E-science

Valdinéia Barreto Ferreira

SciELO Books / SciELO Livros / SciELO Libros

FERREIRA, V.B. E-science. In: *E-science e políticas públicas para ciência, tecnologia e inovação no Brasil* [online]. Salvador: EDUFBA, 2018, pp. 13-30. ISBN: 978-85-232-1865-2.
<https://doi.org/10.7476/9788523218652.0003>.

All the contents of this work, except where otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença [Creative Commons Atribuição 4.0](https://creativecommons.org/licenses/by/4.0/).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia [Creative Commons Reconocimiento 4.0](https://creativecommons.org/licenses/by/4.0/).

E-SCIENCE

O cerne da transformação que estamos vivendo na revolução atual refere-se às tecnologias da informação, processamento e comunicação. A tecnologia da informação é para esta revolução o que as novas fontes de energia foram para as revoluções industriais sucessivas, do motor a vapor à eletricidade, aos combustíveis fósseis e até mesmo à energia nuclear. (CASTELL, 2006, p. 68)

A necessidade de medir como forma de comprovação se faz presente na Ciência em muitas áreas do conhecimento e tornou-se prática constante. Conforme Cesar Junior (2011, p. 7), “a evolução da ciência está profundamente ligada à evolução dos instrumentos que permitem a realização de observações. Em muitos momentos a Ciência precisou aguardar o aparecimento de tecnologias apropriadas de medição dos fenômenos de interesse.” A refutação, aprimoramento e proposição de teorias e pressupostos dependiam dessas tecnologias.

A necessidade de uma tecnologia adicional que suportasse o desenvolvimento das pesquisas e aliviasse o isolamento do pesquisador foi o impulso extra para o aparecimento da *e-Science*; um composto de *hardware*, *software* e um preponderante cunho colaborativo. Esse foi o empurrão necessário para o desenvolvimento de uma metodologia cujo uso intensivo de dados fundamentava a busca para o tratamento dos insumos de um fazer científico abundante. Em seu bojo identifica-se a preocupação crescente pela captura, curadoria e análise dos dados, uma ajuda de peso para que o completo ciclo da pesquisa se concretize.

A *e-Science* caracteriza-se como um novo paradigma da Ciência, a qual marca sua presença e destaca a transição normal e algumas vezes imperceptível no curso evolutivo da Ciência. Apresenta-se paralelamente a outras manifestações científicas, em alguns aspectos, chega até a sobrepor-las.

Na Figura 1, resgatam-se os paradigmas da Ciência segundo Gray e Szalay (2007), os quais ilustram a evolução da Ciência e a enquadram em um período temporal no qual surge a *e-Science*. Realizou-se um passeio pelos primórdios da Ciência e identificaram-se os fenômenos naturais como determinantes do curso seguido, um momento em que a Ciência era eminentemente empírica. A teoria tornou-se predominante no período seguinte com seus modelos e generalizações. A computação ao sobrepor-la, ditou as regras seguintes e passou a conduzir de modo consistente os rumos da Ciência. A atualidade vivencia um período marcado pelo enorme volume de dados, exploração da tecnologia de larga escala e a colaboração entre diversos domínios do conhecimento, características que assinalam a *e-Science*. Esse é um paradigma emergente, o qual influenciou de modo preponderante os rumos tomados no desenvolvimento da pesquisa científica contemporânea.

Figura 1 – Paradigmas da Ciência

Fonte: Adaptado de Gray e Szalay (2007).

Contudo, questões acerca da atribuição de rótulos aos eventos científicos são temas de debate na comunidade científica constantemente, e Bell (2011, p. 11) demonstra preocupação ao afirmar que “disciplinas científicas não podem ser facilmente encapsuladas em alguns poucos números e nomes compreensíveis [...]”. Essa assertiva encontra consonância em sinalizações anteriormente pautadas por Jankowski (2009), ao alertar, para a adoção dos rótulos frequentemente debatidos e contestados, os quais marcaram os diversos eventos históricos da humanidade. As mudanças nem sempre ilustram as reais transformações e muito do que se tem propagado como novo e revolucionário é frequentemente a aplicação de práticas organizacionais e de boa investigação no âmbito da Internet, reflexões que incitam a se buscar fundamentos que permitam identificar a verdadeira natureza de um termo novo e tão potente como a *e-Science*.

