

# FAIR principles: from seed supermarkets to germplasm data centers

Alberto Cámara<sup>1</sup>, Santiago Moreno<sup>2</sup> and Mark D. Wilkinson<sup>1</sup>

<sup>1</sup>*Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Centro de Biotecnología y Genómica de Plantas. Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria-CSIC (INIA-CSIC). Campus Montegancedo 28223 Pozuelo de Alarcón (Madrid), Spain.*

<sup>2</sup>*Departamento de Biotecnología-Biología Vegetal, E.T.S.I. Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Avda. Puerta de Hierro 2-4, 28040 Madrid, Spain*

## Abstract

Ex situ seed storage in germplasm banks (GBs) plays a critical role in biodiversity conservation. However, databases for wild species in GBs, often managed by resource-constrained botanical gardens, are fragmented, utilizing diverse data formats, platforms, and standards with inconsistent accessibility. This fragmentation hinders the integration of data necessary for developing effective conservation strategies for wild species. To address this, we propose applying the FAIR Principles (Findable, Accessible, Interoperable, Reusable) to wild species and crop wild relatives data by creating a federated network of FAIR GB databases. This network would enable seamless cross-resource discovery and analysis, supporting more effective conservation strategies and highlighting the importance of GB data providers. Each database would integrate a FAIR transformation layer comprising semantic models, metadata publication via FAIR Data Points, and three innovations: a query-endpoint matching algorithm, an enhanced Triple Pattern Fragments resolution algorithm for data federation, and a Virtual Platform for data analytics through data visiting.

## Keywords

Data visiting, Triple pattern fragments, Query-endpoint matching, Semantic Models, FAIR Data Point

## 1. Introduction

Germplasm banks around the world store their information in heterogeneous formats and standards, making it hard to integrate with their peers. We propose an underlying-technology-agnostic system that creates machine-understandable representations of the native data that are chosen to be made public.

## 2. Future Work

We plan to work with the government of Spain to make some of their datasets follow the FAIR principles, enhancing the usability of climate, eco-geographical, and other relevant data types that are published by public agencies.

## Acknowledgments

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

---

SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025

✉ alberto.camara.ballesteros@alumnos.upm.es (A. Cámara)

ORCID 0000-0001-5613-9704 (A. Cámara); 0000-0002-6595-8987 (S. Moreno); 0000-0001-6960-357X (M. D. Wilkinson)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).