

O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais

Maira Murrieta Costa

Doutoranda em Ciência da Informação pela Universidade de Brasília. Mestre em Ciência da Informação.

Murilo Bastos da Cunha

Professor titular da Faculdade de Ciência da Informação da Universidade de Brasília. Phd em Information Science pela University of Michigan

<http://dx.doi.org/10.1590/1981-5344/1900>

Contextualiza o surgimento do termo e-science e as tecnologias necessárias para apoiar a pesquisa colaborativa do Século XXI. Argumenta que os dados coletados por instrumentos tecnológicos tais como telescópios, satélites, sensores especializados podem ser considerados como informação científica e, portanto, precisam ser tratados de forma a viabilizar a sua organização, recuperação, difusão e preservação para auxiliar no desenvolvimento de pesquisas futuras. O artigo tem como objetivo discutir o papel do profissional da informação no tratamento dos dados oriundos da e-science. Demonstra que nos Estados Unidos e no Reino Unido os bibliotecários já estão atentos à necessidade de tratamento desse novo tipo de dado. Discute o conceito da e-science no cenário brasileiro.

Palavra Chave: *Biblioteca digital; Bibliotecário; e-science; Ciberinfraestrutura; Dados abertos; Dilúvio de dados; Quarto paradigma; Ciência da Informação; Pesquisa colaborativa.*

The librarian in the treatment of data from the e-science: initial considerations

The article contextualizes the emergence of the term e-science and the technologies needed to support collaborative research of the XXI Century. It argues that data collected by technological instruments such as telescopes, satellites, specialized sensors constitute scientific information and therefore must be treated in order to facilitate their recovery, diffusion and preservation to assist in the development of future research. Discuss the role of the information professional in the treatment of data from e-science. It demonstrates that librarians from the United States and the United Kingdom are already aware of the need for this new type of treatment given. The objective of the study was to identify the basic skills necessary for the information professional who will organize the collection originated from e-science.

Keywords: *Cyber infrastructure; Data deluge; Digital library; e-science; Fourth Paradigm; Information Science; Librarian; Open data; Collaborative research.*

Recebido em 15.10.2013 Aceito em 07.02.2014

1 E-SCIENCE

A evolução da ciência está altamente relacionada com o aprimoramento do instrumental tecnológico que permitiu a realização de observações, como, por exemplo, telescópios, satélites, sensores especializados dentre outros. Para Gordon Bell (2011) as teorias científicas do Século XX foram baseadas em dados geralmente disponíveis em cadernos científicos pessoais e, em alguns casos, na chamada "alta ciência", aquela cuja coleta de dados é altamente dependente dos ensaios realizados em laboratórios, bem como do instrumental tecnológico acima citado – as teorias surgidas desses tipos de atividades foram baseadas, em grande parte, nos dados coletados por esses instrumentos.

Uma questão que emergiu neste início do Século XXI é que, de forma crescente, os dados oriundos de pesquisas são coletados por meio de sensores especializados, telescópios, satélites, ensaios de laboratórios, dentre outros. Green (2011); Fox e Hendler (2011), destacam que a

pesquisa científica será transformada pela criação e disponibilidade de um grande volume de dados.

Sob essas circunstâncias é que Gray (2011, p. 17) afirmou que “e-science é o ponto onde a TI [tecnologia da informação] encontra cientistas”. O fato é que autores (GRAY, 2009; MAYER-SCHONBEERGER, CUKIER, 2013) têm destacado a importância da tecnologia da informação na forma de se fazer ciência. Para esses autores, os desafios tecnológicos incluem a necessidade de melhor captar, analisar, modelar, visualizar e preservar as informações científicas, tornando os sistemas de computação vitais para o moderno ambiente de pesquisa.

Nesse cenário, é prudente refletir sobre a necessidade de um tratamento adequado que viabilize o processo de armazenamento, organização, busca, recuperação e preservação dos dados e das informações geradas a partir desse tipo de pesquisa. Caso contrário os dados coletados podem se tornar ilegíveis ou o que seria mais drástico, se perder em um grande volume de dados, por falta de tratamento técnico adequado.

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*. Hilbert e Lopez (2012) conduziram estudo que revelou que no ano 2000, apenas ¼ da informação armazenada no mundo era digital. Os outros ¾ correspondiam à informação analógica (papel, filmes, vinis, fitas magnéticas). O ano de 2002 marcou o início da era da informação digital, pois o primeiro a ter dados digitais armazenado em uma quantidade maior que os dados armazenados analogicamente. Já em 2007, apenas 7% dos dados armazenado eram analógicos. A previsão dos autores é de que 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”. A questão que se apresenta é como tratar essa proliferação de dados?

Nesse cenário, merece destaque a colocação de Álvaro *et al* (2011) que comenta sobre experimentos em partículas físicas e sobre o *grande colisor de hádrons*¹, conduzidos no Laboratório do CERN, envolvem a colaboração de mais de mil físicos de mais de cem instituições internacionais. Foi estimado que esse projeto gerasse muitos *petabytes*² de dados ao ano. Face ao exposto, é pertinente ressaltar a constatação de Lyman e Varian (2003) de que meros dois *petabytes* equivalem ao conteúdo de todas as bibliotecas universitárias dos Estados Unidos.

No início do Século XXI, a pesquisa colaborativa é descrita como aquela que tem a “capacidade de gerar e armazenar dados em uma escala

1 É o maior acelerador de partículas existente do mundo. É considerado como um dos grandes marcos de engenharia da humanidade. Foi construído pela Organização Europeia para Pesquisa Nuclear (CERN)

2 1 petabyte equivale a 1.024 terabytes. 1 terabyte, por sua vez equivale a 1.024 gigabytes. As medidas referem-se à capacidade de armazenamento de um dado dispositivo.

