



## A Curadoria Digital de Dados Científicos no Campo da Ciência da Informação

**Marcello P. Bax<sup>I</sup>**

<http://orcid.org/0000-0003-0503-3031>

**Liliane Chaves de Resende<sup>II</sup>**

<http://orcid.org/0000-0002-4650-0540>

<sup>I</sup> Universidade Federal de Minas Gerais, MG, Brasil.  
Doutor em Ciência da Informação.  
Professor do PPG-GOC/Escola de Ciência da Informação.

<sup>II</sup> Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas Gerais,, MG, Brasil  
Mestre em Gestão & Organização do Conhecimento  
Professora no IFMG São João Del Rei.

<http://dx.doi.org/10.1590/1981-5344/4306>

*O compartilhamento de dados científicos é primordial para a ciência, pois seu progresso depende, talvez como nunca antes, da colaboração entre os diversos atores e comunidades científicas. Investiga-se, no artigo, sobre a importância da Curadoria Digital de Dados Científicos (CDDC) para a Ciência da Informação (CI) no Brasil, frente à tendência internacional. O método seguiu abordagem quali-quantitativa, de natureza aplicada, exploratória e descritiva. Pesquisa tipo Survey foi aplicada para obter a opinião dos pesquisadores da área no Brasil. As análises revelam que o conhecimento sobre CDDC é atualmente escasso, havendo relativo baixo grau de adesão dos pesquisadores à essa temática. Tais constatações configuram um cenário desafiador para a CI, mas também de oportunidades para o profissional. Constata-se que a área está ainda se adaptando às novas tendências em CDDC. Espera-se contribuir com insumos conceituais para que a CI reflita mais objetivamente sobre um plano de ações concreto capaz de preencher a lacuna*

*teórica e prática constatada pela pesquisa. Poderá adequar os currículos de formação de seus futuros pesquisadores e profissionais para que passem a contribuir efetivamente com novas soluções para os desafios da curadoria digital.*

**Palavras-chave:** *Curadoria Digital. Dados Científicos. Profissional da Informação. Ciência da Informação. Survey. e-Science.*

## **The Digital Curation of Scientific Data in Information Science Field**

*Sharing scientific data is paramount to science, as its progress depends, perhaps as never before, on collaboration between the various scientific actors and communities. The present work investigates the importance of the Digital Curation of Scientific Data (CDDC) for Information Science (IS) in Brazil, in view of the international trend. The method followed a qualitative and quantitative approach of an applied, exploratory and descriptive nature. Survey research was applied to obtain the opinion of researchers in the area in Brazil. Results show that knowledge about CDDC is currently scarce, and researchers have relatively low adherence to this theme. These observations represent a challenging scenario for CI, but also opportunities for professionals. Findings indicate that the area is still adapting to new trends in CDDC. The study intends to contribute conceptual inputs so that the IS area reflects more objectively on a concrete action plan capable of filling the theoretical and practical gap found by the research. IS can tailor the training curricula of future researchers and practitioners to effectively contribute new solutions to the challenges of digital curation.*

**Key-words:** *Digital Curatorship. Scientific data. Information Professional. Information Science. Survey e-Science.*

Recebido em 27.02.2020 Aceito em 27.02.2020

## 1. INTRODUÇÃO

Curadoria digital é a seleção, preservação, manutenção, coleta e arquivamento de ativos (ou recursos) digitais (RUSBRIDGE *et al.*, 2005). A curadoria digital agrega valor aos repositórios digitais para uso presente e futuro, sendo geralmente realizada por arquivistas, bibliotecários, museólogos, cientistas, historiadores e estudiosos em geral. As empresas começam a usar a curadoria digital para melhorar a qualidade das informações e dados em seus processos operacionais e estratégicos (CURRY *et al.*, 2010). A curadoria digital bem-sucedida reduz a obsolescência digital, mantendo as informações acessíveis aos usuários (YAKEL, 2007). O termo "curadoria" se referia a profissionais de museus e bibliotecas, mas hoje vem sendo aplicado também à atividades de gerenciamento de recursos em mídias sociais, incluindo a compilação de imagens digitais, links da web, arquivos de filmes e outros recursos informacionais.

O termo "curadoria digital" foi usado inicialmente nos campos da ciência e das ciências biológicas para diferenciar o conjunto de atividades de curadores de bibliotecas e museus (para agregar valor às suas coleções), daquela sub tarefa mais simples de preservação dos dados, tarefa de arquivamento significativamente mais concisa (DALLAS, 2016). Além disso, o entendimento histórico do termo "curador" exige mais do que um simples "cuidar" da coleção. Espera-se que um curador comande o domínio acadêmico do assunto como parte necessária da avaliação e seleção de ativos e qualquer adição subsequente de valor à coleção por meio da aplicação de metadados (DALLAS, 2016).

A partir do conceito de Digital Curation for Science de Beagrie (2008), emprega-se aqui o termo "Curadoria Digital de Dados Científicos" (CDDC) objetivando reforçar o entendimento de que se trata da "Curadoria Digital" de "dados científicos". A CDDC ganha crescente relevância junto às instituições de pesquisa científica. Internacionalmente, nossa pesquisa autoriza afirmar que a CDDC é, em geral, de responsabilidade de profissionais da informação (bibliotecários, arquivistas e museólogos), que apoiam pesquisadores de domínio específico no gerenciamento de seus dados. No Brasil, contudo, o cenário ainda é embrionário. Existem poucas iniciativas de CDDC no seio das comunidades científicas nacionais. Nota-se, em geral, grande interesse pela CDDC, apesar de não haver ainda políticas que incentivem a sua implantação institucional. Investir em práticas de CDDC pode transformar unidades de informação nacionais, como bibliotecas e arquivos dedicados a apoiar o fazer científico, em um locus estratégico de curadoria.

