



Infraestrutura OpenAIRE: serviço para implementação do Acesso Aberto no Horizonte 2020

Pedro Príncipe Serviços de Documentação da Universidade do Minho pedroprincipe@sdum.uminho.pt

Eloy Rodrigues Serviços de Documentação da Universidade do Minho eloy@sdum.uminho.pt

Resumo

É detalhada a Infraestrutura de Acesso Aberto para a Investigação na Europa - OpenAIRE - que resulta dos desenvolvimentos no âmbito dos projetos europeus OpenAIRE e OpenAIREplus desde dezembro de 2009 até 2014. Os dois projetos que desenvolveram a infraestrutura visaram apoiar a implementação do Acesso Aberto na Europa, fornecendo os meios para promover e realizar a adoção das Políticas de Acesso Aberto da Comissão Europeia, especificamente para o 7º programa-quadro e Horizonte 2020. A infraestrutura consubstancia-se num sistema integrado de informação científica ao serviço da implementação do Acesso Aberto, construído de forma participativa com base num serviço robusto de rede de repositórios em expansão, para interligar publicações científicas, os conjuntos de dados associados e informação dos programas de financiamento.

Palavras-chave: Acesso Aberto, Repositórios, OpenAIRE

OpenAIRE infrastructure: service for implementation of Open Access in Horizon 2020

Abstract

Is detailed the Open Access Infrastructure for Research in Europe – OpenAIRE – resulting from developments in the european projects OpenAIRE OpenAIREplus since December 2009 until 2014. These two projects that developed the infrastructure aimed at supporting the implementation of Open Access in Europe providing the means to promote and realize the adoption of Open Access Policy of the European

Commission, specifically the FP7 and Horizon 2020. The infrastructure is consolidated into an integrated scientific information system for the benefit of the implementation of Open Access, built in a participatory way based on a robust service of repositories network to interconnect scientific publications, the associated data sets and funding programs information.

Key-words: Open Access, Repositories, OpenAIRE

Infraestrutura de Acesso Aberto para a investigação na Europa

A Infraestrutura de Acesso Aberto para a Investigação na Europa - OpenAIRE¹ - resulta dos desenvolvimentos no âmbito dos projetos europeus OpenAIRE e OpenAIREplus desde dezembro de 2009 até ao presente. Genericamente, estes dois projetos visaram apoiar a implementação do Acesso Aberto na Europa, fornecendo os meios para promover e realizar a adoção generalizada das Políticas de Acesso Aberto no espaço da União Europeia, conforme estabelecido no projeto-piloto Acesso Aberto da Comissão Europeia (CE) para o 7º programa-quadro, nas orientações para o Acesso Aberto do Conselho Europeu de Investigação e mais recentemente nas diretrizes de Acesso Aberto no Horizonte 2020.

O desenvolvimento desta infraestrutura, conduzida pela aplicação dos requisitos das políticas Acesso Aberto da Comissão Europeia, teve como ponto de partida a rede de repositórios de publicações científicas e académicas que resultou do projeto europeu DRIVER²). A iniciativa DRIVER, "Digital Repository Infrastructure Vision for European Research", foi um projeto dinamizado por um consórcio financiado pela União Europeia que criou constituiu uma infraestrutura de serviços e uma infraestrutura de dados para instrumentar os recursos e serviços existentes na rede de repositórios. Deste projeto resultaram as diretrizes DRIVER³ que visavam apoiar os gestores de repositórios na exposição de recursos textuais com o protocolo OAI-PMH de forma interoperável e fornecer orientação para os gestores de repositórios definirem as suas políticas locais de gestão de dados e decidirem sobre medidas de melhoria dos serviços prestados.

O arranque do projeto OpenAIRE em dezembro de 2009, para concretização de uma infraestrutura de informação em Acesso Aberto para o espaço europeu de investigação, foi o seguimento lógico e natural dos resultados da iniciativa DRIVER, na assunção de que a rede construída serviria de base para apoiar a implementação das políticas de Acesso Aberto da CE (Castelli, D., Schmidt, B., Rettberg, N., 2013). Com base na mesma arquitetura técnica da rede DRIVER, o projeto OpenAIRE começou a desenvolver a infraestrutura para apoiar a implementação das políticas de Acesso Aberto da CE, evoluindo nos anos de 2013 e 2014 para um sistema de informação científica de maior abrangência para cobrir os diferentes tipos de resultados de investigação.

