

Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica

Jackson da Silva Medeiros*

Sônia Elisa Caregnato**

Resumo O trabalho objetiva apresentar um novo conceito que surge visando ao compartilhamento de dados científicos primários: a e-Science. Este conceito nasce a partir dos novos métodos de se fazer pesquisa, em que grandes quantidades de dados científicos primários são gerados por pesquisas em diversos ramos das ciências e processados e armazenados em grandes centros de dados e repositórios responsáveis pelo gerenciamento que permita a cientistas distribuídos pelo mundo acessar e analisar esses dados a fim de reutilizá-los em suas pesquisas. Finaliza observando que a perspectiva de compartilhar dados entre cientistas pode fazer com que as possibilidades de colaboração se tornem cada vez mais reais, criando ambientes onde as viabilidades de compartilhamento de resultados de pesquisas promovam o desenvolvimento da ciência e tecnologia e permitindo que recursos sejam utilizados de forma a avançar a ciência.

Palavras-chave e-science; compartilhamento de dados científicos; exploração de dados; colaboração científica.

Data sharing and e-Science: exploring a new concept for scientific communication

Abstract The study aims to present a new concept of scientific data sharing. This concept originates from the new methods of doing research, in which large amounts of primary scientific data are generated by scientific research in various branches of science, processed and stored in large data centers and repositories allowing scientists world-wide to access and analyze these data in order to reuse them in their research work. It concludes noting that the prospect of sharing data among scientists can make collaboration become ever more real, creating environments where feasibility of sharing research results promotes the development of science and technology and allows resources to be used in order to advance science.

Keywords e-science; scientific data sharing; data exploration; scientific collaboration.

* Doutorando em Comunicação e Informação na Universidade Federal do Rio Grande do Sul. Mestre em Ciência da Informação pela Universidade Federal Fluminense.. Endereço: Programa de Pós-Graduação em Comunicação e Informação, Rua Ramiro Barcelos, 2705, 2º andar, Santana, 90035-007 - Porto Alegre-RS. Tel: (51) 3308-5116 E-mail: jackson.medeiros@ufrs.br

** Doutora em Ciência da Informação pela Sheffield University. Professora do Departamento de Ciências da Informação e do Programa de Pós-Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul. Endereço: Faculdade de Biblioteconomia e Comunicação, Rua Ramiro Barcelos, 2705, Santana, 90035-007 - Porto Alegre-RS. Tel: (51) 3308-5737 E-mail: sonia.caregnato@ufrgs.br

Introdução

O crescimento da internet, da web, das grandes redes e sistemas de informação tem mudado a forma como as mais diversas atividades humanas estão acontecendo. Uma destas mudanças ocorre devido à possibilidade de colaboração proporcionada por redes distribuídas, favorecendo que pesquisadores possam, entre outros aspectos, realizar estudos de forma a compartilhar e aproveitar dados de pesquisas anteriores.

Os serviços disponibilizados em rede trouxeram grande suporte para o intercâmbio de dados e informações entre pesquisadores e instituições. Não há necessidade de deslocamento de pessoal para acesso a grandes acervos de dados. Embora esse processo não seja mais necessário, a comunicação entre os pares não se tornou mais difícil. Pelo contrário, as tecnologias de serviços online, seja em texto, em áudio e/ou em vídeo, permitem que parcerias que antes seriam impossíveis hoje ocorram de forma instantânea.

O desenvolvimento da ciência e da tecnologia tem permitido diversos avanços internos – tanto na própria ciência e na tecnologia – e externos, que desencadeiam melhorias no cotidiano e na vida das pessoas, envolvendo aspectos sociais e econômicos. A evolução contínua da tecnologia sempre suportou o desenvolvimento científico. A partir da Segunda Grande Guerra e, principalmente, do fim do século passado, as evoluções têm ocorrido com rapidez talvez jamais imaginada, mas com o aporte de questões tratadas há séculos por diversos ramos de cientistas, como matemáticos, físicos, químicos, astrônomos, entre outros. Dentre esses avanços, sobressai o tema e-Science.

E-science pode ser entendida como a infraestrutura que visa permitir que cientistas e pesquisadores possam ter acesso a dados científicos primários distribuídos, utilizando acesso remoto a esses conteúdos, mas promover algo que vai além da estrutura informática, ou seja, é a possibilidade e “a capacidade de acessar, mover, manipular e extrair dados [é] a exigência central dessas novas aplicações das ciências da colaboração” (HEY; HEY, 2006, p. 517).

