

Laguna - FAIR Research Data Infrastructure and Open Science Support Observatory



Washington Segundo Priscila Sena
Brazilian Institute of Information in Science and Technology - IBICT



Abstract

The planning and initial execution of the Laguna Project are described, whose premise is the creation of an open information infrastructure (OII), based on the development of certified data repositories and on the organizing principle of a Data Lake structure that connects to the Ocean of Scientific Object Information Data by systematizing and making available quality information. The infrastructure is designed with connection layers for processing the data collected from the "outside world," the so called data repositories.

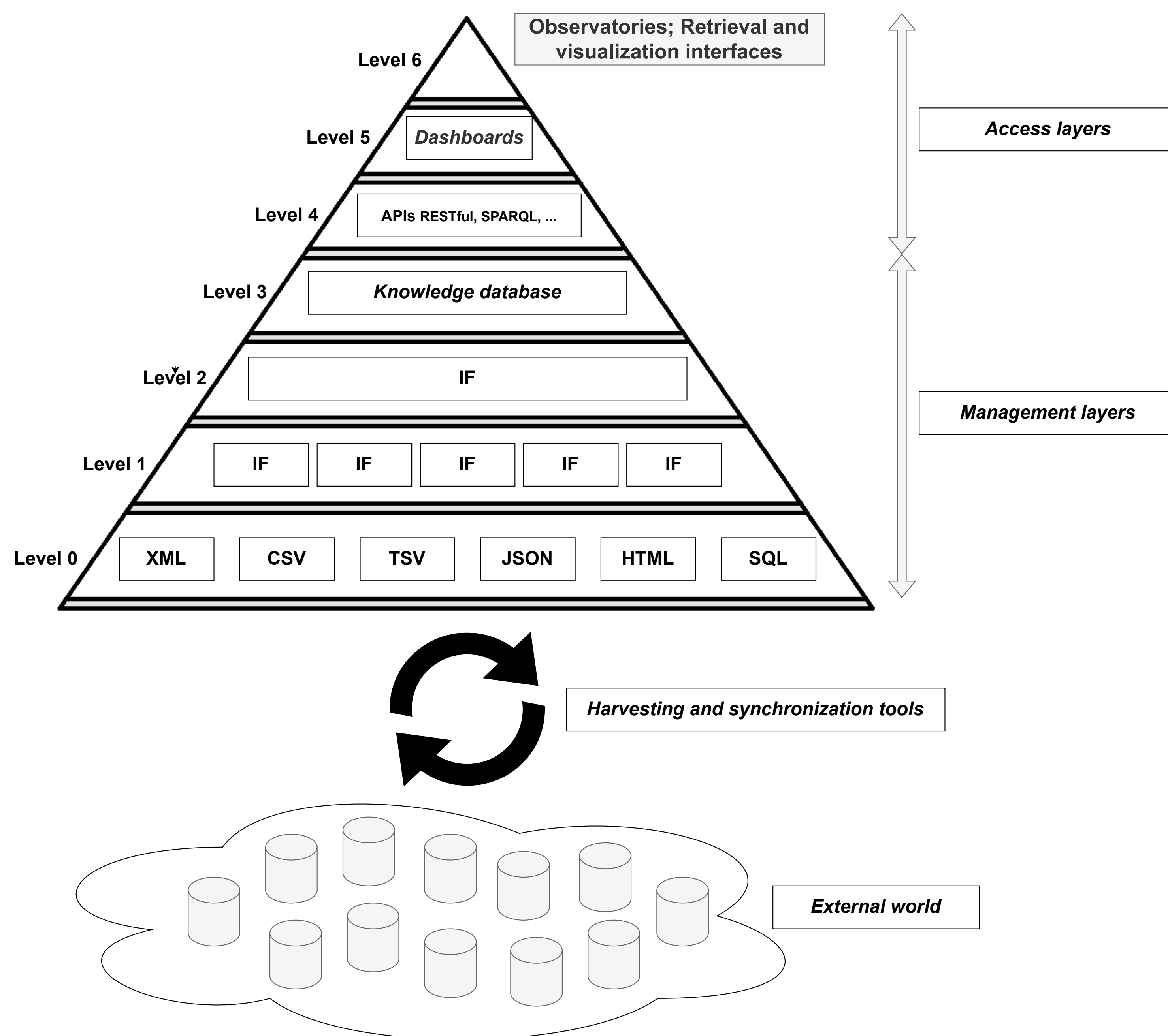
In particular, the Observatory in Support of Open Science in Brazil will map and promote Open Science practices in the country. The goal of its implementation is the monitoring of Open Science practices developed in Brazil and in other countries, as well as the availability of tools that contribute to quantitative and qualitative analyses. It will present a collaborative structure, allowing the scientific community and the population, in general, to develop their academic and technical activities in an agile way, in internal and external cooperation networks, for the production and dissemination of research. Its working dynamics will involve technical/scientific monitoring, information production, dashboards, analysis, and technology development. These will be a model for regional Open Science infrastructures, as well as for international infrastructures in the area.