Nesse contexto, obter opiniões acerca das atividades de CDDC de responsabilidade de pesquisadores da área da Ciência da Informação (CI) é crucial para que se retrate o estado atual de maturidade da área e se

planeje a sua evolução no Brasil. Dessa forma, investiga-se aqui a relevância que CDDC têm para a CI nacional. A pesquisa segue abordagem quali-quantitativa, é de natureza aplicada, do tipo exploratória e descritiva. Utiliza-se de uma pesquisa com Survey, para obter informações da curadoria digital por meio da opinião dos pesquisadores brasileiros da CI.

A Seção 2 apresenta breve panorama da curadoria digital de dados científicos. Em seguida, a Seção 3 explicita os procedimentos da pesquisa. A Seção 4 relata sobre as análises dos dados coletados e discute os resultados. A Seção 5 sintetiza as principais conclusões do trabalho.

## **2. BREVE PANORAMA DA CURADORIA DIGITAL DE DADOS CIENTÍFICOS**

Curadoria constitui o ato ou processo de “cuidar” (HOUAIS, 2001). É semelhante ao trabalho que se faz num museu ou por um curador de arte. Emprestadas as funções desse contexto, a CD é voltada para o ambiente computacional com função específica de cuidar de recursos digitais. Na presente investigação, nos interessa, mais especificamente, examinar a CDDC; i.e., a curadoria de dados coletados e produzidos nos contextos das pesquisas científicas, executadas por pesquisadores de diversos domínios. Como disciplina emergente, seu entendimento difere da curadoria tradicional, embora exiba continuidade no que refere ao valor agregado e à relevância dada ao “material curado”. O termo CD foi mencionado inicialmente no Digital Curation: digital archives, libraries and e-science seminar (Curadoria Digital: arquivos digitais, bibliotecas e seminário de e-science), patrocinado pela Digital Preservation Coalition e pelo British National Space Centre, realizado em Londres em 2001. Uma contribuição desse evento foi estabelecer o termo “curadoria digital” (BEAGRIE, 2008).

A pesquisa científica gera um crescente fluxo de recursos digitais, produzindo dados de vários tipos: imagens, som, animação, tabelas, anotações, cálculos, dados de sensores, satélites, dentre outros. Isso ocorre nas diversas disciplinas científicas, devido à uma constante evolução tecnológica. O uso das tecnologias influencia o aumento e o fluxo da informação digital, sobretudo, no que concerne ao uso do hardware e software para acessar, interpretar e preservar informações digitais (NRC, 2015). Os metadados (dados que descrevem dados) também influenciaram esse incremento, aumentando ainda mais o volume de dados armazenados. Estes devem ser precisos, completos e são essenciais para a análise dos dados. São relevantes para as pesquisas científicas porque descrevem contextos e conteúdos de arquivos de dados. A necessidade de captura, armazenamento, gerenciamento e preservação de dados, sejam eles científicos ou não (pessoais, institucionais ou

governamentais), gera desafios e exigem novas habilidades e conhecimentos em curadoria.

### **2.1.A Curadoria Digital**

A Curadoria Digital é uma disciplina emergente, aplicada e interdisciplinar, com dimensões acadêmicas e profissionais, progredindo a partir da compreensão da dimensão técnica da informação digital como objeto de pesquisa. Ela busca desenvolver estratégias para resolver problemas do fluxo da informação digital. Cuidar desse fluxo é a sua atividade central (HIGGINS, 2018; NRC, 2015). Há diversos profissionais envolvidos nas atividades de curadoria digital, que devem ser engajados e treinados para resolver problemas de uma variedade de domínios para que os objetivos de curadoria sejam atingidos. Diferentes níveis de curadoria devem ser apropriados por diferentes produtores e consumidores da informação digital. A evolução da disciplina partiu de práticas pouco conhecidas, originadas da curadoria tradicional. Hoje, grandes consórcios internacionais estão envolvidos na elaboração e padronização de normas compartilhadas de curadoria. A disciplina evolui em paralelo à e-Science, cuja definição dada por Jim Gray em 2002, vem revolucionando o processo da investigação científica. Para Gray, os especialistas de um determinado domínio, que queiram executar atividades de curadoria, deveriam aprender, também, conceitos e técnicas relevantes utilizadas por bibliotecários como por exemplo técnicas de anotação, descrição e preservação dos dados (NRC, 2015 *apud* GRAY *et al.*, 2002).

### **2.2.Breve Histórico da Curadoria Digital no Mundo**

Um aspecto que contribuiu com a evolução da curadoria digital foi a publicação do Modelo de Referência do Sistema de Informação em Arquivamento Aberto (OAIS) em 2003 pela Organização Internacional para Padronização (ISO). Esse modelo, embora não prescreva diretamente práticas para curadoria digital, propõe uma estruturação de responsabilidades, processos e funções de arquivamento para preservação e acesso a longo prazo das informações. A criação do Centro de Curadoria Digital (DCC) em 2002, no Reino Unido também contribuiu com o desenvolvimento da área de CD. A proposta do DCC era fornecer orientações e formação em padrões, ferramentas e habilidades necessárias para práticas de curadoria digital. Desenvolveu um modelo de ciclo de vida da curadoria digital e vários documentos de treinamento e apoio para a área. Hoje, ainda constitui um centro de referência para o desenvolvimento da curadoria digital. Nos Estados Unidos e Canadá, a Associação de Bibliotecas de Pesquisa (ARL), organização sem fins lucrativos de 126 bibliotecas de pesquisa de instituições de pesquisa, é

um exemplo de iniciativa que “funciona como um fórum para troca de ideias e como agente de ação coletiva” compartilhando missões de pesquisa em CD.

A CDDC, no âmbito internacional, já está em franco desenvolvimento. Vários projetos foram implementados em vários momentos, que embasam o desenvolvimento da área de CDDC.

