14th International Conference on Current Research Information Systems, CRIS2018

Integrating a local CRIS with the PTCRIS synchronization ecosystem

António Luís Lopes*

ISCTE-Instituto Universitário de Lisboa, Lisboa 1649-026, Portugal

Abstract

The scientific community, in particular, and society, in general, should have easy access to the innovation and knowledge generated by scientific research. However, the multiplicity of existing science information systems and institutional repositories does not allow for a simple and fast propagation of the results of scientific activities. In that context, and under the dictum "Add once, reuse multiple times", the PTCRIS initiative was started in 2014 with the goal of creating an integrated ecosystem for the propagation and synchronization of scientific information amongst multiple parties. This ecosystem, which includes the development of the PTCRSync framework, allows for different information systems to share and synchronize information amongst them, thus avoiding the duplication of efforts of the researchers that used to have to deposit the same information in several different systems. This paper describes the integration of the PTCRISync synchronization mechanism with the current research information system of *ISCTE-Instituto Universitário de Lisboa*, Ciência-IUL. The results and the impact of the implemented approach in terms of the quantified benefits for researchers and institutions are also presented.

Keywords: CRIS; interoperability between CRIS; synchronization; PTCRIS; PTCRISync; Ciência-IUL; ISCTE-IUL

1. Introduction

Many studies[1,2,3] have shown that the visibility of research results is of the utmost importance to generate impact and further development of said research. However, the multiplicity of information systems for managing the results of scientific research makes it hard to efficiently share the innovation and generated knowledge of research activities.

* Corresponding author. Tel.: +351-210-464-274.
  E-mail address: alsl@iscte-iul.pt

Guaranteeing the interoperability between multiple systems, thus freeing the researchers of the task of having to insert the same data into different systems (an activity that is prone to errors and duplication), is a step forward in enabling that visibility. In fact, even minor integrations between local and national research systems have shown to be beneficial not only in increasing visibility of research, but also by substantially reducing research administrative costs[4].

Let us consider the example of a researcher from one of *ISCTE-Instituto Universitário de Lisboa* (ISCTE-IUL)'s research centers (ISCTE-IUL is a public University in the city of Lisbon in Portugal). On one hand, the researcher is responsible for depositing her scientific production in the University's current research information system, Ciência-IUL (https://ciencia.iscte-iul.pt), since it is connected to the performance evaluation system. On the other hand, the researcher must also fill in this information in the institutional repository (in order to be picked up by indexing databases). Depending on which research center she is affiliated to, it may be necessary to fill in the same information into the research center's Backoffice for their website. Besides these responsibilities inside the institution, the researcher may have to additionally maintain external systems up-to-date, such as curriculum profiles and funding agencies' application forms. It is clear that this approach for disclosing research results in multiple systems (with varying difficulty or usability) consumes a considerable amount of time that could be spent doing research instead.

1.1. Integration between local systems at ISCTE-IUL

This concern has led to a concerted effort, at ISCTE-IUL, to integrate the multiple information systems in a way that researchers only have to deposit the information a single time. This information will then be shared throughout the remaining systems without the need for any other intervention from the researchers. Currently, researchers only have to fill in their research outputs into Ciência-IUL and this system is then responsible for propagating the information to the performance evaluation system, to the institutional repository and to the websites of the University's multiple research centers. This approach resulted in clear and significant improvements in the researchers' day-to-day activities, by allowing them to optimize the time dedicated to research. For the institutional repository, the integration with Ciência-IUL significantly contributed to increase the number of deposits for articles published in peer-reviewed scientific journals, thus making all of this scientific knowledge available to the society[5,6].

1.2. The global landscape

Although this effort within the systems at ISCTE-IUL has allowed achieving a new level of development and growth in preserving and disseminating the institution's scientific production, the researchers still have the responsibility to maintain up-to-date research-related information in other systems external to the institution. These range from nationwide scientific management systems to international curriculum-based portals and funding agencies' websites for applying to research project or scholarship calls. The same way that at ISCTE-IUL researchers were affected by the multiplicity of systems on which they had to deposit repeated data, at the national and international level, researchers are also faced with large heterogeneity of systems and formats in which the scientific information is filled, which leads to outdated, inconsistent or simply inexistent information for each researcher's scientific production.

