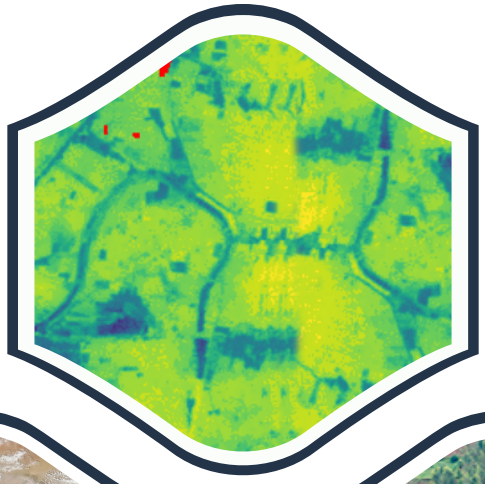
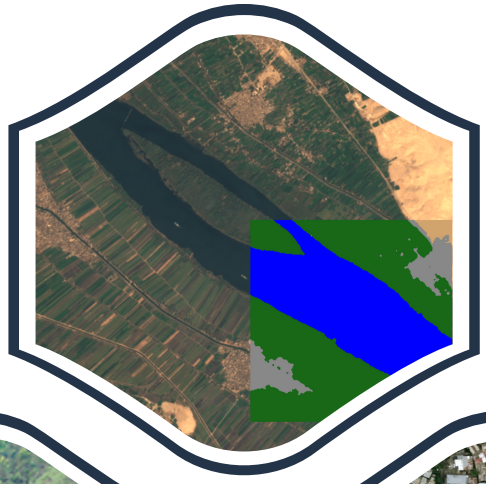
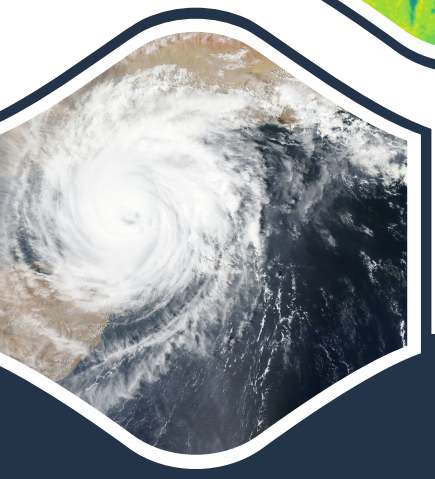
Radiant Earth Foundation

RADIANT EARTH FOUNDATION

2022 ANNUAL REPORT

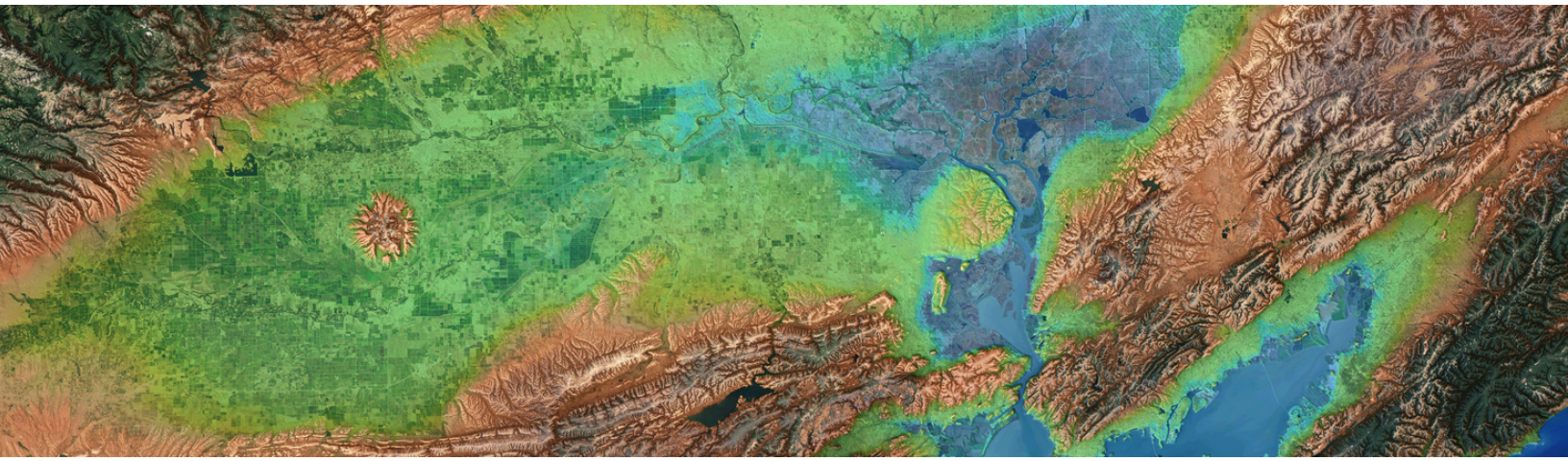





EXPANDING ACCESS TO EARTH SCIENCE DATA

Table of Contents

About Radiant Earth



The future of our species depends on our ability to develop sustainable methods of sharing data. Shared access to data allows people to develop a fact-based and shared understanding of our world, which is necessary to collaboratively address global challenges.

This belief underlies our mission at Radiant Earth Foundation, which is to increase shared understanding of our world by expanding access to geospatial data and machine learning models. Expanding access to shared data is necessary to create a larger and more diverse scientific community, reduce the cost of research, enable reproducibility of research, and foster collaboration among institutions.

Radiant expands access to data in two ways:

- Encouraging the adoption of efficient patterns and best practices for sharing Earth science data on the Internet.
- Providing Radiant ML Hub, a neutral data publishing platform that allows trusted organizations and individuals to share data and machine learning models with one another.