### **2.3 Projetos de Curadoria Digital**

Destacam-se projetos como o Centro de Curadoria Digital (DCC) no Reino Unido, visando identificar as principais habilidades e competências para o gerenciamento de dados de pesquisa. Posteriormente, em colaboração com o JISC e a RIN (Research Information Network - Rede de Informação de Pesquisa) para identificar as principais habilidades e competências para o gerenciamento de dados de pesquisa. O projeto DaMSSI-Assessment, Benchmarking and Classification (ABC) visou promover boas práticas de gerenciamento de dados entre profissionais, identificando uma estrutura curricular e os resultados para treinamento de gerência de dados. O foco desses projetos foi gerenciamento de dados de pesquisa (FENG; RICHARDS, 2018).

Feng e Richards (2018) afirmam que o Instituto de Museus e Serviços de Bibliotecas do EUA financiou três projetos: o projeto DigCCur na Universidade da Carolina do Norte em Chapel Hill (UNC-CH), a nova concentração de curadoria de dados na University of Illinois Urbana-Champaign e o projeto DigIn na Universidade do Arizona. Financiado pela Comissão Européia, também enfatizou a exploração da estrutura curricular para a formação profissional em Curadoria Digital em diferentes ambientes organizacionais, incluindo biblioteca, arquivos e museus (FERNIE; GOW; MALLOY, 2013; FENG; RICHARDS, 2018).

O National Research Council dos EUA, em seu estudo intensivo de curadoria digital, concluiu que:

As demandas por informações digitais prontamente acessíveis, precisas, úteis e utilizáveis de pesquisadores, indústrias intensivas em informações e consumidores expuseram limitações, vulnerabilidades e oportunidades perdidas para ciência, negócios e governo, como resultado da imaturidade e natureza ad hoc de curadoria digital. Há também um esforço por maior abertura e transparência em muitos setores da sociedade. Juntos, esses fatores estão criando uma necessidade urgente de políticas, serviços, tecnologias e conhecimento em curadoria digital. Embora os benefícios da curadoria digital sejam pouco compreendidos e pouco articulados, existem oportunidades significativas para incorporar a curadoria digital profundamente nas práticas de uma organização para reduzir custos e aumentar os benefícios (NRC, 2015, p. 38).

## 2.4. Curadoria Digital na Ciência da Informação

A Curadoria Digital de Dados Científicos (CDDC) é área emergente e que apresenta desafios complexos, entrelaçados em vários domínios e com a participação de vários profissionais, relacionada ao gerenciamento dos dados científicos brutos produzidos pelas pesquisas. A CDDC propõe conferir valor aos dados produzidos pelas pesquisas científicas. Esse valor agregado se refere, principalmente, à manutenção de dados que são impossíveis de recriar, objetivando os mais diversos fins: uso no ensino, validação dos resultados de pesquisa publicados, analisando os que podem ou não serem descartados, mantendo sua integridade e qualidade, para reuso em outras pesquisas (POOLE, 2016).

Kouper (2016) afirma que a CDDC, cresce na interseção entre biblioteconomia, arquivística e tecnologia da informação, e requer uma combinação de habilidades existentes e novas a serem definidas. A competência desejada inclui familiaridades técnicas para uso das tecnologias digitais e ferramentas de CDDC, a capacidade de iniciar e gerenciar projetos, programas e serviços, com fortes habilidades interpessoais e organizações (KOUPEL, 2016).

Para Jørn Nielsen e Hjørland (2014), os dados de pesquisa digitais são diferentes de documentos impressos e sua manutenção requer conhecimento científico do domínio e conhecimento avançado em tecnologia para organizá-los e armazená-los, com vistas à preservação e reutilização (JØRN NIELSEN; HJØRLAND, 2014). Para esses autores, “dados são frequentemente compreendidos como matéria-prima do processamento de informação e aquisição de conhecimento” (JØRN NIELSEN; HJØRLAND, 2014, p.3).

É necessário um conhecimento específico sobre como gerenciar dados de pesquisas em todos os domínios. As várias prioridades, considerações e estimativas a partir das perspectivas de domínio levam a uma exigência de competências mais profundas de gestores e curadores. As tecnologias exigem que os profissionais da informação repensem seu modo de trabalho. Eles devem tecer estratégias e renovar competências para aprimorar suas habilidades no ambiente digital. As habilidades em CDDC são dinâmicas e interdisciplinares, integrando de conhecimentos informática, informação, biblioteconomia, arquivística, bem como outros conhecimentos mais específicos de domínio, relacionados à criação e utilização dos dados científicos. Há severa escassez de profissionais qualificados para atividades de curadoria. A procura por tal profissional aumenta (TAMMARO, 2016; POOLE, 2016; FREITAS, 2017).

Para Tenopir *et al.* (2013), bibliotecários acreditam que os serviços de dados de pesquisa representam uma oportunidade para aumentar o alinhamento entre serviços de bibliotecas e a missão de pesquisa nas instituições acadêmicas. As bibliotecas estão desenvolvendo novos papéis

e implementando novos modelos de serviço. Bibliotecários com funções específicas de tecnologia da informação podem assumir funções de liderança no gerenciamento de dados científicos (TENOPIR *et al.*, 2013; COX, 2015). As habilidades e competências da Ciência da Informação são naturais à curadoria digital. Iniciativas comprovam tal afirmativa. Existe o projeto DataShare do DISC-UK (Data information Specialists Committee) em 2004, das universidades do Reino Unido (Edimburgo, Oxford e Southampton e a London School of Economics). O projeto teve foco nas bibliotecas de pesquisa para o campo de curadoria de dados por meio do apoio a atividades de gerenciamento dos dados e acesso aberto e tecnologias Web 2.0 para repositórios institucionais. A Associação de Bibliotecas de Pesquisa (ARL) nos Estados Unidos em 2007, que congrega em torno de 127 instituições e organizações, representa a força motriz que leva os profissionais de biblioteconomia e ciência da informação a fornecer serviços de curadoria de dados.

