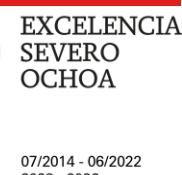


Implementación institucional de una estrategia de Ciencia Abierta basada en datos FAIR: la experiencia del ICN2

Jan Rodríguez Dueñas, Laura Maria Florez Corredor, Robert
Simeon y Andrés Henao Aristizábal



Proyecto María de Guzmán 2024-2025

CONTEXTO

- La **gestión de datos de investigación** es un elemento central de la **Ciencia Abierta**.
- Las **formaciones o talleres** generales suelen tener **poco impacto** en el trabajo real de investigación.
 - Buscamos un **enfoque alternativo** basado en el **trabajo directo con los investigadores**.

OBJETIVOS

- Abordar la **gestión de datos de investigación** como un **pilar** fundamental de la **Ciencia Abierta**.
- Adaptarse a las exigencias de los **principios FAIR** en la gestión de datos de investigación.
- Crear un **modelo operativo y replicable** para la gestión y publicación de datos de investigación.

Plan de Acción

- La **estrategia** se articuló en **tres niveles** complementarios:
- **Institucional:**
 - Diseño de una **estrategia** alineada con los **principios FAIR**.
 - Mejora de las **Capacidades Humanas y Técnicas**.
 - Desarrollo de una **Taxonomía Propia**.
- **Grupos de investigación:**
 - **Trabajo inicial** con **4 grupos piloto**, ampliado a **9 grupos**.
 - **Desarrollo conjunto** de arquitecturas **FAIR** y **DMPs**.
- **Investigadores:**
 - **Apoyo personalizado** adaptado a cada contexto.
 - Trabajo directo con **datasets reales** en **investigación activa**.

Mejora de las Capacidades Técnicas y Humanas

- **Incorporación de un perfil especializado:**
 - **Data Steward & Architect:** lidera la estrategia y el acompañamiento a los Grupos de Investigación.
 - Apoyado por un **ML Engineer** y un **Documentalista**.
- **Revisión de necesidades de almacenamiento.**
 - Análisis detallado de los Grupos y las Divisiones de Soporte a la Investigación: **Déficit estructural** en almacenamiento local seguro y de gran volumen.
 - **500 TB desplegados en el Centro de Procesamiento de Datos del ICN2.**

