
Novo Portal RCAAP: novas funcionalidades

José Carvalho

Universidade do Minho

jcarvalho@sdum.uminho.pt

Nelson Torres

Universidade do Minho

nelson.emmanuel.p.torres@gmail.com

Paulo Lopes

Fundação para a Computação Científica Nacional (FCCN|FCT)

plopes@fccn.pt

Carla Marques

Universidade do Minho

carla.marques@sdum.uminho.pt

Raquel Truta

Universidade do Minho

raquel@sdum.uminho.pt

Resumo

O projeto RCAAP (Repositórios Científicos de Acesso Aberto de Portugal) tem como missão promover, apoiar e facilitar a adoção do acesso aberto ao conhecimento científico em Portugal e armazenar, disponibilizar e preservar a produção científica que lhe for confiada. O Portal RCAAP é a componente central do projeto e tem como objetivo a recolha, agregação e indexação dos conteúdos científicos existentes nos repositórios institucionais das entidades nacionais de ensino superior e outras organizações de I&D.

Ao longo do tempo de vida do projeto verificou-se cada vez mais a necessidade de obter mais informação estruturada dos recursos agregados. Esta carência de informação era evidente para o desenvolvimento de serviços ao nível do Portal RCAAP, desde a integração com as teses e dissertações até aos relatórios do financiamento da FCT. Neste sentido, o projeto tinha como ambição disponibilizar para a comunidade científica uma nova geração de serviços integrados no ecossistema científico nacional,

usando o quadro normativo do projeto PT-CRIS e que pudesse estar alinhado com os interesses dos diversos intervenientes (investigadores, gestores de ciência, instituições, financiadores, governo), mas também alinhado com as práticas e diretrizes internacionais no que diz respeito, por exemplo, à nova geração de repositórios promovida pela Confederação de Repositórios de Acesso Aberto (COAR). Nesta proposta, são apresentadas as melhorias efetuadas ao novo Portal RCAAP, desde a agregação de recursos, validação e transformação, ao enriquecimento e validação dos metadados recebidos dos repositórios e revistas científicas, bem como a integração com outros recursos, como o Ciência Vitae e a gestão das entidades.

Palavras-chave: Portal RCAAP, metadados, La Referência, agregador

The RCAAP Search Portal: new functionalities

Abstract

The RCAAP project (Open Access Scientific Repositories of Portugal) has the mission of promoting, supporting and facilitating the adoption of open access to scientific knowledge in Portugal and storing, making available and preserving the scientific production. The RCAAP Search Portal is the central component of the project and aims to collect, aggregate and index the existing scientific content in scientific journals, institutional repositories of national higher education institutions and other R&D organizations.

Over the lifetime of the project, there has been increasing need for more structured information on aggregate resources. This lack of information was evident for the development of services at the level of the RCAAP Search Portal, from integration with theses and dissertations to FCT funding reports. In this sense, the project's ambition was to provide the scientific community with a new generation of integrated services in the national scientific ecosystem, using the normative framework of the PT-CRIS project and that could be aligned with the interests of the various stakeholders (researchers, institutions, funders, government), but also aligned with international practices and guidelines regarding, for example, the new generation of repositories promoted by the Confederation of Open Access Repositories (COAR).

In this proposal, the improvements made to the new RCAAP Portal are presented, from resource aggregation, validation and transformation, to the enrichment and validation of the metadata received from scientific repositories and journals, as well as the integration with other resources such as Ciência Vitae and management of entities.

Key-words: RCAAP Portal, metadata, La Referência, harvester

Introdução

O projeto RCAAP (Repositórios Científicos de Acesso Aberto de Portugal) tem como missão promover, apoiar e facilitar a adoção do acesso aberto ao conhecimento científico em Portugal e armazenar, disponibilizar e preservar a produção científica que lhe for confiada. Esta missão materializa-se nos seguintes objetivos:

1. Aumentar a visibilidade, acessibilidade e difusão dos resultados da atividade académica e de investigação científica portuguesa, o que é importante para a comunidade académica e científica mas também para a sociedade em geral.

2. Facilitar a gestão e o acesso à informação sobre a produção científica nacional através do registo da produção científica em sistemas de informação específicos – os repositórios institucionais – e da sua agregação no portal RCAAP.

3. Integrar Portugal num conjunto de iniciativas internacionais através da disponibilização de um agregador e de um diretório da produção académica e científica que facilitam a interoperabilidade e a interligação com o crescente número de centros de investigação, organismos financiadores de investigação e instituições de ensino superior que, na Europa e no mundo, dispõem também de ferramentas deste tipo.