Inicialmente, o projeto OpenAIRE visou apoiar e promover a adoção das políticas europeias de Acesso Aberto, nomeadamente as orientações para o Acesso Aberto do Conselho Científico do ERC e o projeto-piloto Acesso Aberto da Comissão Europeia. Nesse sentido, o OpenAIRE estabeleceu a infraestrutura de apoio aos investigadores no cumprimento das orientações de Acesso Aberto, proporcionando um amplo sistema europeu de Helpdesk, baseado numa rede distribuída de pontos de contacto nacionais e regionais em 27 países da União Europeia, e disponibilizando a infraestrutura eletrónica para a rede de repositórios de informação científica por via do portal OpenAIRE.

Para dar continuidade e expandir a infraestrutura gerada no contexto da rede OpenAIRE, foi criado o projeto OpenAIREplus. À semelhança do projeto OpenAIRE, o OpenAIREplus foi um projeto financiado pela Comissão Europeia no âmbito do 7º Programa—Quadro que se iniciou em dezembro de 2011 e se prolongará até final do ano de 2014. No âmbito do alargamento da infraestrutura foi ampliada a rede de países (de 27 para 32 pontos de contacto nacionais), e consequentemente a rede de repositórios e serviços, estando em curso o processo de amplificação da sua missão de implementação das políticas de Acesso Aberto da União Europeia.

Os Serviços de Documentação da Universidade do Minho operam em Portugal enquanto delegação nacional de Acesso Aberto, participando diretamente no projeto com adicionais responsabilidades por via da coordenação das atividades de disseminação para os países da região sul da Europa (Chipre, Grécia, Itália, Malta, Portugal, Espanha e Turquia) e ainda por via da coordenação dos serviços de suporte (sistema de helpdesk) da infraestrutura de informação OpenAIRE.

O OpenAIREplus construiu uma infraestrutura de Acesso Aberto de segunda geração, um sistema de informação científica de maior abrangência para cobrir os diferentes tipos de resultados de investigação, expandindo a rede de repositórios e a base de publicações atingidas. Presentemente o espaço de informação do OpenAIRE, inclui todas as publicações em Acesso Aberto indexadas anteriormente pela infraestrutura DRIVER, e também os metadados de conjuntos de dados científicos (datasets) de repositórios de dados, bem como informação sobre projetos financiados e organizações que realizam investigação.

Complementarmente aos parceiros para desenvolvimento técnico e para a disseminação e rede de implantação nacional, o projeto trabalhou com três comunidades científicas no domínio da gestão de dados científicos: EMBL-EBI (*European Bioinformatics Institute*) para as ciências da vida, DANS (*Data Archiving and Networked Services*) para as ciências sociais e BADC (*The British Atmospheric Data Centre*) para clima e ambiente. No sentido de maximizar o seu alcance enquanto sistema de informação científica e promover o alinhamento e colaboração com outras iniciativas e comunidades relevantes, o OpenAIRE cooperou ainda com várias organizações europeias nos domínios do Acesso Aberto, sistemas de informação para a ciência e comunicação científica e académica, tais como:

EuroCRIS (European Organization for International Research Information), EUDAT (European Collaborative Data Infrastructure), DataCite (International Data Citation Initiative), ORCID (Open Researcher and Contributor ID), COAR (Confederation of Open Access Repositories), LIBER (Ligue des Bibliothèques Européennes de Recherche), SPARC Europe (Scholarly Publishing and Academic Resources Coalition).

Sistema integrado de informação científica e rede de repositórios

O sistema de informação do OpenAIREplus, com a construção de uma infraestrutura participativa para interligar publicações científicas, os conjuntos de dados associados e informação dos programas de financiamento, afirmou-se como um serviço mais robusto baseado numa rede de repositórios em expansão que servirá de suporte à implementação dos requisitos da política Open Access da CE para o Horizonte 2020.

