
Integrar uma rede de repositórios no ecossistema de gestão de ciência nacional

José Carvalho

Universidade do Minho

jcarvalho@sdum.uminho.pt

João Mendes Moreira

FCT/FCCN

jmm@fccn.pt

Cátia Laranjeira

FCT/FCCN

catia.laranjeira@fccn.pt

Fernando Ribeiro

FCT/FCCN

fernando.ribeiro@fccn.pt

Paulo Graça

FCT/FCCN

paulo.graca@fccn.pt

Paulo Lopes

FCT/FCCN

plopes@fccn.pt

Resumo

PTCRIS (Portugal Current Research Information System) é um programa que visa a criação e desenvolvimento sustentado de um ecossistema integrado de informação nacional para apoiar a gestão da investigação, de acordo com os melhores padrões e práticas internacionais.

Para alcançar essa visão, o PTCRIS tem dois objetivos principais. O primeiro é definir um quadro regulamentar com base nas melhores normas e práticas internacionais. O segundo é promover a adoção de tal quadro nos diversos sistemas de informação, tanto a nível nacional (incluindo RCAAP) e local (ao nível institucional).

Este artigo relata o contexto, estratégia e trabalho desenvolvido até agora para tornar a rede repositórios portugueses (RCAAP), com 50 repositórios, compatíveis com o ecossistema de investigação nacional (PTCRIS), mediante a adoção do seu quadro normativo.

Palavras-chave: Repositórios; Rede Nacional; CRIS; Gestão

Integration of a repository network on a national science management ecosystem

Abstract

PTCRIS (Portugal Current Research Information System) is a program that aims at the creation and sustained development of a national integrated information ecosystem to support research management, according to the best international standards and practices.

In order to achieve such vision, PTCRIS has two main goals. The first one is to define a regulatory framework based on the best international standards and practices. The second is to foster the adoption of such framework in the various information systems, both national (including RCAAP) and local (at the institution level).

This paper reports the context, strategy and work developed thus far to make the Portuguese repositories network (RCAAP), with more than 50 repositories, compliant with the national research ecosystem (PTCRIS) by adopting its regulatory framework.

Keywords: Repositories; National Network; CRIS; Management

Introdução

Este trabalho descreve o papel e a importância das normas e diretrizes na integração de repositórios institucionais dentro dos ecossistemas mais amplos para melhor gerir a investigação, bem como para melhor apoiar o acesso aberto, dados abertos e uma ciência aberta.

Este documento relata o contexto, estratégia e trabalho desenvolvido até agora, para tornar a rede portuguesa de repositórios (RCAAP), com 50 repositórios, compatíveis com o ecossistema de investigação nacional (PTCRIS), adotando o seu quadro normativo.

A primeira seção descreve o ecossistema nacional de gestão da investigação (PTCRIS), a rede nacional de repositórios (RCAAP) assim como o quadro normativo PTCRIS.

A segunda secção descreve os desafios, necessidades, cenários, metodologia e as ferramentas utilizadas para implementar rapidamente o quadro normativo PTCRIS na rede de repositórios.

A terceira secção concentra-se no trabalho e nos resultados alcançados no projeto piloto.

1-PTCRIS

O PTCRIS é um programa iniciado em maio de 2014 pela Fundação para a Computação Científica Nacional (FCCN), uma plataforma de alto desempenho para desenvolvimento e teste de aplicações e serviços avançados de comunicações da Fundação para a Ciência e Tecnologia (FCT), a agência pública nacional para a ciência, tecnologia e inovação. O objetivo principal do PTCRIS é assegurar a criação e o desenvolvimento sustentado de um ecossistema nacional integrado de informação de suporte à atividade científica desenvolvida em Portugal, de acordo com as melhores normas e práticas internacionais.

