



VOLUMEN 3 NÚMERO 1 2016

The Revista Internacional de

Ciencia y Sociedad

A e-Science e as novas práticas de produção colaborativa de conhecimento científico

ANDRE LUIZ APPEL
MARIA LUCIA MACIEL
SARITA ALBAGLI

A e-Science e as novas práticas de produção colaborativa de conhecimento científico

Andre Luiz Appel, Universidade Federal do Rio de Janeiro, Brasil
Maria Lucia Maciel, Universidade Federal do Rio de Janeiro, Brasil
Sarita Albagli, Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasil

Resumo: *Contempla uma análise da relação entre novas práticas de produção colaborativa de conhecimento científico e o desenvolvimento e uso de plataformas tecnológicas de amparo ao compartilhamento de dados, movimento conhecido como e-Science, levando-se em consideração as diferentes visões, perspectivas e interesses dos atores atuantes nessas práticas, suas opções de uso e adesão às plataformas de pesquisa de e-Science e as definições quanto aos direitos de acesso e uso dos dados e resultados de pesquisa em tais práticas. O trabalho metodológico envolveu pesquisa e análise documental e bibliográfica, além de estudo empírico, com a realização de entrevistas, com pesquisadores e especialistas brasileiros atuantes em colaborações junto ao Conseil Européen pour la Recherche Nucléaire (CERN). Dentre os principais resultados e considerações estão: (a) a percepção do papel da governança nas formas de organização dos atores e dos grupos de pesquisa para viabilizar a colaboração como uma dimensão significativa no contexto da e-Science; (b) as implicações referentes às condições de financiamento à pesquisa; (c) o novo papel dos dados como ativos importantes nos processos de produção de ciência, e (d) como isso afeta as estruturas de avaliação e mensuração de resultados da pesquisa nesses processos. Em linhas gerais, verificou-se que, mais do que as questões de ordem técnica e tecnológica, são as condições institucionais e culturais que mais interferem nas possíveis novas formas de colaboração científica em torno da e-Science.*

Palavras-chave: *e-Science, uso intensivo de dados, pesquisa colaborativa*

Abstract: *This work aimed to analyse the relationship between the collaborative production of scientific knowledge and the development and use of technological platforms as a support for research data sharing, commonly called the e-Science movement, considering diverse views, perspectives and interests of participants in these practices, their options of use and adherence to e-Science research platforms as well as the rights and conditions of access and use of the data that emerge from these practices. The methodological approach involved a systematic bibliographic analysis complemented by interviews with Brazilian researchers and experts involved in the Conseil Européen pour la Recherche Nucléaire (CERN) collaboration. The main results of the study are: (a) the perception of the importance of governance in the organization of actors and research groups to facilitate collaboration as a meaningful dimension in the context of e-Science; (b) the implications for e-Science regarding research funding options and conditions; (c) the new role of data as important assets in science production processes, and (d) how this new role of data affects the evaluation of Science production and the measurement of search results. In general, it was found that more than the issues of technical nature, is the institutional and cultural conditions that most interfere in possible new forms of scientific collaboration that emerge from e-Science movement.*

Keywords: *e-Science, Data Intensive Science, Collaborative Research*

Introdução

Este trabalho deriva do interesse no estudo da relação entre novas práticas de produção colaborativa de conhecimento científico e o desenvolvimento e uso de plataformas tecnológicas de amparo à pesquisa colaborativa. A emergência de novas plataformas tecnológicas adequadas ao compartilhamento de grandes quantidades de dados e de práticas colaborativas de pesquisa daí derivadas tem sido identificada por diferentes autores sob o nome de *e-Science* – e nomenclaturas similares –, e compreende, primordialmente, aspectos como o uso intensivo de dados na pesquisa científica, o uso de infraestrutura tecnológica para a colaboração entre cientistas, mediados pelas tecnologias.

Com base nesse cenário, foi possível formular alguns questionamentos relacionados ao fenômeno da *e-Science*, assim como possíveis rotas metodológicas que pudessem auxiliar na elucidação e compreensão de tais questões. Buscou-se privilegiar o entendimento de quem são os principais

atores (e seus papéis) envolvidos na *e-Science*, suas formas de interação e organização; como essa interação interfere ou recebe interferências nas/das condições sob as quais são estruturadas as plataformas para desenvolvimento da *e-Science*; e como ocorrem as formalizações dos processos de colaboração em tais plataformas, assim como seus padrões de interoperabilidade. Da mesma forma, buscou-se identificar onde e como são armazenados os dados ou os principais serviços/plataformas de *e-Science*, bem como as formas e condições de uso e compartilhamento de dados e resultados de pesquisas que fazem uso de tais plataformas.

Um outro questionamento compreendeu a necessidade de entendimento de como se manifestam as disputas por prioridade ou recompensa a partir dos resultados de pesquisas desenvolvidas nas plataformas de *e-Science*, ou seja, como se dão os processos decisórios sobre essas e outras questões em que estão presentes os mais variados interesses, sejam eles políticos, econômicos ou científicos.

Nesse sentido, com a realização deste trabalho, objetivou-se analisar os fatores que influenciam as práticas de colaboração científica a partir da introdução da *e-Science*, levando-se em consideração as diferentes visões, perspectivas e interesses dos atores atuantes nessas práticas, suas opções de uso e adesão às plataformas de pesquisa de *e-Science* e as definições quanto aos direitos de acesso e uso dos dados e resultados de pesquisa em tais práticas. A análise deu-se a partir do levantamento do estado da arte do tratamento dessa questão na literatura e a partir do estudo de uma experiência em *e-Science*, mais especificamente, o caso da Organização Europeia para a Pesquisa Nuclear, do francês *Conseil Européen pour la Recherche Nucléaire* (CERN). Para desenvolvimento desse estudo de caso, foram entrevistados pesquisadores e especialistas atuantes em colaborações junto ao CERN e, complementarmente, algumas outras experiências similares de atuação em projetos de *e-Science*.

