

Gestão de dados de pesquisa: uma aproximação entre o modelo de governança de dados proposta por dama e o ciclo de vida de curadoria DCC

Felipe Pires da Silva¹; Rafael Porte da Rocha²

RESUMO

O presente ensaio de pesquisa destina-se a análise de modelos de governança de dados e informação publicados pela DAMA em DMBOK2 e o modelo de Ciclo de Vida de Curadoria Digital publicada pelo DCC. O objetivo da pesquisa é analisar as correlações entre modelos e identificar possíveis oportunidades de complemento entre modelos afim de potencializar o tratamento de dados de pesquisa e informação.

Palavras-chave: dados de pesquisa, curadoria digital, gestão de dados, DMBOK; ciclo de vida DCC;

INTRODUÇÃO

Assim como vários elementos tiveram seu valor em determinadas épocas de nossa sociedade como o sal, o açúcar, o ouro, hoje nos deparamos com um ativo intangível, porém de alto valor agregado mudando definitivamente o dia a dia de pessoas e organizações. Mas como saber se os dados possuem valor? Existe algum modelo para gerenciamento de dados? Podemos aproximar as áreas da informação com as áreas de tecnologia e dados?

Para responder essas perguntas, iremos iniciar uma jornada com o intuito de explorar referências sobre o gerenciamento de dados apresentados pelo DAMA, uma associação global sem fins lucrativos, dedicada ao avanço dos conceitos e práticas de gerenciamento de informações e dados e sua possível contribuição com a estrutura de Ciclo de Vida de Curadoria Digital do Digital Curation Centre (DCC), um centro de especialização em curadoria de informação digital, líder global neste seguimento. Através deste ensaio, poderemos identificar os benefícios que o uso das referências do DAMA poderá trazer à curadoria digital na direção de dados de pesquisa, tendo

¹ Mestrando do Programa de Pós-Graduação em Ciência da Informação – PPGCIN/UFRGS felipepires@live.com

² Professor do Programa de Pós-Graduação em Ciência da Informação – PPGCIN/UFRGS

como base o Ciclo de Vida do DCC (DCC/CVCD). Serão analisadas as estruturas propostas pelo DMBOK e identificadas similaridades que possam vir a apoiar e contribuir com o DCC/CVCD. Os principais objetivos do ensaio serão: Analisar como DMBOK2 contextualiza dado e informação e analisar as estruturas que DMBOK2 apresenta e identificar de que forma estas se enquadram no contexto do DCC/CVCD.

CURADORIA DIGITAL E CICLO DE VIDA DO DCC

Curadoria Digital “trata de manter e agregar valor a um corpo confiável de informações digitais para uso atual e futuro, em outras palavras, é o gerenciamento ativo e a avaliação da informação digital ao longo de todo o seu ciclo de vida” de acordo com Pennock, (2007). Para apoiar a realização da curadoria digital, o Centro de Curadoria Digital (DCC) desenvolveu o Ciclo de Vida da Curadoria Digital. O DCC/CVCD fornece uma visão geral gráfica de alto nível dos estágios necessários para uma curadoria e preservação bem-sucedida dos dados, desde a sua conceituação ou recebimento inicial. O modelo pode ser usado para planejar atividades dentro de uma organização ou consórcio para garantir que todas as etapas necessárias sejam realizadas, cada uma na sequência correta (HIGGINS, 2008).

Segundo pesquisa de Davidson (2016), “em um sentido prático, o modelo ajudou a progredir no desenvolvimento de procedimentos, ferramentas e serviços de apoio para o gerenciamento e compartilhamento de dados de pesquisa dentro das instituições”.

Em sua publicação The DCC Curation Lifecycle Model, o DCC apresenta um modelo de ciclo de vida para a curadoria e preservação dos dados para que organizações tenham uma prática bem-sucedida (figura 1). A abordagem do DCC como se pode observar, mostra o conceito em um círculo contínuo e por camadas para que o dado seja tratado como se estivesse em uma esteira industrial, de forma que há um início, meio e fim.

