

## **Integrar a rede de repositórios RCAAP no ecossistema de gestão de ciência – PTCRIS - Depósitos diretos nos Repositórios a partir de sistemas CRIS**

Paulo Lopes, FCT, plopes@fccn.pt

José Carvalho, Universidade do Minho, jcarvalho@sdum.uminho.pt

Fernando Ribeiro, FCT, fernando.ribeiro@fccn.pt

Paulo Graça, FCT, paulo.graca@fccn.pt

### **Resumo da Proposta**

O PTCRIS (Portuguese Current Research Information System) é um programa desenvolvido pela Fundação para a Ciência e a Tecnologia - Unidade de Computação Científica Nacional, que tem como objetivo o desenvolvimento sustentado de um ecossistema integrado de informação de suporte à atividade científica nacional. O desenvolvimento de tal ecossistema é crucial para garantir, por um lado, a desburocratização e a simplificação dos processos administrativos, por outro, a produção, gestão e acesso a dados fidedignos, completos e atualizados, dos quais depende a efetividade e eficiência dos processos de tomada de decisão.

Os repositórios institucionais são parte importante deste ecossistema. Para garantir a integração destes sistemas, é necessário observar o cumprimento de um quadro normativo que, através da adoção de normas comuns, garanta a interoperabilidade entre os sistemas.

Torna-se assim imperioso intervir no sentido de aumentar o nível de interoperabilidade dos repositórios de forma a facilitar os processos de workflow dos utilizadores nas diversas plataformas do ecossistema.

A equipa do RCAAP tem vindo a trabalhar num conjunto de casos de usos que prevêm cinco níveis de interações: a) A possibilidade de um autor reclamar a sua identidade ou os trabalhos depositados num Repositório; b) A possibilidade de um autor efetuar um depósito num repositório a partir de um sistema de gestão de ciência (CRIS) como, por exemplo, o sistema nacional de gestão curricular; c) a adoção de mecanismos de sincronização da informação; d) a adoção de mecanismos de controlo de autoridade para identificar univocamente autores, organizações e financiamento; e, e) a definição de tarefas de curadoria com o objetivo de atualizar a informação existente nos repositórios de forma a dotá-la de identificadores persistentes de autores, organizações, financiamento ou de vocabulários e semânticas comuns aos restantes sistemas do ecossistema de gestão e ciência.

O trabalho desenvolvido facilita a disseminação da informação científica e alinha-se com a máxima “Introduza uma vez, reutilize múltiplas”, a principal premissa do programa PTCRIS.

Esta comunicação foca-se no caso de uso que prevê a possibilidade de um autor efectuar um depósito num repositório a partir do Sistema de Curricula Nacional, o **CIÊNCIAVITAE**.

Serão apresentados exemplos práticos do trabalho efetuado e abordadas as questões estratégicas relacionadas com esse trabalho (desafios, necessidades, cenários e metodologias).

## Tipo de Proposta

- Comunicação

## Tema da Conferência

- **Gestão de informação de Ciência e Tecnologia**
  - o CRIS – Sistemas de Gestão de informação de Ciência e Tecnologia
  - o Interoperabilidade entre sistemas de informação de apoio à atividade científica e académica
  - o Normas e diretrizes
  - o Identificadores persistentes

### Palavras-chave

Interoperabilidade; normas e diretrizes; repositórios; sistema de currículo

## Audiência

Gestores de repositórios, bibliotecários, gestores de dados de investigação, programadores, gestores de ciência, gestores de sistemas de gestão de ciência

## Proposta

A gestão de ciência e tecnologia tem vindo, ao longo do tempo, a tornar-se cada vez mais complexa. Contribuem para esta complexidade, entre outros fatores, o aumento do número de agentes, instrumentos e a quantidade de informação envolvidos nos ciclos do processo de investigação e dos processos do ensino superior.

