Connecting West and Central African Herbaria Data: A new Living Atlases regional data platform

Sylvain Morin[‡,§], Alice Ainsa[‡,§], Raoufou A. Radji[|], Anne-Sophie Archambeau[¶,§], Hervé Chevillotte[¶], Eric Chenin[¶,§], Sophie Pamerlon[#,§]

‡ MNHN (Museum national d'Histoire naturelle), Paris, France
§ GBIF France (Global Biodiversity Information Facility), Paris, France
| Université de Lomé, Lomé, Togo
¶ IRD (Institut de recherche pour le développement), Paris, France
# OFB (Office français de la biodiversité), Paris, France

Corresponding author: Sylvain Morin (sylvain.morin@mnhn.fr)

Abstract

The label transcription and imaging of specimens in key African herbaria has been ongoing since the early 2000s. Many collections in Benin, Cameroon, Côte d'Ivoire, Gabon, Guinea Conakry, and Togo are now fully transcribed and partially digitized. More than 200 000 transcribed specimens are available with the following distribution:

- Benin: 45 000
- Cameroon: 70 000
- Côte d'Ivoire: 18 000
- Gabon: 70 000
- Guinea Conakry: 5 000
- Togo: 15 000

In April 2021, a BID project was started to deliver a regional data platform of West and Central African herbaria. Biodiversity Information for Development (BID) is a multi-year programme funded by the European Union and led by GBIF with the aim of enhancing capacity for effective mobilization and use of biodiversity data in research and policy in the 'ACP' nations of sub-Saharan Africa, the Caribbean and the Pacific. Our project's funding runs from April 2021 to April 2023.

At this stage of the project, we are working on defining the information technology (IT) architecture (Fig. 1) and selecting the tools that we will be using to achieve our goals. In the talk, we will present our conclusions through architecture schemas and tools demonstrations.

Each of the 6 countries will have its own PostgreSQL database, storing its data. They will also have access to the RIHA data management platform (Réseau Informatique des Herbiers d'Afrique / Digital Network of African Herbaria). This is a web application, developed in PHP, allowing full management of the data by herbarium administrators (Fig. 2).

An Integrated Publishing Toolkit (IPT) will fetch these herbaria data from the databases, create the Darwin Core archives, and connect these data automatically to gbif.org on a periodic basis (Fig. 3).

On the databases, we will use a PostgreSQL view to ease conversion from the RIHA data model to the Darwin Core model. On the IPT, we will create one dataset per country, linked to each PostgreSQL view. The SQL query will be configured to only fetch validated data, depending on the herbarium administrator's validation in the RIHA platform.

The automatic and periodic data transmission to gbif.org is a feature available in the IPT, and recently improved by the GBIF France team, which contributes to the IPT development.

Another part of the automatic data workflow will be to feed a Living Atlases portal for the West and Central African herbaria. This web application will allow public users to search, display and download herbaria data from West and Central Africa (Fig. 4).

Internally, this Living Atlases application will reuse open source modules developed by the Atlas of Living Australia (ALA). The application is mainly written in Java, uses JQuery/Bootstrap for the interface and relies on SolR and Spark in the backend. It has been developed to be easily reusable, by only modifying configuration and doing web customization (HTML / CSS), hiding most of the backend technological complexity.

The automatic data workflow will transfer datasets generated by the IPT, in Darwin Core Archive format, to the Living Atlases portal backend. A technical task orchestrator, yet to be selected, will implement this feature.

Living Atlases subportals, limited to data of one participating country, could be easily set up, leveraging the existing backend resources (Fig. 5).

One of the benefits of the Living Atlases portal is that we can easily deploy additional front end applications with limited data, configured by a filter (here, a filter on the data owner country). Only configuration and web customization (HTML / CSS) are required. All the backend modules, especially the ones storing data, are shared by the multiple front-ends, limiting the hardware consumption and data administration.

The full automation of the workflow will allow this platform to run at a very low maintenance cost for IT administrators. Moreover, adding a new herbarium member from West and Central Africa will be quite easy thanks to the architecture of the Integrated Publishing Toolkit and Living Atlases tools (Fig. 6).

