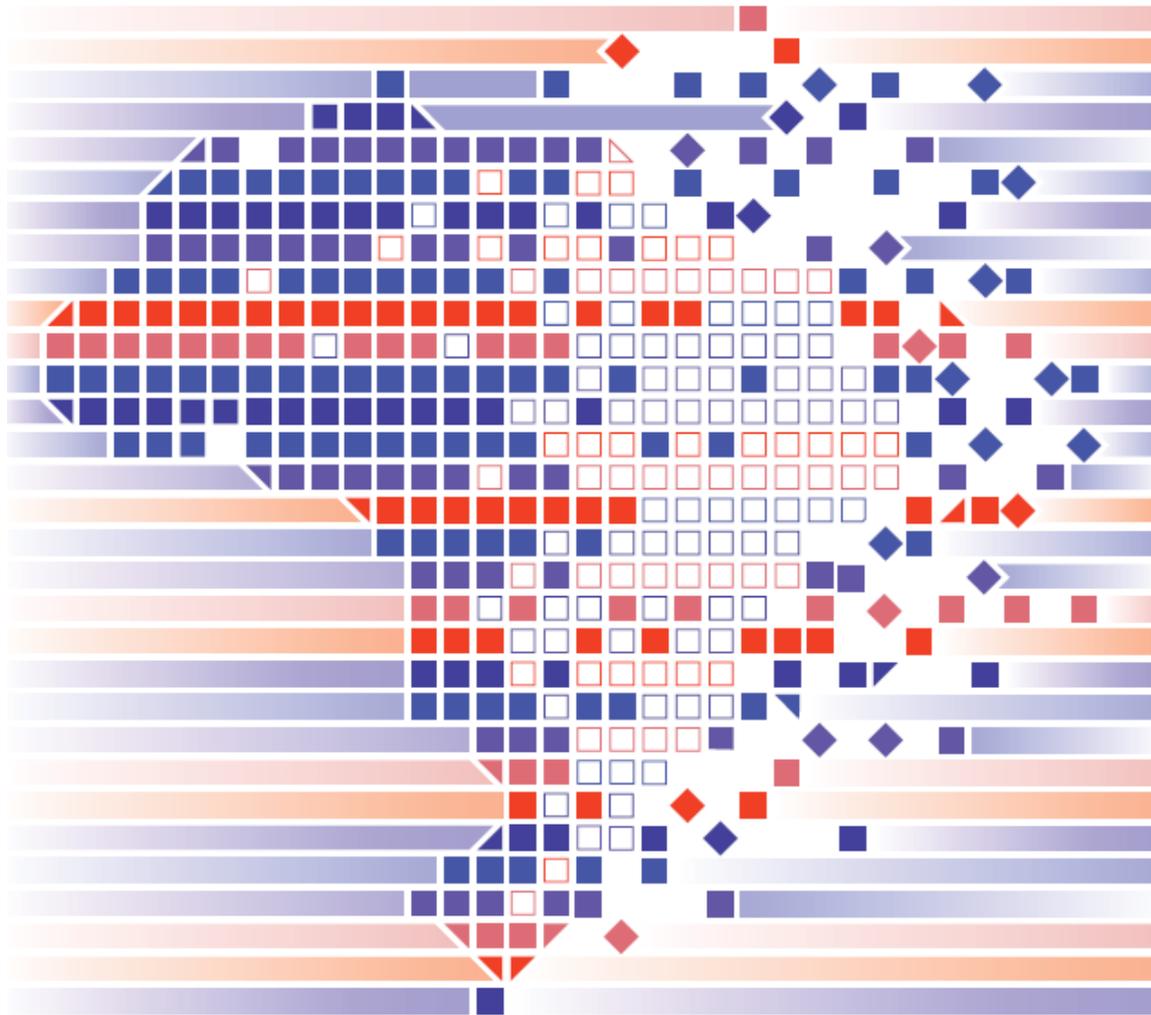


# ACESSO ABERTO A DADOS DE PESQUISA NO BRASIL

Avaliação e Seleção na Curadoria de Dados de Pesquisa



Coordenador

Rafael Port da Rocha (Universidade Federal do Rio Grande do Sul)

Coordenadora adjunta

Sônia Elisa Caregnato (Universidade Federal do Rio Grande do Sul)

Pesquisadores

Universidade Federal do Rio Grande do Sul

Ana Maria Mielniczuk de Moura

Caterina Groposo Pavão

Paula Caroline Schifino Jardim Passos

Rene Faustino Gabriel Junior

Samile Andréa de Souza Vanz

Universidade Federal de Rio Grande

Eduardo Nunes Borges

Adornete de Almeida Martins Junior

**Contato**

[dadosdepesquisa@rnp.br](mailto:dadosdepesquisa@rnp.br)

**Site**

<https://dadosdepesquisa.rnp.br>

**Endereço**

Rua Ramiro Barcelos, 2705 - Campus Saúde

Sala: 106 - Anexo 1

Brasil - Porto Alegre - RS - CEP 90.035-007

Telefone: +55(51)3308.5942

Este relatório é licenciado sobre a licença CC BY - Creative Commons Attribution 4.0 International License.

