



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

50 anos de Ciência da Informação no Brasil:
diversidade, saberes e transformação social

Rio de Janeiro • 25 a 29 de outubro de 2021

XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

GT-8 – Informação e Tecnologia

O PROCESSO DE AGREGAÇÃO DE VALOR AOS DADOS E A CURADORIA DIGITAL

THE PROCESS OF AGGREGATION VALUE TO DATA AND DIGITAL CURATION

Maria Lígia Triques - Universidade Federal de São Carlos - (UFSCar)

Ana Carolina Simionato Arakaki - Universidade Federal de São Carlos - (UFSCar)

Modalidade: Trabalho Completo

Resumo: A curadoria digital tem nos metadados importantes aliados para o funcionamento de suas ações, sem os quais não poderia realizar a manutenção, a preservação e, sobretudo, o processo de agregar valor aos dados. Contudo, o constante aumento do volume e da complexidade dos dados em seus contextos de produção, análise e uso intensivos impõem constantemente novos aspectos a serem considerados que exigem o aprimoramento do planejamento e das estruturas de metadados. Nesse sentido, buscou-se refletir como o uso de metadados, orientados por tecnologias, padrões e princípios, influenciam o processo da curadoria digital de agregar valor aos dados. Assim, realizou-se uma pesquisa exploratória e qualitativa na literatura científica e documental, nacional e internacional, utilizando como método a Análise de Conteúdo, viabilizando a comparação entre as tecnologias, padrões e princípios - que codificam e processam metadados para os sistemas computacionais - e a documentação referente ao *Extended Digital Curation Lifecycle Model* (DCC&U). Os resultados demonstraram que o uso estratégico dos metadados a partir de tecnologias, padrões e princípios possibilitam agregar valor aos dados, pois ampliam o acesso persistente, a verificação de autenticidade, a proveniência dos dados, a experiência de busca e a recuperação para os usuários permitindo multidimensionar o conteúdo disponibilizado em ambientes digitais. Portanto, destaca-se que a curadoria digital acompanha a evolução das soluções tecnológicas para garantir a sustentabilidade e a interoperabilidade frente à obsolescência tecnológica e às mudanças contextuais.

Palavras-Chave: Curadoria digital; Metadados; *Extended Digital Curation Lifecycle Model*; DCC&U.

Abstract: *Metadata are important allies for digital curation actions functioning, without which maintenance, preservation and, mainly, the process of adding value to data would not be possible. However, the constant increase in data volume and complexity in its intensive production, analysis and use contexts constantly impose new aspects to be considered, which require metadata planning and structures improvement. In this sense, it was sought to reflect how the metadata use, guided by technologies, standards and principles influence the digital curation process of data value adding. Thus, an exploratory and qualitative research was carried out in the national and international, scientific and documentary literature, using Content Analysis as a method, enabling the comparison between technologies, standards and principles - that allow metadata coding and processing for the computer systems - and documentation related to the Extended Digital Curation Lifecycle Model*

(DCC&U). The results demonstrated that metadata strategic use based on technologies, standards and principles make it possible to add value to the data, as they expand persistent access, verification of authenticity, the origin of the data, search, and retrieval experience for users, allowing multidimensionality of the content made available in digital environments. Therefore, digital curation follows the evolution of technological solutions to guarantee sustainability and interoperability against technological obsolescence and contextual changes.

Keywords: Digital curation; Metadata; Extended Digital Curation Lifecycle Model; DCC&U.

1 INTRODUÇÃO

Na medida em que os sistemas informacionais se tornam cada vez mais orientados ao processamento e a troca de informações semanticamente, os característicos processos desempenhados nos ambientes informacionais, como geração, armazenamento, preservação, organização, representação, compartilhamento, disseminação, acesso, precisam desenvolver-se de modo estratégico junto com as soluções tecnológicas, ampliando a importância do planejamento e da estruturação padronizada dos dados.

Identifica-se a curadoria digital como uma importante abordagem para os desafios de gerenciamento de dados, tendo em vista que suas ações envolvem além da manutenção e da preservação dos recursos informacionais, o processo de agregar valor aos dados. Nesse processo, os metadados são elementos centrais, possibilitando assegurar características como autenticidade, a confiabilidade, a integridade e a usabilidade dos dados para todo o seu ciclo de vida (HIGGINS, 2008; GILLILAND, 2016), utilizando para isso, tecnologias, padrões e princípios que codificam e processam os metadados para os sistemas computacionais.

Entende-se que a curadoria digital tem como foco além do armazenamento e preservação, a sustentabilidade e a interoperabilidade dos dados para os diferentes contextos de uso. Tanto a sustentabilidade, quanto a interoperabilidade podem ser consideradas questões-chave para os sistemas de informação que lidam com grandes volumes de dados heterogêneos e distribuídos em rede. Assim, considerando os desafios em ambientes digitais, questiona-se como o uso de metadados, pautados por tecnologias, padrões e princípios influenciam o processo de agregar valor na curadoria digital.

