



Eixo 4 : Database and Artificial Intelligence

Data analysis and integration: Galaxy for the Earth System

- Marie Jossé CNRS / Data Terra Research Infrastructure
- Alessandro Rizzo 2 IRD / Data Terra Research Infrastructure

Training duration: 2 days

O Sistema Terrestre lida com uma grande quantidade de dados heterogêneos que são essenciais para o avanço da ciência e para enfrentar os urgentes desafios do desenvolvimento. Além disso, trata-se de um sistema complexo e dinâmico, que abrange as interações entre a atmosfera, os oceanos, a terra e a biosfera. Para enfrentar o desafio de estudar esses componentes interconectados — especialmente no contexto das mudanças climáticas — o Galaxy oferece uma plataforma web aberta, abrangente e sustentável, voltada para a ciência do Sistema Terrestre. Trata-se de um conjunto de ferramentas de TI flexível para pesquisas multidisciplinares e interdisciplinares, fornecendo uma ampla variedade de recursos para visualização, processamento e análise de dados em áreas como oceanografia, ciências atmosféricas e monitoramento terrestre.

Por sua concepção, o Galaxy garante reprodutibilidade e transparência ao registrar metadados, versões de ferramentas e parâmetros, permitindo que pesquisadores compartilhem e publiquem resultados, fluxos de trabalho e visualizações em total alinhamento com as estratégias de Ciência Aberta e os princípios FAIR (Localizável, Acessível, Interoperável e Reutilizável). A plataforma apoia a construção de fluxos de trabalho padronizados e modulares que podem integrar dados e metodologias heterogêneos entre diferentes domínios científicos.

Como parte do projeto europeu financiado FAIR-EASE¹ — “Construindo serviços interoperáveis em Ciências da Terra e ambientais” — uma instância dedicada, o *Galaxy Earth System* (earth-system.usegalaxy.eu), foi implantada para fornecer acesso personalizado a ferramentas, modelos e conjuntos de dados das Ciências da Terra, incluindo dados in situ, de satélite e saídas de modelos. Essa instância está totalmente integrada à Galaxy Training Network

(training.galaxyproject.org), que oferece um extenso catálogo de tutoriais revisados e desenvolvidos pela comunidade, ajudando pesquisadores e estudantes a aprenderem a usar a plataforma de forma eficiente.

Durante esta sessão, demonstraremos como o Galaxy Earth System possibilita fluxos de trabalho analíticos por meio de exemplos práticos, como análises de dados biogeoquímicos oceânicos, observações atmosféricas de vulcões e conjuntos de dados sobre diversidade terrestre. Nosso objetivo é demonstrar como diferentes tipos de dados, de várias disciplinas, podem ser analisados e visualizados juntos em um único Ambiente Virtual de Pesquisa, fazendo do Galaxy uma peça fundamental para uma ciência do Sistema Terrestre aberta, colaborativa e orientada pelos princípios FAIR.

Esta sessão de treinamento de dois dias, organizada pela **Data Terra**, a Infraestrutura Nacional Francesa de Pesquisa dedicada às Ciências da Terra, faz parte da conferência ICID III sobre desertificação e convida os participantes a descobrir e explorar o potencial da plataforma Galaxy Earth System — um ambiente virtual de pesquisa (VRE) projetado para democratizar o acesso a dados e ferramentas das Ciências da Terra, promovendo práticas científicas alinhadas aos princípios FAIR.

Desenvolvido dentro do projeto FAIR-EASE, o Galaxy Earth System permite a integração e análise de dados heterogêneos provenientes das ciências ambientais, tornando-se uma ferramenta ideal para pesquisadores que enfrentam questões complexas como degradação do solo, desertificação e monitoramento de ecossistemas florestais.

