

Caracterização multidimensional do RCAAP

Multidimensional characterization of RCAAP

Caraterización multidimensional del RCAAP

Clarisse Pais | clarisse@ipb.pt

Instituto Politécnico de Bragança. Bragança, Portugal

Paula Odete Fernandes | pof@ipb.pt

UNIAG – Unidade de Investigação Aplicada em Gestão. Instituto Politécnico de Bragança. Bragança, Portugal

Rui Pedro Lopes | rlopes@ipb.pt

Instituto Politécnico de Bragança. Bragança, Portugal

Cristina Mesquita | cmmgp@ipb.pt

Instituto Politécnico de Bragança. Bragança, Portugal

Resumo

A investigação depositada nos repositórios científicos permite fundamentar e suportar o trabalho dos pesquisadores. Estes são plataformas de partilha, integração e preservação de conhecimento que ajudam a divulgar e sistematizar o desenvolvimento científico atual. A importância destas plataformas encontra-se patente nas iniciativas recentes como o Repositório Científico de Acesso Aberto de Portugal (RCAAP) ou o OpenAIRE, que integram e indexam os metadados de centenas de repositórios científicos. O impacto da disponibilização dos resultados da investigação científica, de forma aberta, gera benefícios importantes para a atividade de investigação, potenciando a inovação e a reprodutibilidade da ciência. Neste trabalho propõe-se a análise dos metadados disponíveis no portal RCAAP com o objetivo de caracterizar a investigação por tipo de documento, data, tipo de acesso e idioma.

Palavras-chave: RCAAP; Repositórios científicos; produção científica; caracterização da investigação; acesso aberto.

Abstract

The scientific research deposited in the repositories allows to substantiate and support the work of the researchers. These are sharing, integration and preservation of knowledge platforms that help to disseminate and systematize current scientific development. The importance of these platforms is evident in recent initiatives, such as the Scientific Repository of Open Access of Portugal (RCAAP) or OpenAIRE, which integrate and index the metadata of hundreds of scientific repositories. The impact of the availability of the results of scientific research, in open manner, generates important benefits for the research activity, enhancing innovation and reproducibility of science. In this work we propose the analysis of the metadata available in the portal of RCAAP, in order to characterize the investigation by document type, date, access rights and languages.

Keywords: RCAAP, Scientific repositories; scientific production; characterization of research; open access.

Resumen

La investigación depositada en los repositorios científicos permite fundamentar y soportar el trabajo de los investigadores. Éstas son plataformas usadas para compartir, integrar y preservar el conocimiento, y a su vez, ayudar a divulgar y sistematizar el desarrollo científico actual. La importancia de estas plataformas se encuentra patente en iniciativas recientes como el Repositorio Científico de Acceso Abierto de Portugal (RCAAP) o el OpenAIRE, que integran e indexan los metadatos de cientos de repositorios científicos. El impacto de la disponibilidad de los resultados de la investigación científica, de forma abierta, genera beneficios importantes para la actividad investigativa, potenciando la innovación y la reproducibilidad de la ciencia. En este trabajo, se propone el análisis de los metadatos disponibles en el portal RCAAP con el objetivo de caracterizar la investigación por tipo de documento, data, tipo de acceso y idioma.

Palabras clave: RCAAP; Repositorios científicos; producción científica; caracterización de la investigación; acceso abierto.

Introdução

Em julho de 2008 nasceu a iniciativa nacional de acesso aberto - os Repositórios Científicos de Acesso Aberto de Portugal ([RCAAP](#)), promovido Agência para a Sociedade do Conhecimento ([UMIC](#)), e operacionalizado pela Fundação para a Ciência e a Tecnologia ([FCCN](#)) com o apoio da Universidade do Minho, o qual veio a ser apresentado e lançado oficialmente em dezembro do mesmo ano¹.

Desde então registaram-se progressos assinaláveis tanto na proliferação de repositórios, como na criação de vários [serviços electrónicos](#).

Presentemente o RCAAP é um referencial a nível nacional e internacional e é portanto, um agregador de repositórios institucionais das entidades nacionais de ensino superior, e outras organizações de I&D². Neste momento agrega 126 recursos entre os quais: o portal OASIS do Brasil, um repositório de dados científicos, um repositório comum, repositórios das Universidades, dos Politécnicos, de Laboratórios e Hospitais. Desenvolveu o Serviço de Alojamento de Revistas ([SARC](#)), o Serviço de Alojamento de Repositórios Institucionais ([SARI](#)) e o Serviço de Alojamento de Repositórios de Dados Científicos ([SARDC](#)).