Como em todo fenômeno emergente, muitos questionamentos ainda estão sendo colocados e merecem ser tratados com a devida acuidade. Esses questionamentos são oriundos tanto da novidade do tema quanto do rápido avanço científico e tecnológico existente – uma vez que a e-science está diretamente ligada à tecnologia da informação, acompanha seus avanços de forma paralela.

Essencialmente, é a relação que o homem estabeleceu com a tecnologia¹ ao longo dos anos que permitiu grandes avanços sobre as mais diversas atividades do cotidiano. Como ressalta Solla Price (1976), “temos [...] uma tecnologia altamente científica: as repercussões da ciência modelam nossa vida cotidiana, modelam o destino das nações e a implicação filosófica da Revolução Científica”.

Parece-nos, então, que o potencial informático despendido nesse processo tem construído instrumentos capazes de iniciar o processo de compartilhamento de dados científicos primários²,

¹ Tecnologia aqui se refere à acepção original do termo, ou seja, como o conhecimento que se tem sobre processos, técnicas, métodos etc. que envolvem a atividade humana, como, por exemplo, a ciência.

² Torres-Salinas, Robinson-García e Cabezas-Clavijo (2012) alertam para o fato de poder ser estabelecida uma classificação de dados a partir do tratamento que recebem. Nesse sentido, podem ser categorizados em dados

gerando grandes serviços para que pesquisadores tenham acesso a conteúdos já mapeados e devidamente anotados. No entanto, a vasta empreitada focada na construção de sistemas computacionais deixa de lado, por vezes, os estudos teóricos e conceituais, causando confusões que podem ocasionar conflitos no próprio processo que se propõe estabelecer.

Este trabalho trata de uma proposta exploratória na área de e-Science, buscando delimitar a importância do compartilhamento e reuso de dados científicos primários. Trata de um tema ainda inexplorado nos estudos em Ciência da Informação no Brasil, que nos parece ganhar corpo a partir de atividades desenvolvidas principalmente na Inglaterra e nos Estados Unidos. Seu estudo e amplo debate devem servir como base para um novo paradigma da ciência (GRAY, 2009), onde dados de pesquisa, sejam eles em formato primário (raw data) ou mesmo depois de tratados, possam ser reutilizados por outros pesquisadores a fim de despender menos recursos financeiros e humanos.

Sem a pretensão de trazer respostas prontas, mas sim de vislumbrar um caminho a ser trilhado, o trabalho busca trazer a lume uma breve caracterização da ciência na atualidade, visando mostrar o que Gray (2009) definiu como o quarto paradigma da ciência. Por fim, abre discussões sobre a e-Science, bem como o compartilhamento e reuso de dados científicos primários, visando refletir sobre o tema.

A ciência e a atualidade: breves anotações

Tratar de uma definição de ciência pode tornar-se um tema espinhoso, uma vez que sua definição carece de consenso entre os diversos ramos que tratam de sua existência e de seus processos. No entanto, de forma breve, vamos tratá-la neste trabalho como sendo uma forma de conhecer algo a partir de sistematizações proporcionadas pela utilização de métodos rigorosamente definidos, provendo certa descrição e análise sobre determinada realidade.

Além disso, para Ziman (1979, p. 155), “a ciência é uma forma de conhecimento público”, sendo que seus veículos estão “a serviço do intercâmbio de informações”. Isto é, a ciência é trabalho de uma comunidade, percorrendo desde o processo de sua elaboração, colaborativo em essência, até a validação pública por pares, os quais avaliam o trabalho apresentado aferindo a chancela científica e o repassam a uma comunidade que valida o que foi proposto. Por fim, são necessárias instituições que atuem na função de preservadoras e disseminadoras do conhecimento gerado.

Parece-nos razoável afirmar que, a partir dessas características, uma ciência universal necessita que pesquisadores das mais diversas áreas do conhecimento trabalhem em conjunto, ou seja: pressupõe colaboração. Colaboração pode ser entendida, em linhas gerais, como “trabalho feito em comum com uma ou mais pessoas; cooperação, ajuda, auxílio”, ou, “trabalho, ideia, doação etc. que contribui para a realização de algo ou para ajudar alguém; auxílio”, ou, ainda, “participação numa obra literária, científica etc.” (HOUAISS, 2009). Em termos de ciência, segundo Sonnenwald (2008, p. 645), colaboração é um tipo de “interação que ocorre dentro em

extraídos e não tratados, considerados dados crus, e dados finais de pesquisa. Neste trabalho os denominaremos apenas como dados científicos primários.

um contexto social entre dois ou mais cientistas, facilitando o compartilhamento de significado e de conclusão das tarefas em relação a um objetivo compartilhado mutuamente”.