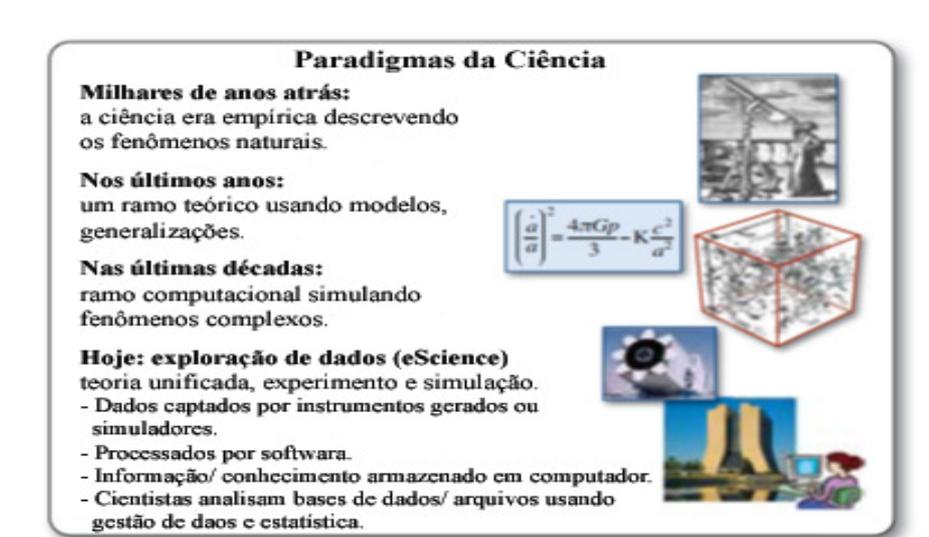
sem precedentes e muito além da capacidade humana de análise” (CESAR JÚNIOR, 2011). Suas características deram origem aos termos: *Big Data* e *e-science*.

Big Data é um 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-Schonbeberger e Cukier (2013) o *big data* representa “uma nova fonte de valor econômico e informação”. A filosofia do *Big Data* é *deixe os dados falarem*.

Mayer-Schonbeberger e Cukier (2013) exemplificam o conceito de *Big Data* lembrando-se do surgimento do vírus H1N1 em 2009. Os autores relatam que pesquisadores da Google analisaram os 50 milhões de termos de busca mais comuns entre os americanos e os compararam com a lista do Centers of Disease Control (CDC). A pesquisa nos termos de busca utilizados no Google revelou onde o vírus estava se espalhando com mais velocidade que o sistema de informações do CDC.

A *e-science*, por sua vez, também é retratada no âmbito da produção de um grande volume de dados e da necessidade de avanço da ciência. Dentre as denominações utilizadas para *e-science*, destacam-se na literatura os termos: ciência orientada a dados, computação fortemente orientada a dados, ciberinfraestrutura ou quarto paradigma (ALVARO *et al*, 2011; CESAR JÚNIOR, 2011; MARCUM, GEORGE, 2010). A diferença parece estar no fato da *e-science* tratar de grande volume de dados no âmbito científico. Neste artigo, optou-se pela utilização do termo *e-science* pelo fato do termo se referir a grande coleta de dados no âmbito científico. Além disso, o termo é o utilizado na literatura americana de Ciência da Informação. A Figura 1 ilustra os paradigmas da ciência na visão de Gray (2007), bem como, retrata a evolução no processo de coleta de dados que culminou com a chamada *e-science* ou dilúvio de dados.

Figura 1 – Paradigmas da ciência



Fonte: Gray (2007) – tradução nossa.

O termo *e-science* foi cunhado no ano 2000 por John Taylor, diretor geral do Conselho do National *e-science* Center no Reino Unido. Caracterizado pela colaboração global de pesquisadores, a *e-science* 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. (MARCUM; GEORGE, 2010; VAZ, 2011)

Segundo Cunha (2010) a *e-science* (ciência eletrônica):

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. (...) 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.

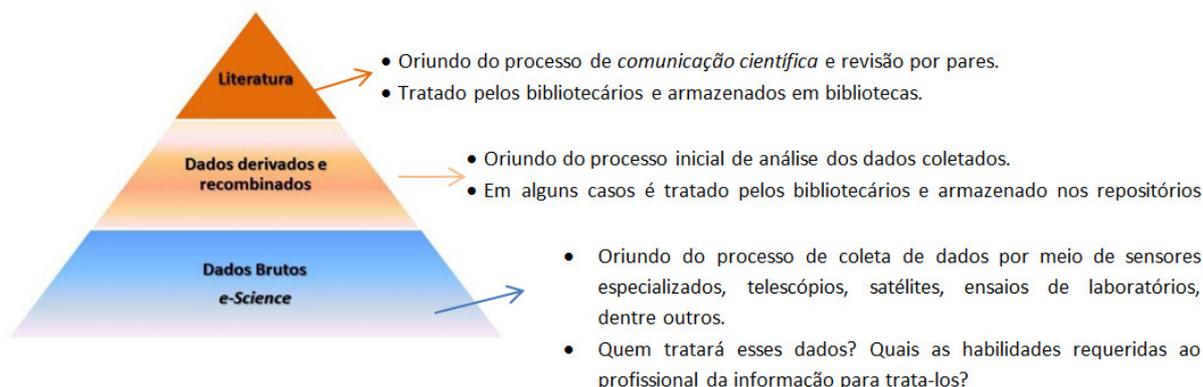
Na Ciência da Informação, a *e-science* traz implicações relevantes sobre a comunicação científica, afinal os dados oriundos da *e-science* são de fato dados científicos primários. Também gera efeitos nos serviços e produtos de informação, bem como afeta diretamente as bibliotecas digitais exigindo reflexões sobre preservação digital e o planejamento das bases de dados.

A partir do exposto, o presente artigo tem como objetivo discutir o papel do profissional da informação no tratamento dos dados oriundos da *e-science*. Para tanto, serão apresentados os cenários de tratamento de dados na Inglaterra e Estados Unidos – países onde a temática é mais avançada -- bem como uma breve análise do contexto brasileiro.

