

Aguilar Gómez, F.; González-Haro, C.; García-Espriu, A.; Caballero de Frutos, I.; García-Díaz, D.; García Díaz, D.; Afán, I.; Flo, E.; López, A. Aplicación de los principios FAIR en el desarrollo de una plataforma de datos de teledetección: aprendizaje del proyecto GOYAS

Aplicación de los principios FAIR en el desarrollo de una plataforma de datos de teledetección: aprendizaje del proyecto GOYAS

Aguilar Gómez, Fernando¹ **González-Haro, Cristina**² **García-Espriu, Aina**² **Caballero de Frutos, Isabel**³ **García-Díaz, Daniel**⁴ **García Díaz, Diego**⁵ **Afán, Isabel**² **Flo, Eva**² **López, Antonio**²

¹ Instituto de Física de Cantabria (CSIC-UC)

² Institut de Ciències del Mar (CSIC)

³ Instituto de Ciencias Marinas de Andalucía (CSIC)

⁴ Centro de Investigaciones sobre Desertificación (CSIC)

⁵ Estación Biológica de Doñana (CSIC)

ORCID: Aguilar Gómez 0000-0001-9462-4831 González-Haro 0000-0003-4602-852X García-Espriu 0000-0001-9793-5547 Caballero de Frutos 0000-0001-7485-0989 García-Díaz 0000-0002-3399-1598 García Díaz 0000-0002-2757-7112 Afán 0000-0003-3874-0461 Flo 0000-0002-4423-4779 López 0009-0004-6954-0184

Correspondencia: aguilarf@ifca.unican.es cgharo@icm.csic.es ainagarcia@icm.csic.es isabel.caballero@csic.es garciad@ifca.unican.es diegogarcia@ebd.csic.es isabel.afan@csic.es evaflor@icm.csic.es antoniolopez@icm.csic.es

RESUMEN

La creciente disponibilidad de datos de teledetección plantea nuevos retos en términos de documentación, trazabilidad, interoperatividad y reutilización. En este contexto, los principios FAIR (Encontrables, Accesibles, Intereoperables y reutilizables, por sus siglas en inglés) se han consolidado como marco conceptual para mejorar la gestión y explotación de datos científicos, facilitando la reproducibilidad y reusabilidad. Sin embargo, su implementación práctica en plataformas operativas sigue siendo un desafío. Este trabajo presenta la experiencia del proyecto GOYAS (Geospatial Open Science Yielding Applications), desarrollado en el marco del proyecto Horizon Europe OSCARS y alineado con la "European Open Science Cloud" (EOSC). GOYAS propone una arquitectura integrada para la producción y publicación de productos derivados de datos satelitales bajo un enfoque FAIR-by-design. La plataforma combina distintos componentes para facilitar la ingesta de datos y publicación. En este artículo se describen todos los apartados técnicos y arquitectura definida, así como el aprendizaje adquirido a lo largo del proyecto. Finalmente, se reflexiona sobre la integración con infraestructuras mayores como ENVRI-HUB (Grupo de Infraestructuras de Investigación Medioambientales europeas) y EOSC como estrategia para aumentar la visibilidad y facilitar el acceso por parte de investigadores, gestores y administraciones públicas.


Palabras clave: *datos, FAIR, EOSC, trazabilidad, reproducibilidad*

Fecha de recepción: 17 febrero 2026 · **Fecha de aceptación:** 19 febrero 2026

Aplicación de los principios FAIR en el desarrollo de una plataforma de datos de teledetección: aprendizaje del proyecto GOYAS

Aguilar Gómez, Fernando ⁽¹⁾, González-Haro, Cristina ⁽²⁾, García-Espriu, Aina ⁽²⁾, Caballero de Frutos, Isabel ⁽³⁾, García-Díaz, Daniel ⁽⁴⁾, García Díaz, Diego ⁽⁵⁾, Afán, Isabel ⁽²⁾, Flo, Eva ⁽²⁾, López, Antonio ⁽²⁾


⁽¹⁾ Instituto de Física de Cantabria (CSIC-UC).