Message from our Executive Director

JED SUNDWALL
Executive Director

As the new Executive Director of Radiant Earth Foundation, I am honored to lead our work to serve the global Earth science community and expand our team's accomplishments in 2022.

Working with global partners in the public and private sectors, Radiant has developed new data products, standards, and tools that have accelerated critical work carried out by the international development community, including crop type monitoring and infrastructure planning. All of our work has been enabled by Radiant MLHub, our neutral data publishing platform for data and machine learning models, and our community-led efforts to enable data access and interoperability through the SpatioTemporal Asset Catalog (STAC) specification.

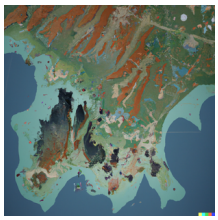
We share the highlights of our impact and multiple testimonials from data users in this report.

The second part of this report discusses our ambitions for 2023 – named the "Year of Open Science" by the White House Office of Science and Technology Policy. Radiant was founded on the principles of open science, intending to expand access to Earth science data and facilitate collaboration across scientific researchers, policymakers, businesses, and other stakeholders.

This year, we plan to open up further and increase the use cases and audiences that we can support with Radiant MLHub and community-led data sharing best practices. Much more work needs to be done to ensure that researchers have access to the data they need to address our world's greatest challenges. To the extent that we can make data more accessible, we can create a larger, more diverse, and more collaborative Earth science community.

2022: Impact and Partners

Key Impacts

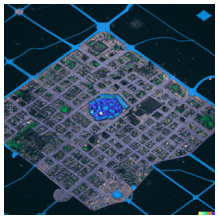


IMPROVED GEODIVERSITY OF TRAINING DATASETS

Added 39 high-quality benchmark training datasets with licenses granting public permission for users to access, use, and share the data. Datasets are from organizations around the world and across all sectors. The geographical distribution identifies the regions from which more data is needed.

ENHANCED DATA INTEROPERABILITY

Unified standards and guidelines for data collection and publication across the geospatial ecosystem to support the interoperability of data and tools.



PRODUCED NEW GEOSPATIAL DATA PRODUCTS

Novel data products, including LandCoverNet, the first multi-mission (Landsat-8, Sentinel-2, and Sentinel-1) global land cover training dataset, and the Replicable AI for Microplanning (Ramp) building footprint training datasets.

Radiant MLHub

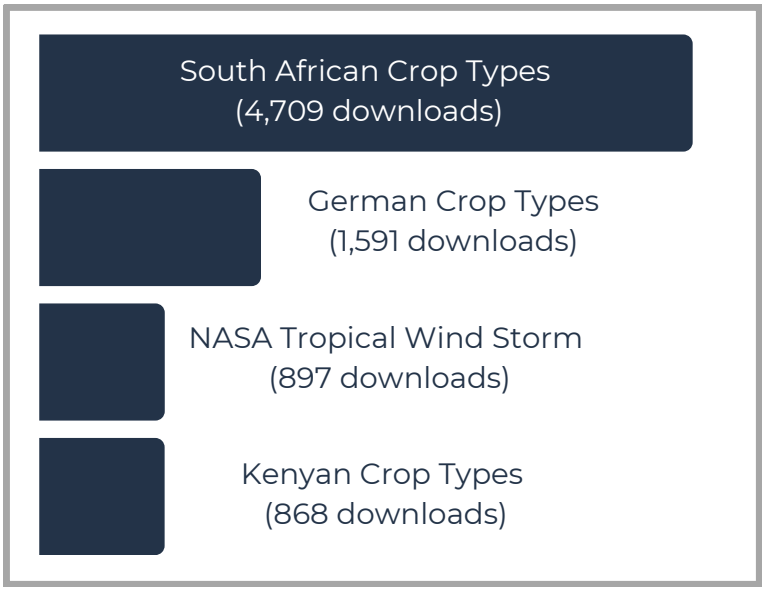
A cloud-based community library for geospatial machine learning projects. It provides the infrastructure that facilitates the publication of Earth observation training datasets and pre-trained machine learning models. The data available on Radiant MLHub are geographically diverse and support global missions like climate change, agriculture, and sustainable cities. Distributed with licenses that give anyone permission to access, use, and publish data and models.

Geospatial Training Datasets

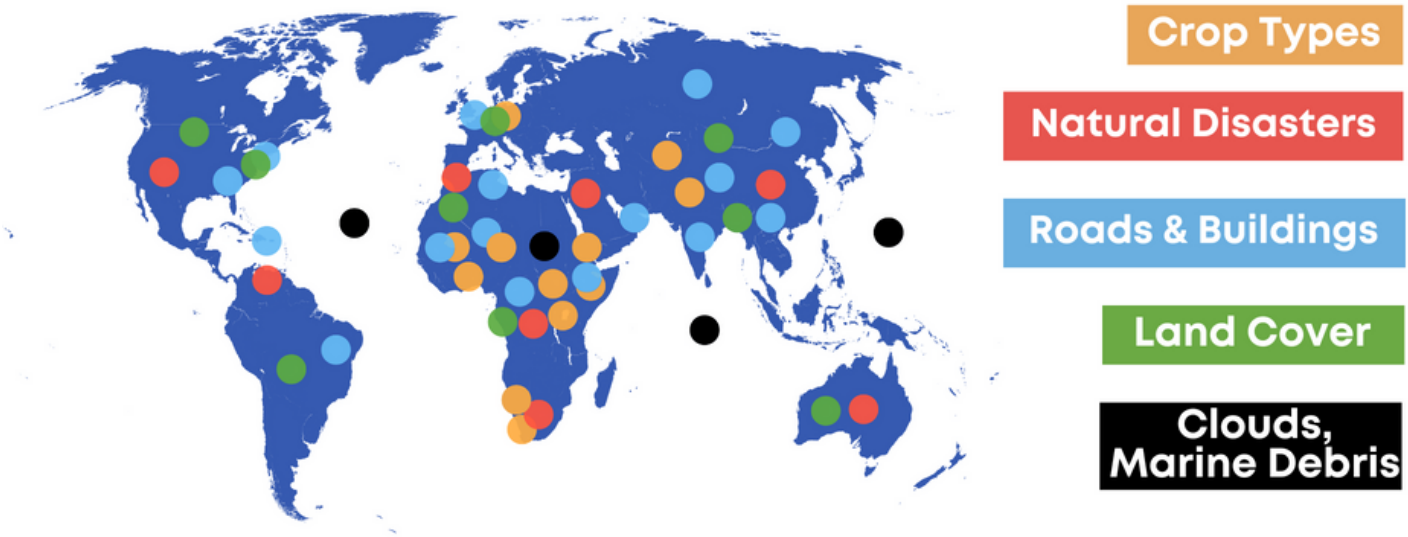
17,123
Data collections downloaded from Radiant MLHub.