O portal OASIS foi integrado no RCAAP em Dezembro de 2010. Esta agregação veio consolidar o memorando de entendimento celebrado entre os dois países e o reforço de conteúdos científicos em língua portuguesa no portal RCAAP.

Neste momento, o RCAAP tem 1 454588 documentos depositados. Portugal tem 424179 e no portal OASIS tem 1 033249 documentos depositados.

Este portal tem muitas potencialidades, fundamentalmente na pesquisa avançada que nos permite refinar ou filtrar a informação a pesquisar. Esta pesquisa pode ser efetuada por: data, entidade financiadora, tipo de acesso, idioma, tipo de documento e ainda nos permite selecionar o recurso que pretendemos. Também podemos fazer pesquisa só nos recursos portugueses ou apenas no portal OASIS.

Neste projeto foram desenvolvidos vários serviços e funcionalidades. Um dos mais importantes é o Validador que é uma ferramenta que permite aos gestores dos repositórios verificar a qualidade dos metadados, garantindo também a interoperabilidade com outros sistemas e fazer cumprir as diretrizes nacionais e internacionaisⁱ. Foi feito um upgrade do validador, integrando assim mais funcionalidades, como é visível na Figura 1.

Figura 1. Validador de repositórios

No validador há vários perfis de validação. Pode-se validar todo o repositório, ou pode-se fazer a validação por coleções.

Perfis de Validação:

Diretrizes DRIVER:

- Verifica se existe pelo menos um título
- Verifica se existe pelo menos um autor
- Verifica se a data existe e está num formato válido
- Verifica se existe pelo menos uma política de acesso
- Verifica se existe a definição de tipo de documento
- Verifica se existe pelo menos uma descrição
- Verifica se existe pelo menos um idioma
- Verifica se o idioma é válido
- Verifica se a política de acesso é aberta (*openAccess*)
- Verifica se o tipo de documento é válido

ⁱ <http://validator.rcaap.pt/validator2/>

Diretrizes OpenAire:

- Verifica se existe pelo menos um título
- Verifica se existe pelo menos um autor
- Verifica se a data existe e está num formato válido
- Verifica se existe pelo menos uma política de acesso
- Verifica se existe a definição de tipo de documento
- Verifica se existe pelo menos uma descrição
- Verifica se existe pelo menos um idioma
- Verifica se o idioma é válido
- Verifica se a política de acesso é válida (*openAccess, embargoedAccess, restrictedAccess, closedAccess*)
- Verifica se o tipo de documento é válido
- Lista os documentos de projectos financiados pela Comissão Europeia

Ficheiros e formatos:

- Verifica se os ficheiros são válidos (usando o FITS) e os seus formatos (*mimetype*) estão no conjunto de formatos *preservation friendly*);

Depósito Legal de Teses e Dissertações (TID)

- Verifica se existe pelo menos um autor;
- Verifica se existe pelo menos uma descrição;
- Verifica se a data existe e está num formato válido;
- Verifica TID em teses/dissertações com data de defesa igual ou superior a 7 de Agosto de 2013;
- Valida o TID em teses/dissertações com data de defesa igual ou superior a 7 de Agosto de 2013;

Portal RCAAP:

- Verifica se existe pelo menos um título; verifica se existe pelo menos um autor;
- Verifica se a data existe e está num formato válido;
- Verifica se existe pelo menos uma política de acesso;
- Verifica se existe a definição de tipo de documento;
- Verifica se existe pelo menos uma descrição;
- Verifica se existe pelo menos um idioma;
- Verifica se o idioma é válido;
- Verifica se a política de acesso é válida (*openAccess, embargoedAccess, restrictedAccess, closedAccess*);
- Verifica se o tipo de documento é válido.

FCT:

- Verifica se o tipo de documento é válido
- Lista os documentos de projetos financiados pela FCT

Acesso embargado:

- Lista todos os documentos que estão embargados no repositório;

Incoerência do Embargo:

- Verifica se a data existe e está num formato válido
- Verifica anomalias na data de embargo.

Este validador ainda permite verificar o acesso ao texto integral, sendo um processo mais demorado³.

A relevância do acesso aberto, dos repositórios e dos portais agregadores, é reconhecida, por aumentarem a visibilidade dos seus autores e instituições, a acessibilidade e a difusão dos resultados da atividade acadêmica e de investigação científica⁴.

Metodologia

A metodologia utilizada consistiu na obtenção de todos os registos disponíveis no portal RCAAP, por intermédio da API RESTful disponibilizada pela plataforma. Os dados foram obtidos em formato JSON e integrados numa base de dados relacional (PostgreSQL). Deste procedimento resultaram 1454113 registos que abrangem todas as tipologias das diretrizes DRIVER. Estas diretrizes encontram-se definidas no âmbito do esquema URI info:eu-repoⁱⁱ.