Pode-se notar, de forma sucinta, que o processo de colaboração ocorre a partir de um processo social entre indivíduos que estão engajados em algum tipo de trabalho, de forma a compartilhar e cooperar para a geração de conhecimento, gerando processos cada vez mais complexos e envolvendo grandes somas de esforços financeiros e intelectuais. Nos últimos anos, principalmente, o processo de colaboração tem se intensificado com as transformações ocorridas na ciência e a facilidade de acesso a outros pesquisadores por meio das tecnologias de informação e comunicação.

O processo de colaboração parece ter ganhado força nos últimos tempos. Esta mudança nas práticas de fazer ciência é observada por Gray (2009), quando descreve que a ciência nasce empírica e vem evoluindo desde os povos primitivos, os quais, inicialmente por meio da magia, faziam ciência como “um meio racional de correlacionar os fatos observados [...] e algumas observações e explicações que seriam gradualmente reunidas, vindo a produzir, um dia, uma visão não-mágica”, passando pelas primeiras ideias sobre astronomia no Egito, a invenção da escrita na Mesopotâmia, a matemática grega etc. Passa-se então para uma ciência capaz de criar e se utilizar de modelos e generalizações para gerar teorias sobre os fatos do mundo para, posteriormente, ser apta a simular fenômenos complexos através de computadores, chegando, por fim, à exploração de dados, ou seja, possibilitar que “os dados sejam capturados por instrumentos ou gerados por simulações antes de serem processados pelo software e para a informação resultante ou o conhecimento ser armazenado em computadores” (GRAY, 2009, p. xix). Isto é sintetizado na figura a seguir.

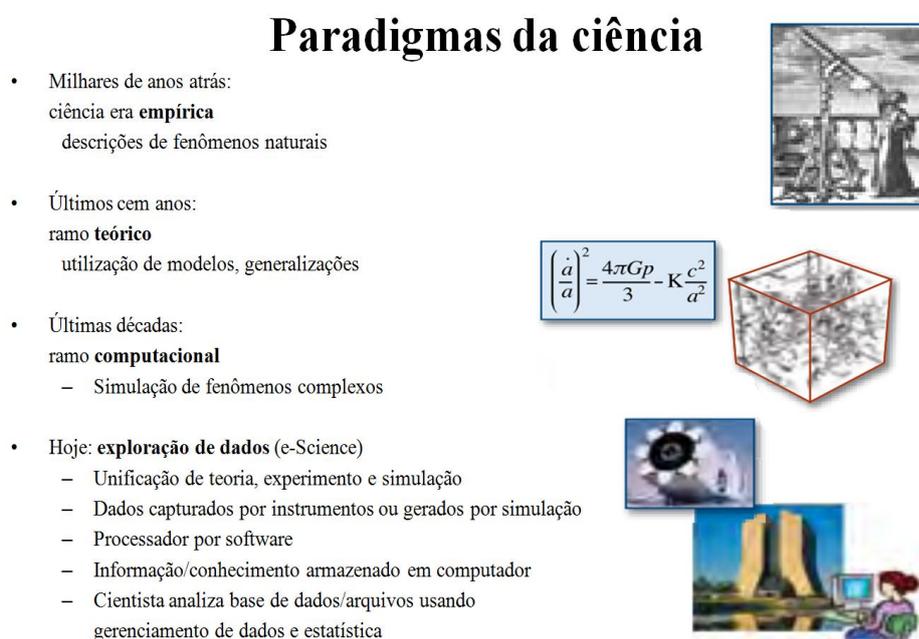


Figura 1: Paradigmas da ciência

Fonte: Traduzido de GRAY (2009, p. xviii)

É possível observar que não necessariamente foi o avanço tecnológico que instituiu o compartilhamento e exploração de dados, mas sim os impulsionou, uma vez que esses procedimentos já eram realizados quando não se dispunha de infraestruturas para abrigar grandes bases de dados, como, por exemplo, através da comunicação informal entre pares (e-mails, sites pessoais etc.).

Meadows, em 1999, já apontava que a existência de redes eletrônicas capazes de suportar a comunicação científica tende a igualar as diferenças existentes entre tipos e níveis de usuários. Assim, “a transmissão por meio de redes pode tornar o conhecimento científico disponível para um público muito maior e de modo mais rápido do que pelos canais tradicionais” (MEADOWS, 1999, p. 113).

