

A preservação digital da e-science: aspirina ou vitamina?

MURILO BASTOS DA CUNHA

MAIRA MURRIETA COSTA

Universidade de Brasília

Faculdade de Ciência da Informação (FCI)

Brasília, DF 70910-900 Brasil

murilobc@unb.br ; mairamurrieta@gmail.com



O surgimento da Big Science

A quantidade de publicações em cada campo do conhecimento, notadamente após a Segunda Guerra Mundial, cresceu, segundo Solla Price (1976), de forma exponencial, duplicando a cada dez ou quinze anos.

Esse fenômeno deu origem a chamada *Big Science*.

A Big Science e Big Data

Hilbert e Lopez (2012) apontaram que:

=> em 2000: apenas $\frac{1}{4}$ da informação armazenada no mundo era digital. Os outros $\frac{3}{4}$ correspondiam à informação analógica (papel, filmes, vinis, fitas magnéticas).

=>em 2002: início da era da informação digital, pois o primeiro a ter dados digitais armazenado em uma quantidade maior que os dados armazenados analogicamente.



A Big Science e Big Data

- em 2007: apenas 7% dos dados armazenados eram analógicos.
- em 2013: o volume de informação armazenada no mundo equivalerá a 1200 exabytes, sendo que destes, apenas 2% será analógico.
- Mayer-Schonberger e Cukier (2013, p. 5) argumentam que a Google “*processa mais de 24 petabytes ao dia, volume milhares de vezes maiores que todo o material impresso na Library of Congress*”.

Como tratar essa proliferação de dados?

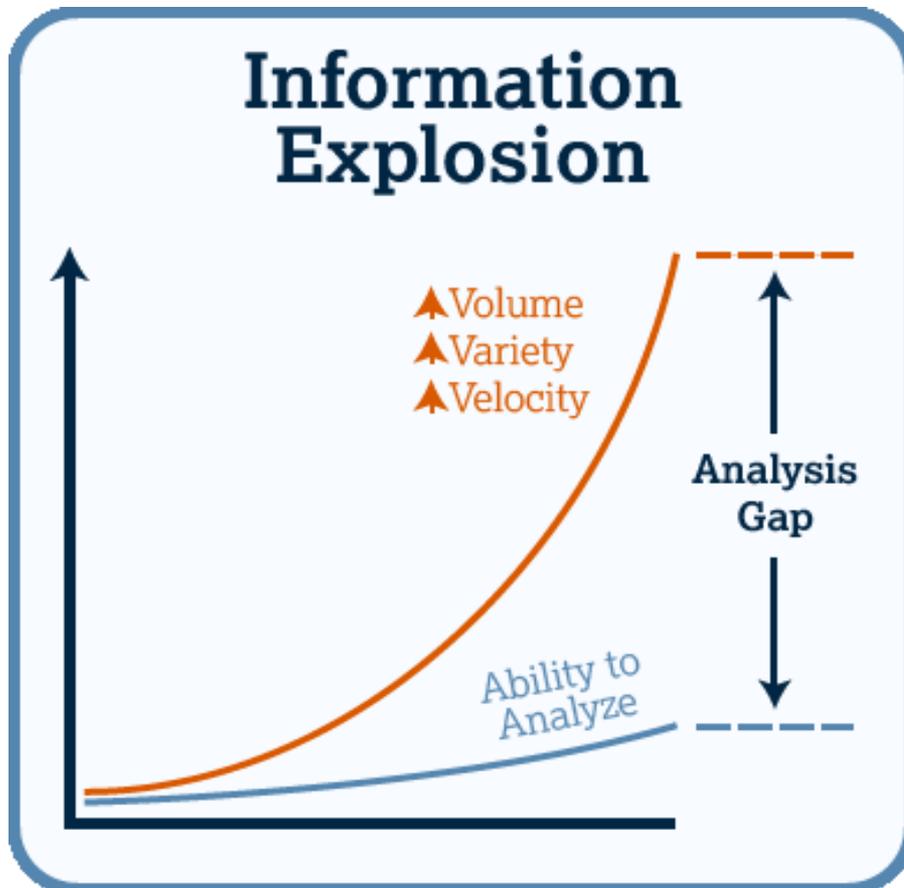
Big Data



- Álvaro *et al* (2011) → os experimentos em partículas físicas conduzidas no Laboratório do CERN, envolvem a colaboração de mais de mil físicos de mais de cem instituições internacionais.
- O CERN gera muitos *petabytes* de dados ao ano. Lyman e Varian (2003) apontam que meros dois *petabytes* equivalem ao conteúdo de todas as bibliotecas universitárias dos Estados Unidos.

