

O Novo Portal RCAAP e o alinhamento internacional

José Carvalho

Universidade do Minho, Portugal

<https://orcid.org/0000-0003-1729-3404>

E-mail: jcarvalho@sdum.uminho.pt

Paulo Graça

FCT-FCCN, Portugal

<https://orcid.org/0000-0002-3503-4812>

E-mail: paulo.graca@fccn.pt

Lautaro Matas

La Referencia, Argentina

<https://orcid.org/0000-0003-0272-1592>

E-mail: lmatas@gmail.com

Paulo Lopes

FCT-FCCN, Portugal

<https://orcid.org/0000-0002-5550-3268>

E-mail: plopes@fccn.pt

João Mendes Moreira

FCT-FCCN, Portugal

<https://orcid.org/0000-0002-9081-2728>

E-mail: jmm@fccn.pt

Raquel Truta

Universidade do Minho, Portugal

<https://orcid.org/0000-0002-6221-9433>

E-mail: raquel@sdum.uminho.pt

Eloy Rodrigues

Universidade do Minho, Portugal

<https://orcid.org/0000-0002-7862-2681>

E-mail: eloy@sdum.uminho.pt

RESUMO

Este trabalho apresenta os desenvolvimentos e roadmap do novo Portal RCAAP desenvolvido com base no software La Referencia. A criação do novo portal teve como principais objetivos o alinhamento das funcionalidades à gestão de entidades (autores, publicações, financiamentos, etc...) mas também adequar-se aos requisitos de um portal nacional de publicações e dados de investigação. Descrevem-se o processo de definição de requisitos, bem como os desenvolvimentos integrados ou a integrar para responder aos requisitos identificados. Finalmente, este trabalho apresenta um caso de sucesso de alinhamento internacional da comunidade que gere infraestruturas de ciência nacionais e iniciativas orientadores das boas práticas como o O,penAIRE.

Palavras-chave: Portal de Pesquisa. Interoperabilidade. Metadados.

The New RCAAP Portal and the international alignment

ABSTRACT

This paper presents the developments and roadmap of the new RCAAP Portal developed based on the La Referencia software. The development of the new Portal had as its main objectives the alignment of functionalities to the management of entities (authors, publications, funding, etc.) but also to meet the requirements of a national portal of publications and research data. We describe the requirements definition process as well as the necessary developments to meet the identified requirements. Finally, this paper presents a successful case of international community alignment that manages national science infrastructures and best practice initiatives such as OpenAIRE.

Keywords: *Search Portal. Interoperability. Metadata.*

El nuevo portal RCAAP y la alineación internacional

RESUMEN

Este artículo presenta los desarrollos y la hoja de ruta del nuevo Portal RCAAP desarrollado en base al software La Referencia. El desarrollo del nuevo Portal tenía como objetivos principales alinear las funcionalidades con la gestión de las entidades (autores, publicaciones, financiación, etc.) pero también cumplir con los requisitos de un portal nacional de publicaciones y datos de investigación. Describimos el proceso de definición de requisitos, así como los desarrollos necesarios para cumplir con los requisitos identificados. Finalmente, este documento presenta un caso exitoso de alineamiento de la comunidad internacional que gestiona las infraestructuras científicas nacionales y las iniciativas de mejores prácticas como OpenAIRE.

Palabras clave: *Portal de búsqueda. Interoperabilidad. Metadatos*

TEXTO

No contexto do projeto RCAAP verificou-se, sobretudo nos últimos anos, cada vez maior necessidade de obter informação mais completa e mais bem estruturada dos recursos agregados, sejam eles revistas ou repositórios. As limitações e insuficiência da informação recolhida e necessidade de mais e melhor informação era evidente para o desenvolvimento de serviços ao nível do Portal RCAAP, desde a integração com as teses e dissertações até relatórios de financiamento da FCT, passando pela ligação ao sistema nacional de gestão curricular e ao Hub do ecossistema (Orcid). Nesse sentido, o projeto tinha como ambição disponibilizar à comunidade científica uma nova geração de serviços integrados no ecossistema científico nacional, usando o quadro normativo do projeto PT-CRIS, 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, houve 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 reconsiderada, 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.

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, a adoção de um software existente, desenvolvido e utilizado pela La Referencia (a Rede Federada de Repositórios Institucionais de Publicações Científicas da América Latina e que inclui 9 nós nacionais), que partilha objetivos comuns surgiu como uma opção natural com mais valias para as diferentes partes. Iniciou-se assim uma colaboração internacional com a celebração de um memorando de entendimento para desenvolvimentos conjuntos do software agregador. Além de ser usado atualmente no contexto do projeto RCAAP, este software está já em uso em nove países da América Latina, pertencentes à rede La Referencia. O novo Portal RCAAP está praticamente igual à versão anterior para o utilizador final, contudo, a grande reformulação foi na maneira como os recursos são agregados e toda essa informação é internamente organizada.