Este trabalho apresenta os resultados de um estudo qualitativo, que buscou discutir o papel dos metadados em uma das abordagens proposta para a curadoria digital, o *Extended Digital Curation Lifecycle Model* (DCC&U), desenvolvido a partir de um dos modelos referência para a área de Ciência da Informação, o *Digital Curation Centre Lifecycle Model* (DCC). Para o desenvolvimento do estudo, realizou-se uma pesquisa exploratória na literatura

científica e documental, nacional e internacional, utilizando como método, a análise de conteúdo, que é definida segundo Bardin (2011) como conjunto de técnicas de análise do conteúdo das comunicações, utilizando procedimentos sistemáticos e objetivos para a sua descrição, tendo em vista a inferência de conhecimentos, por meio de indicadores, quantitativos ou não. O método consiste em três fases: a) Pré-análise - fase inicial de organização, na qual foram escolhidos os critérios e os documentos que serviram para a análise e interpretação dos resultados; b) Exploração do material - fase na qual buscou-se sistematizar e aplicar as decisões definidas durante a pré-análise; c) Tratamento dos dados, inferência e interpretação - fase final, na qual os dados foram validados e comparados, possibilitando a inferência e interpretação dos resultados e das discussões (BARDIN, 2011).

Para o desenvolvimento, utilizou-se a literatura disponível sobre o tema a partir do Portal de Periódicos CAPES, da Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), da *Scientific Electronic Library Online* (SciELO) e do *Google Scholar*. A escolha das publicações permitiu determinar o universo de investigação, definindo as publicações a serem utilizadas e submetidas aos procedimentos analíticos. Desse modo, foi estabelecido um *corpus* de análise, considerando as quatro regras definidas por Bardin (2011) - exaustividade; representatividade; homogeneidade; e pertinência - que permitiram selecionar o material que efetivamente foi analisado em função das relações existentes entre os fundamentos das publicações e o objetivo proposto.

2 AGREGANDO VALOR AOS DADOS NA CURADORIA DIGITAL

Um dos principais objetivos da curadoria digital é proporcionar um conjunto de diretrizes que permitam manter os dados para uso atual e futuro, de modo que eles possam não só se manter acessíveis, mas com qualidade e confiabilidade para serem reaproveitados e reutilizados em diversos contextos de uso. Tal objetivo não é algo simples, pois se ramifica em diversos objetivos específicos. Por isso sua abordagem em geral leva em consideração a constante manutenção de todos os estágios pelos quais os dados passam em sua vida útil.

É comum encontrar a definição de que a curadoria digital envolve mais do que o armazenamento seguro e a preservação dos dados, uma vez que busca agregar valor, aumentando a sua utilidade nos diferentes ambientes digitais (RUSBRIDGE *et al*, 2005; BEAGRIE, 2006; PENNOCK, 2007; SALES; SAYÃO, 2012; OLIVER; HARVEY, 2016; CHOUDHURY; HUANG; PALMER, 2020). Por mais de uma década, percebe-se que os estudos a respeito da

curadoria digital buscaram aprimorar as formas de agregar valor. A definição de agregar valor a algo, seja, um produto, serviço ou recurso, é subjetiva e está intimamente ligada com a aceitabilidade dos agentes envolvidos (cliente/usuário) e com o contexto de sua utilização (CSILLAG, 1995). Contudo, tal processo ainda permanece um desafio diante do constante aumento do volume e da complexidade dos dados e dos seus contextos de produção, análise e uso intensivos (HUANG; LEE; PALMER, 2020). Por isso, as ações da curadoria digital, para agregarem valor, precisam de planejamento e estruturas adequadas pautadas em tecnologias, padrões e princípios.

A curadoria digital tem por base as tecnologias, mais especificamente as tecnologias que têm sido consideradas a chave para o processamento de dados na *Web*, as chamadas tecnologias semânticas. Elas são responsáveis por permitir que a *Web* atinja um maior nível de interação semântica entre suas funções (BERNERS-LEE; HENDLER; LASSILA, 2001). Entre elas, destacam-se o *Uniform Resource Identifier* (URI), o *Extensible Markup Language* (XML), *Resource Description Framework* (RDF), *Resource Description Framework Schema* (RDF-S), *Simple Knowledge Organization System* (SKOS), *Ontology Web Language* (OWL), entre outras.

Desse modo, esforços foram feitos no sentido de ampliar o processamento de dados legíveis para máquinas a partir da implementação e do aprimoramento de soluções que permitissem um maior nível de estruturação semântica às funções sintáticas (VIDOTTI *et al*, 2019), utilizando, para isso, um maior número de metadados. Os metadados geralmente estão na forma de soluções como ontologias ou linguagens formais padronizadas com uma semântica baseada em lógica que admite raciocínio sobre o significado dos dados (HITZLER, 2021), e possibilitam a utilização e reutilização dos dados pelos princípios *Linked Data* e *Linked Open Data*, desenvolvidos como parte da proposta da *Web Semântica*, para conectar dados na *Web* (BAKER *et al*, 2011).