2 A E-SCIENCE NA CIENCIA DA INFORMAÇÃO

De acordo com Mueller (2007, p. 128), "para a ciência avançar não basta que o conhecimento seja publicado (...) o que exige entendimento de todo o processo de produção e comunicação da informação". Em consonância com a autora, Gray (2007) argumentou que todos os dados científicos precisam estar *online*, sejam eles oriundos da literatura ou os dados brutos, conforme ilustra a Figura 2.

Figura 2 – Estrutura dos dados científicos



Fonte: Os autores com fundamento em Gray. (2007)

Ao se tratar dessa nova informação científica e tecnológica, que culminou com o surgimento da *e-science*, Bell (2011, p. 11) comenta que "a origem remota dos dados, assim como o acesso comunitário a dados distribuídos, são apenas alguns dos desafios [da *e-science*]". Para o autor, esses dados devem "permanecer para sempre num estado submetido à curadoria e acessível para o público para análise contínua". Colocação essa, que converge para a questão de acesso aberto à informação científica (*Open Access*).

Evoluindo seu raciocínio, Bell (2011) faz a comparação de que essa permanência de dados é semelhante à permanência de coisas [grifo do autor] mantidas pelas bibliotecas e argumenta:

(...) é exatamente o que as bibliotecas insistem em fazer e têm sempre tentado fazer. A nuvem [grifo do autor] de polarizações magnéticas, codificando dados e documentos na biblioteca digital, vai se tornar o equivalente moderno de quilômetros de estantes de bibliotecas, que conservam papel e partículas de tinta. (BELL, 2011, p. 12)

Em consonância com Bell (2011), Gray (2007) defendeu o estabelecimento de modernos arquivos de dados e documentos, que seriam comparáveis às bibliotecas tradicionais.

3 E-SCIENCE NA CIÊNCIA DA INFORMAÇÃO: contextos mundial e brasileiro

Os dados produzidos pela *e-science* trazem impactos profundos sobre a ciência e, portanto, exige um exame acurado das funções das instituições empenhadas no avanço da ciência e no apoio aos cientistas, dentre elas as bibliotecas.

De acordo com Vaz (2011) o Reino Unido lançou, em 2001, um programa pioneiro que recebeu um aporte financeiro de £250 milhões com o objetivo de estimular a *e-science* em todos os campos de pesquisa. O programa visava:

prover a infraestrutura e facilidades necessárias para a pesquisa colaborativa, acelerar a emergência da próxima geração de padrões de plataforma aberta para serviços globais de informação, resolver os principais desafios em processamento, comunicação, e armazenamento de grandes volumes de dados. (VAZ, 2011, p.10; grifo nosso)

O fato da Inglaterra, já em 2001, ter lançado de forma pioneira, um programa lhe conferiu certa maturidade no tratamento de dados oriundos da *e-science*. Tal afirmação se justifica pelo volume de trabalhos vinculados ao tema pelas Universidades de Oxford e Manchester. Também apresentam trabalhos significativos sobre o tema o Arts and Humanities e-Science Support Centre, o Oxford eResearch Centre, o National e-science Center, o *e-Science Centre*, bem como o *e-Science Core Programme*.

O e-Science Core Programme é gerido pelo Conselho de Pesquisa em Ciências da Engenharia e Física, em nome das comunidades de todos os Conselhos de Pesquisa. Ele tem apoiado o desenvolvimento de tecnologias genéricas, como o *software* conhecido como *middleware* – necessário para permitir que diferentes recursos trabalhem de forma integrada através de redes, bem como criem *grids* computacionais.

Nos Estados Unidos, a National Science Foundation criou, em setembro de 2007, o programa Sustainable Digital Data Preservation and Access Network Partner (*DataNet*), cujo objetivo era desenvolver:

um conjunto de organizações exemplares de infraestrutura nacionais e globais de dados de pesquisa (apelidado *Datanet Partners*) que oferecem oportunidades únicas para as comunidades de pesquisadores para avançar a ciência e/ou engenharia, pesquisa e aprendizagem". (CHOUDHURY, 2010, p. 194).

Para tanto foi criado um prêmio -- o *Data Conservancy* -- que definiu curadoria de dados como: "um meio para coletar, organizar, validar e preservar os dados para que os cientistas possam encontrar novas maneiras de enfrentar os grandes desafios de pesquisa que a sociedade enfrenta" (CHOUDHURY, 2010, p. 194). Ressalta-se que a Sheridan Library conquistou em 2009 um dos prêmios existentes, por meio do programa *DataNet* da National Science Foundation.

Dentre as universidades americanas se destacam os trabalhos vinculados às de Purdue e de Washington. Além disso, percebe-se um interesse de grandes corporações como a Microsoft no tema. A esse respeito merece destaque o fato de Tony Hey, atual vice-presidente da área de pesquisa da Microsoft, ter sido o diretor do *e-Science Core Programme no Reino Unido*.

No Google Acadêmico, ao se pesquisar sobre *e-science* e bibliotecas, os artigos recuperados em destaque são de Tony Hey, o primeiro, *The data deluge: an e-science perspective*, citado por 367 autores e; o segundo: *E-science and its implications for the library community*, citado

por 64 trabalhos. Em função do envolvimento de seu atual diretor de pesquisa com a *e-science*, parece natural que a Microsoft Corporation ter despertado seu interesse para essa nova e crescente área.

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".

Consciente do impacto e das oportunidades para as bibliotecas universitárias, a Association of Research Libraries (ARL) criou uma Força Tarefa *e-Science* (*e-Science Task Force*), em 2006, que definiu o domínio da *e-science*. Essa força tarefa foi seguida por um grupo de trabalho contínuo que teve como missão desenvolver a compreensão dos membros para as mudanças de habilidades profissionais e infraestruturas necessárias para o tratamento de um novo tipo de dado – o oriundo da *e-science*. (SOEHNER; STEEVES; WARD, 2010).

