

# A GRANDE E A PEQUENA CIÊNCIA: análise das diferenças na gestão de dados de pesquisa

## THE BIG AND SMALL SCIENCE: analysis of differences in research data management

Luana Farias Sales<sup>1</sup>  
Luís Fernando Sayão<sup>2</sup>

### RESUMO

A curadoria de dados de dados de pesquisa é geralmente colocada de forma uniforme para toda a ciência. Os perfis que caracterizam a grande e a pequena ciência são frequentemente desconsiderados ou minimizados no planejamento e desenvolvimento de e-infraestrutura de pesquisa e, mais especificamente, de plataformas de gestão de dados. Enquanto na grande ciência há uma uniformização na geração dos dados, nos processos de curadoria e uma demanda imediata para o compartilhamento e grandes investimentos em e-infraestruturas; na pequena ciência, os dados são gerados/coletados por pequenas equipes numa infinidade de laboratórios pertencentes a vários domínios disciplinares, são extremamente heterogêneos e raramente arquivados para o compartilhamento e reuso. Esses dados necessitam de infraestruturas específicas que considerem seus fluxos de geração, metodologias, culturas de compartilhamento e de esquemas de recompensa e sustentabilidade. O presente estudo tem como objetivo analisar essas diferenças tendo como perspectiva o planejamento de plataformas de gestão que possibilite uma maior visibilidade dos produtos de pesquisa da pequena ciência e a integração desses dois universos da pesquisa científica em torno da *eScience*.

**Palavras-chave:** Gestão de dados de pesquisa. Infraestrutura de pesquisa. Pequena ciência. Grande ciência.

### ABSTRACT

The curation of research data is usually uniformly designed for all science. Profiles that characterize the big and small science are often disregarded or minimized in the planning and development of research e-infrastructure and, more specifically, in data management frameworks. While in the big science there is a homogeneity in the generation of data, in the processes of curation and an immediate demand for sharing and great investments in e-infrastructures; in small science the data are extremely heterogeneous, generated/collected by small teams in a multitude of laboratories belonging to various disciplinary domains, and rarely archived for sharing and reuse. These data need specific platforms that consider their generation flows, methodologies, sharing cultures, and reward schemes and sustainability. The present study aims to analyse these differences considering the perspective of the planning of management platforms that allows a greater visibility of small science research products and the integration of these two universes of scientific research in the context of eScience.

**Keywords:** Research data management. Research infrastructure. Big science. Small Science.

*Artigo recebido em 21/06/2019 e aceito para publicação em 28/08/2019.*

- 1 Doutora em Ciência da Informação pelo Programa de Pós-Graduação do convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro, Brasil. Docente permanente do Programa de Pós-Graduação em Ciência da Informação do convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro, Brasil. Coordenadora da Rede de Implementação do GOFAIR Brasil. E-mail: luanasales@ibict.br.
- 2 Doutor em Ciência da Informação pelo Programa de Pós-Graduação do convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro, Brasil. Tecnologista Sênior desde 1980 na Comissão Nacional de Energia Nuclear, Brasil. Docente permanente do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro, Brasil. E-mail: Isayao@cnen.gov.br.

## 1 INTRODUÇÃO

O quarto paradigma científico é descrito na literatura como a reconfiguração dos paradigmas anteriores – ciência experimental, ciência teórica e ciência baseada em simulação - em torno de um mundo rico em dados. São muitas as expectativas que cercam essa nova forma de fazer ciência, dada a possibilidade que a mesma oferece de *insights* revolucionários que vão dos fenômenos relacionados às mudanças climáticas e às descobertas de novas drogas até as metodologias que possibilitam examinar novos ângulos da história e da cultura (BORGMAN, 2012).

Esse novo modelo de ciência aparece mais claramente nos domínios disciplinares como Astronomia, Física e Genômica; nos campos de estudos híbridos que vão surgindo como Astroinformática e Bioinformática e em algumas áreas das Humanidades Digitais. Essas disciplinas têm como características marcantes a utilização de dispositivos e instrumentos de pesquisa complexos e colossais, orçamentos vultosos, escopo nacional e internacional e longa duração dos seus programas, que envolvem grandes equipes de pesquisadores e laboratórios distribuídos em escala planetária, com uma profunda expertise dos seus respectivos domínios disciplinares. Esse segmento da ciência gera e se utiliza de quantidades massivas de dados, num plano convergente ao fenômeno contemporâneo do *big data*.

Nesse contexto de transição, a atenção das agências de fomento, dos formuladores de políticas científicas, das instituições de pesquisa e da sociedade em geral se volta para os segmentos da ciência que baseiam suas metodologias no uso intensivo de dados. Como desdobramento deste desequilíbrio, uma grande quantidade de pequenas coleções de dados geradas na infinidade de laboratórios das universidades e institutos de pesquisa sai do horizonte desses *stakeholders*, provocando uma opacidade em torno dos produtos de pesquisa gerados pelas atividades científicas da chamada pequena ciência (*small science*).

Não obstante a invisibilidade desse segmento da ciência, a longo prazo e de forma surpreendente, seus pesquisadores, que geram e coletam muitas e diferentes formas de dados de alto valor informacional, em todas as áreas do conhecimento, vão coletivamente produzir mais dados do que os pesquisadores das disciplinas da grande ciência, caracterizando o genuíno *big data* científico. Na pequena ciência, por sua autonomia e potencial disruptivo, é onde ocorrem as inovações, as pesquisas interdisciplinares, as bases para a maioria dos artigos científicos e a conexão com os setores produtivos. Porém, grande parte desse estoque informacional não está disponível para o acesso, compartilhamento e reuso, pois, via de regra, permanecem armazenados nos computadores pessoais dos seus criadores, sem nenhuma gestão que os tornem visíveis.

De forma diferente de domínios disciplinares da grande ciência que incorporam nos seus fluxos práticas de disseminação de dados padronizadas, o compartilhamento de dados na pequena ciência não é algo natural e tem muitas dependências que vão de percepções e interesses pessoais até a ausência de e-infraestruturas adequadas, passando pelos obstáculos representados pelos esquemas convencionais de crédito e de recompensa que não valorizam a organização e publicação de dados de pesquisa.

As diferenças nos padrões de gestão, curadoria, compartilhamento e reuso dos dados produzidos por esses dois universos da pesquisa científica atual demandam infraestruturas bem distintas. De uma forma geral, apoiar a pequena ciência requer que as plataformas de dados ofereçam gestão voltada para as especificidades dos fluxos de pesquisa e de geração de dados e da cultura de compartilhamento de cada uma das áreas do complexo mosaico de disciplinas e de comunidades científicas que formam esse segmento da ciência.