Adicionalmente, compreende-se que o modo de fazer pesquisa, ou seja, a metodologia, também “está mudando e estamos no limiar de uma nova era da ciência dirigida por dados” (HEY; HEY, 2006, p. 515).

Esta mudança ocorre, como aponta Marchiori (2002), por diversas formas: (a) avanço tecnológico que suporta o acesso à informação científica; (b) a disponibilidade de conteúdo fortalece a democracia e a sociedade; (c) a capacidade de ter acesso e compartilhar conteúdo tem poder em relação ao setor econômico; (d) conteúdos devem ser organizados e gerenciados; (e) criação, busca, análise e interpretação de informação são essenciais para indivíduos e grupos; (f) necessidades de acesso, consulta e utilização são específicas para cada ator; (g) entender, dominar e gerenciar conteúdo depende diretamente do avanço tecnológico; e (h) o setor informacional faz parte da economia de um país.

A e-Science, como estrutura que visa à colaboração entre cientistas a partir do compartilhamento e gerenciamento de dados científicos primários, parece ganhar corpo, uma vez que é parte essencial de uma descentralização do conhecimento e da aplicação efetiva de recursos públicos em um país com pretensões de avanços significativos em ciência e tecnologia, possibilitando que cientistas de diversos ramos tenham acesso a conteúdo já mapeado.

e-Science: recortando definições

Inicialmente é importante considerar que a e-Science altera fundamentalmente a maneira com que os cientistas realizam seu trabalho, as ferramentas que usam, os tipos de problemas que abordam e a natureza da documentação e da publicação que resulta da sua pesquisa. E-Science requer novas estratégias de suporte à pesquisa e significativo desenvolvimento de infra-estrutura.” (JOINT TASK FORCE ON LIBRARY SUPPORT FOR E-SCIENCE, 2007, p. 6).

Essa nova forma de fazer ciência ocorre, como destaca Gray (2009), porque os tipos de dados, como os resultantes de simulações, por exemplo, geram grande quantidade de conteúdos. Nesse sentido, ressalta o autor, os cientistas em Astronomia, por exemplo, não mais se utilizam de telescópios, mas analisam dados estocados em grandes centros ou bases utilizando-se de computadores pessoais. Assim, ciência e tecnologia atuam de forma imbricada na produção, gerenciamento, acesso e uso de informação. Ou seja, pode-se inferir que a possibilidade de tratar dados, ao invés de ter que coletá-los novamente, possibilita que esforços sejam focados na

análise dos dados já existentes e compartilhados, reduzindo custos e esforços e possibilitando avanços efetivamente mais rápidos.

Dentre as definições que se encontram na literatura sobre e-science³, Hey e Hey (2006) a entendem como um conjunto de ferramentas e tecnologias que devem suportar a ciência em rede. Rambo (2009, p. 159), corrobora com isso, ao afirmar que “um cenário típico de e-science minera dados existentes em busca de padrões e correlações”. Ao enfatizar, novamente, o processo tecnológico, mas notando a presença de outros elementos, Ribes e Lee (2010, p. 232) apontam no sentido de que a e-Science é um nome dado “para as tecnologias de informação em rede de apoio a atividades de investigação científica, como a colaboração de compartilhamento de dados e divulgação dos resultados”.

Pode-se notar três aspectos que tendem a caracterizar as transformações proporcionadas pela e-science: (a) a comunidade de colaboração ampla e interdisciplinar, (b) a coleta, representação e análise de dados dirigidos computacionalmente, e (c) a integração end-to-end (Ribes; Lee, 2010).

Esta última noção permite que seja observado um componente que julgamos essencial à implementação e desenvolvimento da e-science, isto é, além da estrutura física e tecnológica é importante levar em consideração o componente humano e conceitual envolvido nesse fluxo.

Ribes e Lee (2010, p.236) observam que a e-science promete uma diminuição do esforço necessário para a realização de tarefas intensivas ou tarefas tediosas. Os dados devem fluir entre instituições sem a necessidade de renegociar os acordos inter-institucionais, os ciclos de computação devem estar disponíveis sob demanda e a visualização da informação deve facilitar a interpretação e análise científica. Tudo isso é uma forma de automação, ou mais precisamente, uma redistribuição do trabalho entre os seres humanos e tecnologias. (RIBES; LEE, 2010, p. 236).