A identificação de diferentes abordagens sendo empreendidas por instituições isoladas (EUA, Canadá, Reino Unido) para a compreensão do fenômeno da *e-science* incitou a ARL a desenvolver um levantamento, em 2009, com o objetivo de identificar o envolvimento das bibliotecas com a questão do tratamento dos dados oriundos da *e-science*. O instrumento de coleta de dados foi enviado para 123 bibliotecas membros da ARL nos EUA e Canadá. Dentre as indagações do questionário constava a seguinte pergunta: *Serão os bibliotecários aqueles que intervirão e enfrentarão o desafio?* Foram obtidas respostas de 57 bibliotecas membros da ARL, onde se destacou que: 21 bibliotecas afirmaram fornecer infraestrutura ou serviço para *e-science*, 23 bibliotecas afirmaram que estão planejando oferecer esse tipo de serviço e, 13 bibliotecas afirmaram não oferecer suporte para *e-science*. Além disso, o levantamento demonstrou que entre as bibliotecas respondentes, 42% contrataram e 39% planejam contratar membros de equipe com habilidades em *e-science* (SOEHNER; STEEVES; WARD, 2010).

Para Alvaro *et al* (2011), a *e-science* pode fornecer um campo potencial para bibliotecários ramificarem-se para além dos limites das práticas tradicionais de biblioteca. Na visão dos autores, a *e-science* não é prática comum, e em função disso a Biblioteconomia deverá prosseguir neste novo território com cautela.

Luce (2010) argumenta que com visão, investimentos estratégicos e com uma "alavancagem" de sua *expertise* em gestão da informação, as bibliotecas digitais podem se tornar um recurso essencial para o tratamento digital da informação oriunda da *e-science*, que deve estar disponível para a próxima geração de comunidades de pesquisa. Da mesma forma, Soehner, Steeves e Ward (2010, p. 7) comentam que "curadoria de dados, preservação, acesso e metadados são áreas da *e-science* onde as bibliotecas encontram uma afinidade natural".

No Brasil, a problemática dos dados oriundos da *e-science* ainda é pouco trabalhada. A busca bibliográfica, realizada em bases de dados

nacionais e internacionais, revela uma incipiência de estudos que contemplem as contribuições da biblioteconomia e ciência da informação para a *e-science*. O caráter exploratório do estudo pode ser comprovado pela baixa quantidade de artigos publicados tanto no Brasil, como no exterior, conforme demonstra a Tabela 1.

Tabela 1 – Levantamento sobre *e-science*

Base de Dados	Nome da Base	2003/2004	2005/2006	2007/2008	2009/2010	2011/2012	Total de estudos
Internacionais	LISA	11	18	30	35	20	114
	LISTA	10	22	44	51	19	148
	ERIC	3	3	3	5	2	16
Nacionais	BRAPCI	--	--	1	1	2	4
	SCIELO	--	1	5	6	13	25
	Base de Dados ABCDM	--	--	1	2	--	3

Fonte: os autores; dados coletados em junho de 2013.

A pesquisa bibliográfica foi realizada no período de março a junho de 2013 nas seguintes bases de dados internacionais: *Library and Information Science Abstracts* (LISA), *Library e Information Science & Technology Abstracts* (LISTA) e *Education Resources Information Center* (ERIC). No âmbito nacional, a pesquisa foi realizada na *Scientific Electronic Library Online* (SciELO), Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) da Universidade Federal do Paraná e ABCDM, da Universidade de Brasília.

A base de dados LISA foi selecionada por indexar periódicos relacionados com Biblioteconomia e Ciência da Informação e permitir acesso a artigos desde 1969 até a presente data. Já a LISTA foi selecionada por indexar mais de 675 das principais revistas científicas em ciência e tecnologia da informação. A base de dados ERIC fornece acesso ilimitado a mais de 1,3 milhões de registros bibliográficos. Assim, a pesquisa nas três bases de dados citadas conferiu exaustividade à busca sobre o tema *e-science*.

No âmbito nacional, a BRAPCI foi escolhida por ser uma base de dados referencial, que relaciona os 27 títulos de periódicos brasileiros em Ciência da Informação, sejam estes correntes ou não. Para complementar a busca nos periódicos brasileiros também foi realizada uma pesquisa na Base de Dados ABCDM (ex-ABCID), uma base referencial, que cobre os artigos de periódicos das revistas publicadas no Brasil e em Portugal nas áreas de Arquivologia, Biblioteconomia, Ciência da Informação, Documentação e Museologia. Já a SciELO possui a maior coleção *on-line* de periódicos científicos brasileiros.

Apesar do baixo número de trabalhos científicos sobre o tema, já se observa a necessidade de um profissional habilitado a tratar essas informações. A exemplo cita-se o artigo de Tech *et al* (2010) que argumenta sobre a necessidade de um modelo de gestão baseado em *e-science* e *data warehouse* para a aplicação no agronegócio. Outro bom exemplo é o trabalho de Vaz (2011) sobre a importância da EMBRAPA Informática Agropecuária assumir um papel relevante nas iniciativas relacionadas a *e-science* no contexto agropecuário.

Também merecem destaque as oportunidades de emprego de assistente de pesquisa e técnico de programação, divulgadas pelo Laboratório Nacional de Ciência e Tecnologia do Bioetanol, em 29 de julho de 2011 (GF DIVULGA, 2011).

Outra evidência da emergência do tema no Brasil foi a realização do *VI Workshop de e-Science*, no período de 16 a 19 de julho de 2012, na cidade de Curitiba. Em maio de 2013, foi realizado, na cidade de Marília, o *Encontro Internacional de Dados, Tecnologia e Informação*, evento promovido Programa de Pós-Graduação em Ciência da Informação da UNESP. Ressalta-se que está programado para acontecer em outubro de 2014, na cidade de São Paulo, o *IEEE 10th International Conference on e-Science*.

