

Integrating a local CRIS with the PTCRIS synchronisation ecosystem

António L. Lopes¹

ISCTE - Instituto Universitário de Lisboa, Lisboa, Portugal
alsl@iscte-iul.pt

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 synchronisation of scientific information amongst multiple parties. This ecosystem, which includes the development of the PTCRSync framework, allows for different information systems to share and synchronise 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 PTCRSync synchronisation 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.

1 Introduction

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. Let us consider the example of a researcher from one of ISCTE-IUL's research centres¹. On one hand, the researcher is responsible for depositing her scientific productions in the University's current research information system, Ciência-IUL², 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 centre she is affiliated to, it may be necessary to fill in the same information into the research centre'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

¹ISCTE-Instituto Universitário de Lisboa is a public University in the city of Lisbon in Portugal

²<https://ciencia.iscte-iul.pt>

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 centres. This approach resulted in clear and significant improvements in the researchers day-to-day activities, by allowing them to optimise 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[1][2].

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 research-related information up-to-date 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 FCCN³ unit of FCT⁴ has bootstrapped the PTCRIS[5] 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 synchronise 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 this framework in as many information systems as possible, in the scope of local CRIS (individual institutions) as well as at the national and international levels.

One of the main products that resulted from this initiative is the PTCRISync[4] synchronisation framework, a uniform and modular mechanism for synchronising scientific productions between a local system and a central hub in the ecosystem, in this case ORCID⁵[3]. In a typical scenario where this synchronisation framework is used, a researcher is able to export the existing productions⁶ 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.

The PTCRISync framework not only specifies the methods that all ecosystem members must use to ensure the consistency of synchronisation, but also specifies the algorithm that defines the order and the way these methods should be invoked. To facilitate the integration of local

³Fundação para a Computação Científica Nacional, Portugal

⁴Fundação para a Ciência e Tecnologia, Portugal

⁵<https://orcid.org/>

⁶This includes publications and project funding data.

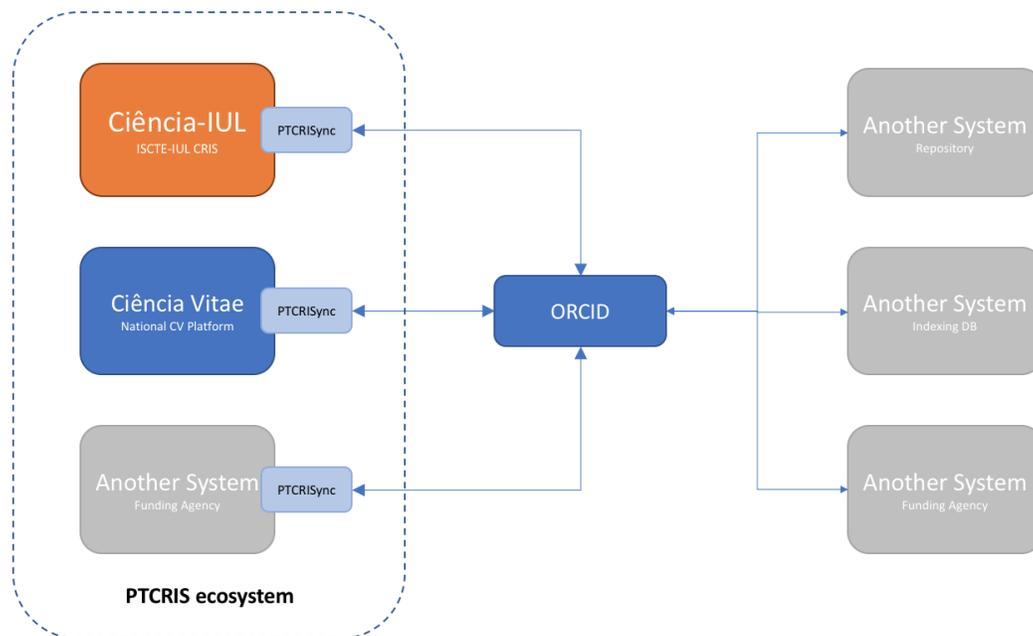


Figure 1: The PTCRIS ecosystem and the connection to the central hub, ORCID

systems with this framework, a software library in Java that complies with the synchronisation algorithm specification⁷. This software library was used directly in Ciência-IUL (since this system was also developed using Java) to allow it to synchronise with ORCID. A SaaS (Software-as-a-Service) version of the PTCRISync synchronisation 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 synchronisation principles

In order for an institution’s 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 synchronisation 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 synchronisation 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 synchronisation process and interact with the researchers when there are decisions that need to be made as a result of the synchronisation.

Another fundamental principle approaches the way the system should execute the two first steps in the synchronisation 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 synchronised data, *i.e.*, in order to guarantee that more recent data that was added/edited from the side of the local

⁷The source code is available at GitHub: <https://github.com/fccn/PTCRISync>

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 PTCRISync⁸, 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 synchronise 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 synchronisation (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 synchronisation mechanism directly into the section of the backoffice responsible for managing productions. Therefore, the creation of this new section for dealing directly with the synchronisation 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 authorisation record 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 synchronisation 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.

Before starting the synchronisation 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 authorisation 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.

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 authorise 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 authorisation, the researcher is redirected back to Ciência-IUL where the synchronisation mechanism 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 *”Synchronise with ORCID”*

⁸This project had the participation of the following institutions in Portugal: FCT—FCCN, Universidade do Minho, ISCTE-IUL and Universidade de Évora.