Para fazer face a esta complexidade e ao seu aumento, têm vindo a ser desenvolvidos e adotados, em todo o mundo, sistemas integrados de ciência, tecnologia e de ensino.

Em Portugal, as Instituições de Ensino Superior (IES) têm feito esforços notáveis na adoção destes sistemas. Contudo, estes sistemas, com diferentes níveis de maturidade, tendem a operar de forma isolada. Neste contexto, em 2014 surge o programa PTCRIS<sup>1</sup> (Portuguese Current Research Information System). O PTCRIS é um programa desenvolvido pela Fundação para a Ciência e a Tecnologia - Unidade de Computação Científica Nacional - FCT - FCCN, que tem como objetivo o desenvolvimento sustentado de um ecossistema integrado de informação de suporte à atividade científica nacional. O desenvolvimento de tal ecossistema é crucial para garantir, por um lado, a desburocratização e a simplificação dos processos administrativos, por outro, a produção, gestão e acesso a dados fidedignos, completos e atualizados, dos quais depende a efetividade e eficiência dos processos de tomada de decisão.

---

<sup>1</sup> <https://ptcris.pt/sobre-o-ptcris/>

O RCAAP<sup>2</sup> é o principal instrumento para a implementação do Open Access (OA) em Portugal. Tem por missão promover, apoiar e facilitar a adoção do acesso aberto ao conhecimento científico em Portugal e tem como objetivo a recolha, agregação e indexação dos conteúdos científicos em acesso aberto (ou acesso livre) existentes nos repositórios institucionais das entidades nacionais de ensino superior e outras organizações de I&D.

Como serviço nacional que agrega a produção científica de vários repositórios institucionais e revistas científicas, o RCAAP constitui-se como peça importante do ecossistema PTCRIS. A adoção pelo RCAAP dos normativos de interoperabilidade definidos pelo PTCRIS são cruciais para a sua integração com outros sistemas de gestão de ciência.

Tendo sempre como ponto de referência o utilizador (autores, investigadores, instituições), o aumento da interoperabilidade na rede RCAAP alinha-se com a visão geral do PTCRIS que defende o princípio “enter once, re-use multiple”.

De forma a promover este princípio no RCAAP, foram identificados cinco blocos de casos de uso (Figura 3) relacionados com necessidades de interação entre os utilizadores e sistemas ou apenas entre sistemas (trocas de informação máquina - máquina):

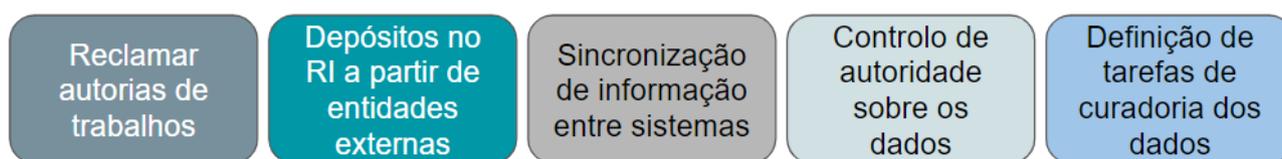


Figura 1 – Blocos de casos de uso de interoperabilidade

Cada um destes blocos prevê um conjunto de ações que promovem a interoperabilidade tendo por base o quadro normativo definido pelo PTCRIS:

**Bloco 1 – Reclamar autorias de trabalhos** – Contempla a possibilidade de um autor: a) relacionar a sua conta Ciência ID (o Ciência ID é um identificador único para investigadores em Portugal) com a sua conta local no seu repositório institucional; b) a possibilidade de reclamar os seus trabalhos no repositório; e, c) a possibilidade de importar trabalhos do RCAAP para o seu Curriculum Vitae.

**Bloco 2 – Depósito no Repositório a partir de entidades externas** – Contempla a possibilidade de se enviar uma publicação para o Repositório a partir de um sistema externo como, por exemplo, o sistema de gestão de currículos (CIÊNCIAVITAE<sup>3</sup>).