É importante observar, de forma resumida na tabela 1, a comparação entre as características da e-science e da pesquisa tradicional. Nela é observada a grande dependência entre ciência e tecnologia, bem como a mudança na atuação dos pesquisadores.

Característica	e-Science	Pesquisa tradicional
Participantes	Diversamente qualificados, equipe de pesquisa distribuída	Pesquisador individual ou pequena equipe local de pesquisa
Dados	Gerados, armazenados e acessíveis de localizações distribuídas	Gerados, armazenados e acessíveis apenas localmente
Computação e Instrumentação	Larga escala ou sob demanda ou acesso à informação compartilhada	Emprego da computação em lote ou emprego do computador ou instrumentos do próprio pesquisador

³ A literatura revela que a *e-Science*, além de ter grafias diferenciadas como eScience, escience, EScience etc., também pode receber o nome de Cyberinfrastructure, cyberscience, eInfrastructure e eResearch. Para fins deste trabalho, utilizamos o termo *e-science*

Rede	Confiança na internet e em mediadores	Não confiante na internet
Disseminação da pesquisa	Via websites e portais especializados	Via publicações impressas ou apresentações em conferências

Tabela 1: Características da e-Science x Características da pesquisa tradicional

Fonte: Traduzido de Appelbe e Bannon (2007, p. 84)

Para que isso seja possível, a National Science Foundation (2007), em seu relatório *Cyberinfrastructure Vision for 21st Century Discovery*, revela que devem existir planos de ação a partir de quatro áreas que atuam de forma interdependente: (i) computação de alta performance; (ii) dados, análise de dados e visualização; (iii) organizações virtuais para comunidades distribuídas; e (iv) desenvolvimento de aprendizado e força de trabalho.

Em (i) é importante que os sistemas computacionais sejam provedores e capazes de lidar com quantidades gigantescas de dados, na casa dos petabytes⁴; (ii) requer investimentos nos próprios dados e no processo de sua geração e compartilhamento, bem como em sistemas de representação capazes de promover esse acesso, como metadados e ontologias; (iii) necessita da aplicação de ferramentas e tecnologias que promovam a colaboração entre pesquisadores, inclusive com o design apropriado para isso; (iv) capacitar profissionais para lidar com essa nova infraestrutura, a qual gera trabalho para suporte, desenvolvimento, estudos de usuários, classificação, design etc. (National Science Foundation, 2007).

No entanto, e-science não pode ser definida apenas como uma grande ciência da computação, mas sim “inclui todos os domínios científicos, como biomedicina e ciências sociais, que compartilham abordagens de pesquisa com as ciências” (Soehner; Steeves; Ward, 2009, p. 11). Com essa aproximação, entendemos que as questões que envolvem aspectos políticos, sociais, autorais, éticos, bem como estudos teóricos e metodológicos para implementação dessa citada infraestrutura, também influem de forma ativa nesse processo. Ora, a simples aplicação de tecnologia em um campo, além de criar soluções temporárias, acarreta o desenvolvimento de soluções individuais e que não mantêm possibilidades de reutilização, influenciando também na interoperabilidade de dados.

Uma vez que a e-science é orientada ao compartilhamento e reuso de dados, é relevante buscar algumas definições sobre essa questão. Este é o propósito da próxima seção.

⁴ 1PB = 1.000.000.000.000.000 bytes ou 1 PB = 1024 TB.

Dados científicos primários: compartilhamento e reuso

Com a possibilidade de contar com uma infraestrutura capaz de subsidiar pesquisas através do compartilhamento e reuso de dados já coletados, como exposto, não cabe ao pesquisador despende um longo tempo no processo de coleta de dados se este procedimento já foi realizado e os dados estão disponíveis. Isto contribui também para a produção do conhecimento científico, uma vez que não há necessidade de retroceder à parte de um processo que já foi executada, bem como evita que recursos já concedidos por agências de fomento sejam duplicados. Em países como Estados Unidos e Inglaterra, por exemplo, “o compartilhamento e gerenciamento de dados têm sido o foco principal entre as agências federais de financiamento, universidades e comunidades científicas” (FANIEL; JACOBSEN, 2010, p. 355).

