

CodexVitae-Sardinia: Harnessing Integrated Data Systems for Biodiversity Conservation in Sardinia

Alessandro Guaraglia^{1,2}, Maria Filippa Addis^{1,3}

¹CodexVitae NPA, Sassari, Italy, ²University of Sassari, Sassari, Italy, ³University of Milan, Milan, Italy

THE PROJECT

Our project aims at **integrating structured and unstructured databases to develop a comprehensive resource on Sardinian Endemisms**. We have identified over **1,000 endemic species** of animals and plants, which have been cataloged in a database, the **Sardinian Endemism Atlas**. This atlas includes more than **50 attributes** for each species, such as identification numbers, taxonomy, and connections to international and specialized databases. Georeferenced occurrences from platforms like **GBIF** (Global Biodiversity Information Facility), **iNaturalist**, and other national specialized portals were systematically correlated. Additionally, we integrated **BOLD DNA barcodes**, biomolecular data, and relevant scientific publications to enhance the dataset's richness. New content has been generated and translated into **Italian**, aiming to increase awareness among citizens and decision-makers regarding the importance of conserving biodiversity in Sardinia. As part of this initiative, we produced reports and established a "Biodiversity Intelligence" dashboard. The Sardinian Endemism Atlas serves multiple purposes and is updated weekly through automated processes. This ensures that users have access to the latest data, facilitating research and conservation efforts related to Sardinia's unique biodiversity.

BACKGROUND

Biodiversity resources and databases have significantly expanded with the establishment of several specialized national and European platforms. A notable advancement in this field is the emergence of **Citizen Science** projects and platforms, which utilize mobile phones, sensors, and increasingly artificial intelligence (AI) to enable users to contribute reports and identifications to databases. To facilitate data exchange between these proprietary systems—some of which are no longer supported, such as **FaunaEuropea.eu**—protocols and standards like **Darwin Core** have been developed. The **Global Biodiversity Information Facility (GBIF)** provides a centralized portal for accessing and searching this data, making it available to the scientific community. However, the lack of an Italian version limits its usability for local users, enthusiasts, and citizens. At the local level, dissemination projects primarily focus on endemic species, particularly plants, macroorganisms, and certain insects. We aim to create a database that highlights the biodiversity of underrepresented organisms, specifically those that are often overlooked due to their **charismatic bias**, cryptic nature, or difficulty in detection. Preliminary findings indicate that many organisms lack recorded occurrences on GBIF or iNaturalist. Some species have detections dating back to the 1980s; while they are present in the **National Biodiversity Database (NNB)**, they are not listed on GBIF. Ultimately, this initiative seeks to establish a unified resource characterized by local biodiversity features and a **Biodiversity Intelligence** dashboard.

