAA). IMPACT has four SAAs currently in effect or being finalized. Under the Google SAA, IMPACT explores the use of commercial Artificial Intelligence (AI) and Machine Learning (ML) techniques to address existing data discovery, access and use challenges with the Frontier Development Lab (FDL), a NASA, SETI Institute, and Trillium Technologies collaboration.

One of the many FDL projects is SpaceML, an open science initiative and toolbox to help the AI community develop an open-source framework that simplifies the discovery of satellite imagery in NASA Worldview. IMPACT joined the SpaceML effort to develop and deliver a new search engine prototype that can perform content-based searches of NASA's WorldView Earth science imagery archive. The pilot was unveiled in April 2021.

For more information about the Google partnership, visit the ESDS Public/Private Partnership website and also explore the NASA-GIBS Imagery Downloader aand the Self Supervised Learner, both products of the SpaceML initiative.

Training Models with ImageLabeler

In May 2021, IMPACT's Machine Learning (ML) team released the open source, web-based tool called ImageLabeler that allows users to create tagged images for use in training image-based ML models for Earth science phenomena. The ImageLabeler is designed to serve two purposes: 1) to provide a catalog of candidate events for scientific investigation, and 2) to gather, in a central location, the data required to train ML algorithms to automatically detect specific Earth events. ImageLabeler provides researchers access to important training sets and also allows for the creation of new training sets. To learn more about how to use ImageLabeler visit the web page.

The ImageLabeler home page

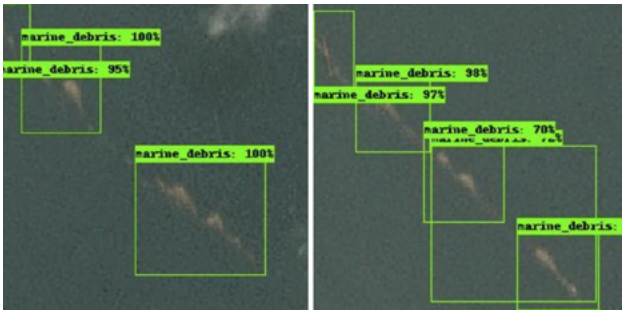
Developing and strengthening partnerships for knowledge exchange and infusing new technology to solve hard problems.

The APT logo

APT Streamlines ATBDs

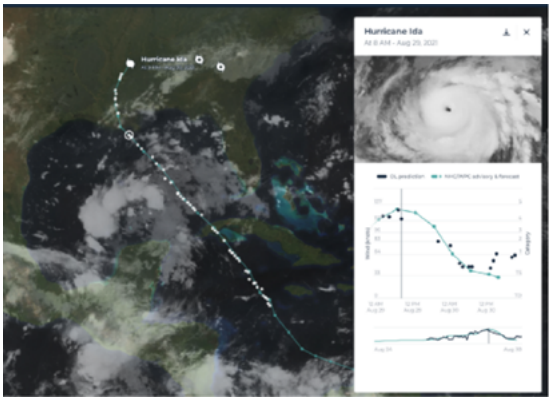
IMPACT and Development Seed have collaborated to design and develop the Algorithm Publication Tool (APT), a web-based application for streamlining the authoring, management, and discovery of Algorithm Theoretical Basis Documents (ATBDs). Historically, these important documents have been difficult to locate with no traditional design or content expectations. The APT is meant to solve both these issues by bringing standardization to ATBDs and making the contents searchable and easily accessible from one location. The APT team has worked with an AGU journal editor to also provide an easier route to journal publication for their ATBD. Approved authors complete an online form to write, edit, review and prepare ATBDs for publication. All users can freely and easily search the APT holdings to find the algorithm descriptions they seek. Routine use of APT is expected to begin in 2022.

Community and Coding with Machine Learning



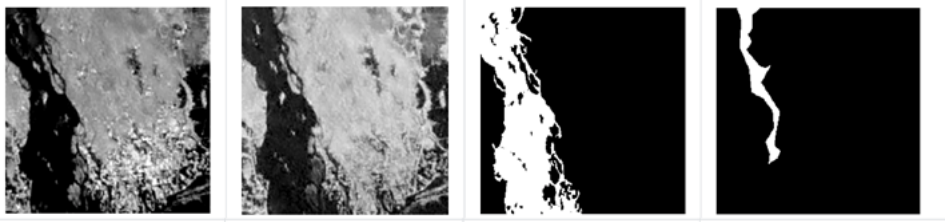
This past year, the IMPACT Machine Learning (ML) team has produced several datasets using ML/AI techniques:

Marine Debris Dataset: Uses commercial Planet data to detect plastic waste/pollution in the ocean.