Torres-Salinas, Robinson-García e Cabezas-Clavijo (2012, p. 175) explicam que dados primários são todos aqueles materiais registrados durante uma pesquisa, reconhecidos pela comunidade científica e que servem para certificar os resultados alcançados. Além destas características, também se especificam que devem provir de uma fonte única e devem ser difíceis ou impossíveis de se obter novamente por serem próprios de um momento ou circunstâncias irreplicáveis de uma forma exatamente igual. (Torres-Salinas, Robinson-García e Cabezas-Clavijo, 2012, p. 175).

Podemos citar como exemplos de dados científicos primários e suas utilizações: mapeamento de populações, dados de transações financeiras coletados por instituições bancárias ou, ainda, dados coletados em redes sociais distribuídas pela rede. Esses dados podem ser utilizados para prevenção de doenças e/ou planos de saúde pública, previsões sobre inflação em determinado período, ou análise de comportamento de usuários para marketing.

O National Science Board (2005), conselho governamental da National Science Foundation, afirma que os dados apresentam diferenças que têm implicações de arquivamento e preservação, podendo ser (a) observacionais, (b) computacionais ou (c) experimentais. Dados observacionais são coletados a partir de análises empíricas, como a atitude de eleitores durante uma votação ou a temperatura de um oceano em um período de tempo específico. Dados computacionais são resultados de simulação ou modelos/modelagens computacionais. Os dados experimentais, por sua vez, são resultados de medições de experimentos controlados realizados em laboratório.

Esses dados são armazenados em grandes centros de dados suportados por uma infraestrutura que possibilita a produção, armazenamento, localização, compartilhamento e reuso. A e-science atua como uma empresa colaborativa onde os esforços não estão destacados em apenas determinada parte do processo científico, mas no seu fluxo contínuo, desde a produção de dados, através de simulações, mapeamentos, experimentações etc., passando pelo processo de gerenciamento, até sua utilização em outras pesquisas e/ou projetos através de compartilhamento.

Em uma definição concisa, Borgman (2012, p. 1060) diz que o compartilhamento de dados é a “liberação de dados de pesquisas para o uso de outros”, sendo que esta liberação pode ocorrer de forma privada ou até mesmo através do depósito em grandes centros de dados. De forma privada, ou informal, pode ocorrer o envio de dados primários por e-mail ou outro serviço pessoal. A partir de repositórios, é possível que uma enorme variedade de pesquisadores com os mais diversos níveis de especialização e que estejam espalhados pelo mundo possam atuar no reuso desses dados em suas pesquisas.

Como apontam Torres-Salinas, Robinson-García e Cabezas-Clavijo (2012), dentre as vantagens conseguidas por meio do compartilhamento de dados está o maior aproveitamento de recursos financeiros investidos em ciência, uma vez que permite o reuso de dados para novas análises, fazendo com que não seja necessário iniciar novos projetos para obtenção de resultados similares. Da mesma forma, atuam contra fraudes, permitindo replicar experimentos e verificar as hipóteses apresentadas. Além do mais, esta prática contribuiria para o crescimento das citações dos autores desses dados. Some-se a isso a já mencionada evolução contínua da tecnologia e a tendência do acesso aberto apoiada por diversos pesquisadores e comunidades científicas.

A ideia desses bancos de dados primários é que seja permitida aos pesquisadores a aceleração de pesquisas a partir da utilização de dados em comum, fazendo com que grandes esforços já despendidos não precisem ser repetidos e permitindo que todos os envolvidos contribuam de forma efetiva para criação de repositórios que permitam o reuso de dados em pesquisas futuras.

O reuso de dados científicos pode ser visto, grosso modo, como a utilização de dados já utilizados, ou seja, são dados utilizados mais de uma vez, mas com intenções originais. Em geral, há três perguntas que cientistas consideram ao avaliar a reusabilidade de dados, de acordo com Faniel e Jacobsen (2010): (i) os dados são relevantes?; (ii) os dados podem ser entendidos?; e (iii) os dados são confiáveis?

Nesse sentido, (i) refere-se à avaliação que os cientistas fazem sobre o nível em que os dados estão para atender seus problemas de pesquisa; (ii) diz respeito à possibilidade de entender os dados, em geral, a partir de anotações de metadados e documentação disponível; (iii) alude à questão do entendimento dos cientistas como membros de uma comunidade, onde desenvolvem um sentido de confiança baseado em competência, honestidade, credenciais, reputação etc. dos produtores originais dos dados (FANIEL; JACOBSEN, 2010).