Requisitos

Esta sessão é aberta a pesquisadores, analistas de dados, estudantes e profissionais técnicos interessados em Ciências do Sistema Terrestre e pesquisa orientada por dados. Não é necessário conhecimento prévio do Galaxy. Os participantes devem levar um notebook com conexão estável à internet, pois todas as atividades práticas ocorrerão por meio de um navegador web, utilizando a plataforma Galaxy Earth System (nenhuma instalação é necessária).

Para a trilha avançada “traga seus próprios dados ou scripts”, no segundo dia, recomenda-se alguma familiaridade com Python, R ou shell script, assim como acesso aos próprios conjuntos de dados ou códigos. Os participantes também são incentivados a criar uma conta de usuário antecipadamente em:

👉 <https://earth-system.usegalaxy.eu>

Isso permitirá salvar dados, compartilhar fluxos de trabalho e acessar ferramentas.

Objetivos do Treinamento

- Apresentar a infraestrutura de pesquisa **Data Terra** e sua integração com o ecossistema **EOSC - European Open Science Cloud** (Ciência Aberta e princípios FAIR);
- Introduzir o **Galaxy** como uma plataforma amigável e de código aberto para projetar, executar e compartilhar fluxos de trabalho compatíveis com os princípios FAIR usando dados do Sistema Terrestre;
- Oferecer uma experiência prática tanto do ponto de vista do **usuário** quanto do **desenvolvedor** da plataforma Galaxy Earth System:
 - Orientar os participantes em exemplos de fluxos de trabalho de análise de dados (ex.: degradação do solo, diversidade florestal), mostrando como adaptá-los às suas próprias questões de pesquisa;
 - Ajudar os participantes a importar seus próprios dados, desenvolver suas próprias ferramentas, e publicar e compartilhar seus resultados e métodos.

Dia 1 - Manhã

- Apresentação geral da infraestrutura de pesquisa **Data Terra**
- Serviços oferecidos pela Data Terra (com foco em **degradação do solo**)
- Projeto **FAIR-EASE** e **EOSC**: promovendo Ciência Aberta e novos serviços digitais
- Apresentação geral do **Galaxy**

Dia 1 - Tarde

- **Lado do Usuário no Galaxy**
 - Apresentação do sistema: como a interface funciona e como ela incorpora os princípios FAIR
 - Como construir fluxos de trabalho, analisar dados e seguir treinamentos (exemplo: fluxo de trabalho sobre degradação do solo e diversidade florestal)
 - Compartilhar e publicar resultados

-
- **Lado do Desenvolvedor no Galaxy**
 - Como construir uma ferramenta
 - Como publicá-la

Dia 2 - Manhã

Dois modos de trabalho em paralelo:

- Traga seus próprios dados e siga um treinamento para aprender a usar o Galaxy em exemplos específicos e desenvolver habilidades práticas
- Traga suas próprias análises ou scripts para analisar dados e construir ferramentas no Galaxy, criando seu próprio fluxo de trabalho

Dia 2 - Tarde

Continuação ou finalização do que foi iniciado pela manhã

Encerramento com feedback: o que foi feito, o que precisa ser melhorado, o que pode ser continuado, etc.

Eixo 4 : Base de Dados e Inteligência Artificial

Análise e integração de dados: Galaxy para o Sistema Terrestre

- Marie Jossé CNRS / Data Terra Research Infrastructure
- Alessandro Rizzo 2 IRD / Data Terra Research Infrastructure

Training duration: 2 days

The Earth System is dealing with a large amount of heterogeneous data that is critical for advancing science and meeting the urgent development challenges. In addition, it is a complex and dynamic system encompassing the interactions between the atmosphere, oceans, land, and biosphere. To address the challenge of studying such interconnected components—particularly in the context of climate change—Galaxy offers an open, comprehensive, and sustainable web-based platform tailored for Earth System science. It serves as a flexible IT toolkit for multidisciplinary and interdisciplinary research, providing a wide array of tools for data visualization, processing, and analysis across fields like oceanography, atmospheric sciences, and land monitoring. By design, Galaxy ensures reproducibility and transparency by capturing metadata, tool versions, and parameters, allowing researchers to share and publish results, workflows, and visualisations in full alignment with Open Science strategies and FAIR principles (Findable, Accessible, Interoperable, Reusable). It supports the construction of standardised, modular workflows that can integrate heterogeneous data and methodologies across scientific domains.