Big Data

novos problemas => armazenamento, indexação e preservação(?)

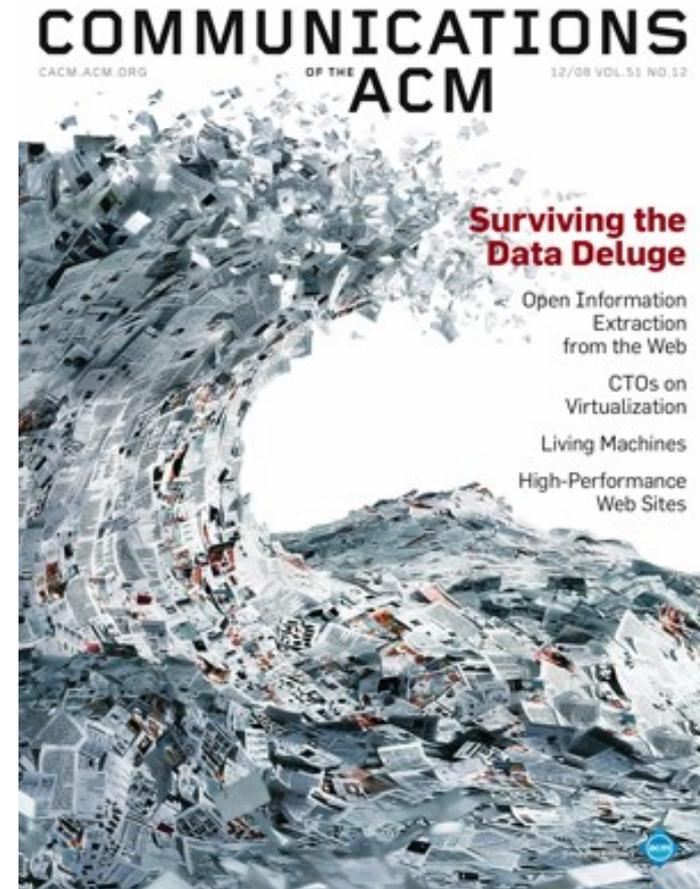


Big Data



- *Big Data* → termo mais amplo, refere-se a um grande volume de dados e o conjunto de soluções tecnológicas para tratar esses dados digitais.
- Relaciona-se com a percepção e compreensão de informações analisadas em grande escala.
- Para Mayer-Schonberger e Cukier (2013) representa “uma nova fonte de valor econômico e informação”.
→ as empresas de TI começam a visualizar grandes negócios nessa área.

A imprensa começa a cobrir Big Data/ Data Deluge/ Info explosion



E-science: uma parte da Big data

E-science/Curadoria de dados/Dados científicos:

→ termo ainda não consolidado, sinônimos: ciência eletrônica, ciência orientada a dados, computação fortemente orientada a dados, ciberinfraestrutura, **curadoria de dados científicos**, **gestão de dados científicos**.

→ gestão de grande volume de dados no âmbito científico.

→ O termo cunhado em 2000, por John Taylor, diretor geral do Conselho do National *e-science* Center (UK).



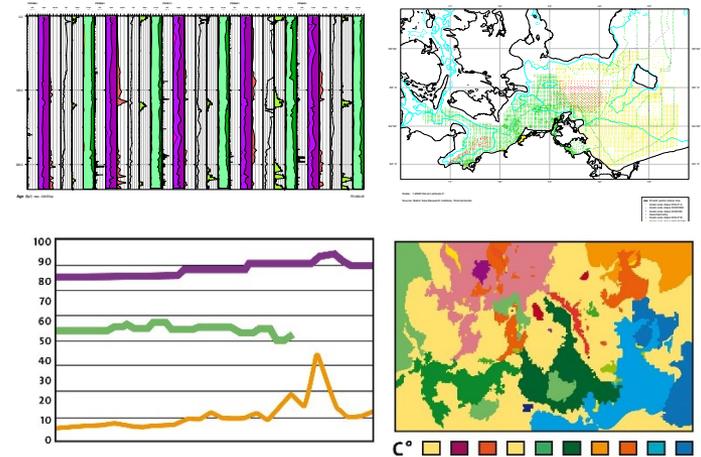
O que é Gestão de dados científicos?