Ainda atualmente, grande parte da *Web* não é propriamente semântica, com dados bem definidos e ligados de forma inteiramente legível para as máquinas (HITZLER, 2021). Apesar disso, as tecnologias disponíveis já permitem evoluir para uma quarta geração da *Web*, que tem sido chamada de *Web Pragmática* e refere-se ao uso de sistemas computacionais mais eficientes e assertivos com a interação humana (VIDOTTI *et al*, 2019). Ou seja, a *Web* se encontra em uma confluência de fases, na qual grande parte dos dados ainda estão configurados como na *Web Sintática*. Mesmo com os avanços tecnológicos,

agregar semântica aos dados ainda permanece um dos grandes desafios não só para a curadoria digital, mas para todas as abordagens que direta ou indiretamente lidam com dados. Uma das principais estruturas para apoiar a construção de redes de relacionamentos semânticos, são as ontologias, que consistem em modelos formais de entidades e relacionamento pertinentes a um domínio do conhecimento (DOERR, 2008). Campos (2009, p. 10), explica que as relações semânticas ao serem explicitadas, “[...] eliminam interpretações dúbias: dois conceitos podem se relacionar de muitas formas diferentes, mas ao explicitar a relação pretendida, o conhecimento é comunicado.” Isso se torna particularmente importante para a curadoria digital por permitir apoiar seus processos a partir da extração e integração de conhecimentos e do raciocínio automatizado (DOERR, 2008).

Além das ontologias, modelos de dados como o RDF proporcionaram a ampliação do uso dos padrões de metadados, com o desenvolvimento dos perfis de aplicação, que reúnem um conjunto de metadados selecionados de um ou mais esquemas a um esquema comum, permitindo combiná-los para adaptarem-se às necessidades específicas (ZENG; QIN, 2016).

Em conjunto com as tecnologias e padrões que são a base das ações da curadoria digital, o emprego de metadados, de qualidade, padronizados e estruturados depende de um planejamento cuidadoso e princípios sólidos (ZENG; QIN, 2016). Tais princípios correspondem, de acordo com Duval *et al* (2002), à modularidade, à extensibilidade, ao refinamento e ao multilinguismo, como apresentado no quadro 1, e estão intimamente ligados à sustentabilidade e à interoperabilidade dos dados.

Quadro 1: Princípios dos metadados

Modularidade	Capacidade de combinação sintática e semântica dos metadados de maneira interoperável de modo que novas estruturas possam ser criadas com base em padrões já estabelecidos.
Extensibilidade	Capacidade de estender o número de metadados conforme as necessidades locais específicas do domínio para a inclusão de novos elementos sem comprometer os já existentes.
Refinamento	Capacidade de escolher o nível de descrição apropriado e suficiente para cada domínio, considerando a inclusão de qualificadores e a escolha de padrões de valores.
Multilinguismo	Capacidade de incluir os aspectos da diversidade sociolinguística na padronização dos metadados, contemplando necessidades gerais e locais por meio de um esquema global.

Fonte: Elaborado pelas autoras com base em Duval *et al* (2002).

Cada princípio refere-se a importantes requisitos a serem considerados para o efetivo funcionamento dos sistemas informacionais digitais. Zeng e Qin (2016, p. 28, tradução nossa) destacam que tais princípios visam que as implementações de metadados sejam “[...] tanto

sustentáveis para a utilização e preservação a longo prazo, quanto interoperáveis para compartilhamento e reutilização”.

A sustentabilidade pode ser entendida como “[...] a habilidade de identificar os recursos necessários para manter o valor de um objeto digital ou de um serviço para os usuários aos quais se destinam.” (MARON; YUN; PICKLE, 2013, p. 12, tradução nossa). Por isso, ela está relacionada com as ações que mantêm a relevância de um recurso para uso futuro. Segundo Zeng e Qin (2016), ter boa qualidade de metadados vinculáveis para suportar as necessidades de gerenciamento, recuperação, navegação, descoberta, uso e reutilização de recursos é fundamental para o efetivo e sustentável acesso às informações. A mesma lógica vale em relação à interoperabilidade, que é a capacidade de vários sistemas comunicarem-se de forma eficaz, o que exige metadados de boa qualidade que possam suportar os processos de mapeamento e busca de dados em diferentes sistemas.

Desde a primeira vez em que o termo curadoria digital foi usado em 2001 (BEAGRIE, 2006), muitos estudos buscaram definir e caracterizar sua prática. Como resultado, diversas abordagens em forma de modelos gráficos foram desenvolvidas para tentar representar a curadoria digital. Mais recentemente, uma revisão de literatura feita por Huang, Lee e Palmer (2020) analisou e comparou vários modelos desenvolvidos, incluindo os de *big data* e de ciclo de vida de dados de pesquisa científica.

