

Pecha kucha

Interoperabilidade e conformidade no ecossistema RCAAP

Interoperability and compliance in the RCAAP ecosystem

Interoperabilidade y conformidad en el ecosistema RCAAP

Paulo Lopes*

Fundação para a Ciência e a Tecnologia (FCT), Portugal

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

CiênciaVitae: <https://www.cienciavitae.pt/portal/4317-2B5C-51C5>

plopes@fccn.pt

Raquel Truta

Universidade do Minho, Portugal

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

CiênciaVitae: <https://www.cienciavitae.pt/portal/6E12-91A4-89D3>

raquel.truta@usdb.uminho.pt

* Autor correspondente.

Resumo

A Ciência Aberta consolidou-se como um paradigma estruturante da investigação científica, promovendo o acesso, a reutilização e a transparência do conhecimento. Em Portugal, este compromisso encontra-se formalizado em instrumentos legais e políticos, incluindo a Política de Acesso Aberto da Fundação para a Ciência e a Tecnologia (FCT), que estabelece requisitos específicos para publicações e dados resultantes de investigação financiada com fundos públicos. A concretização destes princípios depende fortemente da interoperabilidade entre sistemas de informação científica e da qualidade dos metadados produzidos nas fontes primárias. O RCAAP – Repositórios Científicos de Acesso Aberto de Portugal – tem desempenhado um papel central neste ecossistema, promovendo boas práticas de normalização, validação e integração da

informação científica desde o seu lançamento. Este trabalho foi progressivamente reforçado através da adoção de identificadores persistentes, da integração com sistemas nacionais de gestão da ciência e da adoção de guidelines internacionais. No âmbito de um financiamento do Plano de Recuperação e Resiliência (PRR), foram recentemente concluídos dois desenvolvimentos estruturantes para o ecossistema RCAAP. O primeiro consiste na reconstrução do *frontend* do Portal RCAAP, com base no VuFind 10, adotando um modelo orientado a entidades e relações entre publicações, autores, organizações e projetos de financiamento. O segundo corresponde ao desenvolvimento de um novo validador de metadados baseado em OAI-PMH, com perfis de validação alinhados com políticas nacionais e internacionais e com foco na utilização de identificadores persistentes. Estes desenvolvimentos visam reforçar a interoperabilidade, melhorar a qualidade dos metadados, aumentar a visibilidade da produção científica nacional e consolidar o papel do RCAAP como infraestrutura crítica para a Ciência Aberta em Portugal.

Palavras-chave:

Ciência Aberta; Interoperabilidade; Metadados; Identificadores persistentes

Abstract

Open Science has become a central paradigm in scientific research, promoting access, reuse and transparency of knowledge. In Portugal, this commitment is reflected in legal and policy instruments, including the Open Access Policy of the Fundação para a Ciência e a Tecnologia (FCT), which defines specific requirements for publications and data resulting from publicly funded research. Achieving these goals strongly depends on interoperability between scientific information systems and on the quality of metadata produced at primary sources. RCAAP – the Portuguese Open Access Scientific Repositories – has played a key role in this ecosystem, promoting best practices for metadata normalisation, validation and integration since its launch. This work has been progressively strengthened through the adoption of persistent identifiers, integration with national research information systems and alignment with international guidelines. Within the framework of funding from the Recovery and Resilience Plan (PRR), two major developments have recently been completed for the RCAAP ecosystem. The first is the complete redesign of the RCAAP Portal frontend based on VuFind 10, adopting an entity- and relationship-oriented model linking publications, authors, organisations and funded projects. The second is the development of a new OAI-PMH-based metadata validator, offering validation profiles aligned with national and international policies and focused on the use of persistent identifiers. These developments aim to strengthen interoperability, improve metadata quality, enhance the visibility of national scientific output and consolidate RCAAP as a critical Open Science infrastructure in Portugal.

