



## **A CIÊNCIA DOS DADOS: pavimentando o caminho para produção científica no século 21**

### **RESUMO**

Este trabalho tem como objetivo discutir sobre Dados Abertos e Reprodutibilidade, questões emergentes após mudanças de regras e critérios exigidos por revistas de publicações científicas, e também pelos financiadores de pesquisas. O método de pesquisa foi pesquisa bibliográfica e a sites de instituições de ensino e pesquisa. Na introdução discute-se qual o significado disso para as pesquisas no mundo acadêmico. Depois é apresentado um panorama simplificado e comum às Instituições Públicas de Ensino Superior sobre as questões de dados e produção científica. E então sugere-se um projeto estratégico que visualiza essas questões de forma ampla e que possa atender às múltiplas demandas de forma consistente e pragmática. Palavras-chave: Dados Abertos. Ciência Aberta. Reprodutibilidade. Metadados.

### **1 - INTRODUÇÃO**

Temas como “Ciência Aberta”, “Dados Abertos”, vieram com intuito de promover maior disseminação do conhecimento e ampliação das pesquisas. Onde o acesso aos dados passou a ser uma exigência para garantir a reprodutibilidade dos trabalhos, onde as revistas passaram a exigir o envio de dados brutos e os algoritmos (scripts em R, por exemplo) que foram utilizados nas análises. Políticas de publicação de revistas, como também de instituições de fomento, passaram a exigir a abertura e publicação dos dados.

### **2 - REVISÃO DA LITERATURA**

Brokman, M., (2009, p.xiii), diz, que é importante pensar sobre as gerações futuras, assim entendendo o agora, mas as questões a enfrentar no futuro. É muito importante na ciência, pois muitas das descobertas importantes se realizam pelas gerações emergentes. Já Chibeni, S.S, (2004, p.1), entende que o sucesso da ciência, vincula-se ao “método científico” no qual se inclui a publicações de resultados e sua avaliação por pares.

De acordo com Noorden (2014), a quantidade de artigos publicados vem dobrando a cada nove anos, já Larsen e Von Ins (2010), dizem que esses números podem ser maiores, pois os canais de publicação aumentaram e muitos acabam não sendo indexados. Com a competitividade em publicações científicas, emergiram também problemas como a falsificação de dados e de reprodutibilidade, aumentando o debate disso nas publicações científicas.

Marques (2014), diz que esses movimentos por abertura vem ocorrendo a muito tempo, mas só ao se tornarem normas e requisitos exigidos por publicações científicas, que então pesquisadores passaram a reagir. É uma mudança, citando Abel Packer (Co-criador da ScieLO), não tratar apenas de realizar tarefas extras nas pesquisas, é uma mudança bem mais profunda, é uma mudança de paradigma para pesquisas, onde práticas de coletas de dados terão necessariamente de incorporar a organização dos mesmos, de modo que os dados sejam compartilhados em repositórios e que possam ser acessados, entendidos e manipuláveis pelos usuários. Outro ponto importante nesse contexto é que os dados passaram a ser referenciados, dando assim valor da citação aos autores.

Marques (2014), dá como exemplo a *Scientific Data*, publicação online dedicada a descrever conjunto de dados científicos, que foi criada pela *Nature Publishing Group*, grupo responsável pela edição da revista *Nature*. Ou seja, não só os novos dados, mas dados de projetos antigos terão valor, os quais poderão render com citações para seus autores ou grupos de pesquisas. De forma que esforços necessários para organização para reaproveitamento de dados legados poderão ser recompensados com citações e reconhecimento.

### 3 - MÉTODOS ADOTADOS

Como métodos adotados, foi utilizado pesquisa bibliográfica sobre o tema e de sites de universidades públicas e instituições de pesquisas, através do Cria [oasisbr.ibict.br](http://oasisbr.ibict.br) e IbiCT [www.cria.org.br/manager](http://www.cria.org.br/manager),

### 4 - RESULTADOS: PANORAMA SOBRE A PRODUÇÃO CIENTÍFICA NAS IPES

Na figura 1, apresenta-se um diagrama geral simplificado sobre dados, publicações e acervos nas IPES, tanto puramente digitais como físicos.