Jan Rodríguez
Data Steward and Architect



Andrés Henao
Machine Learning Engineer



Robert Simeón
Documentalist

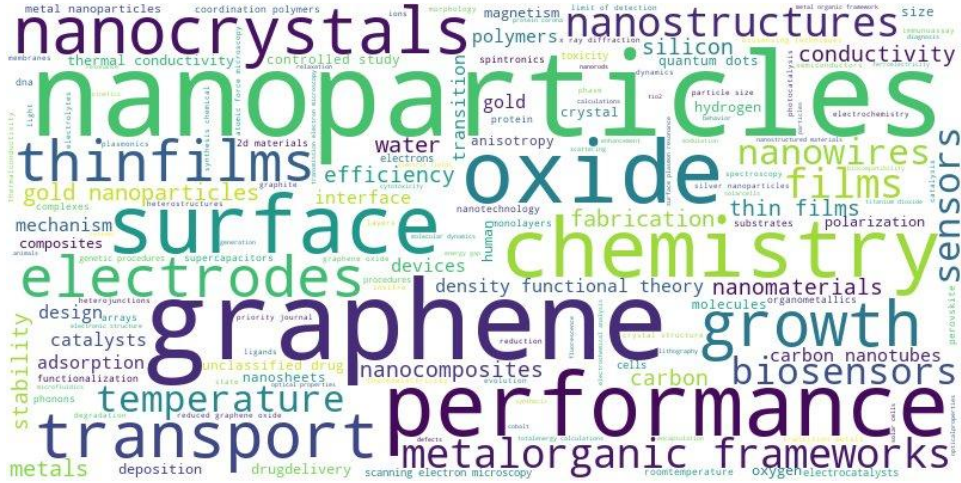


Taxonomía del ICN2

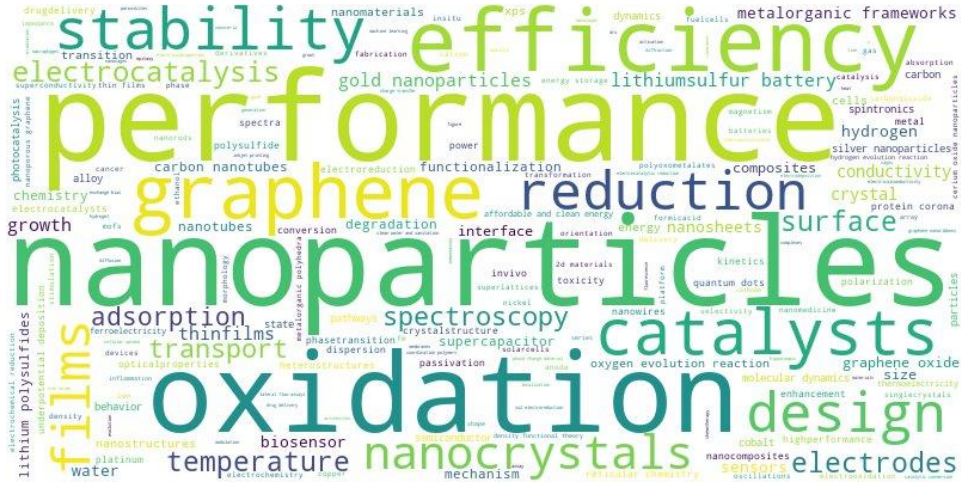
- **Vocabulario controlado** que mejore la capacidad de descubrimiento, la estandarización y la interoperabilidad en toda la producción científica y los conjuntos de datos del ICN2.
- **Metodología:**
 - **Extracción de 28.478 keywords** de publicaciones (2015–2025).
 - **Limpieza y normalización computacional** de los términos.
 - Reducción a un **conjunto curado de 200 términos** institucionales clave.

