

A CONTRIBUIÇÃO DOS FATORES DO CICLO DE VIDA DOS DADOS A FASE DE COLETA DOS DADOS NA ATIVIDADE DE PROTOCOLO DA EMBRAPA SOJA LONDRINA

The contribution of Data Life Cycle factors to the Data Collection phase in the Protocol activity of Embrapa Soja Londrina

Eduardo Kioshi Matsubara¹, Diana Vilas Boas Souto Aleixo^(N)

(1) Universidade Estadual de Londrina, Londrina (PR), eduardo.kmatsubara@gmail.com.

(N) , Universidade Federal do Espírito Santo, Vitória (ES), dianavbsouto@gmail.com.

Resumo:

Com o crescimento de documentos produzidos em meio digital, se faz necessário o uso de sistemas de automação de processos, buscando melhorar as atividades e agilizar os processos dentro da empresa. Nesse contexto surgem modelos de gestão de dados como o Ciclo de Vida dos Dados (CVD), que visam garantir a confiabilidade, integridade, segurança entre outros fatores dos dados, informações e documentos gerenciados eletronicamente. Desta forma, o objetivo do presente trabalho é caracterizar os fatores do Ciclo de Vida dos Dados que podem contribuir com a fase de coleta dos dados na atividade de Protocolo presente no Sistema Eletrônico de Informações (SEI) da Embrapa Soja Londrina. A metodologia se caracteriza como uma pesquisa exploratória e descritiva, com abordagem qualitativa, pesquisa bibliográfica e de campo. Para a pesquisa bibliográfica utilizou-se um levantamento na Brapci, e para pesquisa de campo foi realizada uma entrevista com a arquivista responsável pelo arquivo da Embrapa Soja Londrina. Por meio dos dados, foi possível identificar que a Embrapa Soja Londrina está preocupada com diversos fatores presentes no Ciclo de Vida dos Dados. Conclui-se sobre os fatores do Ciclo de Vida dos Dados, é necessária uma preocupação mais sistematizada com a problemática, que pode contribuir com a adequação da ferramenta Sistema Eletrônico de Informações às exigências arquivísticas.

Palavras-chave: Ciclo de Vida dos Dados; SEI; Coleta de dados

Abstract:

With the growth of documents produced in digital media, it is necessary to use process automation systems, seeking to improve activities and streamline processes within the company. In this context, data management models such as the Data Life Cycle (CVD) emerge, which aim to guarantee the reliability, integrity, security, among other factors, of data, information and documents managed electronically. Thus, the objective of this work is to characterize the Data Life Cycle factors that can contribute to the data collection phase in the Protocol activity present in the Electronic Information System (SEI) of Embrapa Soja Londrina. The methodology is characterized as an exploratory and descriptive research, with a qualitative approach, bibliographic and field research. For the bibliographical research, a survey at Brapci was used, and for field research, an interview was conducted with the archivist responsible for the Embrapa Soja Londrina archive. Through the data, it was possible to identify that Embrapa Soja Londrina is concerned with several factors present in the Data Life Cycle. It concludes on the factors of the Data Life Cycle, a more systematic concern with the problem is necessary, which can contribute to the adequacy of the Electronic Information System tool to archival requirements.

Keywords: Data Life Cycle; SEI; Data Collect.

1. Introdução

A Empresa Brasileira de Pesquisa Agropecuária (Embrapa), criada no ano de 1973, que tem como missão “Viabilizar soluções de pesquisa, desenvolvimento e inovação para a sustentabilidade da

agricultura, em benefício da sociedade brasileira.” (EMBRAPA, 2022, *online*). Para a gestão de seus processos e documentos a Embrapa adota em seus processos arquivístico o Sistema Eletrônico de Informações (SEI), desenvolvido pelo

Tribunal Regional Federal da 4ª Região (TRF4).” (BAHIA, [2022], *online*).

