

## ***E-SCIENCE*, CIÊNCIA ABERTA E O REGIME DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA**

**Sarita ALBAGLI**

Doutora em Geografia

Universidade Federal do Rio de Janeiro (UFRJ)

Pesquisadora do Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT

Professora do Programa de Pós-Graduação em Ciência da Informação – PPGCI/IBICT e

Universidade Federal do Rio de Janeiro –ECO/UFRJ

sarita@ibict.br

**Andre Luiz APPEL**

Mestre em Ciência da Informação

Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT/UFRJ

Professor Substituto na Universidade Federal do Rio de Janeiro (UFRJ)

alappel@gmail.com

**Maria Lucia MACIEL**

Doutora em Sociologia

Université de Paris VII

Professora do Programa de Pós-Graduação em Ciência da Informação – PPGCI/IBICT e

Universidade Federal do Rio de Janeiro –ECO/UFRJ

lucamaciel@gmail.com

### **Resumo**

O trabalho debruça-se sobre as práticas de *e-Science* e as questões que influenciam sua promoção como ciência “aberta”. Para além dos aspectos tecnológicos que caracterizam essa nova forma de fazer ciência, evidencia-se que as questões de ordem institucional (formais e informais) são as que mais interferem no caráter aberto ou proprietário dessas práticas. Situa-se essa questão no contexto de configuração e disputas no âmbito do atual regime de informação em C&T. O trabalho fundamenta-se em levantamento e discussão da literatura sobre o tema, que contempla tanto o debate teórico-conceitual, como relatos de estudos empíricos sobre essas experiências.

**Palavras-chave:** *E-Science*. Ciência Aberta. Regime de Informação. Produção Colaborativa. Propriedade Intelectual.

## **E-SCIENCE, OPEN SCIENCE, AND THE SCIENCE AND TECHNOLOGY INFORMATION REGIME**

### **Abstract**

This paper focuses on E-science and its practices, and on the issues that influence its promotion as “open” science. Beyond the technological aspects that characterize this new way of doing science, it becomes clear that (formal and informal) institutional issues are the more important interferences in the open or proprietary character of these practices. The issue is situated in a context of the configuration and disputes within the present information regime in Science and Technology (S&T). The analysis is based on a discussion of the pertinent literature, examining the theoretical-conceptual discussion as well as a selection of previous empirical studies in the field.

**Keywords:** E-Science. Open Science. Information Regime. Collaborative Production. Intellectual Property.

## 1 INTRODUÇÃO

Este trabalho apresenta parte dos resultados de um projeto de pesquisa que trata dos atuais processos de circulação e apropriação da informação e do conhecimento em ciência e tecnologia (C&T), ante as transformações tecnológicas e institucionais em curso. O foco da pesquisa recai na indagação e investigação sobre como as transformações nas condições contemporâneas de produção e circulação de informação e conhecimento que repercutem nas formas de apropriação social (informação como bem público ou comum) ou privada (informação como mercadoria). Parte-se do suposto de que essas diferentes formas de apropriação são constitutivas dos antagonismos que caracterizam o atual regime de informação em C&T.

Observa-se que, no âmbito das atuais mudanças tecnológicas e político-institucionais e das inovações sociais correlatas, conforma-se uma nova geração de práticas colaborativas em ciência, expressas em noções tais como: redes de conhecimento, co-inovação, co-criação, produção *peer-to-peer*, *crowdsourcing*, inovação aberta, inovação social, *open Science*, entre outras, cada qual com seu significado específico.

Em todas essas, ganham expressão novos tipos de colaboração que se valem do desenvolvimento e difusão das tecnologias de informação e comunicação (TIC), particularmente das plataformas digitais. O crescimento do número de jovens pesquisadores nativos da *web* contribui para ampliar e abrir novas possibilidades de usos das TIC na prática científica. A presença *on-line* alarga a visibilidade das pesquisas, possibilitando a aproximação de iniciativas correlatas.

Esses novos tipos de colaboração mobilizam muito mais do que recursos como videoconferências, *blogs*, *wikis*, redes sociais, sem falar no já tradicional *e-mail*, que são hoje meios correntes para o intercâmbio e a comunicação científica em moldes informais. A alocação de recursos digitais geograficamente distribuídos, o poder de computação (capacidade de processamento e armazenamento) e as tecnologias de visualização de dados têm constituído os principais esforços para a construção de infraestruturas computacionais de suporte à pesquisa colaborativa, no que se tem denominado de *e-Science* (BARJAK et al., 2013).

Considerando esse quadro mais amplo, este trabalho debruça-se especificamente sobre as práticas de *e-Science* e os fatores que influenciam seu papel na promoção de uma ciência “aberta”. Para além dos aspectos tecnológicos que caracterizam essa forma de fazer ciência, se evidenciará que são as questões de ordem institucional (formais e informais) que mais interferem no caráter aberto ou proprietário dessas práticas.

O trabalho fundamenta-se em levantamento, sistematização e discussão da literatura sobre o tema, que contempla, nesse campo, tanto o debate teórico-conceitual quanto relatos de estudos empíricos já realizados sobre experiências nesse campo.

Apresentam-se a seguir conceitos-chave e o contexto de emergência da *e-Science* para então discutir as questões que envolvem a relação da *e-Science* com a noção de ciência aberta, situando essa questão no âmbito da configuração do regime de informação em C&T.