 0000-0001-9462-4831, aguilarf@ifca.unican.es.


⁽²⁾ Institut de Ciències del Mar (CSIC).

 0000-0003-4602-852X, cgharo@icm.csic.es ;  0000-0003-3874-0461, isabel.afan@csic.es


 0000-0002-4423-4779, evafl@icm.csic.es ;  0009-0004-6954-0184, antoniolopez@icm.csic.es

 0000-0001-9793-5547, ainagarcia@icm.csic.es


⁽³⁾ Instituto de Ciencias Marinas de Andalucía (CSIC).

 0000-0001-7485-0989, isabel.caballero@csic.es

⁽⁴⁾ Centro de Investigaciones sobre Desertificación (CSIC).

 0000-0002-3399-1598, garciad@ifca.unican.es

⁽⁵⁾ Estación Biológica de Doñana (CSIC).

 0000-0002-2757-7112, diegogarcia@ebd.csic.es

Resumen: La creciente disponibilidad de datos de teledetección plantea nuevos retos en términos de documentación, trazabilidad, interoperatividad y reutilización. En este contexto, los principios FAIR (Encontrables, Accesibles, Interooperables y reutilizables, por sus siglas en inglés) se han consolidado como marco conceptual para mejorar la gestión y explotación de datos científicos, facilitando la reproducibilidad y reusabilidad. Sin embargo, su implementación práctica en plataformas operativas sigue siendo un desafío. Este trabajo presenta la experiencia del proyecto GOYAS (Geospatial Open Science Yielding Applications), desarrollado en el marco del proyecto Horizon Europe OSCARS y alineado con la “European Open Science Cloud” (EOSC). GOYAS propone una arquitectura integrada para la producción y publicación de productos derivados de datos satelitales bajo un enfoque FAIR-by-design. La plataforma combina distintos componentes para facilitar la ingesta de datos y publicación. En este artículo se describen todos los apartados técnicos y arquitectura definida, así como el aprendizaje adquirido a lo largo del proyecto. Finalmente, se reflexiona sobre la integración con infraestructuras mayores como ENVRI-HUB (Grupo de Infraestructuras de Investigación Medioambientales europeas) y EOSC como estrategia para aumentar la visibilidad y facilitar el acceso por parte de investigadores, gestores y administraciones públicas.

Palabras clave: datos, FAIR, EOSC, trazabilidad, reproducibilidad

Application of FAIR principles in the development of a remote sensing data platform: lessons learned from the GOYAS project

Abstract: *This paper presents the experience of the GOYAS project (Geospatial Open Science Yielding Applications), developed within the Horizon Europe OSCARS project and aligned with the European Open Science Cloud (EOSC). GOYAS proposes an integrated architecture for the production and publication of satellite-derived products under a FAIR-by-design approach. The platform integrates different components to facilitate data ingestion and publication. The article describes the technical decisions adopted, the lessons learned, and the opportunities for producers of innovative, experimental, or highly localized remote sensing products. It discusses the importance of adopting appropriate metadata standards, assigning persistent identifiers to ensure traceability and versioning, and automating data curation and ingestion processes. This article describes the technical details and defined architecture, as well as lessons learnt during the project. Finally, it reflects on integration with larger infrastructures such as ENVRI-HUB (Cluster of Environmental Research Infrastructures) and EOSC as a strategy to increase*

visibility and facilitate access for researchers, decision-makers, and public administrations.