Compreensão conceitual, origem e base fundante da *e-Science*

O termo *e-Science* ou (*eScience*), originário no Reino Unido e predominante no restante da Europa, foi cunhado por John Taylor, no ano de 1999,¹ quando ocupava o cargo de diretor geral do Research Councils UK, conforme Jankowski (2007). Traduzido para o português como e-Ciência, este termo adquiriu um significado que representa a potência da ciência melhorada com o uso intensivo das TICs e sua ampliação em torno de um esforço colaborativo.

Em estudo realizado por Costa e Cunha (2015, p. 128), o qual apresenta a análise de publicações sobre a temática *e-Science* na área da Ciência da Informação, os autores inferem que: após buscas exaustivas nas bases de dados internacionais Education Resources Information Center (ERIC), Library and Information Science Abstracts (LISA), Library e Information Science & Technology Abstracts (LISTA) e Web of Science, “ainda não há trabalho na literatura sobre a área atrelado ao fenômeno da *e-Science* que identifique a sua comunidade, os autores que mais publicam, bem como os países que produzem essa literatura.” Eles observam ainda que “não há um mapeamento a respeito do tema que facilite a identificação dos canais de comunicação utilizados pela comunidade.” O estudo realizado evidência a pertinência de pesquisas relacionadas a *e-Science* entrelaçada à área

1 Segundo Tolle e colaboradores (2011, p. 235), o termo foi cunhado no ano 2000.

da Ciência da Informação, devido a relevância e atualidade do tema e produção na área ainda pequena.

Ao referir-se ao documento do Research Councils UK,² Vaz (2011, p. 20) resalta que o significado da partícula “e” do termo “*e-Science* normalmente não é explicado mas, [...] sem dúvida, começou com o significado de *eletronic* (eletrônico) e agora representa melhor *enhanced* (melhorada) ou *enabled* (habilitada)”. Considera-se pertinente neste estudo, o emprego da partícula “e” com o sentido do termo inglês *enhanced* (melhorada), por refletir de modo apropriado o poder atribuído à Ciência na contemporaneidade. Na pesquisa em curso adota-se o termo *e-Science* com o desejo de resaltar a partícula “e”, a qual simboliza a grande diferença que o eletrônico, melhorado e ampliado tem causado à Ciência nos tempos atuais. Essa característica que faz toda a diferença no contexto de análise.

Atenta-se para a dificuldade formal de se definir o termo *e-Science*, devido a sua precoce vida e falta de consenso na comunidade científica. O melhor caminho para se compreender a definição foi identificar suas origens. Quando mencionado pela primeira vez por John Taylor, este termo foi associado ao reconhecimento do papel cada vez mais importante que a Tecnologia da Informação (TI) representava, no século XXI, para a pesquisa científica intensiva em dados, multidisciplinar e colaborativa. Taylor o designou como o conjunto de ferramentas e tecnologias necessárias para apoiar essa pesquisa (TOLLE et al., 2011).

A inconsistência terminológica do termo *e-Science* está presente na literatura científica e evidencia-se em alguns estudos realizados. Destacam-se dois deles para exemplificar a situação sinalizada e corroborar o que foi pontuado no trabalho de Álvaro e colaboradores (2011, tradução nossa) ao explorarem a compreensão do termo e apresentarem exemplos de trabalhos na literatura relacionados à confusão terminológica existente. Esses autores afirmam que a despeito da variedade de nomes com a qual se denomina a *e-Science* e as questões terminológicas recorrentes, a *e-Science* representa uma grande e significativa mudança.

No trabalho realizado por Gold (2007, [p. 1], tradução nossa), a autora atém-se aos termos *e-Science* e *cyberinfrastructure*, mais detidamente ao segundo. Os identifica-os como recorrentes da mesma temática e intercambiáveis, diferen-

2 O Research Councils UK, conforme Vaz (2011, p. 20), “é uma parceria estratégica entre os Conselhos de Pesquisa do Reino Unido criada em 2002 para permitir que eles trabalhem colaborativamente de maneira mais eficaz [...]”

ciados apenas pela questão geográfica de aplicação. Observa também como a ideia a eles atribuída está no centro das ambições e promessas da Ciência no novo século. A autora manifesta preocupação com os profissionais da informação, especialmente com os bibliotecários que trabalham diretamente com os produtores e usuários dos dados e com os produtos advindos dessa produção. Contribui com uma cartilha para a *cyberinfraestrutura* destinada a bibliotecários e conclui ao pontuar questões como: captura, arquivamento, curadoria, acesso, interoperabilidade, análise, serviços, ciclo de vida, instrumentos políticos e modelos de negócios relacionados aos dados de pesquisa.