## **2 A EMERGÊNCIA DA E-SCIENCE NAS PRÁTICAS DE PESQUISA: CONCEITOS CHAVE**

A partir das últimas décadas do século XX, a ciência passou a valer-se, cada vez mais, de uma vertente computacional, com forte apelo à simulação de fenômenos complexos. Daí derivou-se a exploração de grandes volumes e uso intensivo de dados (*data-intensive*), em busca da interlocução de teorias, experimentos e simulações, dando lugar à *e-Science* (GRAY, 2007; BELL; HEY; SZALAY, 2009). O termo *e-Science*<sup>1</sup> foi introduzido por John Taylor no ano de 2001, quando diretor geral do Conselho de Pesquisa do *Office of Science and Technology* (OST) do Reino Unido, para designar “a ciência crescentemente desenvolvida por meio de colaborações globalmente distribuídas viabilizadas pela Internet, usando grandes coleções de dados, recursos computacionais em escala tera<sup>2</sup> e visualização de alto desempenho” (TAYLOR, 2002 apud FRY; SHROEDER; BESTEN, 2009). A *e-Science* chegou a ser associada à gênese de um quarto paradigma da ciência (GRAY, 2007).

A *e-Science* supõe a construção de uma infraestrutura informática computacional de uso distribuído, capaz de permitir a colaboração à distância de equipes de pesquisa, envolvendo o uso intensivo e o compartilhamento de dados e recursos computacionais. Essa infraestrutura tecnológica tem sido baseada em *Grids*, *Middleware*, *Workbenches*,

---

<sup>1</sup>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; CAREGNATO, 2012, p. 315). Para esta pesquisa, optou-se pelo uso do termo *e-Science*.

<sup>2</sup>1 *Terabyte* é o equivalente a 1024 *Gigabytes*.

*Webservices*, *Virtual Research Environments* (VRE), tecnologias de notação e armazenamento de dados concebidas em padrões como XML<sup>3</sup>, entre outras.

O termo *Grid* (computação em grade) foi utilizado pela primeira vez no final dos anos 1990 (FOSTER; KESSELMAN, 1998), para denotar uma infraestrutura computacional distribuída servindo a atividades de ciência e engenharia avançadas. Os *grids* compreendem 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; KESSELMAN, 1998).

O *Middleware* corresponde ao *software* entre a infraestrutura computacional dos *grids*, distribuída em rede, e as aplicações de uso individual por cientistas, projetos ou instituições, e permite o compartilhamento de recursos entre usuários (SCHROEDER, 2008). Um conjunto de *software* com módulos, funções e aplicações pré-definidas é comumente chamado de *Workbench*, grosso modo a chamada bancada de trabalho.

*Webservice* é um sistema de *software* projetado para suportar interações interoperáveis máquina-a-máquina em uma rede<sup>4</sup>. Possui uma interface 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*.

Essa infraestrutura tecnológica tem a capacidade de exploração de grandes quantidades de dados. Neste aspecto, o termo *e-Science* é mais utilizado na atividade científica do que o termo *big data*. *Big data* compreende a geração, tratamento e análise de grandes volumes de dados que excedem as capacidades convencionais de processamento, e vem sendo explorado também por empresas, governos e outros segmentos interessados em extrair informações a partir de grandes quantidades de dados não estruturados. Já *e-Science* incorpora, além do uso intensivo de dados, a pesquisa científica colaborativa e o uso de recursos compartilhados para sua exploração.

Na *e-Science*, ao menos três ações são tidas como chave quanto ao uso de dados: captura, curadoria e análise (GRAY, 2007). Na captura, os dados são gerados diretamente por instrumentos de captura, tais como sensores de temperatura, de movimento, registros

---

<sup>3</sup>*eXtensibleMarkupLanguage* “é um formato simples, 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, c2010).

<sup>4</sup>World Wide Web Consortium (W3C). **Web Services Glossary**. 11 fev. 2004. Disponível em: <<http://www.w3.org/TR/ws-gloss/#webservice>>. Acesso em: 3 jun. 2013.

de atividades em *sites* da *internet*, entre outros, 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 suas informações para que possam ser intercambiadas com outros cientistas. Para isso, as informações precisam estar representadas de forma algorítmica, por meio de representações padrão de objetos em análise – tais como genes, galáxia, temperatura etc. (GRAY, 2007). 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; SALES, 2012, p. 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.

Nessa direção, Dutton (2008) estudou oito casos de redes de colaboração, visando obter uma visão geral sobre a forma como cada uma dessas iniciativas distribui os esforços na resolução de problemas, buscando identificar, ao mesmo tempo, possíveis indicadores de desempenho para avaliar os resultados de tais iniciativas. A partir dos casos, o autor desenvolveu uma classificação das principais ações que configuram a formação de redes colaborativas em ciência que se valem intensivamente da *Internet*, destacando-se (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 atuação em rede (*social networking*) na *Internet* para facilitar a comunicação entre equipes; 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.

Schroeder (2008) realizou o estudo de três casos<sup>5</sup> que, embora o autor argumente não ser uma amostra representativa de *e-Science*, apresentavam características comuns reconhecidamente da *e-Science*, entre elas: a) a presença de ferramentas para coletar,

---

<sup>5</sup>Virtual Observatory for the Study of Online Networks (VOSON), Enabling Grids for e-Science (EGEE), International Virtual Observatory Alliance (IVOA).

armazenar e manipular grandes quantidades de dados e demais recursos de pesquisa em meio digital, e torná-los acessíveis como recursos de pesquisa compartilhados e distribuídos;

b) disponibilidade de acesso a tecnologias de *grid*, *middleware* e a uma variedade de aplicações.