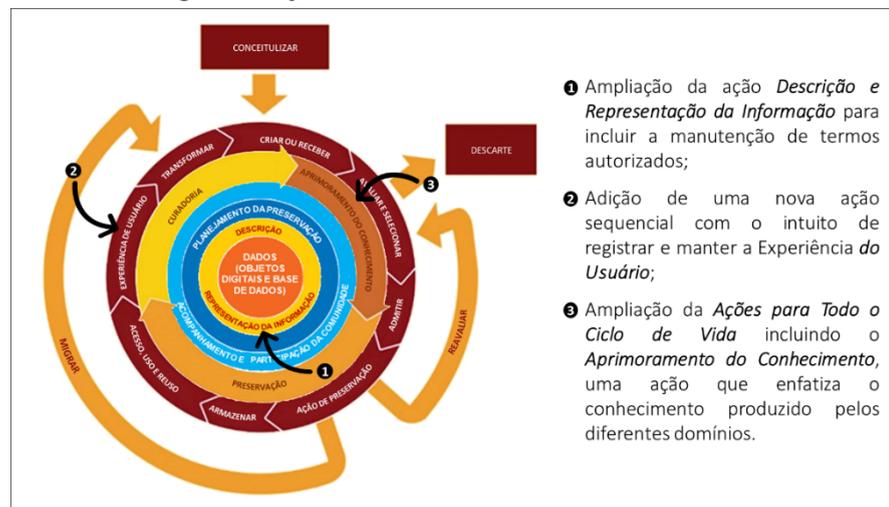
Como é possível observar, alguns modelos deram origem a estruturas genéricas, de tal modo a orientar a prática e os objetivos da curadoria digital para diferentes contextos de uso, como é o caso do *DCC Curation Lifecycle Model*, modelo elaborado pelo *Digital Curation Centre* (DCC), que é referência para o campo e fornece uma visão geral gráfica e de alto nível dos estágios necessários para empreender a curadoria digital (HIGGINS, 2008). Esse modelo apresenta uma abordagem em que as ações estão alinhadas aos estágios do ciclo de vida dos dados, dando ênfase, desse modo, aos metadados (HIGGINS, 2008; OLIVER; HARVEY, 2016).

Diante disso, uma outra proposta de abordagem de curadoria digital buscou dar maior ênfase aos aspectos relacionados à representação dos dados. Desenvolvida pela *Digital Curation Unit* (DCU), vinculado ao *Athena Research Center*, na Grécia, a proposta defendia a necessidade de um estudo mais aprofundado das práticas de curadoria digital voltadas à representação de modo a proporcionar a contínua atualização e enriquecimento dos dados, considerando os diferentes atores envolvidos no processo e as diferenças

disciplinares e contextuais (CONSTANTOPOULOS; DALLAS, 2008; CONSTANTOPOULOS *et al*, 2009).

Como resultado, os pesquisadores do DCU sugeriram um aprimoramento das ações já definidas pelo modelo do DCC, desenvolvendo o *Extended Digital Curation Lifecycle Model* (DCC&U), a partir da união entre o modelo DCC e um modelo próprio do DCU. O DCC&U leva em consideração aspectos que não foram considerados no modelo original, incluindo três propostas, indicadas na figura 1.

Figura 1: Ações acrescentadas no modelo DCC&U



Fonte: Elaborada pelas autoras com base em Constantopoulos *et al* (2009).

A primeira proposta considera que os usuários estão mais ativos, criando e modificando objetos digitais, formando comunidades e interagindo entre si. A segunda proposta se refere à manutenção de termos autorizados, convencionados para denotar conceitos, propriedades e relações relevantes ou instâncias, uma vez que parte do conhecimento especializado sobre o domínio é representado por eles, os quais evoluem ao longo do tempo (CONSTANTOPOULOS *et al*, 2009, p.41, tradução nossa). A terceira e última proposta enfatiza o aprimoramento do conhecimento como um processo fundamental. Isto é, o conhecimento produzido tanto pela pesquisa científica, como pela prática profissional, gera constantemente novos conhecimentos que incidem sobre as entidades, situações e eventos do mundo real representados nos objetos digitais (CONSTANTOPOULOS *et al*, 2009), que, por sua vez, implica diretamente nos processos de abstração e conceitualização.

Tanto o modelo do DCC quanto a sua versão alternativa, o DCC&U, são abordagens “alto nível”, ou seja, que servem como diretrizes para orientar o processo, mas que na prática, acabam sendo adaptadas dependendo das necessidades e dos recursos tecnológicos

e humanos disponíveis. Portanto, é importante considerar as mudanças no contexto para discutir a curadoria digital em quaisquer de suas dimensões de estudo.

3 CURADORIA DIGITAL E O USO DAS TECNOLOGIAS, PADRÕES E PRINCÍPIOS

Para a última etapa do estudo buscou-se a interpretação e desenvolvimento dos resultados e discussões a partir da comparação entre as tecnologias, os padrões, os princípios discutidos e os estudos acerca do *Extended Digital Curation Lifecycle Model* (DCC&U). Com ênfase no contexto e no relacionamento entre os dados, a configuração modelo do DCC&U permite discutir como cada etapa do ciclo de vida de curadoria digital, apresentadas no quadro 2, é influenciada pelos metadados.

Quadro 2: Etapas do *Extended Digital Curation Lifecycle Model* (DCC&U)

Ações para todo o ciclo de vida	Descrição e Representação da Informação; Planejamento da Preservação; Acompanhamento e Participação da Comunidade; Curadoria, Preservação e Aprimoramento do Conhecimento.
Ações sequenciais	Conceitualizar; Criar ou Receber; Avaliar e Selecionar; Admitir; Ações de Preservação; Armazenar; Acesso, Uso e Reuso; Experiência do Usuário; Transformar.
Ações ocasionais	Descartar; Reavaliar; Migrar.