Keywords:

Open Science; Interoperability; Metadata; Persistent Identifiers

Resumen

La Ciencia Abierta se ha consolidado como un paradigma fundamental de la investigación científica, promoviendo el acceso, la reutilización y la transparencia del conocimiento. En Portugal, este compromiso se refleja en instrumentos jurídicos y políticos, incluida la Política de Acceso Abierto de la Fundação para a Ciência e a Tecnologia (FCT), que establece requisitos específicos para las publicaciones y los datos derivados de investigaciones financiadas con fondos públicos. La consecución de estos objetivos depende en gran medida de la interoperabilidad entre los sistemas de información científica y de la calidad de los metadatos producidos en las fuentes primarias. El RCAAP – Repositorios Científicos de Acceso Abierto de Portugal – ha desempeñado un papel central en este ecosistema, promoviendo buenas prácticas de normalización, validación e integración de la información científica desde su creación. Este trabajo se ha reforzado progresivamente mediante la adopción de identificadores persistentes, la integración con sistemas nacionales de gestión de la ciencia y la alineación con directrices internacionales. En el marco de un financiamiento del Plan de Recuperación y Resiliencia (PRR), se han completado recientemente dos desarrollos estructurantes para el ecosistema RCAAP. El primero consiste en la reconstrucción del frontend del Portal RCAAP, basada en VuFind 10, adoptando un modelo orientado a entidades y relaciones entre publicaciones, autores, organizaciones y proyectos financiados. El segundo corresponde al desarrollo de un nuevo validador de metadatos basado en OAI-PMH, con perfiles de validación alineados con políticas nacionales e internacionales y centrados en el uso de identificadores persistentes. Estos desarrollos tienen como objetivo reforzar la interoperabilidad, mejorar la calidad de los metadatos, aumentar la visibilidad de la producción científica nacional y consolidar el papel del RCAAP como infraestructura crítica de Ciencia Abierta en Portugal.

Palabras clave:

Ciencia Abierta; Interoperabilidad; Metadatos; Identificadores Persistentes

1. Introdução

A Ciência Aberta tem vindo a consolidar-se como um paradigma estruturante da investigação científica contemporânea, promovendo o acesso, a reutilização e a transparência do conhecimento científico.

Um pouco por toda a parte têm vindo a ser adotadas políticas de Ciência Aberta. Portugal não é exceção. O compromisso político com a Ciência Aberta é tornado claro em instrumentos como a Resolução do Conselho de Ministros nº 21/2016, de 24 de março, que definiu os princípios orientadores para a implementação de uma Política Nacional de Ciência Aberta, ou o Decreto-Lei nº 63/2019, de 16 de maio, que determina que as instituições de Investigação e Desenvolvimento (I&D) devem contribuir para uma Ciência Aberta, garantindo o acesso livre e aberto ao conhecimento científico e promovendo o envolvimento e interação com a sociedade.

No caso da atividade científica e tecnológica financiada por fundos públicos, torna-se ainda mais relevante assegurar as práticas de Ciência Aberta.

A Fundação para a Ciência e a Tecnologia (FCT), enquanto entidade financiadora do sistema científico e tecnológico nacional, está a adotar instrumentos que promovam a prática da Ciência Aberta no quadro da atividade científica por si financiada (FCT, 2023). Contribuem para esse esforço a adoção da Política de Acesso Aberto a Publicações Científicas Resultantes de Investigação Financiada pela FCT e da Política sobre a Disponibilização de Dados e outros Resultados de investigação científica financiada pela FCT, com as quais se pretende garantir a publicação em acesso aberto de obras resultantes de financiamentos da FCT e a disponibilização dos dados resultantes de investigação financiada pela FCT em acesso aberto.

As vantagens da disponibilização de publicações em Acesso Aberto são amplamente reconhecidas. Os seus benefícios incidem, de forma direta, sobre a investigação, facilitando a interdisciplinaridade e aumentando a visibilidade do conhecimento produzido, mas também causam impacto a nível social e económico, levando a investigação a profissionais de outras áreas, a empresas e a públicos interessados.

Portugal é reconhecido na Europa pela sua política inovadora em matéria de Acesso Aberto, tendo os Repositórios Científicos de Acesso Aberto de Portugal (RCAAP) um papel essencial no aumento da visibilidade do conhecimento produzido pelas universidades portuguesas, assim como na rápida expansão dos repositórios institucionais de Acesso Aberto a nível nacional (RCAAP, 2024).

A concretização destes propósitos depende, em grande medida, da capacidade de assegurar a interoperabilidade entre sistemas de informação científica, permitindo a correta identificação, ligação e reutilização de publicações, autores, organizações e financiamento.

A interoperabilidade assenta na utilização de metadados normalizados e de identificadores persistentes, amplamente reconhecidos como elementos essenciais para a integração de sistemas de gestão de ciência e infraestruturas de agregação (Bilder, 2015).

O RCAAP, desde o seu lançamento em 2008, tem vindo a promover boas práticas de classificação, normalização e validação da informação científica agregada a partir de repositórios e revistas, assegurando o alinhamento com políticas nacionais e internacionais.