Para este fim, PTCRIS visa:

1. Definir o quadro normativo a adotar pelos diversos sistemas de informação
2. Coordenar a integração de sistemas da FCT de acordo com o quadro normativo
3. Coordenar os sistemas externos de integração (nacional e internacional) com a FCT de acordo com o quadro normativo
4. Apoiar e promover a utilização de sistemas PTCRIS dentro da comunidade.

O cumprimento destes objetivos traduz-se em inúmeros benefícios para a comunidade de Ciência e Tecnologia, nomeadamente o acesso facilitado a informação fidedigna, completa e atualizada sobre a atividade científica; a simplificação dos processos de gestão, reporte e tomada de decisão; a otimização do processo de financiamento e o aumento da visibilidade e acessibilidade dos resultados da atividade científica portuguesa, potenciando o seu uso e impacto na comunidade científica nacional e internacional.

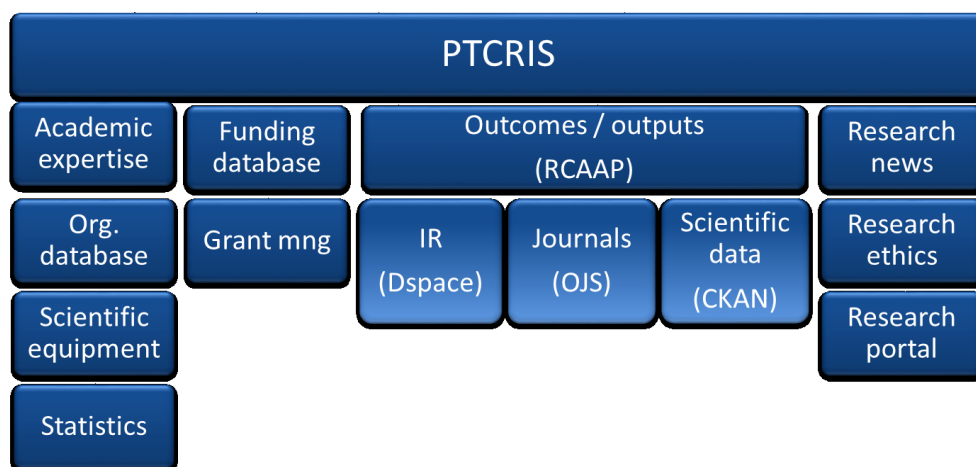


Figura 1: Componentes do ecossistema PT-CRIS

RCAAP

Como se pode observar na Figura 1, os módulos chave do ecossistema PTCRIS são os produtos e os resultados. Em Portugal, estes componentes são coordenados dentro da atividade do projeto RCAAP (Repositório Científico de Acesso Aberto de Portugal).

O RCAAP 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. Um dos principais serviços consiste em armazenar, disponibilizar e preservar a produção científica dos diversos repositórios institucionais a nível nacional. O RCAAP oferece ainda uma ampla gama de serviços para cumprir a sua missão OA. O serviço eletrónico mais conhecido é o serviço de alojamento de repositórios institucionais (SARI), mas há outros serviços prestados pelo RCAAP, tais como: alojamento de revistas científicas (SARC), repositórios de dados científicos (SARDC) e um repositório partilhado (Comum). Os resultados da investigação destes recursos são agregados no Portal RCAAP.

O modelo de serviço pelo qual se optou garante a realização do trabalho comum pela equipa de especialistas do projeto, libertando as instituições para a realização das atividades core. Como resultado da aplicação deste modelo, regista-se um custo de operação reduzido (em resultado das economias de escala e do trabalho especializado), um prazo de implementação de novos serviços também reduzido e permite que os gestores dos repositórios e respetivas equipas se foquem nas atividades nucleares. O projeto RCAAP procura ainda consolidar uma visão holística e integrada do movimento de acesso aberto promovendo ações de divulgação, promoção e formação e desenvolvendo parcerias nacionais e internacionais para a expansão dos serviços disponibilizados.