Fonte: Elaborada pelas autoras com base em Constantopoulos *et al* (2009).

De acordo com Rusbridge *et al* (2005), para alcançar além do uso e preservação momentâneos, a curadoria digital busca agregar valor por meio do contexto e da ligação entre os dados, dando enfoque para a criação e a publicação de dados de maneira a facilitar a reutilização e promover integridade e a integração. Isso ocorre em cada uma de suas ações.

Iniciando pelas 'Ações para todo o ciclo de vida', as quatro ações que dela fazem parte permeiam todo o ciclo de vida, influenciando no planejamento, na coordenação e na execução de todas as etapas de forma contínua. A primeira ação, 'Descrição e a representação da informação', envolve a atribuição de elementos que permitem identificar, entender, controlar e renderizar os dados. Esses elementos são os próprios metadados, que assumem diferentes formas para atender aos requisitos dos sistemas. Além dos vários tipos de padrões genéricos que oferecem funcionalidade que pode ser usada por diferentes domínios - *Dublin Core* (CD), *Metadata Encoding and Transmission Standard* (METS) e o *Preservation Metadata: Implementation Strategies* (PREMIS), entre outros - que podem ser usados a partir de perfis de aplicação para dar suporte a mais de um ação da curadoria digital, há também estruturas próprias para cada domínio ou escopo que ajudam na representação da informação e do conhecimento permitindo a manutenção de termos

autorizados a partir de vocabulários que padronizam conceitos, propriedades e relações, como, por exemplo, os vocabulários *Getty* e os vocabulários *Friend of a Friend* (FOAF).

Na sequência, as ações de 'Planejamento da Preservação' e 'Acompanhamento e Participação da Comunidade' estão ligadas com gestão e administração, determinando as melhores práticas para todo o ciclo. Na primeira ação, cria-se o plano de preservação que deve englobar todas as políticas e as tecnologias, os padrões e os princípios a serem usados no ciclo. Na segunda ação, enfatiza-se o planejamento das atividades referentes aos atores envolvidos, reconhecendo que estes incluem não apenas os guardiões dos dados, mas também aqueles preocupados com o uso e produção de conhecimento gerado por esses dados (CONSTANTOPOULOS *et al*, 2009). Para ambas as ações, verifica-se que os metadados são essenciais para a tomada de decisão que influencia na coordenação e nos resultados de cada ação.

Com base nos dois processos anteriores, as ações de 'Curadoria, Preservação e Aprimoramento do Conhecimento' que estão encadeadas em seu próprio ciclo, tem como foco promover a conservação e a manutenção ao longo do tempo. O que inclui não só as ações que permitem que os dados se mantenham tecnicamente conservados e compatíveis com as tecnologias vigentes, mas também a manutenção dos aspectos semânticos sobre um conjunto de dados e de seu domínio de conhecimento relacionado, para que, desse modo, os dados tenham seu valor intelectual continuamente assegurados visando o interesse de seu público-alvo e os significados pretendidos.

Em relação às 'Ações Sequenciais', estas devem ser cumpridas repetidamente seguindo as melhores práticas sempre que houver dados que necessitem de um contínuo processo de curadoria. A primeira ação é em geral a de 'Conceitualizar', que consiste em planejar a criação dos dados antes de sua publicação ou disponibilização, incluindo o método de captura e as opções de armazenamento, o que envolve o trabalho sinérgico entre modelos conceituais, ontologias e metadados, permitindo antecipar os metadados que serão relevantes ao ciclo de vida útil dos dados. A próxima ação é a de 'Criar ou Receber' e tem direta relação com a ação anterior, envolvendo a criação ou a recepção dos dados juntamente com os vários tipos de metadados necessários ao seu gerenciamento. No caso da recepção, deve-se seguir as políticas, atribuindo, caso seja necessário, metadados mais apropriados.

A ação de 'Avaliar e Selecionar' vem em sequência e diz respeito ao processo de verificação com base nas orientações documentadas, nas políticas e nas exigências legais definidas anteriormente, definindo quais dados serão escolhidos para passar pela curadoria por longo prazo. Em seguida, na ação 'Admitir' (*Ingest*), os dados que são avaliados e selecionados são transferidos para uma entidade custodiadora adequada (repositórios ou *data centers*), garantindo os padrões apropriados, políticas e requisitos legais.

Nas 'Ações de Preservação', busca-se verificar e garantir que os dados permaneçam autênticos, confiáveis e utilizáveis, mantendo sua integridade, o que inclui limpeza de dados, validação, atribuição de metadados correspondentes, bem como a manutenção das estruturas dos dados ou dos formatos de arquivos adequados. A próxima ação é a de 'Armazenar', que consiste em estocar os dados de forma segura, verificando os padrões adequados que devem ser escolhidos por sua relevância para o domínio ou escopo.

