

O IMPACTO DA CURADORIA DIGITAL DOS DADOS DE PESQUISA NA COMUNICAÇÃO CIENTÍFICA

Luana Farias Salesⁱ
Luís Fernando Sayãoⁱⁱ

Resumo: Oferecer acesso aos dados gerados no decorrer das atividades científicas é um desejo cada vez mais relevante para a comunidade científica. Assim como existe uma tendência mundial em dar acesso livre aos resultados de pesquisa por meio da criação de repositórios digitais e da publicação de periódicos livres, existe também uma demanda em torno do acesso livre aos dados gerados pela pesquisa científica. A importância desses dados é cada vez mais patente, visto que eles que dão sustentação aos resultados que serão discutidos nas publicações acadêmicas tradicionais e podem servir de base para novos projetos. Por outro lado, o uso intensivo de computadores, de tecnologias de rede, instrumentação avançada e de simulação na pesquisa científica proporciona o surgimento de uma ciência inteiramente baseada nos fluxos de dados e de conjunto de objetos digitais armazenados em repositórios distribuídos globalmente. Essa nova forma de gerar e disseminar conhecimento é chamada de eScience. Os dados digitais gerados a partir dessa forma de praticar ciência precisam ser tratados e gerenciados de modo que a preservação por longo prazo, o acesso, a autenticidade e o reuso para novas pesquisas desses dados possam ser assegurados. Este trabalho visa apresentar o conceito de Curadoria Digital enquanto uma nova atividade de tratamento e representação da informação que, desenvolvida de forma correta, poderá alterar o ciclo da comunicação científica.

Palavras-chave: Curadoria digital. Reuso de dados científicos. Comunicação científica.

THE IMPACT OF DIGITAL CURATION OF RESEARCH DATA ON SCIENTIFIC COMMUNICATION

Abstract: *Providing access to data generated in the course of scientific research is an increasingly relevant desire to the academic community. Such as there is a worldwide trend to give free access to research results through the creation of digital repositories and the publication of free journals, there is also a demand regarding the open access to data generated by scientific research. The importance of such data is becoming much more evident, since they give support to the results that will be discussed in the traditional scholar publications and which can serve as a basis for further studies. On the other hand, the intensive use of computers, networking technologies, advanced instrumentation and simulation in scientific research allows the emergence of a science based entirely on data flows and set of digital objects stored in globally distributed repositories. This new way of generating and disseminating knowledge is called eScience. The digital data generated from this way of practicing science need to be treated and managed so that long-term preservation, access, as well as the authenticity and reuse for further research can be assured. This paper aims to present the concept of Digital Curation as a new activity for treatment and representation of information that, properly developed, can change the cycle of scientific communication.*

Keywords: *Digital curation. Reuse of scientific data. Scientific communication*



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/)

ⁱ Universidade Federal do Rio de Janeiro (UFRJ) e Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) – lsales@ien.gov.br

ⁱⁱ Comissão Nacional de Energia Nuclear (CNEN) – lsayao@cnen.gov.br

1 INTRODUÇÃO

Pesquisadores e acadêmicos sempre procuraram incorporar os desenvolvimentos tecnológicos na intermediação e intensificação das suas relações de compartilhamento de conhecimentos. Os pressupostos de inovação e de modernidade da ciência, aliados à própria natureza social inerente à construção do saber científico colocam as tecnologias de informação e comunicação (TICs) numa condição determinante para o estabelecimento de novas modalidades de socialização entre pesquisadores. Iniciando com extensões e simulacros virtuais de formas tradicionais de comunicação como correio eletrônico, telefone e reuniões presenciais, as TICs, nos dias atuais, ultrapassam os limites da comunicação e criam novas formas de fazer ciência, como são os laboratórios³ e os experimentos levados a cabo de forma distribuída em escala universal, baseados em redes de computadores.

O que se observa de forma concreta é que as inovações que impactam e criam novos patamares para a comunicação científica nascem no próprio seio das comunidades acadêmicas. Um exemplo significativo desse fenômeno é o protocolo *Open Archives for Metadata Harvesting*⁴, cujas potencialidades contrastam com a sua surpreendente simplicidade. Esse padrão de compartilhamento, interoperabilidade e reuso de metadados deu margem à criação de mecanismos de encurtamento do ciclo de comunicação científica, muitas vezes excessivamente longo em especial para algumas áreas de rápida evolução como a Física e a própria área de Tecnologia da Informação. Ao mesmo tempo, criou espaços virtuais de memória científica para informações sem lugar nos sistemas mais formais, como bases de dados e OPACS, sendo uma alternativa autêntica aos periódicos científicos.

Os pesquisadores têm lançado luz sobre problemas que estavam despercebidos e que, entretanto, têm desdobramentos significativos na estrutura clássica da comunicação científica. Alguns desses problemas são: a gestão de dados de pesquisa, sua preservação, seu reuso e os processos de agregação de valor, cujas metodologias são coletivamente chamadas de curadoria digital. O presente artigo tem por objetivo apresentar esse novo conceito que surge no âmago dos estudos sobre tratamento e recuperação da informação digital, evidenciando como essa nova tipologia informacional, quando bem gerenciada, pode afetar o ciclo da comunicação científica.

³ Laboratório é um termo cunhado por Wulf (1989) para designar um “centro de pesquisa sem paredes, no qual os pesquisadores de um país podem realizar suas pesquisas sem considerar a localização física, interagindo com colegas, acessando instrumentação, compartilhando dados, informações e recursos computacionais, e acessando bibliotecas digitais”

⁴ www.openarchives.org/OAI/openarchivesprotocol.html

2 A COMUNICAÇÃO CIENTÍFICA

O conceito “comunicação científica” foi estabelecido por John Bernal no final dos anos trinta do século XX para designar o processo específico de produção, consumo e transferência da informação no campo científico.

