

formance autonomously. This combination reduces technical barriers and supports expert-driven domain refinement without sacrificing reproducibility.

Open, Interoperable and Extensible by Design

MOEBA-BIO was developed following open research software principles. The framework is fully open source, documented and publicly available through GitHub [L5]. Its architecture encourages extension: new objective functions, representations or evolutionary operators can be integrated without modifying the core system.

To promote interoperability, the framework is distributed through the Maven Central Repository [L6], enabling seamless integration into Java-based research pipelines and larger analytical infrastructures. This design choice supports sustainable reuse, reproducible workflows and incorporation into educational environments.

By clearly separating representation, evaluation and optimisation layers, the framework enhances transparency and traceability. Researchers can inspect, modify and extend each component independently, strengthening methodological openness.

Demonstrated Specialisation in Gene Co-Expression

The extensibility of the framework has already been demonstrated in gene co-expression analysis. By incorporating a regulatory coherence objective derived from inferred gene regulatory networks [3], MOEBA-BIO was specialised to align biclustering results with biologically meaningful modular structures.

Importantly, this domain adaptation did not require redesigning the system. Instead, new objectives were integrated into the existing architecture, illustrating how the framework enables community-driven extension rather than isolated reimplementation.

Towards Sustainable Evolutionary Research Infrastructures

As biomedical datasets grow in complexity and heterogeneity, analytical tools must evolve beyond fixed algorithms. Open frameworks that enable structural learning, adaptive configuration and domain-driven specialisation are essential for sustainable scientific progress.

MOEBA-BIO illustrates how evolutionary optimisation can transition from isolated methodological contributions to reusable, extensible research infrastructure. By embedding modularity, interoperability and expert-guided adaptability into its core design, the framework aligns with broader Open Science priorities centred on transparency, sustainability and collaborative development.

Links:

[L1] <https://khaos.uma.es/>

[L2] <https://itis.uma.es/>

[L3] <https://orkad.univ-lille.fr/>

[L4] <https://www.cristal.univ-lille.fr/>

[L5] <https://github.com/AdrianSeguraOrtiz/MOEBA-BIO>

[L6] <https://kwz.me/hIP>

References:

- [1] A. Segura-Ortiz, et al., “Exhaustive biclustering driven by self-learning evolutionary approach for biomedical data,” *Computer Methods and Programs in Biomedicine*, p. 108846, 2025.
- [2] M. Vandromme, et al., “A biclustering method for heterogeneous and temporal medical data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 2, pp. 506–518, 2020.
- [3] A. Segura-Ortiz, et al., “Multifaceted evolution focused on maximal exploitation of domain knowledge for the consensus inference of Gene Regulatory Networks,” *Computers in Biology and Medicine*, vol. 196, p. 110632, 2025.

Please contact:

Adrián Segura Ortiz

ITIS Software, University of Málaga, Spain

adrianseor.99@uma.es

José Francisco Aldana Montes

ITIS Software, University of Málaga, Spain

jfaldana@uma.es

Reinforcing Open Science in Biodiversity through Semantic Knowledge Graphs

by Yannis Marketakis, Eleni Tsouloucha, Athina Kritsotaki, and Yannis Tzitzikas (FORTH-ICS)

Transforming FishBase into a semantic knowledge graph shows how legacy biodiversity databases can become FAIR, interoperable infrastructures that enable reuse and integration in Open Science.

Open Science is one of the cornerstones of European research policy, promoting transparency, accessibility, reuse, and collaboration across scientific domains. While many scientific communities have made progress in opening data, experience shows that availability alone does not guarantee meaningful integration, reuse, or long-term sustainability. Addressing this challenge requires data infrastructures that go beyond simple access, supporting semantic integration, evolution, and machine-readable access, enabling data to be combined, queried, and reused across domains.

Biodiversity research provides a compelling paradigm for these challenges. FishBase, one of the most widely used global databases on fish species, has long played a crucial role in supporting research, education, and policy. Its contents are openly accessible and continuously curated, making it an important resource for the biodiversity community. Despite its importance, like many other legacy databases, FishBase was originally designed as a standalone information system. As a result, integrating its data with other research resources, performing complex cross-domain queries, or reusing its content in new contexts often requires manual effort.

This experience highlights a broader Open Science challenge: how to support the evolution of well-adopted, widely used open databases into FAIR-by-design data infrastructures. In what follows, FishBase is used as a concrete example to illustrate how semantic knowledge graphs, open APIs, and evolution workflows can reinforce Open Science by enabling interoperability, reuse, and long-term sustainability of biodiversity data, as demonstrated in recent Open Science efforts around biodiversity knowledge graphs [1]. Figure 1 provides an overview of this transformation and its role in enabling integration with external Open Science infrastructures.