Devido às idiosincrasias que caracterizam a grande e a pequena ciência, se identifica uma necessidade determinante de um alinhamento desta ao movimento crescente e global da *eScience*, e de uma integração às suas infraestruturas de gestão e curadoria na direção de um ambiente caracterizado pela diversidade de dados. Considerando essa questão, o avanço no desenvolvimento em e-infraestrutura de pesquisa é dependente da nossa compreensão de como apoiar práticas e necessidades em relação aos dados (CRAGIN *et al.*, 2010) e às suas especificidades.

O problema que se apresenta é que a gestão de dados é geralmente colocada de forma uniforme para toda a ciência e que os perfis que caracterizam a pequena e a grande ciência são geralmente desconsiderados ou minimizados no planejamento e no desenvolvimento das e-infraestruturas de pesquisa. Existem diferenças contundentes que se iniciam nas formas de geração dos dados até o seu arquivamento, na preservação e nas formas de compartilhamento. As metodologias, instrumentos, fluxos de pesquisa, homogeneidade dos dados gerados, volume de investimentos e reconhecimento e recompensa aos pesquisadores condicionam as formas de visibilidade/invisibilidade dos dados gerados pela pequena e grande ciência.

No momento em que as instituições de pesquisa no país começam a desenvolver repositórios institucionais multidisciplinares e com uma frágil organicidade com suas respectivas comunidades, como contribuição a estes empreendimentos, este estudo se propõe a analisar as principais diferenças em termos de homogeneidade, gestão, curadoria, compartilhamento e institucionalização dos dados de pesquisa gerados nos processos científicos da grande e da pequena ciência. Para tal, tomamos como ponto de partida os pressupostos de que diferentes formas de gestão definem os níveis de permanência de longo prazo,

visibilidade, compartilhamento e reuso das coleções de dados; e que peculiaridades da pequena ciência e de suas comunidades devem ser levadas em consideração quando do planejamento e implementação de e-infraestruturas internacionais, nacionais e locais voltadas para as atividades científicas.

## 2 PEQUENA E GRANDE CIÊNCIA: contradições e tensões

Vivemos numa espécie de presente contínuo, sem uma relação orgânica com o passado (HOBBSAWN, 1995) e sem referências para compreender com mais propriedade o presente. Por mais contemporâneo e atual que seja o problema, a sua semente pode estar no passado: é o caso das tensões entre a *big* e a *small science*. As contradições e encontros entre esses universos de prática científica remontam a séculos. Os contornos dessa diferenciação, porém, ficam mais nítidos com o advento dos grandes aparatos científicos apoiados por alta tecnologia como telescópios espaciais e supercolisores de partículas. Nesse contexto, cuja grandiloquência é um componente inevitável para a evolução da ciência, conforme argumentam os defensores da *big science*, os orçamentos dos projetos e programas se iniciam na casa das centenas de milhares de dólares e chegam, ao longo do tempo, ao patamar de bilhões de dólares. O que se observa, porém, é que este volume de investimentos drena os recursos das pesquisas de pequena escala, que são vitais para a criação de novos produtos e empregos e, muito frequentemente, para o avanço da ciência como um todo. Os compromissos com a *big science* paralisam uma parte importante da ciência, contra-argumentam os pesquisadores que defendem o valor da pequena ciência (BROAD, 1990).

Em 1976, Derek de Solla Price mencionava que os investimentos nacionais e internacionais em recursos humanos e financeiros fizeram com que a ciência, repentinamente, representasse um importante segmento das economias dos países que protagonizam essa nova configuração científica. “O aspecto de grande escala da ciência moderna – nova, brilhante e todo-poderosa – é tão evidente que foi criado, para descrevê-la, o expressivo termo ‘Grande Ciência’” (PRICE, 1976, p.2.). Nos dias de hoje, com maior ímpeto, os investimentos e as atenções dos órgãos formuladores das políticas científicas se voltam para esses segmentos da ciência contemporânea. Essa nova configuração colossal que se instala no cenário científico, tem sido comparada às pirâmides do Egito e às grandes catedrais da Europa Medieval (WEINBERG, 1961).

Apesar da contemporaneidade dessa ciência exceder em termos de volume e descobertas às fases anteriores a ponto de estabelecer uma nova era científica, suas origens remontam a séculos, e o seu poder avassalador desperta temores.

A Grande Ciência é tão recente que muitos de nós não recordamos de suas origens. A grande ciência é tão vasta que começamos a nos preocupar com o tamanho do monstro que criamos. A grande ciência é tão diferente da anterior, que nos lembramos, talvez com nostalgia da “Pequena Ciência” que constituiu no passado nossa maneira de viver (PRICE, 1976, p.2)

Neste cenário contencioso há, porém, um certo consenso entre os especialistas quando concluem que os grandes projetos alteram fundamentalmente a ciência contemporânea e que há uma inevitabilidade nas suas ações para viabilizar novas descobertas. Por trás da tendência avassaladora por grandes instrumentos há um fato que os cientistas raramente mencionam: “a maioria das descobertas que podem ser realizadas com instrumentos simples e pouco trabalho já foram feitas, dessa forma, as ferramentas científicas precisam ser grandes e poderosas” (BROAD, 1990, p.2).

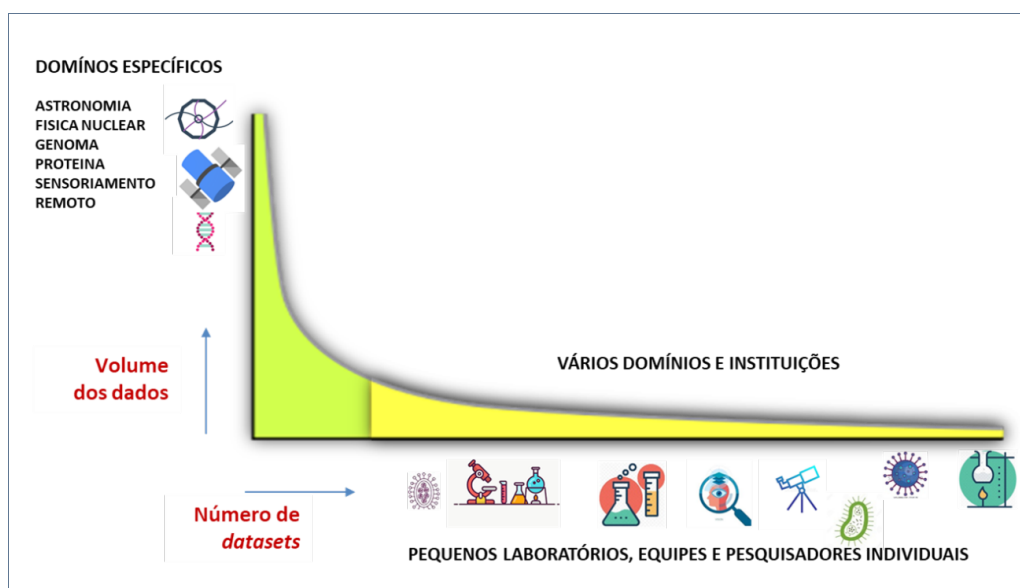
A afirmação de PETSKO (2009, p.1282) de que “nós precisamos da grande ciência, mas o melhor tipo de grande ciência é aquele que apoia e gera um grande número de bons projetos da pequena ciência” está mais próxima da ideia de integração entre esses dois mundos da pesquisa científica. Este fato é enfatizado por uma ciência rica em dados, cujas diversidade e integração transversal e interdisciplinar podem multiplicar a geração de novos conhecimentos. Entretanto, as gestões de dados da pequena e da grande ciência são significativamente diferentes. É o que vemos no decorrer deste ensaio.