Segundo Targino (2000), a história dos estudos sobre comunicação científica surge nos EUA, na década de 1940, como decorrência do crescimento significativo e desordenado da literatura científica. De acordo com a autora, os primeiros estudos tiveram como objetivo central analisar os problemas do uso da informação por cientistas e tecnólogos, configurando os chamados estudos de usuários.

O interesse pelos temas comunicação científica e literatura científica ficou ainda mais evidenciado entre as décadas de 1960 e 1970, quando a disputa acirrada entre as duas potências da época - EUA e URSS (antiga União das Repúblicas Socialistas Soviéticas) - visava mostrar a supremacia científica e tecnológica de cada uma delas. Naquela época, ficava claro que não bastava exclusivamente fazer ciência, também era necessário divulgar o resultado das pesquisas realizadas.

A necessidade de comunicar a atividade científica permitia, além de divulgar os resultados e ter o trabalho avaliado pelos pares, “somar os esforços individuais dos membros das comunidades científicas”, pois possibilitava a troca contínua de informações com seus pares, “emitindo-as para seus sucessores e/ou adquirindo-as de seus predecessores” (TARGINO, 2000, p. 10). Assim, a comunicação científica foi se tornando essencial para todos os pesquisadores.

Menzel (1958 apud KAPLAN; STORER 1968) faz algumas colocações que conduzem às funções da comunicação na ciência, a saber: fornecer respostas a perguntas específicas; concorrer para a atualização profissional do cientista no campo específico de sua atuação; estimular a descoberta e a compreensão de novos campos de interesse; divulgar as tendências de áreas emergentes, fornecendo aos cientistas a ideia da relevância de seu trabalho; testar a confiabilidade de novos conhecimentos diante da possibilidade de testemunhos e verificações; redirecionar ou ampliar o rol de interesse dos cientistas; fornecer *feedback* para aperfeiçoamento da produção do pesquisador.

Ziman (1984, p. 84) afirma que “a ciência é conhecimento público, disponível livremente para todos” e, semelhantemente a esse autor, Storer (1966) coloca que “os

conhecimentos que não estão disponíveis ao público não constituem conhecimento público e assim não podem ser referendados pelo mundo científico.” Dessa forma, foi preciso o estabelecimento de um canal de divulgação “para a certificação do conhecimento científico e para a comunicação autorizada da ciência”, dando-lhe também a “atribuição de confirmar a autoria da descoberta científica (MÜELLER, 2006, p. 27). Como desdobramento, a comunidade científica estabeleceu que as revistas indexadas estariam no centro do sistema tradicional de comunicação científica.

Neste momento, porém, com o monopólio das editoras sobre as publicações científicas, os resultados dessas pesquisas ficam cada vez mais limitados a certo número de pesquisadores ligados às instituições que podem pagar para ter acesso à informação. Esse fato impede que o conhecimento científico – um patrimônio da humanidade – circule livremente e esteja disponível para todos.

O problema do monopólio levou a própria comunidade científica a buscar uma solução que se consolidou por meio do movimento *open access*. Esse movimento de alcance global sugere dois caminhos para o acesso livre à informação científica: a via dourada – que incentiva a publicação de periódicos eletrônicos de acesso livre - e a via verde – que incentiva a implantação de repositórios digitais temáticos e institucionais, bem como o autoarquivamento de *e-prints* nesses repositórios.

Nesse sentido, os periódicos de acesso livre e os repositórios institucionais vêm se constituindo uma alternativa viável para que os resultados da pesquisa não pertençam somente ao cientista, e sim à toda humanidade. A questão que este trabalho traz, no entanto, vai um pouco além do compartilhamento dos resultados da pesquisa, pois considera também que o compartilhamento dos dados científicos possa ser mais uma forma de validar as atividades desenvolvidas no âmbito da ciência, além de otimizar a produção de novos conhecimentos e a indução de novas descobertas.

A questão que se coloca, portanto, é: se compartilhamos os resultados, por que não compartilhamos também os dados que levaram a tais resultados? A “corrida em busca da prioridade da descoberta científica implica originalidade, vista como a capacidade de levar a ciência para frente, de explorar suas potencialidades, de criar alternativas, enfim, de garantir a dinamicidade intrínseca à ciência” (TARGINO, 2000, p. 15). Compartilhar dados de pesquisa de forma consciente e responsável pode ser uma nova forma de mover a ciência e explorar ainda mais suas potencialidades.

3 A IMPORTÂNCIA DOS DADOS DE PESQUISA

A necessidade de se ter dados científicos⁵ tratados e disponíveis para o acesso pelos pesquisadores não é uma questão totalmente nova para a Ciência da Informação. No âmbito da pesquisa sobre primatas, o Museu Paraense Emílio Goeldi, em fins da década de 1980 e início de 1990, desenvolveu o PRIMATAM, projeto ligado ao Núcleo de Primatologia, cujo tratamento dos dados de pesquisa resultou em um catálogo impresso, conforme informações dadas por Suely Marques-Aguiar (2011) através de contato via *e-mail*.