Applying FAIR data principles in practice requires more than just making data available online. It requires data to be described in a structured and semantically explicit manner, accessed through standardized interfaces, and supported by mechanisms that ensure consistency and sustainability as data evolve over time. This also directly supports reproducibility and transparency, since datasets, structures, and relationships can be inspected, queried and recombined in a verifiable and machine-interpretable manner. Thus, FAIR data management is not an afterthought, but a design choice that must be embedded into the data infrastructure and its governance policy.

The experience with FishBase clearly illustrates these requirements. Although FishBase has long been an open and continuously curated resource, its original design as a conventional database limits cross-domain integration and large-scale reuse. Integrating its data with external biodiversity resources, answering complex queries that require combining different pieces of information, or supporting data reuse typically requires manual effort.

The transformation of FishBase into a semantic knowledge graph, carried out within the SemantyFish initiative [1] [L1], addresses these limitations by providing a FAIR-by-design representation of biodiversity data. Species, biological characteristics, ecology, population dynamics, life cycle and history, distribution, and other resources are modelled as semantically defined entities connected through explicit relationships. This makes data machine-interpretable and easier to integrate across domains. To this end, ontologies play a central role, offering shared conceptual models that ensure semantic consistency and interoperability with other biodiversity infrastructures.

Furthermore, access and reuse are further enhanced through open APIs, which expose the semantic knowledge graph via standardized, programmatic interfaces. This layer enables both human-oriented applications and machine-operated services and workflows to access up-to-date data in a structured, standardized, and reusable form, without relying on static exports or other ad hoc solutions.

Finally, evolution workflows ensure that the semantic representation remains aligned with updates to the original FishBase database, without requiring changes to FishBase's existing data management or update policies. As new data are added or existing data are revised, changes are systematically propagated to the knowledge graph, preserving accuracy and

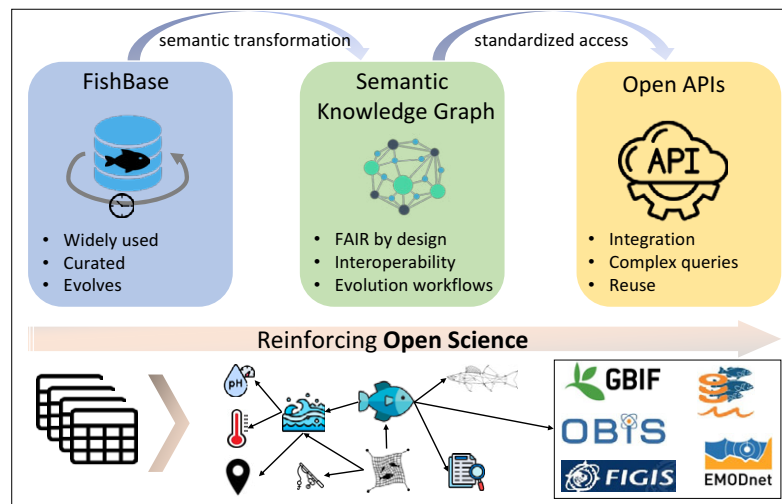


Figure 1: The transformation workflow of FishBase from a relational database into a FAIR-by-design semantic knowledge graph, supported by evolution workflows and open APIs enabling interoperable access and integration with external Open Science infrastructures.

trustworthiness over time. Overall, semantic modeling, knowledge graph construction, open APIs, and evolution workflows demonstrate how a widely used biodiversity database can be transformed into a FAIR, sustainable infrastructure that supports Open Science.

This transformation enables forms of sharing and reuse that were previously difficult to achieve. The explicit representation of entities and relationships allows researchers to formulate complex queries that traverse species, ecological traits, habitats, and references, supporting biodiversity research and analyses that go beyond predefined database views. At the same time, machine-interpretable semantic resources make it possible to programmatically combine FishBase data with external datasets, facilitating automated workflows and interoperability, as demonstrated in cross-domain knowledge graph applications for sustainable aquafood communication [2].

The experience with FishBase illustrates how rethinking data infrastructures can strengthen Open Science practices. Designing for interoperability, reuse, and evolution of existing scientific data resources proved essential for enabling meaningful data sharing without disrupting established practices. Looking ahead, such approaches can support broader perspectives, including cross-domain research by providing sustainable foundations for Open Science.

Link: [L1] <https://semantyfish.github.io/>

References:

- [1] Connecting fish data for open and sustainable science. EuroFish Magazine, Issue 6, 2025. <https://eurofish.dk/connecting-fish-data-for-open-and-sustainable-science/>
- [2] Y. Marketakis, et al “Construction, Access and Application of a Knowledge Base for Sustainable Aquafood Communication”, 14th International Joint Conference on Knowledge Graphs. Heraklion, Greece, 2025.

Please contact:

Yannis Marketakis, FORTH-ICS, Greece
marketak@ics.forth.gr