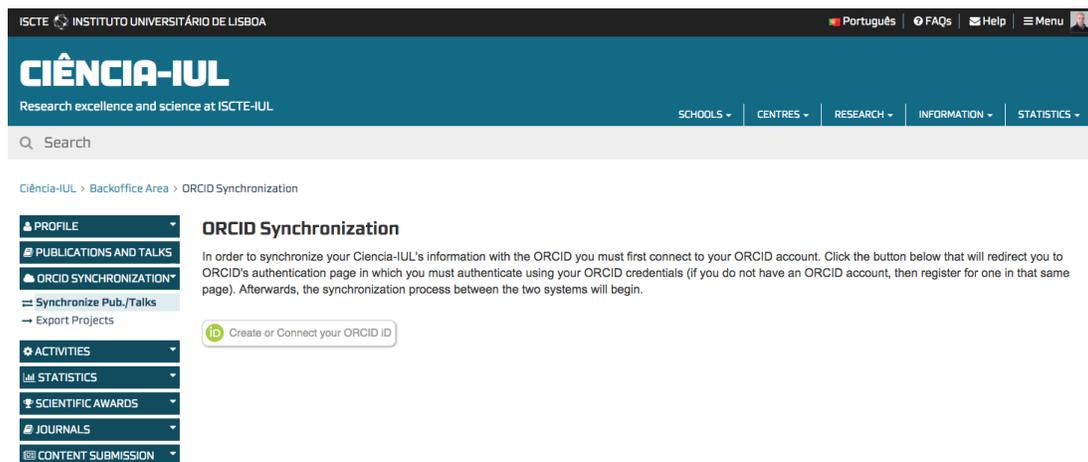


Figure 2: Button to connect to ORCID profile in Ciència-IUL

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 synchronisation mechanism between the Ciència-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 synchronised or not) the system triggers a new run of the synchronisation 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 Vitae⁹. Ciencia Vitae will be the new national curriculum management system in Portugal and, once it is fully functioning, it will already have the same synchronisation 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 Ciència-IUL and synchronised 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 (March 2018), more than 380 researchers have used this feature to synchronise more than 19000 records (including publications and project fundings) with ORCID. This represents around 34% of the total scientific production stored in Ciència-IUL. If we estimate that each record takes an average of 3 minutes to be deposited on one of

⁹<https://www.cienciavitae.pt>

Ciência-IUL > Backoffice Area > Synchronization with ORCID

Synchronization with ORCID Turn Off Synchronization ORCID Profile

To Synchronize (34) **Warnings (0)** **Synchronized (5)** Last Check: 2018-03-15 09:52

This page presents the current differences between your Ciência-IUL profile and your ORCID profile (in the form of notifications). The records that exist in Ciência-IUL but that do not exist in ORCID (or that are not yet synchronized), are presented here with the option "Synchronize with ORCID". The records that exist in ORCID but do not have a corresponding record here in Ciência-IUL, are presented here with the option "Add to Ciência-IUL". Choose which records you want to synchronize with the ORCID system and then click the button "Synchronize". **When you mark a record to be synchronized, any changes you make to that record afterwards in Ciência-IUL will automatically be pushed to ORCID.** Due to the restrictions and limitations of the ORCID system, the synchronization between the records in both systems may not occur immediately. For more informations, check the FAQs.

Select All	Records to Synchronize
<input type="checkbox"/>	<p>Book chapter</p> <p><input type="checkbox"/> Add to Ciência-IUL </p> <p>Executing Semantic Web Services with a Context-aware Service Execution Agent Service-Oriented Computing: Agents, Semantics, and Engineering, 2007 DOI: 10.1007/978-3-540-72619-7_1</p>
<input type="checkbox"/>	<p>Scientific journal paper</p> <p><input type="checkbox"/> Synchronize with ORCID </p> <p>Distributed Coordination of Heterogeneous Agents Using a Semantic Overlay Network and a Goal-Directed Graphplan Planner PLoS One, 2013 Lopes, A. L.; Botelho, L.; DOI (source: author): 10.1371/journal.pone.0062931 ; WoS (source: author): 000319330200014 ; Handle (source: Ciência-IUL): http://hdl.handle.net/10071/5568 ; WoS (source: ORCID): WOS:000319330200014 ; Other ID (source: ORCID): 31727886 ; Web of Science: 000319330200014 ; Scopus: 2-s2.0-84877934803 ; Ciência-IUL ID: ci-pub-10476</p>
<input type="checkbox"/>	<p>Book chapter</p> <p><input type="checkbox"/> Synchronize with ORCID </p> <p>Efficient algorithms for agent-based semantic resource discovery Agents and Peer-to-Peer Computing, 2012 Lopes, A. L.; Botelho, L.; DOI (source: author): 10.1007/978-3-642-31809-2_7 ; Handle (source: Ciência-IUL): http://hdl.handle.net/10071/12036 ; Scopus: 2-s2.0-84865456326 ; Ciência-IUL ID: ci-pub-11953</p>

Figure 3: Difference between the two profiles

these information systems, this feature has allowed the researchers to save more than 57000 minutes (950 hours, or 118 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 anual 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 researchers¹⁰:

	Real Gains (1 system, ORCID)		Potential Gains (Number of systems added to the ecosystem)		
	1		2	3	4
Currently	950 h		1900 h	2850 h	3800 h
380 researchers	14250 €		28500 €	42750 €	57000 €
19000 records					
In the future	1600 h		3200 h	4800 h	6400 h
1350 researchers	24000 €		48000 €	72000 €	96000 €
32000 records					

Table 1: Real and potential gains for this approach

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

¹⁰It is important to note that these numbers do not take into account the need for the work to be replicated by the multiple co-authors of a publication

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 characterised 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 synchronisation 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] 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.
- [2] 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.
- [3] 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.
- [4] João Mendes Moreira, Alcino Cunha, and Nuno Macedo. An orcid based synchronization framework for a national cris ecosystem. *F1000Research*, 4, 2015.
- [5] 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.