O desenvolvimento da rede de repositórios que integra a infraestrutura OpenAIRE, foi suportada pela disponibilização de Diretrizes OpenAIRE, definindo os requisitos que os repositórios deveriam respeitar para integrar a rede. As Diretrizes OpenAIRE foram baseadas e desenvolvidas a partir das Diretrizes DRIVER, já anteriormente mencionadas.

A primeira versão das diretrizes OpenAIRE (versão inicial de agosto de 2010 com atualização em novembro de mesmo ano – versão 1.1), foram estabelecidos para apoiar e acompanhar a execução do projeto-piloto do 7° programa-quadro, pelo que as especificações técnicas fornecidas nas diretrizes eram orientadas para os conteúdos de publicações resultantes de projetos financiados pela CE no quadro do 7°PQ.

Numa segunda fase, em resposta já aos desafios da 2ª geração da infraestrutura OpenAIRE, as diretrizes incluíram especificações no sentido do alargamento a outros financiadores, e adicionalmente à possibilidade de compatibilidade para sistemas agregadores (versão 2.0 das diretrizes publicada em outubro de 2012).

Por último, no ano de 2013, de acordo com a expansão dos objetivos do projeto e tendo em vista uma maior abrangência nos resultados científicos abarcados pela infraestrutura, as diretrizes passaram a constituir-se de forma mais ampla em três conjuntos de orientações. Foram apresentadas diretrizes com os requisitos técnicos específicos⁴: para gestores de repositórios de literatura científica, para gestores de repositórios de dados e para gestores de sistemas CRIS - *Current Research Information Systems*.

Genericamente, as diretrizes OpenAIRE para repositórios destinam-se a orientar os gestores de repositórios e sistemas de informação científica para expor o conteúdo, não só de publicações financiadas pela Comissão Europeia, como também de outras publicações em acesso aberto, independentemente do seu financiamento. Com a implementação destas diretrizes, os gestores de repositórios conferem aos seus repositórios os atributos técnicos para servirem todos investigadores no espaço europeu de investigação para o processo de

cumprimento das exigências de Acesso Aberto da CE no quadro do novo programa para a apoio da Ciência e Inovação na Europa – Horizonte 2020. Podem servir igualmente os objetivos de cumprimento de requisitos de outros financiadores de âmbito nacional ou internacional, como é exemplo a política de Acesso Aberto adotada recentemente pela Fundação para a Ciência e Tecnologia em Portugal, tendo sido já testado com sucesso o serviço para a fundação *Wellcome Trust*.

As diretrizes para gestores de repositórios de literatura científica (*OpenAIRE Guidelines for Literature repositories*⁵) correspondem na prática à versão 3 das diretrizes, na medida em que as duas primeiras versões continham apenas especificações para repositórios de publicações científicas e académicas. Estas diretrizes contêm atualmente as especificações para que os repositórios exponham por via de um conjunto (*set*) OAI-PMH - designado de OpenAIRE - todas as publicações que são resultado de financiamento da CE ou outro organismo financiador, mas também todas as publicações em Acesso Aberto.

Tal como acontece com a maioria dos agregadores de conteúdos científicos, o OpenAIRE usa o protocolo OAI-PMH para agregação dos metadados que devem estar estruturados como formato *Dublin Core* (prefixo de metadados oai_dc). As diretrizes OpenAIRE definem uma sintaxe específica usada nos valores de campos de metadados *Dublin Core* para identificar projetos, financiadores, publicações referenciadas e conjuntos de dados. Esta sintaxe assume a forma de *URIs* e é definida da seguinte forma: *namespace eu-repo*.

As diretrizes para gestores de repositórios de dados (*OpenAIRE Guidelines for Data Archives*⁶) incidem em instruções para que os repositórios exponham os seus metadados, relativos a conjuntos de dados (*datasets*) resultantes de um projeto financiado ou relacionados com um documento, por exemplo, os citados num artigo, de uma forma compatível com a infraestrutura OpenAIRE para dados. Estas diretrizes têm como base o esquema de metadados da *Datacite*⁷.