Com o surgimento das mídias eletrônicas, da mudança nos fluxos da comunicação científica e da rápida obsolescência tecnológica de seus suportes, a necessidade de criação de métodos para tratamento e recuperação de dados científicos toma uma proporção ainda maior. Na área de genoma, por exemplo, o acesso aberto aos dados de pesquisa já é realidade há algum tempo. Desde a década de 1980, o *International Nucleotide Sequence Database Collaboration* (INSDC) mantém o *Genbank*⁶ – um conjunto de bases de dados que trata informações sobre sequenciamento genômico das mais diversas espécies. Os dados são abertos, mas com certo limite, isto é, apenas pesquisadores que fazem parte do consórcio têm acesso ao compartilhamento desses dados. Atualmente, com a *Web* e todo o advento tecnológico, essa base é alimentada pelos pesquisadores que submetem os dados e fazem "anotações" - sobre o andamento da pesquisa, sobre novas conclusões etc. - usando como base uma ontologia chamada *Gene Ontology*. Isso revela não apenas uma preocupação em reunir os dados, mas também em ter dados tratados de modo que seja possível acessá-los e compartilhá-los de forma precisa.

Mas não é exclusivamente a área de Genoma que vem se preocupando com a questão: há um consenso no seio das comunidades científicas de que o acesso aos dados de pesquisa é um imperativo de âmbito global. Esse fato coloca em pauta um problema novo, que é a gestão de dados de pesquisa num mundo digital interligado por redes de computadores, nas quais há um fluxo intenso de dados sendo gerados, processados e compartilhados. A partir desse ponto, instala-se, então, um desafio importante do nosso tempo, que é ao mesmo tempo uma oportunidade significativa e essencial para se conduzir a pesquisa científica nesse século que se inicia (LANNOM, 2011).

⁵ Dados científicos ou dados de pesquisas, conforme definição da OCDE (2007, p.13) podem ser definidos como “registros de fatos usados como fontes primárias na investigação científica e que geralmente são aceitos na comunidade científica como necessários para a validação dos resultados da pesquisa.”

⁶ <http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide>

A Declaração de Berlim sobre o Acesso Aberto ao Conhecimento em Ciências e Humanidades, publicada em 2003, amplia o escopo do que se entende por acesso livre ao definir que: “as contribuições de acesso livre incluem resultados de pesquisas científicas originais, dados não processados e metadados, fontes originais, representações digitais de materiais pictóricos e gráficos e materiais acadêmicos multimídia” (BERLIM, 2003).

Compreendendo a importância do tema, a D-Lib Magazine⁷ – o periódico mais importante no que tange às pesquisas em bibliotecas digitais – publicou, no início de 2011, um número especial sobre dados de pesquisa. Nessa publicação estão endereçadas questões como acesso livre, curadoria digital, aquisição e gestão, qualidade e confiabilidade e as possíveis conexões entre dados de pesquisa e as publicações acadêmicas tradicionais, que oferecem oportunidades para o surgimento de concepções surpreendentes de documentos, como é, por exemplo, o *enhanced publication*.

Enhanced Publication ou publicações ampliadas - como estamos sugerindo chamar - são instâncias de objetos digitais complexos que combinam vários recursos heterogêneos relacionados para uma mesma finalidade científica. Por exemplo, uma tese que, ao ser incluída em um repositório institucional ou temático, agregue os dados de pesquisas utilizados para que ela pudesse ser gerada.

Essa ideia de publicações ampliadas emerge da compreensão de que as publicações tradicionais são limitadas na sua capacidade de incorporar resultados de todo o ciclo do processo de investigação científica. Isso acontece especialmente quando grandes conjuntos de dados são gerados. No modelo vigente de publicação, fica evidente que os textos acadêmicos só podem apresentar os dados de pesquisa de forma condensada.

Pode-se observar o fato promissor de que, crescentemente, os dados de pesquisa estão sendo armazenados em repositórios confiáveis e, gerenciados sob os princípios da curadoria digital, são preservados e mantêm sua capacidade de reuso. Porém, na presente infraestrutura de comunicação científica, esses conjuntos de dados não são conectados às publicações científicas onde são discutidos. A noção que está por trás dos documentos ampliados é precisamente criar pontes que liguem os conteúdos dos repositórios institucionais, ou seja, publicações científicas tradicionais, com os conteúdos dos repositórios de dados.

Assim, o que pode ser observado com clareza é que a própria prática do fazer ciência é reordenada pela intensificação do uso de redes e de computadores na pesquisa científica. Uma das mais notáveis propriedades dos projetos de *eScience* é o uso sem precedentes de conjuntos

⁷ <http://www.dlib.org/dlib/january11/01contents.html>

de dados digitais distribuídos. Disciplinas como Física das Partículas, Química, Geologia e Arqueologia dependem de forma absoluta do uso de ambientes de rede altamente distribuídos, instrumentos automatizados, técnicas de captura de imagens e de programas de simulação.

Essas tecnologias têm impactado ampla e profundamente a forma como os cientistas podem conduzir e disseminar suas pesquisas (VERHAAR, 2008), delineando novos fluxos e definindo patamares inéditos para a comunicação científica, os quais merecem estudos partindo de muitos olhares.

Todavia, o custo-benefício de se manter o acesso e a capacidade de reuso aos dados de pesquisa é difícil de ser mensurado. O valor de um registro pode estar relacionado à possibilidade da reprodutibilidade de um dado experimento onde ele é gerado ou capturado. Algumas pesquisas podem ser fáceis e baratas de se replicar; outras podem ser literalmente impossíveis de se reproduzir (JANSEN, 2006). Assim, o arquivamento eletrônico de dados começa a ser estimulado ativamente pelas agências de financiamento de pesquisa, que demandam mais e mais que os projetos de pesquisa contemplem a submissão dos dados gerados a repositórios confiáveis.