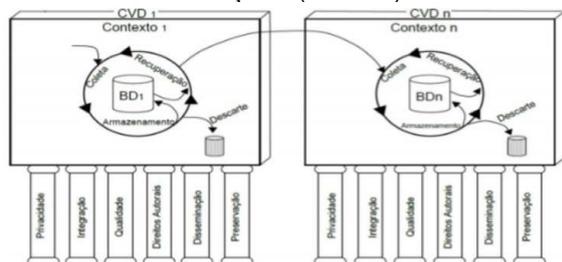
Diante de tal perspectiva, infere-se sobre a importância de modelos que considerem os fluxos de dados, inovando e melhorando o processo de gestão de documentos (BRASIL, 2011). Tal constatação pode auxiliar em questões que abarcam as dificuldades de identificar os contextos de criação dos documentos, ligando-os às suas atividades, considerando o valor primário e o valor secundário de cada documento, e fazer emergir a necessidade de explorar no tocante arquivístico a análise dos dados que perpassam o universo da gestão de documentos.

Nesse sentido, o modelo de Ciclo de Vida dos Dados (CVD) proposto por Sant’Ana (2016), que aborda a necessidade de um modelo que leve a conhecer e compreender todas as fases e fatores do processo de acesso a dados, de maneira a tirar o melhor proveito de cada uma delas, a partir de uma aplicação adequada,

“[...] mantendo-se como ponto central os próprios dados e para tanto se propõe o uso do Ciclo de Vida dos Dados - CVD como forma de evidenciar os diferentes momentos e fatores envolvidos neste processo.” (SANT’ANA, 2016, p. 119).

As fases que compõem o ciclo de vida dos dados são: Coleta, Armazenamento, Recuperação e Descarte (FIGURA 1).

Figura 1 - Ciclo de Vida dos Dados para Ciência da Informação – (CVD–CI)



Fonte: Sant’Ana (2016, p. 123).

De acordo com Sant’Ana (2016), a coleta é o primeiro momento em que os dados serão obtidos, e também onde são levantadas quais serão as necessidades informacionais, as estratégias de como localizar e avaliar, os mecanismos,

metodologias e as ferramentas para conseguir obter os dados, assim podendo possibilitar ao sistema coletar os dados mais relevantes, que posteriormente serão utilizados em alguma pesquisa.

Na fase da Coleta, deve-se “[...] considerar a difícil tarefa de identificar fontes de dados que possam ser úteis para que os usuários possam ter atendidas suas necessidades.” (SANT’ANA, 2016, p. 124).

No Armazenamento “tem-se um enfoque mais tecnológico e se definem aspectos que garantem a reutilização destes dados, por meio de especificações físicas e lógicas sobre como os dados serão registrados em um suporte”. (SANT’ANA, 2016, p. 127).

Na Recuperação “as estratégias e ações passam a ser avaliadas a partir do ponto de vista do responsável por sua manutenção e não daqueles que acessarem estes dados.” (SANT’ANA, 2016, p. 133).

E por fim, no Descarte, “Cabe refletir sobre uma fase em que ocorre o descarte de dados que não são mais necessários ou que estejam acima da capacidade de tratá-los com eficiência para o sistema como um todo.” (SANT’ANA, 2016, p. 135). Ressalta-se que em todas essas fases tem-se a preocupação com os seguintes fatores: preservação, disseminação, direitos autorais, qualidade, integração e privacidade.

No que se refere aos fatores, eles estão presentes em todas as fases e são essenciais para garantir que os dados coletados de fato atendam as demandas para os quais estão sendo coletados, armazenados, recuperados etc (QUADRO 1). Cabe mencionar que em cada a fase os fatores vão apresentar características distintas, isto porque em cada fase eles se manifestam de maneira a garantir que o processo ocorra da melhor forma possível.