Contudo o avanço científico proporciona, cada vez mais, o acesso a dados, causando impactos à eficiência e eficácia das atividades científicas (LEE; STVILIA, 2017). Os bibliotecários de pesquisa devem colaborar com especialistas de domínio, de forma a propor meios de atender as necessidades de curadoria de dados em diferentes domínios, auxiliando em questões de articulação de infraestrutura de pesquisa e atividades correlatas (JØRN NIELSEN; HJØRLAND, 2014, p.3). A importância dos dados científicos está cada vez mais aparente para a CI. As discussões das publicações acadêmicas são sustentadas por dados produzidos nas pesquisas, servindo como insumos para novos projetos.

## **2.5.Desafios e Oportunidades da Curadoria Digital na Ciência da Informação**

A Curadoria Digital de Dados Científicos apresenta diversos desafios para ser implantada e cabe aos profissionais da informação, mobilizar e aprimorar habilidades e competências para atuarem nesse ambiente de atividades relacionadas à tecnologia e pesquisa, sobretudo no desenvolvimento de estratégias para o gerenciamento dos dados científicos a longo prazo. A CDDC surge nesse cenário como desafio, mas também como oportunidade que se coloca para a área CI. As suas atividades constituem o ciclo da comunicação científica, aprimorando, dinamizando metodologias de gestão e preservação de dados científicos em prol de compartilhamento e reuso. Isso acelera novas descobertas de conhecimentos por meio da colaboração entre pesquisadores e com a evolução da ciência.

Como recurso para divulgação da comunicação científica, os repositórios institucionais são considerados como principais instrumentos, disseminando conhecimento e apoiando a criação de novas descobertas,



sobretudo com o movimento da Ciência Aberta que proporcionou a democratização da informação científica. Mas sua criação ainda é recente e seu uso pouco disseminado (SHINTAKU; DUQUE; SUAIDEN, 2015). Assim como os repositórios institucionais, as bibliotecas de pesquisa (research libraries) estão cada vez mais envolvidas em atividades de gerenciamento de dados científicos. Seja pelo aumento considerável dos dados científicos produzidos, seja pelas exigências de instituições de fomento à pesquisa. A necessidade da integração das atividades de CD e atividades dos pesquisadores pode transformar a CDDC como parte da missão dos profissionais da informação. Lee e Stvilia (2017), ao entrevistar profissionais que atuam em repositórios institucionais, descobriram que existem diferentes níveis nas atividades de CDDC executadas em repositórios institucionais. Essas atividades estão ligadas à conceptualização, planejamento, criação, carga e publicação dos dados científicos. E, durante esse fluxo, o pesquisador interage com vários profissionais, entre eles bibliotecários, arquivistas, profissionais de TI (LEE; STVILIA, 2017). Jørn Nielsen e Hjørland (2014) questionam o papel de liderança das bibliotecas de pesquisa nas atividades da CD. Perguntam-se, os autores, se os bibliotecários estão, de fato, habilitados a gerenciar "dados e metadados" (documentos, na visão dos autores) de forma avançada, uma vez que para isso deveria possuir os conhecimentos compatíveis com aqueles dos especialistas de domínios. Afirmam ainda os autores que "a manutenção de conjuntos de dados requer conhecimento científico do domínio e conhecimento tecnológico avançado, a fim de armazenar e organizar os dados com o objetivo de reutilizá-los e preservá-los. Questionam-se ainda os mesmos autores: "os bibliotecários acadêmicos possuem as qualificações necessárias para esta tarefa?" "Existem diferenças principais entre o gerenciamento de dados e documentos?" "Que tipo de programa educacional pode apoiar as atividades de gerenciamento de dados?" Acreditam que os bibliotecários não perceberão ou não poderão descrever objetos (e dados) que estão implícitos nos conjuntos de dados gerados nas pesquisas.

Pensam que esse enriquecimento semântico dos dados somente seria possível pela observação empírica tal como realizada exclusivamente por pesquisadores de domínios específicos (aqueles que entendem o significado e a relevância dos dados).

Por outro lado, alguns autores afirmam que: garantir a disponibilidade dos dados científicos para acesso, uso ou reuso futuro seria prerrogativa da atividade de gerenciamento de dados e que o uso de repositórios digitais ou de bibliotecas de pesquisa para organizar dados científicos é uma decorrência necessária e altamente indicada (TIBBO; HANK, 2015).

### 3. MATERIAIS E MÉTODOS

A pesquisa aplicou questionário que buscou coletar dados junto aos pesquisadores da Ciência da Informação (CI) para elucidar as seguintes questões: 1) qual seria o perfil dos pesquisadores da CI que atuam, no âmbito da temática estudada, em programas de pós-graduação stricto sensu? 2) qual seria o grau de conhecimento desses pesquisadores sobre as atividades que giram em torno do conceito de Curadoria Digital de Dados Científicos e qual seria o nível de envolvimento destes nessas atividades? 3) qual seria a "visão" futura desses pesquisadores a respeito da temática da CDDC?

O método de amostragem utilizado foi a amostragem aleatória simples. Os pesquisadores foram convidados a participar do levantamento mais de uma vez, e todos tiveram a mesma chance de participar do levantamento. A amostra foi auto selecionada, ou seja, composta por pesquisadores que aceitaram o convite em participar do levantamento. Para o cálculo em populações finitas ( $N$ ), ou seja, com tamanho conhecido, utilizaram-se os seguintes critérios: nível de confiança de 95%, ou seja, a probabilidade de acerto para estimativas estatísticas a serem obtidas; e margem de erro máximo permitido de 9% ( $e$ ) nos resultados, ou seja, os parâmetros populacionais em estudo variam em 9% para mais ou para menos em relação às estimativas obtidas (CORREIA; 2003). Considerou-se a taxa de erro nesse valor devido ao cálculo de 25% de taxa de resposta obtida pelos respondentes. Em relação aos parâmetros estatísticos, a margem de erro com esta percentagem reflete a precisão de 95% das opiniões da população-alvo. O cálculo do tamanho da amostra ( $n$ ) foi obtido pela equação:

$$n = \frac{N * n'}{N + n'} ; \text{ sendo } n' = 1 / e$$

Onde  $n$ : é a amostra;  $N$ : é a população-alvo;  $e$ : é o erro estimado da amostragem (CORREIA, 2003).