65 high-quality training datasets

39 new datasets added in 2022

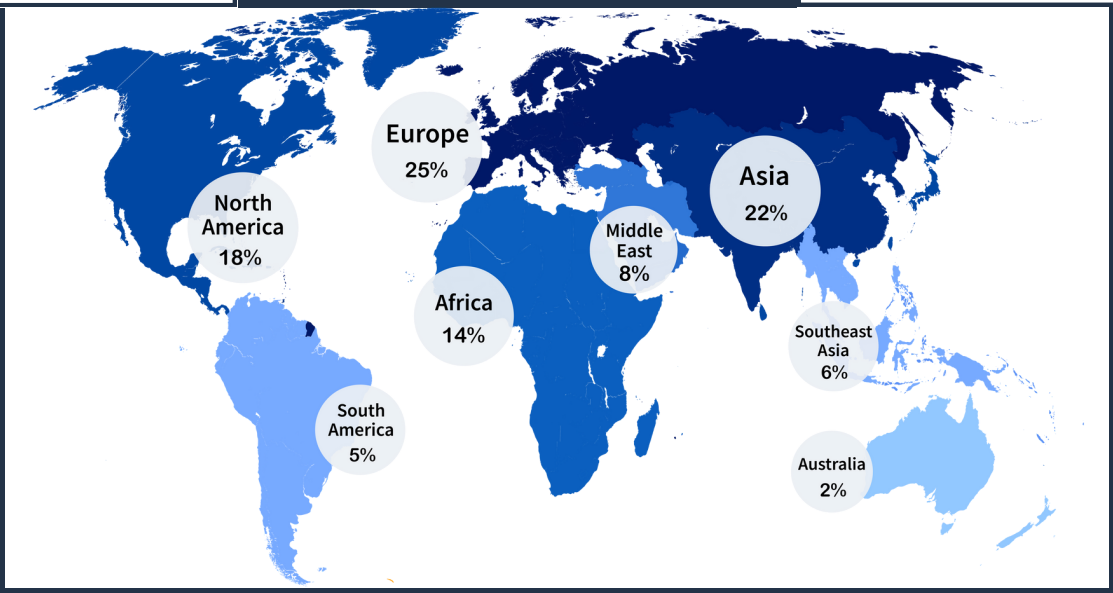


The top three training datasets downloaded from Radiant MLHub.



Available geospatial training datasets.

2,768	6,540	159
New users in 2022	Total users	Countries



Radiant MLHub users per geographical region.

"...it's incredible that Radiant MLHub offers free access to very well-structured datasets available for immediate work without complicated processing or analyzing. What I really like about [it] is the diversity in datasets, which fills the gap and improves the availability of data representative of developing countries."

MOHAMMAD ALASAWDAH

Earth observation and climate data science researcher, Germany

Data Product Highlight: *LandCoverNet*

LandCoverNet is a Human Annotated Global Multi-Satellite Training Dataset for Land Cover Classification divided into regions based on continents

- **586 million pixels** - covering Africa, Asia, Australia and Oceania, Europe, North America, and South America.
- **7 classes** - identifying water, natural bare ground, artificial bare ground, woody vegetation, cultivated vegetation, (semi) natural vegetation, and permanent snow/ice.
- **3 satellite missions** - imagery from Sentinel-1, Sentinel-2, and Landsat 8.
- **Endless mapping** - Enable scientists and practitioners to create high-resolution and up-to-date land cover maps.

Geospatial Machine Learning Models

Our model repository is a collection of machine learning models designed explicitly for geospatial data use. It includes a variety of models for tasks such as crop classification and semantic segmentation. The models in the repository are trained on large satellite imagery datasets and include metadata based on the SpatioTemporal Asset Catalog (STAC) specification that describes the training data associated with it and the model architecture. This helps organizations and individuals to discover the models quickly and to extract insights from the data. New models added to the repo in 2023 include:

Ramp Baseline Model for Building Footprint Segmentation

Deep learning model to detect buildings from satellite imagery and delineates the footprints in low-and-middle-income countries.

AgriFieldNet Model for Crop Detection from Satellite Imagery

Top machine learning model to classify crop types in agricultural fields across Northern India.

Weighted Tree-based Crop Classification Models for Imbalanced Datasets

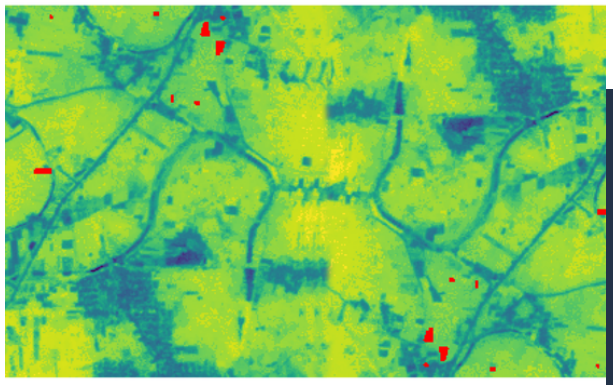
Ensembled weighted tree-based model to classify crop types in agricultural fields across Northern India.