<http://creativecommons.org/licenses/by/4.0/>

Relatório disponível em: **xxxxx**

Como citar:

VANZ, Samile Andrea de Souza; MOURA, Ana Maria Mielniczuk de; PASSOS, Paula Caroline Schifino Jardim; CAREGNATO, Sônia Elisa; PAVÃO, Caterina Groposo; BORGES, Nunes Borges; ROCHA, Rafael Port da; GABRIEL JUNIOR, Rene Faustino; MARTINS JUNIOR, Adornete de Almeida. Avaliação e Seleção na Curadoria de Dados de Pesquisa. Disponível em: **xxxx**

V285 VANZ, Samile Andrea de Souza

Avaliação e Seleção na Curadoria de Dados de Pesquisa./ Samile Andrea de Souza Vanz et al. - Porto Alegre, RS : UFRGS, 2020.  
23 p.

Relatório sobre critérios de avaliação e seleção de dados de pesquisa para curadoria de repositórios de dados abertos, projeto RDP Brasil – Universidade Federal do Rio Grande do Sul; Universidade Federal do Rio Grande; Instituto Brasileiro de Informação em Ciência e Tecnologia;

1. Acesso a dados de pesquisa. 2. Curadoria de dados. I. Título

## Sumário

1	<b>Introdução</b>	5
2	<b>Curadoria de dados de pesquisa</b>	6
3	<b>Avaliação e seleção de dados de pesquisa</b>	9
4	<b>Políticas de avaliação e seleção</b>	11
5	<b>Crterios para seleção e avaliação no contexto brasileiro</b>	15
6	<b>Considerações finais</b>	21
	<b>Referências</b>	22

## **1 Introdução**

O presente relatório compõe um rol de resultados de pesquisa provenientes de um esforço conjunto entre a Rede Nacional de Ensino e Pesquisa (RNP), a Universidade Federal do Rio Grande do Sul (UFRGS) e a Universidade Federal do Rio Grande (FURG), que visa desenvolver pesquisas sobre compartilhamento dos dados coletados, gerados e utilizados pelos pesquisadores brasileiros. Este documento trata da curadoria de dados de pesquisa e das políticas de avaliação e seleção desses dados considerando-se o contexto de repositórios de dados de pesquisa nacionais. Como resultado deste estudo, propõem-se um checklist para avaliação e seleção de dados de pesquisa aplicado à realidade brasileira.

Entende-se que o estabelecimento de políticas para seleção e avaliação de dados de pesquisa pode apoiar os repositórios nacionais no sentido de orientar e justificar as decisões de aquisição, bem como contribuir para agilidade das atividades práticas relacionadas à coleta de dados. Ademais, o estabelecimento de diretrizes nesta área pode auxiliar no treinamento de profissionais em seleção e avaliação de dados e direcionar no desenvolvimento de novos repositórios.

## 2 Curadoria de Dados da Pesquisa

O compartilhamento e reuso de dados de pesquisa é uma prática que vem se consolidando há alguns anos entre a comunidade científica. O tema tem despertado a atenção de pesquisadores, editores de revistas, financiadores e agências de pesquisa e ainda, das instituições onde a pesquisa se desenvolve. A prática de compartilhamento e reuso requer duas frentes de trabalho. Primeiramente, a infraestrutura tecnológica necessária para garantir o ambiente para guarda dos dados - os repositórios. Em seguida, a definição das políticas de seleção, preservação, acesso e reuso destes dados. Neste âmbito estão atividades de curadoria.

A preservação de dados ao longo do tempo requer repositórios digitais confiáveis com governança sustentável e estrutura organizacional, infra-estrutura confiável e políticas abrangentes que deem apoio às práticas acordadas pela comunidade (LIN et al., 2020). Ao se pensar sobre as práticas científicas e as políticas dos repositórios, a decisão acerca do que manter em termos de dados de pesquisa tem sido reconhecida há algum tempo. O gerenciamento de dados de pesquisa e, em particular, a avaliação e seleção vem recebendo atenção nos últimos anos, à medida que crescem o volume e a diversidade de dados (BEAGRIE, 2019). Algumas questões surgem nesse contexto e devem ser respondidas: todos os conjuntos de dados devem ser guardados para sempre? Como saber o valor do dado tendo em vista a linha do tempo? Dados poderiam ser preservados por período determinados? Quais seriam os valores para determinar prazos de guarda? Os conjuntos de dados servem e estão preparados para o reuso e/ou reprodutibilidade? Os dados aceitos possuem qualidade suficiente para atender as necessidades de reuso? Como garantir qualidade ao dado? Preparar o dado para o compartilhamento é um processo caro. Os dados são suficientemente preparados para a replicação?