Os questionários foram enviados, por e-mail, aos 390 membros da população-alvo. Cerca 78 não receberam o questionário pois seus endereços e-mail não foram localizados. Os e-mails foram obtidos por meio de informações públicas, disponibilizadas nas páginas oficiais dos programas de pós-graduação na área da CI ou por meio de artigos publicados em periódicos de acesso aberto. Pela fórmula acima o tamanho ideal da amostra dessa população-alvo teria sido de 198 membros. Obteve-se respostas de 97 pesquisadores, sugerindo uma amostragem representativa e válida, já que superior a 25% da população-alvo

(OLIVEIRA; GRÁCIO, 2005; MARCONI; LAKATOS, 2003). Os sujeitos foram contatados por meio de correio eletrônico (e-mail) enviado individualmente.

Os dados foram coletados de maio a agosto de 2019, por meio de questionário *on-line* auto aplicado, enviado via *e-mail*, contendo o *link* do formulário. Sintetizando, trezentos e noventa (390) pesquisadores foram convidados e 97 responderam. A amostra foi auto selecionada e dependente do número de sujeitos que responderam. A pesquisa demandava uma compreensão em nível básico do conceito de CDDC por parte dos entrevistados. A natureza emergente da temática pode ter sido um fator inibidor da participação de mais sujeitos. Por fim, utilizou-se o SPSS - *Statistical Package for the Social Sciences* – para análise dos dados levantados.

## **4.RESULTADOS E DISCUSSÃO**

Para descrever o cenário nacional da CDDC, foram analisados o perfil do pesquisador; a sua opinião e comportamento acerca das atividades principais do campo; e o que ele considera como tendências futuras de evolução do campo no contexto da Ciência da Informação.

### **4.1.Perfil do Pesquisador na Ciência da Informação**

Com vistas a caracterizar o perfil dos pesquisadores, foram levantados dados pessoais e acadêmicos básicos do pesquisador, além das atividades de CD que ele realiza. Sintetiza-se que o participante tem idade média acima de 40 anos, é em maioria do gênero feminino e possui mais de 10 anos de experiência em pesquisas acadêmicas. Atua, principalmente, em universidades do setor público; e a maioria busca fomento à pesquisa junto aos institutos nacionais. Do total, 58,2% interagem com outras universidades brasileiras, e apenas 32,8% com universidades fora do Brasil (relativamente baixa internacionalização). Quanto à sua formação científica, possui graduação não só em CI, mas é notável a participação de graduação em outras áreas do conhecimento. Poucos dos entrevistados (18%) participaram de cursos de capacitação ou treinamento específicos sobre o tema e quase todos (82%) nunca participaram da elaboração de políticas e normas nacionais do campo.

### **4.2.Opiniões e Atitudes do Pesquisador da Ciência da Informação Relativa à CDDC**

Nesta dimensão perguntou-se a respeito as opiniões, atitudes e comportamentos do pesquisador. Procurou-se observar tais dimensões com o intuito de analisar quais seriam as suas necessidades em CD, bem como quais atividades realizam no campo?

Em síntese, percebeu-se baixa consciência dos pesquisadores sobre a relevância da Curadoria Digital. Mesmo afirmando ter conhecimento de atividades como armazenar, acessar, disponibilizar e reutilizar dados de pesquisa, apenas 46% disseram ter disponibilizado seus próprios dados em algum momento. Além disso, quando gerenciam seus dados de pesquisa somente 24% disponibilizam-nos de fato em repositórios. O percentual de respondentes que afirmaram utilizar repositórios digitais de dados de pesquisa é quase 98%, sugerindo que os pesquisadores têm conhecimento da importância dos repositórios de dados para a pesquisa científica (PAVÃO *et al.*, 2018). Confirma igualmente o interesse entre os pesquisadores em reutilizar dados próprios ou de outrem. Os fatos sugerem o entendimento de que a Ciência da Informação, como outras "é ambiente colaborativo, aberto e digital" (OLIVEIRA; SILVA, 2016, p.11). Porém, mesmo sabendo da tendência do acesso aberto a dados científicos, 84% não os disponibilizam abertamente.

No Brasil, atualmente, constam em torno de oito repositórios de dados científicos, registrados no r3data.org (*Registry of Research Data Repositories Initiative*). Há ainda a proposta de um projeto entre a Rede Nacional de Ensino e Pesquisa (RNP) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) para constituição de um grupo de trabalho juntamente a universidades federal do Rio Grande do Sul e federal do Rio Grande para desenvolver o Acesso Aberto a Dados de Pesquisa (AADP). Segundo Sayão e Sales, a expectativa é a de que repositórios de dados vinculados a instituições acadêmicas criarão oportunidades sem precedentes para acelerar a pesquisa científica (SAYÃO; SALES, 2016).