### **3 A METÁFORA DA CAUDA LONGA DA PESQUISA**

As análises da distribuição da geração de dados pelos segmentos científicos da grande e da pequena ciência são comumente baseadas no conceito estatístico da “cauda longa” (*long tail*), muito utilizado nas áreas de marketing, especialmente de produtos digitais. O que caracteriza essa forma específica de distribuição estatística é que uma pequena parte da população tem muitas ocorrências – a chamada “cabeça”; enquanto uma grande parte tem poucas ocorrências – conhecida como a “cauda longa” da distribuição.

Aplicada à geração de dados da pesquisa, essa distribuição estatística está representada na Figura 1. Na cabeça identificam-se os grandes projetos e programas mais próximos dos padrões da *big science*, com relativamente poucos empreendimentos mas que, no entanto, geram individualmente um volume gigantesco de dados e consomem muitos recursos. Por sua vez, na cauda longa está identificada uma grande quantidade de projetos liderados por cientistas individualmente ou por pequenas equipes e por laboratórios independentes que desenvolvem um grande número de projetos científicos em várias e diferentes universidades e institutos de pesquisa, e em diversos domínios disciplinares. Estes projetos isoladamente geram poucos dados, mas considerados coletivamente produzem um número imenso de coleções que supera em muito as geradas pela cabeça.

**Figura 1** - Distribuição da geração de dados de pesquisa



Fonte: elaborada pelos autores

A característica mais importante dessas pequenas coleções de dados – tipicamente com menos de 1 GB de volume – é a diversidade, que é ao mesmo tempo sua riqueza informacional e um obstáculo relevante para a sua gestão. Trata-se de dados heterogêneos em termos de formatos, tecnologias e tipos; gerados segundo padrões, processos e fluxos específicos que não estão mais amplamente regulados. Conforme sintetiza MacColl (2010) “A pequena ciência é terrivelmente heterogênea e muito mais vasta e gera 2 a 3 vezes mais dados do que a *big science*”. Ela é constituída de dados, usuários, questões de pesquisa e, sobretudo, de metodologias que diferem fundamentalmente das que compõem as bases da grande ciência. Enquanto nos espaços da *big science* as pesquisas dependem de um número relativamente pequeno de grandes coleções de dados que são bem conhecidas e estão arquivadas e preservadas em grandes bancos ou centros de dados; no outro extremo, os pesquisadores da cauda longa produzem coletivamente uma massiva quantidade de dados que está fragmentada e mantida em milhões de arquivos de computador, que são difíceis de descobrir e usar (HEDSTROM; MYERS, 2014).

Não obstante o fato da pesquisa da cauda longa ter suas origens em projetos de pequenas dimensões, as questões científicas que ela trata não refletem uma ciência menos importante ou mesmo uma ciência de menor amplitude, nem os dados que por ela são gerados podem ser desconsiderados num contexto em que a diversidade e a integração de dados são a chave para respostas a novas indagações da ciência. Enquanto as questões a serem respondidas pelos grandes projetos geradores de dados são relativamente bem compreendidas, a dinâmica de produção de conhecimento da cauda

longa se desenrola num ambiente independente e fértil para o desenvolvimento de novas ideias e para a concepção de uma ciência inédita, inesperada e nunca antes tentada. “Parece mais provável que ciência transformadora venha mais da cauda do que da cabeça” (HEIDORN, 2008, p. 282).

Quando se analisa a produtividade acadêmica, pode-se constatar que uma parte expressiva dos produtos de pesquisa publicada em artigos científicos, em forma de dados terciários e de alta densidade, são dados da cauda longa. Esses produtos de pesquisa ainda se manifestam em outros artefatos como patentes, modelos, multimídias, *software*, visualizações e muito mais. “De certa forma, eles [os dados de pesquisa da cauda longa] são peças de um lego que, se forem colocados juntos corretamente, têm o potencial de gerar novos conhecimentos”, ilustram Wyborn e Lehnert (2016, p.1).

É preciso considerar ainda que a disponibilidade e o acesso aos dados da cauda longa é um elo essencial para o princípio da autocorreção da ciência, para a transparência dos fluxos de geração de conhecimento e para verificar, revisar, validar as pesquisas e atenuar os conhecidos vieses das publicações acadêmicas, conforme concluem Ferguson e seus colaboradores (2014). A valorização dos produtos de pesquisa da cauda longa tem como catalisador principal o movimento mundial do acesso aberto e da ciência cidadã. Nessas instâncias se reconhece as potencialidades das coleções de dados produzidas e coletadas pelas universidades “por sua cobertura, seu ciclo de vida, as observações e as variáveis que são únicos” (HEDSTROM; MYERS, 2014, p.1).

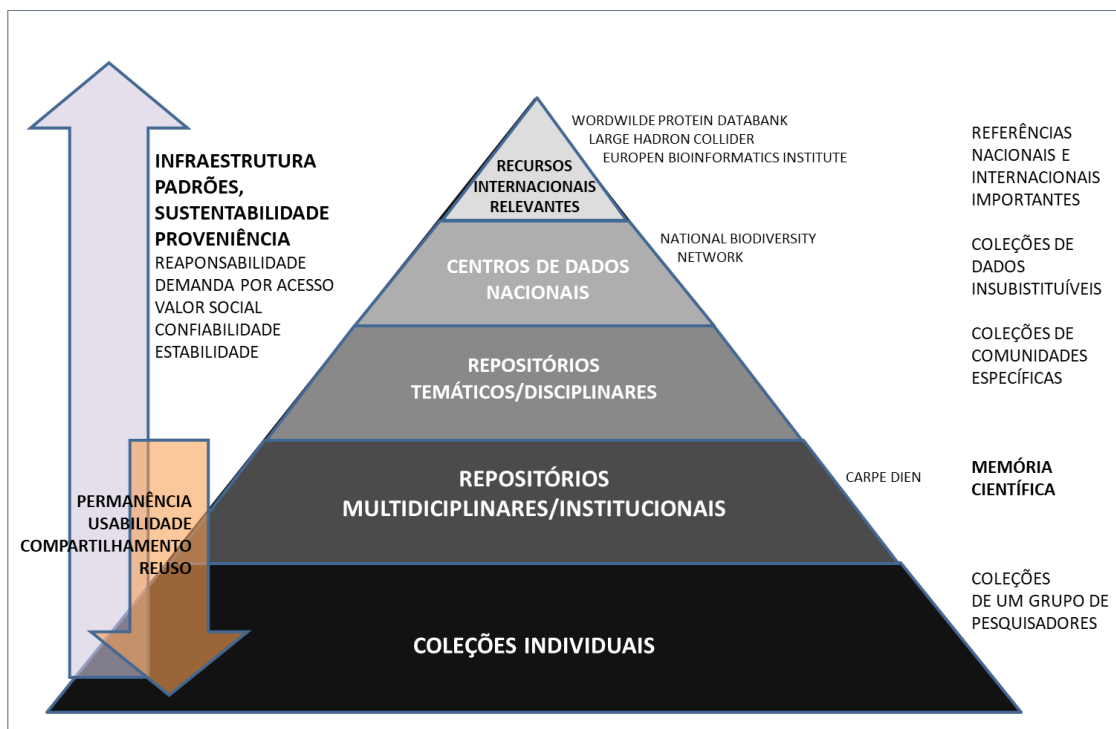
Porém, para explorar o potencial informacional dos dados da cauda longa é preciso compreender as suas distinções em relação aos dados que são gerados pela grande ciência. É isso que analisaremos a seguir.