O Relatório do Projeto *Digital Repository Infrastructure Vision for European Research II* (Driver II), desenvolvido sob os auspícios da Comunidade Europeia, justifica essa preocupação das agências de fomento, enfatizando que o acesso a dados de pesquisa proporciona uma série de vantagens, especialmente quando esses dados estão associados a manuscritos acadêmicos disponíveis *online*. Por exemplo: quando um pesquisador deposita seus dados brutos, ele abre a possibilidade dos seus pares replicá-los e, dessa forma, consegue verificar o que está sendo defendido. Na publicação científica, isso possibilita também que outros pesquisadores reusem os dados, os comparem e os combinem com outros dados, de forma que novas pesquisas possam ser geradas. Outro benefício apontado pelo Relatório é que a curadoria dos dados torna possível traçar a linhagem dos vários produtos dos projetos de *eScience*, dado que esses projetos se desenvolvem por vários estágios, tais como captura de dados, processamento, modelagem e interpretação. Se fosse possível destacar as inúmeras conexões entre os recursos que são produzidos durante os vários estágios do processo científico, isso seria de grande utilidade (VERHAAR, 2008).

No curto período do que se convencionou chamar de era digital, algumas instituições científicas estiveram comprometidas com o desenvolvimento de atividades que pudessem salvaguardar os dados científicos digitais; no entanto, as poucas instituições engajadas nesse processo ainda não estabeleceram práticas e não garantiram os fluxos de recursos que assegurem o completo sucesso da gestão desses dados. O que se observa é que ainda

persistem lacunas críticas e questões de pesquisas em aberto (LEE; TIBBO, 2007; HIGGINS, 2011). Para muitas comunidades acadêmicas, a gestão e o acesso continuado a essa vasta quantidade de dados constituem um grande desafio. Infelizmente, muitos dos dados produzidos, frequentemente a um custo alto, são irremediavelmente perdidos.

Para concluir esta seção, é necessário informar que várias iniciativas importantes, lideradas pelas próprias comunidades científicas, já cumprem papel vital na garantia do acesso livre aos dados de pesquisa e no que se convencionou chamar de curadoria digital. O *Digital Curation Centre* (DCC)⁸ é um exemplo desse tipo de iniciativa que resume e justifica a importância das suas atividades no seu próprio lema: “porque boa pesquisa precisa de bons dados”.

3.1 Curadoria digital de dados científicos

Os conhecimentos e as práticas acumulados na última década em preservação digital e acesso resultaram num conjunto de estratégias, abordagens tecnológicas e atividades que agora são coletivamente conhecidas como “curadoria digital”. Ainda que seja um conceito em evolução, já está estabelecido que a curadoria digital envolve a gestão atuante e a preservação de recursos digitais durante todo o ciclo de vida de interesse do mundo acadêmico nesses dados, tendo como perspectiva o desafio temporal de atender a gerações atuais e futuras de usuários. Pode-se perceber, então, que, subjacente às metodologias utilizadas pela curadoria digital, estão os processos de arquivamento digital e de preservação digital (HIGGINS, 2011; CONWAY, 2011).

O Data Curator Centre (DCC), na sua página *web*, informa que a curadoria digital “envolve a manutenção, a preservação e a agregação de valor a dados de pesquisa durante o seu ciclo de vida”. Registra também que a gestão ativa sobre esses dados reduz as ameaças ao seu valor de longo prazo e minimiza os riscos da obsolescência digital. Além de reduzir a duplicação de esforços na criação de dados de pesquisa, a curadoria reforça o valor de longo prazo dos dados existentes quando os tornam disponíveis para a reutilização em novas pesquisas de qualidade.

Daisy Abbott (2008) amplia um pouco mais a ideia de curadoria digital definindo-a como todas as atividades envolvidas na gestão de dados, desde o planejamento da sua criação – quando os sistemas são projetados – passando pelas boas práticas na digitalização, na

⁸ <http://www.dcc.ac.uk/>

seleção dos formatos, na documentação e na garantia de estarem sempre disponíveis e adequados para serem descobertos e reusados agora e no futuro. A curadoria digital também inclui a gestão de grandes conjuntos de dados para uso diário, assegurando, por exemplo, que eles possam ser pesquisados e que continuem viáveis, ou seja, capazes de serem lidos e interpretados. Nessa perspectiva, a ideia de curadoria digital estende-se para além do controle do repositório que arquiva os recursos e envolve a atenção do criador do conteúdo e dos usuários futuros.

A curadoria digital, em resumo, assegura a sustentabilidade dos dados para o futuro, não deixando, contudo, de conferir valor imediato a eles para os seus criadores e para os seus usuários. Os recursos estratégicos, metodológicos e as tecnologias envolvidas nas práticas da curadoria digital facilitam o acesso persistente a dados digitais confiáveis por meio da melhoria da qualidade desses dados, do seu contexto de pesquisa e da checagem de autenticidade. Dessa forma, a curadoria contribui para assegurar aos dados de pesquisa validade como registro arquivístico, significando que eles podem ser usados no futuro como evidência legal. O uso de padrões comuns entre diferentes conjuntos de dados, proporcionado pela curadoria digital, cria mais oportunidades de buscas transversais e de colaboração. Na ótica financeira, o compartilhamento, o reuso dos dados e as oportunidades de novas análises, além de outros benefícios, valorizam e protegem o investimento inicial para obtenção dos dados.

3.2 Gerenciamento de dados digitais

O DCC oferece um modelo para o ciclo de vida da curadoria que reflete uma visão de alto nível dos estágios necessários para o sucesso do processo de curadoria e de preservação de dados de pesquisa. O modelo proposto pelo DCC - que parte da concepção ou do recebimento dos dados - está orientado para o planejamento das atividades de curadoria nas organizações ou consórcios ajudando a garantir que todos os passos do ciclo serão cumpridos. Entretanto, isso não implica que todas as organizações devam entrar no ciclo no primeiro estágio; na realidade, a operacionalização dos estágios dependerá das áreas de necessidade de cada organização.

