



PortAberta para a Ciência Aberta: portais de investigação da Universidade do Minho e do Instituto Politécnico de Bragança

Raquel Truta^a, Pedro Príncipe^b, Ricardo Saraiva^c, Eloy Rodrigues^d, Clarisse Pais^e

^aUniversidade do Minho, Portugal, raquel.truta@usdb.uminho.pt

^bUniversidade do Minho, Portugal, pedro.principe@usdb.uminho.pt

^cUniversidade do Minho, Portugal, ricardo.saraiva@usdb.uminho.pt

^dUniversidade do Minho, Portugal, eloy.rodrigues@usdb.uminho.pt

^eInstituto Politécnico de Bragança, Portugal, clarisse@ipb.pt

Resumo

Este trabalho visa apresentar o projeto PortAberta – Portal de Investigação para a Ciência Aberta – que tem como promotores a Universidade do Minho e o Instituto Politécnico de Bragança. A iniciativa surge com o propósito de colmatar uma necessidade partilhada de estabelecer um sistema Current Research Information Systems e Portal de Investigação, como ferramentas para apoiar e reforçar a afirmação na investigação e inovação, alicerçada em sistemas de software aberto, sob princípios da Ciência Aberta e alinhado com as iniciativas nacionais, assegurando sustentabilidade à operação e a sua relevância como boa prática a nível nacional e europeu. O PortAberta surgiu com o propósito de disponibilizar *Portais de Investigação* que facilitem o acesso, consulta e utilização da informação da Investigação na Universidade do Minho e Instituto Politécnico de Bragança, quer internamente pelos seus membros e órgãos, quer externamente, promovendo a visibilidade, impacto da atividade e dos resultados de Investigação e Desenvolvimento. Estes Portais de investigação serão operacionalizados através da implementação do software VIVO. A pretensão de apresentar este trabalho no Congresso da BAD prende-se com a vontade de demonstrar o impacto interno em cada uma das instituições, e no impacto externo com a interoperabilidade com serviços do ecossistema PT-CRIS e com o potencial das soluções desenvolvidas para poderem ser adotadas por outras instituições do sistema científico nacional.

Palavras-chave: Portal de Investigação, Comunicação de Ciência, Open Science, VIVO, CRIS, RIM.

Introdução

O projeto *PortAberta – Portal de Investigação para a Ciência Aberta* – resultante de uma parceria entre a Universidade do Minho (UMinho) e o Instituto Politécnico de Bragança (IPB), visa colmatar a inexistência, nas instituições, de um sistema de gestão de informação de investigação, comumente designado de CRIS (Current Research Information Systems), através da criação de um *Portal de Investigação* integrado, aproveitando a longa e relevante experiência das atividades e serviços de Ciência Aberta e de apoio à gestão da informação em operação nestas instituições.

A ciência, tecnologia e o conhecimento gerados pelas atividades de investigação científica nas instituições de ensino superior tem um elevado potencial de impacto positivo na economia e na sociedade, como é amplamente reconhecido. Esses impactos positivos são maximizados quando se

otimiza a gestão da informação (que contribui para a eficiência e a qualidade dos processos de investigação) e se promove a ampla acessibilidade, difusão, transferência e apropriação dos resultados, nomeadamente através das práticas da ciência aberta.

A gestão de informação de investigação e a ciência aberta são, pois, duas importantes e indispensáveis componentes de qualquer estratégia para promover, facilitar e ampliar os benefícios dos investimentos realizados na investigação científica.

Isto mesmo tem vindo a ser reconhecido a nível internacional, e também em Portugal. Os princípios da ciência aberta têm vindo a ser adotados, e muitas vezes acompanhados com políticas que requerem a adoção de práticas de ciência aberta (como o acesso aberto a publicações, a gestão e abertura de dados, etc.), por governos, entidades financiadoras e instituições que realizam investigação.

No que diz respeito à gestão de informação de investigação, um número crescente de instituições científicas e académicas tem vindo a estabelecer, nos últimos anos, *Portais de investigação* (com essa, ou outra designação), onde reúnem e disponibilizam publicamente um conjunto de informações relativas à sua atividade de investigação, nomeadamente o pessoal investigador, os projetos e os resultados da investigação (publicações, patentes e outros resultados). A criação dos *Portais de Investigação* baseia-se na existência de sistemas de informação de investigação, comumente designados CRIS – Current Research Information Systems, ou RIM Systems – Research Information Management Systems, que reúnem, estruturam, inter-relacionam e ligam o conjunto da informação da investigação desenvolvida pela instituição. Um estudo por inquérito¹ realizado no início de 2018 revelou que 58% (das 381 instituições de todas as regiões do mundo que responderam) tinham estabelecido CRISs e 17% estavam em processo de implementação ou contratação.