#### **4 AS ASSIMETRIAS NA GESTÃO DE DADOS**

Nos ambientes científicos há um consenso claro de que o compartilhamento de dados pode ser complexo e custoso e precisa ser reconfigurado segundo estimativas realísticas de demanda e de potencial de reuso que justifiquem os investimentos em curadoria e preservação. As condições de demanda conduzem a diferentes níveis de profundidade de curadoria. Partindo desse pressuposto, o relatório da The Royal Society (2012) apresenta uma representação em forma de pirâmide, em quatro camadas de atividades, que explicita os padrões de gestão de dados de pesquisa e sua relação com as demandas sobre eles. “As camadas de atividades refletem a escala, o custo e o alcance internacional dos dados gerenciados e, em algum grau, a percepção de sua importância” (THE ROYAL SOCIETY, 2012, p.60). Considerando as peculiaridades da pequena ciência, em que seu modelo de gestão deve ser conduzido de forma distinta, acrescentamos à

pirâmide da Royal Society uma nova camada incluindo repositórios temáticos e plataformas disciplinares, no intuito de mostrar as diferenças existentes entre o modelo de gestão em que se apoiam os repositórios institucionais e as plataformas multidisciplinares e os modelos em que se apoiam os repositórios temáticos e plataformas disciplinares. A pirâmide (Figura 2) é, dessa forma, um modelo que representa uma hierarquia de gestão de dados de pesquisa, conforme chamado pelo relatório.

**Figura 2 - Pirâmide de Gestão de Dados de Pesquisa**



Fonte: elaborada pelos autores, baseada em The Royal Society (2012)

No modelo proposto pela The Royal Society, cada uma das camadas exige diferentes aportes financeiros e infraestruturais; apresenta também a profundidade de gestão e curadoria, a estabilidade e a persistência temporal das coleções de dados. No alto da pirâmide estão as coleções massivas de dados gerados pelos grandes programas internacionais de pesquisa como [The Large Hadron Collider](#) e [Worldwilde Protein Data Bank](#). Estas coleções de dados são caracteristicamente padronizadas e relativamente bem curadas por estruturas sofisticadas de gestão e de preservação de dados (WYBORN; LEHNERT, 2016), que podem ser comunitárias, nacionais e internacionais. Esses arcabouços disseminam suas coleções de formas distintas, por exemplo: no *Large Collider* a distribuição está baseada numa complexa rede mundial de distribuição; enquanto o *Protein Data Bank* confia em ferramentas distribuídas para a submissão, curadoria e distribuição dos dados.



Numa situação intermediária, estão os centros e arquivos de dados nacionais com coleções referenciais específicas e alto grau de curadoria e reuso. Essas instalações são gerenciadas por organizações nacionais ou por financiadores independentes como o [Welcome Trust](#). A maior parcela dos aportes financeiros destinada a esses centros é para o propósito específico de atividades de curadoria.

Com níveis variados de curadoria estão os repositórios temáticos ou disciplinares que têm as suas origens em comunidades acadêmicas e oferecem, além dos serviços convencionais de submissão e arquivamento, um elenco de serviços como visualização, modelagem, meta-análises, etc. e muitas vezes possuem instrumentos disciplinares de representação como vocabulários, taxonomia e ontologias.

As plataformas multidisciplinares, por sua vez, oferecem serviços de submissão, descoberta e acesso, porém baixa atividade de curadoria e preservação e uma carteira de serviços limitada devido principalmente a heterogeneidade em termos de formatos, tecnologia e domínios disciplinares dos dados que hospedam. Os repositórios institucionais têm também essas características e armazenam coleções de dados geradas por seus programas de pesquisa. Essas coleções representam a pluralidade de atividades de pesquisas das comunidades de uma instituição, e compõem também a memória acadêmica dessas instituições. Embora muitas delas tenham políticas para dados de pesquisa, elas tendem a ser apenas aconselhamentos genéricos ao invés de um apoio mais intensivo e direcionado, ou uma curadoria ativa com base numa supervisão abrangente que considere toda a gama de dados gerados pelo esforço de pesquisa da instituição. Essa camada apresenta, tipicamente, poucas ações de curadoria, preservação e reuso.

Na base da pirâmide estão os pesquisadores e os grupos de pesquisa individuais. Neste nível estão situados coletivamente o maior volume de dados produzidos pela ciência, a maior parte composta por coleções de dados com o volume menor que 1 GB (HORSTMANN, 2015), cuja metade está armazenada em seus laboratórios de origem. Distante das áreas disciplinares mais estruturadas, onde a norma é tornar os dados disponíveis em bases de dados internacionais ou onde há bases de dados nacionais designadas como repositórios para dados pelas agências financiadoras, os pesquisadores coletam e armazenam seus próprios dados. Nesses ambientes de pesquisa, o compartilhamento de dados se realiza como um ato de confiança entre os colaboradores mais próximos, ou por meio dos *websites* dos projetos ou laboratórios. O esforço de curadoria é geralmente apoiado financeiramente pelas verbas destinadas ao financiamento de projetos e programas do pesquisador. As ferramentas mais comuns utilizadas são os softwares *off-the-shelf* (de prateleira) convencionais, tais como Excell ou MATLAB, que carecem, porém, de funcionalidades para muitas das necessidades para uma curadoria eficiente, uso e sustentabilidade dos dados. Neste território da pequena ciência em constante transição,

se identifica a grande parcela de dados obscuros, sem qualquer gestão que os torne visíveis e aptos a serem compartilhados e reutilizados em outros empreendimentos.