Research objectives

The main objective of the infrastructure is to make available and disseminate resources in multidisciplinary digital repositories of publications and research data, guided by FAIR – Findable, Accessible, Interoperable, and Reusable – principles, in addition to the construction of an Open Science Support Observatory, which will offer subsidies for a better evaluation of science, promoting Open Science practices in scientific research developed in Brazil.

Study methodology

The present study adopted the following methodology to achieve the research objectives. Data is collected from the external world via harvesting and synchronization tools and it is refined by a sequence of treatment layers until be consumed on APIs and user interfaces.



Descriptive statistics

To give an idea of the data volume of data that need to be processed, our main scientific database source, in the Brazilian context, is the Lattes CV Platform (<https://lattes.cnpq.br/>). This source has stored almost **8 million research profiles** (only **225 hundred thousand** have a link with its respective ORCID profile), where, respectively, more than **4 hundred thousand** have a PhD, and approximately **7 hundred thousand** have a master degree as the highest academic qualification. Below are some statistics regarding the outputs that are present in the Lattes CV Platform.

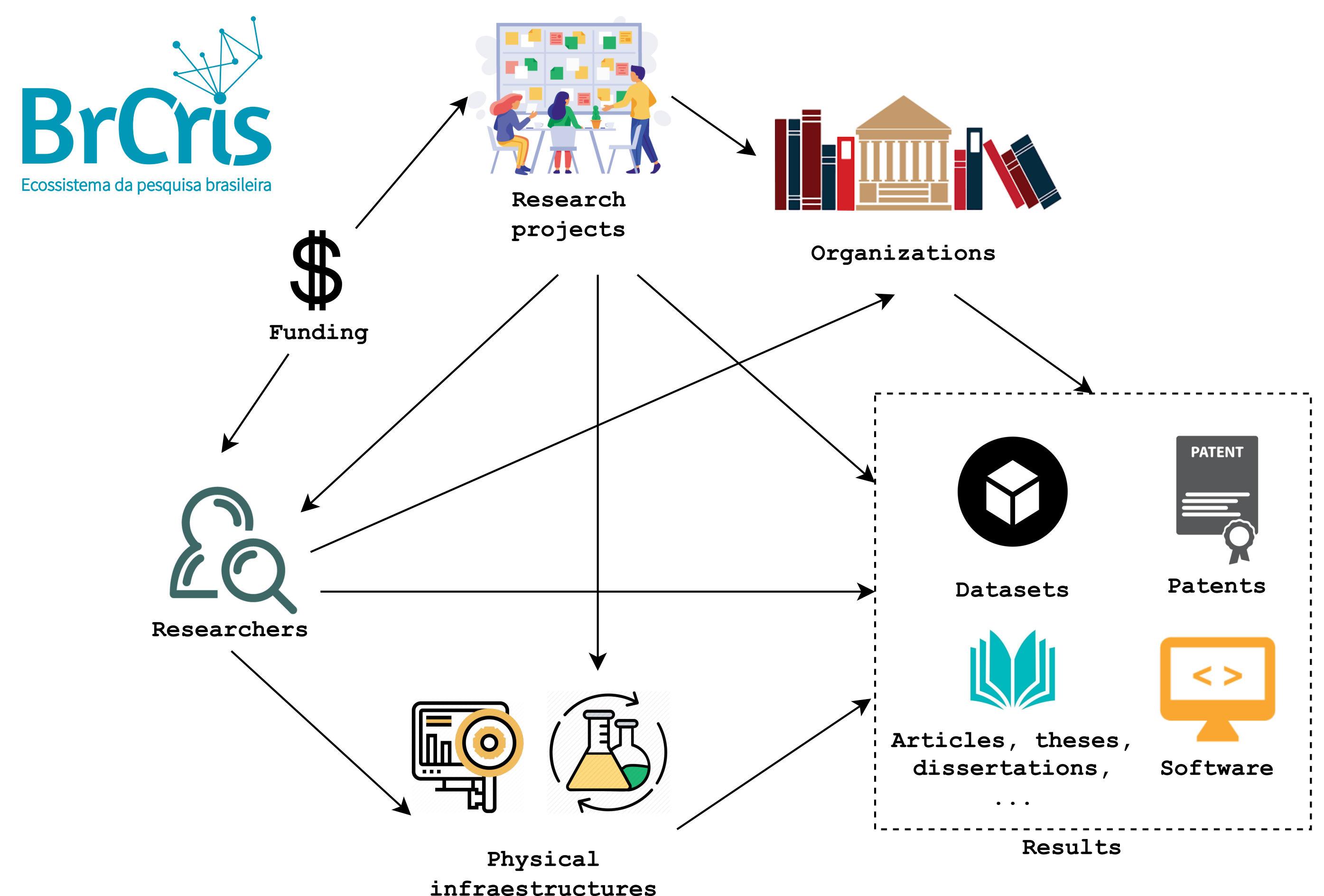
Table 1. Statistics of the Lattes CV Platform