O OpenAIRE adotou o esquema definido pela organização *Datacite*⁸ como base do seu processo de agregação de metadados dos conjuntos de dados de investigação, acreditando optar assim por um esquema agnóstico do ponto de vista da área disciplinar e que possibilita a necessária interoperabilidade com recurso a um número reduzido de propriedades. A implementação das diretrizes para repositórios de dados facilita a criação de publicações enriquecidas e é um contributo decisivo na construção de uma infraestrutura de dados ligados à investigação.

As diretrizes para gestores de sistemas CRIS (*OpenAIRE Guidelines for CRIS Systems*⁹) estão definidas com base no formato normalizado *CERIF-XML*, e no qual assentará o processo de agregação de conteúdos dos sistemas CRIS – Current Research Information Systems. Este processo de agregação de metadados dos CRIS compatíveis com o OpenAIRE é um exemplo de uma troca de dados ponto-a-ponto entre os sistemas CRIS, já que a

infraestrutura OpenAIRE é compatível com a norma *CERIF* e pode ser considerada ela própria um sistema CRIS. Portanto, o formato *CERIF XML* foi a escolha óbvia como padrão para esta interoperabilidade e assim determinado nas diretrizes (Manghi, P., et al. 2012). Estas orientações técnicas especificam essencialmente um subconjunto (*subset*) específico para o OpenAIRE no modelo *CERIF* em termos de elementos de dados e sintaxes particulares do OpenAIRE.

A infraestrutura OpenAIRE está completamente operacional desde dezembro de 2012, e a sua informação é acessível não só através do Portal web OpenAIRE, mas também de forma programática via API OpenAIRE¹⁰.

Na expansão em curso desta infraestrutura de Acesso Aberto para a Europa todos os metadados de publicações anteriormente agregadas na rede DRIVER foram integrados no sistema OpenAIRE. Este processo de fusão da rede DRIVER com o OpenAIRE, ocorrido em 2013 e disponibilizado publicamente por via do Portal web OpenAIRE a partir de abril de 2014, significou um aumento do número de publicações e de repositórios agregados na infraestrutura.

No quadro desta expansão da infraestrutura, foi formalizada e adotada a uma política de agregação de conteúdos¹¹. Nos termos dessa política de conteúdos, o OpenAIRE aceita/agrega na sua infraestrutura:

- Todas as publicações em Acesso Aberto agregadas a partir de um conjunto variado de repositórios institucionais ou disciplinares, portais nacionais ou agregadores e revistas eletrónicas da Europa.
- O OpenAIRE procurará ativamente conteúdo adequado em Acesso Aberto de repositórios não-europeus.
- Sempre que possível, o conteúdo agregado deve estar relacionado com a informação de programas e projetos de financiadores para ajudar a medir o impacto do financiamento.
- Conteúdos em acesso fechado ou restrito só serão aceites quando ligados a financiamento da CE ou outro financiador nacional ou internacional parceiro do OpenAIRE.
- São ainda agregados conjuntos de dados científicos quando são resultados de um de investigação financiada devidamente identificados com um projeto, e/ou quando o conjunto de dados está ligado com uma publicação representada no espaço de informação OpenAIRE.

OpenAIRE: um serviço para implementação do Acesso Aberto no Horizonte 2020

Com base no crescente volume de dados e informações que recolhe e processa, a infraestrutura OpenAIRE disponibiliza serviços de valor acrescentado, tais como serviços integrados de descoberta de informação, interligação de dados, publicações e programas de financiamento e ainda a criação de publicações científicas enriquecidas, a um conjunto diversificado de parceiros e interessados.

Os investigadores e coordenadores de projetos são um dos principais parceiros da infraestrutura OpenAIRE e certamente dos mais interessados nos seus serviços. O OpenAIRE disponibiliza aos beneficiários de financiamento no âmbito do 7° PQ e do Horizonte 2020 um conjunto de serviços e informações para facilitar o cumprimento dos requisitos Acesso Aberto da CE. Desde logo, são disponibilizadas informações detalhadas sobre como cumprir com a política de Acesso Aberto, onde se incluem breves guias informativos e respostas a perguntas frequentes.