Em consulta no *Diretório dos Grupos de Pesquisas do CNPq*, também foi constatada a incipiência de grupos que tratem do tema, apenas nove grupos foram localizados, a partir do termo indexador *e-science*, conforme pode ser observado na Figura 3.

Figura 3 – Resultado de Busca no Diretório de Pesquisa do CNPq



Fonte: Diretório dos Grupos de Pesquisa do CNPq – 19/03/2013

Por outro lado, em pesquisa que utilizou a web como fonte de dados secundários, identificaram-se na USP os seguintes projetos: a) Modelos e métodos de *e-Science* para ciências da vida e agrárias, b) Núcleo de Pesquisa em *e-Science* da USP, Núcleo de *e-Science* de Apoio a Pesquisa na Universidade de São Paulo e c) Transactional model and performance analysis for business processes and *e-science* applications.

Durante esta pesquisa, destacaram-se como iniciativas no Brasil, o eScience Research Network da Universidade de São Paulo (<http://escience.ime.usp.br/index.php>) e o Sistema Nacional de Processamento de Alto Desempenho (<https://www.lncc.br/sinapad/>).

Apesar do notável avanço do Brasil no *ranking* mundial da produção científica, ainda persiste um hiato significativo em relação aos países desenvolvidos. (BRASIL, 2012). Esse fato é constatado nas produções sobre *e-science*, onde o Brasil aparece com apenas três artigos indexados pela LISA e LISTA. As instituições brasileiras que tiveram artigos indexados foram a Universidade Federal do Rio de Janeiro (UFRJ), a Universidade Federal Fluminense (UFF) e a Universidade Federal de Minas Gerais (UFMG), conforme demonstra o Quadro 1.

Quadro 1 – Autores brasileiros com artigos indexados na LISA e LISTA (2003 – 2013)

AUTOR	TÍTULO	PERIÓDICO	ANO	INSTITUIÇÃO
EBECKEN, Nelson F. F.	A process for innovation in a collaborative network	<i>Journal of Information & Knowledge Management.</i> v. 10, n.2, p. 183-192.	2011	COPE/UJRJ
MOURA, Maria Aparecida	Semantic interoperability and semiotic ontology: the construction and sharing of scientific concepts in environments for online collaboration	<i>Informação & Informação</i> n. especial, p. 165 - 179	2011	UFMG
MARCONDES, Carlos Henrique	Knowledge network of scientific claims derived from a semantic publication system	<i>Information Services & Use</i> v. 31, n.3/4, p. 167-176	2011	UFF

Fonte: Os autores.

Merece ser ressaltado o fato do artigo de Moura (2011) ter sido publicado em um periódico brasileiro. O periódico é produzido pelo Departamento de Ciência da Informação, Centro de Educação, Comunicação e Artes da Universidade Estadual de Londrina (UEL).

No contexto da ciência da informação brasileiro, Moura (2011, p. 165) tem estudado a *e-science* a partir de "uma amostra de blogs científicos mantidos por pesquisadores como estratégia para o registro e a divulgação dos resultados parciais de sua pesquisa, sites colaboratórios internacionais e de centros internacionais que apoiam as praticas". Dentre os resultados da pesquisa da autora, merece destaque a *criação do Online Dictionary of E-Science, Cyberculture and Scientific Narratives*³.

Vaz (2011) faz uma comparação entre o cenário de tratamento da *e-science* no Reino Unido e no Brasil. O autor conclui que no Brasil há poucos cientistas que têm conhecimento ou interesse sobre o tema, evidenciando o atraso do brasileiro nesse cenário. Tal situação reforça a necessidade de se fomentar pesquisas nesse tema. Em especial, a necessidade de pesquisas que apontem a contribuição da Ciência da Informação no tratamento desses dados.

O aspecto transversal da Ciência da Informação faz com que, em algum momento, os dados oriundos da *e-science* convirjam para as preocupações da comunicação da informação e, em outros instantes, para questões inerentes à organização da informação. No que diz respeito à comunicação da informação, Sayão e Sales (2012) argumentam que:

(...) dados e informações digitais gerados pelas atividades de pesquisa necessitam de cuidados específicos, tornando-se necessário a criação de novos modelos de custódia e gestão de conteúdos científicos digitais que incluam ações de arquivamento seguro, preservação, formas de acrescentar valor a esses conteúdos e de otimização da sua capacidade de reuso [...]. É nesse ambiente que surge o conceito de curadoria digital de dados científicos.

³ Disponível em: <http://mamoura.eci.ufmg.br/dictionary/>.

Por outro lado, analisando a *e-science* no campo da comunicação científica, destaca-se o trabalho de Medeiros e Caregnato (2012).

Ainda sobre o cenário brasileiro, Cunha (2010) 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” (CUNHA, 2010, p. 10).

Na visão de Cunha (2010), para a biblioteca apoiar os dados da *e-science* 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. Para o autor:

a inclusão dessa nova área pela biblioteca universitária, por meio do que poderia ser chamado de repositório de dados científicos, não será rápida nem tranquila. Ela exigirá o treinamento dos recursos humanos para assumirem as novas funções inerentes à gestão de dados em formatos e assuntos variados, além do tratamento dos documentos e arquivos científicos disponíveis nos laboratórios de pesquisa. (CUNHA, 2010, p. 11).

Durante a 4ª Conferência Nacional de Ciência e Tecnologia, realizada em Brasília, no período de 26 a 28 de maio de 2010, foi argumentado que nenhum país que tenha como objetivo promover o desenvolvimento para se tornar uma potência de inovação tecnológica pode deixar de investir em tecnologias da informação e comunicação. Ressalta-se que para apoiar o fortalecimento da pesquisa & desenvolvimento, está se falando, ainda que indiretamente, no apoio a projetos da *e-science* que permitam um armazenamento, curadoria, recuperação e colaboração de dados para o avanço da pesquisa no Brasil.