Parte importante desse processo de compartilhamento e reuso diz respeito à noção de que após os “dados serem capturados, necessita-se que sejam curados antes que se inicie qualquer tipo de análise de dados” (GRAY, 2009, p. xvii). Ou seja, como a grande quantidade de elementos oriundos de pesquisas faz com que eles se tornem “pesados”, há a necessidade de não se transferir bytes, mas realizar consultas aos que estão estocados em grandes repositórios (GRAY, 2009). Com isto, os centros devem estabelecer, como ressaltam Soehner, Steeves e Ward (2009), estratégias para curadoria, isto é, o tratamento que eles recebem desde a sua conceitualização ou recebimento até o uso e preservação.

A curadoria, neste caso, está ligada aos processos de gerenciamento de dados científicos primários, permitindo que, a partir da preservação e manutenção dos registros existentes em um centro, seja possível agregar valor aos dados de pesquisa e permitir seu compartilhamento e reuso, sendo isto realizado a partir da incorporação de metadados e documentação que os descrevam. No entanto, uma análise prévia mostra que essas questões seguem incipientes, mas mantêm contínua evolução. É essa curadoria que permitirá o compartilhamento de dados científicos primários.

No entanto, não se pode deixar de salientar o que Faniel e Jacobsen (2010) descrevem como sendo uma grande dificuldade encontrada para o reuso de dados: o contexto de produção. Segundo os autores, o contexto reflete as condições ambientais em que os dados foram produzidos, dificultando o reuso por outros cientistas, fazendo com que seja necessário o aprendizado de ferramentas específicas para esse fim. Além disso, o problema da dificuldade de

interlocução entre quem produz e quem necessita de dados e a crença de que o seu fornecimento amplo, por si só, irá garantir a reutilização são outras barreiras enumeradas pelos autores.

Entende-se que a e-science, como todo novo fenômeno, deve ser estudado e analisado de forma detalhada, especialmente no âmbito da Ciência da Informação, para que esse campo continue contribuindo nos processos ligados à produção, armazenamento, gestão e comunicação dos resultados da ciência e tecnologia. O resultado que se espera disso é que os indivíduos envolvidos sejam capazes de promover o avanço da ciência e, por consequência, o avanço de um país.

Considerações finais

Este trabalho representa uma contribuição inicial para promover o estudo da e-Science, tema ainda em estágio inicial principalmente na Ciência da Informação brasileira. Trata-se de um campo ainda pouco explorado e que merece investigações comprometidas e que possam permitir o estabelecimento de relações ainda não bem delineadas, como, por exemplo, os diversos campos do saber que estão envolvidos nesse processo.

O advento da e-Science necessita, sem dúvida, de investimentos que possibilitem a criação de estratégias para a formação de profissionais em alto nível, com capacidade para atuar no gerenciamento de dados primários, assimilando questões contextuais – de domínio –, reconhecendo a existência de um novo paradigma que deve ser observado com a lucidez e atitude de um gerenciador de informação em um mundo dinâmico.

A aceleração da ciência em todos seus âmbitos traz novos desafios que devem ser pensados a partir de um processo sistêmico, ou seja, devem atuar como uma engrenagem capaz de se movimentar a partir de pequenas peças. Peças estas que devem atuar uniformemente e com constante análise e reparo, uma vez que a ação a partir de coleções de dados primários é uma tarefa que envolve complexidade humana e tecnológica, desde o processo de criação dos dados até os sistemas que suportarão o armazenamento, a preservação, a manutenção, o compartilhamento e o reuso destes dados, os quais são responsáveis por uma infraestrutura de acesso.

A perspectiva de compartilhar dados entre cientistas que estão distribuídos por todo o mundo pode fazer com que as interações e possibilidades de colaboração se tornem cada vez mais reais entre instituições, laboratórios etc., criando ambientes via internet onde as discussões crescem e promovem o desenvolvimento da ciência e da tecnologia, permitindo que recursos sejam utilizados de forma a avançar em passos mais largos a partir do reuso de recursos financeiros, humanos, tecnológicos, entre outros, já empregados.

A e-science promove algo que está ligado à constituição da ciência, isto é, maior possibilidade de cooperação, colaboração e interdisciplinaridade. Necessita-se, entre outros quesitos, de esforços e estudos na forma como os dados são liberados, podendo ser dados fortemente estruturados ou que passaram por um processo de curadoria.

Perspectivas de pesquisas futuras estão sendo delineadas, como definir padrões de metadados que sejam compatíveis com os tipos de dados armazenados; investigar a colaboração entre

pesquisadores, fornecedores e utilizadores de dados, visando a construção de um framework capaz de representar uma maneira eficaz de distribuição e reutilização efetiva de dados científicos primários e coletando subsídios para a melhor gestão de dados; analisar a criação, gerenciamento, armazenamento, preservação, compartilhamento e uso, além de estudar questões autorais e éticas em relação ao depósito de dados científicos primários em bases.