Com base nessa abordagem empírica, foram elucidados vários dos questionamentos levantados no início da pesquisa, assim como novos elementos de análise também foram observados, a exemplo do destaque da governança como uma dimensão significativa no contexto da *e-Science*. Destacaram-se, também, implicações referentes às condições de financiamento à pesquisa e às formas de organização dos atores e dos grupos de pesquisa para viabilizar as colaborações, além da observação dos dados como ativos importantes nos processos de produção de ciência, e como isso afeta as estruturas de avaliação e mensuração de resultados nesses processos.

A e-Science e suas dimensões: tecnologias, dados e colaboração

O termo *e-Science*¹ foi introduzido por John Taylor no ano de 2001, enquanto diretor geral do Conselho de Pesquisa do *Office of Science and Technology* (OST) do Reino Unido, para denominar novas formas de colaboração e o trabalho multidisciplinar em áreas-chave da ciência e referindo-se à infraestrutura necessária para possibilitar essas novas formas de trabalho (Hey e Trefethen 2002).

Entre outros fatores que impulsionam a *e-Science*, constata-se a gênese de uma ciência que passou a se valer de uma vertente, quase que exclusivamente, computacional em algumas áreas do conhecimento, com forte apelo à pesquisa por meio da simulação de fenômenos complexos. Vislumbra-se, então, um momento da exploração de grandes volumes de dados, em que se busca a interlocução de teorias, experimentos e simulações, caracterizando esse momento da ciência como *data-intensive* (Gray 2009; Bell, Hey e Szalay 2009).

Esse volume de dados deriva de uma ampla variedade de tecnologias e da colaboração entre pessoas, realizada em larga escala. A aplicação de sensores, equipamentos de leituras e medições automatizadas, experimentos com conjunto de dados retroalimentáveis, sistemas de informação e bases de dados alimentadas por cientistas que registram e realizam anotações em dados de pesquisa são apenas alguns exemplos de possíveis fontes geradoras de dados. O trabalho na formulação de modelagens ou metodologias para análise desses dados também intensifica os fluxos de colaboração entre instituições e cientistas.

¹ 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* (Medeiros e Caregnato 2012, 315). Para esta pesquisa, optou-se pelo uso do termo *e-Science*.

A armazenagem e a análise de tais dados requerem, por sua vez, a preexistência de uma infraestrutura computacional robusta, expansível e, preferencialmente, que possa ser organizada ou acessada de forma distribuída, para que os cientistas possam contribuir, por exemplo, na análise dos dados a partir de qualquer parte do planeta, desde que disponham de acesso à internet.

Em linhas gerais, baseando-se em tais evidências conceituais, é possível apontar ao menos três fatores básicos para o desenvolvimento da *e-Science*:

- a) a construção de uma infraestrutura computacional para uso distribuído ou para processamento de larga escala;
- b) a produção e o uso intensivo de dados; e
- c) a colaboração entre cientistas, grupos de cientistas ou instituições, pelo compartilhamento de esforços, dados e/ou recursos computacionais.

Schroeder (2008) reforça essa ideia apontando como possíveis dimensões do conceito de *e-Science* o uso de tecnologias computacionais para o compartilhamento distribuído de recursos digitais na pesquisa científica ou acadêmica e a promoção da colaboração global, não se aplicando apenas à questão do uso da internet, mas considerando também a dimensão institucional da *e-Science*. Tal argumentação amplia o horizonte conceitual da *e-Science*, acrescentando ou reforçando as dimensões sociais, políticas e econômicas do fenômeno, para além da faceta tecnológica.

A implementação de infraestrutura tecnológica para *e-Science* é baseada na construção e disponibilização de *Grids*, *Middleware*, *Workbenches*, *Webservices*, *Virtual Research Environments* (VRE), tecnologias de notação e armazenamento de dados concebidas em padrões como XML², entre outros elementos técnicos.

O termo *Grid*, computação em grade, foi utilizado pela primeira vez no final dos anos 1990 (Foster e Kesselman 1999), para denotar uma infraestrutura computacional distribuída servindo a atividades de ciência e engenharia avançadas. À época, a concepção dessa ideia foi direcionada pela vontade de se utilizar um *cluster*³ de recursos de computação distribuídos como um metacomputador. O nome *grid* é derivado da analogia com a rede/grade de energia elétrica (do inglês *electricity power grid*), compreendendo a ideia de que o poder de computação poderia estar disponível para qualquer pessoa em qualquer lugar (Hey e Trefethen 2005) e por um baixo custo.

Os *grids* compreendem a disponibilização de uma infraestrutura de *hardware* e *software* que fornece acesso a recursos computacionais de alto nível (com alto poder de processamento de dados e informações), confiáveis, consistentes, abrangentes e de baixo custo (Foster e Kesselman 1999). A questão da confiabilidade está ligada ao fato de os usuários requererem altos níveis de previsibilidade, sustentabilidade e desempenho a partir dos recursos presentes nos *grids*. Esses fatores são associados, em geral, à velocidade de banda larga, ao poder de computação e aos serviços de *software* e segurança. A consistência envolve a implementação de serviços, interfaces e operações padrão, para que todos os usuários tenham um mesmo entendimento das formas de uso dessas tecnologias. A abrangência requer que os serviços oferecidos via *grid* sejam acessíveis a partir de qualquer ambiente em que se deseje implementá-los. O baixo custo é um fator que se desenvolve na medida em que as tecnologias se tornam amplamente aceitas e utilizadas.

O *Middleware* corresponde ao *software* disponível entre a infraestrutura computacional dos *grids*, distribuída em rede, e as aplicações de uso individual dos cientistas ou projetos e instituições, e permite que cada usuário possa compartilhar seus recursos ou ter acesso aos recursos de outros usuários (Schroeder 2008). De forma literal, *middleware* corresponde ao *software* que “está no meio” (*middle*), fazendo a ponte entre as aplicações e a rede (Hey e Trefethen 2005). Um conjunto de *software* com módulos, funções e aplicações pré-definidas é comumente denominado *Workbench*, grosso modo chamada de bancada de trabalho.