## 5 AS PRINCIPAIS DIFERENÇAS: uniformidade, gestão, compartilhamento e institucionalização

Patrick Bryan Heidorn, em artigo publicado em 2008, faz uma análise ainda bastante atual sobre as diferenças entre os dados gerados pela cabeça e pela cauda longa da ciência. Essas diferenças têm uma influência determinante no tipo de gestão, curadoria, compartilhamento e reuso, e, sobretudo na carteira de serviços oferecidos pelos sistemas de informação. Partindo dessa análise e incluindo novos elementos, na tentativa de aprofundar o estudo comparativo, chegamos ao Quadro 1, que sintetiza as diferenças que são analisadas logo em seguida.

**Quadro 1** - Dados da cabeça e da cauda longa: principais diferenças

		CABEÇA	CAUDA LONGA
UNIFORMIDADE	DIVERSIDADE	homogêneos	heterogêneos
	GERAÇÃO/COLETA	instrumentos automatizados	gerados/coletados manualmente
	PROCEDIMENTOS	padronizados	específicos
GESTÃO	CURADORIA	Centralizada/institucionalizada	Individual
	REPOSITÓRIOS DIGITAIS	Disciplinares ou referenciais	Institucionais ou Multidisciplinares
	PRESERVAÇÃO	Preservados	Não preservados
	ARMAZENAMENTO	Sistemas de Storage	Computadores pessoais/dispositivos portáteis
	ESTRUTURAÇÃO	Banco de dados	Planilhas
COMPARTILHAMENTO	ACESSO	Acesso aberto/distribuído	Obscuro ou protegido
	REUSO	Imediato/globalizado	Episódico/entre a equipe
INSTITUCIONALIZAÇÃO	FINANCIAMENTO	Fluxo contínuo/Apoio internacional	Por projeto
	RECONHECIMENTO/RECOMPENSA	SIM	NÃO

Fonte: elaborado pelos autores, baseado em HEIDORN (2008)

## 5.1 Uniformidade das coleções de dados

O grande volume de dados coletados ou gerados pelos grandes projetos da cabeça são caracteristicamente mais homogêneos, pois num contexto de projetos altamente coordenado há uma concordância prévia dos pesquisadores sobre que tipo de dados serão gerados, e como eles serão formatados e armazenados para o acesso posterior. Isto acontece em parte, como enfatiza Heidorn (2008), porque existe uma demanda imediata por outros pesquisadores internos e externos ao projeto considerado. Há uma comunicação/transmissão direta. O grande volume e o nível de homogeneidade dos dados têm ainda como causa a instrumentação automatizada. Apesar da grande variedade de instrumentos usados entre projetos, no escopo do mesmo projeto os mesmos instrumentos tendem a ser usados. O uso de procedimentos padronizados resulta em práticas mais uniformizadas. A padronização dos processos de pesquisa, que caracteriza a cabeça, torna muito mais fácil o armazenamento das coleções de dados em bancos de dados estruturados e permite oferecer serviços mais diversificados que vão além do acesso como visualização, análises e modelagens.

Em contraste com a situação da cabeça, a coleta manual de dados dos projetos dos pequenos laboratórios da cauda longa não é uniforme. Cada projeto define seu próprio formato de dados, o que pode inviabilizar a estruturação das coleções em bases de dados. “Com frequência, o maior nível de estruturação alcançada na cauda é a codificação em planilhas eletrônicas”, observa Heidorn (2008, p.288). Muitas das características dos dados da cauda longa são dependentes da área disciplinar específica onde o dado é gerado/coletado. Como registra o relatório do eInfrastructure Reflection Group - e-IRG (2016), a diversidade/heterogeneidade tem muitas dimensões, entre elas, as seguintes:

- **Volume:** os dados de pesquisa da cauda longa podem variar em ordem de magnitude de uns poucos *bytes* em ASCII até a escala de *exabytes*, em grandes instalações experimentais;
- **Formato:** este item inclui a diversidade de dados no formato original e as reformatações e o uso frequente de formatos proprietários e disciplinares;
- **Estrutura:** a capacidade de referir-se a um modelo de dados explícito e implícito não é garantida;
- **Complexidade:** os dados podem ser compostos de diferentes elementos, podem ter diferentes versões ou variar com o tempo.

## 5.2 Gestão das coleções de dados

Dando continuidade à sua análise, Patrick Bryan Heidorn (2008) nos informa que no âmbito dos projetos e programas localizados na cabeça da distribuição, se torna mais econômico adotar estratégias de curadoria centralizadas para os dados, posto que os pesquisadores envolvidos estão trabalhando juntos na obtenção dos dados e parece natural que eles sejam reunidos em uma única estrutura lógica.

No que diz respeito às infraestruturas de arquivamento das coleções de dados, Rodrigues e Saraiva (2010) identificam dois tipos: o primeiro tipo caracterizando a cabeça considerando domínios disciplinares, cujos fluxos incluem práticas estabelecidas de registro e compartilhamento de dados, como Astronomia e Física. “Nestes domínios pode-se dizer que o paradigma vigente é a base de dados e não o repositório de dados” (RODRIGUES; SARAIVA, 2010, p. 24). Exemplos significativos para esses casos são o [Genbank](#) ou o [Protein Data Bank](#). As bases de dados são instaladas em estruturas tecnológicas próprias e com modelos de dados específicos que proporcionam uma curadoria próxima da comunidade de pesquisadores; o segundo tipo se refere à cauda da distribuição e está representado pelos repositórios institucionais, implantados em decorrência do movimento de livre acesso, do desejo, por parte das instituições de pesquisa, de registrar por completo a sua memória acadêmica e das políticas mandatórias das agências de fomento em relação ao compartilhamento e reuso dos dados, resultado dos projetos por elas financiados. Contudo, tipicamente, os dados da cauda longa não são encaminhados para qualquer infraestrutura formal de arquivamento, por conseguinte, recaem, na maior parte, no desuso e na obscuridade. Dados da cauda longa existem em todas as disciplinas, mas muito frequentemente estão armazenados em computadores individuais ou em algum servidor da universidade acompanhados de um conjunto mínimo de metadados ou sem nenhum metadado e nenhuma documentação (e-IRG, 2016).

