

Home / Search For Events / GEO-OPEN-HACK-2024: Big Geospatial Data Hackathon With Open Infrastructure and Tools

GEO-OPEN-HACK-2024: Big Geospatial Data Hackathon with Open Infrastructure and Tools



Milutin Milenkovic
Research Scholar (NODES)

Event details

📅 24 - 28 Jun 2024

📍 Laxenburg, Austria


[Advancing Systems An...](#) [Novel Data Ecosystem...](#) [Open-Earth-Monitor Cy...](#) [Summer schools](#)

GEO-OPEN-HACK-2024 is a comprehensive and informative event designed for advanced geo-coders to explore various open tools and approaches for upscaling geospatial analysis on open High-Performance Computing (HPC) infrastructure.

The event is organised by the [International Institute of Applied Systems Analysis \(IIASA\)](#) in collaboration with [Spatial Ecology](#). This hackathon delves into advanced cutting-edge open techniques, tools, and best practices for efficiently handling and processing vast amounts of geospatial data. Participants will gain hands-on experience in leveraging HPC resources and geo-tools for tasks such as geospatial data preprocessing, spatial modeling and analytics, and visualization.

Hackathon highlights

- **Introduction to Big Geospatial Data:** Understanding the challenges and opportunities presented by large-scale geospatial datasets.
- **High-Performance Computing Basics:** Familiarization with HPC systems, queuing system, parallel processing, and optimization techniques
- **Open Tools and Workflows:** Techniques and tools for geospatial data processing and spatial analytics for applications like remote sensing, GIS, and environmental change monitoring.
- **Modern Geo-analytics:** Exploring emerging trends and technologies in the field, such as machine learning and cloud-based geospatial analytics and visualization.
- **Parallel Computing:** Harnessing the power of parallel and distributed computing for speed and efficiency for geospatial analysis.
- **Performance Tuning:** Strategies to optimize ML models and workflows for HPC environments.
- **Case Studies:** Real-world examples of successful big geospatial data projects on HPC systems.
- **Scalability and Big Data Challenges:** Addressing issues related to data volume, velocity, variety, and veracity in geospatial analysis.

| Agenda GEO-OPEN-HACK 2024 | | | | | | |
|---|--|---|--|--|---|---|
| Time | Sunday 23-Jun-24 | Monday 24-Jun-24 | Tuesday 25-Jun-24 | Wednesday 26-Jun-24 | Thursday 27-Jun-24 | Friday 28-Jun-24 |
| 08.00-08.30 | | | | | Canadian Pancake Breakfast IIASA Schlossrestaurant | |
| 08.30-09.00 | | Get together IIASA WODAK meeting room | Get together IIASA WODAK meeting room | Get together IIASA WODAK meeting room | | Get together IIASA WODAK meeting room |
| 09.00-09.30 | | Keynote Raymond Oonk, SURF | Keynote Tom Hengl, OGH | Keynote Jan Verbesselt & Dainius Masiliunas | Keynote Edzer Pebesma, Uni Münster | Keynote Wolfgang Wagner, TU Wien & EODC |
| 09.30-10.00 | | Giuseppe Amatiulli Geo-Processing with HPC | Pieter Kempeneers Geo-python with HPC | Keynote Patrick Griffiths, ESA | Claus Michele & Valentina Premier openEO | Anne Fouilloux & Tina Odaka Pangeo |
| 10.00-12.30 | | | | Antonio Fonseca ML with HPC | | |
| 12.30-13.30 | | Lunch break IIASA Gvishianhi meeting room | Lunch break IIASA Gvishianhi meeting room | Lunch break IIASA Gvishianhi meeting room | Lunch break IIASA Gvishianhi meeting room | Lunch break IIASA Gvishianhi meeting room |
| 13.30-14.00 | | Big Data Story Giuseppe Amatiulli | Big Data Story Milutin Milenković & Florian Hofhansl | Big Data Story Yu-Feng Ho & Johannes Heisig | Big Data Story Valentina Premier | Big Data Story Leandro Parente |
| 14.00-17.30 | | Working with your data | Park Visit Laxenburg Park --- START: 16.00h Working with your data | Working with your data | Working with your data | Open floor sessions OEMC and other open initiatives. OEMC data platform & feedback. Where to implement my big data problem? |
| 18.00-21.00 | Ice Breaker Event Rest. Stöckl im Park Prinz-Eugen-Strasse 25 1030 Vienna | | | Social Event IIASA Schlossrestaurant Herzog Albrecht-Strasse 1 2361 Laxenburg | | |
| <div> This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No.101059548</div> <div>© IIASA</div> | | | | | | |

