

IBICT/ REDE CARINIANA
Curso Técnico de Curadoria Digital
27 /11/2014

Bibliografia sobre Ciência Aberta

Anaiza C. Gaspar
anaizagaspar@gmail.com

Guia de Fontes de Informação: Grupos de Pesquisa - Rede Cariniana

- Ciência Aberta - Big Data - Gestão de Dados de Pesquisa - Curadoria Digital
- Literatura comentada, resumos traduzidos, pode ser acessada online, realizado download, de acordo com as leis de copyright de cada publicação.

<http://cariniana.ibict.br/index.php/guia-de-fontes>

Ciência aberta - Ciência cidadã - Ciência transparente

- “Ciência aberta é um termo guarda-chuva, que envolve múltiplos níveis e escopos de abertura”, (ABLAGI; APPEL; MACIEL, 2014).
- Valor do dado de pesquisa - dinamismo às atividades de CT&I;
- Maior abertura e participação da sociedade
“Passaporte digital para pesquisadores” - redes de pesquisa e novas plataformas de serviços - alunos, cientistas, professores, cidadãos acessando recursos científicos de alta qualidade.

(data-intensive), e-Science (GRAY, 2007; BELL; HEY; SZALAY, 2009).

- *e-Science* - John Taylor , 2001 “ ciência crescentemente desenvolvida por meio de colaborações globalmente distribuídas viabilizadas pela Internet, usando grandes coleções de dados, recursos computacionais em escala tera e visualização de alto desempenho”

Relação entre política científica e opinião pública (IRWIN, 2001)

- Reino Unido - crise da vaca louca e alimentos geneticamente modificados -
- Consulta Pública sobre construção social da ciência - evolução das Biociências - ir além da mera sloganização da ciência e da democracia.
- <http://pus.sagepub.com/content/10/1/1.short?rss=1&ssource=mfc>.

E-Science resgata pesquisadores do 'dilúvio' de dados (Hey e Trefethen , 2003)

- " e- Science" - ideia subjacente ao “dilúvio iminente de dados científicos” – nova geração de experimentos, simulações, sensores e satélites - motores de busca e ferramentas de software de mineração de dados, grande número de anotações - metadados de proveniência, de conteúdos, métodos e processo de coleta de coleta dos dados, condições operacionais de manipulação e assim por diante.
- [http://eprints.soton.ac.uk/257648/1/The Data Deluge.pdf](http://eprints.soton.ac.uk/257648/1/The_Data_Deluge.pdf).
-

eScience (Sales; Sayão, 2012)

DOI:10.5007/1518 - 2924.2012v17nesp2p118.

- “Uso intensivo de computadores, tecnologias de rede, instrumentação avançada e simulação na pesquisa científica proporciona o surgimento de uma ciência inteiramente baseada nos fluxos de dados e de conjunto de objetos digitais armazenados em repositórios distribuídos globalmente”.

<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p118>

Dados Científicos (*Research Data*) criados /usados na pesquisa

- Dados para processamento científico: dados atmosféricos - previsão meteorológica; dados de sensores - monitoração.
- Dados resultados do processamento automático de objetos - histogramas de cor / fotos imagens.
- Dados-objeto da pesquisa científica, contribuições de redes sociais - estudos sociológicos (RODRIGUES & SARAIVA, 2010).

Dados abertos da pesquisa científica

“(...)aqueles que estão disponíveis gratuitamente na internet pública, permitindo a qualquer usuário baixar, copiar, analisar, reprocessar, fazer a captura por software ou utilizá-los para qualquer outra finalidade, sem outras barreiras financeiras, legais ou técnicas além daquelas que dizem respeito à própria internet.” <http://pantonprinciples.org> Extrato de Panton Principles: Principles for Open Data in Science.

Nível macro das políticas públicas

- “Novas oportunidades de pesquisa são exercidas sobre os conjuntos de dados existentes, incluindo os produtos da chamada pequena ciência, gerando novas descobertas, sem passar pelo processo de coleta de novos dados, caro e dispendioso em recursos e tempo (National Science Board, 2005).
- <<http://www.nsf.gov/pubs/2005/nsb0540/>>

Neelie Kroes, Agenda Digital da Comissão Europeia (Roma, 2012):

- “A União Europeia lidera o movimento do Acesso Livre que continuará no programa Horizonte 2020, começando em 2014 com o e- infraestrutura, (...) sistemas de *helpdesk* regulares e interfaces *web* para tornar as redes de repositórios mais amigáveis e transparentes e incluir serviços avançados de busca, vincular publicações a dados experimentais, processar e visualizar dados digitais, aplicativos de armazenamento e preservação de longo prazo”.
- <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/allea-declaration-1.pdf>

políticas públicas e regulamentos internacionais

"Dados abertos da pesquisa científica devem ser facilmente detectáveis, acessíveis, avaliáveis, inteligíveis, úteis, e sempre que possível interoperáveis com padrões de qualidade específicos."