No segundo exemplo, representado pelo estudo de Schottlaender (2010, p. 99, tradução nossa), a questão conceitual dos termos *e-Science* e *cyberinfraestrutura* é pontuada. O autor é muito incisivo ao distingui-los terminologicamente e critica o fato desses termos serem considerados intercambiáveis ou carregados do mesmo significado. Salienta que não os define como tal e não fará uso dessa prerrogativa para tratá-los em seu estudo e considera a *cyberinfraestrutura* como um meio para atingir o fim, que é a *e-Science*. Para esse autor, esses termos são dependentes, mas não iguais, como destaca em sua assertiva: “[...]’ ou seja, *cyberinfrastructure* é condição para o desenvolvimento e realização da *e-Science*. Esta última não pode, literalmente, ser realizada sem a primeira; e sem a última, a primeira é menor do que uma solução em busca de um problema.”

São posições opostas entre os autores que apresentam os conceitos de *e-Science* e *cyberinfraestrutura* e essas representam o quadro indefinido que caracteriza a compreensão da comunidade científica sobre a temática e evidenciam claramente a necessidade de muitos estudos para que uma posição seja contemplada na questão. Somente para ilustrar a situação são apresentados outros termos que se inserem no contexto e representam a questão maior que os congrega, bem como as mudanças ocorridas na prática científica com o uso intensivo dos artefatos tecnológicos.

Esses termos carregam preferências, linhas de pesquisa, objetivos etc, e são identificados como: ciência orientada por dados (*data-driven science*); computação fortemente orientada por dados (*data-intensive computing*); mineração de dados (*data mining*); quarto paradigma (*fourth paradigm*); dos dados ao conhecimento (*from data to knowledge*); ciência com uso intensivo de dados (*data-intensive science*); pesquisa

eletrônica (*e-Research*). (CESAR JUNIOR, 2011; JANKWSKI, 2009). Segundo Whitmire (2013, tradução nossa), o termo *e-Research*, inclusive, no contexto bibliotecológico, foi definido pela Association of Research Libraries (ARL) (2013) como:

uma forma de rede colaborativa, de larga escala e computacionalmente intensiva de pesquisa e de bolsas de estudo para todas as disciplinas, incluindo todas as ciências físicas e naturais, relacionadas e aplicadas às disciplinas tecnológicas, biomédicas, ciências sociais e as humanidades digitais.

Essa definição, segundo Soehner e colaboradores (2010, p. 3), assemelha-se com a que foi atribuída ao termo *e-Science* pela própria ALA, em 2010, o que reforça a imprecisão terminológica e provoca confusão quanto ao entendimento dos conceitos na área. Fica evidente a necessidade de estudos que indiquem como os pesquisadores das diversas áreas de conhecimento acolhem este conceito, sem descartar, também, a necessidade de um vocabulário que melhor o defina. Contribui-se, desse modo, com construção de um referencial terminológico para a área da Ciência da Informação. Como sinalizado antes, estes termos representam as mudanças em curso na Ciência e resgatam outros mais antigos que foram precursores desse contexto atual como: *Big Science*, *Cyberscience* e mais recentemente o *Big Data*, o *Open Data* e o *Open Science*.

Independentemente da terminologia adotada, faz-se necessário registrar que tanto a *e-Science* quanto a *cyberinfrastructure* foram iniciativas originárias de políticas governamentais, a primeira no Reino Unido e a segunda nos Estados Unidos, com a intenção de fortalecer o desenvolvimento da alta computação e da tecnologia de informação e comunicação (JANKWSKI, 2007).

A *e-Science* surgiu da necessidade urgente de se enfrentar o “dilúvio de dados”, popularizou uma nova metodologia de pesquisa, desenvolvida em diferentes lugares, com uma história ainda imprecisa, denominações e grafias distintas, flexíveis e ainda tênues, mas com o objetivo comum de se usar tecnologias de computação em rede para melhorar a colaboração e os métodos inovadores de investigação (WHITMIRE, 2013). Conforme Levy (1999, p. 13) o termo “segundo dilúvio” das informações foi atribuído a Roy Ascott e ilustra perfeitamente o “dilúvio de dados” tão utilizado atualmente.