O estabelecimento de sistemas CRIS foi acompanhado por um esforço de normalização e interoperabilidade, nomeadamente através da adoção de um modelo de dados (tendo sido especificado o modelo CERIF - Common European Research Information Format:

<https://www.eurocris.org/cerif/main-features-cerif> - que se transformou num standard de facto), do uso de identificadores persistentes e da definição de vocabulários e taxonomias.

O estabelecimento de sistemas CRIS tem surgido como uma estratégia para melhorar a gestão institucional da investigação, existindo uma forte relação entre estes sistemas, o acesso aberto e a ciência aberta, nomeadamente entre os CRIS e os repositórios institucionais, que estão geralmente interligados, e frequentemente mesmo integrados.

Também em Portugal, nos últimos anos, têm vindo a ser criados ou estabelecidos CRIS em algumas instituições do ensino superior. No entanto, quer a nível local na maioria das instituições, quer a nível nacional continua a verificar-se uma forte dispersão (muitas vezes acompanhada de repetição/replicação desnecessária) de informação em vários sistemas, a inexistência ou não adoção de quadros normativos e boas práticas que facilitem a interoperabilidade e integração da informação e dos sistemas, gerando dificuldades, ineficiências e desperdícios. Reconhecendo a importância de endereçar este problema a FCT-FCCN estabeleceu a iniciativa PT-CRIS (<https://ptcris.pt/>), a criação do CIÊNCIA ID (<https://www.ciencia-id.pt/>) e o desenvolvimento do CIÊNCIAVITAE (<https://cienciavitae.pt/>). À semelhança do que acontece em outros países, as iniciativas de gestão de informação de ciência foram desenvolvidas em articulação com a agenda do acesso aberto e da ciência aberta, na qual Portugal tem investido ao longo de mais de dez anos, nomeadamente através da iniciativa RCAAP - Repositórios Científicos de Acesso Aberto de Portugal (www.rcaap.pt).

Neste trabalho daremos a conhecer o âmbito e objetivos do projeto, refletindo sobre o alcance do modelo de implementação projetado. Iremos incluir a discussão dos primeiros resultados e da realização das principais etapas já concretizadas do PortAberta, operação de modernização administrativa que se iniciou em julho de 2021 e que tem a duração de 24 meses.

Metodologia

A Universidade do Minho e o Instituto Politécnico de Bragança têm realizado percursos similares, e em vários momentos em conjunto ou em colaboração, no que diz respeito ao acesso aberto e à ciência aberta (sendo ambas instituições reconhecidas nestes domínios) e encontram-se ambas perante a necessidade de estabelecer os seus portais de investigação, enquadrados nessa experiência.

Os portais de investigação para a Ciência Aberta da Universidade do Minho e do Instituto Politécnico de Bragança estão em desenvolvimento no quadro da operação *PortAberta*, de modernização administrativa (SAMA) e com financiamento no quadro do surgiu com o propósito de disponibilizar *Portais de Investigação* que facilitem o acesso, consulta e utilização da informação da Investigação na UMinho e IPB, quer internamente pelos seus membros e órgãos, quer externamente, promovendo a visibilidade, impacto da atividade e dos resultados de I&D. Pretende constituir-se como uma ferramenta para apoiar e reforçar a afirmação na investigação e inovação, alicerçada em sistemas de software aberto, sob princípios da Ciência Aberta e alinhado com as iniciativas nacionais, assegurando sustentabilidade à operação e a sua relevância como boa prática a nível nacional e europeu. Com o seu desenvolvimento almeja-se ainda:

- Criar um Portal agregador dos produtos e resultados de Investigação, ligando a informação de recursos humanos, financeira e bibliométrica, para expor e conferir visibilidade aos perfis dos docentes e investigadores das instituições;
- Alavancar os sistemas de apoio à gestão da informação de ciência existentes para maximizar o retorno do investimento realizado nas duas instituições neste domínio e minimizar os custos da criação e operação de um CRIS;
- Fortalecer a centralidade dos repositórios institucionais na gestão de informação científica e académica e para a avaliação e progressão de carreiras na UMinho e IPB;
- Melhorar a gestão institucional da investigação através de sistemas mais articulados, promovendo a visibilidade das suas atividades contribuindo para a sua competitividade e capacidade de atrair e gerir financiamento para a investigação;
- Reforçar o potencial de interoperabilidade dos sistemas de gestão de informação institucionais, adotando normas internacionais, modelos de dados e vocabulários normalizados e identificadores persistentes;
- Promover boas práticas de gestão e abertura de dados de investigação com recurso a ferramentas de repositórios de dados, planos de gestão e cadernos de laboratório abertos, integrando conjuntos de dados de investigação como produtos para a comunicação de ciência da UMinho e IPB.

O projeto apresenta-se com duas fases, a primeira dedicada ao diagnóstico, com a realização de estudos técnicos, de diagnóstico e prospetivos para definição das especificações técnicas, análise de lacunas de interoperabilidade e identificação das soluções tecnológicas a implementar no desenvolvimento dos componentes do PortAberta. A segunda fase de implementação, com a conceção, instalação, configuração do sistema para Portais de Investigação da UMinho e IPB e desenvolvimento dos serviços de suporte à Gestão de Dados de Investigação e serviços de valor acrescentado dos repositórios de publicações académicas e científicas.

De forma sumária, a metodologia em curso para primeira fase de diagnóstico e implementação do sistema de informação para os portais, tem sido a seguinte:

1. Especificação dos requisitos do sistema e gestão de entidades de informação.
2. Análise de lacunas de interoperabilidade dos sistemas de gestão da informação da UMinho e IPB.
3. Instalação, configuração e parametrização das instâncias VIVO nas duas instituições.

4. Integração de fontes de informação internas e externas do ecossistema PTCRIS.
5. Desambiguação das entidades de informação.
6. Curadoria da informação e atualizações para maior interoperabilidade nos sistemas.

Resultados e trabalho em curso

Os trabalhos preliminares do *PortAberta* consistiram na análise de diferentes sistemas que acomodassem o desenvolvimento de *Portais de investigação*. Com base na prospeção realizada, a escolha recaiu no sistema VIVO, um software de código aberto, apoiado e desenvolvido em comunidade sob a égide da Lyrasis, para representar a atividade académica. Outro aspeto relevante para a seleção do VIVO foi a sua relação com a comunidade e entidades de sistemas de gestão de informação académica, como são o CASRAI, EuroCRIS e o ORCID.

O VIVO é um software de código aberto, apoiado pela Comunidade Lyrasis, para representar a atividade académica e científica de investigadores, instituições, centros de investigação e projetos.

Apoia a edição, pesquisa, navegação e visualização da atividade académica e científica individual e institucional, encorajando o registo académico e a avaliação do impacto dos resultados de investigação. Todos os componentes são de código aberto e a instituição possui & controla os seus dados, com perfis simplificados que recolhem dados de um número ilimitado de fontes e que inclui editor de perfis melhorado.

Na etapa subsequente à seleção da plataforma, realizou-se um processo de identificação dos requisitos do VIVO para definir as normas a utilizar e suportar na gestão de entidades, tendo por base o CERIF. Neste modelo distinguem-se entidades nucleares (Person, OrganizationUnit e Project), de resultado (ResultPublication, ResultPatent e ResultProduct), de segundo nível (Funding, Facility, Equipment, Prize, CV, Expertise, Qualification, Citation, Metrics, Event, PostalAddress, and ElectronicAddress) e de ligação (relações entre entidades de núcleo, resultado e 2^o nível). Como características, consideram-se o multilinguismo (suporte de nomes, títulos, descrições, palavras-chave, resumos, etc.) e a semântica (permite representação de tipos de relação, vistas de aplicação, de assunto, esquemas de classificação ou mapeamento entre esquemas).

Foram realizados os estudos de diagnóstico nas duas instituições para identificação das potenciais fontes de dados para integração no sistema e gestão das entidades de informação, tendo sido concretizada no estudo *E2.1 - Especificação dos requisitos do sistema e gestão de entidades* disponível publicamente. Seguiu-se um levantamento das lacunas de interoperabilidade dos sistemas de gestão de informação internos à UMinho e ao IPB tendo por base os tipos de entidades conceptuais, as características CERIF e principalmente o modelo de dados da última versão do sistema VIVO.