Declaração de Ministros de Ciência do G8 em 12 de junho de 2013

[https://www.gov.uk/government/news/g8-science-ministers-statement.](https://www.gov.uk/government/news/g8-science-ministers-statement)

G8+06

- Australia, Brazil, Canada, China, France, Germany, India, Italy, Japan, Mexico, Russia, South Africa, United Kingdom, United States.
- RDA – Aliança para Dados de Pesquisa – padrões e diretrizes para infraestrutura global de dados de grandes projetos nacionais.
- Mecanismo de financiamento à pesquisa estratégica - provedores de serviços e infraestruturas de rede - pesquisadores e suas organizações.
- <<https://europe.rd-alliance.org/Pages/Home.aspx>>).

EUA -Escritório de Política Científica e Tecnológica da Casa Branca (2013)

- memorando enviado às principais agências de fomento estabelecendo o acesso aberto a resultados de pesquisa financiada com recursos público, incluindo a oferta dos dados primários em repositórios, salvo restrições de confidencialidade e privacidade pessoal.

iCORDI – Colaboração internacional de infraestrutura de dados de pesquisa

- Fórum para convergência de infraestruturas de dados globais emergentes (experts em políticas de gestão de dados).
- Plataforma de coordenação UE-EUA para interoperabilidade das infraestruturas de dados científicos nos dois continentes.
- <https://www.epcc.ed.ac.uk/projects-portfolio/icordi-international-collaboration-research-data-infrastructure>

EUDAT – Infraestrutura Europeia de Serviços de Dados (1)

- Serviços com recursos de armazenamento compartilhados, distribuídos por 15 países europeus - de uso geral e repositórios de dados específicos da comunidade científica.
- Infraestrutura de dados para múltiplos usuários: comunidades de pesquisa e indivíduos.
- Soluções para encontrar, compartilhar, armazenar, replicar, dados da pesquisa primária e secundária.
- <http://www.eudat.eu/services>

Eudat – Serviços (2)

- B2DROP – identificação do dado, armazenamento e sincronização com grandes arquivos de dados – múltiplos usuários
- B2SHARE –armazena e compartilha dados de pesquisa em pequena escala de diversos contextos - comunidade científica e cidadãos.
- http://www.eudat.eu/system/files/B2Services-presentation_0.pdf

Eudat – Serviços (2)

- B2SAFE – serviço de gerenciamento de dados de vários domínios administrativos, abrigando repositórios departamentais com políticas definidas pela comunidade sobre seus dados de pesquisa.
- B2STAGE - serviço de transferência de conjuntos de dados de pesquisa entre os recursos de armazenamento e EUDAT (HPC) espaços de trabalho de computação de alto desempenho

OpenAIRE - Política de Open Access na Europa

- Projeto piloto (3 anos) da Comissão Europeia: infraestrutura para pesquisadores, helpdesk, rede distribuída em 27 países.
- Comunidades temáticas: requisitos, práticas, incentivos, fluxos de trabalho, modelos de dados e tecnologias para depósito, acesso e manipulação de conjuntos de dados de pesquisa em combinação com publicações científica.
- http://openaccess.sdum.uminho.pt/?page_id=8

Ciência aberta / transparente: objetivo de longo prazo

- As práticas contemporâneas de "e-science" são coerentes com o ideal da "ciência aberta"?
- Normas e práticas de abertura são vitais para o trabalho das comunidades científicas modernas, mas há restrições técnicas e institucionais.
- Diferenças conceituais entre e-science e ciência aberta: presunção indevida de que a e-science promove a colaboração científica aberta global.
- (SCHROEDER, R.; DAVID, P. A.; DEN BESTEN, M. L. 2008) <<http://dx.doi.org/10.2139/ssrn.1317390>> pdf.

Dados primários

- Massa de dados brutos e informações primárias que, uma vez analisada e interpretada, serve de base para as conclusões do artigo científico.
- Em vez de começar do zero o pesquisador pode reusar dados produzidos por outros para validar dados ou chegar a novas conclusões com os dados citados, isto é, abordando novos pontos de vista.

Reutilização (I)

- A primeira forma pode ser vagamente definida como "referência". Inclui casos em que os pesquisadores usam os resultados existentes como contexto vital para os dados por eles próprios coletados ou gerados, seja para ajudar a análise ou para assegurar a qualidade do seu próprio trabalho.
- Em outros casos, os pesquisadores podem reanalisar os dados existentes para testar a replicabilidade de certos resultados - uma pedra fundamental de garantia de qualidade em pesquisa.