Corroborando com esse aspecto, o antigo Ministro da Ciência Tecnologia e Inovação - Aluísio Mercadante, ao apresentar em 2012, a Estratégia Nacional de Ciência, Tecnologia e Inovação (ENCTI), comentou que “o principal desafio que o Brasil terá de enfrentar se quiser se transformar em um País efetivamente desenvolvido, com uma economia eficiente e competitiva é preparar-se para a sociedade do conhecimento” (BRASIL, 2012, p. 9). Ao se falar em uma sociedade competitiva, retomamos o conceito de sociedade do conhecimento, cenário em que o tratamento de dados da *e-science* torna-se fundamental para o desenvolvimento do País.

4 O bibliotecário e o tratamento de dados da *e-science*

Ao longo do texto acima se pode ver a necessidade de um novo perfil de profissional, apto para trabalhar no contexto virtual e com a

preocupação de preservar informações digitais. Abordar um campo novo como a *e-science* prevê este novo contorno de profissional.

Face ao exposto, cabe aos profissionais da informação entenderem seus papéis nesse importante contexto de mudanças. Pesquisas precisam ser desenvolvidas para clarificar qual a necessidade de informação científica e tecnológica que propicie o avanço tecnológico do Brasil. Os profissionais da informação precisam compreender esse novo cenário de dilúvio de dados e como podem contribuir para uma gestão eficiente da informação.

Na visão de Tenopir, Birch e Allard (2012) a biblioteca, em conjunto com escritórios de pesquisa do *campus* universitário, é o local ideal para apoio à gestão de dados oriundos de pesquisa acadêmica.

A Association of College & Research Libraries (ACRL) identificou, em 2012, que o envolvimento biblioteca na curadoria de dados é uma das dez maiores tendências em bibliotecas universitárias.

Para Choudhury (2010, p. 194) "(...) uma das mais importantes oportunidades para as bibliotecas tornarem-se parceiras no desenvolvimento da ciberinfraestrutura relaciona-se com curadoria de dados". A autora relembra que os bibliotecários da Sheridan Library da Johns Hopkins University colaboraram por anos com astrônomos para melhor entender os requisitos de curadoria de dados.

O diálogo entre os bibliotecários e os astrônomos resultou em uma maior compreensão da natureza transformadora da ciência intensiva de dados, especialmente no que se refere a novas formas de publicação, pesquisa e aprendizagem. Além disso, o diálogo revelou ideias e observações importantes sobre repositórios institucionais no contexto da curadoria de dados (CHOUDHURY, 2008; 2010).

É interessante ressaltar que a School of Information da University of Michigan já oferece uma disciplina de pós-graduação, ministrada pelo Professor Carl Lagoze, cujo título é *Data Curation*. De acordo com a ementa da disciplina⁴, curadoria de dados é "a atividade de gerenciamento de dados desde o seu ponto de criação para garantir que eles estão disponíveis para a descoberta e reutilização no futuro".

Outra iniciativa americana foi a da Berkeley School of Information que criou em 2013 o curso *on-line* multidisciplinar – Master of Information and Data Science, dentre os objetivos do curso, destaca-se o de preparar os alunos para lidarem com as últimas ferramentas e métodos para identificar padrões de dados e obter insight a partir desses dados.

Ross Harvey (2010), em seu livro introdutório sobre o tema, argumenta que curadoria digital "aborda toda a gama de processos aplicados a objetos digitais ao longo do seu ciclo de vida". Nesse sentido, ao autor entende que curadoria digital trata-se de um conceito mais amplo que o arquivamento digital e a preservação digital. O autor afirma que:

⁴ Disciplina SI 754: Data Curation. Fall 2012, ministrada pelo professor Carl Lagoze. Arquivo Disponível em: <<http://www.si.umich.edu/class/fall-2012/data-curation>>. Acesso em: 3 mar. 2013.

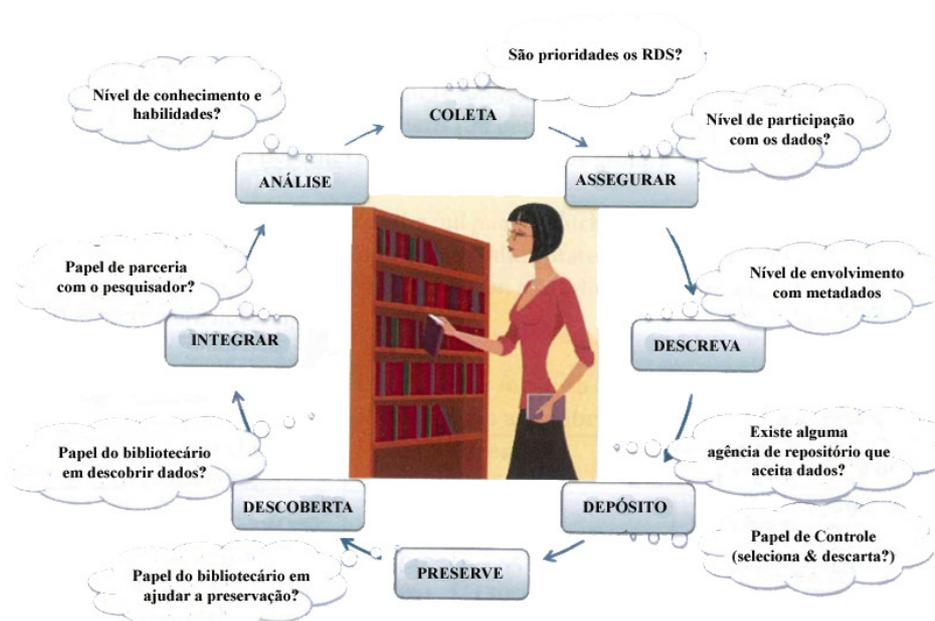
a curadoria digital diz respeito à gestão ativa de dados durante o tempo em que ele continua a ser: acadêmico, científico, pesquisável, administrável e/ou de interesse pessoal, com objetivos de apoiar sua reprodutibilidade, reutilização e agregando valor a esses dados. Gerenciando-os do momento de sua criação, até que eles sejam determinados como não úteis e garantindo a sua acessibilidade em longo prazo, assim como a sua preservação, autenticidade e integridade. (HARVEY, 2010, p. 8).