Para Pavão *et al.* (2018) a prática de compartilhamento de dados não é comum na ciência brasileira. Os dados obtidos confirmam essa visão que é justificada com argumento, dentre outros, de que "no Brasil falta de instrumentos e diretrizes uniformes para a preservação e segurança de dados de pesquisa". De Oliveira e Da Silva (2016) alegam que o avanço no arcabouço normativo-legal brasileiro, referente aos repositórios institucionais, é lento e que fatores restritivos estão além de questões tecnológicas e técnicas; assentam-se, principalmente, em âmbitos políticos, legais, econômicos e culturais. As evidências em relação à disponibilização dos dados também corroboram a afirmativa de Pavão *et al.* (2018) pelo baixo percentual de respondentes, 12,8% que afirmam disponibilizar seus dados. Esse valor é, ainda, inferior ao percentual de respondentes que reutilizam dados de pesquisa, 18,3% (OLIVEIRA; SILVA, 2016; PAVÃO *et al.*, 2018). Portanto, para avançar com a CDDC, são necessários esforços em várias instâncias, envolvendo, principalmente, "ação e decisão diferenciadas, internas e externas à ciência, que vão desde o pesquisador individual e equipes de pesquisa até

o nível macro das políticas públicas e das regulações internacionais, passando pelo nível mesmo das instituições científicas e agências de fomento” (ALBAGLI, 2015, p. 21).

Acerca das técnicas relacionadas às atividades de CDDC sobre descrição de dados científicos, Poole (2016) esclarece que estudos apontam que pesquisadores não possuem familiaridade com a criação ou documentação de metadados, e que não conseguem prever as necessidades daqueles que reutilizariam seus dados. Ele também reforça que metadados estruturados são essenciais para o compartilhamento e reutilização dos dados (POOLE, 2016). Por isso os serviços de curadoria de dados científicos desempenham um papel essencial para apoiar todos os pesquisadores, especialmente da CI (que podem prestar apoio às outras áreas), no gerenciamento dos dados científicos produzidos nas pesquisas (TRIPATHI; SHUKLA; SONKAR, 2017). Contudo, mais que em outras áreas, a situação na CI é crítica, pois os dados mostram que não há, ainda, participação do pesquisador em questões básicas, mesmo tendo conhecimento da sua importância. Com efeito, 93% concordam que a CDDC representa oportunidade de carreira diferenciada para o profissional da informação. Percebe-se claramente a necessidade de ampliar e aprofundar a formação desse profissional, para que assuma a nova missão, colocada pela CDDC.

Mesmo que, para 77% dos sujeitos, o índice de citação dos artigos aumente com a disponibilização dos dados utilizados, 70% desses sujeitos ainda mantêm seus dados privados, em computadores pessoais. Vale notar que, segundo 91% dos entrevistados, as questões éticas são as mais importantes no campo. Com relação a ferramentas e padrões, para 46% dos entrevistados, o *software* de repositório mais utilizado é o DSpace. Dezoito por cento (18%) dos pesquisadores têm como referência, dentre os modelos de gerenciamento de dados científicos, o modelo do DCC; quatorze por cento (14%) têm o OAIS como referência. Entretanto, é alto o número de pesquisadores que afirmam não utilizar plataformas tecnológicas para armazenar (47%) e para reuso dos dados (51%). É alto também o percentual de pesquisadores que não sabem dizer (52%) qual padrão de metadados utilizam para descrever seus dados científicos. Tais resultados sugerem que é necessária e urgente a conscientização sobre o tema, bem como buscar a maior imersão destes pesquisadores nas práticas de gestão de dados científicos e de atividades de CD. Somente assim poderão eles contribuir com o desenvolvimento da curadoria digital na pesquisa brasileira.

### 4.3. Estado Atual e Tendências da CDDC na Opinião do Pesquisador da Ciência da Informação

Ao serem solicitados a opinar a respeito da CDDC no Brasil, as respostas dos pesquisadores foram bastante variadas, gerando a nuvem de palavras mais frequentes da Figura 2.

**FIGURA 2** – Nuvem de palavras mais frequentes



Fonte: Dados de pesquisa

Sobre o estágio de evolução da CDDC no Brasil, os pesquisadores afirmaram que a CDDC está "... na sua forma embrionária, dando os primeiros passos". Afirmações de que a CDDC está "precisando de mais estudos que a desenvolvam" constituem uma amostra desse cenário. Um dos respondentes menciona iniciativa em que ela está:

*"...em estágio embrionário, pois ainda não há uma política nacional para a gestão e compartilhamento de dados científicos. Porém, merece ser comentado que, em novembro de 2018, o Ministro Kassab (MCTIC) criou o grupo de trabalho para elaborar uma minuta de Decreto para a ciência aberta no Brasil, espaço onde se têm discutido questões complexas, como, por exemplo, papel das agências de fomento, acesso a recursos internacionais para desenvolvimento de pesquisa, disponibilização de dados que envolvem soberania nacional em repositórios internacionais. O fato é que a curadoria de dados passará a ser uma necessidade a partir do momento em que existir uma diretriz nacional – e fomento da CAPES /CNPq. A partir desse momento, as agências de fomento no Brasil passarão a exigir uma Plano de Gestão de Dados para o pesquisador".*

Portanto, de acordo com os pesquisadores da CI, há ainda um longo caminho a ser percorrido para que a CI assuma o campo como fazendo parte da sua missão. Contudo as respostas indicaram que a CDDC no Brasil, e não apenas na CI, está "em sua fase de expansão, apesar de não ser amplamente discutida dentro dos espaços institucionais, como também em disciplinas específicas" e "se tornando um assunto que

*necessita de atenção dos pesquisadores, pois a organização e acesso aos dados abertos de objetos digitais é uma necessidade eminente”.*

## **5. CONSIDERAÇÕES FINAIS**

O *survey* apresentado investigou sobre o grau de importância da temática "Curadoria Digital de Dados Científicos" (CDDC) para o campo da Ciência da Informação (CI) no Brasil, levantando e descrevendo detalhadamente o cenário nacional. Buscou-se explorar de forma empírica e aprofundada como a CI, através de seus representantes pesquisadores, percebe o seu papel frente a este fenômeno e também como ela está acompanhando a tendência internacional na área.