Reutilização (II)

- Na segunda forma os pesquisadores podem realizar novas análises sobre os dados existentes para realizar a sua própria pesquisa original. Os dados podem ser agregados para criar novos e enormes conjuntos de dados, que podem então ser analisados para novas ideias usando técnicas computacionais através de uma rede associada e distribuída.

Programa FAPESP, 2014 - Pesquisa em eScience,

- Organizar, classificar e garantir acesso ao gigantesco volume de dados gerados continuamente em todos os campos de pesquisa, a fim de extrair novos conhecimentos e fazer análises abrangentes e originais.
- Requisitos para financiamento: pesquisa em computação com outras áreas do conhecimento que contribua para criação de metodologias e modelos de repositórios de dados

Ciência transparente

- Março/2014, as sete revistas científicas *PLoS* (Public Library of Science) – Biblioteca Pública de Ciências - condicionaram a aceitação de novos artigos à divulgação pelos autores, em repositórios públicos, dos dados de pesquisa
- <http://revistapesquisa.fapesp.br/2014/04/24/ciencia-transparente/>

Theodora Bloom - *PLoS Biology*, *PLoS Computational Biology* e *PLoS Genetics*

- “Nosso ponto de vista é simples. Garantir o acesso aos dados subjacentes deve ser parte intrínseca do processo de publicação científica”
- 30 mil artigos publicados em 2013; criadas ao longo da década de 2000; modelo inovador: artigos online, acesso aberto; corpo de revisores de primeira linha - alcançaram um fator de impacto comparável aos de publicações tradicionais.

PLoS Medicine

Fator de impacto de 15,2 em 2012

- Em média, cada um de seus artigos publicados entre 2010 e 2011 teve 15,2 citações em periódicos indexados em 2012.
- *Nature Medicine* - fator de impacto de 24,3.
- “A *PLoS* - referência internacional – dissemina a ideia do depósito dos dados e cria uma demanda adicional para repositórios e modelos” (Abel Packer)

Desafios

- Apontar em que banco ou repositório estão os dados primários (os arquivos do próprio pesquisador não vale) – onde revisores do artigo ou outros pesquisadores precisem avaliá-los.
- Dados primários abastecem tabelas e análises estatísticas publicadas no artigo e são indispensáveis para que outros pesquisadores consigam reproduzir os mesmos achados.
- Dados protegidos por razões de segurança ou de privacidade não estão incluídos na exigência.

Reprodução de experimentos

- O compartilhamento de dados exige conhecer a origem e entender como foram produzidos, associando aos resultados os métodos, algoritmos ou técnicas adotadas, e ainda ter acesso ao *software* necessário para processá-los, o que torna o processo bastante complexo. Sem isso, pode não ser possível reproduzir o experimento original ou reutilizar o dado em outra pesquisa

Princípio da reprodutibilidade das pesquisas

- É o dinamismo mais importante para a criação dos repositórios de dados de pesquisa. Uma quantidade não desprezível de descobertas científicas acaba não sendo confirmada após sua publicação, por problemas que incluem erros e fraudes, mas que também se estendem a falsos resultados positivos ou negativos obtidos de boa fé.

Pesadelo das empresas farmacêuticas e de biotecnologia

- O problema assombra pesquisadores e revistas científicas, obrigados a cancelar publicação de trabalhos cujos resultados soavam promissores.
- Pesquisadores de uma empresa de biotecnologia, a Amgen, constataram que era possível reproduzir apenas seis entre 53 estudos considerados “marcos” na pesquisa do câncer (The Economist).
- <http://revistapesquisa.fapesp.br/2014/04/24/ciencia-transparente/>

Ações de preservação

- Devem ser incluídas juntamente com as ações de registro dos dados de pesquisa em repositórios. Preservação de dados trata das atuais estratégias globais realizadas em colaboração com os proprietários de dados e prestadores de serviços, entidades e instituições que têm o objetivo de garantir a preservação de dados e conhecimentos em longo prazo.

Dados primários não estão disponíveis ao longo do tempo

- Artigo da revista *Current Biology*: as informações que servem de base a artigos científicos vão se perdendo ao longo do tempo.
- 516 artigos da área de ecologia publicados entre 1991 e 2011 – os artigos publicados nos dois anos anteriores estavam acessíveis, mas os publicados anteriormente caíam a uma taxa de 17% ao ano.
- Cedo ou tarde, o *software* vai ficar obsoleto. A mídia de armazenamento aparece no mercado e em seguida desaparece.

<http://revistapesquisa.fapesp.br/2014/04/24/ciencia-transparente/>.

Fim da 1a Parte