→ Refere-se à coleção de instrumentos e tecnologias necessárias para apoiar a pesquisa científica do Século XXI – intrínseca à **natureza colaborativa e multidisciplinar**, bem como pelo **grande volume de dados** produzidos que precisam estar **disponibilizados em rede**.

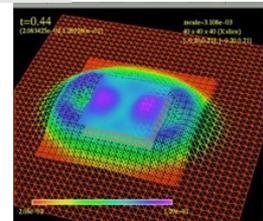
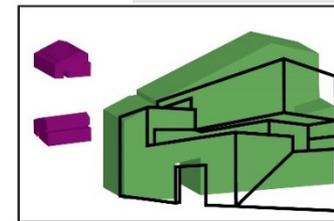


Tipos de Dados Científicos

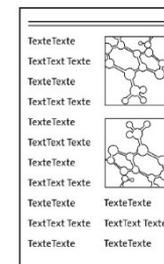
- Instrumentos de medidas
- Experimentos/observações
- Imagens, vídeo e áudio
- Documentos textuais, planilhas, bases de dados
- Dados quantitativos
- Resultados de levantamentos
- Transcrições de entrevistas
- Dados de simulação, modelos, software
- Slides, artefatos, espécimes, amostras
- Diários, livros de laboratório, desenhos....



Research Data



3D Objects

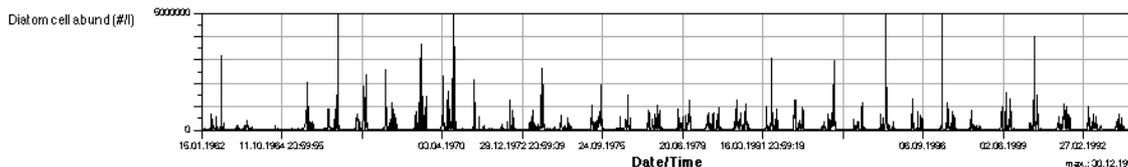


Simulation

```

<state>
  <language>de<
  <position/>
  <length/>4</
  <name>peter</
  ~<favourites>
    ~<favourite>
  
```

Software



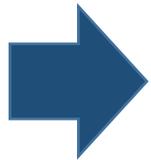
Grey Literature

Gestão de Dados científicos e a biblioteca

Num contexto pré-digital as bibliotecas:

→ desempenharam papéis na coleta, difusão e preservação dos documentos científico e tecnológicos;

→ ajudaram os cientistas a utilizarem as potencialidades do ciclo da Informação Científica e Tecnológica (ICT).



A World Wide Web, proposta por Berners-Lee, em 10 novembro de 1990, já completou 24 anos!

Gestão de Dados científicos e a biblioteca

A Web

- provocou enormes efeitos nos sistemas de comunicação científica e da informação digital.
- dinamizou os sistemas da informação científica e tecnológica (ICT) e trouxe novas indagações:

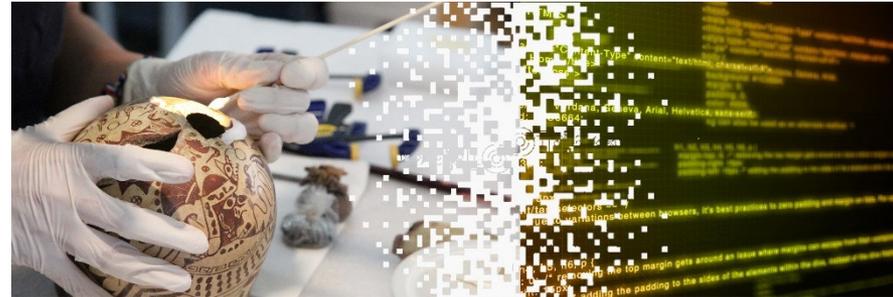
Novas indagações surgiram:

- Como podemos **criar e comunicar o conhecimento** gerado nesse contexto tão dinâmico?
- Como vamos **preservar** a comunicação científica e a ICT geradas nesse contexto?
- Quais serão os novos papéis da biblioteca na execução da **gestão do conhecimento** dentro da academia?

Gestão de Dados científicos e a biblioteca

Cunha (2010) comentou que:

“tende a crescer no futuro (...) Essa área, geralmente composta por **bases de dados numéricos** e os diferentes conjuntos de **resultados das pesquisas** realizadas nos institutos, faculdades e departamentos, **nunca foi objeto de preocupação** por parte da biblioteca universitária”.