² *eXtensible Markup Language* “é um formato simples de marcação, baseado em texto, para representar informações estruturadas: documentos, dados de configuração, livros, transações, faturas e muito mais. É derivado de um formato padrão mais antigo denominado SGML (ISO 8879), a fim de ser mais apropriado para o uso na Web” (Quin 2010).

³ Conjunto de computadores conectados.

Webservice é um sistema de *software* projetado para suportar interações interoperáveis máquina-a-máquina em uma rede⁴. Apresenta uma interface descrita em um formato processável por máquina/computador, ou seja, é uma camada intermediária que permite a comunicação entre o *software* ou sistemas presentes nos *grids*.

A segunda dimensão significativa da *e-Science* – e um dos objetos da implementação da infraestrutura anteriormente descrita – compreende a exploração de grandes quantidades de dados que são geradas ao longo de atividades de pesquisa. O fenômeno da geração de grandes volumes de dados vem sendo recorrentemente identificado como *big data*, dados que excedem as capacidades convencionais de processamento dos sistemas de bases de dados. Esses dados variam em volume, velocidade de crescimento e variabilidade (sofrendo transformações), dificultando sua estocagem nas arquiteturas tradicionais de banco de dados, sendo que, para se obter algum valor a partir dos mesmos, existe a necessidade de se apelar para processos alternativos de armazenagem e processamento (Dumbill 2012).

O fenômeno do *big data* vem sendo explorado, além do campo científico, também por empresas e demais segmentos que trabalham com dados e informação, com finalidades que extrapolam a pesquisa científica. Os dados podem ser coletados a partir de textos de discussões e interações em redes sociais, *logs*⁵ de servidores da internet, sensores de tráfego, imagens de satélite, transações bancárias, dados do mercado financeiro, entre outros (Dumbill 2012), com as mais diversas finalidades, tais como estudos preditivos, estudos de comportamento dos consumidores, estudos de planejamento e gestão urbanos etc.

No campo da ciência, o termo *e-Science* é mais comumente utilizado, pois por si só já denota o uso intensivo de dados, e sua finalidade é estritamente a pesquisa científica. Outro fator que diferencia a *e-Science* da simples exploração de *big data* está na pesquisa colaborativa e no uso de recursos compartilhados para a exploração de dados, tal como se discute a seguir.

Na *e-Science*, ao menos três ações são tidas como chave quanto ao uso de dados: captura, curadoria e análise (Gray 2009). No que diz respeito à captura, vislumbram-se processos nos quais os dados são gerados diretamente por instrumentos de captura ou por meio de simulações computacionais prévias, processados por *software* e armazenados em bancos de dados. Nesse processo, o papel de muitos cientistas é o de codificar as suas informações para que as mesmas possam ser intercambiadas com outros cientistas. Para isso a informação precisa estar representada de forma algorítmica, por meio de representações padrão de objetos em análise – tais como genes, galáxia, temperatura etc. (Gray 2009). A curadoria compreende a “gestão atuante e a preservação de recursos digitais”, durante seu ciclo de vida, “tendo como perspectiva o desafio temporal de atender a gerações atuais e futuras de usuários” (Sayão e Sales 2012, 184). A análise envolve o uso de tecnologias computacionais distribuídas e compartilhadas, os *grids*, as quais permitem o uso de recursos tecnológicos ou habilidades analíticas de cientistas posicionados em diferentes localidades.

Sobre os aspectos que dizem respeito à gestão e ao amadurecimento de soluções práticas voltadas para o uso e compartilhamento de dados científicos, algumas questões emergem, tais como: onde os dados são armazenados? Como são organizados? Quem tem acesso? Quem paga pela infraestrutura de armazenagem, compartilhamento e análise desses dados? (Ohno-Machado 2012).

As iniciativas de *e-Science* estão frequentemente associadas ao desenvolvimento de uma ciência aberta (*open Science*), principalmente quanto à disponibilização e manutenção de bases de dados abertos, de acesso público, que subsidiem o trabalho de pesquisa tanto no âmbito individual como no colaborativo (Albagli, Appel e Maciel 2014). A formação desses conjuntos de dados é, em geral,

⁴ World Wide Web Consortium (W3C). 2004. “Web Services.” *Web Services Glossary*. Acesso em 3 jun. 2013. <http://www.w3.org/TR/ws-gloss/#webservice>.

⁵ Um arquivo de *log* contém um registro de eventos gerados por *software* ou a saída de um dispositivo de *hardware*. A maioria dos arquivos de *log* são salvos em um formato de texto simples, permitindo sua visualização em um editor de texto básico. Podem também ser gerados por servidores *Web*, instaladores de *software*, ou uma variedade de aplicações. Os dados armazenados em um arquivo de *log* podem ser utilizados para solucionar problemas de *software* ou para a geração de relatórios (Christensson 2010).

decorrente de resultados de projetos de pesquisa financiados com recursos públicos, sendo disponibilizados para acesso público após um período de carência.

Buscando-se garantir a reprodutibilidade de posteriores estudos, os conjuntos de dados são “congelados” no momento da publicação e raramente sofrem modificações. Os novos resultados são incorporados como novas versões, servindo a pesquisas posteriores (Sánchez-Artigas e García-López 2010). Ohno-Machado (2012) acrescenta que a reprodutibilidade é de difícil realização, pois demanda os mesmos ambientes de *software* usados nos experimentos originais, requerendo instalações e configurações extensivas. A ideia do congelamento subentende a necessidade de os dados serem salvos ou publicados com um nível máximo de detalhamento e padronização, dispensando a necessidade de alterações posteriores.

No caso das ciências biomédicas, em especial, Ohno-Machado (2012) salienta que muitos periódicos dessa área ainda não estão equipados para implementar processos de revisão ou manutenção de grandes bases de dados anotados⁶ ou aplicações de *software*, e pequenos grupos de pesquisa podem não ter recursos para manutenção de dados, metadados e recursos de *software* desenvolvidos para uso primário *in loco*. O acesso a dados gerados em outras localidades é igualmente difícil, pois mesmo que existam repositórios públicos, uma grande variedade de dados não está corretamente representada, tornando-se pouco útil. Nas ciências da saúde, existe ainda a falta de padronização dos dados advindos de diferentes instituições, além da necessidade de ações que permitam a proteção da privacidade institucional e pessoal (de pacientes), tal como um quadro de políticas que garantam níveis apropriados de consentimento e concordância com as regulações estabelecidas nos âmbitos institucional, estadual, federal e internacional (Ohno-Machado 2012).