Keywords: data, FAIR, EOSC, trackability, reproducibility

1. INTRODUCCIÓN

La teledetección ha experimentado un crecimiento exponencial en volumen, resolución y frecuencia de adquisición de datos. Más allá de los datos originales, los productos de datos derivados, proporcionando distintos tipos de variables y estimaciones, son cada vez más utilizadas en el ámbito científico y por las administraciones públicas. Sin embargo, la mera disponibilidad de estos datos no garantiza su reutilización efectiva. La falta de documentación estructurada, información de procedencia, versionado explícito o identificadores persistentes dificulta la reproducibilidad científica y la trazabilidad de los productos derivados, así como su utilización, por lo que es necesario automatizar y protocolizar el preprocesado (Vu-Duc *et al.*, 2025). Los principios FAIR proporcionan un marco conceptual para abordar estos desafíos (Wilkinson *et al.*, 2016). No obstante, su implementación práctica requiere decisiones técnicas y organizativas que deben adoptarse desde la concepción del sistema de información. El proyecto GOYAS surge precisamente con este objetivo: diseñar una plataforma de productos de datos de teledetección que integre los principios FAIR desde su arquitectura inicial, bajo un enfoque FAIR-by-design. GOYAS se desarrolla en el contexto del proyecto Horizon Europe OSCARS y en alineación con la European Open Science Cloud (EOSC), buscando no solo la producción de datos FAIR sino también su integración en ecosistemas europeos de ciencia abierta.

2. MATERIAL Y MÉTODOS

La arquitectura de GOYAS se basa en cuatro elementos integrados (Fig. 1). Primero, un catálogo de metadatos basado en una versión personalizada de GeoNetwork, con énfasis en transparencia estructural del producto. Segundo, GeoServer como motor de publicación de servicios OGC (WMS, WCS, WFS). Tercero, un pipeline automatizado de ingestión y generación de metadatos. Por último, un sistema de gestión de identificadores persistentes (PIDs). Esta integración permite cubrir simultáneamente los cuatro pilares FAIR.

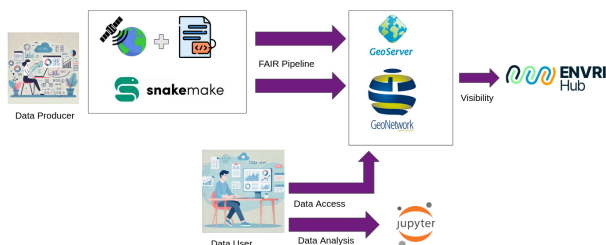


Figura 1. Arquitectura y componentes

2.1. Estándares de metadatos

Uno de los aprendizajes fundamentales es la necesidad de emplear un estándar de metadatos capaz de recoger información estructural, contextual y administrativa. En el caso de productos derivados satelitales, no basta con describir la cobertura espacial y temporal; es necesario documentar: variables físicas incluidas en el producto, algoritmos y versiones empleados, parámetros de procesamiento, dependencias de datos de entrada e información de calidad y validación.

La adopción de estándares interoperables permite que esta información sea interpretable tanto por humanos como por máquinas, facilitando la reutilización y la integración en infraestructuras mayores. En particular, dentro del proyecto GOYAS el estándar utilizado es el ISO 19139, es decir, la codificación XML del modelo conceptual definido en ISO 19115 (metadatos para información geográfica). Es compatible con los perfiles del reglamento europeo de metadatos INSPIRE. Este estándar permite la definición de diversos detalles sobre el proceso de producción y documentación concreta sobre la calidad de los productos y las variables que contienen. La Figura 1 muestra una captura de cómo se ven representados estos metadatos en el sistema.

2.2. PIDs y trazabilidad

La asignación de identificadores persistentes constituye un elemento central del diseño. Cada producto maduro recibe un PID que permite: distinguir versiones, garantizar la citabilidad, facilitar la trazabilidad entre producto, metadatos y servicios y mantener referencias estables en el tiempo.

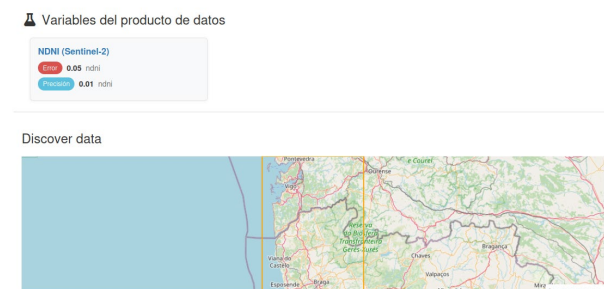


Figura 2. Vista de producto de datos con identificación de variables (precisión, error) y cobertura geográfica.