Na sequência, as ações de 'Acesso, Uso e Reuso' visam promover o acesso ao público-alvo para o uso ou para o reuso dos dados, seja pela publicação de forma aberta ou a partir do controle de acesso ou procedimentos de autenticação de usuários. A ação seguinte, 'Experiência do Usuário', consiste em registrar a interação entre usuários e dados, bem como os efeitos dessa interação a longo prazo para o aprimoramento das ações. Na ação de 'Transformar' ocorre a criação de novos dados a partir do conjunto original, o que ocorrer tanto pela migração para formatos diferentes, como por meio da criação de subconjuntos derivados desses dados, atendendo às necessidades do público-alvo e aos requisitos, padrões e diretrizes de preservação.

Além dessas ações que são sequenciais, existem as 'Ações Ocasionais', que podem ocorrer em decorrência de alguma situação que gere a necessidade de mudar o curso do processo. Como a ação de 'Descartar', que ocorrer caso os dados não sejam aprovados na ação 'Avaliar e Selecionar' e assim sendo, eles são eliminados de acordo com as políticas e diretrizes documentadas e com os requisitos legais estabelecidos. Esses dados eliminados podem ser transferidos para outro local de guarda ou destruídos, sempre considerando os requisitos legais, mediante a natureza dos dados. A outra ação que pode ocorrer é de 'Reavaliar', que acontece quando os dados, ao passarem pelas 'Ações de Preservação', não atendem aos requisitos, padrões e diretrizes e então retornam para a ação de 'Avaliação e Seleção', na qual ocorrerá a decisão se esses dados serão enviados para o 'Descarte'. Caso não sejam eliminados, eles podem passar pela ação de 'Migrar', na qual são transformados,

adequando-se a formatos diferentes que estejam em conformidade com o ambiente de disponibilização e armazenamento ou que diminuam o risco de obsolescência tecnológica.

Desse modo, verifica-se que a curadoria digital considera a integração e reutilização de dados para poder atender aos requisitos de gerenciamento e manutenção em seu ciclo de ações, o que reflete diretamente no uso dos diversos tipos de metadados, e consequentemente em diferentes estruturas de metadados. Portanto, o uso de estruturas que sejam aderentes a princípios (DUVAL *et al*, 2002; ZENG; QIN, 2016) configura-se como um importante requisito principalmente para a ação de ‘Descrição e Representação da Informação’, influenciando em maior ou menor grau todas as ‘Ações Sequenciais’ e as ‘Ações Ocasionais’.

No caso da modularidade, tal princípio tem importância por permitir criar estruturas novas usando como base padrões já estabelecidos, podendo reutilizar os metadados existentes, o que otimiza e enriquece principalmente as ações de ‘Conceitualizar’ e ‘Criar ou Receber’. Em relação à extensibilidade, ampliar o número de metadados de acordo com as necessidades específicas do domínio tem maior influência em ‘Ações de preservação’ e em “Transformar” que são duas etapas nas quais ocorre a inclusão de novos metadados com a preocupação de não comprometer os existentes.

O refinamento, na medida em que estabelece o grau de detalhamento, ou seja, a granularidade a ser considerada, impacta principalmente a ação de ‘Conceitualizar’ permitindo definir o nível de descrição apropriado e suficiente para cada situação ao longo do processo. Também afeta as ações ‘Criar ou Receber’ e ‘Avaliar e selecionar’, na medida em que o nível de descrição se torna um requisito para eleger os dados. Por último, o multilinguismo refere-se à capacidade de criar um esquema global de metadados que inclua os aspectos da diversidade sociolinguística de cada domínio e escopo, o que enriquece os dados e otimiza a sua utilização afetando não só as ações de ‘Conceitualizar’ e ‘Criar ou Receber’, mas principalmente as ações de ‘Acesso, Uso e Reuso’, contemplando necessidades gerais e contextos específicos.

Diante dessa discussão, é possível entender que o processo de agregar valor na curadoria digital está relacionado a assegurar a validade e a utilidade dos dados de modo que eles não apenas sejam devidamente identificados e recuperados, mas sobretudo para que, dessa forma, possam fomentar ainda mais conexões dentro do seu domínio de

conhecimento para adequação a qualquer que seja seu propósito ao longo do tempo. O quadro 3 apresenta tais características.

Quadro 3: Características dos dados.

Autenticidade	Natureza, propriedade ou condição do que é autêntico e legítimo.	Controles de acesso; uso de metadados apropriados e de identificadores persistentes; Análise de fluxo de dados, como verificação para garantir que os dados não tenham sido corrompidos ou adulterados.
Confiabilidade	Qualidade de algo que é confiável, fiável.	Manutenção dos objetos digitais, realizando limpeza de dados, validação e atribuição de metadados;
Integridade	Estado ou característica de algo que está inalterado, completo em suas partes.	Proteção por controle de autoridade, verificando constantemente a adequação à aplicação.
Usabilidade	Capacidade de permitir, facilitar e otimizar o uso.	Implementação de sistemas adequados; inclusão de metadados para compreensão contextual e gerenciamento sistematizado.

Fonte: Elaborada pelas autoras com base em Higgins (2009).