Looking Further: A crop type classification model for fields

Model to classify crop types for each agricultural field based on the field as well as on its surroundings across Northern India.

A Spatio-Temporal Deep Learning-Based Crop Classification Model for Satellite Imagery

Model to classify fields by crop type during the growing season across western Kenya.



Data Competitions

Radiant Earth organizes and supports high-impact data competitions to encourage data scientists around the world to use satellite data and machine learning to solve global problems such as climate change and sustainable development. By curating and easing access to Earth data, and hosting data challenges, Radiant aims to spur innovation and collaboration in the field of Earth observation in support of better decisions for the planet. Radiant organized two challenges and supported one other.

$35,000
in cash prizes for data challenges in 2022

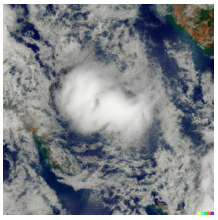
1,949
Data scientists from around the world participated

AGRIFIELDNET INDIA CHALLENGE

To classify crop types in agricultural fields across Northern India using multispectral observations from the Sentinel-2 satellite mission. The competition training dataset is generated by Radiant Earth, using the ground reference data collected and provided by IDinsight's Data on Demand team.

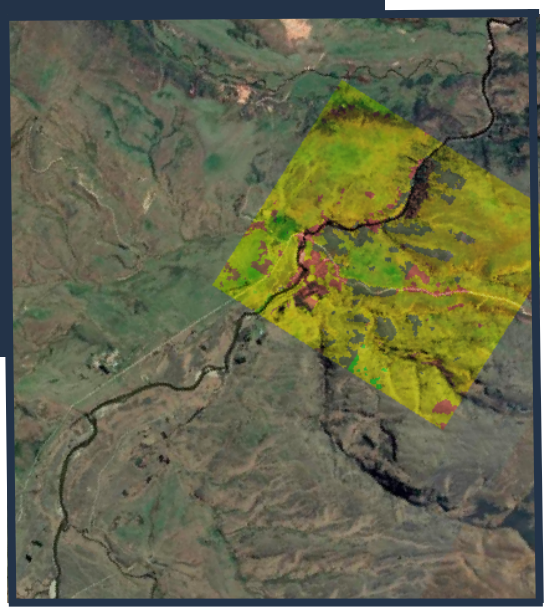
NASA HARVEST FIELD BOUNDARY DETECTION CHALLENGE

To classify crop field boundaries in Rwanda using multispectral observations collected by PlanetScope. The NASA Harvest Rwanda field boundary competition training dataset was curated by Radiant Earth. This challenge ends in 2023.



CLOUD COVER DETECTION CHALLENGE

In partnership with Microsoft AI for Earth, the goal of this challenge is to most accurately detect cloud cover in multispectral satellite imagery from Sentinel-2 satellite imagery. Radiant provided the training dataset, which was curated through a crowdsourcing contest in 2020 and validated by human annotators.



STAC SPEC

The SpatioTemporal Asset Catalog (STAC) specification is an open set of guidelines for describing and organizing Earth science data into a hierarchical structure, making it easier for discovery, access, and use. This standard is actively developed by a community of organizations and individuals. It reached stable version 1.0. in 2021.

Highlights for 2022 include:

- STAC API v.1.0.0-rc2 release (stable version expected in 2023)
- STAC and OGC API Features added for meteorological data
- STAC plug-in 1.0.1 for QGIS
- Azavea's Franklin supports the latest STAC API specification

STAC Ecosystem

Organizations supporting the STAC ecosystem.

STAC Browser

The STAC Browser is an open-source tool allowing users to search for geospatial data based on various criteria such as data type and location. Users can also view metadata and other information about the data, such as image resolution. The STAC Browser is essential as more and more organizations are publishing their geospatial data with STAC metadata. In 2022, the STAC Browser version 3 was released and implemented. This version is designed to support the STAC API, with additional features such as:

- Allowing to filter by temporal extent, spatial extent, Collection identifiers, and Item identifiers
- Pagination support for APIs, both in lists of Collections and Items
- Visualization of Asset metadata
- A new sidebar to quickly browse through the Catalogs, Collections, and Items
- Client-side support for cloud-optimized GeoTiffs (COGs)

PySTAC

PySTAC is a Python library working with the STAC specification. It allows developers to easily read and write STAC catalogs and items and provides a consistent way to access and manipulate the data and metadata in a STAC catalog. The PySTAC version released in 2022 makes it easier to navigate, updated many tutorials, and added support for the Storage Extension, and the filtering links by media type.

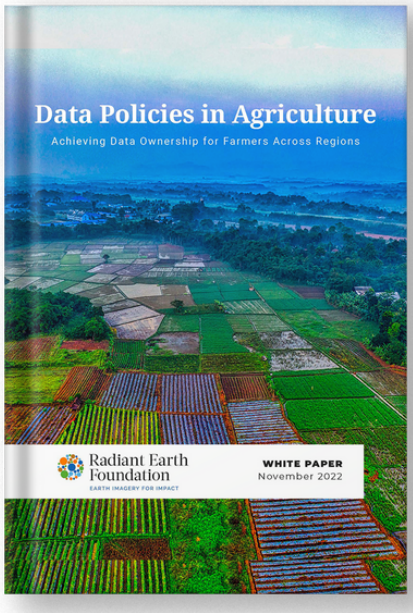
Organizations supporting the STAC ecosystem.