El uso de PIDs no se limita al dataset final, sino que se concibe como parte de una estrategia más amplia de gestión de ciclo de vida del dato. Esta aproximación resulta esencial para entornos EOSC, donde la interoperatividad y persistencia son requisitos clave. Asimismo, gracias al estándar de metadatos elegido, es posible definir ciertos procesos de preprocesado o curación para indicar los pasos de generación del producto. Esto también se representa en el sistema de información (Fig. 3).

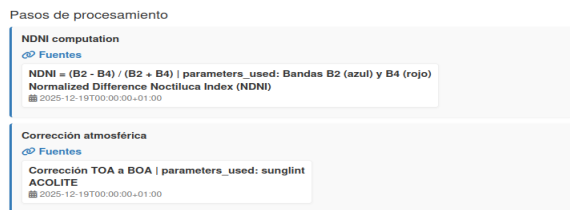


Figura 3. Vista de metadatos con información de procedencia

2.3. “Pipeline” de ingestión de datos

La producción de datos FAIR no puede depender exclusivamente de procesos manuales que pueda llevar a errores o inconsistencias. GOYAS incorpora un pipeline automatizado que genera metadatos estructurados como subproducto del procesamiento. Está basado en la librería de Python “Snakemake” y su misión principal es asegurarse que cada paso de la ingestión de datos sea completado de forma correcta. A partir de un fichero de configuración en el que se expresan todos los detalles de un producto de datos, este sistema valida la coherencia entre los archivos y prepara la estructura adecuada que permite la elaboración automática de los metadatos. Al estar basada en un sistema de transacción paso a paso, el “pipeline” asegura que todos los pasos se realizan de la forma esperada y, en caso de error, se paraliza el proceso de ingestión, que puede ser reanudado en el mismo punto una vez se ha solucionado (ejemplo: inconsistencia en los metadatos). Además, al tratarse de un sistema integrado, facilita la publicación coordinada en catálogo y servicios, de modo que garantiza esta consistencia. La automatización reduce errores humanos, mejora la consistencia y disminuye la barrera de entrada para nuevos productores de datos, que no necesitan conocer el estándar de metadatos complejo para poder publicar. Este enfoque es especialmente relevante en escenarios de producción continua de productos derivados. Una vez finaliza la ingestión de los datos, el propio sistema Geonetwork puede validar el cumplimiento del estándar. Sólo en ese caso, se puede asignar un identificador persistente al conjunto de datos.

3. PRODUCTOS DE DATOS

El proyecto GOYAS está formado por un consorcio de cuatro institutos del CSIC productores de datos de teledetección. Dentro del proyecto se están haciendo disponibles distintos productos, entre los que se encuentran: datos de calidad de agua en embalses, información de inundación del Parque Nacional de Doñana, batimetría derivada de satélite en zonas someras, información de la calidad del agua del Mar Menor y variables oceanográficas. El objetivo es que todos los productos estén representados de forma homogénea, con información similar de procedencia, detalles de variables y basado en estándares aceptados internacionalmente. A través de la plataforma, disponible en goyas.csic.es, se permite el acceso libre a usuarios para la descarga de datos e imágenes en distintos formatos como GeoTIFF o NetCDF, así como

la conexión a través de estándares de la OGC. Gracias al uso de identificadores persistentes, los datos pueden ser directamente citados en caso de reutilización.