No caso em que os dados são depositados, os repositórios disponíveis são os institucionais ou os multidisciplinares de organizações externas – como por exemplo o [Dryad](#). Ambos os tipos têm uma abordagem genérica e dificultam a curadoria, o acesso e a oferta de serviços mais específicos e de maior valor agregado (CRAGIN *et al.*, 2010). A gestão e curadoria desses dados é bastante complexa na medida em que faltam aos laboratórios ferramentas tecnológicas, informacionais (por exemplo, metadados disciplinares) e referenciais (por exemplo, manuais) e, ainda, políticas institucionais que garantam a estabilidade, persistência, contexto e encontrabilidade dos dados. Além disso, a natureza heterogênea e fragmentada desses dados exige estratégias diversificadas para sua gestão. É preciso

considerar também a ausência de esquemas de reconhecimento da autoria e políticas de recompensa pelo trabalho de organização e disseminação das coleções de dados, que têm forte impacto sobre o compartilhamento e reuso, mas que também desencadeiam a falta de interesse dos pesquisadores em divulgar dados além dos limites profissionais mais próximos, e, principalmente, dados sobre hipóteses não confirmadas e sobre resultados negativos, recursos reconhecidos como de grande valor para a ciência contemporânea (SAYÃO; SALES, 2019). Em contraste, na *big science*, notadamente para os dados observacionais e referenciais, há um compromisso claro com a tríade curadoria, arquivamento e preservação, que inclui os pressupostos de armazenamento seguro e ambientes confiáveis que garantam o acesso no tempo e no espaço.

### 5.3 Compartilhamento: acesso e reuso

Tornar um conteúdo que foi criado para uma audiência, útil para outra, é um problema complexo, porque cada disciplina tem seu próprio vocabulário, estrutura de dados e práticas de pesquisa e formulam questões de formas distintas usando sua própria terminologia. Isto coloca um desafio importante para os serviços de curadoria, que é criar descrições e representações, ferramentas e serviços que tornem viável o compartilhamento entre diferentes audiências (BORGMAN, 2007).

O reuso das coleções de dados está no cerne do sucesso da gestão de dados de pesquisa, posto que coletar, catalogar e adicionar valor, arquivar dados de pesquisa e compartilhar só ganha sentido se estas ações tiverem como perspectiva o reuso no seu domínio disciplinar e, num grau mais avançado, a ressignificação dessas coleções em outros domínios. “O verdadeiro valor dos dados não é determinado pelo custo de recolhê-lo, mas pela economia incorrida por alguém que não precisa reuni-lo novamente” (HEIDORN, 2008, p. 290). O potencial de reuso é o primeiro parâmetro de avaliação dos dados de pesquisa (BORGMAN, 2007). Entretanto, não há uma conexão direta entre compartilhamento e reuso. Compartilhar os dados, publicando-os na *web*, por exemplo, não significa que eles terão potencial de reuso. Para isso é necessário garantir a interpretabilidade desses recursos por meio de documentação e de um conjunto de metadados, com ênfase nos metadados disciplinares. Esses elementos garantem a semântica, estrutura e contexto às coleções de dados.

No mundo da *big science*, cada domínio disciplinar tem, geralmente, seus próprios dispositivos semânticos como esquema de metadados, ontologias, taxonomias e padrões de anotação, que garantem um alto grau de interpretabilidade das coleções de dados e de precisão na recuperação. As consultas

dirigidas às coleções disciplinares, tipicamente, permitem aos usuários recuperar registros individuais com maior rigor. “Por exemplo, um pesquisador pode recuperar uma sequência de nucleotídeos em um gene particular de uma espécie particular ou individual no GenBank” (HEIDORN, 2008, p.289). Em contraste, a maioria dos repositórios institucionais – onde tipicamente são publicados os dados da cauda longa – tem que gerenciar muitos e diferentes formatos e descreve o conjunto de registros ao invés de registros individuais, limitando a recuperação às coleções inteiras como foram originalmente depositadas pelos autores. Isso implica na necessidade de processamento manual extra por parte de pesquisadores que precisam selecionar registros específicos.

O acesso aos dados da *big science* é caracteristicamente aberto e a disseminação é distribuída em escala planetária por meio, por exemplo, de computação em grade, e o reuso é imediato, orientado para uma demanda bem identificada. De forma distinta de campos como Genômica e Astronomia, que tendem a ter padrões nas práticas relacionadas ao intercâmbio de dados, o compartilhamento na pequena ciência não é comum e nem esperado. Ele funciona como uma indústria artesanal onde os dados são intercambiados baseados nas relações profissionais e na comunicação pessoal. Há, portanto, uma invisibilidade na cauda longa da pesquisa, especialmente em relação às coleções de dados geradas nesse segmento da ciência. “Estudos recentes indicam que mais de 50% das descobertas científicas não aparecem na literatura publicada, ao invés disso, residem nas gavetas e nos computadores pessoais dos pesquisadores” (FERGUSON *et al*, 2014, p. 1443).

Não obstante o relativo grau de negligência e de opacidade das atividades de pesquisa da cauda longa, quando comparada com a visibilidade e os cuidados dedicados à grande ciência, algumas iniciativas nacionais e internacionais concentram sua atenção em compreender a ciência gerada nos pequenos laboratórios. Essas iniciativas incluem pesquisas, instalações, grupos de interesse e ações governamentais, tendo a cauda longa como principal objeto de interesse (e-IRG, 2016).

## **5.4 Institucionalização e sustentabilidade**

Considerações sobre a persistência do acesso aos dados de pesquisa e da sustentabilidade das plataformas de gestão, como os repositórios de dados, não podem ser encaradas como extensões ou algo acessório aos projetos e programas de pesquisa, cujo fim está previamente determinado nos seus cronogramas. A facilitação do acesso, a gestão e a preservação desses dados requerem planejamentos orçamentários específicos e um suporte financeiro apropriado. Essa constatação tem

origem na própria natureza da gestão de dados de pesquisa que é um processo que se desenrola indefinidamente. Isto implica que o fluxo de fundos para a gestão das coleções de dados com potencial de reuso deve se compatibilizar com o ritmo dessa continuidade, o que parece óbvio, mas que na prática é frequentemente negligenciado (SAYÃO; SALES, 2013). Dessa forma, persiste como condição crítica para um futuro de longo prazo, para os dados de pesquisa, o reconhecimento de que a alocação contínua de recursos é um passo fundamental para os processos de gestão e curadoria. No entanto, é preciso observar que assegurar a sustentabilidade econômica de um conjunto de dados de pesquisa ultrapassa a mera alocação de recursos. Na opinião de Erway e Lavoie (2012), o processo envolve a utilização eficiente destes recursos e a alavancagem de parcerias e colaboração no sentido de se alcançar uma economia de escala. Isto pode significar, na prática, que o estabelecimento de arranjos institucionais abrangentes e organicamente comprometidos sejam essenciais na sustentabilidade das coleções de dados de pesquisa de valor contínuo.

A análise de Heidorn (2008) considera a distribuição de fundos por parte das agências de fomento à pesquisa. O autor destaca que no escopo dos grandes projetos, as agências públicas e privadas frequentemente dão especial atenção aos dados resultantes dos projetos por elas financiados. Essas agências procuram assegurar o máximo de impacto para os seus investimentos em pesquisa, e para tal, se torna cada vez mais comum a exigência de planos de compartilhamento de dados e a obrigatoriedade de acesso aberto aos dados para toda a comunidade científica.