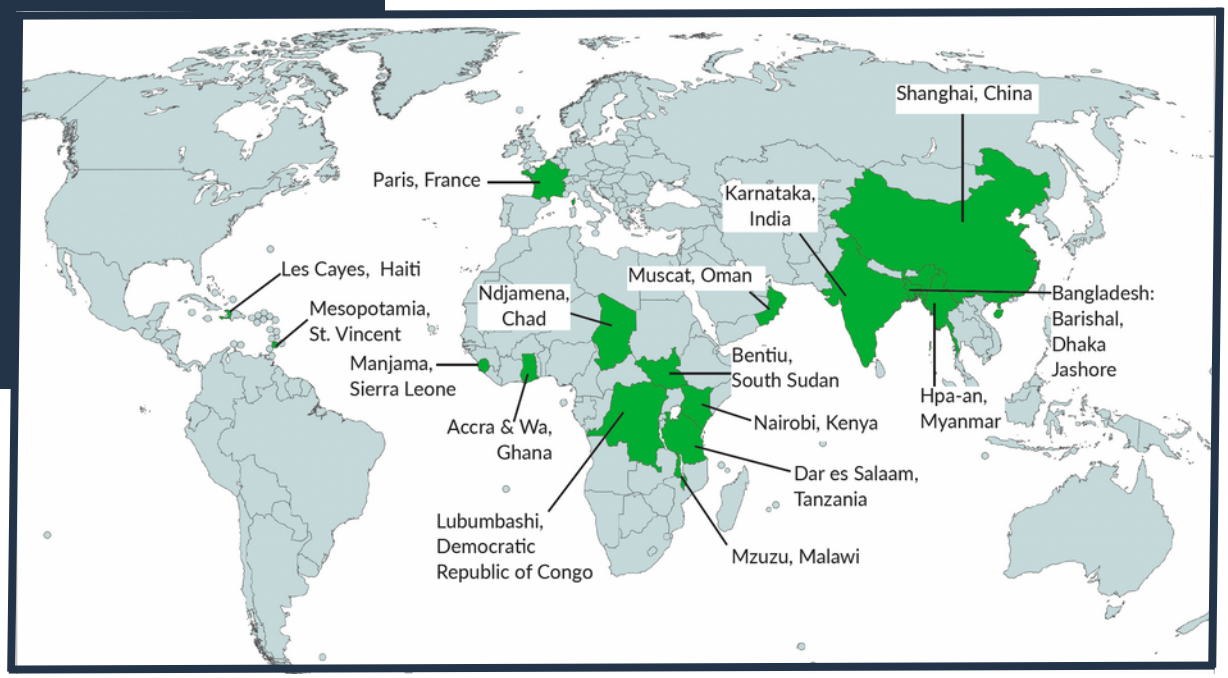
Partnerships

TETRA TECH

Enabling Crop Analytics at Scale

Radiant Earth continued its collaboration with the Enabling Crop Analytics at Scale (ECAAS) initiative that is managed by Tetra Tech with funding from the Bill & Melinda Gates Foundation. In 2022, this partnership resulted in a global data competition to classify crop types in agricultural fields across Northern India. In addition, Radiant published a white paper examining how high-income countries currently address smart farming's resulting data ownership issues and identifies ways to preemptively address relevant data ethics issues and policies for low- and middle-income countries.





Ramp building footprint training datasets around the world.

DevGlobal

Radiant Earth partnered with DevGlobal on the Replicable AI for Microplanning Open Buildings Model (Ramp) project to produce a collection of building footprint training datasets. Each dataset covers a specific region or city, as well as a diverse range of geographies.

A baseline model created to utilize the datasets has been dockerized and cataloged to promote its discoverability to a wider user group via Radiant MLHub.

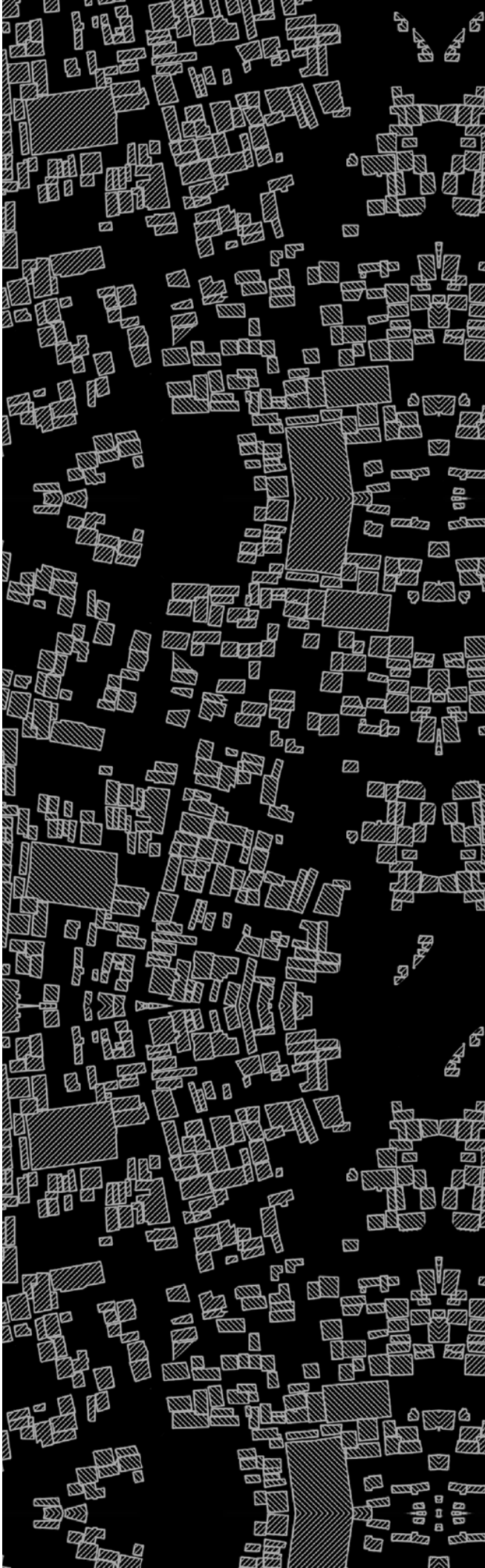
TaQadam B.O.T

Radiant Earth partnered with TaQaDam and B.O.T. to label the image chips for the Ramp project building footprint data. Twenty-two areas of interest around the world were identified and labeled equalling 101,312 chips.