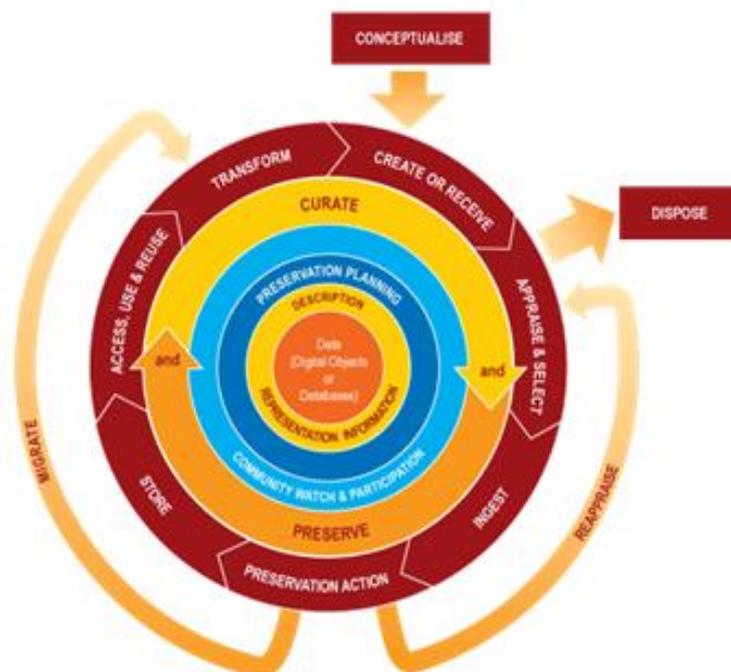
No contexto deste relatório, curadoria digital é compreendida de acordo com as ideias de Pennock (2007): "manter e agregar valor a um corpo confiável de informação digital para uso atual e futuro: em outras palavras, é o gerenciamento ativo e a avaliação das informações digitais durante todo o seu ciclo de vida."

Entende-se que curadoria digital, preservação digital e acesso relacionam-se da seguinte forma:

A curadoria expressa melhor a necessidade de atividades coordenadas de preservação e acesso aos ativos gerenciados. Embora a preservação e o acesso fossem anteriormente considerados funções díspares, agora são vistos corretamente como complementares: preservação destinada a fornecer acesso ao conteúdo gerenciado ao longo do tempo, enquanto o acesso depende da preservação em um determinado momento. A curadoria também conota melhor o enriquecimento contínuo de conteúdo gerenciado; isto é, a intenção de curadoria não é apenas manter o conteúdo originalmente adquirido, mas também agregar valor a ele. Todas essas atividades são relevantes durante todo o ciclo de vida digital, idealmente começando antes da criação de um ativo e durando até sua disposição final (CALIFORNIA DIGITAL LIBRARY, 2010)

Curadoria digital está associada ao ciclo de vida do objeto. O DCC Curation Center especifica ações para o ciclo de vida da preservação curadoria digital, classificando-as em ações sequenciais, ações ocasionais e ações envoltivas (sempre presentes, independentemente da fase), conforme descrito no quadro 1. A figura 1 apresenta graficamente, em ciclo, essas ações.

Figura 1 - Ciclo de Vida da Curadoria Digital



Fonte: DCC

Quadro 1 – Ações do Ciclo de Vida da Curadoria Digital do DCC

Tipo de ação	Ação
Ações sequenciais do ciclo	<ul style="list-style-type: none"> <li>● <b>Conceitualizar:</b> conceber e planejar a criação de dados, incluindo método de captura e opções de armazenamento</li> <li>● <b>Criar ou receber:</b> criar dados, incluindo metadados administrativos, descritivos, estruturais e técnicos.</li> <li>● <b>Avaliar e selecionar:</b> Avaliar dados e selecionar para curadoria e preservação a longo prazo.</li> <li>● <b>Ingestão:</b> transferir os dados para um arquivo, repositório, centro de dados ou outro custodiante. Aderir a orientações documentadas, políticas ou requisitos legais</li> <li>● <b>Ação de preservação:</b> Realizar ações para garantir a preservação e guarda a longo prazo dos dados.</li> <li>● <b>Armazenar:</b> armazenar os dados de maneira segura, seguindo os padrões relevantes.</li> <li>● <b>Acessar, usar e reutilizar:</b> garantir que os dados estejam diariamente acessíveis tanto para usuários designados quanto para reutilizadores.</li> <li>● <b>Transformar:</b> criar novos dados a partir do original</li> </ul>
Ações contínuas	<ul style="list-style-type: none"> <li>● <b>Descrever e representar a informação:</b> Atribuir metadados administrativos, descritivos, técnicos, estruturais e de preservação</li> <li>● <b>Planejar a preservação:</b> planejar a preservação durante todo o ciclo de vida da curadoria digital.</li> <li>● <b>Participar e vigiar a comunidade:</b> vigiar as atividades da comunidade e participar do desenvolvimento de padrões, ferramentas e software adequados e compartilhados</li> <li>● <b>Curar e preservar:</b> estar ciente e realizar ações administrativas e de gerenciamento planejadas para promover a curadoria e a preservação durante todo o ciclo de vida da curadoria</li> </ul>
Ações ocasionais	<ul style="list-style-type: none"> <li>● <b>Descartar:</b> descartar os dados que não foram selecionados para curadoria e preservação a longo prazo, de acordo com políticas, orientações ou requisitos legais documentados.</li> <li>● <b>Reavaliar:</b> Retornar dados que falhem nos procedimentos de validação para posterior nova avaliação e seleção</li> <li>● <b>Migrar:</b> migrar dados para um formato diferente.</li> </ul>