A definição de Harvey (2010) apresentada acima traz a tona reflexões profundas sobre a Ciência da Informação, a Biblioteconomia e, principalmente, a Arquivologia. Afinal, a ciência que trata da autenticidade, integridade e a acessibilidade em longo prazo é a Arquivologia, cujos princípios arquivísticos, na visão de Rosseau e Couture (1998) são: o princípio da proveniência, o da territorialidade e a abordagem das três idades. Arelados a esses três princípios são tratadas as características dos documentos, dentre elas a autenticidade, imparcialidade etc. Não é objeto de este artigo exaurir as contribuições arquivísticas no tratamento dos dados oriundos da *e-science*, mas é certo que há necessidade de um profissional multidisciplinar, dentre eles o arquivista, com visão holística para o tratamento dos dados da *e-science*.

Sayão e Sales (2012, p. 182) atentos as grandes mudanças que trazem consigo a definição de novos papéis para os profissionais que trabalham com informação, comentam sobre o surgimento do "cientista de dados" e argumentam "(...) apesar de não ser uma carreira de contornos bem definidos e de reconhecimento óbvio, a sua contribuição é fundamental para um diálogo bem sucedido entre as partes envolvidas." Entendem-se aqui como partes envolvidas, as profissões que procuram tratar a informação, ou seja, bibliotecários, arquivistas, cientistas da computação, engenheiros de *softwares* dentre outros.

O momento mostra-se oportuno tanto para profissionais da informação, como profissionais de tecnologia da informação dentre outros. Os limites dessa nova área ainda não estão definidos, por isso faz-se necessário que o bibliotecário posicione-se como um profissional que apresenta capacidades para lidar com o tratamento de dados oriundos *da e-science*. Como exemplo dessa capacidade, cita-se o ciclo do bibliotecário, proposto por Tecnopir, Birch e Allard (2012, p. 12), ilustrado na Figura 4.

Figura 4 – Bibliotecário ponderado **se** possui as habilidades para oferecer pesquisa de dados.



Fonte: Tecnopir, Birch, Allard (2012, p. 12) – tradução nossa.

5 Considerações finais

Certamente, a análise do fenômeno envolve aspectos legais, como, por exemplo, por quanto tempo armazenar dados produzidos no âmbito das ciências da saúde, ciências ambientais dentre outras? Nesse aspecto, a teoria arquivística pode trazer importantes contribuições. Outro aspecto relevante se refere a portabilidade dos dados armazenados, ou seja, sua preservação digital. Qual a infraestrutura tecnológica adequada para armazenar esse volume de dados de forma a garantir sua preservação? No que diz respeito ao tratamento técnico da informação, como se dará a curadoria desses dados? Onde as teorias de classificação, catalogação e indexação podem contribuir nesse cenário? Sob a perspectiva do usuário dos dados, a situação parece mais complexa – Como o pesquisador poderá utilizar ou reutilizar esses dados? Qual o limite da utilização do dado produzido por outro centro de pesquisa?

A partir do exposto neste artigo, é possível afirmar que as discussões sobre *e-science* no âmbito mundial da ciência da informação encontram-se em estágio inicial, sendo que no Brasil, especificamente, encontra-se numa fase embrionária. Pesquisas precisam ser desenvolvidas para identificar como os dados produzidos por estudos conduzidos nas universidades federais, nos institutos federais, centros de pesquisa ou laboratórios de pesquisa estão sendo tratados.

O bibliotecário, principalmente aquele que exerce suas atividades em universidades ou instituições científicas e tecnológicas, precisa se perguntar: *como organizar esses dados produzidos nas universidades e centros de pesquisa? Quais serão as habilidades requeridas para o profissional da informação que deseja trabalhar com organização de dados*

de e-science no Brasil? Como organizar esses dados de forma a disponibilizá-los em rede? Quais as necessidades de informações dos pesquisadores que trabalham com esses dados? Como irei me capacitar para gerir e organizar esses dados?

Merece reflexão por parte dos profissionais da informação brasileiros:

Qual o papel que o IBICT tem no sentido de estimular ações estratégicas para que a formação do curador digital seja uma realidade?

O que as escolas de biblioteconomia e ciência da informação podem fazer para acelerar programas específicos para a formação do curador digital?

As respostas às indagações acima não podem demorar e exigem ações imediatas. A existência do curador digital no contexto brasileiro é, certamente, um dos fatores do sucesso e sobrevivência da *e-science* em nosso país.

Referências

ABBOT, Daisy. *What is digital curation*. Edinburgh, UK: Digital Curation Center, April 2008. Disponível em < <http://www.era.lib.ed.ac.uk/handle/1842/3362> > Acesso: 30 jan. 2013.

ALVARO, Elsa *et al.* E-science librarianship: field undefined. *Issues in Science & Technology Librarianship*, Chicago, n. 66, p. 28-43, Summer 2011.

BELL, Gordon. Prefácio. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Org.). *O quarto paradigma: descobertas científicas na era da e-science*. São Paulo: Oficina de Textos, 2011. p. 11- 15.

BRASIL. Ministério da Ciência, Tecnologia e Inovação. *Estratégia nacional de ciência, tecnologia e inovação 2012 - 2015: balanço das atividades estruturantes de 2011*. Brasília, 2012.