The datasets are stored in Radiant MLHub, and each dataset includes accompanying metadata and documentation to improve the search and discovery of the data. These datasets are valuable in applications such as health, disaster response, and climate mitigation efforts.

Radiant Earth partnered with NASA Harvest to generate and curate a labeled training dataset for agricultural field boundary detection from multispectral satellite imagery. We generated the training dataset together with TaQadam through a team of annotators and curated it in a cloud format. A data competition was organized to crowdsource winning machine learning models based on this dataset. The competition ends in 2023 and the winning models will be published on the Radiant MLHub geospatial model repository.

Radiant Earth is guiding the Government of Vanuatu's National Statistics Office (VNSO) in effectively collecting ground data that can be combined with annotated labels to create a training dataset for a baseline coconut tree detection model.



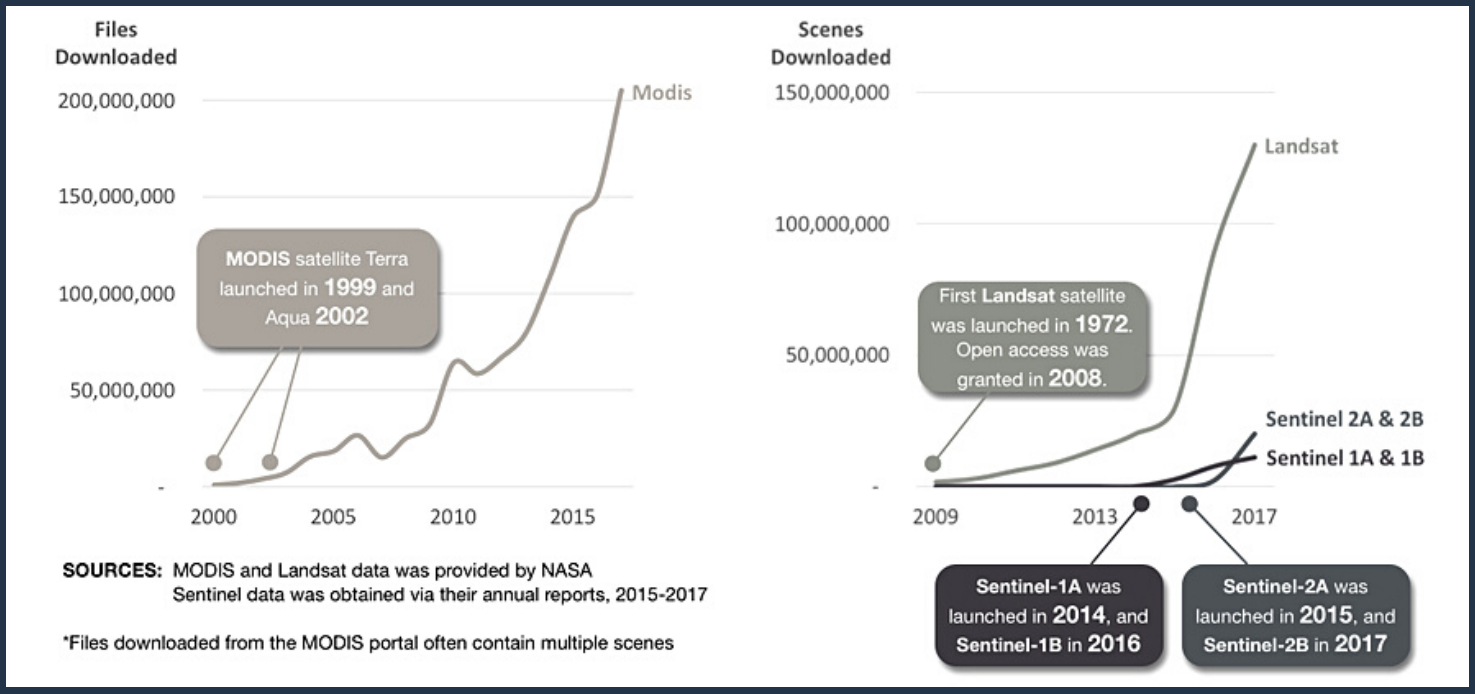
2023: Future Outlook and Growth Opportunities

Making more data more available to more people.

As we enter the Year of Open Science, we're giving some extra thought to what "open" means and why it matters. When Tim Berners-Lee [first proposed "the WorldWideWeb project,"] the sharing of information was core to his philosophy:

> The project started with the philosophy that much academic information should be freely available to anyone. It aims to allow information sharing within internationally dispersed teams, and the dissemination of information by support groups.

Berners-Lee's vision was quite prescient. The volume of data shared on the web today is astonishing. Indeed the volume and complexity of data that people can access over the Internet is so great that it's frequently overwhelming. It's quite common for organizations to launch a data portal or put something on GitHub and dust their hands off, satisfied that they've done their part to "be open". We enter 2023 recognizing that we'll never be "done" with open science, but that it will always be a continual process. The framework we use to guide us on this continual journey is to make *more data more available to more people*.