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

- 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 desenvolvidas ou ainda em desenvolvimento 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). 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 a seguir:

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. 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 ou não um Ciência Vitae.

VALIDAÇÃO

A validação da qualidade dos metadados tem sido uma aposta importante no projeto e neste processo está também considerada. A validação permite identificar quais trabalhos 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 o 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 permitirá o 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 em face da linguagem de programação a usar na integração e obedece à especificação OpenAPI (OAS). Nesse contexto, foi mantida a versão 1 anterior e está em desenvolvimento uma versão 2 com informação mais rica e pormenorizada.

Ao nível do interface OAI-PMH, ele foi totalmente recriado usando a mesma aplicação usada no DSpace, o XOAI. 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.

DIRETRIZES OPENAIRE 4

Com a disponibilização das novas diretrizes da iniciativa OpenAIRE, toda a informação recebida pelo Portal RCAAP é mapeada para os tipos de documento, tipos de acesso, versões, idiomas, etc. do contexto das novas guidelines. Esse processo permitirá ao Portal RCAAP apresentar já essa informação de acordo com as novas diretrizes.

INTEGRAÇÃO COM CIÊNCIA VITAE

Com o aumento da qualidade e pormenor da informação gerida pelo portal, torna-se possível agora o desenvolvimento de integrações mais efetivas e automatizadas.

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

- Pesquisa e importação do Portal RCAAP para o CV

Permitindo ao autor pesquisar informação disponível no Portal RCAAP, agregada por qualquer tipo de recurso (repositório ou revista) e adicionar esse trabalho à lista de publicações, possibilitando depois, por exemplo, sincronizar com o seu perfil Orcid.

- Enriquecimento de um autor com identificador (em desenvolvimento)

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 registado. Se o autor já possuir um currículo e permitir a partilha de dados com o Portal RCAAP, então ele 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 (em desenvolvimento)

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 determinada publicação do repositório possui uma relação direta com o autor que possui 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.

Essas 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, o financiamento. Na 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.

Contudo, este conjunto de novos requisitos obrigou à reformulação da arquitetura existente e estão em curso atualmente a definição e testes do novo modelo baseado noutras aplicações e modelos de dados como o DSpace 7 ou modelo CRIS. Além dos aspetos já considerados, a nova versão do software La Referência permitirá independência tecnológica relativamente ao uso dos componentes como a base de dados ou sistema de indexação.

NOVOS INTERFACES

Todas essas 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 a mesma informação (gestores de ciência).

Assim, foi desenvolvida uma pesquisa integrada das várias entidades (publicações, projetos e autores), ainda em testes, foi 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. As alterações estarão visíveis aquando da disponibilização pública da entidade autores.

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, estará 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.

CONCLUSÕES

As necessidades identificadas numa fase inicial deram origem a muitas alterações de fundo, até de paradigma quanto à forma de gerir ciência nestes contextos, contrariando até o conceito de quanto mais simples, mais interoperável, sendo que nesse contexto particular que se definiu, quanto mais complexo, mas organizado, mais interoperável se torna. Para que a última opção seja viável, é necessário adotar regras, normas, protocolos e métodos alinhados com a comunidade e de acordo com os restantes serviços do ecossistema. Devem também todos seguir o mesmo nível de evolução técnica e de gestão para poder integrar-se adequadamente.

As 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 essas 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.

Todo esse processo envolveu outras tarefas periféricas como a definição de modelos de dados, mapeamentos de esquemas de metadados, a definição de um modelo de partilha de informação entre os nós La Referencia e RCAAP, o desenvolvimento de roadmaps e documentação conjunta, assim como a partilha de código e de desenvolvimentos.

O processo de desenvolvimento do software La Referencia no contexto do projeto RCAAP tem sido muito enriquecedor, pois permite a partilha dos problemas e a respetiva resolução deles em contextos muitos similares.

Atualmente, os desenvolvimentos e mais valias previstas no contexto do software La Referencia estão muito dependentes da informação de base dos repositórios, sendo que só poderá disponibilizar todo o seu potencial quando os próprios repositórios também disponibilizarem a informação descrita sob a forma de entidades, como as diretrizes OpenAIRE 4 preconizam. Este processo permitiu ainda alinhar um conjunto de regras, práticas e caminhos futuros relativamente às diferentes redes, La Referencia, RCAAP e OpenAIRE.