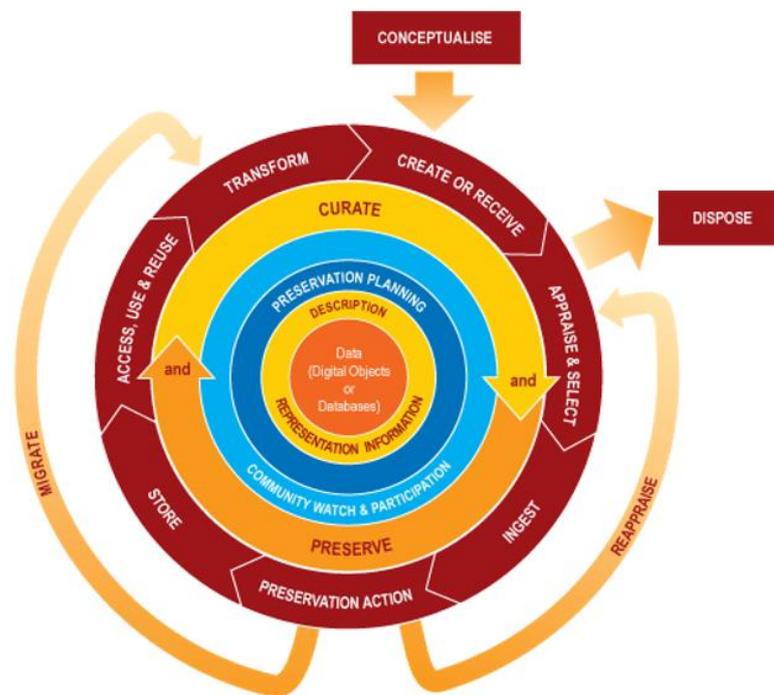


Figura 1: Ciclo de Vida da Curadoria Digital de DCC. Fonte Higgins (2008)

CONHECENDO O DAMA E O DMBOK2

DAMA Internacional é uma associação global sem fins lucrativos, dedicada ao avanço dos conceitos e práticas de gerenciamento de informações e dados. Seu objetivo principal é promover a compreensão, o desenvolvimento e a prática de gerenciamento de dados e informações como ativos corporativos essenciais para apoiar a organizações.

Para materializar os mais de trinta anos de estudos, a DAMA International publica em 2009 o DMBOK, Guide to the Data Management Body of Knowledge, (MOSLEY et al, 2009) um livro de práticas e experiências compartilhadas pelos melhores profissionais da indústria. O DMBOK se encontra em sua segunda edição (DAMA 2017), abordando novos temas como “Big Data”, inserindo um capítulo ressaltando a importância da ética no trabalho com dados.

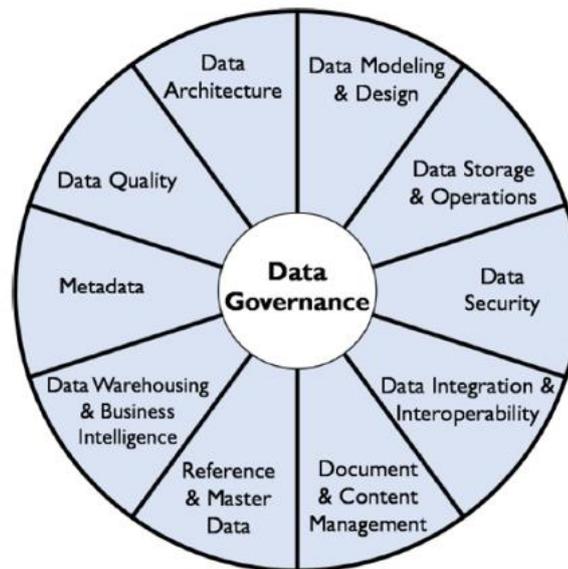


Figura 2- The DMBOK2 Framework – DAMA Wheel. Fonte DAMA, 2017

Os capítulos do DMBOK2 abordam os temas: gestão de dados, ética na manipulação de dados, governança de dados, modelagem e design de dados, segurança de dados, integração e interoperabilidade, documentos e conteúdos, referência e dados mestre, *data warehousing* e *business intelligence*, gestão de metadados, qualidade de dados, big data e ciência de dados, avaliação da maturidade da gestão de dados, organização da gestão de dados e expectativa dos papéis e gestão de dados e gestão de mudança organizacional.