Contribui-se também, com a apresentação do cenário de atuação profissional e acadêmica dos pesquisadores da CI em CDDC. Acredita-se ser possível, a partir deste estudo, iniciar uma discussão mais fundamentada em dados acerca de estratégias para incrementar a relevância dessas atividades para esses profissionais, e, conseqüentemente, para a própria Ciência da Informação.

Inferiu-se, em relação aos dados coletados e à opinião dos pesquisadores, que já há uma movimentação para adesão à curadoria de dados científicos. Há o entendimento de que a CDDC tem por missão, mais do que simplesmente disponibilizar e preservar os dados, cuidar também da modelagem conceitual dos estudos e do enriquecimento semântico dos dados gerados pelos estudos. O levantamento permite afirmar, contudo, que é baixa a atuação dos pesquisadores sobre o tema. Constatou-se claramente ainda que práticas de CD por esses pesquisadores são quase que inexistentes. Portanto, atesta-se que não há suficiente envolvimento com a CDDC que a posicione como um pilar da sua missão. Assim, concluiu-se, de forma específica, ser ainda necessária uma mudança evolutiva considerável na formação disciplinar teórica, prática e técnica desses pesquisadores. Tal conclusão é ainda mais relevante para o fortalecimento da CI enquanto área de conhecimento como um todo, pois são esses mesmos pesquisadores que formam os futuros profissionais da informação que irão atuar nas práticas profissionais e em serviços técnicos de informação.

Uma limitação desta pesquisa é que a participação dos pesquisadores da CI nacional poderia ter sido ainda maior numericamente e também mais abrangente. Acredita-se, contudo, que os objetivos inicialmente propostos, mesmo que preliminarmente, foram atingidos. Como trabalhos futuros, sugere-se que sejam realizadas novas pesquisas que procurem gerar informações ainda mais detalhadas para subsidiar e aprofundar esta discussão, complementando, corroborando ou refutando os resultados alcançados até aqui. O levantamento detalhado das informações descritivas do cenário da CI frente à CDDC permitirá que se

estabeleçam estratégias de evolução do tratamento deste tema pela área da CI no país. Permitirá ainda discutir, de forma embasada empiricamente, sobre lacunas da formação disciplinar do profissional da informação, tanto em nível de graduação quanto pós-graduação.

Outra sugestão é investigar temas mais específicos no domínio da CDDC, tais como: (1) a integração, o reuso, a publicação e a preservação de dados científicos; e (2) o uso de ferramentas de repositórios de dados e não apenas textos (artigos, monografias, teses e dissertações). Informações sobre tais atividades e ferramentas permitirão que a CI reflita sobre estratégias e meios para conscientizar a comunidade acadêmica para a importância da CDDC, bem como sobre a importância de se disponibilizar os dados produzidos em suas pesquisas. Acredita-se que tal discussão permitirá que a Ciência da Informação contribua, de forma efetiva, para melhorar as taxas de reuso de dados e condições de reprodução dos resultados de pesquisas científicas em todas as áreas de conhecimento, assim como de seus próprios pesquisadores.

## Referências

ALBAGLI, S. Ciência aberta em questão. *Ciência aberta, questões abertas*. Brasília: Ibict, p. 9-25, 2014. Disponível em:

<http://ridi.ibict.br/handle/123456789/910>. Acesso em: 7 jan.2018.

BEAGRIE, N. Digital curation for science, digital libraries, and individuals. *International Journal of Digital Curation*, v. 1, n. 1, p. 3-16, 2008.

Disponível em: <https://doi.org/10.2218/ijdc.v1i1.2>. Acesso em: 26 out.2018.

CORREIA, M. S. B. B. *Probabilidade e estatística*. 2ª ed. Belo Horizonte: PUC Minas Virtual, 2003.

COX, A. *et al.* Librarians as partners in research data service development at Griffith University. *Program: electronic library and information systems*, v. 49, n. 49, p.440-460, 2015. Disponível em: <https://doi-org.ez367.periodicos.capes.gov.br/10.1108/PROG-02-2015-0013>. Acesso em: 02 ago.2018.

CURRY, E.; FREITAS, A.; O'RIÁIN S. *The role of community-driven data curation for enterprises*. In: WOOD; D. Linking enterprise data, Springer, Boston, MA, 2010. p. 25-47. Disponível em: [https://doi-org.ez367.periodicos.capes.gov.br/10.1007/978-1-4419-7665-9\\_2](https://doi-org.ez367.periodicos.capes.gov.br/10.1007/978-1-4419-7665-9_2). Acesso em: 06 mai.2019.



DALLAS, C. Digital curation beyond the “wild frontier”: a pragmatic approach. *Archival Science*, v. 16, n. 4, 421-457, 2016. Disponível em: <https://rdcu.be/b1fKm>. Acesso em: 06 mai.2019.

FENG, Y.; RICHARDS, L. A review of digital curation professional competencies: theory and current practices. *Records Management Journal*, v. 28, n. 1, p. 62-78, 2018. Disponível em: [https://doi-org.ez367.periodicos.capes.gov.br/10.1108/RMJ-09-2016-0022](https://doi.org.ez367.periodicos.capes.gov.br/10.1108/RMJ-09-2016-0022). Acesso em: 02 ago.2018.

FERNIE, K.; GOW, A.; MOLLOY, L. *Digital Curator Vocational Education Europe: overview of the DigCurV project*, 2013.

FREITAS, C. O futuro é hoje: perfis e competências dos profissionais da informação para a curadoria digital. In: *Encontro Curadoria Digital – Estratégias e experiências: atas*, p. 28-39, 2017. Disponível em: <http://hdl.handle.net/10316/45735>. Acesso em: 26 out.2018.

HIGGINS, S. Digital curation: the development of a discipline within information science. *Journal of Documentation*, v. 74, n. 6, p. 1318-1338, 2018. Disponível em: <https://doi.org/10.1108/JD-02-2018-0024>. Acesso em: 22 nov.2018.