Os elementos chaves do modelo são: dados, objetos digitais e bases de dados. No centro do ciclo de vida da curadoria está o dado digital, que é qualquer informação codificada em formato binário. A ideia de dado inclui: os objetos digitais simples, que são aqueles compostos por um único arquivo, identificador e metadados; e os objetos digitais complexos,

que por sua vez são formados pela combinação de outros objetos digitais, formando uma unidade discreta, como é, por exemplo, uma página *web*. As bases de dados são definidas como coleções estruturadas de registros ou de dados armazenados em sistemas de computadores.

As ações que completam a totalidade do ciclo de vida incluem: a descrição e a representação da informação, efetivada pela atribuição de metadados – administrativos, técnicos, estruturais e de representação – de acordo com os padrões apropriados; a definição de um plano de preservação cujo espectro englobe todo o ciclo de vida da curadoria digital; a manutenção do monitoramento sobre as atividades das comunidades envolvidas, bem como participação no desenvolvimento de padrões que possam ser compartilhados, de ferramentas e de *software* adequados ao problema; o alerta contínuo e a efetivação das ações administrativas e gerenciais planejadas para a curadoria e preservação que abranjam todo o ciclo de vida da curadoria.

A sequência de ações do modelo de ciclo de vida da curadoria digital proposto pelo DCC tem os seguintes estágios:

- **Conceituar** – conceber e planejar a criação do dado, incluindo os métodos de captura e as opções de armazenamento.

- **Criar e receber** – criar o dado incluindo o elenco de metadados necessários à sua gestão e compreensão, ou seja, metadados administrativos, descritivos, estruturais e técnicos; os metadados de preservação podem ser também incluídos no momento da criação do dado.

- **Avaliar e selecionar** – avaliar o dado e selecionar o que será objeto dos processos de curadoria e de preservação por longo prazo; manter-se aderente às práticas, às políticas pertinentes e às exigências legais.

- **Capturar** – transferir o dado para um arquivo, repositório, centro de dados ou outro custodiante apropriado.

- **Ação de preservação** – promover ações para assegurar a preservação de longo prazo e a retenção do dado de natureza oficial; as ações de preservação devem assegurar que o dado permaneça autêntico, confiável e capaz de ser usado enquanto mantém sua integridade. Essas ações de preservação incluem: a limpeza do dado e a sua validação, a adição de metadados de preservação; adição de informação de representação e a garantia de estruturas de dados ou formatos de arquivos aceitáveis.

- **Armazenar** – armazenar o dado de forma segura, mantendo a aderência aos padrões relevantes.

- **Acessar, usar e reusar** – assegurar que o dado pode ser cotidianamente acessado tanto pela sua comunidade alvo quanto pelos demais usuários interessados no reuso do dado; isso pode ser realizado na forma de informação publicada disponível publicamente; controle de acesso robusto e procedimento de autenticação podem ser aplicados.

- **Transformar** – criar novos dados a partir do original, por exemplo, pelo processo de migração para diferentes formatos ou pela criação de subconjuntos - realizada por meio de seleção ou formulação de consultas – derivando novos resultados que podem ser publicados.

O DCC estabelece também estágios que são aplicados ocasionalmente:

- **Eliminar** – eliminar o dado que não foi selecionado para curadoria e preservação de longo prazo de acordo com políticas documentadas, diretrizes e exigências legais.

- **Reavaliar** – retornar ao dado cujos procedimentos de avaliação foram falhos para nova avaliação e possível seleção.

- **Migrar** – migrar os dados para um formato diferente; isso pode ser feito no sentido de compatibilizá-lo com o ambiente de armazenamento ou para assegurar a imunidade do dado em relação à obsolescência de hardware e de software.

Assim, como se pode verificar, a disseminação da curadoria digital de dados poderá levar a uma nova alteração do ciclo da comunicação científica.

4 ALTERAÇÃO DO CICLO DA COMUNICAÇÃO CIENTÍFICA

No âmbito da Ciência da Informação, vários modelos de ciclo de comunicação científica já foram explorados. Lancaster(1977), Jordan(1973), King e Bryant(1971) são exemplos de autores que desenvolveram estudos a respeito.

Em geral, esses primeiros modelos visavam representar os processos envolvidos desde a elaboração do documento pelo pesquisador, passando pelas diversas fases de avaliação, publicação, tratamento, disseminação, obtenção até a assimilação e o uso para criação de novas pesquisas e, conseqüentemente, novos documentos. Esse ciclo pode ser visualizado no esquema a seguir (Figura 1), apresentado por Rodrigues (2008) na 3ª Conferência Open Access, na Universidade do Minho.

Acesso Limitado = Impacto Limitado

(Traduzido e adaptado de Harnad)



Figura1: Modelo da Comunicação Científica Tradicional (RODRIGUES, 2008)

Vickery(1999) publicou um modelo que “engloba em seus pressupostos a transferência por meio de canais eletrônicos” (BENCHIMOL, 2009, p.44), fazendo um mapeamento que incluía a transferência da informação científica em meios eletrônicos, abrangendo desde a comunicação informal (*e-mails*, listas de discussão etc) até a comunicação formal (OPACS, periódicos eletrônicos etc.) (VICKERY, 1999; PINHEIRO, 2003).