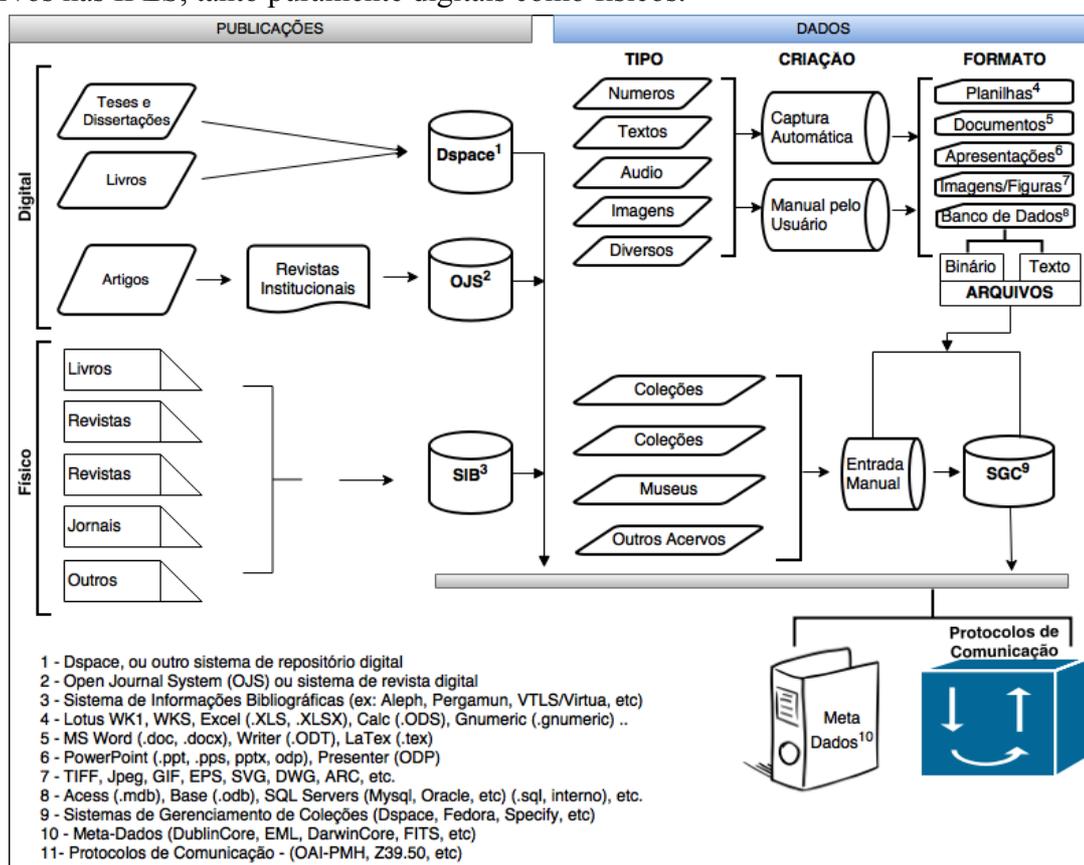


Figura 1 -Diagrama Geral Dados e Publicações Científicas nas IPES

Fonte: criação do próprio autor.

Neste contexto encontram-se acervos digitais de publicações (teses, dissertações, etc) disponibilizados através de portais, os quais foram implantados por instituições de ensino e pesquisa, para busca, divulgação e organização de trabalhos científicos, para atender demandas internas das instituições ou de órgãos de fomento como a Capes por exemplo. Esses portais armazenam arquivos de textos, imagens (fotos, mapas, desenhos), áudios e vídeos, nos formatos PDF, DOC, JPG, MP3, MP4, AVI, entre outros. Já para publicações em meio físico, encontrou-se sistemas de informações bibliográficas (SIB) de licença fechada e sem integração entre instituições, podendo os mesmos possuir alguma interação com sistemas de gestão de acervos digitais como Dspace, entre outros, através de links.

No que confere a dados digitais produzidos por pesquisas, verificou-se pelo ([cria.org.br/manager](http://cria.org.br/manager)), um maior uso de arquivos como planilhas, depois banco de dados locais e poucos sistemas com metadados padronizados. Verificou-se que os indexadores, são os principais sistemas de informações com metadados em uso são os próprios para coleções biológicas, como de plantas, peixes, etc, que importam de aplicativos como Bhrms (flora) ou Specify (de uso geral) e outros que nem sempre tem serviços de metadados. Dentre esses in-

dexadores temos como exemplo, o nacional Cria (ex:<http://cria.org.br/>) e internacional (<http://www.i-marine.eu>).

De modo geral podemos elencar as seguintes questões sobre onde estamos e o que pensar sobre agora e para o futuro:

- a digitalização é um caminho sem volta, sendo ubíquo para fazer trabalhos científicos.
- organização, indexação e informações sobre dados e publicações passam a ser pontos-chaves e diferenciais para as instituições universitárias
- as instituições devem incorporar as melhores práticas para atender as demandas de seus usuários e a integração com toda a sociedade, num mundo cada vez mais aberto, transparente e conectado.

Dentro das práticas que podemos sugerir temos:

1. Uso de práticas como gestão de processos (ex: ISO9000, ISO1400, etc) de forma que a obtenção de dados seja sistematizada e documentada registrando-se metadados.
2. facilitar a localização de informações, com serviços inteligentes que forneçam informações de dados relacionados, comentários e facetas.
3. reduzir o número de locais de consultas através de uso de índices agregadores.
4. uso de interfaces responsivas a computadores, tablets e celulares.
5. uso de padrões abertos de arquivos, para garantir o acesso deles a longo prazo e atender não só as demandas atuais, mas estar capacitado para uma evolução constante, sendo capaz de criar ou introduzir novas práticas e tecnologias.

## 5 - CONCLUSÃO

Com isso podemos entender que, pesquisadores e instituições terão que se adaptar, afim que incorporem práticas já na fase de elaboração, e de forma contínua para projetos de pesquisa. Para isso é importante que esforços conjuntos entre instituições sejam celebrados, a fim de consolidar ações estratégicas para atender a essas demandas.

## 6 - REFERÊNCIAS

BROCKMAN, M. What's Next?: Dispatches on the Future of Science. 2009.

CHIBENI, S. O que é ciência. **Instituto de Filosofia e Ciências Humanas da Unicamp**, 2004.

LARSEN, P. O.; VON INS, M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. **Scientometrics**, v. 84, n. 3, p. 575–603, set. 2010.

MARQUES, F. Ciência transparente. **Pesquisa FAPESP**, n. 218, p. 54–58, 2014.

NOORDEN, R. VAN. Global scientific output doubles every nine years. **Nature News Blog**, 2014.

## **SCIENCE OF DATA: paving the way for scientific production in the 21st century**

**ABSTRACT:** This work aims to discuss open data and reproducibility, issues that emerge after changes of rules and criteria required by scientific publications journals, as well changes of rules required by research financial supporters. As method was used bibliographic research. The introduction talks about the mean of these for research in the academic world. Then are showed a simplified and common panorama about data and scientific production issues in higher education public insitutions. And then proposes a strategic project which see these issues in large view and could meet the multiple requirements in a consistently and pragmatically way.

**Keywords:** Open data. Open Science. Reproducibility. Metadata.