Keywords

data portal, ALA, Atlas of Living Australia, IPT, Integrated Publishing Toolkit, GBIF, data workflow

Presenting author

Sylvain Morin

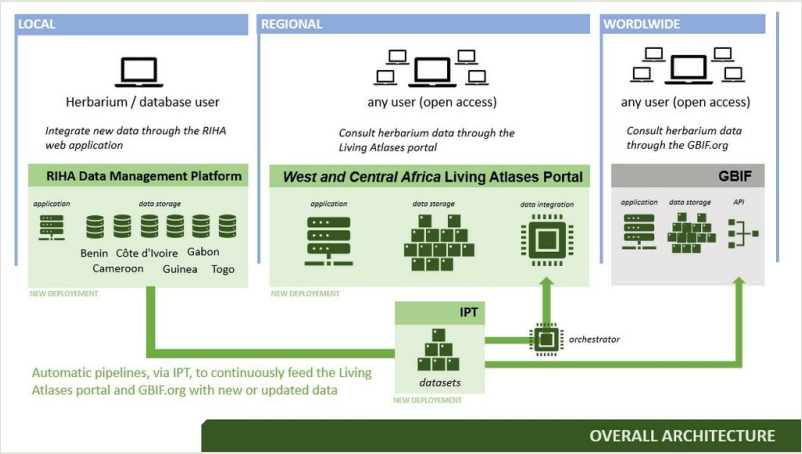## Presented at

TDWG 2021

Conflicts of interest

Figure 1.
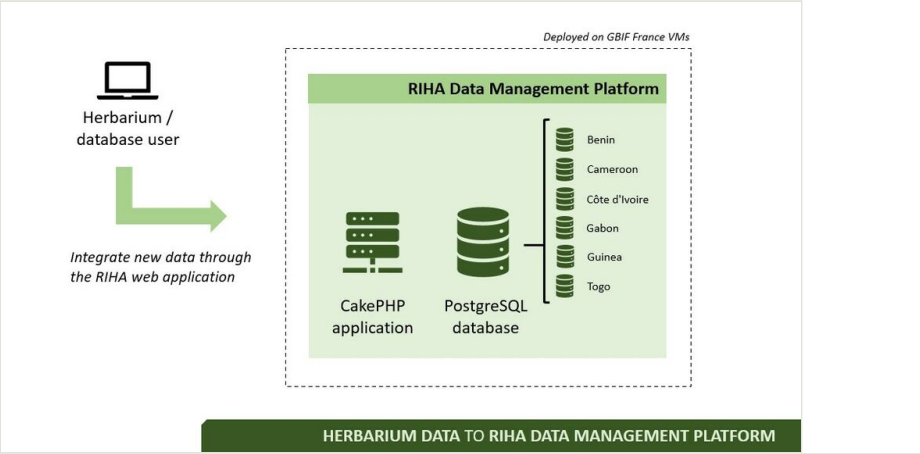
Overall architecture

Figure 2.

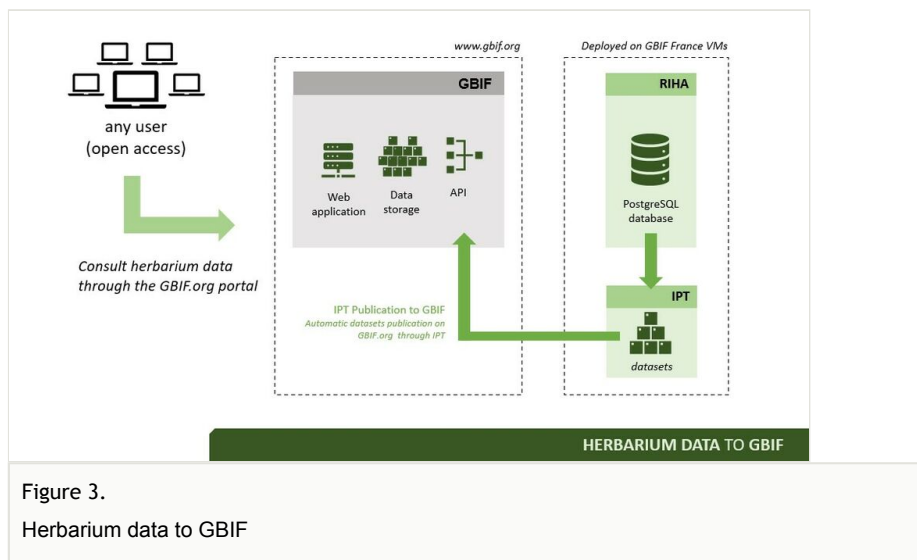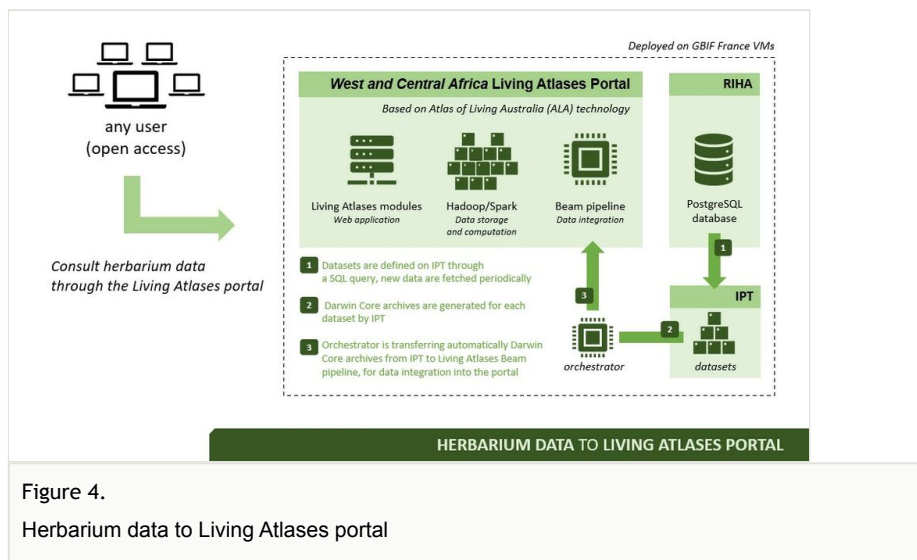Herbarium data to RIHA data management platform.