Quadro 1 - Fatores presentes em cada fase do Ciclo de Vida dos dados

FATORES	FASES			
	COLETA	ARMAZENAMENTO	RECUPERAÇÃO	DESCARTE
Privacidade	Verificação das fontes utilizadas, para que não haja quebra de privacidade;	Relaciona-se com a identificação de quem terá acesso ao registro, não apenas para consulta, mas para modificação e exclusão;	Preocupa-se com uma estrutura de recuperação que garanta a privacidade dos envolvidos com a conteúdo disponibilizado;	Tem relação com a possibilidade dos indivíduos retirarem seus dados da base caso ou queiram ou seja necessário;

Integração	Se preocupa em como a fase de coleta vai identificar e validar os atributos responsáveis pela identificação única de cada registro;	Refere-se às definições sobre como serão acessados os dados, assim como uma camada de interação que se articule de maneira a ocorrer uma integração. Ou, a adoção de formatos abertos que garanta acesso direto.	Garantir a integração dos registros recuperados com outros com os quais se relacionem;	Preocupa-se com como a exclusão de registros de uma base pode afetar registros em bases integradas;
Qualidade	Relaciona-se com a confiabilidade, obtendo informações tais como procedência e mecanismos de coleta etc.;	Preocupa-se com a manutenção da integridade física e lógica do registro durante seu armazenamento;	Os registros a serem recuperados devem ter qualidade garantida, principalmente no que se refere à interação com o usuário;	A manutenção de informações registradas com detalhamento sobre a eliminação, de forma que futuramente possam ser identificados os contextos afetados pela eliminação.
Direito autoral	Preocupa-se com a responsabilidade da fonte, vinculando-a aos dados coletados;	Visa registrar as informações, para segurança das instituições que respondem pelo registro e pelo dispositivo de armazenamento;	Refere-se à clareza sobre quem pode acessar os dados;	Relacionada à manutenção de informações sobre autoria após o descarte, de maneira a possibilitar a identificação da origem dos dados mesmo depois de descartados.
Disseminação	Preocupa-se com a viabilidade de acesso, favorecendo sua localização e interpretação;	Prevê forma de acesso aos registros, estabelecendo elementos que facilitem sua localização;	Preocupa-se com elementos que permitam sua localização;	Preocupa-se com a eliminação de dados que mesmo não relacionados ao conjunto possam prejudicar a localização deste posteriormente.
Preservação	Diz respeito ao estabelecimento de dados adicionais sobre características que auxiliem na preservação do registro em questão.	Recursos que possibilitem a interpretação do registro no futuro, independentemente dos dispositivos originais de armazenamento.	Relaciona-se a questões de sua interpretação, que devem ser possíveis em momentos distintos de recuperação.	Se preocupa com a preservação de dados que mesmo não apresentando utilidade podem ser necessários futuramente.

Fonte: Elaborado pelos autores com base em Sant'Ana (2016).

Infere-se que todos os fatores se manifestam em todas as fases. Em destaque está a fase de descarte, na qual apenas os fatores integração e privacidade são considerados relevantes para sua aplicação, uma vez que nesta fase os dados serão eliminados, assim, os aspectos que foram relevantes na coleta, armazenamento e demais fases, serão menos relevantes.

Considera-se que o CVD não é um modelo aplicado atualmente na área da Arquivologia, e seu desenvolvimento volta-se para o tratamento de dados.

Tendo as informações apresentadas, a questão que se destaca é: "Quais os fatores do CVD envolvidos na coleta dos dados na atividade de Protocolo?"

A relevância do tema considera essencial debater sobre o tratamento de dados, visando facilitar o processo de gestão arquivística, cujas tecnologias de informação ampliaram, e as quais a gestão tradicional de documentos já não abarca tais exigências.

Diante de tal perspectiva, o presente trabalho refere-se a um recorte do Trabalho de Conclusão de Curso intitulado "OS FATORES ENVOLVIDOS NO PROCESSO DE COLETA DE DADOS NO SEI DA EMBRAPA SOJA LONDRINA" (MATSUBARA, 2020), o qual teve como foco analisar os fatores envolvidos no processo de coleta dos dados a partir da atividade de Protocolo presente no SEI da Embrapa Soja Londrina.

Ademais, acrescenta-se a importância de discorrer sobre as iniciativas arquivísticas que estão em desenvolvimento no setor público e quais as contribuições que estas podem trazer a outras instituições, com enfoque na Embrapa Soja, instituição pública selecionada como universo desta pesquisa.

2. Objetivo

A pesquisa teve como objetivo caracterizar os fatores do CVD que podem contribuir com a fase de coleta dos dados na atividade de Protocolo presente no SEI da Embrapa Soja Londrina.