Sobre a disponibilização de dados abertos para a pesquisa científica, David, Besten e Schroeder (2006) argumentam que o crescimento dos estoques de tais dados depende da sua disponibilização como ‘bens públicos’ (*public goods*), que permitam o acesso e o compartilhamento irrestritos. Segundo os autores, esse panorama contrasta com as ações de controle e restrição do acesso às informações que geralmente são requeridas em favor da obtenção de benefícios materiais privados pela posse de conhecimentos científicos e tecnológicos, como os registros de propriedade intelectual. Para David, Besten e Schroeder (2006), em meio a um regime de propriedade intelectual, em que se busca a exploração comercial e rentável de descobertas e invenções, os investimentos de capital privado em pesquisa e desenvolvimento (P&D) são realizados com base na expectativa de obtenção de lucro. No entanto, lembram os autores, boa parcela dos conhecimentos utilizados pelo setor privado advém de pesquisas oriundas de financiamento público.

Outra dimensão da *e-Science*, a colaboração científica, é definida por Sonnenwald (2007, 645, tradução nossa) como “a interação que se situa em um contexto social entre dois ou mais cientistas, a qual facilita a partilha do propósito e da ação de execução de tarefas, respeitando a um objetivo superordenado e compartilhado mutuamente”. A autora lança mão de características de contextos científicos descritas na literatura para categorizar ou classificar a colaboração a partir de categorias disciplinares, geográficas ou organizacionais.

De acordo com a autora, novos tipos de colaboração estão sendo e ainda podem ser gerados ou estimulados a partir do uso ou implementação de tecnologias da informação e da comunicação (TICs), especialmente quando os cientistas não podem ou não devem estar alocados/fixados geograficamente (Sonnenwald 2007). A alocação de recursos digitais geograficamente distribuídos, o poder de computação (capacidade de processamento), as tecnologias de visualização de dados e a capacidade de armazenamento são colocados como os principais esforços para a construção de TICs como infraestrutura de suporte à pesquisa colaborativa (Barjak et al. 2013).

Em meio à gênese de inúmeras ferramentas e plataformas colaborativas da internet, e com finalidades igualmente múltiplas, Dutton (2008) destaca, no âmbito da pesquisa científica, o conceito de “organização rede”. Segundo o autor, tal conceito é útil para distinguir esta forma organizacional daquelas mais formais ou localmente/fisicamente baseadas, sendo que a organização em redes cola-

⁶ Um dado anotado recebe uma descrição, comentário, marcação ou um conjunto de descritores (metadados, ou dados sobre dados) que facilitam a sua organização, arquivamento e recuperação, além do seu uso por outros pesquisadores.

borativas promove configurações dinâmicas, e na maioria das vezes inter-organizacionais e inter-pessoais.

Dentre as ações que mediam a formação de redes colaborativas e iniciativas que privilegiam o avanço da pesquisa científica por meio da internet, pode-se destacar (Dutton 2008):

- a) a ação de compartilhar, como a habilidade de criar objetos e documentos conectados por meio de uma rede distribuída, desse modo reconfigurando como e com quem determinadas informações são compartilhadas;
- b) a ação de contribuir, incluindo a habilidade de empregar aplicações de social networking da internet para facilitar a comunicação entre grupos; e
- c) a ação de cocriar, a qual implica a habilidade de colaborar por meio de redes para facilitar o trabalho cooperativo em direção a objetivos comuns, desse modo reconfigurando o sequenciamento, composição e definição de papéis dos colaboradores.

Por outro lado, as infraestruturas tecnológicas para pesquisa colaborativa podem ser usadas para facilitar a formação de ‘clubes fechados’, embora localizados de forma distribuída – incluindo laboratórios governamentais engajados em projetos secretos de defesa e equipes corporativas de P&D que trabalham com dados e materiais proprietários, cujos resultados são guardados como segredos comerciais. Um sintoma disso é o fato de que muitos pesquisadores empregam *software* proprietário e instrumentos patenteados e publicam suas descobertas em revistas científicas que cobram taxas elevadas de acesso aos artigos, embora divulguem suas descobertas frequentemente e por completo e colaborem livremente com colegas em bases informais e não contratuais (David, Besten e Schroeder 2006).

Sobre as implicações das tecnologias no processo da comunicação científica, tanto formal quanto informal, Meyer e Schroeder (2009) argumentam que esse processo vem ocorrendo crescentemente na internet, impulsionado pela praticidade e rapidez de acesso à informação quando comparado aos recursos tradicionais de busca por informação. Esse fator também pode ser explicado, de acordo com os autores, pelo crescimento do número de jovens pesquisadores, nativos da *web*. Com base em Garvey e Griffith (1967), argumentam que os próprios cientistas tendem a criar elementos que preencham suas necessidades de informação não satisfeitas pelos canais tradicionais existentes, fortalecendo a comunicação científica informal na internet, por meio de *blogs*, *wikis*, redes sociais etc. A presença *online* também amplia a visibilidade das pesquisas, possibilitando a aproximação de iniciativas correlatas.

Práticas de colaboração em e-Science: o caso do CERN

O CERN se estabeleceu como organização no ano de 1954, na cidade de Genebra, na Suíça, em uma região de fronteira com a França, tendo como objetivo inicial o estudo do interior do átomo, de onde deriva o termo ‘nuclear’ presente no nome da organização. Atualmente, os interesses de estudo do CERN concentram-se na física de partículas, ou o estudo dos constituintes fundamentais da matéria e das forças que agem sobre ela (CERN 2014a).