Figure 3.

Herbarium data to GBIF

Figure 4.

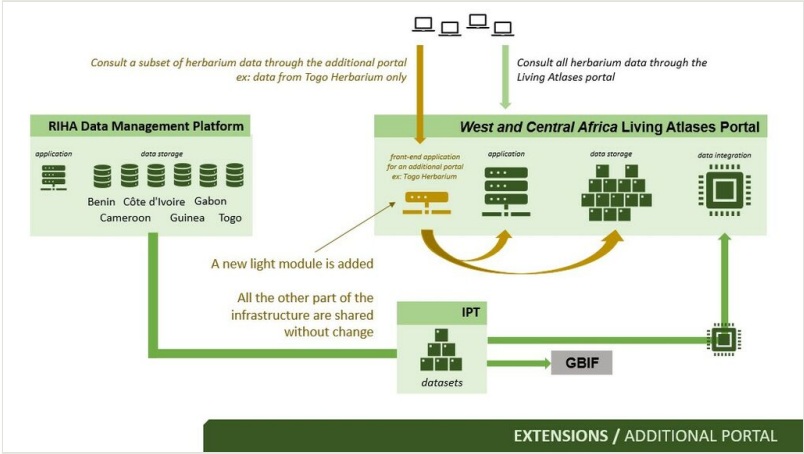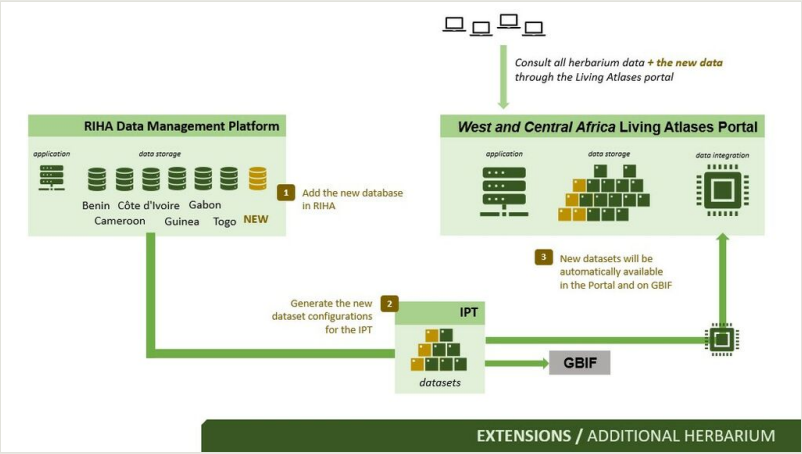Herbarium data to Living Atlases portal

Figure 5.

Extensions / Additional Portal

Figure 6.

Extensions / Additional herbarium