3. Procedimentos Metodológicos

A pesquisa caracteriza-se como exploratória e descritiva, que fez uso da pesquisa bibliográfica e de campo para desenvolver as etapas que viriam a cumprir o objetivo proposto.

Como universo de pesquisa, elegeu-se o arquivo da unidade da Embrapa Soja na cidade de Londrina, sendo que no setor realizou-se uma visita e a aplicação da coleta de dados por meio da entrevista. Na sequência apresenta-se a amostra utilizada.

A amostra foi composta por uma funcionária da instituição, de maneira a representar um subconjunto de pessoas que compõem o todo, possibilitando a realização da coleta e a análise dos dados obtidos.

O levantamento bibliográfico foi realizado na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (Brapci), com os seguintes termos: SEI,

SIGAD¹ e Ciclo de Vida dos Dados. Os trabalhos recuperados foram comparados, e permitiu a análise e discussão junto aos dados obtidos na visita e entrevista realizada na Embrapa com a arquivista responsável pelo arquivo da Embrapa/Londrina.

A entrevista ocorreu a partir de um roteiro que teve como foco identificar como identificar como ocorre a atividade de Protocolo (as etapas e processos compreendidos).

Para isso, enviou-se um e-mail marcando o horário mais adequado para a entrevista. Após o aceite da instituição, foi apresentado à arquivista o roteiro de entrevista, sendo a entrevista em si realizada no Arquivo da Embrapa, no dia 01 de outubro de 2019. A entrevista ocorreu com base em um roteiro estruturado e gravada por meio de áudio com o consentimento da participante obtido por meio da assinatura do Termo de Consentimento Livre e Esclarecido. Durante a realização da entrevista pôde-se também observar e coletar algumas informações em campo, uma vez que a visita permitiu acessar o ambiente estudado.

4. Resultados

A entrevista realizada permitiu caracterizar os fatores do CVD que podem contribuir com a fase de coleta dos dados na atividade de Protocolo presente no SEI da Embrapa Soja Londrina a partir da visão da arquivista que aqui será identificada como “Participante”.

Sobre a **privacidade**, a Participante mencionou que este fator se apresenta principalmente no que concerne às informações institucionais, ao sigilo dos documentos (compreendido como algo relativo a segredos de um negócio), a propriedade intelectual, documento preparatórios, às questões éticas e às informações pessoais. Tal resposta é justificada por meio da fala de Santana (2016, p. 125) que argumenta que para garantir a privacidade é necessário “[...] identificar, nas fontes utilizadas, aspectos que possam configurar quebra de privacidade de pessoas ou instituições

relacionadas aos dados que estão sendo coletados [...]”. Na análise, é possível compreender que o fator privacidade é levado em conta no processo de coleta de dados realizados pelo SEI.

Ainda sobre a privacidade e sua relação com os dados, Rockembach (2020, p. 104) diz que o “[...] cruzamento de dados e o acesso sem controle podem expor os indivíduos a situações complicadas nos meios pessoal e profissional.” A exposição pode ocorrer devido as falhas de segurança, vazamento de dados e venda ilegal de informações pessoais, casos assim são noticiados com frequência e reforçam a necessidade de regulamentação sobre os limites no acesso e publicação dos dados coletados e processados por organizações.

A respeito do fator **integração**, a Participante mencionou que o SEI não interage com outros sistemas da instituição. A Participante mencionou que, para o SEI fazer interoperabilidade com outro SEI, ele precisa de um sistema que se chama “barramento”, de forma que os órgãos possam conversar a nível de sistema, e realizar o trâmite de documentos com outros órgãos governamentais que usam a plataforma de forma eletrônica, sem a necessidade de imprimir tudo e reinserir no sistema. Tal questão remete à **interoperabilidade**, a qual Frigo (2012) argumenta que algumas questões que devem ser consideradas são o formato dos documentos digitais - que devem ser possíveis de leitura em diferentes sistemas e não se limitar a softwares proprietários -, assim como saídas que possibilitem a leitura desses documentos por diferentes sistemas em um longo período – já que a longa duração desses documentos o exigem.