O Portal RCAAP é a componente central do projeto e tem como objetivo a recolha, agregação e indexação dos conteúdos científicos existentes nos repositórios institucionais das entidades nacionais de ensino superior e outras organizações de I&D.

Ao longo do tempo de vida do projeto verificou-se cada vez mais a necessidade de obter mais informação estruturada dos recursos agregados. Esta carência de informação era evidente para o desenvolvimento de serviços ao nível do Portal RCAAP, desde a integração com as teses e dissertações até aos relatórios do financiamento da FCT. Neste sentido, o projeto tinha como ambição disponibilizar para a comunidade científica uma nova geração de serviços integrados no ecossistema científico nacional, usando o quadro normativo do projeto PT-CRIS e que pudesse estar alinhado com os interesses dos diversos intervenientes (investigadores, gestores de ciência, instituições, financiadores, governo), mas também alinhado com as práticas e diretrizes internacionais no que diz respeito, por exemplo, à nova geração de repositórios promovida pela Confederação de Repositórios de Acesso Aberto (COAR).

Por outro lado, havia a necessidade de desenvolver novas funcionalidades para todos os intervenientes no sentido de reutilizar a informação de vários serviços integrados.

Finalmente, a relação com os investigadores deveria ser revista e repensada, quer na disponibilização de serviços de valor acrescentado, quer no desenvolvimento de aplicações com interfaces e conceitos uniformes, fáceis de entender por qualquer indivíduo para que possam vislumbrar uma visão integrada dos sistemas.

Contudo, estas linhas de orientação necessitavam de outro ingrediente fundamental para que se possam desenvolver estes tipos de serviços, e neste caso as limitações do tipo de metadados partilhados pelos repositórios e revistas estava bastante aquém do necessário e desta forma foi necessário considerar também este aspeto prático que tem implicações diretas na gestão dos recursos científicos, quer sejam repositórios institucionais, revistas científicas ou repositórios de dados.

Em certa medida, foi necessário repensar o conceito de portal agregador tal como ele existe, e desenvolver uma nova visão funcional no contexto da gestão de ciência em Portugal.

Após o desenvolvimento de um estudo para delinear os requisitos do novo Portal RCAAP, surgiram algumas linhas de orientação para os seus desenvolvimentos no sentido de suportar vários esquemas de metadados e cumprir com novas diretrizes e requisitos, como os identificadores persistentes de autores.

Das várias possibilidades apontadas no relatório, o uso de um software existente, usado no serviço La Referência, que partilha objetivos comuns surgiu como uma opção natural com mais valias para as diferentes partes. Deste modo, iniciou-se uma colaboração internacional através da celebração de um memorando de entendimento para desenvolvimentos conjuntos do software agregador. Além de vir a ser usado no contexto do projeto RCAAP, este software está já em uso em nove países da América latina, pertencentes à rede La Referência.

Considerando a base de desenvolvimento já estabelecida, foram definidas três linhas de orientação para o seu desenvolvimento:

- Adoção de normas e standards internacionais;
- Promoção de uma arquitetura modular e independente;
- Integração do quadro normativo PT-CRIS.

Os desenvolvimentos em curso do novo Portal RCAAP apresentam já um conjunto de funcionalidades que descrevemos de seguida:

Gestão do Processo de Agregação totalmente reformulado

Um dos componentes essenciais do novo Portal é o processo de agregação que permite que vários esquemas de metadados sejam integrados (oai-dc; xoai; openaire4). Por outro lado, após a sua agregação, os metadados são uniformizados para o modelo de dados local do serviço e após esse processo, seguem um workflow definido e configurável para cada tipo de recurso (revista, repositório) que apresentamos de seguida:

- Transformação

Este processo de transformação dos metadados permite efetuar correções de elementos de metadados e identificar estruturas conhecidas, como ligações de licenças Creative Commons que são integradas num campo próprio internamente. Permite ainda transformar valores conhecidos das diretrizes DRIVER ou OpenAIRE para os COAR access types e COAR document types definidos no âmbito das diretrizes OpenAIRE 4.0.

– Enriquecimento

O processo de enriquecimento é composto por vários módulos já existentes e aos quais podem ser adicionados outros. Para já existe a extração do texto integral dos trabalhos para melhorar o processo de pesquisa, a extração dos identificadores de autores e a confirmação se possuem um Ciência Vitae.