A estrutura dos dados foi analisada por inspeção visual para averiguar os campos mais significativos e o formato do texto que contêm. Foi decidido incluir os campos de *subjects* (associado às palavras chave de cada trabalho), que descrevem os temas, áreas ou assuntos tratados pelo depósito, *sources*, que contém o título das revistas, conferências ou livros onde o trabalho foi publicado, *contributors*, que contém o nome dos autores, *dates*, com as datas de publicação, *languages*, que indica a língua em que o trabalho se encontra redigido, *rights*, que descreve o tipo de acesso e *country*, com o país.

Verificou-se, pela inspeção visual, que a estrutura do texto em cada um dos campos é muito diversificado, principalmente, nos campos de *contributors*, *sources* e *subjects*. Os campos de *dates*, *languages*, *rights* e *country* são mais uniformes, apesar de, no entanto, ainda existir conteúdo que não se encontra bem enquadrado.

Relativamente ao restante texto, este foi inicialmente normalizado, passando todo o texto a letra minúscula. Considera-se, assim, que as palavras têm o mesmo significado independentemente de estarem escritas em letra maiúscula ou minúscula. De seguida, foram eliminados alguns conectores de discurso e as palavras mais comuns, tais como determinantes, pronomes e proposições. Por último, cada palavra foi substituída pela sua raiz, para generalizar o seu significado. Após normalização, foi feita uma contagem de termos por área científica, calculando-se a sua frequência absoluta. A mesma técnica foi usada para identificar pares de palavras mais frequentes.

Posteriormente, procedeu-se à análise dos campos: *dates*, *languages*, *rights* e *country* e *type*. O objetivo da análise é caracterizar a investigação disponível nos repositórios indexados pelo RCAAP, pelo que se recorreu à construção de uma base de dados que permitisse a produção de tabelas de frequência absolutas e relativas para realizar uma análise descritiva exploratória.

ii <http://info-uri.info/registry/OAIHandler?verb=GetRecord&metadataPrefix=reg&identifier=info:eu-repo/>

Resultados

Tendo por base a recolha de dados realizada e já explicada no ponto anterior, de seguida pretende-se apresentar alguns resultados que merecem alguma reflexão. De salientar que o Brasil apresentou um total de 1.451.273 documentos depositados, no RCAAP, e Portugal 42.082 documentos depositados, correspondendo a uma percentagem de 77,5% e 22,5%, respetivamente. Esta diferença deve-se ao facto das diferenças geográficas e demográficas, havendo portanto, mais pesquisadores envolvidos e mais instituições de ensino superior.

Na tabela seguinte torna-se evidente que os documentos classificados, para o tipo de acesso: *Desconhecido* são os que apresentam o peso relativo mais elevado, aproximadamente 74%, seguindo-se o *Acesso Aberto* com 21% e o *Acesso Restrito* com 4,2%. O *Acesso Embargado* e o *Acesso Fechado* não apresentam valores elevados, podendo dizer-se mesmo, quase sem expressão significativa.

Tabela 1. *Tipo de Acesso*

Tipo	Peso Relativo
Desconhecido	73,6%
Acesso Aberto	20,9%
Acesso Restrito	4,2%
Acesso Embargado	0,8%
Acesso Fechado	0,5%
Total	100,0%

O portal RCAAP, que agrega instituições de ensino superior e/ou centros de investigação portugueses e brasileiros, foi também propósito do presente estudo verificar em que linguagem é que se encontravam escritos os documentos depositados, por país. Assim, pela análise aos resultados apresentados na tabela seguinte e tendo por referência o top 3 da língua em que são escritos os documentos, pode concluir-se que: globalmente o português é a linguagem mais utilizada, correspondendo a 74%, refletindo-se este comportamento para Portugal (69%) e para o Brasil (76%); seguem-se os documentos escritos em inglês, registando já um peso significativo (cerca de 24%), o que se torna compreensível uma vez que é uma língua comumente aceite no campo científico, o que permite uma maior visibilidade aos documentos científicos e dos pesquisadores; de realçar que em Portugal a percentagem apresentada para os documentos escritos em inglês é superior quer comparada com o Brasil quer com o total; o espanhol é a 3.^a língua mais utilizada, o que é intuitivo pela proximidade que os dois países detêm com países que têm como língua materna o espanhol e pela colaboração/cooperação transfronteiriça existente.