Hurricane Dataset: Uses a collection of GOES Longwave Infrared Radiation (IR) data and HURDAT-2 wind intensities to estimate a hurricane's track and intensity.

Flood Detection Dataset: Uses SAR imagery acquired from ESA Sentinel 1A/1B missions to identify flood regions labeled by Subject Matter Experts (SMEs).

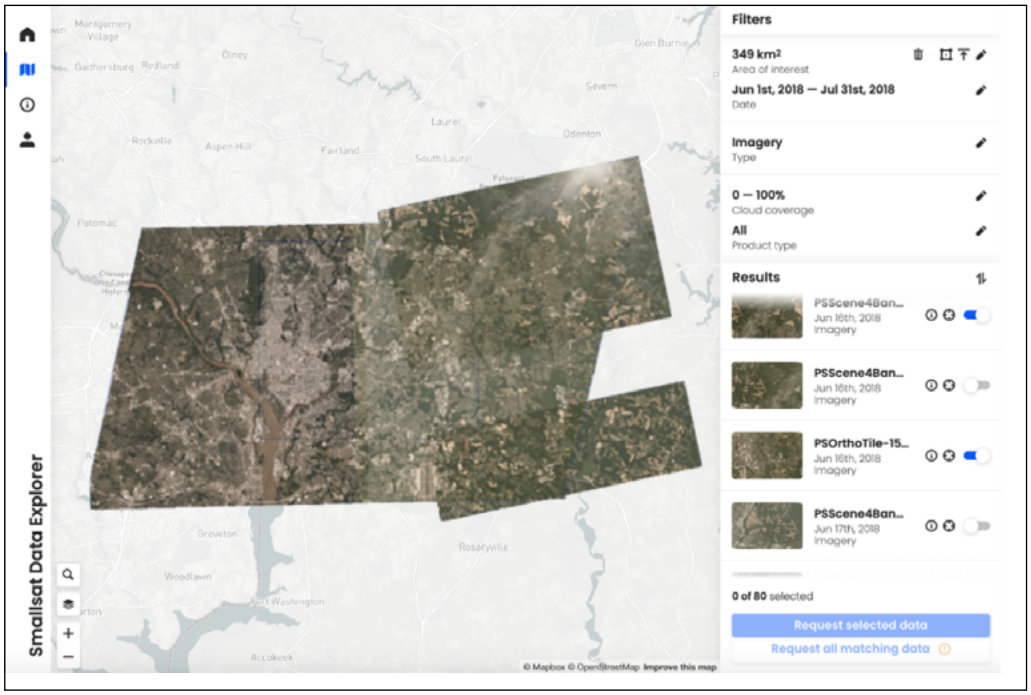


(a) Sentinel-1 image (Grayscale: VV) | (b) Sentinel-1 image (Grayscale: VH) | (c) Reference flood data (Grayscale) | (d) Reference water body data (Grayscale)

Bringing Commercial Data to Researchers

The Commercial Smallsat Data Acquisition (CSDA) Program was established to acquire commercial satellite data and make it available for research and applications by the NASA Earth science community. The CSDA data team, located at Marshall Space Flight Center, supports the program by managing all acquired data after vendor acquisition. The team also develops procedures and services to support the search, discovery, distribution of commercial data. Current commercial vendors include Planet Labs, Inc., Spire Global, Inc, Maxar and Teledyne-Brown Engineering, Inc. The data product access varies by vendor and remains restricted to U.S. Government-funded researchers for science use only. The CSDA data team received greater than a four-fold increase in the total number of access requests after recent end-user license agreement improvements allowing for broader data use.

Additional details on the data available, restrictions on data use, and how to access the data can be found on the CSDA website. Any user can explore the CSDA datasets through the Smallsat Data Explorer.



Screenshot of the Smallsat Data Explorer (SDX)

Learn More About IMPACT

Follow IMPACT's blog.
Gain an overview of IMPACT within Earth Science Data Systems.
Visit the IMPACT project website for more about IMPACT projects and access to tools.
IMPACT-related content is shared via NASA Earthdata's official account: @NASAEarthData.