Quadro Normativo PTCRIS

Para garantir que todos os componentes do PTCRIS funcionem sem problemas, um quadro regulamentar comum teve de ser definido. Com base em estudos e análises de *state-of-the-art* foram adotadas as seguintes normas e diretrizes:

1. Interoperabilidade de Dados – um dos principais resultados da PTCRIS é conseguir a integração efetiva entre elementos do ecossistema. Para este fim, tanto o perfil de aplicação (com base em CERIF-XML), como o formato de dados, a semântica e classificação de dados são essenciais;
2. Identificadores únicos – usando identificações exclusivas, persistentes e internacionais para investigadores / docentes (ORCID), organizações (ISNI / Ringgold), etc.

3. Sincronização de dados – Os dados do ecossistema residem atualmente em vários sistemas nacionais distribuídos. Para garantir a consistência dos dados, os protocolos de sincronização devem ser utilizados e neste contexto é sugerido o PTCRISync. O PTCRISync é definido com um protocolo para apoio à sincronização de publicações entre sistemas por intermédio do sistema ORCID. O seu principal benefício é o de evitar muito do trabalho repetitivo necessário para manter o perfil do autor atualizado em diferentes plataformas. O modo de funcionamento consiste num mecanismo que exporta para o ORCID todas as publicações selecionadas para sincronização pelo utilizador num sistema CRIS. Posteriormente todas as alterações efetuadas a estas publicações serão automaticamente exportadas para o ORCID, mantendo-se desta forma a informação no ORCID permanentemente atualizada. Quando novos identificadores únicos (DOIs, Handles, etc) são adicionados a uma publicação no ORCID que esteja em modo de sincronização, estes serão automaticamente adicionados à publicação respetiva no sistema CRIS;

4. Privacidade e Proteção de Dados – Para construir indicadores de confiança depende da existência de dados abrangentes, completos, consistentes e confiáveis. No entanto, a necessidade de coletar, gerir e utilizar esta informação pode estar em desacordo com os direitos de proteção de dados.

2– O Desafio

O desafio era integrar os serviços RCAAP, nomeadamente repositórios institucionais, no ecossistema de gestão da investigação português (PTCRIS). Isto teve de ser feito respeitando alguns princípios: código aberto, não disruptivo e com uma abordagem sustentável.

Para atingir este fim foram considerados dois cenários: S1) adaptar soluções existentes; S2) mudança da rede de repositórios institucionais (IR) para um software, tanto quanto possível, compatível com o quadro normativo PTCRIS.

Depois de algumas análises tornou-se claro que a solução mais rentável e sustentável era a utilização do DSpaceCRISⁱ como solução que está em conformidade com muitos requisitos do quadro normativo PTCRIS regulamentar, ou seja, integra com o sistema ORCID e inclui o formato CERIF-XML.

A rede de repositórios RCAAP tem 50 repositórios, sendo 28 deles geridos em regime de SaaS. Para testar esta abordagem, decidiu-se executar um piloto com três instituições. Desta forma proporciona-se um ambiente seguro para testar a abordagem proposta.

Atualmente, está prevista uma abordagem metodológica que irá guiar o projeto piloto e preparar as ferramentas, integrações e métodos que irão apoiar as três instituições piloto.

Um dos desafios deste projeto é manter o "trabalho em série" e a "cadeia produtiva" que nos permitem hoje gerir todos os serviços RCAAP (incluindo os 28 repositórios) com 2 homens-mês.

Além dos aspetos metodológicos, deverão ainda ser mantidos os níveis de compatibilidade com o depósito legal de teses e dissertações, a monitorização da política da FCT, assim como integrações com outros serviços e sistemas integrados na rede RCAAP.

Resultados

O projeto piloto em desenvolvimento no projeto RCAAP relativamente à implementação de três instâncias do DSpaceCRIS foi iniciado através de uma candidatura aberta para a qual foram selecionadas as instituições: Universidade do Algarve, Universidade Aberta e a Egas Moniz – Cooperativa de Ensino Superior CRL.