Os pequenos projetos não atraem esta supervisão, de forma que os dados podem acabar armazenados em lugares inapropriados, protegidos contra o acesso e a perspectiva mais ampla da comunidade científica. No âmbito da cauda longa, raramente, os orçamentos dos projetos preveem o tratamento dos dados para o compartilhamento e reuso, e as instituições não estão comprometidas e nem preparadas tecnicamente para o tratamento desses ativos informacionais para disseminação em maior escala. Os dados disseminados são aqueles que são publicados de forma resumida, incluídos em artigos, teses e demais publicações acadêmicas. Heidorn (2008) inclui na sua análise a diferença dos dados na cabeça e na cauda longa considerando também o impacto que o acesso tem na carreira do pesquisador. Para os dados provenientes da cauda longa, a única definição de sucesso, segundo o autor, é publicação em periódicos de uma representação abstrata e reduzida dos dados na forma, por exemplo, de gráficos, tabelas ou estatísticas. Na grande ciência, por sua vez, a gestão de dados é frequentemente objeto de interesse acadêmico e social, como desdobramento, os esquemas de ascensão profissional e de recompensa voltados

para os pesquisadores reconhecem o sucesso da gestão de dados primários como forma de viabilizar que os dados possam ser efetivamente utilizados.

Diante desse ambiente altamente diversificado e das diferenças algumas vezes contundentes desses dois territórios da ciência, as infraestruturas voltadas para dados de pesquisa – tecnológicas, gerenciais, políticas e informacionais – precisam considerar as exigências da cabeça e da cauda longa da pesquisa sob a perspectiva da integração e da formação de uma ecologia de dados. As recentes discussões sobre o *big data* e a emergência da tecnologia de nuvens impulsionam o surgimento de visões de e-infraestrutura de amplo espectro que podem beneficiar as pesquisas internacionais e interdisciplinares. O sucesso dessas visões depende criticamente da amplitude de suas ações em prol da implementação dos imperativos de uma pesquisa posicionada holisticamente, que respeite os fluxos das pesquisas locais e que assegure que os projetos e desenvolvimentos de infraestruturas de dados considerem todas as diferenças existentes. Este tema será tratado a seguir.

## 6 IMPRESCINDÍVEL DIVERSIDADE DOS DADOS

Enquanto o *Big Data* científico compreende um vasto espectro de dados estruturados e não estruturados, nos quais se identifica uma tendência na direção da homogeneidade e da padronização (BORGMAN, 2015), no lado oposto desse espectro, a cauda longa da pesquisa é descrita como um território que realça a variedade de dados em diversas dimensões: estrutura, disciplinas, complexidade, formatos, volume e aplicação na pesquisa (HEIDORN, 2008; e-IRG, 2016). Na integração desses estoques informacionais, “é necessário compreender como diferentes blocos de construção podem se combinar para a construção de um ecossistema [de dados] relevante e operacional” (e-IRG, 2016).

Os dados da cauda longa, caracterizados pela extrema heterogeneidade e transitoriedade, precisam se integrar à homogeneidade e à permanência da grande ciência formando uma ecologia ou diversidade de dados. Isto porque nem sempre a grande ciência, definida por predicados estáveis, é o modelo mais adequado para algumas das áreas mais avançadas e inovadoras da pesquisa científica.

Na maioria das vezes, integrar dados formando uma diversidade de dados transversalmente rica, estabelece modelos eficientes de geração de conhecimento como em alguns segmentos da Astronomia, que articula dados provenientes dos grandes observatórios com os dados coletados em projetos de pequenos laboratórios terrestres. “Muitos resultados do [Programa Copernicus](#) vão exigir que dados



provenientes de satélites e dados obtidos in-situ, geralmente coletados por programas orientados para fins específicos, sejam usados em conjunto” (e-IRG, 2016, p.12).

Generalizando, a diversidade de dados pode ser considerada o princípio organizador mais inovador da pesquisa científica. Num extremo do espectro da ecologia de dados está o *big data* científico, endereçando o aumento exponencial de geração e disponibilidade de dados homogêneos e padronizados, oferecendo múltiplos desafios para a pesquisa científica. No outro extremo, a cauda longa, sublinhando o potencial criativo e inovador proporcionado pela variedade em termos de estrutura, domínio disciplinar, nível de complexidade, contextos, formatos, volume, localização e de aplicação prática. Essa diversidade quando organizada cria um ecossistema científico, ressaltando a importância das relações entre os dados da Grande e da Pequena Ciência.

Os dados da cauda longa são muito complicados para gerenciar e compartilhar, mas estão por todos os lugares, se explorados apropriadamente, criam oportunidades extraordinárias de otimizar o potencial da Ciência Aberta e da Ciência Cidadã. (e-IRG, 2016; HEDSTROM, MYERS, 2014, p.1).

## 7 À GUIA DE CONCLUSÃO

As universidades e os institutos de pesquisa, principalmente por meio de ações que partem das suas bibliotecas, começam a oferecer serviços, infraestruturas e a delinear políticas para apoiar a gestão e o arquivamento de dados de pesquisa gerados por seus pesquisadores e estudantes de pós-graduação, mais proximamente identificados com a pequena ciência. Há nessas iniciativas uma demanda explícita dos pesquisadores, cuja origem está, principalmente, nas exigências mandatórias das agências de fomento e nas políticas de publicação dos periódicos científicos. Essas exigências começam a se sobrepor aos obstáculos pessoais, culturais e infraestruturais que se interpõem ao pleno compartilhamento e reuso dos dados de pesquisa da cauda longa.

As iniciativas mundiais em torno da pesquisa da cauda longa abrem janelas para que a ciência contemporânea comece a compreender que é hora de voltar a atenção para os dados gerados por este domínio da ciência. Esses dados são componentes imprescindíveis na formação de uma ecologia de dados que tenha como predicado a integração da entropia da cauda longa à homogeneidade da *eScience*, impulsionando a inovação, a pesquisa interdisciplinar e a geração de empregos. Nessa perspectiva, o acesso às coleções de dados da cauda longa pode ter um impacto significativo no desenvolvimento da ciência.

Entretanto, a heterogeneidade dos dados da cauda longa implica na necessidade de gestão e curadoria personalizadas e que considere parâmetros tecnológicos, culturais e sociais exclusivos e talhados às especificidades de cada situação. Estas condições devem ser refletidas nos projetos de plataformas institucionais e multidisciplinares que estão sendo delineados neste momento. Num plano convergente, a cauda longa deve ser beneficiada também por todos os desenvolvimentos que estão sendo colocados em prática para solucionar as exigências de e-infraestrutura voltadas para a ciência como um todo. Isto inclui ações no campo político, tecnológico, ético e legal, que devem ser reorientadas para contemplar os produtores e usuários de dados da pequena ciência.