As telecomunicações geram esse novo dilúvio por conta da natureza exponencial, explosiva e caótica de seu crescimento. A quantidade bruta de dados disponíveis se multiplica e se acelera. A densidade dos links entre as informações aumenta vertiginosamente nos bancos de dados, nos hipertextos e nas redes. Os contratos transversais entre os indivíduos proliferam de forma anárquica. É o transbordamento caótico das informações, a inundação de dados, as águas tumultuosas e os turbilhões da comunicação, a cacofonia e o psitacismo ensurdecido das mídias, a guerra das imagens, as propagandas e as contrapropagandas, a confusão dos espíritos.

Ao retomar a questão conceitual observa-se que diferentemente do termo *e-Science*, o qual é relativamente novo, o termo *cyberinfrastructure*, mais usado nos Estados Unidos, tem mais tempo de aparecimento e seu emprego inicial é atribuído a Richard A. Clarke e Jeffrey Hunkeyr, do National Science Foundation (NSA), em entrevista concedida em 1998, nos Estados Unidos (CLARKE; HUNKER, 1998). Esses termos são identificados como sinônimos, em alguns casos, e se referem ao uso de tecnologias de computação em rede para melhorar a colaboração e os métodos inovadores de investigação, como foi destacado nos exemplos extraídos da literatura científica.

Acredita-se ser mais apropriado observarem-se as colocações e adequá-las entre as várias proposições, de acordo com a realidade de cada país. A Ciência pode atestar conceitos e, por essa razão, definições ou terminologias são importantes. Entretanto, os resultados advindos do fazer científico proporcionarão à sociedade um ganho de maior vulto enquanto a terminologia se define com o amadurecer e aprofundamento da temática e realização de pesquisas. Devido ao pouco tempo de existência da *e-Science*, ainda decorrerá muito tempo para que esse amadurecimento ocorra.

As características atribuídas aos dois termos atestam que o foco na pesquisa é atribuição da *e-Science*, justamente por ser considerada como finalidade do processo, ao passo que a *cyberinfrastructure* é mais abrangente e robusta por representar o meio necessário ao processo de construção do conhecimento científico. Inicialmente a *e-Science* foi usada em áreas como as Ciências Naturais e Biológicas, mas atualmente não se restringe apenas a essas. Ambas,

cyberinfrastructure e *e-Science* incorporaram maior ênfase nos recursos de super-computação e inovação devido ao grande volume de dados envolvidos nos esforços de pesquisa.

A *e-Science* agrega-se à infraestrutura, a qual fornece aos cientistas ferramentas que apresentam um escopo para coletar, armazenar, interpretar e analisar a rede de dados produzida e disseminá-la para outros grupos de trabalho. Conforme Soon e Park ao citar Hine, a *e-Science* tem sido definida como “usos da infraestrutura de Tecnologia de Informação e Comunicação (TIC) para o armazenamento de dados científicos, análises e realização de trabalho colaborativo, também conhecida nos EUA como as *cyberinfraestruturas* e no Reino Unido como *e-Science*.” (SOON; PARK, 2009, tradução nossa).

Esse modelo significa o elo entre uma tendência da Ciência para utilização da tecnologia e o compartilhamento de saberes. A prática colaborativa, característica da *e-Science*, desempenha papel fundamental ao aliar a tecnologia por essa desenvolvida ao movimento do homem em direção à descoberta, intrínseco à sua natureza (E-SCIENCE, 2012b).

A tecnologia, como peça chave no processo, desempenha seu papel, proporciona a operacionalização dos projetos pretendidos e garante a disponibilidade e acessibilidade da informação científica e tecnológica. Considerada por Jim Gray como o quarto paradigma, a *e-Science* pode ainda ser definida como a utilização de ferramentas computacionais para analisar grandes conjuntos de dados científicos (E-SCIENCE, 2012a).

Analisa-se, na próxima seção deste livro, a atual situação relativa à infraestrutura científica, tecnológica e de inovação, em termos de *e-Science*, e espera-se poder associá-la ao contexto brasileiro, representado pelos INCTs da área de Nanotecnologia. Isso configuraria a evidência de que se está caminhando, efetivamente, para o envolvimento real e para a frutífera inserção do Brasil na agenda da *e-Science* mundial.

Modelo de infraestrutura e-Science

Conceber um modelo que represente a infraestrutura e-Science requer fundamentos que a identifique e permita referência e comparação com outras es-

truturas. A alocação de recursos e todo arsenal necessário para a sustentação da prática colaborativa contemporânea caracteriza esse tipo de infraestrutura, sobre o qual Albagli e colaboradores (2013, grifo do autor) expõem:

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, Webservices, Virtual Research Environments (VRE), tecnologias de notação e armazenamento de dados concebidas em padrões como XML3, entre outras.