Em segundo lugar, o OpenAIRE oferece aos investigadores serviços e ferramentas de apoio ao cumprimento da obrigação de depósito das publicações financiadas em repositórios. No portal OpenAIRE é disponibilizada uma ferramenta para localizar um repositório institucional adequado e iniciar o processo de depósito. Para o caso do investigador interessado não encontrar um repositório apropriado (institucional ou temático), o OpenAIRE faculta o repositório Zenodo¹² para depositar as publicações e conjuntos de dados. Finalmente, o OpenAIRE integrou no seu portal um serviço para que os autores "reclamem" uma publicação como resultado de um projeto financiado, ligando a publicação disponível no Portal OpenAIRE ou localizada via DOI (*Digital Object Identifier*) ou ORCID com o projeto de que são beneficiários.

Para além dos serviços relacionados com informação de suporte e processo de depósito em repositórios, o OpenAIRE também oferece aos investigadores e coordenadores de projetos um conjunto de ferramentas relevantes para a divulgação dos seus resultados de investigação e para elaboração de listagem de publicações e relatórios de acompanhamento dos projetos. Todas as publicações automaticamente agregadas a partir dos repositórios compatíveis ou manualmente ligadas a projetos diretamente no portal, que estão corretamente identificadas com projetos do 7°PQ ou Horizonte 2020 são apresentadas no numa página web específica para cada projeto. Cada projeto tem assim uma página própria no portal OpenAIRE, destacando a informação do projeto, a listagem de publicações, os conjuntos de dados associados e alguns dados estatísticos.

As publicações ligadas a projetos e representadas na página do respetivo projeto no portal OpenAIRE podem ser incorporadas em *sites* externos, ou reutilizados para efeitos de disseminação ou monitorização, com recurso aos seguintes meios:

- A lista de publicações pode ser automaticamente incorporada no portal de projetos da CE - CORDIS¹³;
- A lista de publicações disponível para efeitos de relatórios de acompanhamento do projeto pode ser gerada em formato html ou descarregada em formato CSV;
- A lista de publicações do projeto pode ainda ser incorporada em sites externos, por exemplo do próprio projeto, das instituições envolvidas ou dos investigadores, por via de uma aplicação web (javascript) providenciada pela "App box" do OpenAIRE.

Os gestores de repositórios são outros dos parceiros relevantes para o OpenAIRE, e para os quais o projeto disponibiliza informação, ferramentas e serviços. Em primeiro lugar, é facultado a todos os gestores de repositórios e sistemas para gestão de investigação e publicações científicas e académicas, um vasto conjunto de informações e serviços para apoiarem os investigadores das suas instituições a cumprirem com os requisitos do Horizonte 2020 e 7° PQ.

Para além desta vasta quantidade de informações sobre o Open Access, o OpenAIRE fornece aos gestores de repositórios um guia detalhado sobre como tornar os seus repositórios compatíveis com a infraestrutura seguindo as diretrizes OpenAIRE. Para apoiar o processo de compatibilização e integração na infraestrutura é disponibilizada uma ferramenta de validação e registo de repositórios, útil para validar a compatibilidade com a diretrizes OpenAIRE, mas também com o formato Dublin Core. Este validador OpenAIRE¹⁴ é o modo simples para registar um novo repositório na infraestrutura ou atualizar informações dosrepositórios que já estão integrados e fornecem dados. Por último, são oferecidos ainda aos gestores de repositórios serviços web de informação sobre projetos financiados pela CE, a fim de facilitar a ligação das publicações aos projetos associados. Para facilitar a identificação dos projetos no processo de depósito, o OpenAIRE faculta a informação devidamente atualizada dos projetos, através de *addons* ou *plugins* às plataformas de repositórios mais populares.

Os gestores de ciência ou administradores de investigação podem ser igualmente beneficiários da informação de suporte disponível no portal OpenAIRE, fundamentalmente para apoiarem os investigadores das suas instituições com base nos guias disponibilizados, mas também por via da ferramenta de listagem de publicações disponível para relatório final ou de acompanhamento dos projetos. A informação detalhada sobre as políticas e requisitos de Acesso Aberto da CE, bem como sobre questões de direitos de autor, disponibilizadas pelo OpenAIRE são um recurso útil aos gestores de ciência para desenvolveram a sua atividade nas instituições de investigação e ensino.