2. The PTCRIS initiative and the PTCRISync framework

In this context, and under the dictum "Add once, reuse multiple times", the *Fundação para a Computação Científica Nacional* (FCCN) unit of *Fundação para a Ciência e Tecnologia* (FCT) in Portugal has bootstrapped the PTCRIS[7] initiative with the goal of creating an ecosystem that connects multiple national (and in a way, international) science management systems such that the researcher only has to deposit the information in one of the connected systems that will then synchronize the information across the remaining systems. To fulfil this vision, PTCRIS has two main goals. First, to define a regulatory framework based on the best international standards and best-practices related to scientific data representation and information sharing procedures. And second, to promote the adoption of
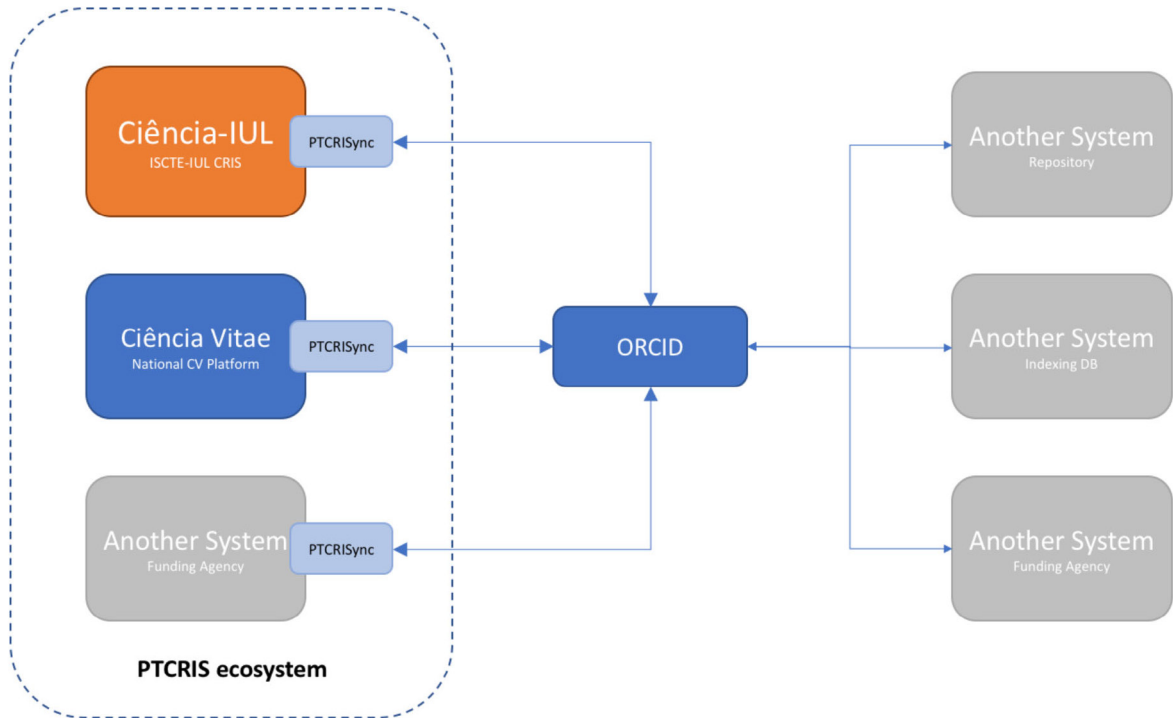
Figure 1 - The PTCRIS ecosystem and the connection to the central hub, ORCID

this framework in as many information systems as possible, in the scope of local CRIS (individual institutions) as well as the national and international levels.