Os estudos sobre os perfis das pesquisas e os fluxos de geração de dados na cauda longa são particularmente importantes para o país, que está atualmente investindo na implantação de infraestruturas de dados tipificadas como institucionais e multidisciplinares. Porém, sem ligação orgânica com as comunidades científicas e com suas demandas, em termos de curadoria e de serviços, corre-se o risco desses empreendimentos se transformarem em cemitério de dados ou em arquivos vazios. As bibliotecas de pesquisa, que estão muito próximas dos pesquisadores da cauda longa e de seus experimentos e laboratórios, têm papel importante e precisam se reinventar - ancoradas, porém, no seu conceito secular - para criar arcabouços sustentáveis, talhados nas necessidades de seus usuários e que confirmem aos produtos de pesquisa da cauda longa a visibilidade necessária para o compartilhamento e reuso.

## REFERÊNCIAS

BORGMAN, Christine L. **Big Data, Little Data, no Data: scholarship in the networked world**. London : The MIT Press, 2015.

BORGMAN, Christine L. **Scholarship in the Digital Age: information, infrastructure, and the internet**. London : The MIT Press, 2007.

BORGMAN, Christine L. The conundrum of sharing research data. **Journal of the Association for Information Science and Technology**, v. 63, n. 6, p. 1059-1078, June 2012. Disponível em: <<https://dl.acm.org/citation.cfm?id=2222887>>. Acesso em: 07 ago. 2018.

BROAD, William J. Big science: is it worth the price? A periodic look at the largest new research projects; Heavy Costs of Major Projects Pose a Threat to Basic Science. **The New York Times**, May 27, 1990. Disponível em: <<https://www.nytimes.com/1990/05/27/us/big-science-it-worth-price-periodic-look-largest-new-research-projects-heavy.html>>. Acesso em: 29 out. 2018.

CRAGIN, Melissa H. et al. Data sharing, small science, and institutional repositories. **Philosophical Transactions of the Royal Society A**, n. 368, p. 4023–4038, 2010. Disponível em: <<https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2010.0165>>. Acesso em: 29 out. 2018.

INFRASTRUCTURE REFLECTION GROUP. **Long Tail of Data**: e-IRG Task Force Report. Hague: e-IRG, Sept. 2016. Disponível em: <<http://e-irg.eu/documents/10920/238968/LongTailOfData2016.pdf>>. Acesso em: 03 out. 2018.

ERWAY, Ricky; LAVOIE, Brian. **The Economics of Data Integrity**. Ohio : OCLC, 2012. Disponível em: <<https://www.oclc.org/content/dam/research/publications/library/2012/erway-dataintegrity.pdf>>. Acesso em: 21 set. 2018.

FERGUSON, Adam R. et al. Big data from small-data: data sharing in the 'long tail' of neuroscience. **Nature Neuroscience**, v. 17, n. 11, p. 1442-1447, Nov. 2014. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4728080/>>. Acesso em: 07 ago. 2018.

HEDSTROM, Margareth; MYERS, Jim. **SEAD**: finding the long tail lost in big data. 2014 Disponível em: <<http://sead-data.net/sites/default/files/pubs/NDS-White-Paper-The-Long-Tail-Lost-in-Big-Data-2.pdf>>. Acesso em: 07 ago. 2018.

HEIDORN, Patrick Bryan. Shedding light on the dark data in the long tail. **Library Trends**, v. 57, n. 2, p. 280-299, 2008. Disponível em: <[https://www.researchgate.net/publication/49175975\\_Shedding\\_Light\\_on\\_the\\_Dark\\_Data\\_in\\_the\\_Long\\_Tail\\_of\\_Science](https://www.researchgate.net/publication/49175975_Shedding_Light_on_the_Dark_Data_in_the_Long_Tail_of_Science)>. Acesso em: 23 out. 2018.

HOBBSAWM, Eric. **Era dos extremos**: o breve século XX - 1914-1991. São Paulo: Companhia das Letras, 1995.

HORSTMANN, Wolfram. Beyond the big data: the long tail of research. In: E-IRG WORKSHOP, 2015, Riga. **Anais...** . Apresentação em Power Point. Disponível em: <<http://e-irg.eu/documents/10920/288074/2+Wolfram+Horstmann.pdf>>. Acesso em: 01 out. 2018.

MacCOLL, John. The role of libraries in data curation. In: RLG PARTNERSHIP ANNUAL MEETING, June 2010, Chicago. **Anais...** Disponível em: <<http://slideplayer.com/slide/723056/>>. Acesso em: 04 out. 2018.

PETSKO, Gregory A. Big science, little science. **EMBO Rep**. v. 10, n. 12, p. 1282, Dec. 2009. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2799218/>>. Acesso em: 03 set. 2018

PRICE, Derek J. de Solla. **O desenvolvimento da ciência**. Rio de Janeiro: Livros Técnicos e Científicos, 1976, 96p.

RODRIGUES, Eloy; SARAIVA, Ricardo. **Os repositórios de dados científicos**: estado da arte. Porto: RCAAP, 2010. Disponível em: <<https://repositorio-aberto.up.pt/bitstream/10216/23806/2/44632.pdf>>. Acesso em: 07 ago. 2018.

SAYÃO, Luis Fernando; SALES, Luana Farias. A ciência invisível: os dados da cauda longa da pesquisa científica. In: DIAS, G. A.; OLIVEIRA, B. M.J. F. **Dados científicos**: perspectivas e desafio. João pessoa: EdUFPB, 2019. No Prelo.

SAYÃO, Luis Fernando; SALES, Luana Farias. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência**

**da Informação**, v. 6, n. 1, 2013. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000014157/0f3127bfd6475f45e6950e1e0da20371>>. Acesso em: 30 out. 2018.

THE ROYAL SOCIETY. **Science as an open enterprise**. London: The Royal Society Science Policy Centre, 2012. Disponível em: <<https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>>. Acesso em: 23 out. 2018.

WEINBERG, Alvin M. Impacto of large-scale science on the United States. **Science**, v.134, n. 3473, July 1961. Disponível em: <[http://www.andreasaltelli.eu/file/repository/Weinberg\\_Big\\_Science.pdf](http://www.andreasaltelli.eu/file/repository/Weinberg_Big_Science.pdf)>. Acesso em: 03 set. 2018.

WYBORN, Leslie; LEHNERT, Kerstin. Exploiting the long tail of scientific data: making small data BIG. In: ERESEARCH AUSTRALASIA CONFERENCE, 10-14 Oct. 2016, Melbourne, Australia. **Anais...** Melbourne, Australia, 2016. Disponível em: <[https://eresearchau.files.wordpress.com/2016/03/eresau2016\\_paper\\_88.pdf](https://eresearchau.files.wordpress.com/2016/03/eresau2016_paper_88.pdf)>. Acesso em: 03 set. 2018.