Não havia dúvidas de que a tecnologia e os recursos eletrônicos afetariam o ciclo da comunicação científica, principalmente em termos de velocidade e acesso às informações, mas algumas fases do processo encontraram uma barreira: a questão do embargo ao acesso aos documentos feito pelas grandes editoras. Em busca de uma solução para o problema, as iniciativas que incentivavam a autosubmissão de *e-prints*, bem como a publicação em periódicos de acesso livre possibilitaram uma primeira alteração no ciclo tradicional da comunicação científica.

No esquema apresentado a seguir (Figura 2), também apresentado por Rodrigues (2008) na mesma conferência pode-se observar o encurtamento no tempo de acesso ao resultado da pesquisa. Além disso, a disseminação desse resultado se intensifica, já que um número maior de pessoas tem acesso à publicação.

Impacto e acesso à investigação maximizado pelo “auto-arquivo”

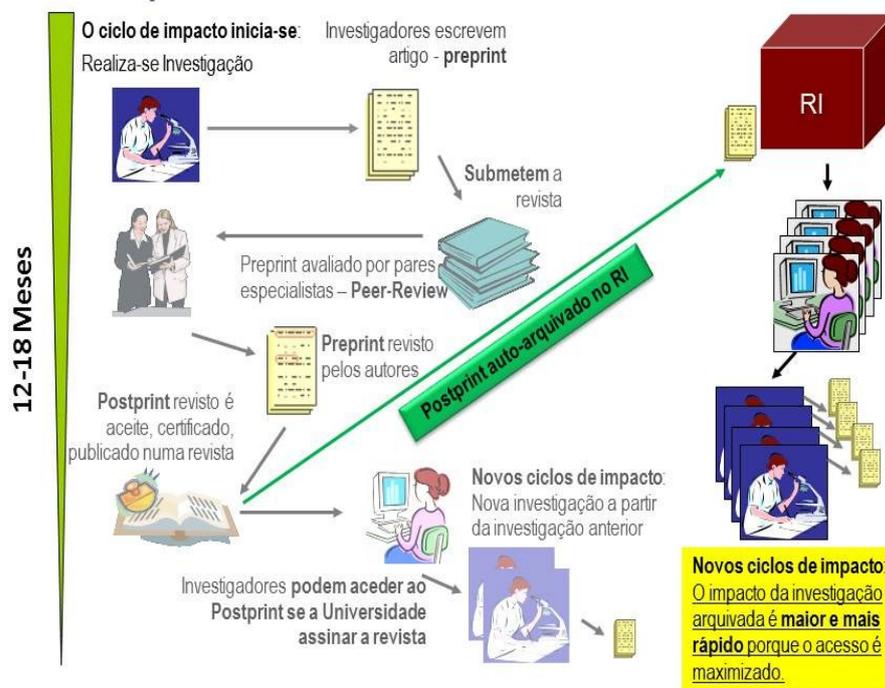


Figura 2: Modelo tradicional alterado pelo Acesso Aberto (RODRIGUES, 2008)

Atualmente, em um novo ambiente de pesquisa, quase que inteiramente baseado em redes, computadores e processamento distribuído de conjunto de dados, a inclusão de uma nova etapa chamada curadoria digital incorpora-se como uma peça chave no ciclo tradicional de comunicação científica baseado em artigos de periódicos. A curadoria digital abre a possibilidade de mudança no ciclo da comunicação científica, adicionando uma nova dinâmica, na medida em que estabelece metodologias de gestão e de preservação de dados de pesquisa com o foco no reuso dos dados disponíveis em meio digital.

A situação vigente, todavia, é a seguinte: quando, por exemplo, um estudante de doutorado conclui a sua pesquisa e esta é registrada na forma de um documento que conhecemos por tese, temos aí somente um retrato parcial dos conteúdos intelectuais gerados no desenrolar de anos de trabalho. Via de regra, os dados de pesquisa que dão sustentação à tese adormecerão armazenados em computadores e mídias pessoais que inexoravelmente serão tragados pela obsolescência tecnológica, da mesma forma como muitos dados foram simplesmente “jogados fora” quando ainda existiam somente na versão impressa.

Esses dados perdidos custaram, muitas das vezes, anos de investimentos financeiros e pessoais e constituem uma parte essencial do trabalho de pesquisa. Quem perde com isso é

geralmente o próprio pesquisador, que não pode reutilizar seus próprios dados e, é claro, a sociedade, que tem que arcar com o ônus da duplicação de esforços e com a desaceleração do desenvolvimento científico causado pela necessidade de se replicar experimentos científicos já desenvolvidos anteriormente.

Quando o compartilhamento de dados científicos realiza-se por meio das técnicas de curadoria digital, passa a fazer parte do ciclo da comunicação científica, tem lugar então uma nova alteração no fluxo da comunicação. Uma relação inédita se estabelece entre pesquisadores, na medida em que um pesquisador deposita toda a confiança nos dados levantados por outro pesquisador e os têm como patamar para o desenvolvimento de seus novos projetos.

Dessa forma, a curadoria digital reduz o ciclo da comunicação científica quando oferece aos pesquisadores dados de pesquisa prontos e confiáveis para o reuso. Isso significa dizer que a curadoria disponibiliza dados tratados, acompanhados por metadados semânticos e estruturais – que assegurariam o seu significado e a reconstrução de sua apresentação correta - e metadados de preservação - que mantêm sua integridade, precisão e autenticidade desses dados.