METHODS

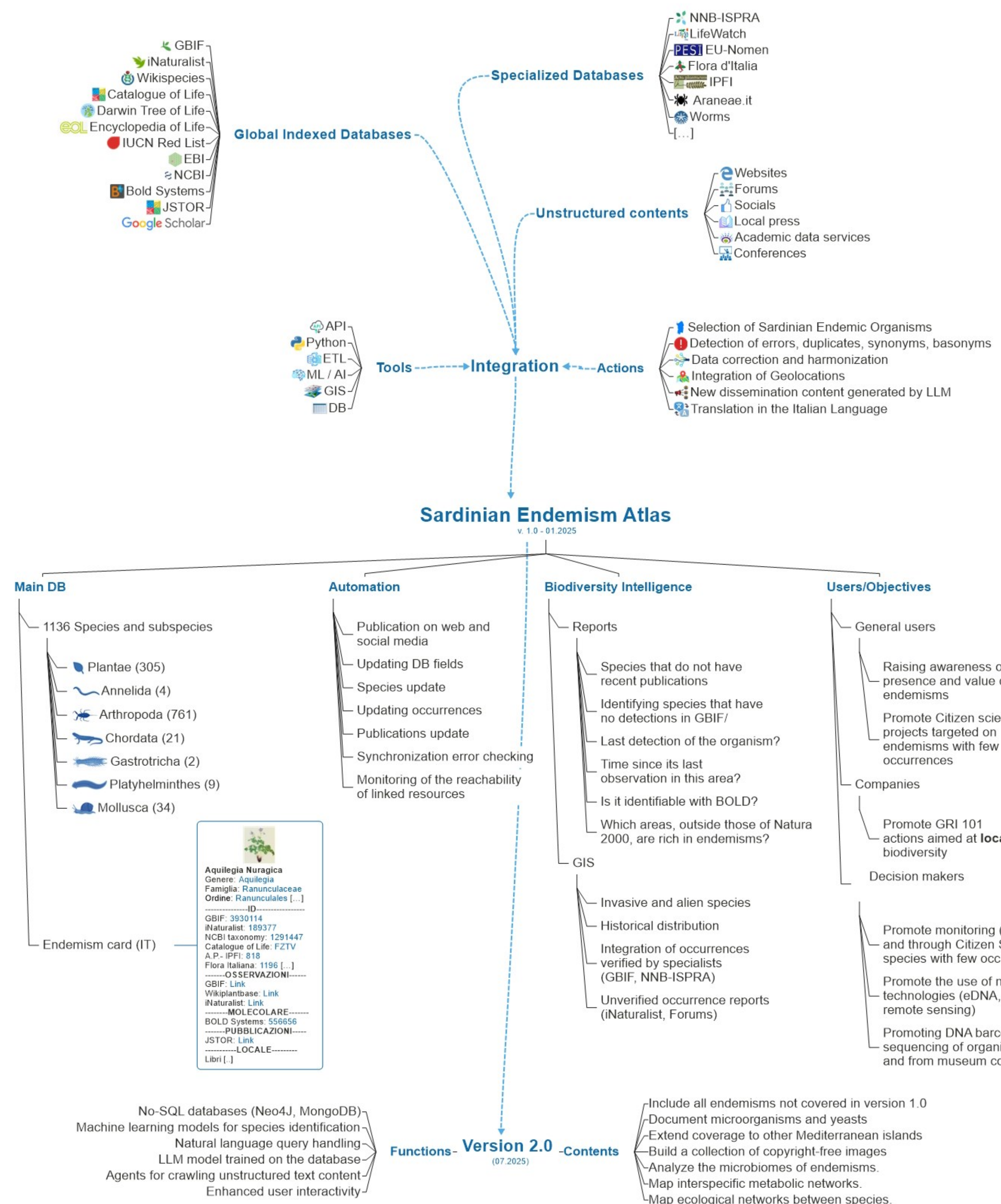
The updated list of endemisms was compiled through consultation of the Checklist of Italian Fauna, the Checklist of Italian Flora, and various specialized databases. Several recently discovered or newly classified species were integrated. The lists were compared with the species and occurrences available on GBIF.

Data integration, comparison, and transformation were managed using APIs, Python, and ETL pipelines. The database of endemics and the development website database are hosted on an Apache/MySQL server. Consistency checks and updates are performed via APIs and Python agents.

The versions of the programming languages, server, and database used are as follows:

- Apache HTTP Server** (version 2.4.58, Apache Software Foundation, USA)
- MySQL** (version 8.0.35, Oracle Corporation, USA)
- PHP** (version 8.2.12, The PHP Group, USA)
- Python** (version 3.12.1, Python Software Foundation, USA)

SARDINIAN ENDEMISM ATLAS



PRELIMINARY RESULTS

A total of 1136 endemic taxa (species and subspecies) exclusive to or strongly characteristic of Sardinia have been identified. The taxonomic distribution obtained with this approach (as of 12 November 2024, subject to updates) is as follows:

- 305 Plantae
- 4 Annelida
- 761 Arthropoda
- 21 Chordata
- 2 Gastrotricha
- 34 Mollusca
- 9 Platyhelminthes

It was observed that many species, particularly among arthropods and cave-dwelling organisms, have not been recorded in decades. For each species, unique identifiers (IDs) and links to key national and international taxonomic databases have been identified and harmonized.

The data has been stored in a database, and dynamic agents have been developed to update information (e.g., occurrences, publications) via APIs.

A website has been set up at <http://codexvitae.it/sardinia>. A dynamic page has been created for each organism, designed to be automatically updated with data from the main database.

The release of version 1.0 is scheduled for **January 2025**.

REFERENCES

1. Checklist della Fauna d'Italia, <https://www.lifewatchitaly.eu/iniziativa/checklist-fauna-italia-it/>
 2. Checklist della Flora d'Italia, <https://floritaly.plantdata.it/index.php>

1. National Biodiversity Network (NNB) - ISPRA, <https://www.nnb.isprambiente.it/en/find-data>
 2. GBIF Global Biodiversity Information Facility, <https://www.gbif.org>