"What Radiant MLHub has created sets an example of accessible and usable datasets. Easy access means that we can use them for fast prototyping ideas and even evaluate our data's quality assurance processes."

VICTOR FARAGGI
Computer Science Engineer, Chile,

More Data

In one sense, "more data" isn't a problem. As more satellites are launched and new sensors are invented, we expect to have more than enough data to work with for a long time. But still, we want more. Radiant has historically focused on Earth imagery. In fact, our official legal name is "Open Imagery Network." We're already working on ways to expand Radiant ML Hub to support other types of data that our community needs, such as vector data, point clouds, and even tabular data.

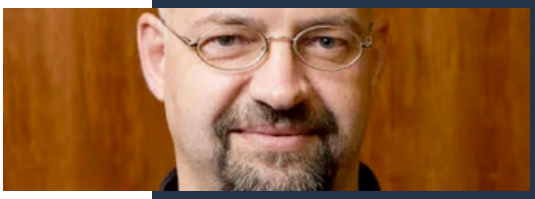
We also know that "data begets data". Members of our community frequently access open data and use it to produce their own derived products that they want to share with others. We will make it easier for our community members to publish their own data products on Radiant ML Hub in 2023.

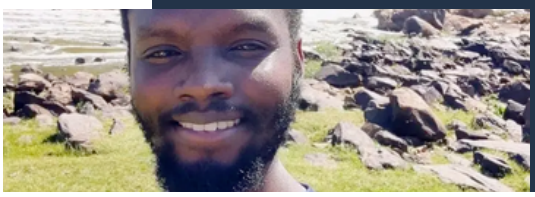
More Available

Making data more available is the thrust of our work on STAC and with the cloud-native geospatial community more broadly.

Is it really fair to say that data is "open" it takes a day to download it? Or if you need more than a terabyte of available storage to work with it? Or if there's no documentation for it in the language you speak? And even if you can use a bit of open data with open source code, does it matter if that open source code only runs on a computer that you can't afford? In short: making data available for download is not enough! Making data available requires thinking about the needs and capacity of your users.

There is a lot of exciting work already underway to make massive planetary-scale datasets in a variety of formats easier to work with. This year, we plan to apply some of the lessons we've learned from rapid adoption of the STAC specification to improve the availability of many other types of Earth science data.





More People

As we said at the beginning of this report, the future of our species depends on our ability to develop sustainable methods of sharing Earth data. I recognize that this is a big claim, so here's why I believe it's true.

Our ability to respond effectively to global crises is contingent on widespread access to trustworthy and accurate data about our planet. Whether we're confronting climate change or a pandemic, we need to make policy decisions at all levels of society, and shared access to data helps us make those decisions collaboratively. Perhaps more significantly, it allows us to assess the impact of those decisions collaboratively.

While policy changes at public and private institutions have made significant inroads in improving access to data, more needs to be done to make Earth data work for all of us. In particular, large environmental datasets have traditionally only been available to institutions with large computing and storage infrastructure, putting them out of reach of underrepresented communities that are likely to be most impacted by climate change.

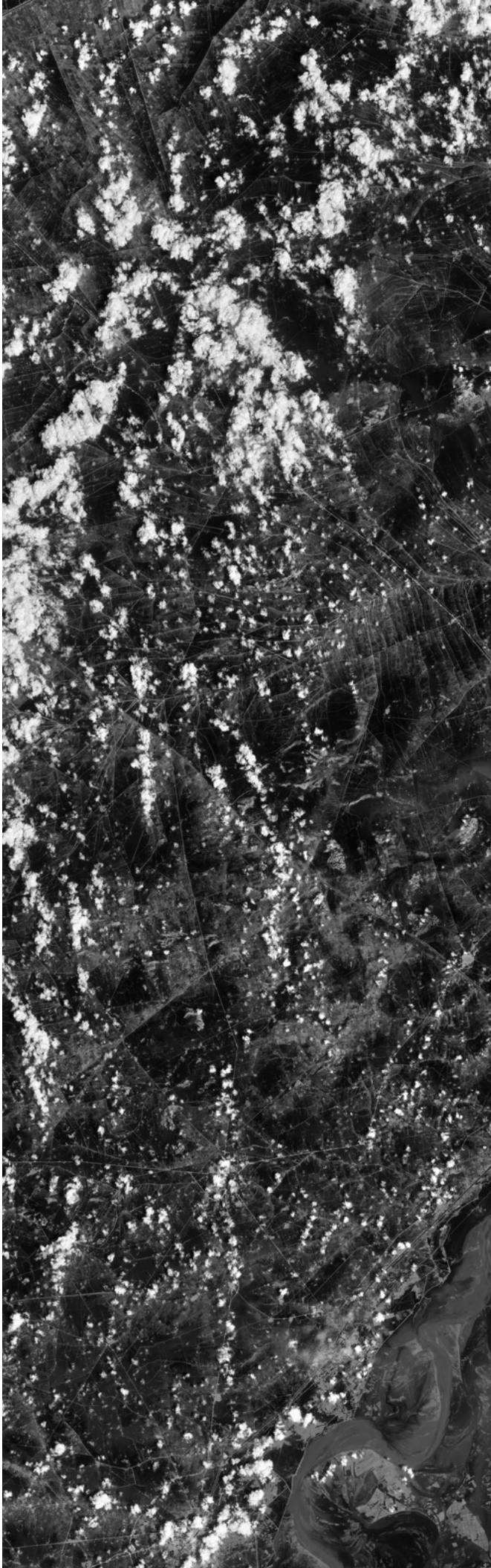
The impact of our work to make more data more available will be cut short if we aren't also deliberate about creating a larger and more diverse community of users who can work with the data to inform decisions about their communities. The peril of failing to do this is summarized well in a paper titled, "The co-development of models with expert judgement suppresses model diversity and underestimates risk," by Erica Thompson, who recently published the book, *Escape from Model Land*.