4. ECOSISTEMAS EUROPEOS

Un objetivo estratégico del proyecto es su alineación con infraestructuras mayores como EOOSC y ENVRI-HUB. La agregación a estos ecosistemas permite: Incrementar la visibilidad, facilitar el descubrimiento transdisciplinar, mejorar la confianza institucional y garantizar sostenibilidad a largo plazo. Asimismo, GOYAS incluye demostradores basados en notebooks Jupyter que permiten a usuarios explorar y reproducir análisis directamente sobre los datos publicados. Este componente contribuye tanto a la accesibilidad como a la reutilización efectiva. La automatización de los procesos de ingestión y publicación de datos facilita y agiliza la disponibilidad. Esto permite el uso de datos en situaciones de emergencia, para la toma de decisiones o en la gestión de entornos naturales. A diferencia de otros repositorios y portales de datos ([3, 4, 5]), GOYAS se enfoca en datos de teledetección experimentales: productos derivados de satélite u otras fuentes remotas que suelen ser discontinuos o generados con nuevos métodos de investigación.

5. DISCUSIÓN

Nuestra experiencia previa en la producción de datos y en la participación en iniciativas internacionales para la interoperatividad de datos y servicios nos ha permitido diseñar un sistema que tiene a la reproducibilidad y la transparencia como concepto principal. Gracias al proyecto GOYAS hemos podido generar un sistema FAIR desde el inicio, facilitando esta reutilización de datos. A lo largo del proyecto hemos ido superando distintos retos y dificultades, que nos han llevado a desarrollar distintas reflexiones y adquirir cierto aprendizaje. Creemos que el concepto FAIR debe integrarse desde el diseño arquitectónico, no añadirse a posteriori. De este modo se toman decisiones estructurales que ayudaran a la producción de datos y a mejorar su uso. El esquema de metadatos a utilizar debe ser suficientemente expresivo para explicar todos los detalles de cómo se ha creado el producto de datos, las fuentes originales, los procesos involucrados, etc. Esto implica que la estructura del documento de metadatos puede ser completa y el número de campos a rellenar alto. Sin conocer a priori el estándar puede ser costoso entender qué información debe ir en cada apartado, por lo que creemos que sustituir un formulario o un documento XML por una lista puede resultar más sencillo para productores de datos. Esta información estructurada beneficia a posibles usuarios de los datos, que pueden identificar claramente la utilidad de la información. Los identificadores persistentes son esenciales para la trazabilidad y el versionado. Es la mejor forma que tenemos de indicar un uso y citar fuentes. La automatización del pipeline es clave para la sostenibilidad y para facilitar a los productores de datos una ingestión sencilla. Por último, la integración en infraestructuras mayores amplifica el impacto científico y social. La experiencia demuestra que implementar FAIR implica decisiones técnicas

concretas que afectan a la arquitectura, los flujos de trabajo y la gobernanza del dato. La plataforma GOYAS propone una nueva arquitectura integrando distintos componentes que facilitan la producción de los datos siguiendo los principios FAIR, adoptando distintas buenas prácticas de la ciencia abierta que hacen los productos más transparentes y reproducibles.

6. AGRADECIMIENTOS

Este trabajo ha sido desarrollado en el marco del proyecto GOYAS, financiado a través del programa OSCARS y alineado con la European Open Science Cloud (EOSC). También se ha financiado por la Agencia Estatal de Investigación (AEI) del Ministerio de Ciencia, Innovación y Universidades, proyecto PROFECIA (PID2023-152114OB-I00), correspondiente a la convocatoria de Proyectos de Generación de Conocimiento.

7. REFERENCIAS

- Vu-Duc, A., et al. (2025). A comprehensive survey of super-resolution remote sensing image datasets: Evolution, challenges, and future directions. *IEEE Access*, 13, 145350–145372. <https://doi.org/10.1109/ACCESS.2025.3599535>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 1–9. <https://doi.org/10.1038/sdata.2016.18>
- Roncella, R., et al. (2023). Publishing NextGEOSS data on the GEOSS Platform. *Big Earth Data*, 7(2), 413–427. <https://doi.org/10.1080/20964471.2022.2135234>
- Centro Nacional de Información Geográfica (CNIG). (2026). Centro de Descargas del CNIG. <https://centrodedescargas.cnig.es>
- European Space Agency (ESA). (2023). Copernicus Data Space Ecosystem. <https://dataspace.copernicus.eu>