A partir da realização de entrevistas junto a pesquisadores sobre as condições de adoção de infraestruturas de pesquisa de *e-Science*, Voss et al. (2010) identificaram alguns fatores que influenciam a participação em práticas colaborativas com base em tais ferramentas, dentre eles:

- a) a familiaridade e a disponibilidade para capacitação no uso dessas tecnologias, bem como o entendimento e o diálogo entre usuários e desenvolvedores;
- b) a facilidade ou dificuldade de estabelecimento de diálogos interdisciplinares, considerando a diversidade de práticas, metodologias e linguagens científicas das diferentes áreas do conhecimento envolvidas na colaboração;
- c) as vantagens ou desvantagens quanto ao compartilhamento de dados;
- d) a adequação de políticas e de mecanismos de financiamento em ciência para o apoio a formas de produção colaborativa;
- e) as condições técnicas locais, incluindo a disponibilidade de infraestrutura tecnológica e o suporte institucional para aquisição de equipamentos de alta computação para a pesquisa;
- f) os dispositivos legais e contratuais para o compartilhamento de dados.

A adoção de plataformas computacionais na colaboração científica não se reduz então ao componente tecnológico. Mais importantes são os novos usos que implicam transformações nos métodos e estruturas lógicas da pesquisa e, logo, em seus resultados, em um processo de aprendizado e inovações contínuos.

### **3 CIÊNCIA ABERTA E *E-SCIENCE* NO CONTEXTO DO REGIME DE INFORMAÇÃO EM C&T**

A relação entre ciência aberta e *e-Science* é aqui compreendida no contexto de configuração do atual regime de informação em ciência e tecnologia.

A concepção de regime de informação assume relevância, no campo da Ciência da Informação no Brasil, a partir dos trabalhos de Bernd Frohman (1995), Maria Nélida González de Gómez (2002) e Sandra Braman (2004).

Frohman (1995, p.5), baseando-se nos trabalhos de Bruno Latour, Michel Callon e Michel Foucault, definiu regime de informação como “qualquer *sistema* ou *rede* mais ou menos estável, no qual a informação flui por meio de determinados canais – de produtores específicos, via estruturas organizacionais específicas, a consumidores ou usuários específicos”<sup>6</sup>. Para o autor, “descrever um regime de informação significa mapear os processos agonísticos que resultam em tentativas de estabilização e conflitos entre grupos sociais, interesses, discursos e até artefatos científicos e tecnológicos”<sup>7</sup> (FROHMAN, 1995, p.5).

Em uma perspectiva análoga, González de Gómez (2002, p. 28), referenciando-se, além dos autores citados, no “modo de informação” de Mark Poster<sup>8</sup>, entende por regime de informação “o modo de produção informacional dominante em uma formação social”, aí incluídos os sujeitos, as regras, os meios e recursos de informação. A autora considera que um regime de informação se define mais pelas *ações de informação* dos atores sociais (WERSIG; WINDEL, 1985) do que por uma intencionalidade prévia.

Sandra Braman (2004), por sua vez, fundamentando-se na teoria do regime no campo das relações internacionais, refere-se à emergência de um regime global de política de informação (*global information policy regime*). Sua ênfase recai, sobretudo no arcabouço legal que cristaliza e explicita formalmente as políticas de informação.

Na concepção de regime de informação adotada neste trabalho, enfatizam-se dois aspectos centrais às abordagens de Frohman e González de Gómez: o reconhecimento do papel das práticas informacionais (ações de informação) para além da dimensão institucional formal; e o reconhecimento do regime de informação como sendo um campo de disputa e conflito, do mesmo modo que de negociação e estabilização.

Desta perspectiva, considera-se, neste trabalho, que o atual regime de informação em C&T, em âmbitos global e nacional brasileiro, caracteriza-se por um duplo movimento. De um lado, coloca-se o imperativo de expansão contínua da produção científica e

---

<sup>6</sup> “[...] any more-or-less stable *system* or *network* in which information flows through determinable channels— from specific producers, via specific organizational structures, to specific consumers or users [...] Radio and television broadcasting, film distribution, academic publishing, libraries, transborder data flows, the emerging infobahn: these are all nodes of information networks, or elements of specific régimes of information.” Tradução nossa.

<sup>7</sup> “[...] describing a *régime* of information means charting the agonistic processes that result in tentative and uneasy stabilizations of conflicts between social groups, interests, discourses, and even scientific and technological artifacts.”

<sup>8</sup> Poster, Mark. “The mode of information and the cultures of the internet- a conversation with Mark Poster”. **Consumption, Markets and Culture**, v. 3, n. 3, p.195-213, 2000.

tecnológica, como requisito para o desenvolvimento e a competitividade, e como condição para formulação de respostas adequadas à complexidade e urgência dos atuais desafios globais – ambientais, econômicos, sociais. Nos processos de inovação e criatividade, demonstram-se cruciais as dinâmicas colaborativas, em muito facilitadas pelas tecnologias da informação e comunicação (TIC) e as novas plataformas digitais, que propiciam a produção conjunta, o compartilhamento e a circulação ampliada de informação e conhecimento, por meio de redes sociotécnicas.

De outro lado, assiste-se à concentração – espacial, social e organizacional – e à privatização da informação e do conhecimento, sobretudo daqueles considerados estratégicos e/ou lucrativos. Nessa direção atuam o recrudescimento dos instrumentos de proteção de direitos da propriedade intelectual<sup>9</sup> desde a década de 1980, bem como a crescente mercantilização e tentativa de controle dos espaços de interação digital (a *internet* principalmente) e dos bens imateriais e intangíveis que por eles circulam – informação, conhecimento, cultura. Essas formas de regulação mostram-se, por sua vez, inadequadas e ineficazes ante a crescente dificuldade de se introduzirem barreiras legais (artificiais) à ampla difusão e ao compartilhamento que os novos meios técnicos proporcionam.