The hackathon will start with an icebreaker/get-together event on Sunday evening (18.00-21.00h). On Wednesday evening, a social event will be organized. For more details on topics covered in the morning sessions, please click below:

Monday: Geo-Processing with HPC

- VRT format for tile splitting and tile re-aggregation
- Embedding Python and R in Bash.
- HPC architecture and queue systems
- Geo-Processing with Slurm queue system
 - Single and multiple core processing using a simple job
 - Single and multiple core processing using many simple jobs
 - Single and multiple core processing using job-array
- Working with Slurm using GRASS
 - Building MAPSET on the RAM
 - Using GRASS commands that incorporate multi core processing

Tuesday: Geo-python with HPC

- Geospatial data processing with pyjeo
 - Introduction to the pyjeo data model
 - Combining pyjeo with other open source libraries
 - Numpy
 - xarray
 - Machine learning applied to geospatial data (optional)
 - Feature extraction based on labeled data
 - Model training based on labeled data
 - Model prediction on raster dataset
 - Access data cubes from a STAC compliant catalog (optional)
- Upscaling workflows: high throughput computing (HTC)
 - Processing on a cluster with tiled images
 - Parallel and distributed computing with Dask

Wednesday: ML with HPC

- ML-python: Sckit-learn/Pytorch on a supercomputer/HPC (MLPClassifier, MLPRegressor, etc.)
 - Training a CNN model on the cluster.
 - Multicore processing and GPU usage.
- Using and finetuning large models for satellite image datasets via [HuggingFace](#)
 - Intro to Foundation Models
 - Training computer vision models via Huggingface: [ViT](#), [MAE](#), [Prithvi](#) ([IBM/NASA model](#))
 - Potentially use the Satellite image dataset

Thursday: openEO

OpenEO is a standardized API for cloud computing in EO. This training will teach the concepts behind openEO and how to use it to carry out a typical EO workflow on a cloud platform.

- The concepts of openEO
 - Cloud computing in EO - why standards are needed
 - Where can you find cloud platforms that support openEO
- OpenEO in action
 - Learn how to use openEO for cloud processing
 - Create your own openEO workflow to solve a typical EO question.

Friday: Pangeo

The Pangeo training will be delivered as a mix of presentations, hands-on exercises, group activities and group discussions (including showcasing successful projects based on the Pangeo ecosystem).

- Unlocking Pangeo for Geospatial Data Science
 - Introduction to the Pangeo and its significance in big geospatial analysis
 - Different ways to setting up a Pangeo environment for geospatial applications
 - Accessing and loading large geospatial datasets using Pangeo
 - Understanding data storage options & best practices
- Scalable geospatial analysis with Pangeo
 - Leveraging Pangeo for scalable geospatial analysis
 - Exploring parallel computing capabilities of Pangeo using Dask and Xarray

- Applying machine learning models to geospatial data within the Pangeo environment

Less

Keynotes

Monday, 9.00-9.30 am

- HPC and current European developments: European data spaces, EuroHPC, and DestinE; *Raymond Oonk, SURF (Collaborating University Computing Facilities), The Netherlands*

Tuesday, 9.00-9.30 am

- Open Earth Monitor Cyberinfrastructure (OEMC): modern open source solutions for HPC and open science; *Tom Hengl, OpenGeoHub Foundation, The Netherlands*

Wednesday, 9.00-9.30 am

- Evolution of BFAST functions for characterizing land change using satellite image time series; *Jan Verbesselt, Belgian Science Policy Office, Belgium & Dainius Masiliunas, Wageningen University, The Netherlands*