Gestão de Dados científicos e a biblioteca

“Uma definição ampla da ciência eletrônica significa que para apoiá-la necessariamente estão sendo incorporada uma série de atividades e serviços.

Tal apoio exige o desenvolvimento, a coordenação e investimentos em vários setores da biblioteca para criar um sistema onde, certamente, estarão envolvidos a segurança dos dados, a **preservação**, o acesso e o controle dos metadados”. (Cunha, 2010).

Gestão de Dados científicos e a biblioteca

- A literatura norte-americana já revela uma preocupação dos bibliotecários com esse novo cenário.
- Luce (2010, p. 3) argumenta que para as “bibliotecas universitárias a evolução gradual da *e-science* provoca **desafios profundos** e, ao mesmo tempo, proporciona às bibliotecas uma **oportunidade** de **redefinir seus papéis** e agregar valor ao seu **portfólio de serviços**”.

Situação nos EUA e Canadá

- 2006: a Association of Research Libraries (ARL) criou uma Força Tarefa *e-Science* (*e-Science Task Force*).
- 2009: a ARL realizou um levantamento com o objetivo de identificar o envolvimento das bibliotecas com a questão do tratamento dos dados oriundos da *e-science*. (Soehner; Steeves; Ward, 2010).



ASSOCIATION OF
RESEARCH LIBRARIES

Situação nos EUA e Canadá

- Levantamento obteve respostas de **57** bibliotecas:
- **21** fornecem infraestrutura ou serviço para gestão de dados científicos,
- **23** bibliotecas estavam planejando;
- 13 bibliotecas não ofereciam.
- O estudo mostrou que entre as bibliotecas respondentes, 42% contrataram e 39% planejam contratar membros de equipe com habilidades em gestão de dados científicos (Soehner; Steeves; Ward, 2010).



ASSOCIATION OF
RESEARCH LIBRARIES

Gestão de Dados científicos no Brasil

- Temática é pouco trabalhada.
- Vaz (2011) mostrou que EMBRAPA Informática Agropecuária deve assumir um papel relevante nessa área.
- Eventos específicos:
 - 2012: *VI Workshop de e-Science*, Curitiba (PR).
 - 2013: *Encontro Internacional de Dados, Tecnologia e Informação*, Marília (SP)
 - 2014 (Out.): *IEEE 10th International Conference on e-Science*, São Paulo (SP).
 - 2014 (Nov.): tema incluído no Seminário de Bibliotecas Digitais

Gestão de Dados científicos no Brasil

Cunha (2010, p. 10) já havia comentado que o acervo de dados oriundos da *e-science* tende a crescer no futuro e argumenta:

→ “como a biblioteca começa a tomar para si a responsabilidade de **gestão do conhecimento** gerado no campus, ela agora precisa conhecer os conteúdos e as estruturas desses recursos informacionais hospedados nos **laboratórios e gabinetes docentes**”.



Gestão de Dados científicos no Brasil

- “Para a biblioteca gerir os dados científicos terá que incorporar uma série de atividades e serviços.
- Além disso, a biblioteca terá que fazer investimentos na segurança de dados, **preservação**, acesso e controle de metadados”.

Preservação digital: aspirina ou vitamina?

Na área de gestão é comum planejar um negócio para:

→ melhorar a longo prazo (**modelo vitamina**)

→ para dar solução a um problema existente, uma dor de cabeça (**modelo aspirina**).

Nos próximos anos os especialistas em preservação digital disporão de suas próprias **aspirinas** para as numerosas **instituições que esqueceram de fazer os seus deveres digitais**. (Llueca, 2014).

Aspirina ou vitamina?