É nesse contexto, a partir das últimas décadas do século XX, que se observa o acirramento do antagonismo entre uma concepção de ciência aberta<sup>10</sup> (*open Science*), que ganha proeminência a partir dos trabalhos de Robert Merton<sup>11</sup>, e o paradigma que iria privilegiar uma ciência proprietária. É a partir daí, também, que aumentam as pressões sobre os ambientes acadêmicos e universitários para o patenteamento e o retorno financeiro dos resultados das atividades de C&T, buscando-se estabelecer aparatos institucionais e legislação pertinente a esses objetivos.

---

<sup>9</sup>Segundo a Organização Mundial da Propriedade Intelectual (OMPI), propriedade intelectual é hoje definida de modo bastante abrangente, incluindo direitos relativos às obras literárias, artísticas e científicas, às interpretações dos artistas intérpretes e às execuções dos artistas executantes, aos fonogramas e às emissões de radiodifusão, às invenções em todos os domínios da atividade humana, às descobertas científicas, aos desenhos e modelos industriais, às marcas industriais, comerciais e de serviço, bem como às firmas comerciais e denominações comerciais, à proteção contra a concorrência desleal e todos os outros direitos inerentes à atividade intelectual nos domínios industrial, científico, literário e artístico.

<sup>10</sup>Ciência aberta é, por sua vez, um termo guarda-chuva, que envolve múltiplos níveis e escopos de abertura, remetendo tanto a um sentido pragmático, no sentido de permitir maior dinamismo às atividades de CT&I, quanto um sentido democrático, no sentido de permitir maior abertura e participação da sociedade.

<sup>11</sup>Merton (1973), retomando seu ensaio *The Normative Structure of Science* (1942), define como normas chave da atividade científica o comunalismo, o universalismo, o desinteresse e o ceticismo organizado, às quais foi posteriormente agregada a originalidade.



Consistentemente com o que já se assinalara em relação ao regime global de informação em C&T, observa-se aqui a existência de uma disputa entre duas perspectivas distintas, que expressam condições diferenciadas de acesso a informações em C&T, desde dados brutos até artigos científicos; e que irão representar, em maior ou menor medida, dois modos de desenvolvimento da *e-Science*. Por um lado, têm-se os esforços pelo amplo compartilhamento dos resultados da pesquisa científica, convergindo para diferentes iniciativas e movimentos de acesso aberto (como os de acesso livre à informação e publicação científica e os de promoção do *software* de código aberto). Por outro, colocam-se interesses favoráveis a pesquisas de cunho proprietário e que são movidas principalmente pela lucratividade.

Põe-se, portanto, em questão: em que medida a *e-Science* implica o acesso aberto aos recursos científicos? Embora exista um grande número de ferramentas da *e-Science* para suporte a projetos de natureza aberta e distribuídos, alavancando novas formas de colaboração científica e de compartilhamento de informações e materiais de pesquisa, isto não significa que toda a pesquisa colaborativa identificada como *e-Science* será qualificada como *open Science*.

O significado de uma *e-Science* aberta envolve tanto a dimensão dos dados, quanto a dimensão da infraestrutura, aí incluído o *software*. Nessas dimensões atuam fatores institucionais, normativos, culturais e técnicos.

Esses diferentes aspectos abrangem instâncias de ação e decisão diferenciadas, que vão desde o pesquisador individual até o nível macro das políticas públicas e das regulações internacionais, passando pelo nível meso das instituições de pesquisa e agências de fomento. Trata-se, portanto, de distintos âmbitos de governança, que mantêm suas especificidades, ao mesmo tempo que se influenciam mutuamente. Esses âmbitos envolvem tanto arranjos institucionais formais, quanto informais. Frequentemente a governança das instâncias meso (institucional) e micro (no âmbito das práticas individuais dos pesquisadores e das equipes de pesquisa) estão desconectadas. A existência de diferentes instâncias e mecanismos de regulação gera complexidade e eventualmente barreiras para a desejada abertura no âmbito da *e-Science*. Seu sucesso dependerá em grande medida da concordância sobre normas e procedimentos de participação e de autoria, que passam pelo estabelecimento de mecanismos de governança previamente acordados entre as partes.

### 3.1 Dados

As ações de *e-Science* costumam privilegiar a disponibilização e manutenção de bases de dados abertos, de acesso público, que subsidiem o trabalho de pesquisa tanto no âmbito individual como no colaborativo, dentro do espírito da ciência aberta. A formação desses conjuntos de dados é, em geral, decorrente de resultados de projetos de pesquisa financiados com recursos públicos, sendo disponibilizados para livre acesso após um período de carência. No entanto, mesmo projetos acadêmicos de *e-Science*, cuja intenção é trabalhar com dados abertos, falham em um ou mais aspectos, especialmente no que diz respeito às normas de divulgação de informações (DAVID; BESTEN; SCHROEDER, 2006).

No que se refere ao compartilhamento de dados científicos, em projetos de *e-Science*, colocam-se então questões de várias ordens, tais como: onde os dados serão armazenados? Como serão organizados? Quem terá acesso? Quem irá pagar pela infraestrutura de armazenamento, compartilhamento e análise desses dados? (OHNO-MACHADO, 2012). A esse respeito, David, Besten e Schroeder (2009, p. 3) assinalam ainda:

- a) em que proporção e quão rapidamente as informações sobre procedimentos e dados de pesquisa são divulgados?
- b) o quão completos os dados e procedimentos são documentados – de forma a que não estejam apenas acessíveis, mas também utilizáveis para quem não participa diretamente do grupo do projeto?
- c) em que termos e a partir de quanto tempo pesquisadores externos poderão ter acesso a material e dados resultantes do projeto?
- d) em que momento a propriedade intelectual é considerada (pelos membros dos projetos ou pelas instituições de origem)? Haverá licenciamento para usuários externos ou será uma base de uso exclusivo?
- e) os direitos de uso desses recursos “protegidos” estarão condicionados ao pagamento de *royalties*?
- f) os acordos de transferência de materiais impõem o pagamento de taxas (para as pesquisas celulares, reagentes, espécies) substancialmente mais altas por parte de pesquisadores externos?