Fonte: DCC Curation Lifecycle

### 3 Avaliação e Seleção de Dados de Pesquisa

Nesta seção, serão abordados os conceitos de avaliação e seleção. O que os arquivistas chamam de 'avaliação' é frequentemente referido fora da profissão de arquivo como 'seleção' ou 'aquisição' e está intimamente ligado a um repositório ou política institucional de desenvolvimento de coleções. A avaliação é a função mais nobre, o núcleo central da prática arquivística contemporânea (WHYTE; WILSON, 2010), o que também é ressaltado por Couture (2005), que afirma que a avaliação é uma das funções mais importantes da prática arquivística contemporânea, e as decisões dela resultantes afetam outras atividades como o recolhimento, o arranjo, a descrição, a acessibilidade e a preservação dos documentos. Para o autor, a avaliação pode ser considerada como o ato de julgar os valores primários e secundários de documentos e estabelecer o período no qual eles retêm este valor, em um contexto que respeite as relações essenciais entre uma dada instituição (ou pessoa) e os documentos que eles criaram no curso de suas atividades (COUTURE, 2005, p. 83, tradução nossa). O Arquivo Nacional do Reino Unido (TNA) define avaliação como: “o processo de distinguir registros de valor contínuo daqueles sem valor adicional, para que este último possa ser eliminado”.

No ciclo de vida da curadoria digital do DCC, **avaliação e seleção** compreendem o processo de avaliar os materiais, para decidir quais manter a longo prazo, quais manter por um tempo, e quais descartar (HIGGINS, 2012). **Avaliar e selecionar** envolve adaptar para o mundo digital práticas de avaliação e seleção já desenvolvidas para acervos arquivísticos e bibliotecas.

Existe um consenso considerável sobre os critérios utilizados na avaliação e seleção de materiais não digitais. A prática de avaliação de arquivo usa o conceito de valor de arquivo, que é obtido considerando-se o valor administrativo (utilidade para a condução dos negócios), valor fiscal (utilidade para negócios financeiros), valor legal (digno para a realização de negócios legais), valor intrínseco (natureza inerente e significância do artefato), valor probatório (valor como registro das origens, funções e atividades do criador do registro) e valor informativo (utilidade do conteúdo para fins de pesquisa mais gerais) (TIBBO, 2003, p. 29-30).

Na opinião de Harvey (2007), a seleção de materiais para preservação a longo prazo em bibliotecas concentra-se na manutenção de itens físicos em seus formatos originais e aplica cinco critérios principais: valor probatório, valor estético,

valor de mercado, valor associativo e valor da exposição. O autor comenta que critérios adicionais podem ser aplicados, como condição física, recursos disponíveis, uso e significado social.

A possibilidade de armazenamento com custos mais baixos não diminuiu a importância das atividades de **avaliação e seleção**. Segundo Higgins (2012), é uma falácia que todo o material digital possa ser mantido sem a necessidade de avaliação porque o armazenamento é barato, e continua a se tornar cada vez mais barato. O autor pontua cinco razões principais para realizar a avaliação dos dados da pesquisa:

1. Reduz a quantidade de material que deve ser gerenciado ou custodiado a longo prazo, permitindo que os recursos possam ser direcionados para os materiais que possuem valor a longo prazo.
2. Facilita a capacidade de manter o acesso intelectual ao material, criar e anexar metadados, indexar e armazenar logicamente, para que os dados possam ser pesquisados e recuperados com rapidez e eficiência.
3. Garante que atividades de preservação possam ser realizadas de uma maneira simples e organizada para assegurar melhor a longevidade dos dados.
4. Limita os custos de armazenamento e manuseio do material.
5. Garante que as obrigações legais de armazenamento e acesso a dados sejam cumpridas.

A avaliação é o processo pelo qual alguns registros são selecionados para preservação, outros (grande maioria) são considerados de valor insuficiente para justificar a guarda permanente. A seleção deve ser guiada por políticas locais e comunitárias e por requisitos. O processo usado para decisões de seleção deve ser transparente e responsável, não pode ser baseado em visualizações individuais sobre possíveis necessidades futuras de pesquisa. Desta forma, comunidades de pesquisa e instituições precisam desenvolver e concordar com um conjunto de critérios objetivos para avaliar o significado dos conjuntos de dados de pesquisa a longo prazo.

