

XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - XXV ENANCIB

GT 8 – Dados, Informação e Tecnologia

FASES DO CICLO DE VIDA DOS DADOS NO SISTEMA DE INFORMAÇÃO SOBRE A BIODIVERSIDADE BRASILEIRA

DATA LIFE CYCLE PHASES IN THE BRAZILIAN BIODIVERSITY INFORMATION SYSTEM

Amanda Garcia Gomes – Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp)
Campus de Marília

Iris Alves Miranda Negrão – Universidade Federal do Pará (UFPA)

Enne Rebeca Silva de Freitas – Universidade Federal do Pará (UFPA)

Fernando de Assis Rodrigues – Universidade Federal do Pará (UFPA)

Modalidade: Trabalho Completo

Resumo: o acesso a dados requer uma infraestrutura adequada para coleta, armazenamento e recuperação. Nesse contexto, foi selecionado o Sistema de Informação sobre a Biodiversidade Brasileira, voltado à gestão e ao compartilhamento de dados sobre a biodiversidade em nível nacional e internacional. Esta pesquisa teve como objetivo identificar as fases do Ciclo de Vida dos Dados presentes nesse sistema, com ênfase nas etapas de coleta, armazenamento e recuperação, além de suas formas de interação. A pesquisa é de natureza básica, abordagem qualitativa e caráter exploratório. Os resultados foram obtidos por meio de um instrumento de análise baseado no modelo do Ciclo de Vida dos Dados da Ciência da Informação, destacando características e relações entre as fases analisadas. As considerações finais indicam que os procedimentos metodológicos empregados podem subsidiar futuras análises em outros repositórios, contribuindo para o aprimoramento das práticas de gestão, interoperabilidade e reutilização de dados sobre a biodiversidade.

Palavras-chave: ciclo de vida dos dados; dados de pesquisa; repositório de dados.

Abstract: access to data requires an adequate infrastructure for collection, storage, and retrieval. In this context, the Brazilian Biodiversity Information System was selected to manage and share biodiversity data at the national and international levels. This research aimed to identify and describe the phases of the Data Life Cycle present in this system, emphasizing the stages of collection, storage, and retrieval, in addition to their forms of interaction. The research is basic in nature, qualitative in approach, and exploratory in nature. The results were obtained using an analysis tool based on the Information Science Data Life Cycle model, highlighting the characteristics and relationships between the phases analyzed. The final considerations indicate that the methodological procedures used can support future analyses in other repositories, contributing to improving management practices, interoperability, and the reuse of biodiversity data.

Keywords: data life cycle; research data; data repository.

1 INTRODUÇÃO

A geração de conhecimento científico contemporâneo precisa de uma infraestrutura informacional capaz de armazenar, trocar e acessar conjuntos de dados de pesquisa (Conrad *et al.*, 2024). Nesse sentido, iniciativas nacionais e internacionais promovem a integração de

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

dados de pesquisa, com o objetivo de ampliar o acesso por outras instituições. Por meio desse compartilhamento, é possível fomentar o intercâmbio de informações científicas, fortalecendo a cooperação entre diferentes agentes de pesquisa (Li *et al.*, 2022).

O intercâmbio de dados por instituições de pesquisa contribui para o desenvolvimento científico e tecnológico de países envolvidos neste processo, devido ao potencial de gerar novas soluções, tecnologias, além de gerar alternativas para o enfrentamento de problemas sociais. Este último pode reduzir impactos ambientais e aumentar a preservação de ecossistemas por meio da análise baseada em dados.

Dados de pesquisa oriundos de projetos, de programas e de redes científicas têm o potencial de transformação social, quando são usados para tomada de decisão. Contudo, sem um sistema de informação que possa permitir a coleta, o armazenamento e a recuperação de dados de pesquisa, especialmente para o reuso, a capacidade de modificação social fica reduzida, se comparado ao processo verificado em outros países. Em especial, é importante enfatizar que a infraestrutura da ARPANet (precursora da Internet) é fruto de uma demanda militar, mas também científica, pois permitiu um menor tempo de resposta nas trocas de mensagens e dados entre universidades participantes (Packard, 2023).

Diante da necessidade de desenvolvimento tecnológico, o acesso a dados sobre sustentabilidade na região Amazônica requer uma infraestrutura informacional capaz de disseminar dados de pesquisa. Entretanto, considerando elementos tecnológicos e organizacionais para esses dados, a estruturação de sistemas de informação carece de aprofundamento sobre os momentos pelos quais os dados passam por transformações até a disponibilização, um problema que motiva esta pesquisa.

A partir disso, foi selecionado um sistema de informação sobre a biodiversidade, iniciativa do governo brasileiro, que direciona esforços para o compartilhamento de dados de pesquisa: o Sistema Brasileiro sobre Biodiversidade (SiBBr). O SiBBr é uma plataforma online que disponibiliza um sistema de informação integrando dados, metadados descritivos e informações sobre a biodiversidade e os ecossistemas, originários de diferentes fontes, tornando-os acessíveis para usos diversos (SiBBr, 2024).