Tabela 2. *Linguagem dos documentos*

Língua	Portugal		Brasil		Total	
	n	%	n	%	n	%
Português	287.328	69,4%	783.017	76,1%	107.0345	74,2%
Inglês	123.221	29,7%	220.309	21,4%	343.530	23,8%
Espanhol	3.716	0,9%	24.999	2,4%	28.715	2,0%
Total	414.265	100,0%	1.028.325	100,0%	1.442.590	100,0%

Dos documentos analisados e em estudo verifica-se que há 649.789 (45,9%) do tipo *article*, 394.258 (27,8%) *masterThesis*, 155.068 (10,9%) *doctoralThesis*, 111.799 (7,9%) *conferenceObject*, 70.591 (5%) *bachelorThesis*, 16.469 (1,2%) *book*, 14.594 (1%) *bookPart*. Uma análise mais pormenorizada verifica-se, na Tabela 3, e através de uma análise comparativa entre países e entrando em linha de conta com o total de cada país, que no OASIS do Brasil os resultados mais elevados se registam ao nível de documentos como *article* (cerca de 51%), *doctoralThesis* (aproximadamente 13%) e *bachelorThesis* (6,3%). Já para Portugal as demais classificações de documentos registam valores bastante elevados, quando comparados com o Brasil, observandose mesmo que ao nível do registo de *dataset* e *patente*, apenas Portugal apresentou registos.

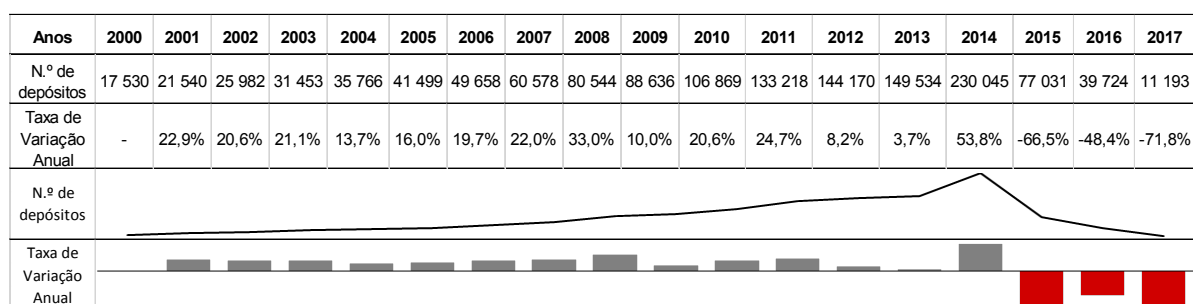
Tabela 3. *Tipo de documento*

Documento	Portugal		Brasil		Total	
	n	%	n	%	n	%
article	127.871	32,9%	521.918	50,8%	649.789	45,9%
masterThesis	153.056	39,4%	241.202	23,5%	394.258	27,8%
doctoralThesis	22.056	5,7%	133.012	12,9%	155.068	10,9%
conferenceObject	47.524	12,2%	64.275	6,3%	111.799	7,9%
bachelorThesis	6.288	1,6%	64.303	6,3%	70.591	5,0%
book	14.993	3,9%	1.476	0,1%	16.469	1,2%
bookPart	12.425	3,2%	2.169	0,2%	14.594	1,0%
dataset	3.963	1,0%	0	0,0%	3.963	0,3%
patent	103	0,03%	0	0,0%	103	0,0%
Total	388.279	100,0%	1.028.355	100,0%	1.416.634	100,0%

De assinalar que ainda há muitos registos que não adotaram a norma de classificação de tipologia ou que foram incorretamente depositados. Há muita diversidade de termos como, por exemplo, *publishedVersion*, *preprint*, *Thesis*, *Peer-reviewed Article*, *finalReport* ou mesmo *Article évalué par des pairs*, *Artículo revisado por pares*, *annotation* ou outros.

Tendo como suporte os resultados apresentados na Figura 2, é notório um crescimento significativo e exponencial até ao ano de 2014, registando-se mesmo uma variação anual positiva cerca de 54%, de 2013 para 2014. Até 2014, as variações positivas registaram-se entre os 10% (em 2009) e os 54% (em 2014) e 33% (em 2008). Por outro lado, nos últimos 3 anos, começou-se a assistir a um decréscimo significativo, em 2015, a taxa de variação anual negativa que rondou os 67%, e em 2016 a variação anual negativa foi de 48,4%. Em 2017, e à data de 20 de setembro de 2017, a taxa de variação anual negativa registada ascendeu a 71,8%.