One of the main products that resulted from this initiative is the PTCRISync[8] synchronization framework, a uniform and modular mechanism for synchronizing scientific productions between a local system and a central hub in the ecosystem, in this case ORCID[9] (https://orcid.org). In a typical scenario where this synchronization framework is used, a researcher is able to export the existing productions (including publications and project funding data) from a local CRIS onto ORCID (leaving these records available to be picked up by other systems connected to the ecosystem) as well as importing productions from ORCID that were added by other systems. The main goal is to achieve a level of a fully-shareable ecosystem of connected information systems through the use of a central hub, as depicted by Figure 1. Interoperability approaches between multiple systems have been tried before[4,10], but they are usually one-time or one-way exporting mechanisms that do not consider the need to constantly update one's curriculum or scientific profile.

The PTCRISync framework not only specifies the methods that all ecosystem members must use to ensure the consistency of synchronization, but also specifies the algorithm that defines the order and the way these methods should be invoked. To facilitate the integration of local systems with this framework, a software library in Java that complies with the synchronization algorithm specification was developed (the source code is available at GitHub: https://github.com/fccn/PTCRISync). This software library was used directly in Ciência-IUL (since this system was also developed using Java) to allow it to synchronize with ORCID. A SaaS (Software-as-a-Service) version of the PTCRISync synchronization mechanism is under development. This will allow other systems to integrate with the PTCRIS ecosystem without having to use the Java software library.

2.1. General synchronization principles

In order for an institution's research information system to be able to be a part of the PTCRIS ecosystem, it has to obey the principles established by the PTCRISync framework. This guarantees the consistency of the synchronization between all the members of the ecosystem. One of the fundamental principles is based on the way the system interacts with the user in order to trigger the synchronization process. Each member system must be able to collect the necessary information from the users (namely their ORCID identifier and which productions they wish to import/export from/to ORCID), then trigger the synchronization process and interact with the researchers when there are decisions that need to be made as a result of the synchronization.

Another fundamental principle approaches the way the system should execute the two first steps in the synchronization mechanism: exporting and importing. The *exporting* step (which consists of sending the information marked as to be synced in the system onto ORCID) should be executed before the *importing* step to ensure the consistency of the synchronized data, *i.e.*, in order to guarantee that more recent data that was added/edited from the side of the local system is reflected on ORCID's side. Therefore, when the *importing* step is executed next, we already know that the data collected from ORCID's side contains the most up-to-date data, preventing the occurrence of inconsistencies in the data.

3. Integrating Ciência-IUL with PTCRISync

In the context of the pilot project for the development of PTCRISync, ISCTE-IUL has integrated its current research information system, Ciência-IUL, with the PTCRIS ecosystem. This integration allows for all researchers in the institution to automatically synchronize their scientific production with ORCID and thus with any other system that may be a part of the same ecosystem.

In order to ensure the successful integration of Ciência-IUL with the PTCRIS ecosystem, there was a need to make some adjustments in the system. First off, a new section in the researchers' Backoffice dedicated to the ORCID synchronization (through the PTCRISync mechanism) was created. This new section was added because the set of guidelines in the PTCRIS ecosystem for the way member systems should interact with the users - not only at the applicational level, but also at the usability and user experience level - were too specific to allow integrating the synchronization mechanism directly into the section of the Backoffice responsible for managing productions. Therefore, the creation of this new section for dealing directly with the synchronization process allowed for the system to fully comply with framework guidelines.

The PTCRISync framework requires that certain information is maintained on the side of each local system, namely the ORCID account authorization token and the record of which productions are to be synced for each researcher. In order to be able to monitor the system, it is also recommended to keep a record of statistics regarding the use of the synchronization mechanism, by registering the number of imported/exported productions for each researcher. Therefore, it was necessary to create the data models that allow to store the referred information in Ciência-IUL.

In order to interact with the PTCRISync framework, it was also required to make the translation between Ciência-IUL's data schema (for publications and projects) and the framework's schema (which is actually the ORCID schema). Considering that most systems nowadays use some form or subset of a well-known standard for depicting scientific productions, this translation is usually straightforward.