O esquema abaixo (Figura 3) pode ilustrar essa nova perspectiva:

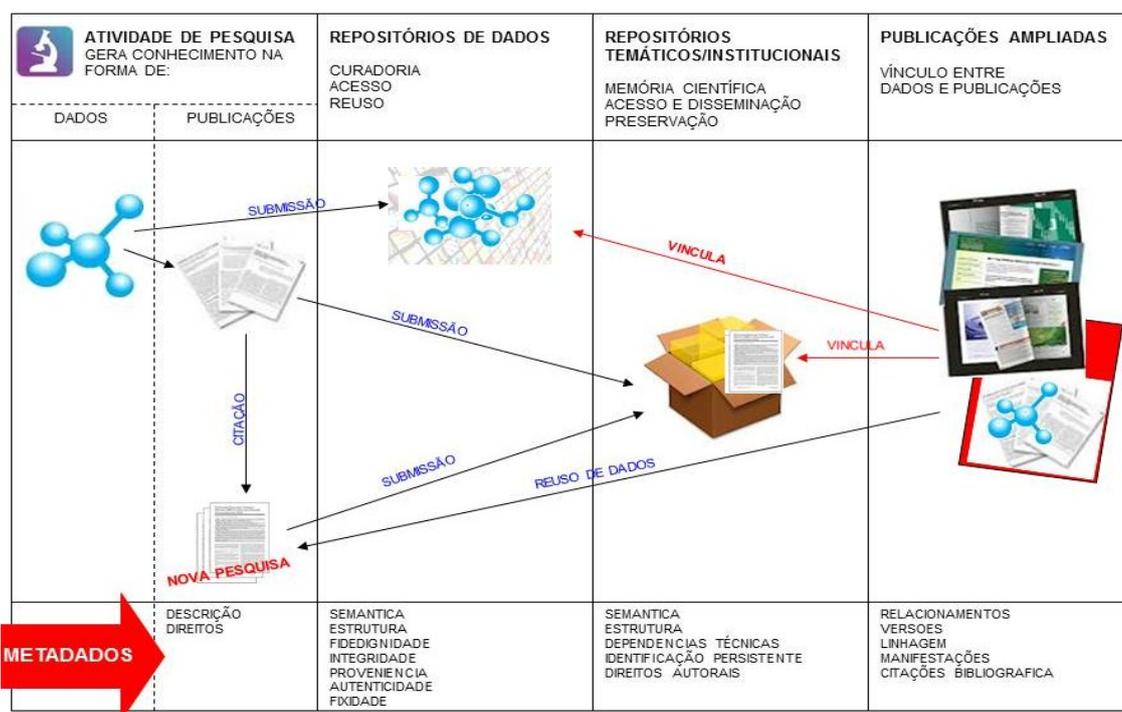


Figura 3: Modelo da comunicação científica tradicional alterado pelo reuso de dados científicos

Assim, o terceiro modelo pode ser explicado da seguinte forma: primeiramente, um pesquisador gera dados durante o desenvolvimento de sua pesquisa; em seguida, esses dados são depositados em um Repositório 1, que foi chamado no esquema de Repositório de Dados representado na Figura 3; esses dados são mantidos em acesso fechado por um tempo. Simultaneamente, o pesquisador submete os resultados de sua pesquisa em forma de artigo a um periódico científico que, quando aprovado pelos pares, terão também os dados avaliados, como no ciclo tradicional da comunicação científica. Posteriormente, com artigo avaliado e aprovado, a terceira etapa é a da submissão do *pos-print* em um Repositório 2 (Institucional ou Temático), que interoperando com o Repositório 1 gerará um terceiro tipo de documento chamado “documento ampliado”, ou seja, exatamente a soma do documento avaliado com os dados que o geraram. Nesse momento o acesso é totalmente aberto, permitindo que um segundo pesquisador, ao acessar o documento, tenha acesso também aos dados produzidos, podendo reutilizá-los, encurtando, assim, o tempo gasto para reproduzir tal estudo, dar continuidade ao mesmo com outro enfoque, além de verificar a veracidade dos resultados publicados. Como já mencionado, o encurtamento do ciclo agora se encontra a partir do reuso dos dados científicos para geração de novos resultados.

Vale ressaltar que a alteração que se dá no novo ciclo da comunicação científica não está apenas no âmbito do encurtamento do mesmo, mas também na esfera da qualidade da informação acessada, tendo em vista que a partir da proposta da curadoria digital pode-se ter acesso a dados científicos que antes não eram disseminados, impossibilitando a replicação dos procedimentos que levaram ao resultado de determinada pesquisa.

5 CONSIDERAÇÕES FINAIS

O presente artigo veio apresentar um conceito emergente no mundo da pesquisa científica que está relacionado ao tratamento e preservação dos dados científicos e que pode provocar impactos importantes no ciclo da comunicação científica: a curadoria digital. As teorias e as práticas de curadoria digital tornam-se essenciais no contexto da ciência baseada em dados, que é para onde se deslocam rapidamente os padrões atuais de pesquisa.

Em um novo ambiente de pesquisa, movido pelo uso intensivo de dados gerados pelo avanço vertiginoso dos instrumentos científicos ou por sistemas de simulação, um novo paradigma para a ciência, a *eScience*, vai se delineando. Nesse contexto altamente

virtualizado, quase que inteiramente baseado em redes, computadores e processamento distribuído, o ciclo de vida da curadoria digital integra-se como uma peça essencial no ciclo tradicional de comunicação científica.

Assim como se debate a questão do acesso livre aos periódicos acadêmicos, criando-se novos padrões de comunicação científica - mais ágeis, mais dinâmicos e organicamente mais próximos das comunidades científicas - atualmente fica claro que é preciso estender o movimento de livre acesso também aos dados científicos. Esses recursos constituem uma fatia importante do estoque de conhecimento acumulado pelo trabalho de pesquisa e fazem parte da memória científica.