2. Trabalho efetuado

Essas práticas foram reforçadas com o surgimento do programa PTCRIS em 2014. Este programa promove a integração e interoperabilidade dos sistemas de informação sobre a atividade científica em Portugal, através da adoção de normas comuns e da disponibilização de serviços de valor acrescentado que oferecem uma visão holística da ciência em Portugal.

Em 2016, iniciou-se um trabalho relacionado com a integração do RCAAP com a plataforma Ciência Vitae. Este trabalho incidiu na associação de identificadores únicos como Open Researcher and Contributor ID (ORCID) e Ciência ID a autores com o objetivo de facilitar a circulação de informação entre as duas plataformas: (a) utilizadores do Ciência Vitae passaram a poder importar as suas publicações a partir do Portal RCAAP; e, (b) Utilizadores do Ciência Vitae passaram a poder enviar para o repositório da sua instituição produções que declararam no seu CV e que não se encontravam depositadas.

Foram também adotadas as orientações do OpenAIRE (OpenAIRE, 2023) e configurada uma integração com a base de dados de financiamento do OpenAIRE para relacionar artigos científicos com projetos financiados.

Em 2022, com a adoção da nova versão do software para repositórios DSpace7 no serviço SARI, foi introduzido nos repositórios o conceito de entidades relacionadas. Passou a ser possível, desde que a informação esteja presente nos registos depositados, relacionar autores com publicações, autores com organizações, publicações com organizações, organizações com projetos financiados, projetos financiados com publicações e/ou autores. Estas relações são a base para, por exemplo, monitorizar a produção científica de determinado autor, de determinada organização ou a produção científica produzida no âmbito de um projeto financiado (fundamental para monitorizar a política de acesso aberto da FCT).

Todos estes desenvolvimentos ao nível dos repositórios foram também acompanhados por desenvolvimentos no sistema agregador (Portal RCAAP) de forma a possibilitar que o Portal expusesse estas relações nas suas interfaces de API (essencial para garantir a troca de informação entre sistemas).

Este trabalho conheceu desenvolvimentos recentes através de uma oportunidade de financiamento no âmbito do Plano de Recuperação e Resiliência (PRR). Teve como objetivo modernizar o ecossistema RCAAP, reforçando a sua capacidade de assegurar interoperabilidade, conformidade com políticas de acesso aberto e qualidade dos metadados agregados. Para isso, foram concretizados dois eixos estruturantes.

2.1. Reconstrução completa do *frontend* do Portal RCAAP com base no VuFind 10

Foi concebida uma nova interface para o Portal RCAAP, recorrendo à tecnologia VuFind 10, que introduz um modelo de navegação semântico e relacional. Este novo *frontend* permite:

- Organizar e apresentar a informação científica de forma orientada a entidades, destacando ligações explícitas entre:
 - autores (com possibilidade de integração de identificadores ORCID);
 - organizações (com inclusão de ROR - Research Organization Registry e identificação precisa das afiliações);
 - projetos de financiamento;
 - publicações e conjuntos de dados (*datasets*).
- Melhorar processos de descoberta e filtragem, oferecendo experiências de pesquisa mais intuitivas e alinhadas com modelos internacionais;
- Garantir maior consistência e uniformização da visualização dos metadados agregados pelos repositórios, contribuindo para uma perceção mais clara da produção científica nacional.

O desenvolvimento incluiu trabalhos de adaptação do VuFind às especificidades do RCAAP, criação de novos componentes visuais, definição de relações entre entidades e extensões ao modelo de dados interno.

2.2. Desenvolvimento de um novo validador de metadados baseado em OAI-PMH

O novo validador foi totalmente redesenhado, tanto ao nível tecnológico como conceptual, para responder às atuais exigências da Ciência Aberta. Entre os principais avanços destacam-se:

- Reimplementação completa do processo de colheita OAI-PMH, permitindo validações mais robustas, diagnósticos mais detalhados e deteção de problemas estruturais na origem.
- Criação de perfis de validação distintos e configuráveis, alinhados com:
 - Política de Acesso Aberto da FCT;

- Requisitos do Plano S para financiamentos;
- *OpenAIRE Guidelines*;
- Validação específica para conjuntos de dados (*datasets*) e respetivos esquemas;
- Verificação da presença e correta utilização de identificadores persistentes (Digital Object Identifier (DOI), ORCID, ROR, Handle);
- Geração automática de relatórios de conformidade, com indicação clara de requisitos obrigatórios, recomendados e opcionais;
- Adoção de um modelo extensível, permitindo evoluções futuras, integração de novos perfis de validação e adaptação às mudanças das políticas europeias.