– Validação

A validação da qualidade dos metadados tem sido uma aposta importante no projeto e neste processo está também considerada. Esta validação permite identificar quais os trabalhos que não cumprem com determinadas regras definidas no âmbito das condições de agregação do Portal, informando depois o gestor do recurso sobre os erros de validação identificados.

Após este processo de validação, os trabalhos sem erros são integrados no Portal RCAAP para efeitos de pesquisa e disponibilizados nos diferentes interfaces OAI-PMH e API.

A API do Portal RCAAP é pública e permite acesso a três métodos distintos: documentos, projetos e autores. O output da interface é JSON, sendo possível solicitar os pedidos em XML ou JSONP. Sendo uma API REST, é agnóstica face à linguagem de programação a usar na integração e obedece à especificação OpenAPI (OAS). Neste contexto, foi mantida a versão 1 anterior e desenvolvida uma versão 2 com informação mais rica e pormenorizada. Está ainda a ser desenvolvida uma versão com exposição dos dados no formato CERIF.

Ao nível do interface OAI-PMH, este foi totalmente recriado usando a mesma aplicação usada no DSpace, o XOAI. Esta aplicação teve origem no plugin desenvolvido no âmbito do projeto RCAAP para um interface OAI-PMH estendido (oai extended addon). Além de disponibilizar os mesmos conjuntos de informação, tem um novo conjunto que permite a exposição do link para o texto integral dos registos nos repositórios.

Integração com Ciência Vitae

A integração com o serviço de currículos Ciência Vitae possui para já duas vertentes:

Enriquecimento de um autor com identificador

Para cada registo de um recurso local, em que o autor possui identificador, como por exemplo o ORCID, o Portal RCAAP vai questionar o Ciência Vitae para o caso do autor estar já

registado. Se o autor já possuir um currículo e permitir a partilha de dados com o Portal RCAAP, então este poderá recuperar informação básica do autor como o seu Ciência-ID e guardar no portal esse e outros identificadores associados.

Enriquecimento das ligações entre autor e publicações

No caso do autor de uma publicação possuir um CV no Ciência Vitae, o Portal vai consultar as publicações identificadas no currículo desse autor e notificar o gestor de repositório associado a esse autor de que uma determinada publicação do repositório possui uma relação direta com o autor que possui um determinado identificador. Este processo permitirá aos repositórios locais enriquecer, para já manualmente, os seus registos e aproveitar a informação curada pelo autor no seu currículo para atualizar o repositório institucional.

Estas duas integrações são apenas alguns exemplos que podem já ser desenvolvidos, existindo agora a possibilidade de criar novas relações e integrações entre estes e outros serviços que partilhem os mesmos princípios.

Gestão de Entidades

A grande mudança conceptual do novo Portal RCAAP assenta no modelo de dados adaptado para a gestão de diversas entidades. Até agora o Portal organizava e disponibilizava na prática uma entidade, as publicações. Com o tempo foi adicionada uma nova, os projetos financiados pela FCT. Nesta nova versão, foi adicionada a possibilidade de gerir também os autores e foi estendida a descrição dos recursos (repositórios e revistas) para que também eles possam ter uma caracterização mais pormenorizada.

Novo Interface

Todas estas alterações estruturais no Portal RCAAP obrigaram a uma reformulação do seu interface gráfico e funcionalidades disponíveis para os utilizadores, quer para quem pesquisa informação (investigadores), como para quem gere essa mesma informação (gestores de ciência).

Deste modo, foi desenvolvida uma pesquisa integrada das várias entidades (publicações, projetos e autores), redefinida a página de cada registo para apresentar informação mais personalizada e pormenorizada. No que diz respeito aos autores, foi criada uma página semelhante à já existente dos projetos mas neste caso com a lista de publicações e informação básica do autor.

A página de cada recurso integrado no Portal foi também alterada, possuindo agora indicadores do número de documentos agregados e a sua tipologia de acesso associada ao longo do tempo. Inclui ainda uma lista de projetos identificados em cada recurso.

Finalmente, está disponível para os gestores de repositórios um *dashboard* com os relatórios de agregação (com possibilidade de filtrar por regras de validação) e a informação dos enriquecimentos disponíveis entre um trabalho e o autor com identificador.

Todas estas mudanças no Portal RCAAP implicam também uma reformulação das condições de agregação do Portal para motivar as alterações locais nos recursos e tirar partido das novas funcionalidades. Com estas alterações, foi também criada a possibilidade de no futuro incluir novos módulos de enriquecimento, regras de validação e integrações, não só com o serviço Ciência Vitae, mas outros que possam facilitar o processo de gestão de ciência no contexto nacional.