Enquadrada na metodologia de trabalho, foi desenvolvida uma instância master que agrega todos os desenvolvimentos e alterações ao código base da plataforma e que depois é instanciado nas 3 instalações ou mais no futuro.

A metodologia adotada incluiu uma fase de preparação, com instalações e configurações técnicas mas também de contacto inicial com as instituições para definição das tarefas, expectativas e informação necessária.

Posteriormente avançou-se para a importação da informação dos repositórios existentes para o DSpaceCRIS no sentido de reaproveitar toda a informação existente.

Foram consideradas duas fases de organização da informação, uma primeira fase com foco nas publicações, pessoas, organizações e projetos, e uma segunda focada nos eventos, revistas, e outro tipo de entidades que possam vir a ser criadas.

A grande alteração face à plataforma DSpace é a possibilidade de descrever diversas entidades, implicando por si só um esforço e uma curadoria de informação mais cuidada e pormenorizada. No sentido de otimizar o trabalho, foi considerada a utilização de identificadores externos e internacionais para criar os índices das entidades em cada instância do DSpaceCRIS. Deste modo, essa informação poderá ser interpretada fora do contexto do repositório.

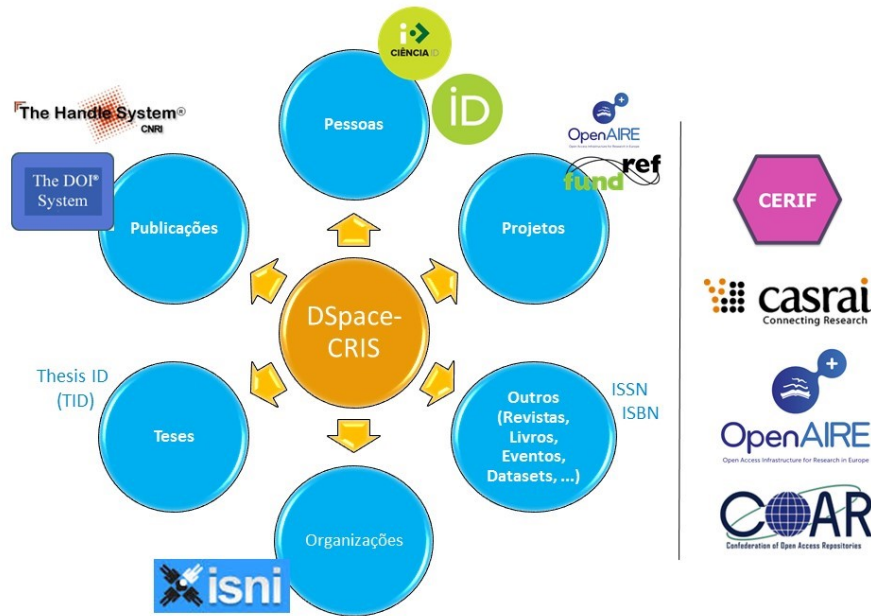


FIGURA 1 – IDENTIFICADORES DAS INSTÂNCIAS NO DSPACECRIS

Aliado aos identificadores, foram considerados vocabulários e diretrizes internacionais para a organização e partilha de informação do DSpaceCRIS.

No decorrer do processo foram identificados alguns aspetos negativos e positivos relevantes para a implementação de um repositório com DSpaceCRIS:

- **Curadoria de Informação** – É necessária a correção e uniformização de vários campos como os autores, projetos, nomes de revistas, editoras assim como lidar com situações em que um campo pode ter diversos usos conforme o tipo de trabalho descrito.
- **Bugs do Sistema** – Considerado o sistema nas versões testadas, existem ainda diversos bugs que foram resolvidos e outros por resolver, denotando ainda alguma falta de maturidade do projeto. De realçar que a equipa gestora do projeto DSpaceCRIS colaborou na resolução dos problemas.
- **Funcionalidades não implementadas** – Algumas funcionalidades não estão ainda implementadas ou não existem na sua forma final e otimizada como o interface CERIF–XML como interface máquina–máquina.
- **Versatilidade** – Considerado normalmente um aspeto positivo, o facto de a ferramenta ser altamente configurável torna também difícil a sua estabilização, principalmente com algumas lacunas de documentação.