#### **4 Políticas de Avaliação e Seleção**

Diversas instituições adotam políticas para a gestão dos dados de pesquisa por elas produzidos ou que estão sob sua custódia. No entanto, poucos desses planos especificam os parâmetros para estimar o valor dos conjuntos de dados e as ações a serem adotados no sentido de definir o tempo de sua preservação, reavaliação ou descarte, quando adequado.

Assim, a especificação de uma política para avaliação e seleção de dados de pesquisa tem se mostrado uma iniciativa necessária, mesmo frente à perspectiva tecnológica que aponta para a capacidade de armazenamento cada vez maior a um custo decrescente, o que permitiria manter todos os dados indefinidamente. Harvey (2007) adverte que essa abordagem se limita a preservação dos bits e que não é apropriada à preservação do conhecimento. Ele esclarece que não é prático ou mesmo desejável manter o acesso a todos os dados permanentemente e que é necessário manter um conjunto de dados com base em seu significado e valor permanente, para que se possa compreendê-los no futuro.

Além disso, o tamanho dos dados de pesquisa implica em uma tarefa de curadoria assustadora. Com os instrumentos científicos recentemente desenvolvidos e o crescente uso de simulações por computador, uma equipe de pesquisa pode gerar muitos terabytes de dados por dia. Os curadores de dados enfrentam o gerenciamento na magnitude de petabytes e muito além (NDSA Agenda Working Group, 2020).

Whyte e Wilson (2010) especificaram sete critérios gerais que embasam a derivação de critérios mais específicos em uma política de avaliação e seleção de dados de pesquisa. São eles:

- a) Relevância: o conteúdo do recurso atende às funções da instituição e cumpre os requisitos estabelecidos pela instituição de pesquisa ou pelo órgão de financiamento, incluindo qualquer aspecto legal para preservação dos dados além de seu uso imediato;
- b) Valor científico ou histórico: os dados são científica, social ou culturalmente importantes, ou seja, pode-se prever o uso futuro pelo valor da pesquisa;

- c) Exclusividade: o recurso é a única ou a mais completa fonte de informação sobre o tema e está em risco de perda se não for aceito;
- d) Potencial de redistribuição: a confiabilidade, a integridade e a usabilidade dos arquivos de dados podem ser determinadas, estes são recebidos em formatos que atendem aos critérios técnicos especificados e propriedade intelectual ou questões éticas são abordadas;
- e) Não replicabilidade: não seria viável replicar os dados ou a replicação seria financeiramente inviável;
- f) Viabilidade econômica: os custos para gerenciar e preservar o recurso podem ser estimados e são justificáveis a partir de evidências de possíveis benefícios futuros, além disso, o financiamento está previsto, quando apropriado;
- g) Documentação completa: as informações necessárias para facilitar a descoberta, acesso e reutilização são abrangentes e corretas; incluindo metadados sobre a proveniência do recurso e o contexto de sua criação e uso.

O UK Data Service, conforme consta no documento intitulado Collection Development Selection and Appraisal Criteria (UK Data Service, 2018), apresenta critérios semelhantes aos postulados por Whyte e Wilson (2010), acrescentando, entretanto, os seguintes itens:

- a) Novidade: os dados contêm informações que foram solicitadas pelos usuários ou representam lacuna na coleta atual;
- b) Valor internacional: os dados contêm informações de interesse da comunidade científica internacional.
- c) Necessidade de replicação: os dados e recursos representam resultados necessários para replicar ou revisar as pesquisas.

A esses sete critérios citados, pode ser acrescido por pelo menos mais um, anteriormente identificado por Harvey (2007), a vulnerabilidade. Ela é determinada pela necessidade de medidas especiais para ler ou acessar os dados ou pelas condições e idade da mídia na qual eles se encontram. Segundo Eaker (2016), os curadores do repositório precisam determinar se os custos adicionais para oferecer acesso diferenciado compensam os benefícios de ofertar aqueles dados.

O UK Data Service (2018) também apresenta critérios para não publicar os dados, que são os seguintes: a existência de questões legais e éticas como direitos autorais e proteção que impediriam o uso pleno dos dados; a falta de materiais contextuais que permitam a reutilização dos dados; e a aplicação de formatos antigos, que tornam as informações ilegíveis ou de difícil de conversão e recuperação.

Além de atender a critérios específicos para cada contexto, os dados de pesquisa poderão ser classificados em categorias que irão definir o tipo de curadoria conforme o tempo e o modo de disponibilização. Abaixo, apresenta-se cinco categorias de curadoria empregadas nas coleções de dados do UK Data Service, conforme consta no documento intitulado Collection Development Selection and Appraisal Criteria (UK Data Service, 2018).

Categoria 1: conjuntos de dados selecionados para curadoria de longo prazo e disponibilizados para acesso on-line ou download.

Categoria 2: conjuntos de dados selecionados para gerenciamento de curto prazo. Mesmo não sendo selecionados para preservação por longo prazo, eles receberão cópias de segurança (preservação em nível de bit), serão disponibilizados e poderão ser recuperados por meio de ferramentas de acesso on-line ou em software de repositório interno.