A autenticidade e a confiabilidade podem ser garantidas pelos processos que certificam a proveniência dos dados, como na ação de ‘Avaliar e selecionar’, que vai analisar se os dados são aderentes às orientações documentadas, bem como às políticas e às exigências legais. Outras ações são as ações de ‘Admitir’ e ‘Armazenar’, que levam em consideração formatos e estruturas de dados aderentes a padrões.

Já a integridade, que tem direta influência no grau de autenticidade e a confiabilidade de um recurso - portanto também sendo influenciada pelas ações de ‘Avaliar e selecionar’, ‘Admitir’ e ‘Armazenar’ - é impactada principalmente pela ação ‘Conceituar’, na qual ocorre o planejamento de quais dados farão parte da curadoria digital, e pela ação de ‘Criar ou Receber’, na qual os dados considerados necessários serão criados ou recebido de fontes externas (repositórios ou *data centers*), considerando os tipos de metadados. Por isso, a ação de ‘Transformar’, também exercerá influência na integridade, pois ela tem a finalidade de criar dados a partir dos originais, seja para adequá-los a uma determinada aplicação ou para mantê-los sustentáveis ao longo do tempo de modo a contemplar as necessidades de gerenciamento e manutenção.

Em relação à usabilidade, a ação sequencial que mais pode afetar essa característica é a de ‘Acesso, uso e reuso’, que considera as condições dos dados para viabilizar acesso, tanto pela publicação de forma aberta, levando em conta direitos e licenças, como a partir do controle de acesso por procedimentos de autenticação. Percebe-se, assim, que devido à complementaridade desses conceitos e da própria interdependência das ações da curadoria digital, mais de uma ação pode influenciar cada característica dos dados, evidenciando o fato

de que as quatro características são afetadas pela 'Ação de preservação', que tem como intuito garantir que os dados permaneçam autênticos, confiáveis e utilizáveis, mantendo a respectiva integridade.

4 CONSIDERAÇÕES FINAIS

O estudo acerca da curadoria digital permite discutir uma abordagem que considera diferentes variáveis tecnológicas e sociais para criar estratégias e procedimentos, visando, além da preservação, do armazenamento e do acesso contínuo, a garantia de autenticidade, confiabilidade, integridade e usabilidade para as diferentes possibilidades de acesso, uso e reuso dos dados. Nessa perspectiva, o estudo possibilitou desenvolver discussões teóricas e sistematizadas baseadas na análise e interpretação da bibliografia sobre a relação da curadoria digital com os metadados. A proposta do modelo DCC&U de acrescentar à curadoria digital uma camada adicional de representação dos aspectos semânticos dos objetos digitais evidencia a ênfase em considerar, além dos metadados que garantam sua descrição e recuperação, também aqueles que garantam a persistência e a validade dos aspectos relacionados ao seu domínio de conhecimento. Isso não significa que anteriormente tais aspectos não eram considerados. O próprio conceito de metadados já implica na ideia de agregar semântica. A questão-chave, contudo, é que na medida em que aumentam os processos e demandas computacionais, torna-se necessário fornecer continuamente um maior nível semântico, e, sobretudo, pragmático aos dados para que estes sejam capazes de gerar e integrar conhecimentos em diversos contextos.

Logo, evidencia-se que o emprego de metadados de qualidade, padronizados e estruturados no processo de curadoria digital e o seu uso estratégico a partir de tecnologias, padrões e princípios possibilitam agregar valor aos dados, pois ampliam o acesso persistente, a verificação de autenticidade, a proveniência dos dados, a experiência de busca e recuperação para os usuários, bem com a garantia de sustentabilidade e de interoperabilidade frente à obsolescência tecnológica e às mudanças institucionais e contextuais.

Portanto, percebe-se que a busca por agregar valor motivou os pesquisadores do DCU a sugerirem uma atualização do modelo do DCC e, assim como houve essa motivação no final da década passada, verifica-se atualmente estudos como o de Huang, Lee e Palmer (2020) e Choudhury, Huang e Palmer (2021) que propõem a análise das ações da curadoria

digital tendo em vista um conjunto de recomendações para atualizar o modelo DCC. Os autores consideram os recentes fatores que impactam a pesquisa e os serviços de dados, como o desenvolvimento da área de Ciência de Dados, os avanços em aprendizado de máquina, o crescimento de fluxos de trabalho integrados, bem como os aspectos relacionados com justiça, responsabilidade e transparência de dados e algoritmos. As recomendações estão relacionadas principalmente à aplicação disciplinar baseada em contexto, aos tipos de dados e à infraestrutura, envolvendo adoção de abordagens orientadas a processos, fluxos de trabalho e tecnologias necessárias. Ainda destacam a necessidade de identificar mecanismos explícitos para a tomada de decisão, considerando todas as funções e atividades de curadoria, os fatores de custo, políticas e transparência de dados e algoritmos (CHOUDHURY, HUANG, PALMER, 2021).

Assim, destaca-se que, o fato de a curadoria digital estar apoiada no uso de metadados, a coloca em uma posição de constante aprimoramento e estudo, visando assim, acompanhar, não só a evolução das soluções tecnológicas, mas também as transformações sociais que apoiam a organização e representação proveniente dos metadados.