As part of the EU funded project FAIR-EASE¹ - “*Building interoperable Earth Science and environmental services*”, a dedicated instance—Galaxy Earth System (earth-system.usegalaxy.eu)—has been deployed to provide tailored access to Earth science tools, models, and datasets, including in-situ, satellite, and model outputs. It is fully integrated with the Galaxy Training Network (training.galaxyproject.org), which offers a vast catalogue of reviewed tutorials developed by the community, helping researchers and students to learn how to use the platform efficiently. During this session, we will showcase how Galaxy Earth System enables analytical workflows through practical examples such as analysis on biogeochemical ocean data, atmospheric volcanic observations, and land diversity datasets. Our aim is to demonstrate how multiple data types, from various disciplines, can be analysed and visualised together within a single Virtual Research Environment, making Galaxy a cornerstone for open, collaborative, and FAIR Earth system science.

This two-day training session, organised by Data Terra, the French national Research Infrastructure dedicated to the Earth Sciences, as part of the ICID III conference on desertification, invites participants to discover and explore the power of the Galaxy Earth System platform — a virtual research environment (VRE) designed to democratise access to Earth system data and tools, while promoting FAIR scientific practices.

Developed within the FAIR-EASE project, Galaxy Earth System enables integration and analysis of heterogeneous data from the environmental sciences, making it an ideal tool for researchers tackling complex issues like land degradation, desertification, and forest ecosystem monitoring.

Requirements

This session is open to researchers, data analysts, students, and technical staff interested in Earth system sciences and data-driven research. No prior knowledge of Galaxy is required. Participants should bring a laptop with a stable internet connection, as all hands-on activities will take place through a web browser on the Galaxy Earth System platform (no installation needed). For the advanced “bring your own data or scripts” track on Day 2, some familiarity with Python, R, or shell scripting is recommended, as well as access to your own datasets or code. Participants are also encouraged to create a user account in advance on <https://earth-system.usegalaxy.eu> to enable data saving, workflow sharing, and tool access.

Training Objectives

- Present the Data Terra research infrastructure and its integration with the EOSC - European Open Science Cloud, ecosystem (Open Science and FAIR principles).
- Introduce Galaxy as a user-friendly, open-source platform to design, execute and share FAIR-compliant workflows on Earth system data.
- Provide hands-on experience on both the user and developer sides of Galaxy Earth System:
 - Guide participants through examples of data analysis workflows (e.g. land degradation, forest diversity), and how to adapt them to their own research questions.
 - Help participants learn to bring their own data, develop their own tools, and publish and share their results and methods.

Day 1 Morning:

- General presentation of Data Terra RI
- Services of Data Terra (focus on Land degradation)
- Fair-Ease project and EOSC: promoting Open Science and new digital services
- Focus general presentation of Galaxy

Day 1 Afternoon:

- Galaxy user side:
 - presentation of earth system how the interface works, focus Fair principles
 - how to build workflows, analyse data, and write or follow a training (example on land degradation, forest diversity workflow)
 - Share & publish

-
- Galaxy Dev side
 - how to build a tool
 - how to publish it

Day 2 Morning:

2 ways of working in parallel:

- Bring your own data and follow a training to learn how to use Galaxy on some specific trainings and get some Galaxy skills
- Bring your own analysis, scripts to analyse data and build galaxy tools to make your own workflow

Day 2 Afternoon:

Continue/ finish what was started in the morning

Wrap-up feedback on what was done, what needs improvement, what is there to continue, ...