Em seu capítulo sobre Gestão de Dados, o DMBOK2 apresenta estruturas de gestão de dados (Data Management Frameworks), que trata sobre um conjunto de interdependências no qual a governança de dados faz parte, respeitando os objetivos, atividades e responsabilidades individuais para cada finalidade.

DMBOK Framework é uma estrutura que contém três modelos que ajudam a definir as áreas de conhecimento: o DAMA Wheel, Hexagon e o Context Diagram. Para realizarmos um recorte na pesquisa, o modelo discutido será o DAMA Wheel (figura 2) que possui as funções de gestão de dados que são devidamente expandidos e explicados ao longo dos capítulos do DAMA-DMBOK2.

GESTÃO DE DADOS NO CONTEXTO DO DMBOK2

Diante da nova economia, o dado tornou-se um ativo de valor. Ele é a matéria prima para a criação de novos produtos em qualquer área para a sociedade. O dado em si já é um ativo que pode possuir preço, seja ele bruto ou tratado, extraído ou comprado. Como todo ativo de valor, precisa-se gerenciá-lo. Processos e a clareza das necessidades estratégicas e técnicas devem estar no centro das discussões quando o assunto for a gestão dos dados.

A gestão de dados passa por princípios importantes que devem estar claros principalmente sobre a importância do tratamento deste ativo. Alguns princípios da gestão de dados estão descritos no DAMA-DMBOK2 (DAMA,2017):

- entender que ele é um item de valor e que possui uma propriedade única;
- requerer o compromisso das lideranças no gerenciamento de dados;
- conhecer os requisitos de negócio;
- gerenciar a qualidade dos dados;
- fazer com que metadados gerencie os dados;
- planejar o gerenciamento dos dados;
- orientar as decisões de tecnologia da informação através dos requisitos de gerenciamento de dados;
- possuir diversos conhecimentos em diferentes perspectivas da gestão de dados;
- compreender que dados possuem ciclo de vida distintos e precisam ser tratados individualmente;
- gerenciar riscos associados aos dados.

PROCEDIMENTOS METODOLOGICOS E RESULTADOS PRELIMINARES

O estudo realizado segue o método de pesquisa bibliográfica, com base em literatura científica e técnica em curadoria digital, gestão de dados e, mais especificamente, no DCC/CVCD e nas referências expostas pelo DMBOK2. A análise realizada traz a compreensão da relevância do dado na perspectiva do DMBOK2, isto

é, uma visão de gestão em torno dos objetivos estratégicos envolvendo as áreas técnicas e de negócio das organizações com foco em dados e informação. A análise da forma com que as estruturas que DMBOK2 se enquadram no contexto do DCC/CVCD envolverá verificar como a Curadoria Digital, em seu ciclo de vida, pode ser inserida na gestão de dados de uma organização, na perspectiva que DMBOK2 visa a gestão de dados correntes e permanentes da organização e DCC/CVCD visa realizar a curadoria digital dos dados permanentes. Para cada área de conhecimento de DMBOK2, serão investigadas ações pertinentes do DCC/CVCD e analisadas de que forma essas ações poderiam ser contempladas no sentido de trazer Curadoria Digital à gestão de dados tratada por DMBOK2. O quadro 1 resume os resultados preliminares. O texto a seguir apresenta os resultados preliminares do estudo.

DADO E INFORMAÇÃO EM DMBOK2

Quando se deseja definir o elemento “dado”, existem vários conceitos espalhados por todas as áreas do conhecimento. Em seu berço, a tecnologia da informação, dados são compreendidos como informações armazenadas em um formato digital (embora os dados não se limitem a informações que foram digitalizadas e os princípios de gerenciamento de dados se apliquem aos dados capturados em papel, bem como em bancos de dados) conforme DAMA (2017).