HOUAISS, A.; VILLAR, M.; DE MELLO FRANCO, F. M. *Dicionário Houaiss da língua portuguesa*, 2001.

HULLEY, S. B. *et al. Delineando a pesquisa clínica-4*. Artmed Editora, 2015.

JØRN NIELSEN, H.; HJØRLAND, B. Curating research data: the potential roles of libraries and information professionals. *Journal of Documentation*, v. 70, n. 2, p. 221-240, 2014. Disponível em: <https://doi.org/10.1108/JD-03-2013-0034>. Acesso em: 26 out.2018.

KOUPER, I. Professional participation in digital curation. *Library & Information Science Research*, v. 38, n. 3, p. 212-223, 2016. Disponível em: <https://doi.org/10.1016/j.lisr.2016.08.009>. Acesso em: 26 out. 2018.

LEE, D. J.; STVILIA, B. *Practices of research data curation in institutional repositories: A qualitative view from repository staff*. PloS one, v. 12, n. 3, 2017. Disponível em: <https://doi.org/10.1371/journal.pone.0173987>. Acesso em: 22 nov.2018.

MARCONI, M. A.; LAKATOS, E. M. *Fundamentos de metodologia científica*. 5. ed. São Paulo: Atlas, 2003.

NRC. NATIONAL RESEARCH COUNCIL *et al.* Preparing the workforce for digital curation. Local: National Academies Press, 2015. Disponível em: <http://www.nap.edu/catalog/18590/preparing-the-workforce-for-digital-curation>. Acesso em: 02 ago.2018.

OLIVEIRA, E. F. T.; GRÁCIO, M. C. C. Análise a respeito do tamanho de amostras aleatórias simples: uma aplicação na área de Ciência da Informação. *Revista Ciência da Informação*, v. 6, n. 3, p. 1-11, 2005. Disponível em: [www.brapci.inf.br/index.php/article/download/11612](http://www.brapci.inf.br/index.php/article/download/11612). Acesso em: 25 mai.2019.

OLIVEIRA, A. C. S.; SILVA, E. M. Ciência aberta: dimensões para um novo fazer científico. *Informação & Informação*, v. 21, n. 2, p. 5-39, 2016. Disponível em: <http://www.uel.br/seer/index.php/informacao/article/view/27666>. Acesso em: 6 mai.2019.

PAVÃO, C. M. G. *et al.* Acesso aberto a dados de pesquisa no Brasil: repositórios brasileiros de dados de pesquisa: relatório, 2018. Porto Alegre, RS: UFRGS, 2018. Disponível em: <http://hdl.handle.net/10183/185138>. Acesso em: 6 maio. 2019.

POOLE, A. H. The conceptual landscape of digital curation. *Journal of Documentation*, v. 72, n. 5, p. 961-986, 2016. Disponível em: <https://doi.org/10.1108/JD-10-2015-0123>. Acesso em: 26 out.2018.

RUSBRIDGE, C. *et al.* The digital curation centre: a vision for digital curation. 2005 IEEE International Symposium on Mass Storage Systems and Technology. *Anais...* Sardinia, Italy, IEEE, 2005. p. 31-41. Disponível em: <https://doi.org/10.1109/LGDI.2005.1612461>. Acesso em: 26 out.2018

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação, Londrina*, v. 21, n. 2, p. 90-115, 2016. Disponível em: [www.brapci.inf.br/index.php/article/download/48979](http://www.brapci.inf.br/index.php/article/download/48979). Acesso em: 20 abr.2018.

SHINTAKU, M.; DUQUE, C.; SUAIDEN, E. J. Análise da adesão às tendências da Ciência pelos repositórios institucionais brasileiros. *InCID: Revista de Ciência da Informação e Documentação*, v. 6, n. 2, p. 148-169, 2015. Disponível em: <https://doi.org/10.11606/issn.2178-2075.v6i2p148-169>. Acesso em: 20 abr.2018

TAMMARO, A. M. Heritage curation in the digital age: Professional challenges and opportunities. *International Information & Library Review*, v. 48, n. 2, p. 122-128, 2016. Disponível em: <https://doi.org/10.1080/10572317.2016.1176454>. Acesso em: 26 Out.2018.

TENOPIR, C. *et al.* Academic librarians and research data services: preparation and attitudes. *IFLA Journal*, v. 39, n. 1, p. 70-78, 2013. Disponível em: <https://doi.org/10.1177/0340035212473089>. Acesso em: 26 Out.2018.

TIBBO, H. R.; HANK, C. Digital Data Curation Essentials for Data Scientists and Data Curators and Librarians. *In: 15th ACM/IEEE-CS Joint Conference on Digital Libraries, 2015, Knoxville, Tennessee, USA. Proceedings [...]* New York, NY, USA, ACM, 2015. p. 293-294. Disponível em: <https://doi.org/10.1145/2756406.2756928>. Acesso em: 26 Out.2018.

TRIPATHI, M.; SHUKLA, A.; SONKAR, S. K.. Research data Management Practices in university libraries: a study. *DESIDOC Journal of Library & Information Technology*, Dehli, v. 37, n. 6, 2017, p. 417-424. Disponível em: [https://www.researchgate.net/publication/321212270\\_Research\\_Data\\_Management\\_Practices\\_in\\_University\\_libraries\\_A\\_study/link/5a64776daca272a1581cf67f/download](https://www.researchgate.net/publication/321212270_Research_Data_Management_Practices_in_University_libraries_A_study/link/5a64776daca272a1581cf67f/download). Acesso em: 26 Ago.2018.

YAKEL, E. Digital curation. *OCLC Systems & Services: International digital library perspectives* 23, n. 4, 335-340, 2007. Disponível em: <https://doi.org/10.1108/10650750710831466>. Acesso em: 20 Abr.2018.