CÉSAR JÚNIOR, Roberto Marcondes. Apresentação à edição brasileira. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Org.). *O quarto paradigma: descobertas científicas na era da e-science*. São Paulo: Oficina de Textos, 2011. p. 7- 8.

CHOU DHURY, Sayeed. Data curation: an ecological perspective. *College and Research Library News*, Chicago, v. 71, p. 194-196, Apr. 2010.

_____. Case Study in Data Curation at Johns Hopkins University. *Library Trends*, Urbana, Fall, 2008. Disponível em <<https://www.ideals.illinois.edu/bitstream/handle/2142/10669/choudhury.pdf?sequence=2>>.

CONFERÊNCIA NACIONAL DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO PARA O DESENVOLVIMENTO SUSTENTÁVEL, 4., 2010, Brasília. *Anais eletrônicos...* Brasília, DF: CGEE, 2010. Disponível em <<http://www.cgEE.org.br/publicacoes/livroazul.php>> Acesso em: 19 mar. 2012. (Livro Azul).

CUNHA, Murilo Bastos da. A biblioteca universitária na encruzilhada. *DataGramaZero: Revista de Ciência da Informação*, v. 11, n. 6, dez. 2010. Disponível em: <http://dgz.org.br/dez10/Art_07.htm>. Acesso em 25 jul. 2012.

FOX, Peter, HENDLER, James. E-science semântica: o significado codificado na próxima geração de ciência digitalmente aprimorada. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Org.). *O quarto paradigma: descobertas científicas na era da e-science*. São Paulo: Oficina de Textos, 2011.

[GF DIVULGA] Vagas em e-Science: Laboratório Nacional de Ciência e Tecnologia do Bioetanol. Blog Antenado: nada definido, de tudo um pouco. Postado por Graça Freitas em 29 jul. 2011. Disponível em: <http://www.temosvagasdeemprego.com.br/2011/07/fwd-sbc-l-vagas-em-e-science.html>. Acesso em: 20 jul. 2012.

GRAY, Jim. eScience: a transformed scientific method. Palestra apresentada no Conselho Nacional de Pesquisa dos Estados Unidos (NRC-CSTB). Mountain View, Califórnia, 11 jan. 2007. Disponível em: <http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt> Acesso em 30 ago. 2012.

GREEN, Daron. Infraestrutura-científica: introdução. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Org.). *O quarto paradigma: descobertas científicas na era da e-science*. São Paulo: Oficina de Textos, 2011. p. 129-130

HARVEY, Ross. *Digital curation*. London: Facet, 2010. 225p. (How-to-do-it Manual, n. 170).

HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Org.). *O quarto paradigma: descobertas científicas na era da e-science*. São Paulo: Oficina de Textos, 2011. 261 p.

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*, Los Angeles, v. 6, p. 956-979, Apr. 2012.

LYMAN, Peter; VARIAN, Hal R. *How much information 2003?* Berkeley, Ca: 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 jul. 2012.

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.

MARCONDES, Carlos Henrique *et al.* Ontologias como novas bases de conhecimento científico. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 13, n. 3, p. 20-30, set./dez. 2008.

MARCUM, Deanna B.; GEORGE, Gerald (Ed.). *The data deluge: can libraries cope with e-science?* Santa Barbara, California: Libraries Unlimited, 2010. 139 p.

MEDEIROS, Jackson da Silva; CAREGNATO, Sônia Elisa. Compartilhamento de dados de e-science: explorando um novo conceito para a comunicação científica. *Liinc em Revista*, Rio de Janeiro, v. 28, n.2, p. 311-322, set. 2012.

MOURA, Maria Aparecida. Interoperabilidade semântica e ontologia semiótica: a construção e o compartilhamento de conceitos científicos em ambientes colaborativos *online*. *Informação & Informação*, Londrina, v. 16, n. especial, p. 165-179, jan./jun. 2011.

MUELLER, Suzana Pinheiro Machado. O crescimento da ciência, o comportamento científico e a comunicação científica: algumas reflexões. *Revista da Escola de Biblioteconomia da UFMG*, Belo Horizonte, v. 24, n. 1, p. 63-84, jan./jun. 1995.

MUELLER, Suzana Pinheiro Machado. Literatura científica, comunicação científica e ciência da informação. In: TOUTAIN, Lídia Maria Batista Brandão (Org.). *Para entender a ciência da informação*. Salvador: EDUFBA, 2007. p. 125-144.

ROUSSEAU, Jean Ives; COUTURE, Carol. *Os fundamentos da disciplina arquivística*. Lisboa: Dom Quixote, 1998. 356 p.

SAYÃO, L. F.; SALES, L. F. Curadoria geral: um novo patamar para a preservação de dados digitais de pesquisa. *Informação & Sociedade*, João Pessoa, v. 22, n. 3, p. 179-191, set./dez. 2012.

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. Documento disponível em: <http://www.arl.org/bm~doc/escience_report2010.pdf>. Acesso em: 18 jul. 2012.

SOLLA PRICE, Derek. *O desenvolvimento da ciência: análise histórica, filosófica, sociológica e econômica*. Rio de Janeiro: Livros Técnicos e Científicos, 1976.

MAYER- SCHONBERGER, Victor; Cukier, Kenneth. *Big data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana*. Rio de Janeiro: Elsevier, 2013

TECH, A. R. B. *et al.* Um modelo de gestão baseado em e-science e data warehouse para aplicação no agronegócio. *Archivos de Zootecnia*, Córdoba (Espanha), v. 59, n. 226, p. 161-168, 2010.

TENOPIR, C.; BIRCH, B.; ALLARD, S. *Academic libraries and research data services: current practices and plans for the future*; White Paper. Chicago, IL: Association of College & Research Libraries, 2012.

VAZ, Glauber José. *E-Science na Embrapa*. Campinas: Embrapa Informática Agropecuária, 2011. 58p. (Documentos, 117).