No CERN existe um conjunto de detectores, que são estruturas responsáveis pela coleta de dados derivados das colisões que ocorrem no *Large Hadron Collider* (LHC), um túnel de 27 km de extensão em que são realizadas experiências de aceleração de partículas. Cada detector é formado por uma série de subdetectores, cada qual projetado para permitir a observação de propriedades particulares ou de tipos específicos de partículas. Incluem dispositivos de rastreamento que revelam a trajetória de uma partícula; calorímetros que retêm, absorvem e medem a energia de uma partícula; e detectores de identificação de partículas, que usam uma variedade de técnicas para definir a identidade de uma partícula (Cern 2014b). Dentre os principais detectores estão o ATLAS (*A Toroidal LHC Apparatus*), CMS (*Compact Muon Solenoid*), ALICE (*A Large Ion Collider Experiment*) e o LHCb (*LHC beauty*). As colaborações do CERN, iniciadas por volta dos anos 1990, foram formadas para permitir o estudo dos dados gerados por cada um dos experimentos conduzidos nos detectores, os quais correspondem a diferentes ‘tipos de física’ ou fenômenos de interesse.

A seguir, é feito um relato sobre como ocorrem as colaborações no CERN e iniciativas de *e-Science* similares, com base nas entrevistas realizadas na pesquisa empírica.

As entrevistas foram realizadas junto a um grupo de pesquisadores de instituições brasileiras – nas quais estão implementados *grids* conectados ao CERN – que já estiveram ou estão envolvidos diretamente em colaborações do CERN, investigando como se dá o processo de colaboração em suas interações com instituições e com outros pesquisadores, e levantando o ponto de vista desses pesquisadores sobre os seus papéis ou suas funções nesse contexto. Como complementação, optou-se também pela entrevista com um segundo grupo de pesquisadores não relacionados diretamente ao CERN, mas com atuação em outros projetos ou iniciativas relacionadas à *e-Science*, buscando-se captar suas experiências com a *e-Science* ou com estudos sobre pesquisa colaborativa, sob a perspectiva do tratamento dos dados ou da infraestrutura computacional.

Foram entrevistados oito pesquisadores, entre outubro de 2013 e janeiro de 2014. Cinco desses pesquisadores são da área de física e três são das áreas de ciências da computação ou engenharias. A entrevista foi pautada por um roteiro, explorando conceitos ou temas como a atuação no CERN, as condições de participação nas colaborações, os procedimentos para publicação e atribuição de créditos de resultados alcançados pela colaboração e benefícios ou barreiras/dificuldades decorrentes da colaboração brasileira e o uso de plataformas ou recursos da *e-Science* junto ao CERN.

Dos entrevistados que atuam ou colaboram diretamente com o CERN, todos apontaram que iniciaram essa colaboração por volta do início da década de 1990, sendo que as formas de colaboração ou atuação envolviam, primordialmente, a realização de análises de dados, a atuação em serviços da colaboração – como a identificação de partículas e aprimoramento de *software* –, desenvolvimento de detectores e sistemas a estes associados e sua preparação e atualização visando coletas de dados futuras, participação em reuniões presenciais e por videoconferência, além de ações locais de construção e teste de sistemas de detecção e de aceleração e de manutenção de infraestrutura computacional para atendimento à colaboração.

Dos entrevistados que não atuam diretamente no CERN, um deles indicou participação em projetos ou ações de mensuração e avaliação da participação de pesquisadores em colaborações científicas. Outro indicou a participação em iniciativas para o desenvolvimento de instrumentação científica – eletrônica para instalações/laboratórios científicos, incluindo circuitos integrados, aceleradores etc. – com base em licenças *Open Hardware* (*'hardware aberto'*), desenvolvidas pelo CERN. E o terceiro mencionou a participação em núcleo de pesquisa em computação e matemática, que realiza parceria com grupos ou projetos de pesquisa de outras áreas voltados para a pesquisa com a utilização de ferramentas da *e-Science*, em especial nas áreas de biologia e medicina, no trabalho de integração de bancos de dados de genoma humano.

Sobre as formas ou condições de participação no CERN, observou-se que a organização conta com um Conselho Superior, que institui sua instância deliberativa superior e que é composto de membros associados, ou seja, países que ingressam pelo pagamento de uma cota financeira, proporcional ao seu Produto Interno Bruto (PIB). Verificou-se também a existência de uma instância científica, formada por integrantes das colaborações, sendo que cada uma dessas é independente, possuindo estatuto próprio, e composta por diferentes conselhos/*boards* (editorial, executivo etc.). Tais conselhos são responsáveis pelas decisões acerca dos investimentos financeiros, do direcionamento dos esforços para o trabalho técnico e científico e do investimento em infraestrutura.

Sobre a questão da geração, das condições de acesso e de uso dos dados gerados pelas colaborações, verificou-se que os dados estão essencialmente acessíveis aos membros com experiência e conhecimento dos processos de geração e análise desses dados, identificados como membros 'qualificados' – aqueles cujos nomes figuram na lista de autores de artigos publicados pela colaboração. Evidenciou-se, também, que o acesso a grandes volumes de dados (dados brutos ou *'crus'*) depende de justificativa e requisição, sendo esta uma condição essencial para preservação do sistema. Destacou-se, também, pela fala de um dos entrevistados, que pesquisadores "de fora" das colaborações CERN não dispõem de expertise ou ferramentas/estrutura para a análise dos dados, e que, portanto, os dados só são úteis no contexto das colaborações/do CERN.

Observou-se que, para algumas áreas do conhecimento, as condições de acesso, regras de uso e segurança são bem estabelecidas e dimensionadas e que, no caso do Brasil, sua atuação na proposição de soluções para gestão de dados de pesquisa lhe traria vantagens estratégicas, frente a outros países, em termos de avanços científicos.

Sobre o processo de partilha e avaliação dos resultados das colaborações, observou-se que o mesmo ocorre, primordialmente, por meio de crédito aos autores pagantes, que têm seus nomes divulgados na lista de autores nas publicações das colaborações. Cada colaboração requer o pagamento, por cada pesquisador, de uma taxa/contribuição anual chamada de Manutenção e Operação (M&O), derivada do rateio dos custos operacionais dos experimentos, como água, eletricidade etc. O rateio é proporcional ao número de pessoas de cada instituição atuantes na colaboração, sendo que pesquisadores-autores, cujos nomes aparecem nas publicações em nome da colaboração, correspondem a um valor cheio (100%), pós-doutorandos 75% e estudantes não pagam.