Suposições são indicadas em torno da concepção de uma infraestrutura *e-Science* e, conforme Andronico e colaboradores (2011, p. 156, tradução nossa), atualmente ela é denominada como *e-infrastructure*, ou seja, “um ‘método científico’ que prevê a adoção de plataformas digitais de ponta conhecidas como infraestruturas eletrônicas em todo o processo da ideia à produção do resultado científico.” Segundo os autores citados esta infraestrutura eletrônica pode ser conceitualmente representada por três camadas:

- parte inferior, composta pelos instrumentos científicos e experimentos que fornecem grande quantidade de dados;
- em seguida, a camada de rede, centros de processamento de dados em rede e *software middleware* como a “cola” dos recursos; e
- o terceiro e mais alto nível que inclui pesquisadores que realizam suas atividades independente da localização geográfica, interagem com os colegas, compartilham e acessam os dados.

A definição de infraestrutura *e-Science* anteriormente apresentada possui consonância com a suposição de Albagli e colaboradores (2013). A seguir, apresenta-se um modelo de infraestrutura *e-Science* baseado nos elementos obtidos e compilados, na Figura 2.

Figura 2 – Modelo de Infraestrutura e-Science

Fonte: Elaboração da autora baseada em Andronico e colaboradores (2011).

As práticas colaborativas contemporâneas, especialmente as realizadas nos domínios de ponta, necessitam de uma infraestrutura potente a qual permita o fluxo rápido e eficiente dos dados em constante transmissão. Essa infraestrutura “requer um ambiente para o desenvolvimento eficiente de novos programas e recursos para rapidamente compartilhar os programas entre os membros dentro de uma comunidade de pesquisa.” (USAMI et al., 2011, p. 851, tradução nossa).

Neste cenário, a infraestrutura agregada à *e-Science* tornou-se a base em torno da qual os cientistas e tecnólogos estão trabalhando para alcançar novas descobertas e avanços para melhorar áreas nos diversos campos do conhecimento e desenvolver alternativas para lidar com a enorme quantidade de dados produzidos. Entretanto, Hey e Hey (2006, p. 517, tradução nossa) enfatizam que:

é importante ressaltar que a *e-Science* não é uma nova disciplina científica, em seu próprio direito: a *e-Science* é um atalho para o conjunto de ferramentas e tecnologias necessárias para apoiar a ciência, colaborativa em rede. Toda infraestrutura da *e-Science* se destina a capacitar os cientistas a fazer suas pesquisas de forma mais rápida, melhor e diferente.

A estrutura de rede, um elemento indissociável do pensamento contemporâneo, foi reconhecida por Roure e Hendler (2004) como a infraestrutura da *e-Science*. Eles salientam que também a gestão de informação, e não apenas de computação de alto desempenho ou instrumentação científica especializada, seria necessária para cumprir as metas de um programa com a amplitude que os programas de *e-Science* demandam.

Uma plataforma *e-Science* está cercada de um conjunto de elementos que viabilizam sua execução. Leva-se em consideração tanto os recursos computacionais, compostos por todo arcabouço tecnológico cercado pelo *hardware*, *software* e redes que suportam e gerenciam os dados produzidos quanto pelo capital humano com toda sua *expertise* e saber, e as instituições que fomentam e são responsáveis pelo abrigo e manutenção de toda infraestrutura envolvida.

Conforme Hey e Trefethen (2006) o termo “*Grid*” foi usado pela primeira vez em meados da década de 1990 para designar uma infraestrutura de computação distribuída para a ciência e engenharia avançada. Essa infraestrutura ocupava-se tanto do *hardware* como do *software* para apoiar os principais passos tipicamente envolvidos de captura, armazenamento, manutenção, análise e visualização dos dados, utilizando-se, desse modo, dos esforços dessa computação em grade, realizada quando uma rede permite que os recursos de cada computador sejam compartilhados com toda e qualquer máquina do sistema (CESAR JUNIOR, 2011; E-SCIENCE, 2012b).

Ao configurar-se como elo entre uma poderosa tendência da Ciência – o compartilhamento de saberes e fazeres e a infraestrutura fornecida pela Ciência da Computação – a *e-Science* alia o movimento do homem em direção à descoberta, intrínseco à sua natureza, à tecnologia por ele desenvolvida. O registro recente dessa metodologia, de aproximadamente 16 anos, e uma história em construção desafiam o pesquisador a conhecê-la melhor e a buscar referências mais precisas sobre seu aparecimento, definindo uma agenda motivadora e instigante para a pesquisa.