Para Sant’Ana (2016, p. 125) a integração na fase de coleta deve se preocupar com aspectos relacionados a “[...] identificação e validação dos atributos que serão responsáveis pela identificação unívoca de cada registro [...] e seus correspondentes nas outras entidades [...] para que a integração possa ser garantida.”

Sobre a **qualidade** e procedência dos dados coletados, a participante menciona que é difícil avaliar dados externos à instituição, já que eles não podem

¹ Sistema Informatizado de Gestão Arquivística de Documentos (SIGAD).

estabelecer padrões de avaliação a esses dados, uma que a coleta acontece de forma manual. Sant'Ana (2016) menciona que é essencial que elementos que permitam a percepção sobre a qualidade de dados coletados estejam presentes, entre eles a garantia de integridade física e lógica, de forma que ele possa ser utilizado futuramente.

A respeito dos **direitos autorais**, a participante menciona que de forma geral os dados não envolvem direito autoral, exceto quando pertencem a outra organização. A Participante da entrevista citou um exemplo que pode ocorrer quando alguma organização solicita da Embrapa uma amostra de um determinado material, que pode ser tanto uma semente, quanto alguma tecnologia. Conforme a participante a Embrapa pode fornecer essa amostra, só que para isso é feito um acordo de transferência de material.

Sobre os **direitos autorais**, é essencial manter atenção ao responsável da fonte de maneira que os direitos a ele garantidos por meio de legislação sejam respeitados, sendo necessário consultar as informações sobre direito de acesso aos dados, assim como a vinculação financeira de usos (SANT'ANA, 2016). Ainda que a participante mencione que os direitos autorais não são recorrentes nos documentos da instituição que atua, é sempre importante criar padrões para o caso de ser necessário se apliquem, evitando problemas futuros com questões legais.

A **disseminação** também é um dos fatores apresentados no CVD, sobre esse aspecto a participante relata que a disseminação é realizada por meio de tramitação de processos, que ocorrem de duas formas, a primeira quando o documento não é sigiloso, em que ele é tramitado para a área ou setor e todos os colaboradores podem ter acesso a este documento e às informações que estão sendo tramitadas. E, a segunda, quando o documento é considerado sigiloso, que a disseminação ocorre por meio de tramitação individual. Ou seja, quando o documento não é sigiloso, ele tramita por setor, quando é sigiloso, só é liberado por meio de atribuição de processo, para as pessoas que podem ter acesso àquela informação. Pode-se observar a partir

da fala da Participante, que a privacidade se faz presente durante a presença deste fator, pois há a preocupação com os dados que são considerados sigilosos na instituição.

Quanto ao último fator, que se trata da **preservação**, a participante relatou que a instituição desenvolve alguns procedimentos para a preservação de dados e documentos eletrônicos. Um desses procedimentos refere-se, segundo ela, ao software utilizado para a edição dos documentos arquivísticos. Por exemplo, o SEI só permite a inserção de documentos externos quando estes são editados por meio de *software* livre.

Sobre o uso de **softwares livres**, Frigo (2012, p. 32) argumenta que “[...] as especificações produzidas e mantidas por órgãos normalizados são mais seguras e tem mais garantia de disponibilidade por longo prazo, sendo assim, são os mais recomendados para preservação digital.” Como exemplo ele menciona o *software Libre Office*, cujo formato é o *Open Document Format (ODF)*, cuja escolha está relacionada com a independência de fornecedores privados, autonomia do usuário e transparência, visto que esse padrão é livre e não exige pagamento de royalties.

Outros cuidados mencionados pela Participante referem-se à organização dos documentos. Segundo ela, é essencial garantir que tudo esteja devidamente classificado, sendo necessário frequentemente fazer uma avaliação, gerar estatísticas para auxiliar na gestão do ciclo de vida dos documentos, assim como descartar alguns documentos que já cumpriram o ciclo. O SEI permite o descarte dos documentos e garante a segurança das informações descartadas, assim como melhor aproveitamento das informações em todas as fases do processamento.

Sobre a **preservação**, Sant'Ana (2016, p. 126) argumenta sobre sua importância na gestão eletrônica.