	Jornal Articles	Conference Articles	Patents
with DOI	9 million 4.5 million	17 million 3 hundred thousand	75 thousand

Results and discussion

The Brazilian Open Science Ecosystem

From the Laguna infrastructure the research ecosystem of Brazil has been organized in the BrCris system (<https://brcris.ibict.br/vivo>).



Roughly speaking, this information system is based on a VIVO instance, in Kibana dashboards and in the LA Referencia Platform (<https://www.lareferencia.info>). From BrCris, the Open Science Support Observatory is emerging as a subset of outputs, datasets and visualizations.

Conclusions

The project will apply advanced computational methods for the processing, organization, and analysis of information to obtain data sets that are FAIR. Thus, it covers the stages of selection; transformation, linking, and enrichment; organization and indexing; and retrieval and visualization.

What is already known about this subject?

- Datalakes, Current Research Information Systems (CRISs) and Observatories are themes widely explored in the literature.
- Brazil is a continental country with great differences in the social development between geographic regions, which have a great impact in the policies and priorities in scientific research.

What does this study add?

- Despite all the difficulties, Open Science is already a reality in Brazil, Latin America and the world.
- Monitoring indicators and studying metrics of science evaluation is a key process in the promotion of Open Science.
- Using data lakes for processing scientific data and exporting it to a CRIS will help to answer a plenty of questions in the science evaluation process.

Practical implications

- We are almost launching BrCris as a production service.
- The Laguna infrastructure will help to improve quality and quantity of information that is aggregated by BrCris, promoting visualizations that will compose the Open Science Support Observatory.

Acknowledgements

We would like to thank all the team that is involved in the Laguna and BrCris's related projects. Namely: Adilson Pinto, André Lanna, André Corrêa, Caterina Pavão, Daniel Sundfeld, Fábio Canto, Flávio Silva, Henrique Rodrigues, Ingrid Pacheco, Jesiel Silva, Jesús Mena-Chalco, Joaquim Filho, J. Edilson Filho, Lautaro Matas, Leonardo Vitoriano, Mariana Salgueiro, Mateus Dias, Milena Silva, Nayara Gonzaga, Rapahel Boas, Raulivan Silva, Renan Batista, Renan Oliveira, René Gabriel Júnior, Samile Vanz, Sérgio Freitas, Tales Moreira, Thiago Nóbrega, Tiago Oliveira, Verônica Santos, Ulisses Silva and Vivian Silva.

Author Portfolio Website



Scan this QR code