Taxonomía del ICN₂

PERIODO 2015-2022

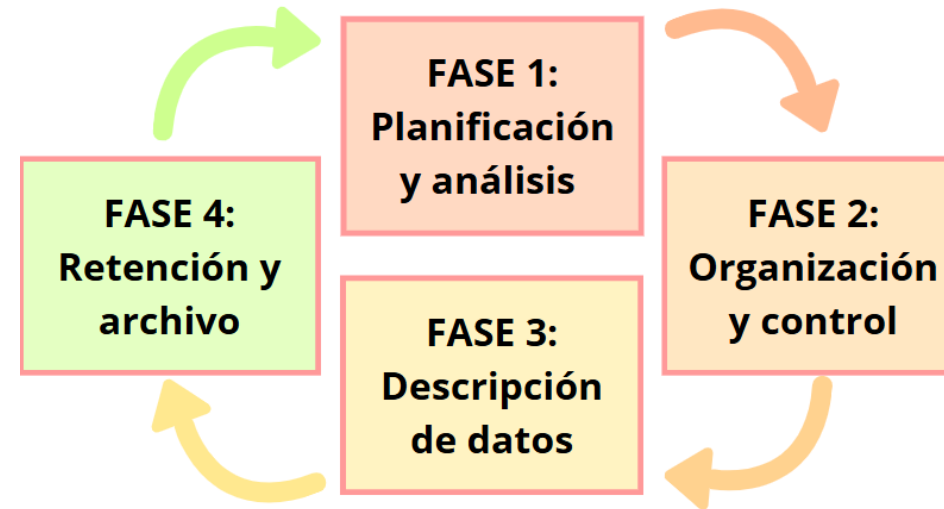


PERIODO 2023-2025

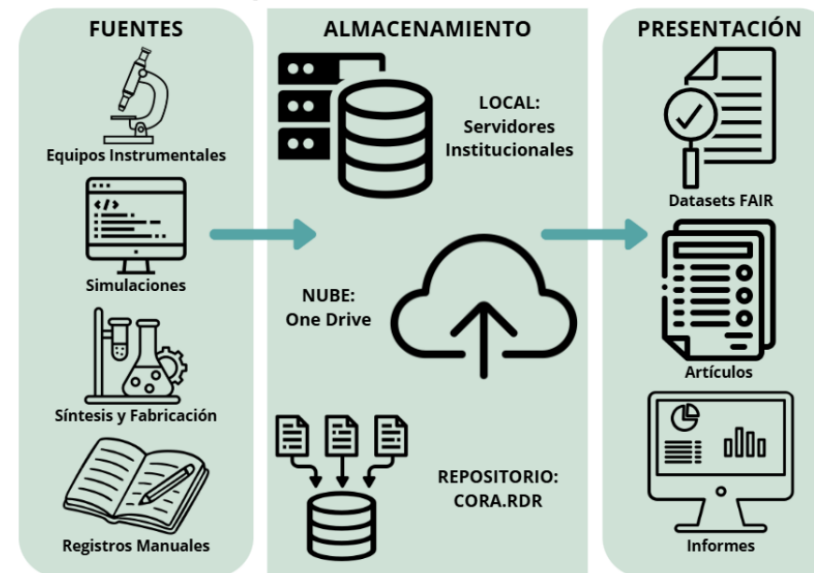


Definición de Arquitecturas

Definición de Arquitecturas












Arquitectura de Datos del ICN2



Trabajo con los Grupos de Investigación

- **Enfoque iterativo:** Permite validar y afinar un modelo flexible de arquitectura de datos, adecuado a sus necesidades.
 - Se desarrolló una arquitectura de datos con cada grupo, incluyendo flujos de trabajo, metadatos, estructuras de archivos y preparación de datasets.

Jordi Arbiol Advanced Electron Nanoscopy  4D-STEM, EELS/EDX and high-volume imaging; metadata design Dataset published 1	Jose A. Garrido Advanced Electronic Materials and Devices  Analytics for microscopies, biointerfaces and sensors Datasets published: 3	Aitor Mugarza Atomic Manipulation & Spectroscopy  STM, AFM, synchrotron spectroscopy and simulations	Mónica Lira Nanostructured Materials for Photovoltaic Energy  Photovoltaic materials; full synthesis traceability Datasets published: 3	Arben Merkoçi Nanobioelectronics and Biosensors  Biosensing and diagnostics; development of a pioneering RAW-data protocol.	Neus Bastús Inorganic Nanoparticles  Colloidal nanoparticle synthesis; creation of a cerium-oxide synthesis database.	Javier Rodríguez-Viejo Thermal Properties of Nanoscale Materials  Nanocalorimetry; FAIR handling of thermodynamic datasets. Datasets published: 1	Sergio Valenzuela Physics and Engineering of Nanodevices  Initial FAIR architecture for device-fabrication and spintronics measurements	Josep Nogués Magnetic Nanostructures  FAIR plan for synchrotron, microscopy, magnetometry Documentation structures for multidisciplinary integration
---	---	--	--	--	--	--	--	---

Trabajo con el Personal Investigador

- El desarrollo paralelo de **Data Management Plans (DMPs)** específicos por proyecto.
- Sesiones de formación en **curación de datasets** para investigadores.
- Análisis de los cuellos de botella encontrados.
- Creación de **plantillas adaptables** para acelerar futuras implementaciones.
 - **DMPs**, Archivos **README**, **Diccionarios** de Datos, **Análisis** de Gestión de Datos de los Grupos, **Arquitecturas** de Datos, **Presentaciones** para Grupos, **Scripts para Automatizar** procesos.
- Desarrollo de **módulos formativos** específicos para necesidades recurrentes, cómo la Formación continua en gestión de datos y la Formación en IA.

Resumen de Resultados

- **9 grupos** implementando **Arquitecturas** de datos FAIR.
- **8 datasets FAIR publicados** en el repositorio institucional + **6 datasets** adicionales **preparados** para su publicación.
- **Desarrollo y implementación** en procesos reales de **guías y plantillas institucionales** para la gestión de datos.
- Primera versión de la **Taxonomía Institucional**.

Conclusión

- La adopción efectiva de buenas prácticas de gestión de datos no depende solo de la infraestructura, sino de integrar estándares, liderazgo experto y trabajo directo con los investigadores desde el inicio del proceso científico.

Próximos Pasos

- **Proyecto María de Guzmán 2026-2027.**
- **Expansión integral:** implementación de las arquitecturas de datos en la totalidad de Grupos de Investigación.
- **FAIR by Design:** incluir a las Divisiones de Soporte a la Investigación (con equipos de alta tecnología que ayudan a producir los datos).
- **Compartir en abierto las plantillas y guías institucionales,** para facilitar su reutilización por parte de otros grupos e instituciones y reforzar el compromiso con la Ciencia Abierta y las prácticas FAIR.

Gracias por su atención

- © 2026 Jan Rodríguez Dueñas, Laura Maria Florez Corredor, Robert Simeon y Andrés Henao Aristizábal.
- Algunos derechos reservados.
- Esta presentación se distribuye bajo la licencia Creative Commons CC BY-SA 4.0 Internacional.
- <https://creativecommons.org/licenses/by-sa/4.0/>