Segundo Jim Gray, a *e-Science* é o ponto onde a TI encontra cientistas (GRAY; SZALAY, 2007). A tecnologia, como peça chave no processo, desempenha seu papel ao proporcionar a operacionalização dos projetos pretendidos e ao garantir a disponibilidade e acessibilidade da informação científica e dos dados

primários produzidos com a captura, curadoria e análise. O tratamento desses insumos com a interveniência dos recursos tecnológicos da infraestrutura *e-Science* é ansiada por Bell (2011, p. 12), o qual acredita “que logo mais chegará um tempo em que os dados viverão para sempre em mídia arquivada – igualzinho ao armazenamento em papel – e será acessível publicamente a homens e máquinas numa nuvem”. Descrita também como o quarto paradigma, conforme indicado anteriormente, a infraestrutura que compõe a *e-Science* pode ainda ser definida como as ferramentas computacionais utilizadas para analisar grandes conjuntos de dados científicos aliados à colaboração impulsionada pelos pesquisadores. Conforme Mundie (2011, p. 232),

Essa infraestrutura de computação aperfeiçoada tornará possível a verdadeira biblioteca digital global, onde todo o ciclo de vida da pesquisa acadêmica – da formulação à publicação – terá lugar num ambiente eletrônico e estável disponível para todos, de forma aberta. Durante o desenvolvimento das ideias científicas e sua publicação subsequente, os cientistas poderão interagir entre si virtualmente – compartilhando dados, fluxos de trabalho e pesquisa.

Ao resgatar a palestra que foi proferida por Jim Gray, em 11 de janeiro de 2007, no National Research Council/Computer Science and Telecommunications Board, (NRC-CSTB), na qual ele expande sua visão da Ciência com uso intensivo de dados, Tolle e colaboradores (2011, p. 235) apresentam as sete áreas-chave indicadas por Gray para ação a serem consideradas pelas agências de fomento. São elas:

- estimular tanto o desenvolvimento de ferramentas de *software* como o apoio a essas ferramentas;
- investir em ferramentas em todos os níveis da pirâmide de recursos;
- estimular o desenvolvimento de Sistemas de Gerenciamento de Informações Laboratoriais (SGIL) genéricos;
- estimular pesquisa nas áreas de gerenciamento de dados científicos, análise de dados, visualização de dados e novos algoritmos e ferramentas;
- estabelecer bibliotecas digitais para apoiar outras ciências, assim como a Biblioteca Nacional de Medicina apoia as Biociências;

- estimular o desenvolvimento de novos modelos de publicação e novas ferramentas de autoria de documentos;
- estimular o desenvolvimento de bibliotecas digitais de dados que contenham dados científicos (não apenas metadados) e apoiem a integração com a literatura publicada (TOLLE et al., 2011, p. 235).

Existem muitas iniciativas de projetos *e-Science* distribuídas no mundo inteiro e fomentados por agências internacionais e conselhos que se encarregam da manutenção e suporte para o desenvolvimento efetivo. Destacam-se programas financiados pela National Science Foundation (NSF), nos Estados Unidos e os Research Councils UK (RCUK), no Reino Unido, cujas iniciativas na área datam dos anos 2000. Os centros de *e-Science* na Holanda, Suécia e Austrália são outros exemplos internacionais nas mais variadas áreas de conhecimento (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2015).

A *Amazon*,³ o *Disc*,⁴ o *MapReduce*⁵ do *Google*, o *Hadoop*⁶ e o *Dryad*⁷ são exemplos de infraestruturas emergentes de computação em nuvem e plataformas de computação de nova geração que estão voltadas para o gerenciamento e processamento de grandes volumes de dados. Parastatidis (2011, p. 176, grifo do autor) acrescenta que:

empresas como o *Google*, o *Yahoo* e a *Microsoft* estão demonstrando que é possível agregar enormes volumes de dados de toda parte na *web* e armazená-los, gerenciá-los e indexá-los, para depois construir experiências atraentes para os usuários em torno disso tudo.

3 A *Amazon* oferece serviços de infraestrutura de nuvem abrangentes e altamente especializados. Disponível em: <<http://aws.amazon.com>>.