Artigo recebido em 18/07/2012 e aprovado em 06/09/2012.

Referências

APPELBE, B.; BANNON, D. E-research - paradigm shift or propaganda?. *Journal of Research and Practice in Information Technology*, v. 39, n. 2, p. 83-90, May 2007. Disponível em: <<http://www.jrpit.acs.org.au/jrpit/JRPITVolumes/JRPIT39/JRPIT39.2.83.pdf>>. Acesso em: 28 maio 2012.

BORGMAN, C. L. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, v. 63, n. 6, p. 1059-1078, June 2012. Disponível em: <<http://onlinelibrary.wiley.com.ez45.periodicos.capes.gov.br/doi/10.1002/asi.22634/pdf>>. Acesso em: 17 jun. 2012.

FANIEL, I. M.; JACOBSEN, T. E. Reusing scientific data: how earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work*, v. 19, p. 355-375, 2010. Disponível em: <<http://www.springerlink.com.ez45.periodicos.capes.gov.br/content/n14087vn56336218/fulltext.pdf>>. Acesso em: 27 maio 2012.

HEY, T.; HEY, J. E-science and its implications for the library community. *Library Hi Tech*, v. 24, n. 4, p. 515-528, 2006.

HOUAISS, A. *Dicionário eletrônico Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2009.

GRAY, J. Jim Gray on escience: a transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). *The fourth paradigm: data-intensive scientific discovery*. Washington: Microsoft Research, 2009. Disponível em: <<http://language.log.ldc.upenn.edu/myl/JimGrayOnE-Science.pdf>>. Acesso em: 20 maio 2012.

JOINT TASK FORCE ON LIBRARY SUPPORT FOR E-SCIENCE. *Agenda for developing e-science in research libraries final report and recommendations to the scholarly communication steering committee, the public policies affecting research libraries steering committee, and the research, teaching, and learning steering committee*. 2007. 26 p. Disponível em: <http://www.arl.org/bm~doc/ARL_EScience_final.pdf>. Acesso em: 09 jun. 2012.

MARCHIORI, P. Z. A ciência e a gestão da informação: compatibilidades no espaço profissional. *Ciência da Informação*, v. 31, n. 2, p. 72-79, maio/ago. 2002. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/download/159/138>>. Acesso em: 28 maio 2012.

NATIONAL SCIENCE BOARD - NSB. The elements of the digital data collections universe. In: _____. *Long-lived digital data collections enabling research and education in the 21st century*. Cap. 2. Disponível em: <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540_4.pdf>. Acesso em: 09 jun. 2012.

NATIONAL SCIENCE FOUNDATION. *Cyberinfrastructure vision for 21st century discovery*. 2007. Disponível em: <<http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>>. Acesso em: 13 jun. 2012.

PRICE, D. S. *A ciência desde a Babilônia*. São Paulo: EDUSP, 1976. 189 p. (O homem e a ciência; v. 2).

RAMBO, N. E-science and biomedical libraries. *Journal of the Medical Library Association*, v. 97, n. 3, p. 159-161, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2706433>>. Acesso em: 15 abr. 2012.

RIBES, D.; LEE, C. P. Sociotechnical studies of cyberinfrastructure and e-research: current themes and future trajectories. *computer supported cooperative work*, v. 19, n. 3-4, p. 231-244, 2010. Disponível em: <<http://www.davidribes.com/storage/Ribes%20Lee%20-%20Cyberinfrastructure%20Studies.pdf>> Acesso em: 08 jun. 2012.

SONNENWALD, D. H. Scientific collaboration. *Annual Review of Information Science and Technology*, v. 42, n. 1, p. 643-681, 2008.

TORRES-SALINAS, D.; ROBINSON-GARCÍA, N.; CABEZAS-CLAVIJO, A. Compartir los datos de investigación em ciência: introducción al data sharing. *El profesional de la información*, v. 21, n. 2, p. 173-184, mar./abr. 2012.

YAKEL, E. Digital curation. *OCLC Systems & Services*, v. 23, n. 4, p. 335-340, 2007.

ZIMAN, J. M. *Conhecimento público*. São Paulo: EDUSP, 1979. 164 p. (O homem e a ciência; n. 8).