IMPACT launched a brand new seminar series in 2021, 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. Hannah Kerner: "Advancing Global Food Security and Sustainable Development Goals with Machine Learning and Earth Observations"
- Sukriti Sharma: "Watson NLP Library and Demo"
- Dr. Dennis Gannon: "Building a Tiny Knowledge Graph with Bert and Graph Convolutions"
- Dr. Sundar Christopher: "From Pixels to Products: An Overview of Satellite Remote Sensing"

Recent IMPACT Publications

"Improving Discovery and Use of NASA's Earth Observation Data Through Metadata Quality Assessments"

"Science Storms the Cloud"

"Exploring the Use of PlanetScope Data for Particulate Matter Air Quality Research"

"On Measuring and Employing Texture Directionality for Image Classification"

"Advancing Machine Learning Tools for Earth Science Workshop [Technical Committees]"

"Call to Action for Global Access to and Harmonization of Quality Information of Individual Earth Science Datasets"

"From Open Data to Open Science"

"Daytime and Nighttime Dust Event Segmentation using Deep Convolutional Neural Networks"

Open Source Library for Metadata QA

IMPACT's Analysis and Review of Common Metadata Repository (ARC) team has developed an open source, python-based library called pyQuARC. The ARC team conducts quality assessments of NASA's metadata records in the CMR, and their framework consists of quality criteria for a series of automated and manual metadata checks focused on correctness, completeness, and consistency. In addition to basic validation checks, pyQuARC flags opportunities to improve or add contextual metadata information in order to help the user connect to, access, and better understand relevant data products. pyQuARC also ensures that information common to both the data product (i.e. collection) and the file-level (i.e. granule) metadata are consistent and compatible. As open source software, pyQuARC can be adapted and customized to allow for quality checks unique to different needs, and demonstrates the ARC team's commitment to open science and open source. Explore the GitHub page for pyQuARC for coding resources.

Student Spotlight

Several undergraduate and graduate students of the University of Alabama in Huntsville (UAH) support various IMPACT projects part-time as Research Assistants or Graduate Research Assistants (GRAs). IMPACT is grateful for their hard work and dedication to the success of our projects. Many thanks to all our students.

Ashlyn Shirey: graduate student working with ADMG

Lucia Alonso Guzman: graduate student working with ADMG and CSDA

Jillian Ethridge: undergraduate student working with ADMG

Ankur Kumar: graduate student working on CSDA

Meilia Tecson: undergraduate student working on CSDA

Suraj Regmi: graduate student working on KnowledgeGraph

Satkar Dhakal: graduate student working on KnowledgeGraph

Georgios (George) Priftis: graduate student working as subject matter expert for Marine Debris and PM2.5

Manisha Khatri: graduate student who worked on PM2.5

IMPACT hosted several interns virtually over the summer of 2021:

Damian Ugalde, undergraduate double major at Cal Poly Pomona, worked on CSDA

Nia Asemota, undergraduate double major at the New York University – Polytechnic School of Engineering, worked on CSDA

Bethany Kuo, undergraduate double major at University of Maryland - College Park, worked on KnowledgeGraph

Anisha Kabir, undergraduate student at UC Santa Barbara, worked on ESSI trend analysis

Tony Park, undergraduate student at the University of Michigan — Ann Arbor, worked on the SMD AI project

Congratulations Corner

Congratulations to IMPACT team members **Dr. Stephanie Wingo** and family who welcomed a baby girl, Eyla Mae Wingo, in September 2020 and **Elizabeth Fancher** and family who welcomed a baby girl, Lily James Fancher, in December 2020.

Congratulations to these IMPACT UAH graduates:

Shelby Bagwell graduated with a Bachelor's degree in Earth system science from the University of Alabama in Huntsville in spring 2021, and has joined IMPACT full-time to work with ADMG and ARC.

Prasana Koirala graduated with a Master's degree in computer science in summer 2021 and joined the IMPACT dev team to work on the KnowledgeGraph, ImageLabeler, SNWG, and CSDA projects.

Essence Raphael graduated with a Master's degree in atmospheric science in summer 2021 and then joined IMPACT's ARC and DCD projects.

Jeanné le Roux graduated with a Master's degree in Earth system science in December 2020 and leads the ARC project in addition to adding her skills to DCD and other projects.

Doing More with Natural Language

The **GCMD Keyword Recommender (GKR)** has been incorporated into the Metadata Management Tool for seamless use by human curators so they can accurately tag data with useful keywords. GKR also provides recommendations for science keywords related to data at the time of ingest.

The initial version of **BERT-E**: an Earth science specific language model in the Hugging Face model hub has been released. This model is further trained on top of scibert-base using masked language modeling loss (MLM). The corpus is roughly 100,000 Earth science-based publications.