Essas questões dizem respeito diretamente aos mecanismos de governança dessas equipes e projetos, isto é, às formas de gestão e resolução de conflitos e de poder.

Ohno-Machado (2012) identifica três formas de compartilhamento que vêm sendo praticadas em projetos de *e-Science*: a) *download* de dados; b) acesso e análise de dados remotos; c) importação de um conjunto de ambiente de *software*.

No caso do *download* de dados, usuários responsáveis pela busca identificam fontes de dados relevantes a partir de servidores distribuídos ou centralizados e os baixam diretamente para computadores locais, do tipo estação de trabalho – de uso individual. Tal processo incorre em alguns problemas, tais como a geração de quantidades massivas de dados, necessidade de uma alta frequência de acesso aos servidores para a atualização dos dados e necessidade de banda larga de alta velocidade.

No modelo em que usuários acessam e analisam dados remotamente, ainda persiste a necessidade de conexão com a *internet*. Contudo, as responsabilidades dos usuários são reduzidas no caso de perda, roubo ou danificação de um computador o que, em outros casos, provocaria a perda de dados.

O modelo em que usuários importam todo um conjunto de ambiente de *software* é similar ao modelo anterior. No entanto, ao invés de os usuários usufruírem de recursos computacionais externos, eles fazem o uso de máquinas virtuais (*virtual machines* – VMs) que permitem a manipulação de dados locais (armazenados no computador do usuário), assim como dados de *sites* remotos. Esse modelo também permite a formação de redes de centros de colaboração, mesmo que políticas institucionais não permitam o compartilhamento de dados no nível individual.

No contexto da *e-Science*, devem ser, portanto, considerados os mecanismos de distribuição, o volume e a heterogeneidade dos dados. A questão da heterogeneidade, em particular, desempenha um papel chave devido à grande variedade de plataformas, fontes de dados, sistemas, formatos, interfaces, políticas, abordagens etc., ao mesmo tempo que amplia a complexidade no ambiente de trabalho e requer esforços para a busca de interoperabilidade e padronização (FIORE; ALOISIO, 2011).

Para 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 (SÀNCHEZ-ARTIGAS; 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 repetitivas e

demoradas. A ideia do congelamento subentende a necessidade de os dados serem salvos ou publicados com um detalhamento e padronização mais extensivos possíveis, dispensando a necessidade de alterações posteriores.

Em diversas áreas, como no caso das ciências biomédicas citado por Ohno-Machado (2012), as equipes não estão suficientemente equipadas para implementar processos de revisão ou manutenção de grandes bases de dados anotados<sup>12</sup> ou aplicações de *software*, sendo que 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 úteis. Existe ainda a falta de padronização de dados advindos de diferentes instituições, além da necessidade de ações que permitam a proteção da privacidade institucional e pessoal (no caso, por exemplo, de pacientes), além de um arcabouço político-institucional que garanta um quadro regulador adequado e aceito nos âmbitos organizacional, local/regional, nacional e internacional (OHNO-MACHADO, 2012).

### 3.2 Infraestrutura

Um dos aspectos centrais do desenvolvimento das práticas de *e-Science* refere-se à abertura de seus instrumentos e compartilhamento de uma infraestrutura comum, o que vem sendo objeto de controvérsias.

A chamada *ciberestrutura* abrange: artefatos tecnológicos (computadores, estoques, redes, *software*); serviços tecnológicos para redes; serviços de pessoas e organizações (compartilhamento de expertise para desenvolver e operar os artefatos tecnológicos e assistir seus usuários finais) (ATKINS et al., 2003 apud FRY; SHROEDER; BESTEN, 2009).

Abertura envolve aqui quatro grandes eixos: no acesso, no uso, na inexistência de encargos e na utilização de padrões não restritivos, não exclusivistas, isto é, abertos. “Em síntese, abertura significa ‘compartilhado’ ” (FRY; SHROEDER; BESTEN, 2009, p. 8). O compartilhamento dessa infraestrutura comum é considerado um fator crítico para a ampliação de uma massa crítica, na comunidade científica, capaz de dar continuidade ao desenvolvimento de ferramentas e serviços relevantes às práticas de *e-Science*. Não se trata

---

<sup>12</sup>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.

tão somente do acesso a artigos científicos ou mesmo a recursos primários (como bases de dados), mas, sobretudo do compartilhamento e acesso a códigos de *software*, inclusive de modo a poder modificá-los e, assim, contribuir para seu desenvolvimento.

Além dos aspectos regulatórios e institucionais, o estabelecimento de uma infraestrutura comum e aberta, em projetos de *e-Science*, enfrenta também dificuldades técnicas e operacionais, especialmente a compatibilidade e interoperabilidade entre sistemas computacionais (alguns dos quais privados e comerciais), bem como entre padrões, ontologias e tipos de metadados.

Infraestruturas para pesquisa colaborativa podem ser usadas também 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 pesquisa e desenvolvimento (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; SCHROEDER, 2008).

Ressalta-se que a noção de abertura deverá incorporar novos elementos, que justamente contribuem para novos desenvolvimentos, como “liberação para livre uso”, “documentação permitindo posterior desenvolvimento” e “produção por pares” (FRY; SHROEDER; BESTEN, 2009).