Os dados são um meio de representação, significam coisas diferentes de si mesmo (Chisholm 2010). Podemos também dizer que assim com um dado de jogo, um dado como representação possui facetas distintas do mesmo elemento (e que podem ter interpretações dependendo do seu lado). As organizações focadas em extrair valor sobre dados, precisam compreender e criar padrões para manter a consistência de seus registros e vencer os desafios da diversidade de elementos e fontes para lidar com o imenso volume, variedade e velocidade dos dados (DAMA, 2017).

Assim como a definição de “dados”, também há uma discussão em torno das diferenças conceituais sobre dado e informação. Dado tem sido chamado de “matéria

prima da informação” e Informação tem sido chamada de “dado em um contexto” segundo (DAMA, 2017). Como descreve DMBOK2, uma pirâmide frequentemente é utilizada para representar a relação entre os elementos dados, informação, conhecimento e sabedoria (figura 3).

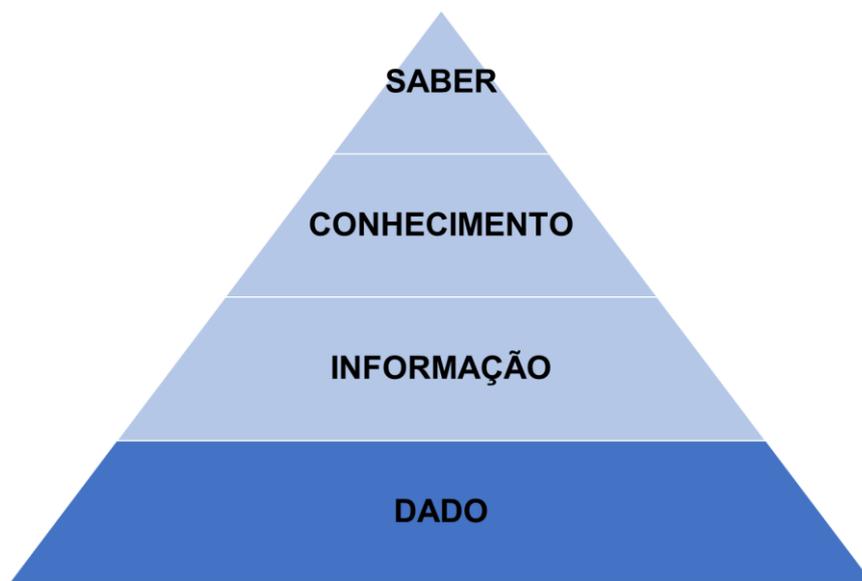


Figura 3 – Pirâmide de representação Fonte: Elaborada pelo autor

Como podemos observar, o dado é a base da pirâmide, o elemento fundamental que alicerça e torna capaz a geração da informação, do conhecimento e do saber. Através desse ponto de partida, pode-se compreender, organizar e identificar propósitos sobre o uso massivo de dados. Por muitas vezes as organizações buscam informações a respeito de determinados assuntos com objetivos específicos, mas não os conseguem de forma consistente pois não estão com o alicerce bem consolidado (dados). Também o contrário pode acontecer, uma organização pode ter muitos dados e não saber que tipo de informação procurar diante da diversidade e do volume extraído e, se há necessidade de mantê-lo por algum tempo. A Essa visão está alinhada com a CI, como podemos observar em Barreto,

que analisa essa pirâmide na perspectiva de fluxos e estoques com que se ocupam quando se vivencia a condição da informação. Segundo Barreto (2002),

O saber tem, quando se nasce, uma condição de vida igual ao *labor*; o conhecimento surge pela conquista, pelo *trabalho* e é inserido nas práticas de uma *ação* de inteligência com a realidade. Na base da pirâmide existe, assim, um estoque de fatos, ideias e produtos da sensibilidade humana, institucionalizados ou não, que se transformam em outro estoque de informação.

AS ABORDAGENS DAS ÁREAS DE CONHECIMENTO DO DAMA E DCC

O quadro 1 resume os estudos iniciais que visam aproximar DMBOK2 e DCC/CVCD na perspectiva de que a Curadora Digital é parte da gestão de dados desenhada por DMBOK2. O quadro traz, na coluna 1, as áreas de conhecimento de DAMA/DMBOK Wheel (figura 2), com uma breve descrição resumida de DMBOK2 (DAMA,2017). A coluna 2 apresenta a análise preliminar de como a Curadoria Digital (no quadro, abreviada para CD) poderia ser considerada em cada área de conhecimento de DMBOK2.