Na Tabela a seguir, é possível visualizar um resumo do número de pesquisadores brasileiros que participam de colaborações no CERN. Tais dados alimentam um relatório destinado a um consórcio brasileiro para permitir o financiamento das participações na colaboração com o CERN, denominado Rede Nacional de Física de Altas Energias (Renafae), o qual foi instituído pela Portaria n. 321 de 28 de maio de 2008 do Ministério da Ciência e Tecnologia e Inovação⁷, para permitir, além de outros objetivos, o financiamento dos custos de Manutenção e Operação (M&O).

Tabela 1: Autores brasileiros pagantes da taxa de Manutenção e Operação no CERN

<i>Experimento</i>	<i>2010</i>	<i>2011</i>	<i>2012</i>	<i>2013</i>	<i>2014</i>
ALPHA	2	2	2	2	2
ALICE	6	7	5	5	8
ATLAS	10	10	10	11	11
CMS	17	17	17	22	31
LHCb	12	12	18	18	20
Total	47	48	52	58	72

Fonte: Ministério da Ciência e Tecnologia (Brasil). 2013. “Rede Nacional de Física de Altas Energias.” ATA do Workshop Anual da RENAFEA e Reunião do CTC. Acesso em 1 set. 2013. http://www.cbpf.br/~renafae/atas/ATA_wshop-2013.pdf.

Dentre os procedimentos e critérios para publicação de resultados de pesquisa realizadas no CERN, destaca-se que toda publicação divulgada em nome de uma determinada colaboração (com sua respectiva lista de autores) deve passar antes pelo conhecimento de todos os membros e pela revisão de comitês ou comissões dessa colaboração. Todas as regras e diretrizes de trabalho das colaborações, incluindo os procedimentos para divulgação e avaliação dos resultados de pesquisa, são extensivamente documentadas e acessíveis a todos os membros, por meio de plataformas Wiki, com acesso restrito a pesquisadores atuantes no CERN.

Quanto ao aspecto da propriedade intelectual e licenciamento, não se observou consenso ou a existência de uma diretriz comum a partir dos relatos dos pesquisadores. Percebeu-se maior preocupação com os resultados científicos em relação aos tecnológicos, além de baixa preocupação com a exploração comercial imediata das descobertas, ou com a sua restrição. De acordo com os entrevistados, empresas fornecedoras de tecnologias, materiais e soluções para o CERN encontram a oportunidade de testarem essas tecnologias sob as condições extremas de desenvolvimento derivadas dos experimentos. Destacou-se também que soluções em regime aberto (*open source software, open hardware* etc.) suprem carências tecnológicas por meio do desenvolvimento compartilhado.

⁷ Publicada no Diário Oficial da União, n. 101, de 29 de maio de 2008.

Considerações finais

Um dos resultados mais significativos alcançados no decorrer deste trabalho foi a elucidação das principais dimensões presentes no contexto das colaborações em *e-Science*, a saber: a construção de uma infraestrutura computacional para uso distribuído ou para processamento de larga escala; a produção e o uso intensivo de dados; e a colaboração entre atores da ciência, pelo compartilhamento de esforços, dados e/ou recursos computacionais. A compreensão de tais dimensões enriqueceu a proposta de pesquisa, possibilitando a construção de categorias de análise que orientaram a pesquisa empírica.

Percebeu-se, ao longo da pesquisa, que as tecnologias de informação e comunicação (TICs) expandiram os limites das colaborações. Contudo, elas também ampliaram o leque de possíveis barreiras ou restrições para o acesso aos dados ou demais recursos mobilizados nas diferentes colaborações. As TICs também impactaram a ampliação dos níveis de avaliação das colaborações. Enquanto na fase anterior à emergência da *e-Science* as colaborações ou grupos de pesquisa eram avaliados unicamente com base nas publicações, o advento da *e-Science* tem motivado a proposição de novos indicadores, sendo que muitos deles ainda estão sendo estudados e aperfeiçoados e alguns grupos têm mais ou menos sucesso na sua implementação.

Com relação à infraestrutura necessária à viabilização da *e-Science*, observou-se que as colaborações se desenvolvem em torno de problemas complexos, e essa complexidade é diretamente relacionada à quantidade de dados que esses problemas científicos geram ou demandam. Existe uma produção de dados em larga escala, associada à natureza científica do problema abordado, e esse processo traz novos desafios às colaborações, como a questão da armazenagem – possibilidade de estocagem e uso eficaz dos dados em momentos posteriores –, a questão do processamento e da análise, sendo que estes últimos processos devem ocorrer de forma distribuída ou a partir de diferentes localidades.

Sobre a interoperabilidade dos dados e dos processos, verificou-se que esta ocorre com maior frequência no plano interno das colaborações, o que as favorece no sentido de colocá-las em uma posição privilegiada na busca por financiamento, ou seja, não faz sentido ou torna-se inviável a mobilização ou a replicação de recursos e esforços fora das colaborações.

Verificou-se também que os dados, assim como os artigos e demais tipos de publicações científicas, tornam-se ativos importantes da pesquisa, reconfigurando a busca por resultados e os esforços das colaborações em *e-Science*.

Evidenciou-se que a documentação e a definição de diretrizes que regulem o acesso e as formas de uso dos dados devem ser extensivas e conhecidas por parte de todos os integrantes das colaborações. Por outro lado, comunidades ou pesquisadores que não estejam formalmente vinculados às colaborações podem encontrar dificuldades no acesso aos dados ou à documentação, seja por limites técnicos das colaborações, seja pela dificuldade de compreensão e uso desses recursos. Diante disso, evidenciou-se ainda que a percepção de restrições não se resume a questões comerciais, de registro de propriedade intelectual ou cobrança de *royalties*, mas diz respeito também às dificuldades de acesso inerentes aos processos de uso e análise dos dados gerados pelas colaborações.