Categoria 3: conjuntos de dados selecionadas apenas para "entrega" aos usuários finais por meio de uma interface do repositório, consistindo, por exemplo, em dados de terceiros acessados por APIs / serviços da Web.

Categoria 4: conjuntos de dados selecionados apenas para "descoberta", ou seja, são dados de terceiros que não serão formalmente incluídos no repositório, mas

serão passíveis de recuperação via registro de metadados no repositório, permitindo que os dados sejam encontrados mais facilmente.

Categoria 5: esta categoria se refere apenas à preservação de conjuntos de dados que não estão no escopo do repositório. Os dados classificados nesta categoria poderão ser movidos para outras categorias, se necessário.

Os critérios acima elencados são norteadores, mas as políticas são implementadas de acordo com as necessidades locais. É necessário também considerar as partes envolvidas no processo de avaliação e seleção, assim como as necessidades dos produtores e dos consumidores e estabelecer as responsabilidades dos profissionais e das instituições.

Em suma, uma política de avaliação e seleção é fundamental para se ter um entendimento claro dos tipos de conjuntos de dados que podem fazer parte do repositório, dirimindo incertezas e estabelecendo limites e responsabilidades das partes envolvidas. Essa clareza ajuda também na determinação de como promover os serviços do repositório, identificando onde concentrar os esforços de divulgação (EAKER, 2016).

No entanto, mesmo políticas claras apresentam limites. Elas descrevem pontos-chave, por isso precisam ser complementadas por critérios específicos e procedimentos para sua implementação. Neste sentido, checklists são ferramentas que permitem operacionalizar as políticas, como tratado na próxima seção.

## **5 Seleção e avaliação em repositórios de dados brasileiros: proposta de checklist**

Um dos elementos fundamentais acerca da decisão de disponibilização e preservação de dados de pesquisa é a exigência por parte dos financiadores da pesquisa. No Brasil as agências de fomento tem implementado recomendações e algumas condições.

O CNPq informa em seus editais a responsabilidade dos pesquisadores, suas equipes e instituições em manter, sempre que possível, os resultados da pesquisa, dados e coleções à disposição de outros pesquisadores para fins acadêmicos. Os artigos científicos resultantes dos projetos apoiados pela agência deverão ser publicados, preferencialmente, em periódicos de acesso público e depositados, em conjunto com os dados científicos e com todo material suplementar relacionado, em repositórios eletrônicos de acesso público (CONSELHO..., 2020).

Na FAPESP o Plano de Gestão de Dados vem se tornando um componente obrigatório na fase de submissão de um projeto. Para determinadas modalidades e chamadas, o documento “Plano de Gestão de Dados” faz parte dos anexos obrigatórios de uma proposta submetida à FAPESP. O Plano reúne um conjunto de informações básicas – quais dados serão produzidos pelo projeto, restrições de compartilhamento, como serão compartilhados e como serão preservados (FAPESP, 2020). A FAPESP orienta a redação de um texto de até duas páginas, contendo as seguintes informações:

- a) Descrição dos dados e metadados produzidos pelo projeto - por exemplo, amostras, registros de coleta, formulários, modelos, resultados experimentais, software, gráficos, mapas, vídeos, planilhas, gravações de áudio, bancos de dados, material didático e outros.
- b) Quando aplicável, restrições legais ou éticas para compartilhamento de tais dados, políticas para garantir a privacidade, confidencialidade, segurança, propriedade intelectual e outros.
- c) Política de preservação e compartilhamento (por exemplo, compartilhamento imediato ou apenas após a aceitação da publicação associada). Período de carência (antes do compartilhamento) e período durante o qual os dados serão preservados e disponibilizados.

d) Descrição de mecanismos, formatos e padrões para armazenar tais itens de forma a torná-los acessíveis por terceiros. Esta descrição pode incluir o uso de repositórios e serviços de outras instituições.

O checklist apresentado no Quadro 2 tem por base os documentos do Digital Curation Centre and Australian National Data Service Guide, o UK Data Service Collection Development Selection and Appraisal Criteria, o NERC Data Value Checklist, o DCC Checklist of Appraising Research Data, o Cambridge PrePARe Project, e o University of Bristol Research Data Evaluation Guide. O checklist está organizado em critérios obrigatórios e critérios importantes. Orienta-se que, quando houver resposta positiva a pelo menos um dos critérios obrigatórios os dados sejam selecionados para preservação. Em relação aos critérios importantes, orienta-se que quando for possível responder "Sim" a pelo menos uma das perguntas de cada seção os dados provavelmente sejam encaminhados para seleção para preservação. Os demais critérios são considerados de suporte (itens origem, condições, Requisitos de armazenamento e preservação, limitações de acesso, Limitações técnicas). Nesses casos, a seleção e preservação ocorrerá caso seja possível responder "Sim" à maioria das perguntas.