4 Data-Intensive Super Computing (DISC). Disponível em: <www.pdl.cmu.edu/DISC>.

5 *MapReduce* “é um modelo de programação e aplicações associadas para processamento e geração de grandes conjuntos de dados” (DEAN; GHEMAWAT, 2004, p. 1). Disponível em: <<http://labs.google.com/papers/mapreduce.html>>.

6 *Hadoop* é uma plataforma de *software*, em *Java*, de computação distribuída voltada para *clusters* e grandes conjuntos de dados. Disponível em: <hadoop.apache.org>.

7 *Dryad* é uma infraestrutura que permite que um programador use os recursos de um *cluster* de computadores ou de um centro de dados para a execução de programas de dados paralelos. (MICROSOFT, 2016). Disponível em: <<https://www.microsoft.com/en-us/research/project/dryad/>>.

Essas iniciativas em *e-Science* são investimentos para a construção de ambientes favoráveis e que deem sustentação às atribuições que uma infraestrutura desse porte carrega. Além das iniciativas sinalizadas, inserem-se as do Japão nesse contexto de implementações. Apresenta-se, como exemplo, a iniciativa japonesa descrita por Usami e colaboradores (2011, p. 849, tradução nossa), o Projeto *REsources liNKage for E-science* (RENKEI), sucessor do projeto *National Research Grid Initiative* (NAREGI). A principal atividade desenvolvida pelo grupo que compõe o projeto foi estender a rede *middleware* NAREGI com a intenção de oferecer suporte em diferentes ambientes de rede. As especificidades desse projeto foram destacadas a seguir:

O RENKEI é um esforço para desenvolver um ambiente de *software* onde os pesquisadores podem criar e operar facilmente em comunidades em vários ambientes de rede. Especificamente nós estamos realizando desenvolvimento e pesquisa de grade *middleware* para utilizar o poder de computação federada, arquivos, bancos de dados etc., a partir de ambientes de rede diferentes, de ambientes de rede de nível laboratório de pequena escala, para ambiente de grade de nível central em super-computador de larga escala e outros ambientes de rede de países estrangeiros (USAMI et al., 2011).

Uma representação da estrutura que compõe o projeto RENKEI e a relação entre os grupos que o compõem foi apresentada na Figura 3.

Os grupos do RENKEI que desenvolvem atividades de P&D foram identificados conforme Usami e colaboradores (2011, p. 849, tradução nossa), da seguinte forma:

- federação de recursos de computação e aplicações de hospedagem de Organizações Virtuais;
- compartilhamento de arquivos e federação de catalogação de arquivos;
- federação de Bases de Dados e sistemas de gerenciamento de ID;
- API para várias grades *middleware*;
- avaliação e colaboração com os usuários.

Figura 3 – Projeto RENKEI – Suporte para ambientes de rede

Fonte: Usami e colaboradores (2011, p. 850).

Iniciativas desse porte evidenciam que se faz necessário, na esfera nacional, projetos que estejam voltados para uma infraestrutura complexa que suporte as novas formas do fazer científico, com os compartilhamentos dos dados empíricos produzidos e a colaboração entre os pesquisadores das diversas áreas do conhecimento. Conforme Vaz (2011, p. 11), “no Brasil, mais de uma década depois do início do programa do Reino Unido, há poucos cientistas que têm conhecimento ou interesse sobre o tema e tampouco há um programa nacional integrado”.

Contudo, os pesquisadores brasileiros, concentrados em São Paulo, começam a perceber a necessidade de engajamento nesse contexto e começam a intensificar os estudos e aplicações nesta temática. Atenta a essa necessidade e em parceria com a Ciência da Computação e Engenharia (CS&E), a Fundação de Amparo à

Pesquisa do Estado de São Paulo (FAPESP) organizou duas oficinas com as metas identificadas como G1 e G2, em agosto e novembro de 2009.

G1) para identificar novos rumos na Ciência da Computação (CS) para avançar a pesquisa na área, no estado de SP, e mecanismos para implementar iniciativas relacionadas;

G2) para investigar como essas instruções poderão, ao mesmo tempo ajudar no avanço de outros programas relevantes da FAPESP, com relevantes, com oportunidades de colaboração de pesquisa identificáveis com CS (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2014, tradução nossa).