Outro desafio que se mostrou presente durante o estudo das mudanças nas colaborações foi a questão da escala ou dimensionamento. Percebeu-se uma mudança de escala, tanto na quantidade de pessoas envolvidas nas colaborações, na distância a ser percorrida/transposta para o intercâmbio de recursos, assim como na escala econômica ou de financiamento. Evidenciou-se que cada projeto ou iniciativa voltada para a colaboração em *e-Science* deve ser planejada com uma provisão de recursos para uma perspectiva temporal muito mais longa. As fases de início, meio e fim de cada projeto tornam-se difusas e com isso as fontes de recursos devem ser contínuas, já que não mais é possível indicar precisamente o período de duração de um projeto ou quando os primeiros resultados começarão a surgir.

Com o advento da *e-Science* e das colaborações em larga escala, a governança passa por uma mudança significativa nas práticas científicas. Sua importância se faz presente no processo de organização dos atores da colaboração, na forma como definem seus papéis e na forma como esses ato-

res compartilham seus recursos e os resultados alcançados pelas colaborações. Nesse sentido, evidencia-se a necessidade de aprofundar o estudo de novas formas de reconhecimento pela atuação individual nas práticas colaborativas, ou seja, de como melhor identificar e avaliar os diferentes níveis ou os diferentes modos de atuação e contribuição, para que possam ser considerados na partilha de resultados ou na correção de distorções.

Tais preocupações não são significativas em projetos colaborativos locais ou com menor número de integrantes, em meio aos quais a existência de um processo de tomada de decisão centralizado pode até ser um fator favorável. Porém, com a transição para a colaboração em redes, a centralização das decisões torna-se inviável e seu alcance cada vez mais limitado. Neste sentido, a maneira como se constrói a governança pode auxiliar na definição conjunta dos objetivos a serem alcançados, na definição dos propósitos das colaborações, assim como na explicitação de como se dão as relações de poder entre os diferentes atores partícipes das colaborações e como alinhá-los para permitir que a colaboração se auto governe e se auto avalie.

Dentre as mudanças que afetam as práticas dos cientistas brasileiros, decorrentes da sua inserção nas colaborações em *e-Science*, pode-se destacar a ampliação do leque de capacitação para a pesquisa, com o desenvolvimento de habilidades e competências que só podem ser alcançadas pela vivência e prática nas colaborações.

Verificou-se também que há uma demanda dos pesquisadores para a ampliação da partilha de responsabilidades quanto às tarefas e decisões burocráticas das iniciativas de colaboração, como a provisão de recursos e a necessidade de implementação de escritórios para a elaboração de projetos, por exemplo – o que também pode demandar novas práticas e ações no campo da governança.

Em linhas gerais, foi possível observar a convergência dos questionamentos inicialmente levantados e dos temas abordados, apresentando-se como principal desafio a carência de estudos e publicações brasileiras sobre o assunto. Como perspectivas futuras de estudo, destaca-se ainda a potencialidade de desenvolvimento de um arcabouço para análise e estudo da governança em *e-Science*, com foco na abertura dos processos de geração e tratamento de dados derivados de pesquisas colaborativas, ou de iniciativas por parte da comunidade científica brasileira alinhadas a esses processos, garantindo-lhe um posicionamento estratégico no campo das colaborações em *e-Science*.

Agradecimentos

A pesquisa contou com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brasil – Bolsa Demanda Social (DS).

REFERÊNCIAS

- Albagli, S., Appel, A. L., & Maciel, M. L. (2014). E-Science, ciência aberta e o regime de informação em ciência e tecnologia. *Tendências da Pesquisa Brasileira da Pesquisa Brasileira em Ciência da Informação*, 7(1). Recuperado de <http://inseer.ibict.br/ancib/index.php/tpbci/article/view/124/166> em 26 out. 2014.
- Barjak, F., Eccles, K., Meyer, E. C., Robinson, S., & Schroeder, R. (2013). The emerging governance of e-Infrastructure. *Journal of Computer-Mediated Communication*, 18(2), p. 1-24. doi:10.1111/jcc4.12000
- Bell, G., Hey, T., & Szalay, A. (2009). Beyond the data deluge. *Science*, 323(5919), pp. 1297-1298. doi:10.1126/science.1170411
- Christensson, P. (2010). LOG File. *The PC glossary: definitions of computer and internet terms*. Recuperado de <http://pc.net/glossary/definition/logfile> em 7 jun. 2013.
- Conseil Européen pour la Recherche Nucléaire (CERN). (2014a). *About CERN: What is the universe made of? How did it start? Physicists at CERN are seeking answers, using some of the world's most powerful particle accelerators*. Recuperado de <http://home.web.cern.ch/about> em 2 fev. 2014.
- Conseil Européen pour la Recherche Nucléaire (CERN). (2014b). *How a detector works: Just as hunters can identify animals from tracks in mud or snow, physicists identify subatomic particles from the traces they leave in detectors*. Recuperado de <http://home.web.cern.ch/about/how-detector-works> em 2 fev. 2014.
- David, P. A., Besten, M., & Schroeder, R. (2006 Dec. 4). How open is e-Science?. *IEEE 8th International Conference on e-Science and Grid Computing*, Amsterdam. doi:10.1109/E-SCIENCE.2006.85
- Dumbill, E. (2012). Getting up to speed with big data. O'Reilly Media (Ed.). *Big data now: 2012 edition* (p. 3-18). Sebastopol USA: O'Reilly Media. Recuperado de <http://www.oreilly.com/data/free/big-data-now-2012.csp> em 9 nov. 2014.
- Dutton, W. H. (2008). Collaborative network organizations: new technical, managerial and social infrastructures to capture the value of distributed intelligence. *Oxford Internet Institute DPSN Working Paper Series*, 5. doi:10.2139/ssrn.1302893
- Foster, I. & Kesselman, C. (1999). Computational grids. In Foster, I. & Kesselman, C. (Eds.). *The Grid: blueprint for a new computing infrastructure* (p. 15-51). Los Altos, USA: Morgan Kaufmann. Recuperado de http://www.elsevierdirect.com/companions/9781558609334/appendices/Chapter_02.pdf em 28 maio 2013.
- Garvey, W. D. & Griffith, B. C. (1967). Scientific communication as a social system. *Science*, 157(3792), pp. 1011-1016. doi:10.1126/science.157.3792.1011
- Gray, J. (2009). eScience: a transformed scientific method. In Hey, T., Tansley, S., & Tolle, K. *The fourth paradigm: data-intensive scientific discovery*. Redmond, USA: Microsoft Research.
- Hey, T., & Trefethen, A. E. (2005). Cyberinfrastructure for e-Science. *Science*, 308(5723), p. 817-821. doi:10.1126/science.1110410
- Hey, T., & Trefethen, A. E. (2002). The UK e-Science core programme and the Grid. *Future Generation Computer Systems*, 18(8), p. 1017-1031. doi:10.1016/S0167-739X(02)00082-1
- Medeiros, J. da S., & Caregnato, S. E. (2012). Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. *Liinc em Revista*, 8(2), pp. 311-322. doi:10.18225/liinc.v8i2.488
- Meyer, E. T., & Schroeder, R. (2009). The world wide web of research and access to knowledge. *Journal of Knowledge Management Research and Practice*, 7, pp. 218-233. doi:10.1057/kmnp.2009.13
- Ohno-Machado, L. (2012). To share or not to share: that is not the question. *Science Translational Medicine*, 4(165), pp. 165cm15. doi:10.1126/scitranslmed.3004454