Quanto aos termos sob os quais cada integrante pode ingressar ou deixar um projeto de *e-Science*, David, Besten e Schroeder (2006, p. 3) enumeram as seguintes questões:

- a) quem está autorizado a juntar-se à colaboração?
- b) todos os participantes estão autorizados a receber acesso total às bases de dados do projeto e outras ferramentas-chave de pesquisa?
- c) o quão fácil ou difícil é para novos ingressantes definirem agendas distintas de investigação no contexto do projeto em andamento, e qual o nível de controle que os mesmos detêm sobre novas descobertas?

d) quais restrições são colocadas (formalmente e informalmente) com relação ao uso de dados, informações e conhecimentos na posse de pesquisadores antes dos mesmos deixarem a colaboração?

### 3.3 O aparato de proteção de direitos de propriedade intelectual (DPI)

Direitos de propriedade intelectual desempenham também um importante papel no âmbito da *e-Science*, incidindo sobre as informações científicas e as ferramentas computacionais, tais como textos, bases de dados, metodologias, *software*, entre outros.

Muitas das controvérsias sobre propriedade intelectual no contexto da *e-Science* não diferem daquelas no âmbito das práticas tradicionais de pesquisa. O compartilhamento via plataformas digitais traz, no entanto, novos ingredientes a essa problemática.

Um caso emblemático é o *Projeto Genoma Humano* (1990-2003), que envolveu mais de 5 mil cientistas de vários países no esforço de sequenciamento genético do ser humano. Ao longo do projeto, deu-se acirrada disputa entre os que, de um lado, contando com recursos públicos, pressionaram pelo livre acesso às informações resultantes da pesquisa; e, de outro, os que realizaram e patentearam seu próprio sequenciamento, ou ainda que capturaram dados de repositórios tornados públicos pelo *Projeto Genoma Humano*, misturando com seus próprios dados e comercializando as novas bases assim obtidas. Isto levantou a necessidade de novos arcabouços reguladores:

Apesar do fracasso na provisão comercial de dados genômicos, a captura privada de dados nessas instâncias alertou os pesquisadores sobre o perigo de simplesmente disponibilizar produtos comuns no domínio público; tal estratégia pode impedir o patenteamento de um produto particular, mas não impede a captura em alguma outra forma fechada, proprietária (BURK, 2006, p. 9).

Exemplo disso é a base de dados GenBank<sup>13</sup>, criada no decorrer do Projeto Genoma Humano para permitir o acesso, pela comunidade científica, a informações sobre sequências de DNA a partir de uma fonte atualizada e abrangente. Não há restrições para o uso e distribuição dos dados disponíveis no GenBank, o que não impede, por outro lado, que determinadas sequências ou porções de informações sejam patenteadas ou licenciadas por parte dos seus depositantes. Contudo, a validade de tais reivindicações não pode ser aferida

---

<sup>13</sup>Disponível em: <<http://www.ncbi.nlm.nih.gov/genbank/>>.

pelos gestores da base, já que as mesmas estão atreladas às regulações e jurisdições de seus países de origem.

As formas de colaboração da *e-Science* têm sido afetadas de modo particular pelos sistemas de *copyright*, patentes e segredos comerciais. Mais recentemente, novas formas *sui generis* de proteção de bases de dados vêm sendo estabelecidas, particularmente na União Europeia, regulando a extração de informações de coleções proprietárias. Como assinalado por Burk (2006, p. 9), “leis de *copyright* e de proteção de bases de dados podem de algum modo ampliar-se na *e-Science*, dada sua dependência de grandes coleções de dados e de *software* necessários para acessar e processar tais conjuntos de dados”.

Um aspecto do aparato legal que afeta mais particularmente a *e-Science* refere-se às discrepâncias institucionais entre distintas jurisdições nacionais dos participantes. A *e-Science*, que nasceu a partir de iniciativas de âmbito primordialmente nacional, vem hoje ganhando forte estímulo para sua internacionalização. Isto, por sua vez, suscita o desenvolvimento de políticas reguladoras (muitas das quais restritivas) de acesso a recursos computacionais acadêmicos presentes nos *grids*, a discrepância entre diferentes regimes de direito de propriedade intelectual, a formação de barreiras comerciais (como limitações de acesso em bases de dados e bibliotecas digitais de conteúdos acessíveis apenas para assinantes) e a necessidade de tradução de *software* ou tecnologias para permitir a interação de equipes situadas em diferentes contextos (SCHROEDER, 2008).

Cientistas localizados em diferentes jurisdições territoriais podem gerar conjuntamente novos dados e materiais, assim como é possível o acesso quase que imediato a dados produzidos em outras jurisdições. Mas a legislação sobre direitos de propriedade, uso e alocação desses materiais pode se diferenciar substancialmente entre essas jurisdições. “Determinar quais leis das diferentes jurisdições deveriam prevalecer é um exercício difícil e frequentemente incerto” (BURK, 2006, p. 10).

Apesar das pressões pela harmonização da legislação internacional de propriedade intelectual – o Acordo TRIPs, no âmbito da Organização Mundial do Comércio (OMC) é apontado como a iniciativa mais proeminente nesse sentido –, prevalecem importantes diferenças entre os arcabouços jurídicos nacionais nessa área. Assim, ainda que conflitos interjurisdicionais existam de longa data, eles vêm se aprofundando com a configuração de um ambiente global de cooperação facilitado pelas plataformas digitais.

Ante essas dificuldades, desenvolvedores de projetos de *e-Science* recorrem, em alguns casos, ao estabelecimento de contratos para reger os termos entre as partes de um projeto colaborativo, a exemplo do que se faz no mundo empresarial. Alerta-se, no entanto, que tais contratos serão sempre incompletos e limitados em sua aplicação, ante a possível prevalência de normas estabelecidas em arcabouços legais mais abrangentes (BURK, 2006).