Before starting the synchronization mechanism between Ciência-IUL and ORCID through the PTCRISync framework, the system must first establish the connection between the researcher's profile in Ciência-IUL and the ORCID profile. This step is not executed by the PTCRISync framework and it is the sole responsibility of each local system to execute this step, in order to retrieve the authorization token required to operate under the name of the researcher (when applying changes to the ORCID profile). In this case, this step is triggered by showing the button to connect to the researcher's ORCID profile, as depicted in Figure 2.
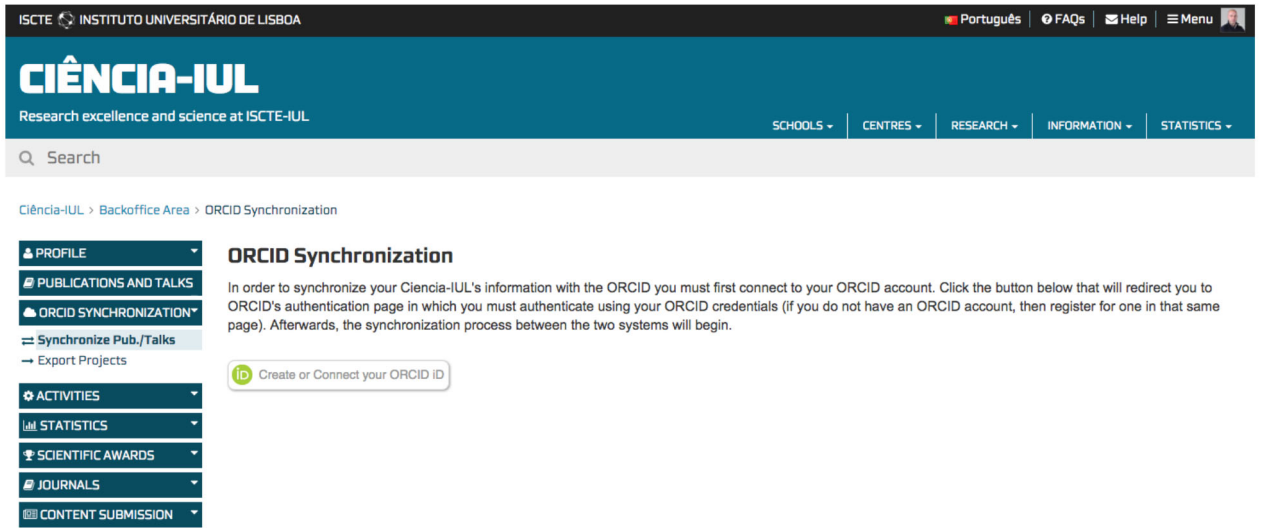
Figure 2 - Button to connect to ORCID profile in Ciência-IUL