REFERÊNCIAS

BAKER, T. *et al.* **Library linked data incubator group final report**. W3C Incubator Group Report, 2011. Disponível em: <http://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>. Acesso em: 29 abr. 2021

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2011.

BEAGRIE, N. Digital curation for science, digital libraries, and individuals. **The International Journal of Digital Curation**, Edinburgh, v. 1, n. 1, p. 3-16, 2006. Disponível em: <http://www.ijdc.net/article/view/6>. Acesso em: 28 abr. 2021.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic *web*. **Scientific American**, p. 29-37, 2001. Disponível em: <https://www.ida.liu.se/~evabl45/files/bernersLee01Semantic.pdf>. Acesso em: 29 abr. 2021

CAMPOS, J. A. G. **Análise conceitual sobre as relações semânticas em Ciência da Informação**: contribuições para o desenvolvimento de ontologias. 2009. 135 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: <http://hdl.handle.net/1843/ECID-7V4PLA>. Acesso em: 27 abr. 2021.

CHOUDHURY, S.; HUANG, C.; PALMER, C. L. Updating the DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, Edinburgh, v. 15, n. 1, p. 1-12, 2020. Disponível em: <https://doi.org/10.2218/ijdc.v15i1.721>. Acesso em: 27 abr. 2021.

CSILLAG, J. M. **Análise do Valor**: metodologia do valor: engenharia do valor, gerenciamento do valor, redução de custos, racionalização administrativa, aumento do valor percebido pelo

cliente, a empresa vista como sistema, melhoramento contínuo. 4. ed. São Paulo: Atlas, 1995.

DOERR, M. **DCC Digital Curation Manual: instalment on ontologies**. Edimburgo, Escócia: Digital Curation Manual. 2008. Disponível em: <http://hdl.handle.net/1842/3341>. Acesso em: 27 abr. 2021.

DUVAL, E. *et al.* Metadata principles and practicalities. **D-Lib**, v. 8, n. 4, 2002. Disponível em: <http://www.dlib.org/dlib/april02/weibel/04weibel.html>. Acesso em: 29 abr. 2021

GILLILAND, A. J. Setting the stage. *In*: BACA, Murtha (Ed.). **Introduction to metadata**. Los Angeles, CA: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/setting-the-stage/>. Acesso em: 27 abr. 2021.

HIGGINS, S. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, Edinburgh, v. 3, n. 1, p. 134-140, 2008. Disponível em: <http://dx.doi.org/10.2218/ijdc.v3i1.48>. Acesso em:

HIGGINS, S. DCC DIFFUSE Standards Frameworks: a standards path through the Curation Lifecycle. **The International Journal of Digital Curation**, Edinburgh, v. 4, n. 2, p. 60-67, 2009. Disponível em: <https://doi.org/10.2218/ijdc.v4i2.93>. Acesso em: 27 abr. 2021.

HITZLER, Pascal. A review of the semantic *web* field. **Communications of the ACM**, v. 64, n. 2, p. 76-83, 2021. Disponível em: <https://doi.org/10.1145/3397512>. Acesso em: 21 abr. 2021.

HUANG, C.; LEE, J. S.; PALMER, C. L. **DCC curation lifecycle model 2.0: Literature review and comparative analysis**. Seattle, WA: University of Washington Information School, 2020. Disponível em: <http://hdl.handle.net/1773/45392>. Acesso em: 21 abr. 2021.

MARON, N. L.; YUN, J.; PICKLE, S. **Sustaining our digital future: institutional strategies for digital content**. New York: Ithaka S+R, 2013.

OLIVER, G.; HARVEY, R. **Digital curation**. London: Facet Publishing, 2016.

PENNOCK, M. Digital curation and the management of digital library cultural heritage resources. **Local Studies Librarian**, v. 25, n. 2, p. 3-7, 2006. Disponível em: http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lsl-curation_mep.pdf. Acesso em: 29 abr. 2021.

RUSBRIDGE, C. *et al.* The digital curation centre: a vision for digital curation. *In*: IEEE INTERNATIONAL SYMPOSIUM ON MASS STORAGE SYSTEMS AND TECHNOLOGY, Sardinia, Italy, 2005, p. 31-41. **Proceedings [...]**. Sardinia, Italy: Mass Storage and Systems Technology Committee of the IEEE Computer Society, 2005. Disponível em: 10.1109/LGDI.2005.1612461. Acesso em: 29 abr. 2021.

SAYÃO L. F. SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: estudos**, v. 22, n. 3, p. 179-191, 2012. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/92680>. Acesso em: 29 abr. 2021

VIDOTTI, S. A. B. G. *et al.* *Web, web* semântica e *web* pragmática: um posicionamento da Arquitetura da Informação. **Informação & Sociedade**, v. 29, n. 1, 2019. Disponível em: 10.22478/ufpb.1809-4783.2019v29n1.44358. Acesso em: 21 abr. 2021.

ZENG, M. L.; QIN, J. **Metadata**. 2.ed. Chicago, IL: ALA Neal-Schuman, 2016.