As oficinas representaram os primeiros movimentos em busca da inserção da temática *e-Science* entre os pesquisadores de diversos domínios do conhecimento, localizados no estado de São Paulo, nas quais se realizaram discussões iniciais em torno de um esforço colaborativo para se buscar alternativas que levassem ao alcance das metas G1 e G2. Fizeram parte do grupo pesquisadores da área de computação e de outros domínios do conhecimento, os quais desenvolveram projetos nas áreas da “medicina, bioenergia, mudanças climáticas, biodiversidade, astronomia, física de partículas, desenvolvimento urbano e bibliotecas digitais (Brasileira).” Esses pesquisadores concluíram pela relevância da temática, que essa poderia proporcionar ganhos para a pesquisa nos diversos domínios (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2014).

Em continuidade à sua ação, a FAPESP realizou o *Latin American e-Science Workshop* 2013, em maio de 2013, seminário que buscou reunir estudantes e pesquisadores da Europa, América do Sul e do Norte e da Oceania para discutir e apresentar as implementações nessa área (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2013).

Ainda em 2013, a FAPESP criou o Programa FAPESP de Pesquisa em *eScience*. “O principal objetivo do programa era encorajar abordagens novas, ousadas e não convencionais para pesquisa de ponta, multidisciplinar, integrando grupos de pesquisa em Computação e em outras áreas.” (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2015).

O primeiro edital da FAPESP a contemplar projetos em *e-Science* foi lançado em dezembro de 2013. Para isso, realizou o investimento de R\$ 4 milhões e promoveu o apoio a projetos nas seguintes áreas: “Ciências Agrárias; Artes, Humanidades e Ciências Sociais; Bioinformática; Ciências Biológicas e Ciências da Saúde; Engenharia e Física; Clima e Ciências da Terra; e prática e educação em *eScience*.” (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2015).

A primeira seleção do programa contemplou quatro projetos em pesquisas espaciais, agropecuária, arqueologia e engenharia. A característica do fomento teve-se a Projetos Temáticos com pesquisadores de uma ou mais instituições, com duração de 60 meses, ou Auxílio Regular à Pesquisa, projeto individual, com duração de 24 meses. A contratação de profissionais da área de Computação é a contrapartida institucional do Programa *eScience*, em que se insere como exigência a política de gestão de dados, com as especificidades que a cercam. Em 2015, lançou-se nova chamada para propostas ao Programa *eScience* da FAPESP (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2015).

Outras iniciativas identificadas são as realizadas pelo Núcleo de Pesquisa em *e-Science* da Universidade de São Paulo (NAP-USP), responsável pelo projeto *scriptLattes*.⁸ A Biblioteca de dados Geoespaciais⁹ da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) lançou um sistema de armazenamento e consulta de conteúdo geoespacial, o qual busca ajudar a tornar menos árdua a tarefa de implementação dessa prática no Brasil. Essas iniciativas são muito bem vindas e um mapeamento mais apurado permitiria identificá-las e ampliar sua divulgação (BIBLIOTECA, 2013).

Identifica-se, nas iniciativas pautadas, a intenção de associar pesquisas sobre infraestrutura de informação e os seus respectivos domínios. Percebe-se que a busca por um melhor entendimento sobre esse conceito é latente e a associação

8 *scriptLattes* é um sistema de código aberto que foi desenvolvido para a compilação automática de dados referentes a produções bibliográficas, técnicas, artísticas, de orientações, projetos de pesquisa, prêmios e títulos, grafo de colaborações, mapa de geolocalização, coautoria e internacionalização de um conjunto de pesquisadores cadastrados na Plataforma Lattes (MENA-CHALCO; CESAR JUNIOR, 2009).

9 O acesso à Biblioteca de dados Geoespaciais pode ser realizado pelo endereço eletrônico: <<http://geo.cnpma.embrapa.br>>.

da *e-Science* a área da Nanotecnologia é recorrente. Em relatório realizado pela Lloyd's Register Foundation (LR) em 2004, encontra-se a indicação de que a Nanotecnologia terá um grande impacto sobre quase todos os setores incluindo, energia, transporte, fabricação, medicina, tecnologia da informação e telecomunicações. Sua associação com os *Big Data* é frequente ao passo dessa parceria ser denominada como a próxima revolução digital (RODRIGUEZ, 2015).

A associação entre pesquisas sobre infraestrutura de informação e os respectivos domínios cobertos pela *e-Science*, práticas colaborativas e políticas de fomento são pertinentes. Observa-se neste conjunto que a *e-Science* é considerada a “base para a pesquisa colaborativa e em rede” (IZIQUE, 2013, p. 2), um elemento importante para a realização do “experimento” que se define como fazer científico contemporâneo. O entendimento da construção desse fazer permite um trânsito seguro sobre seus domínios e uma preocupação nos capítulos seguintes.