Quadro 1 – Ações do Ciclo de Vida de DCC e DMBOK2

DMBok2	Análise quanto ao CVCD
Governança de Dados Planejamento, supervisão e controle sobre o gerenciamento e uso de dados.	Ciclo de Vida Da Curadoria Digital Monitorar a Comunidade - Curar e preservar <ul style="list-style-type: none"> Inclusão da CD na Governança de Dados Desenvolvimento da cultura organizacional em CD Promover políticas, requisitos, normas e padrões de CD
Arquitetura de Dados Definindo o modelo (blueprint) para gerenciamento de ativos de dados	Descrição e Representação da Informação Planejamento da Preservação - Monitorar a Comunidade <ul style="list-style-type: none"> Incluir, na Arquitetura de Dados Corporativa, modelos de descrição e representação para os dados de acordo com requisitos de CD
Modelagem e Desenho de Dados Descobrir, analisar e definir os requisitos de dados; representar e comunicar os requisitos de dados em "modelo de dados".	Descrição e Representação da Informação Planejamento da Preservação <ul style="list-style-type: none"> Modelar os dados que serão preservados a longo prazo de acordo com requisitos de CD Planejar, construir e manter o modelo de dados da CD
Referência e Dados Mestres Gerenciar dados compartilhados para reduzir a redundância e garantir melhor qualidade por meio da padronização de valores de dados.	Descrição e representação da informação Planejamento da Preservação <ul style="list-style-type: none"> Preservar a longo prazo dados em conformidade com a Gestão de Referência e Dados Mestre da organização Incluir referências e dados mestres da CD
Metadados Planejamento, implementação e controle para acesso a metadados integrados de alta qualidade	Descrição e representação da informação Planejamento da Preservação <ul style="list-style-type: none"> Descrever os dados que serão preservados longo prazo de acordo com requisitos de CD Incluir, na Gestão de Metadados, os metadados de CD
Armazenamento e Operações de Dados O design, implementação e suporte de dados armazenados, para maximizar seu valor	Dados: Base de Dados - Planejamento da Preservação <ul style="list-style-type: none"> Desenvolver estratégias para preservar a longo prazo dados de bases de dados da organização Ciclo de Vida da Preservação Digital Monitorar a Comunidade, Armazenamento <ul style="list-style-type: none"> Incluir, na Gestão de Armazenamento e Operação de Dados da organização, a gestão da CD
Gestão de Data Warehouse e Business Intelligence Fornecer dados de suporte à decisão.	Dados - Base de Dados Planejamento da Preservação <ul style="list-style-type: none"> Desenvolver estratégias para preservar a longo prazo data warehouses e data marts
Documentos e conteúdo Planejamento, implementação e controle para o gerenciamento do ciclo de vida de dados e informações em qualquer forma ou meio.	Dados: Objetos Digitais - Planejamento da Preservação <ul style="list-style-type: none"> Desenvolver estratégias para preservar a longo prazo textos, imagens, sons, sites, combinações de objetos Ciclo de Vida da Preservação Digital - Monitorar a Comunidade <ul style="list-style-type: none"> Incluir, na Gestão de Documentos e Conteúdos da organização, documentos e conteúdos da CD
Qualidade dos Dados Planejamento, implementação e controle que aplicam técnicas de gestão da qualidade aos dados	Planejamento da preservação - Avaliação e Seleção <ul style="list-style-type: none"> Incluir a CD na Gestão de Qualidade de Dados da organização Utilizar de critérios de qualidade para avaliar e selecionar dados a serem preservados a longo prazo
Segurança de dados Definição, planejamento, desenvolvimento e execução de políticas e procedimentos de segurança	Planejamento da preservação - Acesso e Uso <ul style="list-style-type: none"> Incluir CD na Gestão da Segurança de Dados da organização. Gerir riscos na preservação digital. Acessar e usar dados preservados a longo prazo em conformidade com a Segurança de Dados
Integração e interoperabilidade de dados Gerenciar o movimento e a consolidação de dados dentro e entre aplicativos e organizações.	Ciclo de Vida de Curadoria Digital <ul style="list-style-type: none"> Desenvolver a CD integrada e interoperável com os ambientes e aplicativos da organização Planejamento da Preservação <ul style="list-style-type: none"> Desenvolver estratégias para preservar a longo prazo dados integrados e interoperáveis

Elaborada pelos autores.