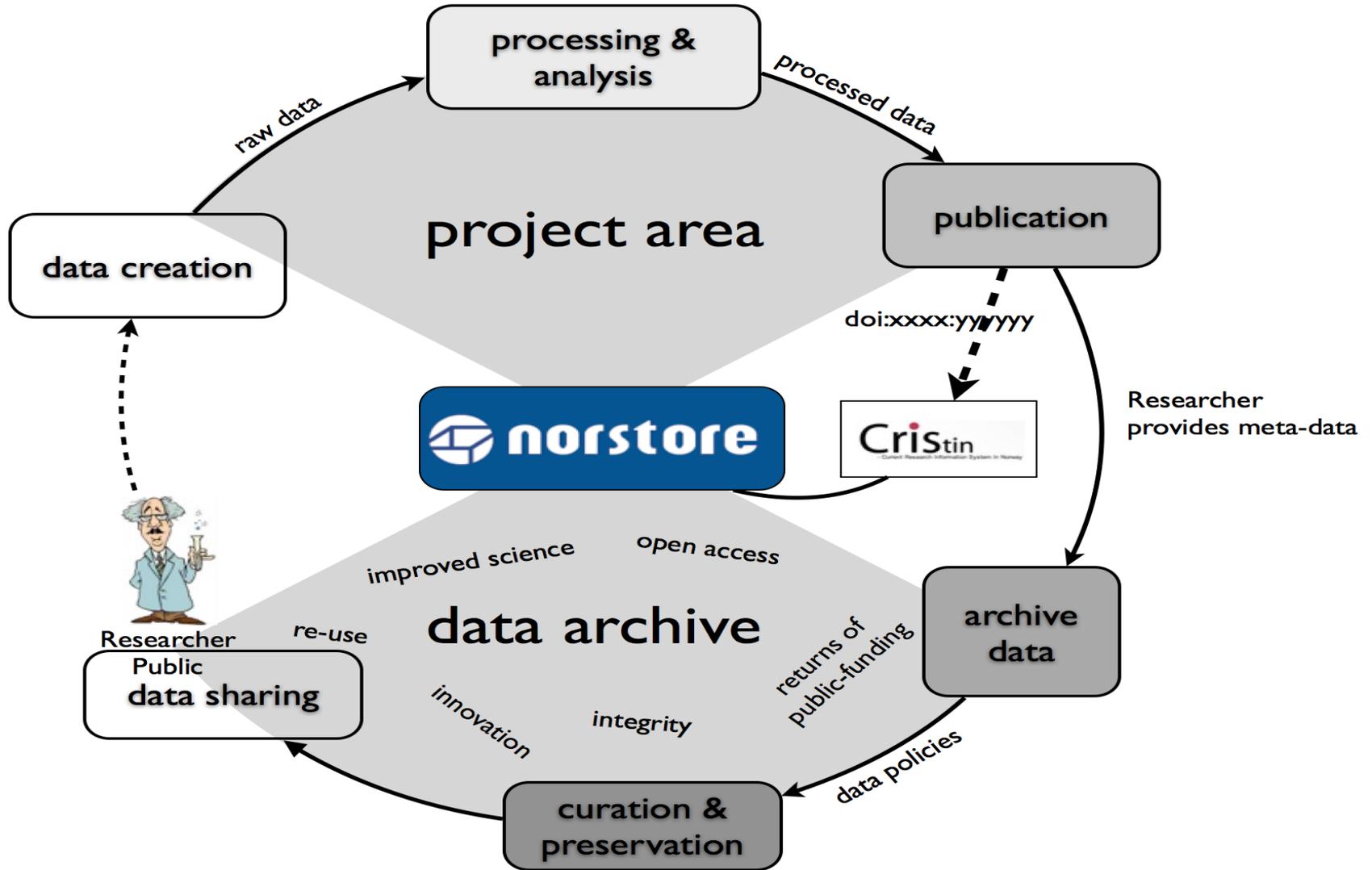
- Apesar de ser inquietante o futuro, o potencial de **desaparecimento** ou **inacessibilidade dos documentos digitais** é visto com receio justificado.
- Hoje a **preservação digital** passa a ser vista como uma disciplina **vitamina**: se usar políticas de preservação viverás melhor!
- Se tomarmos **mais vitaminas** iremos precisar de **menos aspirinas!**