A preservação dos dados coletados e que eventualmente venham a ser armazenados, também podem exigir que dados adicionais sejam incluídos nos pré-requisitos definidos para a coleta, proporcionando que estes dados possam ser identificados de forma mais ampla e incorporando,

inclusive, informações sobre eventuais características de dispositivos que tenham sido utilizados como fonte dos dados, permitindo que estes dados sejam não só preservados mas também utilizados, mesmo após inevitáveis alterações em suas estruturas e constituições semânticas advindas de evolução nos dispositivos, tais como aumento de acurácia e ou precisão, com eventuais, e muito prováveis, melhorias nos níveis de granularidade de informações.

Sobre a presença dos indicadores do CVD no processo de coleta realizado no arquivo da Embrapa, foi possível perceber que de alguma maneira todos os indicadores foram contemplados nos processos mencionados pela Participante. Ainda que alguns fatores como direitos autorais não se apliquem com frequência aos documentos coletados, produzidos e armazenados pela instituição, ele aparece como uma das preocupações que a instituição possui em manter os dados de maneira adequada.

É possível pensar que, mesmo que alguns indicadores possam ser melhor aplicados no processo arquivístico da instituição e melhor estabelecidos no SEI enquanto sistema de gerenciamento eletrônico, de forma geral os fatores foram identificados e mostram que o arquivo em questão já se preocupa com aspectos essenciais da gestão de informações e dados em meio eletrônico.

Sobre o objetivo, que era identificar como o CVD poderia contribuir com a fase de coletas do SEI, é possível responder que muitos dos fatores já têm sido aplicados de forma simples no arquivo estudado, porém, é importante mencionar que os fatores estabelecidos por Sant'Ana (2016) podem contribuir para o aprimoramento da aplicação no arquivo e sistema em questão estudado, assim como em outras instituições e sistemas de gerenciamento.

5. Conclusão ou Considerações Finais

No que diz respeito aos fatores presentes no CVD e suas contribuições para a coleta de dados em um SEI, entende-se que muitos dos fatores presentes já compõem o processo de coleta de dados e

protocolo do SEI. Entretanto, compreende-se que são necessários estudos mais profundos que analisem a presença dos fatores também em outros processos do SEI, assim como aprimoramento da aplicação dos fatores visando o melhoramento da gestão arquivística.

Reflete-se que a sistematização de aplicação dos fatores do CVD pode contribuir para as adequações do SEI ao e-ARQ. Brasil, assim como potencializar as ferramentas de gestão eletrônica, permitindo uma maior integridade, confiabilidade, qualidade, segurança e autenticidade aos documentos digitais.

6. Referências

BAHIA. **O que é SEI**. [2022]. Disponível em: <http://www.portalseibahia.saeb.ba.gov.br/o-que-e-o-sei>. Acesso em: 20 ago. 2022.

BRASIL. CONSELHO NACIONAL DE ARQUIVOS. CÂMARA TÉCNICA DE DOCUMENTOS ELETRÔNICOS. **e-ARQ Brasil**: modelo de requisitos para sistemas informatizados de gestão arquivística de documentos. Conselho Nacional de Arquivos, 2011. Disponível em: <http://www.siga.arquivonacional.gov.br/images/publicacoes/e-ARQ.pdf>. Acesso em: 23 ago. 2022.

EMBRAPA. **Missão, visão e valores**, 2022. Disponível em: <https://www.embrapa.br/missao-visao-e-valores>. Acesso em: 13 set. 2022.

FRIGO, D. **Preservação Digital**: um subsídio para o Centro de Artes e Letras da UFSM. 161 p. Dissertação (Mestrado em Patrimônio Cultural) - Universidade Federal de Santa Maria, Centro de Ciências Sociais e Humanas, Santa maria/ RS, 2012.

ROCKEMBACH, M. Estudos de usuários de arquivo e os desafios da Lei Geral de Proteção de Dados. **Acervo**, v. 33, n. 3, p. 102-115, 2020.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da

informação. **Informação & Informação**, v. 21, n. 2, p. 116-142, 2016.