Wednesday, 9.30-10.00 am (**remote**)

- Streamlining operationalisation of EO based solutions for the Green Transition using interoperability best-practices; *Patrick Griffiths, European Space Agency, Italy*

Thursday, 9.00-9.30 am (**remote**)

- openEO: open science for Earth Observation compute platforms; *Prof. Edzer Pebesma, Münster University, Germany*

Friday, 9.00-9.30 am

- Global scale analysis of Sentinel-1 satellite data at the Earth Observation Data Centre; *Prof. Wolfgang Wagner TU Vienna & Earth Observation Data Centre GmbH, Austria*

Big Data Stories

Monday, 13.30-14.00

- Global Flow Discharge: Modeling Upper Stream Contribution with a ML Feamework; *Giuseppe Amatulli, Yale University & Spatial Ecology, USA*

Tuesday, 13.30-14.00

- Citizen Science Mobile App for Measuring Trees; *Milutin Milenković & Florian Hofhansl, The International Institute for Applied Systems Analysis, Austria*

Wednesday, 13.30-14.00

- Accessing Big Vector Data on the cloud using Arrow Parquet; *Yu-Fend Ho, OpenGeoHub Foundation, The Netherlands & Johannes Heisig, Münster University, Germany*

Thursday, 13.30-14.00

- Leveraging Satellite Data for High-Resolution Snow Monitoring: Scaling with openEO; *Valentina Premier, Eurac Research, Italy*

Friday, 13.30-14.00

- Optimizing Global Time Series ML Predictions: From Modeling to Production Deployment; *Leandro Parente, OpenGeoHub Foundation, The Netherlands*

Hackathon requirements

This is an **advanced-level** hackathon, ideal for early-career researchers, scientists, and professionals interested in unlocking the full potential of big geospatial data by harnessing the computational power of open HPC systems. Participants will leave with valuable insights and practical skills to tackle their geospatial challenges at scale.

The hackathon is aimed at individuals with masters or doctoral qualifications. It will provide them an opportunity to scale-up their own spatio-temporal modeling and data analysis projects. Hackathon participants should have intermediate bash and Python skills, basic ML-python knowledge, and a strong desire to learn command line tools for massive geo data processes. R users are also welcome if they feel comfortable with command line operations in bash and python environments.

Basic concepts of GIS, such as familiarity with rasters/vectors, overlays, buffering etc, and basics of statistics, such as mean standard, deviation, residuals, as well as python and bash syntax will be assumed as given. We will share tips and tricks on massive data processing, and therefore a strong bash and Python base will be essential grasping nuanced concepts that can boost your geo-analysis.

A developer committee will evaluate the pre-registration assessment and select 25 applicants on **one** of the following criteria:

1. Applicant's git repository (Github, Gitlab, Bitbucket) for an evaluation. The following will be assessed:

- Use of Python code and bash (including GDAL) for txt and geo-data analysis. Laptop-executed scripts are sufficient, i.e. no need to demonstrate code for cluster processes.
 - Use of scikit-learn python (or similar library) for ML applications, in other words we want people that think in a multidimensional way.
2. Any evidence of bash and Python used by advanced R programmers, e.g. Git repositories, coding samples or relevant course certificates etc. Moreover fundamentals of ML are also needed.
 3. Completion of the “[Geocomputation & Machine Learning for environmental applications](#)” course and workshop organized by [Spatial Ecology](#).

Registration information

- On-site registration fee 200 Euro
- Online registration fee 50 Euro

This is a not-for-profit event, and registration fees serve only to cover hospitality costs (social events, lunches, coffee breaks, etc.). **Accommodation and travel costs are the responsibility of the participants.**

Online participants will be able to ask questions and receive live assistance by a course instructor. However, troubleshooting will be limited due to logistical constraints. Lectures will be recorded and made available for asynchronous viewing, allowing participants in distant time zones to participate as well. Links to the recordings will be shared only with those who register as online attendees.