Saber as respostas para não tomar aspirina

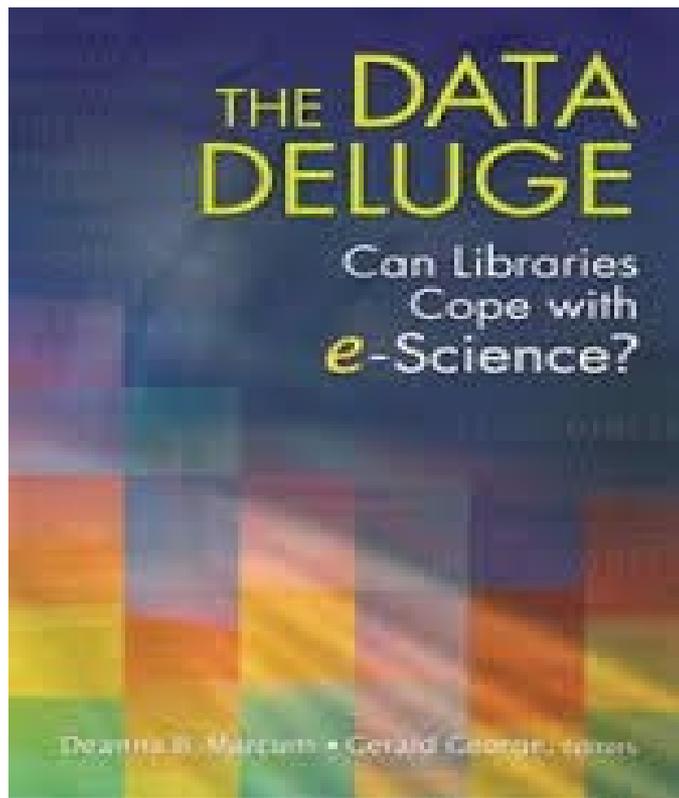
- Quais são os serviços de gestão de dados científicos necessitados pelas comunidades científicas?
- Quais os desafios para atender os variados grupos de pesquisa da instituição?
- Que **estrutura organizacional** melhor responde para atender esses grupos?
- Que **habilidades/conhecimento** o bibliotecário e a biblioteca precisam para atender esses grupos?
- **O que preservar? Como preservar?**
- Que modelos seriam ideais para a biblioteca trabalhar com o **Centro de Informática** e as **unidades acadêmicas**?

Ciclo de vida do dado científico



Gestão de dados científicos: uma nova fronteira da preservação digital.

A hora chegou!



Perguntas? Obrigado!



Referências

- ALVARO, Elsa *et al.* E-science librarianship: field undefined. **Issues in Science & Technology Librarianship**, n. 66, p. 28-43, Summer 2011.
- BERNERS-LEE, Tim. WorldWideWeb: Proposal for a HyperText Project. 12 November 1990. Disponível em: <http://www.w3.org/Proposal> Acessado em: 24 de fev. 2014.
- CUNHA, Murilo Bastos da. A biblioteca universitária na encruzilhada. **DataGramZero – Revista de Ciência da Informação**, v. 11, n. 6, dezembro 2010. Arquivo disponível em < http://dgz.org.br/dez10/Art_07.htm > Acesso em 25 julho 2012.
- HILBERT, Martin; LOPEZ, Priscila. How to measure the worlds technological capacity to communicate, store and compute information part I: results and scope. **International Journal of Communication**, v. 6, p. 956-979, April 2012.
- Lluca, C. Reseña del libro "Preservación digital". *Revista Española de Documentación Científica*, 37 (1): e039, 2014. Disponível em: <http://redc.revistas.csic.es/> Acessado em: 3 março 2014.
- LUCE, Richard E. Grand challenges and new roles for the twenty-first-century research library in an era of e-science. In: MARCUM, Deanna B.; GEORGE, Gerald (Ed.). **The data deluge: can libraries cope with e-science?** Santa Barbara, California: Libraries Unlimited, 2010. Cap. 1.
- LYMAN, Peter; VARIAN, Hal R. **How much information 2003?** Berkeley, California: University of California at Berkeley, 2003. 100 p. (Relatório produzido pelos estudantes da Escola de Gestão da Informação e Sistemas da Universidade da Califórnia em Berkeley). Disponível em <<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/index.htm>> Acesso em 18 julho 2012.

Referências

- MAYER- SCHONBERGER, Victor; Cukier, Kenneth. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.
- SOEHNER, Catherine; STEEVES, Catherine; WARD, Jennifer. **E-science and data support services**: a study of ARL member institutions. Washington, DC: Association of Research Libraries, 2010. Disponível em:
- <http://www.arl.org/bm~doc/escience_report2010.pdf> Acesso em 18 julho 2012.
- SOLLA PRICE, Derek. **O desenvolvimento da ciência**: análise histórica, filosófica, sociológica e econômica. Rio de Janeiro: Livros Técnicos e Científicos, 1976.
- VAZ, Glauber José. **E-Science na Embrapa**. Campinas: Embrapa Informática Agropecuária, 2011. 58p. (Documentos, 117)