

Developing a national science & technology funding registry in Portugal

The Portuguese scientific system is relatively recent and, mostly, supported by a few non-interoperable research information management systems with which researchers have to interact multiple times throughout their professional activity. The investment in the Portuguese scientific system has been fruitful and soon, it became obvious that to achieve sustainable growth and consolidate the achievements made so far, it was imperative to develop an integrated ecosystem to manage data on the Portuguese scientific activity. Launched in 2014, PTCRIS [1] is a program aiming at creating and sustaining the development of such integrated and coherent ecosystem at the national level (Figure 1) [2]. By promoting a wide and effective integration of several key information systems, PTCRIS will contribute to alleviate the administrative burden that lies on researchers and streamline management and scientific production processes through access to an authoritative, complete and reliable source of information.

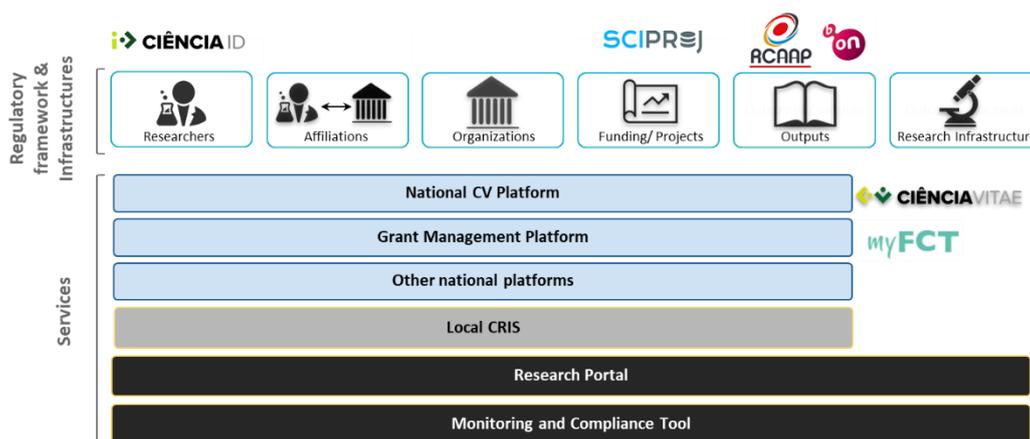


Figure1. PTCRIS ecosystem

Funding information is one of the pillars of PTCRIS. More often than not, scientific outputs involve funding of different stakeholders: researchers, the organization to which the researcher is affiliated to, and the material resources necessary for its execution. In spite of this, it is still very challenging to associate scientific outputs to all the entities that financially support the project, and everyone involved. In Portugal, this problem is further aggravated by the fact that funding information is dispersed throughout multiple databases, most of them hand crafted, not interoperable and with no support documentation. These databases are constituted and managed by different entities making it impossible to follow the life cycle of a funding record, since its attribution to the scientific product generated by it. Respecting the universality of scientific knowledge, it is necessary to have a structure that interconnects all contributors of the scientific system, which will allow a simple and quick answer to “Who?”, “Where?” and “How?”. To this end, the PTCRIS has developed SciPROJ - a new national database that aggregates funding records (national and international) that support science and technology activities carried out in Portugal.

The information available at SciPROJ is organized into four major groups: funding, people, projects and organizations. The constitution of these data groups derives from the adoption of the OpenAIRE data profile [3] and the respective adaptation to the

reality of the portuguese scientific-academic system. The model for sharing these data respects CERIF [4].

The SciPROJ data comes mostly from third-party systems, namely the funding entities issuing the records. SciPROJ's enrichment is based on an information extraction, transformation and loading (ETL) process. This information is then reconciled with authoritative information from other source systems, contributing to the completeness and quality of the recorded data. One of the authoritative source databases is the Identify Database, a Ringgold initiative [5], which aggregates more than half a million records from organizations (academic, business and scientific) worldwide.

To ensure unequivocal recognition of the funding awarded in national science and technology ecosystem and its relation to other entities (organizations, people, projects) a pilot study was launched aiming at adopting a system of unique and persistent grant identifiers, in accordance with national and international best practices.

Pursuant to PTCRIS aims, a REST API service was developed to support sharing and reuse of SciPROJ's information as open data. Currently, SciPROJ is integrated with Dimensions, a platform that offers linked research data and a wide range of solutions for discovery and analytics [6]. Also, SciPROJ with a service of the PTCRIS ecosystem - **CIÊNCIAVITAE**, the new scientific curriculum management system [7]. In fact, nearly 12 000 funding records have been automatically imported to **CIÊNCIAVITAE**, thereby alleviating the user from the need to introduce these data manually. In 2022, more key integrations are on the pipeline, namely with OpenAIRE [8], Clarivate [9], RCAAP [10] and the European Data Portal [11]. Besides this use cases, SciPROJ will also be the foundation for an Open Access monitoring and compliance tool and a business intelligence system that aims to support decision and strategic planning in the context of scientific research information.

The development of SciPROJ is an example of excellence in materializing the interoperability between information systems based on the PTCRIS regulatory framework.

References

- [1] "PTCRIS - Portuguese Current Information System," [Online]. Available: www.ptcris.pt.
- [2] J. M. Moreira, A. Cunha e N. Macedo, 6 July 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/26308833/>.
- [3] "OpenAIRE Guidelines for CRIS Manager," [Online]. Available: <https://guidelines.openaire.eu/en/latest/cris/index.html>.
- [4] "Common European Research Information Format," [Online]. Available: <https://eurocris.org/services/main-features-cerif>.
- [5] "Ringgold," [Online]. Available: <https://www.ringgold.com/>.
- [6] "Dimensions," [Online]. Available: <https://www.dimensions.ai/>.
- [7] "CIÊNCIAVITAE," [Online]. Available: www.cienciavitae.pt.
- [8] "OpenAIRE," [Online]. Available: <https://www.openaire.eu/>.
- [9] "Clarivate," [Online]. Available: <https://clarivate.com/>.
- [10] "RCAAP," [Online]. Available: www.rcaap.pt.
- [11] "European Data Portal," [Online]. Available: <https://data.europa.eu/en>.