As she says in the paper:

> All “climate decisions” are also political decisions about which industries to support or restrain, which goals to prioritise, which voices to amplify or to ignore. All “climate decisions” are also moral decisions about whose lives matter; what species matter; what levels of risk we are prepared to live with and accept on behalf of future generations. In framing climate decisions as technical decisions primarily to be answered by modelling studies, it is imperative to consider the political and ethical dimensions of that framing and what interests are served by doing so.

Increasing the size and diversity of our audience is the most consequential aspect of our work and, we believe, the most difficult.

We certainly won't be able to do it on our own, so please get in touch if you're interested in helping us reach more people.



Meet the Team

Board of Directors

JERRY JOHNSTON
(BOARD CHAIR)

Specialist Leader, Deloitte

ANNE HALE MIGLARESE
(BOARD TREASURER)

Program Executive Officer,
Saildrone / Founder, Radiant Earth

CLAUDIA JUECH
(BOARD) SECRETARY

Vice President of Data and Society
Patrick J. McGovern Foundation

HAMED ALEMOHAMMAD
(BOARD ADVISOR)

Geospatial Director and Professor
Clark University

SIVES GOVENDER

Research Group Leader, Spatial
Information Systems, CSIR South
Africa

JED SUNDWALL

Executive Director, Radiant Earth
Foundation

PETER RABLEY

Co-founder and Managing
Partner, Place

KASS GREEN

President, Kass Green & Associates

Staff

JED SUNDWALL
Executive Director

HAMED ALEMOHAMMAD
Executive Director and Chief Data Scientist (former)

ABBA BARDE
Data Scientist (former)

ALEX RICE
Full Stack Developer (former)

AMBER MYERS
Business Development Manager (former)

DANIEL NWAEZE
Data Scientist

GREGORY ESSUMAN
Geospatial Data Engineer (former)

JON DUCKWORTH
Tech Lead (former)

KENDALL SMITH
Geospatial Data Engineer (former)

KEVIN BOOTH
Engineering Manager

LOUISA DIGGS
Marketing and Communications Manager

MAGDALENA BENZA
Chief Data Scientist (former)

MICHELLE ROBY
Geospatial Software Engineer

RENEE PIESCHKE
Technical Manager

Fellows

CINDY LIN
Cornell University, Policy Fellow

MATTHIAS MOHR
OpenEO and the University of Münster, Technology Fellow

CHRIS HOLMES
Planet, Technology Fellow

"[Radiant MLHub] It's a game-changer . . . The fact that most of the data is collected for Africa is amazing, as this is a continent that is generally underrepresented. This kind of data infrastructure is empowering for researchers in developing countries. It equalises the playing field and allows us to participate in international research and derive solutions applicable to our own countries."

RENATA THIEDE
Ph.D. Candidate, University of Pretoria, South Africa

Funders and Sponsors

Programs and Operations

Data Competitions

- University of Maryland, NASA Harvest
- USDA-Foreign Agricultural Service
- USAID
- IDInsight
- Microsoft AI for Earth

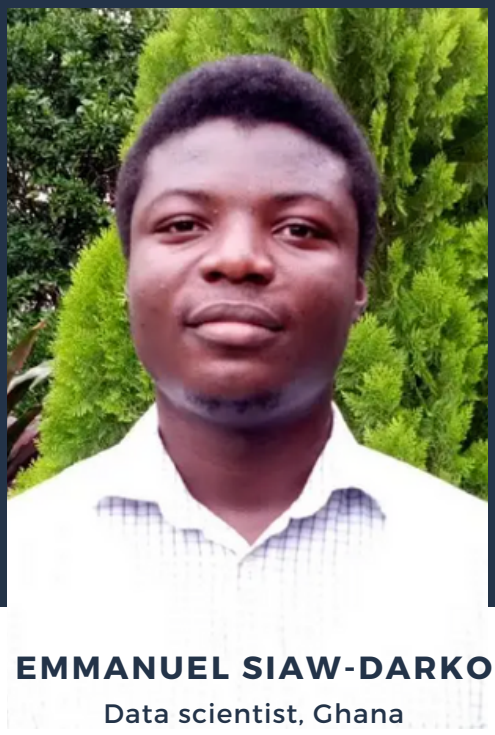
SpatioTemporal Asset Catalog

- Arturo
- Development Seed
- Digital Earth Africa
- Element 84
- Geoscience Australia
- Hydrosat
- Microsoft AI for Earth
- Planet
- Radiant Earth Foundation
- Sparkgeo
- Toitū Te Whenua Land Information New Zealand

Collaborators

- AI4EO
- Azavea
- DevGlobal
- DrivenData
- Endeva
- IDinsights
- Lacuna Fund
- Let's B.O.T.
- Locus Charter
- NASA IMPACT
- PLACE
- Planet
- RCMRD
- Sinergise
- TaQadam
- TetraTech
- The Munich Technical University (TUM) & the German Aerospace Center (DLR)
- Zindi

"[Radiant MLHub] is a great tool for developing local solutions. Having access to open data that I can use to experiment with solutions for the problems in my community is essential. The settlement that I live in experiences a lot of floods. As a data scientist, I feel compelled to help fix the problem. . . An open data infrastructure such as Radiant allows me to do just that."



EMMANUEL SIAW-DARKO
Data scientist, Ghana

Radiant Earth Foundation

We thank you for your continued support in our efforts to to make more data more available to more people.

Contact

Radiant Earth Foundation
740 15th St NW, Suite 900,
Washington DC 20005
202-596-3603

www.radiant.earth
hello@radiant.earth
@ourradiantearth