Podemos observar em DMBOK um enfoque muito grande no levantamento de necessidades de dados e o desenvolvimento e manutenção de soluções que atendam a essas necessidades. No DCC/CVCD isso está relacionado a planejar, descrever, conceitualizar, criar e receber, avaliar e selecionar, na perspectiva de que essas soluções devam ser analisadas no contexto de sua preservação a longo prazo. Em todas as ações do DMBOK encontramos uma correlação com o DCC/CDCD e pode ser aprofundado em pesquisas mais detalhadas. Conforme relatado no quadro 1, percebe-se que a Curadoria Digital pode trazer à gestão de dados os requisitos, os avanços, os estudos, os padrões, os planejamentos e as ações desenvolvidas para Curadoria Digital. Observa-se que as análises desenvolvidas no quadro 1 tendem a apresentar estratégias, requisitos e/ou ações de Curadoria Digital a serem consideradas no desenvolvimento de cada área de conhecimento da gestão de dados, no sentido de ampliar o escopo e os requisitos dessas áreas de conhecimento no que diz respeito à Curadoria Digital.

CONSIDERAÇÕES FINAIS

Tendo em vista as argumentações expostas no ensaio, podemos observar uma aproximação direta do DAMA com a responsabilidade da governança dos dados e informações com abordagem estratégica e focada em organizações para fins de expansão de técnicas aplicadas no contexto da tecnologia da informação e gestão dos dados e informação. Na perspectiva do DCC, conseguimos captar um modelo referencial teórico que apresenta práticas consistentes da curadoria e preservação digital de forma a atender os meios físicos e digitais pautados em uma esteira contínua levando o dado para o centro das ações.

Neste estudo podemos perceber que há uma oportunidade de aprofundamento nos modelos DAMA e DCC que poderá abrir portas de complementos de técnicas de governança sobre os dados aliados ao processo de curadoria digital. Observa-se, a partir desse estudo preliminar, uma necessidade em aproximar a Curadoria Digital à

Gestão de Dados, em que Curadoria Digital, com suas técnicas, padrões e avanços, passa a ser parte e contribuir com a gestão de dados.

Diante dos estudos preliminares, faz-se necessário a continuidade da pesquisa em busca da identificação de processos que podem estar contidos em um modelo e que pode ser complementar a outro e vice-versa. A contribuição da pesquisa poderá aprimorar estruturas e modelos vigentes em prol da potencialização das atuais tecnologias de curadoria e preservação de dados e informação direcionados à dados de pesquisa.

REFERÊNCIAS

MOSLEY, M. et al. The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMBOK Guide). DAMA International, 2009.

DAMA. DAMA-DMBOK: Data Management Body of Knowledge. DAMA International, 2017.

HIGGINS, S. The DCC Curation Lifecycle model. **The International Journal of Digital Curation**, Pittsburgh, v. 3, n. 1, Aug. 2008. DOI: [10.1145/1378889.1378998](https://doi.org/10.1145/1378889.1378998)

PENNOCK, M. Digital curation: A life-cycle approach to managing and preserving usable digital information. **Library & Archives**, v. 1, n. 1, p. 1-3, 2007.

DAVIDSON, Joy. Fostering open science practice through recognising and rewarding research data management and curation skills. In: Bisto, C. and Raju, R. (eds.) LIS Education and Research in a Dynamic Information Landscape: Proceedings of the Library and Information Studies Centre 75 Years Commemorative Conference. University of Cape Town Libraries: Cape Town, p. 63-75, 2016.