Contribution ID: **47**   Type: **Presentation**

FAIR-by-design Pipelines to ensure reproducibility and transparency of innovative remote-sensing Data Products

Tuesday 14 October 2025 11:30 (11 minutes)

The rapid growth in global data production, particularly from remote sensing and Earth observation, has created significant opportunities to address pressing global challenges such as climate change, biodiversity loss, and sustainable resource management. Open data from satellites, drones, and environmental sensors, although increasingly available, often require complex integration due to diverse formats, large data volumes, and heterogeneous sources. Effectively exploiting these datasets needs automated processes that rely heavily on detailed metadata. This proposed presentation will detail the advances and developments of the GOYAS project and its platform, which emphasizes adopting metadata standards and automated processing pipelines to ensure a FAIR (Findable, Accessible, Interoperable, and Reusable) data lifecycle, generating novel remote-sensing-based products.

Introduction
Earth Observation and environmental sciences are experiencing significant growth in available data, not only in terms of volume but also in terms of quality at various resolutions (spatial, spectral/radiometric, temporal) and through new instruments and sensors. Moreover, recent developments in techniques, methods, and Artificial Intelligence (AI) algorithms facilitate the estimation and production of new physical, chemical, or biological georeferenced variables. Addressing global issues such as climate change and ecosystem management requires integrating diverse datasets. For instance, AI-based algorithms for estimating chlorophyll concentration in freshwater reservoirs depend on in-situ data to train and validate new models. Remote-sensing data pose challenges due to their format variability, large volumes, and complexity. Global initiatives such as Copernicus and Digital Earth Australia standardize data products derived from Earth observation missions, facilitating user access. However, incorporating new algorithms or variables into these platforms can be complex, particularly for variables that cannot be produced globally. Nevertheless, innovative and experimental AI-derived data products are beneficial for various stakeholders. Ensuring FAIR compliance of these datasets is crucial as it enables their reuse, validation, and reproducibility. Achieving this requires robust data lifecycle management supported by comprehensive metadata standards.

FAIR Data Life Cycle in GOYAS project
The OSCARS project aims to enable Open Science services adhering to FAIR principles, interconnected with Research Infrastructures and the European Open Science Cloud (EOSC). Under the OSCARS umbrella, the GOYAS project is developing an Open Science service linked to ENVRI (European Environmental Research Infrastructures). The GOYAS platform employs the ISO19139 metadata standard, comprehensively addressing descriptive, administrative, and structural requirements for remote sensing data.

- `Descriptive metadata ensures datasets are discoverable by incorporating unique persistent identifiers a`
- `Administrative metadata provides essential context regarding data creation, quality assurance, licensing`
- `Structural metadata, including format details, encoding, and logical attributes, enable interoperability`

GOYAS systematically documents these metadata categories, ensuring data are ready for integration, interpretation, and validation, both manually and through automated pipelines. Metadata also includes quality details such as accuracy and expected errors, enhancing transparency and helping users in evaluating data suitability.

FAIR-by-design Pipeline
The GOYAS infrastructure incorporates data products from four Spanish research institutes belonging to CSIC (Spanish National Research Council): Doñana Biological Station, Institute of Marine Sciences of Andalusia, Institute of Marine Science, and the Physics Institute of Cantabria. Product types are diverse, including

flood data from Doñana National Park, satellite-derived bathymetries, oceanic data (temperature, salinity), and freshwater quality indicators, among others. The lifecycle of these products involves multiple complex steps—data collection, corrections, preprocessing, curation, processing, and ingestion. The FAIR-by-design approach adopted by GOYAS documents all components necessary for reproducibility. This pipeline systematically records each dataset action, from initial collection through final outputs, ensuring transparency and reproducibility. It references preprocessing algorithms using persistent identifiers and explicitly describes performed actions. Additionally, the pipeline integrates FAIR assessment tools (such as FAIR EVA) to conduct tests based on Research Data Alliance FAIR indicators (RDA FAIR Maturity Group). Consequently, new data products are published automatically only after all preceding steps have achieved FAIR compliance.

The primary users of the GOYAS platform include researchers developing innovative, experimental, or very localized remote sensing data products not suited for integration into large-scale portals like Copernicus. Additionally, the produced data are valuable to environmental scientists, policymakers, and public administrations for informed decision-making on environmental issues. This structured, FAIR-compliant pipeline significantly supports interdisciplinary research, environmental monitoring, resource management, and decision-making.

The proposed presentation will provide an overview of the GOYAS project, showcasing the FAIR-by-design workflow, including metadata management, automated data integration, processing, and FAIR validation. It will also include technical details of the underlying technologies and illustrate practical applications of the platform using available data products. This transparent and automation-supported approach is crucial for effectively adopting FAIR principles, and the GOYAS platform could serve as a model for similar initiatives.

Primary author: AGUILAR GÓMEZ, Fernando (IFCA-CSIC)

Co-authors: Ms GARCÍA-ESPRIU, Aina (ICM-CSIC); Dr GONZÁLEZ-HARO, Cristina (ICM-CSIC); Dr GARCÍA-DÍAZ, Daniel (IFCA-CSIC)

Presenter: AGUILAR GÓMEZ, Fernando (IFCA-CSIC)

Session Classification: Presentations Session 3: Rigorous, responsible and reproducible science in the era of FAIR data and AI

Track Classification: SciDataCon2025 Specific Themes: Rigorous, responsible and reproducible science in the era of FAIR data and AI