To register, all participants must:

1. Complete the skills assessment and the pre-registration form **here**. - **closed**
2. Wait for our approval before proceeding with fee payment.
3. Pay the registration fee.

You will be considered registered only when we have received your completed pre-registration form, and we have sent you a confirmation email, and the registration fee has been paid in full.

The pre-registration form will be open until **March 31. 2024**.

Payment information: the payment options are explained [here](#). Please wait first to receive the approval (see step 2 above). The deadline to pay the registration fees is until **May 15. 2024**.

Refund policy: A written request for registration cancellation must be e-mailed to the hackathon organizers. If registration cancellation is requested by April 30. 2024, the full registration fee will be refunded. If a cancellation request is submitted after **April 30. 2024**, we will retain 80% of the fees unless a candidate from the waiting list is able to replace the withdrawing candidate.

Visa requirements: On-site participants are responsible for ensuring they have the correct documentation to enter Austria. Please check your visa requirements and visa issuance waiting times, ensuring that you are able to travel to Austria on a Schengen visa in time for June 2024. Check [here](#) for countries that need a Schengen visa. Please note that the hackathon organizers are not authorized to assist with the VISA process beyond providing the Invitation Letter.

Logistics

The hackathon logistics information can be found in this [document](#).



Teachers and supervisors

[Giuseppe Amatiulli](#) - Yale University, US; Spatial Ecology, UK

[Antonio Fonseca](#) - Yale University, US; Spatial Ecology, UK

[Anne Fouilloux](#) - SIMULA, Norway

[Tina Odaka](#) - CNRS-Ifremer-IRD-UBO, France

[Pieter Kempeneers](#) - JRC, Italy

[Francesco P. Lovergine](#) - CNR-IREA, Italy

[Michele Claus](#) - EURAC, Italy

[Valentina Premier](#) - EURAC, Italy

[Tom Hengel](#) - OpenGeoHub, The Netherlands

[Leandro Parente](#) - OpenGeoHub, The Netherlands

[Yu-Feng Ho](#) - OpenGeoHub, The Netherlands

[Dainius Masiliunas](#) - WUR, The Netherlands

[Johannes Heisig](#) - Münster University, Germany

[Milutin Milenković](#) - IIASA, Austria

Scientific Advisory Panel

- [Milutin Milenković](#) - IIASA, Austria
- [Giuseppe Amatulli](#) - Yale University, US; Spatial Ecology, UK

- [Pieter Kempeneers](#) - JRC, Italy
- [Antonio Fonseca](#) - Yale University, US; Spatial Ecology, UK

Steering Committee

- [Milutin Milenković](#) - IIASA, Austria
- [Giuseppe Amatulli](#) - Yale University, US; Spatial Ecology, UK
- [Tushar Sethi](#) - Spatial Ecology, UK

Acknowledgments

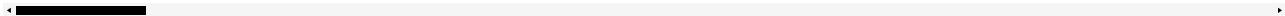
GEO-OPEN-HACK-2024 is an initiative under the Open-Earth-Monitor Cyberinfrastructure (OEMC) project aiming to lower the barrier and transfer knowledge to users dealing with big geospatial data analytics. The OEMC has received funding from the European Union's Horizon Europe research and innovation programme ([grant agreement No. 101059548](#)). GEO-OPEN-HACK-2024 is also an initiative in the framework of NSF-funded POSE project TI-2303651: Growing GRASS OSE for Worldwide Access to Multidisciplinary Geospatial Analytics.



Upcoming Events

01 June 2024

Shaping the Future of Rural Africa: Online Training to Navigate the Water-Energy-Land Nexus



Subscribe

Keep up to date with our latest news, events, and other IIASA activities. [Sign up](#).

[Media](#)

[Employment](#)

[How to find us](#)

[Whistleblowing Reporting System](#)

[Cookies](#)

[Terms of use](#)

Follow us on:



International Institute for Applied Systems Analysis (IIASA) | Schlossplatz 1 | A-2361 Laxenburg, Austria

T: (+43 2236) 807 0

info@iiasa.ac.at