Quadro 2: Checklist para seleção e avaliação de dados de pesquisa  
aplicado à realidade brasileira

**Critérios obrigatórios**

Responder "Sim" a qualquer uma das perguntas abaixo resulta automaticamente na seleção para preservação

**Considerações legais / estatutárias****Sim****Não**

Existe uma razão legal para reter os dados?

Existe alguma razão para acreditar que os dados possam ser usados em litígios, inquéritos públicos, investigações policiais, ou qualquer relatório ou artigo que possa ser legalmente contestado?

Os dados são o produto do financiamento de agências públicas e sustentam uma produção de pesquisa publicada?

Existem obrigações contratuais que exigem a preservação dos dados?

**Políticas****Sim****Não**

A política de Dados de Pesquisa do financiador de pesquisa exige que os dados sejam mantidos?

Os dados serão citados em uma publicação com uma política que exige que os dados sejam disponibilizados?

Aplica-se alguma orientação específica da disciplina que solicite a preservação dos dados?

**Critérios importantes**

Responder "Sim" a pelo menos uma das perguntas de cada seção abaixo provavelmente deve resultar em uma seleção para preservação.

**Valor de reutilização****Sim****Não**

O arquivo não é uma cópia de dados disponíveis em um repositório. Os dados são únicos e / ou impossíveis de reproduzir, como por exemplo, registram eventos que não podem ser recriados?

Os dados têm apelo amplo e é provável que sejam do interesse de outras pessoas (por exemplo, uma ampla faixa geográfica ou temporal ou um foco interdisciplinar)?		
É provável que os dados tenham valor acadêmico especial (por exemplo, representam uma descoberta histórica) ou estabelecem um novo precedente importante que provavelmente será seguido por outros (por exemplo, envolve uma nova técnica de processamento de dados)?		
<b>Contexto da pesquisa</b>	<b>Sim</b>	<b>Não</b>
É provável que os dados sejam citados / referenciados em uma publicação acadêmica?		
Os dados agregam valor a alguma coleta de dados significativa, contribuindo para uma coleção preexistente?		
Os dados apoiam projetos e estão alinhados com as tendências atuais de pesquisa?		
É provável que os dados se alinhem fortemente às tendências futuras de pesquisa? Isso deve ser inferido, com base em evidências do valor atual, como as taxas de citação existentes.		
<b>Critérios de suporte</b> Responder "Sim" à maioria das perguntas abaixo deve resultar em seleção para preservação.		
<b>Origem</b>	<b>Sim</b>	<b>Não</b>
A reprodução dos dados é cara ou difícil?		
Os dados têm sua integridade original? (por exemplo, não processado e foi armazenado com segurança desde que foi gerado)		
O arquivo é a cópia de referência (definitiva) dos dados?		
<b>Condições</b>	<b>Sim</b>	<b>Não</b>

Os dados possuem metadados suficientes? (por exemplo, uma descrição no nível do catálogo, uma descrição de como os dados são organizados, documentação de como e por que os dados foram criados e um guia sobre como usá-los)		
Os dados são de qualidade adequada para depósito em um Data Center ou outro repositório? (ou seja, os dados são controlados pela qualidade, bem organizados, legíveis e íntegros)		
Existem, proporcionalmente, dados mais valiosos		
<b>Requisitos de armazenamento e preservação</b>	<b>Sim</b>	<b>Não</b>
Os dados podem ser armazenados (ou seja, arquivados) sem requisitos excepcionais?		
Os dados podem ser preservados de uma forma utilizável (isto é, permanecem adequados à finalidade) sem quaisquer requisitos excepcionais?[SV1]		
Existe financiamento para financiar a preservação (pela equipe de pesquisa, por uma instituição anfitriã ou por um data center) desses dados em particular?		
<b>Limitações de acesso</b>	<b>Sim</b>	<b>Não</b>
Se houver dados pessoais envolvidos, foi obtido o consentimento informado dos sujeitos da pesquisa para arquivamento e reutilização de dados?		
Se for necessária a aprovação de um Comitê de Ética, há evidências de que este procedimento foi seguido?		
A natureza dos dados sugere restrições ao compartilhamento, acesso e reutilização? (por exemplo, conjunto de dados envolve dados políticos ou de saúde confidenciais)		

Os dados estão livres de quaisquer termos e condições que limitariam o acesso? (por exemplo, requisitos de licença de banco de dados, acordos comerciais que proíbem a reutilização)		
<b>Limitações técnicas</b>	<b>Sim</b>	<b>Não</b>
Os dados estão em um formato técnico aceitável para depósito em um data center?		
Os dados são utilizáveis sem nenhum software / hardware especializado?		
Se "Não" à pergunta acima, o software / hardware especialista necessário está prontamente disponível?		
É possível gerar versões diferentes dos dados para aumentar o valor de reutilização (por exemplo, criar formatos de arquivo alternativos)?		