Finalmente, os financiadores de ciência são também beneficiários dos resultados dos projetos OpenAIRE e OpenAIREplus. Os desenvolvimentos alcançados no contexto de serviços e ferramentas para a implementação das políticas de Acesso Aberto da CE no quadro do Horizonte 2020 e anteriormente do 7° PQ, podem ser direcionados e realizados

para outros financiadores, como foi aliás testado para a fundação *Wellcome Trust*. Portanto, os serviços de identificação dos projetos associados a publicações e dados, bem como as ferramentas de estatísticas e monitorização da produção científica em resultado de projetos financiados, podem ser reconfiguradas para as necessidades de outros financiadores de cariz nacional ou internacional, e deste modo contribuir para um melhor e maior desenvolvimento do Acesso Aberto no mundo.

O OpenAIRE será o serviço chave para suporte à realização do mandato de Acesso Aberto para Horizonte 2020. Para esse efeito foi já aprovado um novo projeto - OpenAIRE 2020 - que funcionará durante 42 meses a partir de janeiro de 2015, e irá continuar e expandir a infraestrutura de comunicação científica do OpenAIRE.

Combinando o potencial da rede OpenAIRE, quer da rede de repositórios, quer da rede humana de secretariados nacionais, com as capacidades técnicas do sistema integrado de informação científica, o OpenAIRE 2020 irá disponibilizar um serviço robusto de suporte às políticas de Acesso Aberto da CE, tanto ao mandato relativo às publicações, quanto ao projeto-piloto de Dados Abertos do Horizonte 2020. Finalmente, para garantir a sustentabilidade a longo prazo da infraestrutura de informação em Acesso Aberto, o OpenAIRE2020 irá desenvolver os necessários estudos e procedimentos no sentido de se estabelecer como uma entidade legal para assegurar com confiabilidade e continuidade o suporte a todos os beneficiários da infraestrutura OpenAIRE no quadro da União Europeia.

Bibliografia

Castelli, D., Schmidt, B., Retterberg, N. (2013). *Setting the Default to "Open": Realizing a Vision through Participatory Repository Infrastructures in Europe. In* Uma Década de Acesso Aberto na UMinho e no Mundo. Braga, SDUM. 978–989–98704–1–3.

Houssos, N., et al. (2014). *OpenAIRE Guidelines for CRIS Managers: Supporting Interoperability of Open Research Information through Established Standards.* Procedia Computer Science, Elsevier, Volume 33.

Manghi, P., et al. (2012). *The Data Model of the OpenAIRE Scientific Communication e-Infrastructure.* Metadata and Semantics Research, Communications in Computer and Information Science, Springer, Vol. 343.

OPENAIRE. *OpenAIRE Guidelines Wiki* [Em linha] [Consult. Set. 2014]. Disponível na Internet: <URL: http://guidelines.openaire.eu>

Príncipe, P., et al. (2014). *OpenAIRE Guidelines: Supporting Interoperability for Literature Repositories, Data Archives and CRIS.* Procedia Computer Science, Elsevier, Volume 33.

Notas

- ¹ www.openaire.eu
- ² www.driver-repository.eu
- ³ http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final__PT.pdf
- ⁴ O conjunto das 3 diretrizes OpenAIRE está disponível numa wiki em http://guidelines.openaire.eu
- $^{5}\ \underline{https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_Literature_repositories}$
- ⁶ https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_Data_Archives
- ⁷ http://schema.datacite.org
- 8 http://www.datacite.org
- $^{9}\ \underline{https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_CRIS}$
- 10 A documentação da API OpenAIRE está disponível em http://api.openaire.eu
- 11 https://www.openaire.eu/openaire-s-content-acquisition-policy/document-details
- 12 Disponível em <u>www.zenodo.org</u>
- 13 Disponível em http://cordis.europa.eu
- ¹⁴ Disponível em http://validator.openaire.eu