Questões como: quem serão os responsáveis pela curadoria desses dados? Que tipos de pesquisas deverão conceder os dados? Como adquirir esses dados? Quem poderá ter acesso? Qual o nível de qualidade e confiabilidade? Ou qual o nível de transparência? São questões que merecem ser discutidas amplamente.

No âmbito da área de Representação e Tratamento da Informação, questões relacionadas ao estabelecimento de padrões de metadados, bem como o controle de vocabulário também merecem ser estudadas e esclarecidas com mais detalhes. Devem-se incluir, também, tópicos ligados aos padrões de preservação e acesso aos dados digitais. Enfim, há uma porta aberta para uma gama de estudos sobre esse novo desafio que se coloca para a Ciência da Informação.

REFERÊNCIAS

ABBOT, Daisy. **What is digital curation?**. Digital Curation Center, 2008. Disponível em: <http://www.era.lib.ed.ac.uk/bitstream/1842/3362/3/Abbott%20What%20is%20digital%20curation_%20_%20Digital%20Curation%20Centre.doc> Acesso em: 20 dez. 2011.

BENCHIMOL, Alegria Célia. **Informação o objeto etnográfico**: percurso interdisciplinar no museu paraense Emílio Goeldi. Dissertação (Mestrado em Ciência da Informação). IBICT/UFF. Niterói, RJ. 2009.

BERLIN Declaration on open access to knowledge in the sciences and humanities. Berlin, 2003. Disponível em: <http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf> Acesso em: 20 dez. 2011

CONWAY, Esther et al. Curating scientific research data for the long term: a preservation analysis method in context. **The International Journal of Digital Curation**, n. 2, v. 6, 2011.

HIGGINS, Sarah. Digital curation: the emergence of a new discipline. **The International Journal of Digital Curation**, v. 6, n. 2, 2011. Disponível em: <<http://www.ijdc.net/index.php/ijdc/article/view/184>>. Acesso em: 20 dez. 2011.

JANSEN, Hans. Permanent access to electronic journals. **Information Services & Use**, v. 26, 2006. Disponível em: <<http://iospress.metapress.com/content/7drby91r8t4gf8ap/fulltext.pdf>> Acesso em: 10 nov. 2010.

JORDAN, Michael P. Expanding the invisible college. In: Annual Meeting Los Angeles, 36, 1973. Los Angeles. **Anais...**, Los Angeles, 1973.

KAPLAN, Norman; STORER, Norman W. Scientific communication. In: SILLS, David L. **International Encyclopedia of the Social Sciences**, New York: The Macmillan Co & The Free Press, 1968. v. 13. p. 112-117.

KING, Donald W; BRYANT, Edward C. **The evaluation of information services and products**. Washington: Information resources, 1971.

LANCASTER, F.W. **The measurement and evaluation of the library services**. Airlington: Information Resources, 1977.

LANNOM, Laurence. Research Data. **D-Lib Magazine**, v. 17, n. 1/2, Jan. / Feb. 2011. Disponível em: < <http://www.dlib.org/dlib/january11/01editorial.html>>. Acesso em: 20 dez. 2011.

LEE, Cristopher; TIBBO, Helen. Digital curation and trusted repositories: steps toward success. **Journal of Digital Information**, v. 8, n. 2, 2007. Disponível em: <<http://journals.tdl.org/jodi/article/viewArticle/229/183>> Acesso em: 20 dez. 2011.

MARQUES-AGUIAR, Sueli. Bases de dados científicos de primatas. [mensagem pessoal] Mensagem recebida por: <lfs@gmail.com>. em: 05 dez. 2011.

MUELLER, Suzana Pinheiro Machado. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ci. Inf.**, Brasília, v. 35, n. 2, 2006.

OCDE. **Principles and guidelines for access to research data from public data**. 2007. Disponível em: <<http://www.oecd.org/dataoecd/9/61/38500813.pdf>> Acesso em: 17 fev. 2012.

PINHEIRO, Lena Vânia. Comunidades científicas e infra-estrutura tecnológica no Brasil para uso de recursos eletrônicos de comunicação e informação na pesquisa. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 62-73, 2003.

RODRIGUES, Eloi. As universidades e o open access: apresentação para dirigentes universitários. In: CONFERÊNCIA OPEN ACCESS, 3, 2008 Braga. **Apresentação...** Braga: Universidade do Minho: 15 e 16 dez. 2008. Disponível em: <<http://confoa08.sdum.uminho.pt/apresentacoes/Apresentação%20RCAAP2.pdf>> Acesso em: 21 fev 2011.

STORER, Norman. W. **The social system of science**. New York: Holt, Hinehardt and Winston, 1966. 180 p.

TARGINO, Maria da Graça. Comunicação científica: uma revisão dos elementos básicos. **Inf. & Soc.** João Pessoa, v.10, n.2, p.37-85, 2000.

VERHAAR, Peter. **Report on object models and functionalities**. DRIVER, 2008. Disponível em: <https://openaccess.leidenuniv.nl/bitstream/handle/1887/16018/Report_on_Object_Models_and_Functionalities.pdf?sequence=2> Acesso em: 20 dez. 2011.

VICKERY, Brian. A century of scientific and technical information. **Journal of Documentation**, v. 55, n. 5, p. 476-527, Dec. 1999.

WULF, W. The national collaboratory. In: _____. **Towards a national collaborator**: Unpublished report of a National Science Foundation invitational workshop, Rockefeller University, New York. 1989.

ZIMAN, F. **An introduction to science studies**: the philosophical and social aspects of science and technology. Cambridge: Cambridge University, 1984. 203 p.