Fonte: Elaborado pelos autores com base nos checklists do Digital Curation Centre and Australian National Data Service Guide, UK Data Service Collection Development Selection and Appraisal Criteria, NERC Data Value Checklist, DCC Checklist of Appraising Research Data, Cambridge PrePARE Project, University of Bristol Research Data Evaluation Guide.

A aplicação desse checklist deverá oferecer uma base sólida para a decisão da equipe responsável pela curadoria de dados nos repositórios de dados abertos.

## **6 Considerações finais**

As instituições de pesquisa e universidades tem trabalhado na criação de normativas para seus pesquisadores e repositórios para depósito e preservação dos dados. O desenvolvimento destes repositórios é um processo técnico, no entanto, seu uso envolve decisões importantes sobre a guarda e preservação dos dados gerados no âmbito do repositório. Para apoiar estas decisões, diversas instituições internacionais têm disponibilizado aos seus pesquisadores e gestores checklists para seleção e avaliação de dados, de modo a facilitar a interpretação e a definição acerca da guarda de um determinado conjunto de dados. Nesse sentido, propõe-se um checklist baseado em documentos internacionais, que identifica requisitos importantes ao contexto de pesquisa brasileiro.

Entre os checklists analisados e que serviram de base para a presente proposta está o do Digital Curation Centre and Australian National Data Service Guide, o UK Data Service Collection Development Selection and Appraisal Criteria, o NERC Data Value Checklist, o DCC Checklist of Appraising Research Data, o Cambridge PrePARE Project, e o University of Bristol Research Data Evaluation Guide. Os documentos internacionais direcionaram a proposta que ora se apresenta, que sugere-se, seja avaliada após ser utilizada por repositórios brasileiros temáticos, multidisciplinares e institucionais.

## Referências

BEAGRIE, NEIL. **What to keep**: a JISC research data study. 2019. 64 p.

CALIFORNIA DIGITAL LIBRARY. UC3Curation Foundations. UC Curation Center / California Digital Library, 2010. Disponível em: <https://confluence.ucop.edu/download/attachments/13860983/UC3-Foundations-latest.pdf?version=1>

COLE, Gareth; EVANS, Jill; LLOYD-JONES, Hannah. **Selecting data**: what to keep, what to delete? Exeter: University of Exeter, 2013. Disponível em: <http://hdl.handle.net/10036/4427>

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. 2020. Disponível em: <http://dadosabertos.cnpq.br/> Acesso em: 08 jun. 2020.

EAKER, C. Selection and appraisal of digital research datasets. In Kellam, L. & Thompson, K. (Eds.). *Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago: American Library Association, 2016. Disponível em: [https://trace.tennessee.edu/cgi/viewcontent.cgi?article=1018&context=utk\\_libpub](https://trace.tennessee.edu/cgi/viewcontent.cgi?article=1018&context=utk_libpub)

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. Plano de Gestão de dados. 2020. Disponível em: <http://www.fapesp.br/gestaodedados/> Acesso em: 08 jun. 2020.

HARVEY, R. Appraisal and Selection. In: S.Ross, M.Day (eds). **DCC Digital Curation Manual**. 2007. Disponível em: <http://www.dcc.ac.uk/resource/curation-manual/chapters/appraisal-and-selection>

HIGGINS, Sarah. Lifecycle of data management. In: Graham Pryor. **Managing Research Data**. London: Facet Publishing, 2012.

LIN, Dawei et al. The TRUST Principles for digital repositories. **Scientific Data**, v. 7, n. 144, 2020.

NATURAL ENVIRONMENT RESEARCH COUNCIL. NERC Data Value Checklist. [2019]. 3 p. Disponível em: <https://nerc.ukri.org/research/sites/data/policy/data-value-checklist/>

NDSA Agenda Working Group. 2020 NDSA Agenda. Charlottesville, Center for Open Science, 2020. DOI 10.17605/OSF.IO/BCETD.

PENNOCK, Maureen. *Digital Curation: a Life-Cycle Approach to Managing and Preserving Usable Digital Information*, 2007. Disponível em [http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch\\_curation.pdf](http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf)

PrePARE Project Team. **Selecting what to keep and what to bin (Checklist)**. Cambridge: Cambridge University Library, 2012. 1 p. Disponível em: <http://www.dspace.cam.ac.uk/handle/1810/243754>

UK DATA SERVICE. Collections Development Selection and Appraisal Criteria, 2018. Disponível em: <https://www.ukdataservice.ac.uk/media/455175/cd234-collections-appraisal.pdf> . Acesso em: 20 jun. 2020.

UNIVERSITY OF BRISTOL. Research Data Service. **Research Data Evaluation Guide**. Bristol: 2018. 6 p. Disponível em: <http://www.bristol.ac.uk/staff/researchers/data/writing-a-data-management-plan/>

WHYTE, A.; WILSON, A. How to Appraise & Select Research Data for Curation. Edinburgh, DCC, 2010. disponível em: <http://www.dcc.ac.uk/resources/how-guides/appraise-select-research-data>