Figura 2. *Evolução anual total do n.º de depósitos e respetiva Taxa de Variação Anual*



A quebra registada a partir de 2015, dos depósitos nos repositórios, pode dever-se a um acréscimo das atividades administrativas e burocráticas por parte dos docentes e pesquisadores e o preenchimento múltiplo

de várias plataformas que não são interoperáveis com os repositórios. Em Portugal nalgumas instituições de ensino superior a avaliação docente é interoperável com os repositórios, pelo que os docentes/pesquisadores apenas se preocupam com o depósito nos momentos de avaliação. Embora o RCAAP, tenha introduzido formas de depositar os documentos nos repositórios com o identificador DOI, há muitas publicações científicas, como artigos em conferências que não têm esse identificador e portanto torna-se trabalhoso para os docentes/pesquisadores depositarem essa informação, levando-os ao abandono dessa tarefa.

Considerações finais

Com o presente estudo concluiu-se que ainda persistem erros de classificação da tipologia documental, embora têm sido publicadas orientações para minimizar esses erros, mas que seguidas pelos docentes, pesquisadores e editores de metadados. Assim, torna-se ainda impossível fazer uma análise comparativa correta e pormenorizada entre os diferentes repositórios existentes.

Na base de dados utilizada, no presente estudo, verificou-se que ainda existe informação que pode vir a ser tratada posteriormente e os erros detetados podem ser comunicados aos gestores dos repositórios para que possam harmonizar a informação.

Porém, o RCAAP torna-se de vital importância para a comunidade científica, já que a sua filosofia e o seu princípio assenta no acesso aberto. Tem vindo a afirmar-se e a ganhar visibilidade e reconhecimento a nível nacional e internacional⁴. O reconhecimento nacional foi visível com o Decreto-Lei n.º 115/2013, que determina que as teses de doutoramento, bem como outros trabalhos que nos termos da lei substituem estas no processo de obtenção do grau de doutor, bem como as dissertações de mestrado, estão sujeitas ao depósito obrigatório de uma cópia digital num repositório integrante da rede do Repositório Científico de Acesso Aberto de Portugal, operado pela Fundação para a Ciência e Tecnologia, I.P. Posteriormente serviu como instrumento de suporte na Política sobre Acesso Aberto a Publicações Científicas resultantes de Projetos de I&D Financiados pela FCT.

Em 2015, saiu o Despacho n.º 14167, que aprova a lista de formatos autorizados para efeitos de depósito no Repositório Científico de Acesso Aberto de Portugal (RCAAP) dos documentos a que se refere o artigo 50.º do Decreto-Lei n.º 74/2006, de 24 de março.

Com este Governo o acesso aberto e a ciência aberta ganhou nova dimensão com a publicação da Resolução do Conselho de Ministros n.º 21/2016, que define os princípios orientadores para a implementação de uma Política Nacional de Ciência Aberta. Nesta resolução o RCAAP volta a ser referido como sendo um sustentáculo da investigação em Portugal.

Ainda, o facto de não existir uma harmonização de alguns metadados, a integração OASIS/RCAAP, ajuda a sistematizar o desenvolvimento científico atual, como também leva a que transferência do conhecimento se torne mais visível através de parcerias e protocolos de colaboração entre instituições, empresas e outros *stakeholders* e materializa-se na visibilidade que os pesquisadores outorgam aquando de contactos com instituições/entidades públicas ou privadas interessadas no conhecimento gerado e que contribui para efeitos de externalidade (por exemplo patentes) da investigação produzida, em vários sectores da economia⁵.

Bibliografia

1. RCAAP. Projecto RCAAP. [Online].; 2017 [cited 2017 9 14. Available from: <http://projecto.rcaap.pt/index.php/lang-pt/sobre-o-rcaap/enquadramento>.
2. RCAAP. Repositórios Científicos de Acesso Aberto de Portugal. [Online]. [cited 2017 10 14. Available from: <https://www.rcaap.pt/>.

3. RCAAP. Validador de Repositórios. [Online].; 2017 [cited 2017 09 14. Available from: <http://validator.rcaap.pt/validator2>.
4. Carvalho J, Moreira J, Saraiva R. O RCAAP e a evolução do Acesso Aberto em Portugal. [Online]. Braga: Universidade do Minho; 2013 [cited 2017 10 14. Available from: <http://hdl.handle.net/1822/27919>.
5. Pais C. O repositório como serviço de informação na instituição: visibilidade e impacto, práticas de monitorização e validação da informação. [Online]. Lisboa: Universidade Nova de Lisboa; 2017 [cited 2017 10 14. Available from: <http://hdl.handle.net/10198/14469>.