Contudo, e tal como a primeira análise à ferramenta revelou, existem bastantes aspetos positivos a considerar:

- **Qualidade dos Metadados** – A alteração da informação descritiva dos recursos para o formato CERIF–XML em vez de Dublin Core permite e obriga a um maior cuidado na curadoria dos metadados pois neste caso a informação está ligada entre as várias instâncias.

Implica também mais tempo para que toda essa informação seja registada, mas a termo compensa pois o sistema implementa mecanismos internos de índices que ajudam no preenchimento e ligação da informação.

– **Informação pormenorizada** – O alargamento do esquema de metadados leva a que cada instância seja descrita com mais pormenor, permitindo análises mais finas e relatórios sobre a informação mais ricos, quer para o gestor, quer para os diferentes níveis de gestão na instituição (departamentos, centros de investigação ou instituição como um todo).

– **Perfil de Investigador** – O perfil do investigador é um conceito muito importante neste contexto pois existe uma relação direta entre as publicações do autor e o seu perfil, permitindo-lhe ter um papel mais ativo na construção do seu próprio perfil.

– **Mais métricas** – Como cada entidade possui indicadores e métricas próprias, permite uma monitorização mais alargada, quer das publicações, mas também da página do investigador, da página do projeto ou da página do departamento.

– **Metodologia de importação de DSpace** –> DSpaceCRIS – Este projeto piloto permite ainda definir uma metodologia, otimizada e testada para a importação da informação existente num repositório institucional com DSpace para uma instância do DSpaceCRIS. Assim que totalmente terminada, permitirá a sua aplicação a 28 instâncias do projeto SARI assim como será partilhada pela comunidade para que outras instituições possam usar essa mesma metodologia para a importação e gestão da informação na nova instância DSpaceCRIS.

– **Integrar o quadro normativo PTCRIS** – Com esta iniciativa, é mais um passo dado na consolidação da integração dos repositórios portugueses no quadro normativo PTCRIS. Sendo os repositórios sistemas base onde é organizada, gerida e validada a informação científica das instituições, será sempre o ponto inicial para enriquecer os dados que serão depois usados noutros contextos como os portais agregadores, sistemas de currículos ou o próprio ORCID.

Face ao decorrer das atividades do piloto DSpaceCRIS, será necessário melhorar alguns aspetos na plataforma no sentido de estar totalmente compatível com o quadro normativo. Contudo, para tal é necessário o contributo da comunidade para que seja melhorada a documentação e sejam afinadas configurações e comportamentos que se enquadrem nos requisitos da comunidade.

Conclusão

Os repositórios institucionais são parte importante do ecossistema de investigação. Para integrar sem problemas os repositórios institucionais num ecossistema mais amplo, estes devem cumprir com um quadro normativo. Por outro lado, nos últimos anos, os sistemas de gestão de publicações como os repositórios tornaram-se insuficientes para atender às necessidades dos gestores de informação. Além de publicações, as informações sobre autores, afiliações, financiamento e infraestruturas de investigação são cada vez mais importantes.

Nesse sentido, o desenvolvimento de sistemas integrados como o DSpaceCRIS, alinhados com orientações, diretrizes, e identificadores internacionais são fundamentais para definir uma base de informação validada, perto da fonte autoritativa que permitirá alimentar o conceito de CRIS nacional como o PTCRIS.

Com este trabalho, iremos contribuir para a integração com os sistemas externos da web e, conseqüentemente, proporcionar um melhor apoio do conceito de Ciência Aberta.

ⁱ <https://wiki.duraspace.org/display/DSPACECRIS/DSpace-CRIS+Home>