- Quin, L. R. E. (2010). XML essentials. *World Wide Wibe Consortium*. Recuperado de <http://www.w3.org/standards/xml/core> em 2 jul. 2013.
- Sánchez-Artigas, M. & García-López, P. (2010). eSciGrid: A P2P-based e-science Grid for scalable and efficient data sharing. *Future Generation Computer Systems*, 26(5), pp. 704-719. doi:10.1016/j.future.2009.05.013
- Sayão, L. F., & Sales, L. F. (2012). Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. *Informação & Sociedade: Estudos*, 22(3), pp. 179-191. Recuperado de <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/12224/8586> em 28 jan. 2013.
- Schroeder, R. (2008). e-Sciences as research technologies: reconfiguring disciplines, globalizing knowledge. *Social Science Information*, 47(2), pp. 131-157. doi:10.1177/0539018408089075
- Sonnenwald, D. H. (2007). Scientific collaboration. In Cronin, B. (Ed.). *Annual review of information science and technology* (pp. 643-681). Medford: Information Today. doi:10.1002/aris.2007.1440410121

SOBRE OS AUTORES

Andre Luiz Appel: Bacharel em Gestão da Informação pela Universidade Federal do Paraná (2010) e Mestre em Ciência da Informação pelo Instituto Brasileiro de Informação em Ciência e Tecnologia em convênio com a Universidade Federal do Rio de Janeiro (2014). Atualmente é professor substituto na Universidade Federal do Rio de Janeiro (UFRJ), presta apoio técnico-especializado junto ao Portal de Publicações Eletrônicas da Universidade do Estado do Rio de Janeiro (Uerj) e atua como colaborador em grupo de pesquisa da Universidade Federal do Paraná (UFPR). Tem experiência na área de Ciência da Informação, atuando principalmente nos seguintes temas: gestão da informação, sistema eletrônico de revistas, metodologias em gestão da informação, produção colaborativa em CT&I.

Maria Lucia Maciel: cursou graduação em Ciências Sociais/Sociologia na Universidade de Brasília, Mestrado em Sociologia na Université Libre de Bruxelles (1981) e Doutorado em Sociologia na Université de Paris VII (1986). Coordenadora do Laboratório Interdisciplinar sobre Informação e Conhecimento - LIINC; diretora do Instituto Ciência Hoje. Tem experiência na área de Sociologia, com ênfase em Sociologia do Desenvolvimento e Sociologia do Conhecimento, atuando principalmente nos seguintes temas: ciência e tecnologia para o desenvolvimento; era da informação e do conhecimento; conhecimento e inovação, produção e circulação de ciência.

Sarita Albagli: Pesquisadora do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Professora do Programa de Pós-Graduação em Ciência da Informação - PPGCI/IBICT-UFRJ. Pesquisadora 1C do CNPq. Coordenadora do Laboratório Interdisciplinar sobre Informação e Conhecimento (Liinc) e editora da Liinc em Revista. Líder do Grupo de Pesquisa Informação, Conhecimento e Mudança Sociotécnica. Graduação em Ciências Sociais (UFRJ), Mestrado em Engenharia de Produção - Política e Gestão de Ciência e Tecnologia (UFRJ) e Doutorado em Ciências - Geografia (UFRJ). Estágio Pós-Doutoral Senior, Visiting Senior Fellow, London School of Economics and Political Science (LSE). Foi Coordenadora do PPGCI/IBICT-UFRJ (2011-2013). Coordenou o GT5 Política e Economia da Informação da Ancib (2007-2008; 2009-2010). Coordenou o GT22 Sociedade da Informação e do Conhecimento do Congresso Brasileiro de Sociologia (2007; 2009). Atua na área de Ciência da Informação, com ênfase nos seguintes temas: informação, conhecimento e inovação social; informação, conhecimento e poder; políticas de informação, ciência, tecnologia e inovação; informação, conhecimento e dinâmicas socioambientais; produção colaborativa em ciência, tecnologia e inovação.

La **Revista Internacional de Ciencia y Sociedad**

ofrece un espacio para el diálogo interdisciplinar, sobre el pasado, el presente y el futuro de la ciencia y sus relaciones, con la sociedad.

Los temas de interés para la revista abarcan las ciencias particulares, es decir, la biología, la química, la física, la astronomía, las matemáticas, etc., en su relación con la historia, la filosofía, la sociología, la economía, la política científica, la educación, la tecnología, a innovación, el periodismo, la religión o el sector militar.

La **Revista Internacional de Ciencia y Sociedad**

es una revista académica sujeta a revisión por pares, y acepta textos en español y portugués.

ISSN: 2340-9991