O projeto PortaAberta tem em curso várias outras atividades, das quais podemos destacar a instalação, a configuração e parametrização das instâncias VIVO nas duas instituições, desambiguação das entidades, desenvolvimentos com foco na interoperabilidade dos sistemas das fontes de dados institucionais, identificação de melhorias nos repositórios de publicações e de dados de investigação, adoção de boas práticas e ferramentas inovadoras no domínio do uso de identificadores persistentes, da gestão de dados de investigação (GDI) e da publicação inovadora através de serviços *overlay*.

Nos processos de GDI o foco será na disponibilização de ferramentas para elaboração de Planos de Gestão de Dados (PGD) e na utilização de cadernos de laboratório eletrónicos na atividade dos centros de investigação. O PortAberta irá dotar as instituições de uma ferramenta online para elaboração de PGD, com possibilidade de parametrização de modelos institucionais ou de unidades, publicação dos PGD em repositórios e a sua integração no repositório de dados e no Portal de Investigação. No caso dos cadernos de laboratório, serão identificadas e testadas ferramentas de Electronic Lab Notebook que se configurem como a solução adequada para um ambiente de Ciência Aberta nos centros de investigação das duas instituições.

Conclusão

No PortAberta, procurou-se estabelecer objetivos realistas, que possam ser atingidos até ao final da operação, no máximo até final de 2023, considerando que uma parte significativa dos serviços e ferramentas só estarão disponíveis para utilização na fase final da operação.

A pretensão de apresentar este trabalho no Congresso da BAD prende-se com a vontade de demonstrar o impacto interno em cada uma das instituições, mas sobretudo em tornar visível no sistema científico nacional, permitindo melhorar a disseminação e visibilidade dos recursos, atividades e resultados da investigação desenvolvida nas instituições no conjunto das instituições públicas do sistema científico nacional. Sendo um aspeto central da operação a integração e interoperabilidade com os serviços e sistemas do ecossistema do PT-CRIS, acreditamos que o PortAberta contribuirá para a adoção e reforço desses serviços e sistemas, maximizando o investimento já realizado nessas infraestruturas, não apenas pela sua adoção em duas relevantes instituições do norte do país, mas porque as soluções desenvolvidas no PortAberta podem vir a ser adotadas/adaptadas por outras instituições do sistema científico nacional.

Referências bibliográficas

Bryant, Rebecca, Clements, A., Castro, P., Cantrell, J., Dortmund, A., Fransen, J., Gallagher, P., & Mennielli, M. (2018). *Practices and Patterns in Research Information Management: Findings from a Global Survey*. OH: OCLC Research. <https://doi.org/10.25333/BGFG-D241>

Castro, P.; Shearer, K.; Summann, F. (2014). The gradual merging of repository and CRIS solutions to meet institutional research information management requirements. *Procedia Computer Science*, 33, 39–46. <http://doi:10.1016/j.procs.2014.06.007>

Ivanović, Dragan; Ivanović, L.; Bojana, D.S. (2014). Multi-interoperable CRIS Repository. *Procedia Computer Science*, 33, 86–91. <http://doi:10.1016/j.procs.2014.06.014>

JOINT, Nicholas (2008) - Current research information systems, open access repositories and libraries. *Library Review*, 57(8), 570–575. <http://doi:10.1108/00242530810899559>

Relatório PortAberta (2021). E2.1 - Especificação dos requisitos do sistema e gestão de entidades. <https://repositorium.sdum.uminho.pt/handle/1822/75515>

Relatório PortAberta (2021). E2.2 - Relatório de análise de lacunas de interoperabilidade dos sistemas de gestão de informação da UMinho e IPB. <https://repositorium.sdum.uminho.pt/handle/1822/76790>

<https://wiki.lyrasis.org/display/VIVODOC112x/VIVO+1.12.x+Documentation>

<https://eurocris.org/cerif/main-features-cerif>

<https://ptcris.pt/>

¹ Bryant, Rebecca, Anna Clements, Pablo de Castro, Joanne Cantrell, Annette Dortmund, Jan Fransen, Peggy Gallagher, and Michele Mennielli. 2018. *Practices and Patterns in Research Information Management: Findings from a Global Survey*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/BGFG-D241>