**Bloco 3 – Sincronização** – Prevê a adoção do mecanismo de sincronização PTCRISync nos repositórios e no Portal RCAAP. Este mecanismo já funciona em alguns sistemas do ecossistema, como é o caso da Plataforma Ciência Vitae. Através deste mecanismo, o ORCID funciona como um Hub central que permite fazer circular a informação entre os diversos sistemas que adotem este método de sincronização. Ao referenciar a sua produção científica (no Repositório, no seu perfil ORCID ou no seu CV) o utilizador pode fazer circular essa informação entre os vários sistemas sem ter a necessidade de reintroduzir a informação em cada um dos sistemas.

<sup>2</sup> <http://projeto.rcaap.pt/index.php/lang-pt/sobre-o-rcaap/enquadramento>

<sup>3</sup> <https://cienciavitae.pt/mais-informacao/>

**Bloco 4 – Controlo de autoridade** – Prevê a introdução de mecanismos de controlo de autoridade para autores, instituições e financiamento. Foi adicionada nos repositórios a possibilidade de relacionar um autor com o os seus identificadores (Ciência ID e/ou ORCID) no ato do depósito de uma publicação.

**Bloco 5 – Tarefas de curadoria** – Define tarefas de curadoria a executar com o objetivo de enriquecer as publicações existentes nos repositórios de forma a dotá-las com informação sobre identificação de autores, instituições e financiamento.

Esta comunicação foca-se no caso de uso do bloco 2 que prevê a possibilidade de um utilizador do CIÊNCIAVITAE poder optar por enviar para o repositório documentos inseridos manualmente ou importados de outras fontes (ORCID, SCOPUS; WoS, ...) que não se encontrem depositados.

Para esses casos foi criada a possibilidade do utilizador enviar essa informação para o seu repositório. O desenvolvimento desta funcionalidade teve em conta os seguintes requisitos:

- permitir a possibilidade de se enviar produções para os RIs;
- ser capaz de identificar o que já está depositado e o que é passível de enviar para o repositório (para evitar a criação de registos duplicados);
- proporcionar um wizard associado ao processo de depósito que cubra mecanismos de autenticação entre sistemas, a seleção do repositório de destino, a seleção da coleção de destino, o carregamento do documento que suporta o depósito, o tipo de acesso a esse documento e a licença associada a esse depósito.
- ser capaz de manter um estado dos documentos enviados para depósito (Depositado; Recusado; Depósito em processo de aprovação).

Em termos técnicos optou-se pela solução disponibilizada pelo software utilizado nos repositórios baseada em métodos de envio através do protocolo Sword<sup>4</sup>. Este protocolo não suporta métodos de autenticação baseados num Identity Provider o que impediu tirar partido dos mecanismos de single sign on proporcionados pela autenticação Ciência ID implementada, tanto do lado do CIÊNCIAVITAE, como do lado dos Repositórios.

Por forma a facilitar o processo e a abrangência da funcionalidade, optou-se pela utilização de utilizadores e coleções genéricas em cada um dos repositórios. Isto agiliza o processo pois não obriga os utilizadores a mais um processo de autenticação.

A utilização da funcionalidade passa por três fases:

**A - requisitos prévios** - Existem três requisitos prévios associados à funcionalidade de depósito:

- O primeiro diz respeito à disponibilização da própria funcionalidade que só está disponível para os utilizadores onde seja possível identificar a sua filiação. Esta identificação é feita através da consulta da tabela de filiações introduzida pelo utilizador durante a construção do seu CV. Paralelamente, caso esta informação não tenha sido facultada, o sistema tenta inferir a filiação do utilizador através do identificador Handle das suas publicações ou através do domínio do endereço de mail associado à conta do utilizador;
- Outro dos requisitos prévios diz respeito à possibilidade de se associar informação sobre financiamento no acto do depósito. Os utilizadores podem fazer esta associação aos seus depósitos se, previamente, declararam no seu CV os projetos financiados em que participaram;
- O último requisito prévio tem a ver com a forma como o CIÊNCIAVITAE identifica o que já se encontra depositado. A funcionalidade de depósito só está disponível para as publicações que não se encontram depositadas. O sistema faz esta identificação recorrendo às publicações que possuem um identificador