Outra alternativa tem sido o uso criativo de licenças do tipo das adotadas para *software* livre e de código aberto (*open source*), como as licenças *copyleft*, e também as licenças Creative Commons. Esses tipos de licenças têm um caráter “viral”, garantindo a manutenção do acesso aberto dos produtos e licenças derivados da cópia, adaptação e redistribuição do que foi originalmente licenciado. A cultura e o modelo de licenciamento *open source* vêm influenciando fortemente esse debate no campo da *e-Science*.

Um exemplo desse tipo de licença foi adotado pelo *Projeto HapMap*<sup>14</sup> que, com isto, pretendeu estabelecer, como condição para o acesso a dados livremente disponíveis pelo projeto, a concordância dos usuários em não patentear aplicações derivadas desses dados, bem como em compartilhá-los apenas com aqueles que aceitassem tais condições. Estabeleceu-se ainda que quaisquer patentes obtidos desses usos derivados devessem obedecer a essas mesmas condições.

#### **4 CONSIDERAÇÕES FINAIS**

Adeptos da ciência aberta advogam que a ampla disseminação das informações e conhecimentos resultantes da atividade de pesquisa promove o aumento dos estoques de conhecimento público, propiciando não apenas a ampliação dos índices gerais de produtividade científica e de inovação, como também das taxas de retornos sociais dos investimentos em ciência e tecnologia. Além disso, tem-se demonstrado historicamente que é no compartilhamento e na abertura, de modo coletivo e não individual, que ocorre a criatividade e a inovatividade, valendo-se crescentemente das infraestruturas de conexão e interação, ou seja, das redes (ALBAGLI; MACIEL, 2012).

A necessidade de resolução de problemas de alta complexidade e os elevados custos da pesquisa têm movido boa parte dos pesquisadores a buscarem colaboração aberta,

---

<sup>14</sup>O Projeto Internacional HapMap é de uma parceria de cientistas e agências financiadoras do Canadá, China, Japão, Nigéria, Reino Unido e Estados Unidos para desenvolver um recurso público que ajude pesquisadores a encontrarem genes associados a doenças humanas e às possíveis respostas para fármacos.



frequentemente por meios interpessoais e informais, a despeito dos limites macro e meso-institucionais. A formalização de redes de colaboração interinstitucionais enfrenta barreiras que, frequentemente, levam ao engessamento da pesquisa e do intercâmbio de conhecimentos e informações, na contracorrente da agilidade hoje propiciada pelas novas plataformas de informação e comunicação.

Persiste, por outro lado, a rivalidade e a competitividade desencadeadas pela disputa de prioridade na revelação de descobertas e invenções, estimulando a busca pelo reconhecimento individualizado e minando muitas vezes iniciativas de cooperação.

O recente desenvolvimento da *e-Science* insere-se neste quadro. Como ficou demonstrado, o estímulo para a construção e o uso de plataformas e tecnologias na pesquisa apresenta novos desafios para a organização de equipes, que envolvem uma pluralidade de questões, incluindo as de ordem técnica e tecnológica; as regulatórias (particularmente as de proteção de direitos de propriedade intelectual) e os arranjos e mecanismos de governança que reconheçam e contribuam para lidar com as diferentes visões, culturas e interesses dos participantes. A perspectiva de compartilhamento aberto seria então crucial para o próprio desenvolvimento da *e-Science* enquanto tal.

Trata-se, assim, de situar as novas práticas da *e-Science* no contexto de reconfiguração do regime de informação em C&T. Como já assinalado, nesta análise, longe de se restringir à dimensão institucional formal, deve-se reconhecer o regime de informação como um campo de disputa e conflito, assim como de negociação e de estabilização. Para além de uma perspectiva institucionalista estrita – foco exclusivo no arcabouço legal –, é preciso levar em conta o processo social no qual as ações e práticas informacionais efetivamente ocorrem.

Dessa perspectiva, não se trataria de dois regimes antagônicos no campo da *e-Science*, mas de *um* regime cuja configuração está sendo pautada pela disputa entre distintas perspectivas. Até que ponto a *e-Science* irá de fato encorajar e possibilitar novas formas de colaboração e compartilhamento de informação e conhecimento nos moldes de uma ciência aberta? Dependerá dos desdobramentos concretos dessas disputas no plano das condições e resultados das práticas colaborativas, que, ao final, irão ou não contribuir para modificar os arcabouços institucionais vigentes.

## REFERÊNCIAS

ALBAGLI, S.; MACIEL, M.L. Informação, conhecimento e democracia no Capitalismo Cognitivo. In: COCCO, G.; ALBAGLI, S. (Org.). **Revolução 2.0 e a crise do capitalismo global**. Rio de Janeiro: Garamond, 2012.

ALBAGLI, S.; MACIEL, M.L. Informação, poder e política: a partir do Sul, para além do Sul. In: ALBAGLI, S.; MACIEL, M.L. (Org.). **Informação, conhecimento e poder**: mudança tecnológica e inovação social. Rio de Janeiro: Garamond, 2011.

BARJAK, F.; ECCLES, K.; MEYER, E. T.; SCHROEDER, R.; ROBINSON, S. The emerging governance of e-Infrastructure. **Journal of Computer-Mediated Communication**, v. 18, n. 2, p. 1-24, Jan. 2013. Disponível em: <<http://dx.doi.org/10.1111/jcc4.12000>>. Acesso em: 26 fev. 2013.

BELL, G.; HEY, T.; SZALAY, A. Beyond the data deluge. **Science**, v. 323, n. 5919, p. 1297-1298, 6 Mar. 2009. Disponível em: <<http://dx.doi.org/10.1126/science.1170411>>. Acesso em: 25 mar. 2013.