After clicking the "Create or Connect your ORCID iD" button, the researcher is redirected to the ORCID authentication page where she has the chance to authorize Ciência-IUL's access to the account by signing into ORCID. If the researcher hasn't previously created an account, at this point, the opportunity to create an account is given. After having given authorization, the researcher is redirected back to Ciência-IUL where the synchronization mechanism
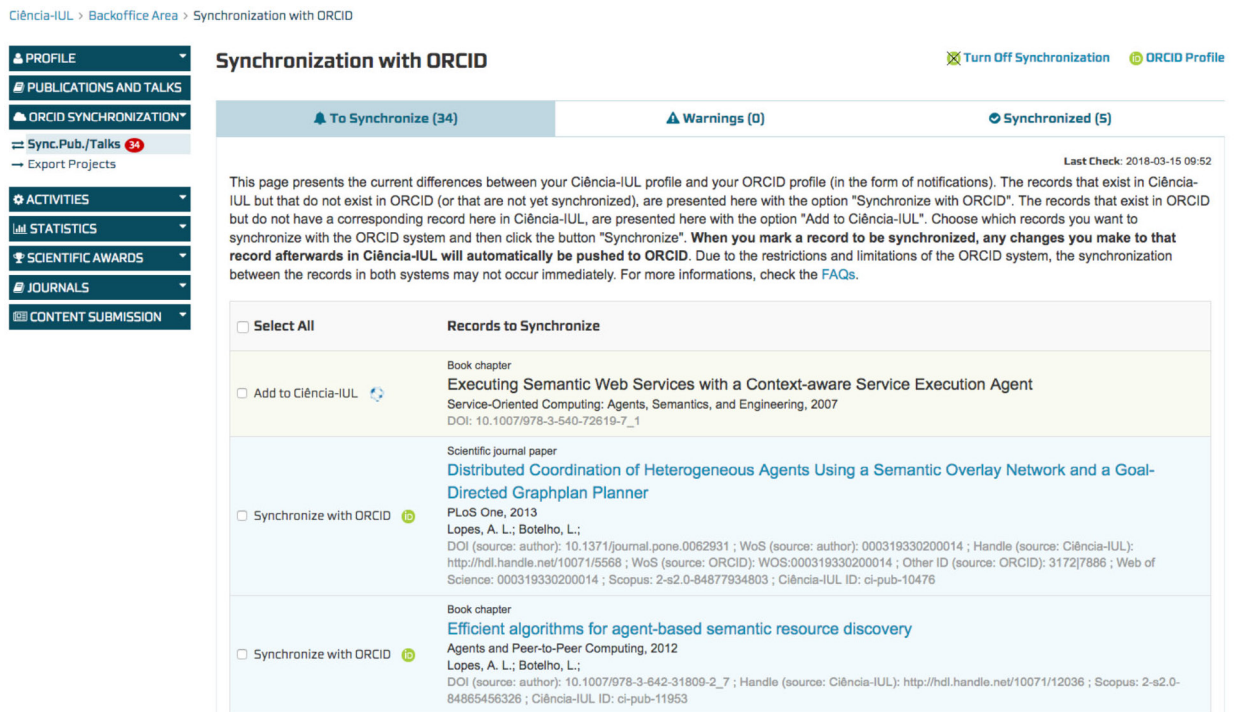

Figure 3 - Difference between the Ciência-IUL profile and the ORCID profile

between the two systems is triggered and the differences between the two profiles are presented (see Figure 3):

- Publications that exist in ORCID but that do not exist in Ciência-IUL. These are the ones showing with the suggestion "Add to Ciência-IUL"
- Publications that exist in Ciência-IUL but that do not exist in ORCID. These are the ones showing with the suggestion "Synchronize with ORCID"

As depicted in Figure 3, the researcher has the possibility to decide what to do by (un)checking the notifications that resulted from the comparison done by the synchronization mechanism between the Ciência-IUL profile and the ORCID profile of the researcher. After the researcher selects and confirms which actions should be taken (regarding which productions should be synchronized or not) the system triggers a new run of the synchronization mechanism in order to actually transfer the corresponding records from one system to the other (according to the decisions by the researcher).

3.1. More systems in the PTCRIS ecosystem

Ciência-IUL started being a part of the PTCRIS ecosystem in February 2017 and has, since then, allowed researchers at ISCTE-IUL to save considerable time while updating their scientific production information in ORCID. The next system that will become a part of the PTCRIS ecosystem is *Ciencia Vitae* (https://www.cienciavitae.pt). *Ciencia Vitae* will be the new national curriculum management system in Portugal and, once it is fully functioning, it will already have the same synchronization mechanism that allows for researchers from ISCTE-IUL to update their scientific curriculum at *Ciencia Vitae* without the need to manually deposit the information all over again (if they have previously added that data in Ciência-IUL and synchronized it with ORCID).

4. Results

We can estimate the time and financial impact that this approach has had at ISCTE-IUL since February 2017 (the time when we first integrated with the PTCRIS ecosystem) and that will have in the future as new systems are added to the ecosystem and more researchers start using this feature. At the current date (October 2018), more than 430 researchers have used this feature to synchronize more than 21000 records (including publications and project funding) with ORCID. This represents around 66% of the total scientific production stored in Ciência- IUL.