---

<sup>4</sup> <http://swordapp.org/about/>

Handle. Uma vez que a composição do Handle possui um código que permite identificar o Repositório, todas as publicações que possuam um Handle são identificadas como depositadas e não é permitido voltar a depositá-las. Para as publicações sem Handle, o sistema procede a uma verificação, via API do Portal RCAAP, para tentar ver se se encontram depositadas.

**B - wizard de depósito** - Cumpridas as condições descritas anteriormente, a partir das publicações não depositadas é possível lançar a funcionalidade de depósito. Esta ação invoca um wizard composto por seis passos:

- **Passo 1** - Seleção do repositório - A seleção é feita a partir de uma listagem sendo que, por defeito, aparece destacado o Repositório, ou Repositórios, associados à filiação previamente declarada;
- **Passo 2** - Após a seleção do repositório, é pedido ao utilizador que selecione a coleção onde o registo deve ser depositado;
- **Passo 3** - Ficheiros e tipo de acesso - Neste passo o utilizador indica o tipo de acesso ao documento (acesso aberto, acesso embargado, acesso restrito) e anexa o ficheiro, ou ficheiros, com o texto integral da publicação;
- **Passo 4** - Informação sobre projetos - Trata-se de informação facultativa, previamente declarada pelo utilizador na construção do seu CV;
- **Passo 5** - Licença de depósito - Neste passo o utilizador deverá ler e concordar com a licença associada ao depósito;
- **Passo 6** - Validação e conclusão do depósito - Este último passo destina-se a que o utilizador possa editar e corrigir qualquer um dos passos anteriores de forma a concluir o depósito:

Depois de concluir o depósito o utilizador recebe uma mensagem a informar que o processo está concluído e que o mesmo será validado pelo gestor do Repositório antes de ser disponibilizado publicamente. Na listagem de produções a publicação fica com a indicação de “Depósito em validação”.

**C - Tarefas de validação** - As tarefas de validação ocorrem do lado do Repositório após o depósito estar concluído. Se a informação prestada estiver correta o Gestor do Repositório valida o depósito sendo o utilizador notificado, automaticamente, dessa validação. Na listagem de produções a publicação passa do estado de “Depósito em validação” para “Depositado”. Caso por qualquer motivo o depósito não seja validado, o utilizador recebe uma notificação com essa informação e a causa da não validação. Na listagem de produções a publicação passa para o estado de “Depósito rejeitado”. Nestes casos o utilizador fica com a possibilidade de repetir o processo de depósito corrigindo o que não estava correto.

## Resultados do trabalho efetuado

A disponibilização desta funcionalidade começou a ser implementada em produção a partir de janeiro de 2020 e ainda decorrer. A receção por parte das instituições foi bastante boa e a sua utilização tem vindo a fazer-se de forma gradual, à medida que a mesma é divulgada internamente e os utilizadores alertados para esta possibilidade. A 15 de março, numa fase ainda precoce da disponibilização da funcionalidade, verificavam-se um total de 346 depósitos com proveniência no CIÊNCIAVITAE distribuídos pelos seguintes repositórios:

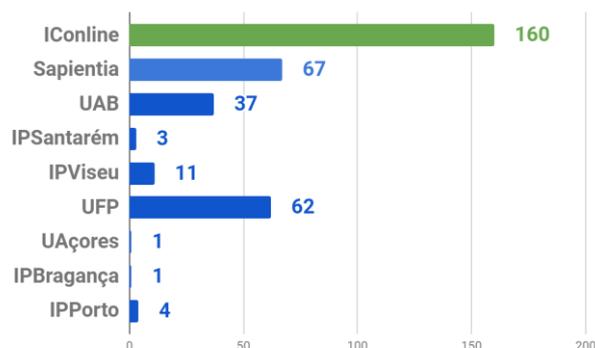


Figura 2 - Depósitos nos repositórios a partir do CIÊNCIAVITAE

## Trabalho futuro

O plano de trabalhos para os anos 2020 / 2021, prevê, sobre este tópico, as seguintes atividades principais:

- Continuar a parametrização da funcionalidade de depósito nos repositórios da rede RCAAP que a pretendam e que tenham condições técnicas para a receber;
- Iniciar os desenvolvimentos da mesma funcionalidade no ambiente do DSpace7. Esta nova versão do software para repositórios suporta uma utilização exaustiva de métodos de gestão via APIs. Um desses métodos irá suportar mecanismos de autenticação baseados num Identity Provider. Com isto será possível promover mecanismos de single sign on entre o CIENCIAVITAE e os Repositórios e, nesses casos, dispensar a autenticação e a coleção genérica.

## Conclusões

O objetivo de integrar a rede RCAAP no ecossistema de gestão de ciência PTCRIS tem esbarrado em inúmeras dificuldades, nomeadamente com o facto de as funcionalidades dos softwares de gestão de repositórios e revistas estarem concentradas na gestão de publicações, sem estarem preparados para gerir outras entidades importantes do ecossistema de gestão de ciência como é o caso dos autores, das organizações ou do financiamento. De forma a ultrapassar estas dificuldades foram adotadas duas grandes linhas de ação:

- Por um lado, trabalhar conjuntamente com a comunidade responsável pelos principais softwares utilizados pelo RCAAP de forma a influenciar o seu ciclo de desenvolvimento para que os mesmos integrem funcionalidades de gestão de novas entidades e adotem normas internacionais de classificação dos dados;
- Por outro, modificar e adaptar os softwares, ferramentas e processos em utilização de forma a que possam responder às necessidades de interoperabilidade desejadas.

Pretende-se que este trabalho represente a visão de uma rede organizada de repositórios institucionais que combina um alinhamento nacional e internacional em relação às diretrizes e organização das informações dos diferentes sistemas. Os desenvolvimentos focam-se na reutilização da informação, não apenas para o gestor de repositórios, mas também, e sobretudo, para o próprio utilizador, integrando e permitindo a sincronização de informações de diferentes fontes. Por fim, introduz o conceito de gestão de novas entidades, para além das publicações científicas, através dos diferentes sistemas.

Espera-se que a adoção de normativos nacionais e internacionais comuns aos diversos sistemas de gestão de ciência e a gestão destas novas entidades (autores, organizações e financiamento) pelos repositórios tragam benefícios a diversos níveis:

- Para os autores - Fácil acesso a informação relevante sobre oportunidades de financiamento, pessoas, parceiros e/ou concorrentes. Permite identificar projetos, publicações e infraestruturas científicas. Gestão simplificada da sua informação através da introdução da informação uma vez e reutilização em múltiplas em sistemas do ecossistema PTCRIS. Agilização da compilação de dados para relatórios a partir de informação previamente introduzida no ecossistema.
- Para as organizações - Facilitação da atividade de reporting uma vez que a integração de sistemas facilita a reunião de informação autoritativa, completa e fidedigna diversa sobre a prática científica tornando mais célere a produção de relatórios. Facilitação do acesso a indicadores/medidas do desempenho permitindo uma avaliação do desempenho de investigadores /instituições mais fácil e rápida. Otimização dos processos de financiamento uma vez que a existência de informação autoritativa, completa e fidedigna e atualizada sobre a atividade de investigação permite otimizar, quer do ponto de vista de concessão, quer de obtenção, o processo de financiamento.