BRAMAN, Sandra (Org.). **The emergent global information policy regime**. Houndsmills, UK: Palgrave Macmillan, 2004.

BURK, D. L. Intellectual property in the context of E-Science. **Minnesota Legal Studies Research Paper**, n. 6-47???, 2006. Disponível em: <<http://ssrn.com/abstract=929479>>. Acesso em: 1 jul. 2013.

DAVID, P. A.; BESTEN, M. L. den; SCHROEDER, R. **Will e-Science be open Science?** [S.l.][s.n.]14 Dec. 2008. Working paper series. Disponível em: <<http://dx.doi.org/10.2139/ssrn.1317390>>. Acesso em: 1 jul. 2013.

DUTTON, W. **Collaborative network organizations: new technical, managerial and social infrastructures to capture the value of distributed intelligence**. Oxford: Oxford Internet Institute, 17 Nov. 2008. DPN Working Paper Series, n. 5. Disponível em: <<http://dx.doi.org/10.2139/ssrn.1302893>>. Acesso em: 14 jun. 2012.

FIORE, S.; ALOISIO, G. Special section: data management for eScience. **Future Generation Computer Systems**, v. 27, n. 3, p. 290–291, Mar. 2011. Disponível em: <<http://dx.doi.org/10.1016/j.future.2010.08.012>>. Acesso em: 3 maio 2013.

FOSTER, I.; KESSELMAN, C. Computational grids. In: \_\_\_\_\_ (Ed.). **The grid**: blueprint for a new computing infrastructure. Los Altos, CA: Morgan Kaufmann, 1999. cap. 2., p. 15-51. Disponível em: <[http://www.elsevierdirect.com/companions/9781558609334/appendices/Chapter\\_02.pdf](http://www.elsevierdirect.com/companions/9781558609334/appendices/Chapter_02.pdf)>. Acesso em: 28 maio 2013.

FROHMAN, Bernd. Taking policy beyond information science: applying the actor network theory for connectedness – information, systems, people, organizations. In: ANNUAL CONFERENCE OF THE CANADIAN ASSOCIATION FOR INFORMATION SCIENCE, 23., 1995, Edmonton. **Proceedings...** Edmonton, Alberta [s.n.] 7-10 June 1995. Connectedness: Information, Systems, People, Organizations. Disponível em: <http://instruct.uwo.ca/faculty/Frohmann/actor.htm>. Acesso em: jan. 2010.

FRY, J.; SCHROEDER, R.; BESTEN M. den. Open science in e-science: contingency or policy? **Journal of Documentation**, v. 65, n. 1, p. 6-32, 2009. Disponível em: <[www.emeraldinsight.com/0022-0418.htm](http://www.emeraldinsight.com/0022-0418.htm)>. Acesso em: 1 jul. 2013.

GONZÁLEZ DE GÓMEZ, Maria Nélide. Novos cenários políticos para a informação. **Ciência da Informação**, Brasília, v.31, n.1, p.27-40, jan. 2002.

GRAY, J. eScience: a transformed scientific method. Transcrição de palestra ministrada por Jim Gray no Conselho Nacional de Pesquisa (EUA), 11 Jan. 2007. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm: data-intensive scientific discovery**. Redmond: Microsoft Research, 2009.

MEDEIROS, J. da S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, v. 8, n. 2, jul./dez. 2012. Disponível em: <<http://revista.ibict.br/liinc/index.php/liinc/article/view/488>>. Acesso em: 29 dez. 2012.

MERTON, R. K. **The sociology of science: theoretical and empirical investigations**. Chicago: University of Chicago, 1973.

OHNO-MACHADO, L. To share or not to share: that is not the question. **Science Translational Medicine**, v. 4, n. 165, Dec. 2012. Disponível em: <<http://dx.doi.org/10.1126/scitranslmed.3004454>>. Acesso em: 15 mar. 2013. (Seção Comentários).

QUIN, L. R. E. **XML essentials. WORLD WIDE WIBE CONSORTIUM (W3C)**. [S.l.][s.n.] c2010. Disponível em: <<http://www.w3.org/standards/xml/core>>. Acesso em: 2 jul. 2013.

SÀNCHEZ-ARTIGAS, M.; GARCÍA-LÓPEZ, P. eSciGrid: a P2P-based e-science grid for scalable and efficient data sharing. **Future Generation Computer Systems**, v. 26, n. 5, p. 704-719, May 2010. Disponível em: <<http://dx.doi.org/10.1016/j.future.2009.05.013>>. Acesso em: 3 maio 2013.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, v. 22, n. 3, p. 179-191, set./dez. 2012. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/12224/8586>>. Acesso em: 28 jan. 2013.

SCHROEDER, R. e-Sciences as research technologies: reconfiguring disciplines, globalizing knowledge. **Social Science Information**, v. 47, n. 2, p. 131-157, Jun. 2008. Disponível em: <<http://dx.doi.org/10.1177/0539018408089075>>. Acesso em: 22 maio 2013.

SONNENWALD, D. H. Scientific collaboration. In: CRONIN, B. (Ed.). **Annual Review of Information Science and Technology**. Medford: Information Today, 2007. v.41, cap. 14. p. 643-681. Disponível em: <<http://dx.doi.org/10.1002/aris.2007.1440410121>>. Acesso em: 18 out. 2012.

VOSS, A.; ASGARI-TARGHI, M.; PROCTER, R.; FERGUSON, D. Adoption of e-Infrastructure services: configurations of practice. **Philosophical Transactions of Royal Society A**, London, v. 368, n. 1926, p. 4161-4176, Aug. 2010. Disponível em: <<http://dx.doi.org/10.1098/rsta.2010.0162>>. Acesso em: 5 mar. 2013.