O desenvolvimento foi acompanhado de testes com repositórios e revistas científicas reais do RCAAP, ajustando perfis e regras segundo os desafios identificados na prática.

3. Resultados esperados

Espera-se que os desenvolvimentos realizados no âmbito do ecossistema RCAAP produzam um impacto significativo na interoperabilidade, na qualidade dos metadados e na capacidade de visibilidade da produção científica nacional. Ao alinhar os mecanismos de validação com padrões internacionais e com as políticas de Ciência Aberta, o RCAAP ficará mais bem preparado para integrar e sincronizar a informação dos repositórios nacionais com sistemas europeus e globais, como o OpenAIRE, os registos de identificadores persistentes e os serviços de gestão da investigação. Essa interoperabilidade reforçada facilitará tanto a agregação como a disseminação da informação, tornando os repositórios portugueses mais compatíveis com infraestruturas internacionais e promovendo um fluxo mais transparente entre os diferentes sistemas envolvidos na comunicação científica.

A introdução de um validador de metadados mais robusto deverá contribuir para um aumento consistente da qualidade, rigor e completude dos registos produzidos nas fontes primárias. O mecanismo de validação funcionará como um instrumento de diagnóstico contínuo, capaz de identificar problemas estruturais e semânticos nos metadados, orientar as equipas de gestão dos repositórios para boas práticas e estimular a adoção sistemática de identificadores persistentes. Com isso, a informação agregada pelo RCAAP tornar-se-á mais fiável, precisa e capaz de sustentar processos de monitorização de políticas públicas, auditorias científicas ou análises bibliométricas.

A reformulação do *frontend* do Portal RCAAP permitirá também uma melhoria substancial da visibilidade da produção científica nacional, ao adotar uma abordagem orientada a entidades que dá maior contexto às relações entre autores, instituições, projetos e publicações. Esta forma de navegação possibilitará uma exploração mais profunda e significativa da informação científica,

destacando redes de colaboração, vínculos institucionais e dinâmicas de investigação financiada. A experiência de descoberta torna-se, assim, mais intuitiva e completa, beneficiando tanto a comunidade científica como decisores políticos e o público geral.

Além disso, a modernização do ecossistema RCAAP terá um impacto direto no reforço da conformidade com as políticas de Ciência Aberta. O novo validador e os mecanismos de análise associados permitirão monitorizar mais eficazmente o cumprimento dos requisitos definidos pela FCT, pelas orientações do Plano S e pelas recomendações europeias relativas a publicações e dados. Com esta capacidade, os repositórios poderão ajustar mais facilmente os seus processos, garantindo que a produção científica financiada com recursos públicos cumpre as exigências de acesso aberto e de identificação persistente.

Por fim, todos estes resultados deverão contribuir para consolidar o papel do RCAAP como infraestrutura nacional crítica no domínio da Ciência Aberta. A combinação de maior interoperabilidade, melhor qualidade dos metadados, aumento da visibilidade da produção científica e reforço da conformidade política posicionará o RCAAP como um elemento central na arquitetura digital da ciência em Portugal e exportador de ciência, capaz de responder às necessidades presentes e de se adaptar às necessidades futuras.

Agradecimentos

Este trabalho foi desenvolvido no âmbito do Programa Nacional de Ciência Aberta e Dados Abertos de Investigação (PNCADAI), financiado pela medida RE-C05-i08 – Ciência Mais Digital do Programa de Recuperação e Resiliência (PRR).

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

CRedit – Contribuições dos Autores

Paulo Lopes | Concetualização, Escrita – redação original, Escrita – revisão e edição
Raquel Truta | Escrita – revisão e edição

Referências

Bilder, G. L. (2015). *Principles for open scholarly infrastructure (Version 1)*. Figshare.
<https://doi.org/10.6084/m9.figshare.1314859>

Fundação para a Ciência e a Tecnologia [FCT]. (2023). *Política de Acesso Aberto da FCT*.
<https://www.fct.pt/financiamento/politicas-de-acesso-aberto>

OpenAIRE. (2023). *OpenAIRE guidelines for literature repositories (Version 4)*.
<https://guidelines.openaire.eu/en/latest/literature/index.html>

RCAAP. (2024). *Sobre o RCAAP*. <https://projeto.rcaap.pt/sobre-rcaap>