In previous usability tests done in our CRIS, we observed that users, on average, took a bit more than 3 minutes to add a new production to their profile. So, if we estimate that each record takes an average of 3 minutes to be deposited on one of these information systems, this feature has allowed the researchers to save more than 63 000 minutes (1 050 hours, or 131 work-days) adding information into ORCID.

As more systems are added to the ecosystem, the more time is saved adding this information into the new systems. If we consider that the average annual income for a researcher in Portugal is 29000€ (the working hour would correspond to 15€), these are the real and potential gains just for ISCTE-IUL's researchers:

Table 1. Real and potential gains for this approach.

	Real Gains	Potential Gains		
	(1 system, ORCID)	(Number of systems added to the ecosystem)		
Currently	**1 050 h**	2 100 h	3 150 h	4 200 h
380 researchers; 19 000 records	**15 750 €**	31 500 €	47 250 €	63 000 €
In the future	1 600 h	3 200 h	4 800 h	6 400 h
1 350 researchers; 32 000 records	24 000 €	48 000 €	72 000 €	96 000 €

It is important to note that these numbers do not consider the need for the work to be replicated by the multiple co-authors of a publication.

The values depicted in Table 1, which represent both real and potential gains for the institution, are a clear motivation for other institutions to also integrate their systems into the PTCRIS ecosystem. If we perform this analysis at the national level, considering the number of researchers and records that would need to be synced between multiple systems, these gains could reach to half a million euros.

5. Conclusion

The global landscape for sharing scientific information is characterized by the multiplicity of information systems and the duplication of efforts by researchers when depositing research output information into these systems. The PTCRIS project, with its synchronization framework, PTCRISync, aimed to change that landscape by creating guidelines and regulations based on international standards and best-practices regarding representation and sharing of scientific information. The adoption of this regulatory framework by ISCTE-IUL's current research information system has proven to be effective in saving time and resources that can now be used for what really matters: research and generation of scientific knowledge. With the adoption of this framework by other national and international systems, we can build a new landscape in which sharing scientific information is fast and efficient, allowing us to really have open science to the world.

References

1. Lawrence, S., 2001. Free online availability substantially increases a paper's impact. Nature, 411(6837), p.521.
2. Harnad, S. and Brody, T., 2004. Comparing the impact of open access (OA) vs. non-OA articles in the same journals. D-lib Magazine, 10(6).
3. Antelman, K., 2004. Do open-access articles have a greater research impact?. College & research libraries, 65(5), pp.372-382.
4. Klausen, M.H., 2017. Even minor integrations can deliver great value–A case study. Procedia Computer Science, 106, pp.153-159.
5. Maria João Amante, Susana Lopes, Bruno Marçal, and Teresa Segurado. A interoperabilidade entre o repositório e um sistema CRIS: o caso do ISCTE-IUL. Cadernos BAD, (2):83–93, 2014.
6. Maria João Amante, Susana Lopes, Bruno Marçal, and Teresa Segurado. Partilhar ciência: o repositório do ISCTE-IUL e a interoperabilidade com um sistema CRIS. In 12º Congresso Nacional BAD, 2015. Associação Portuguesa de Bibliotecários Arquivistas e Documentalistas, 2015.
7. André Santos, José Carlos Ramalho, Miguel Ferreira, Hélder Silva, Luis Faria, João Mendes Moreira, Eloy Rodrigues, and José Carvalho. PT-CRIS: Um miradouro sobre o universo científico nacional. In 12º Congresso Nacional BAD, 2015. Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas (APBAD), 2015.
8. João Mendes Moreira, Alcino Cunha, and Nuno Macedo. An ORCID based synchronization framework for a national CRIS ecosystem. F1000Research, 4, 2015.
9. Laurel L. Haak, Martin Fenner, Laura Paglione, Ed Pentz, and Howard Ratner. Orcid: a system to uniquely identify researchers. Learned Publishing, 25(4):259–264, 2012.
10. Dragan, I., Lidija, I. and Bojana, D.S., 2014. Multi-interoperable CRIS repository. Procedia Computer Science, 33, pp.86-91.