Dada a relevância do SiBBr como repositório de dados sobre biodiversidade em âmbito nacional e internacional, e considerando a importância de compreender suas práticas de gestão de dados, este estudo tem como objetivo identificar as fases do Ciclo de Vida dos Dados

presentes no sistema, com ênfase nas etapas de coleta, armazenamento e recuperação, bem como analisar as formas de interação adotadas.

Delimita-se a análise às fases de coleta, armazenamento e recuperação, compreendidas como componentes do processo de acesso a dados de pesquisa no SiBBr, a partir da instrumentalização do Ciclo de Vida dos Dados (Sant'Ana, 2016). Dessa forma, surge a seguinte questão: como as fases do Ciclo de Vida dos Dados estão presentes no SiBBr e de que maneira ocorrem as interações no processo de acesso aos dados de pesquisa disponibilizados pelo sistema?

2 MARCO TEÓRICO

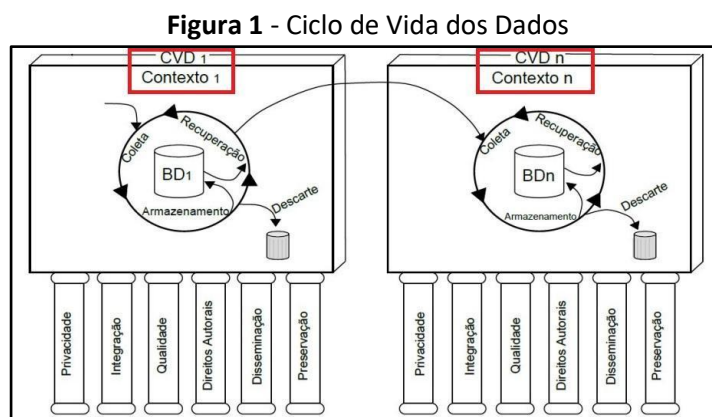
Diversos modelos de Ciclo de Vida dos Dados (CVD) têm sido propostos no cenário internacional, de modo geral, o CVD pode ser definido como um *framework* ou modelo que descreve as fases de gestão de dados, desde a coleta até a reutilização ou descarte, visando garantir qualidade, e boas práticas no gerenciamento de dados. A sua utilidade está na capacidade de atuar como instrumento de apoio aos pesquisadores em diferentes estágios de suas pesquisas científicas.

De acordo com Sayão e Sales (2015) há iniciativas internacionais como *Curation Lifecycle Model*, *Combined Lifecycle Model*, *Research Data Lifecycle* e *Data Lifecycle* são amplamente utilizadas por pesquisadores e instituições de pesquisa. Além desses, outros modelos são citados por Rodrigues (2021) como o *DataONE Lifecycle framework* (inclui etapas como Planejar, Coletar, Assegurar a Qualidade, Descrever, Preservar, Descobrir, Integrar, Analisar e Reutilizar) voltado para dados científicos, o *Curation Lifecycle* (envolve Conceptualizar, Criar ou Receber, Avaliar e Selecionar, Ingerir, Preservar, Armazenar, Acessar, Usar e Reutilizar e Transformar) para a curadoria digital e gestão de dados, e por fim, o *Data Documentation Initiative* (Planejamento, Coleta, Processamento, Distribuição, Arquivamento e Reutilização) para dados de ciências sociais.

No contexto brasileiro, destaca-se o Ciclo de Vida dos Dados para a Ciência da Informação, proposto por Sant'Ana (2016), como contribuição teórica adaptada às especificidades da área. Neste modelo de Sant'Ana (2016) propõe quatro fases: Coleta, Armazenamento, Recuperação e Descarte, complementadas por fatores transversais como privacidade, integração, qualidade, direitos autorais, disseminação e preservação que sustentam todas as fases do modelo.

Assim, reconhecendo a contribuição teórica deste modelo voltado para a Ciência da Informação, adotou-se o CVD com base em Sant'Ana (2016) para atingir o objetivo desta pesquisa. Nessa perspectiva, a fase Coleta corresponde ao momento de elaboração de estratégias para localizar e avaliar os dados, definindo as ferramentas necessárias. Na fase de Armazenamento, define-se os dados que serão disponibilizados, bem como da estrutura física e lógica utilizada para o registro.

A fase de Recuperação consiste no processo que viabiliza que os dados sejam encontrados, acessados e compreendidos por humanos ou por máquinas. Por último, a fase de Descarte é o momento que ocorre a limpeza ou desativação da base de dados, podendo incluir a exclusão completa do registro ou apenas de um atributo específico (Sant'Ana, 2016). Além disso, existem fatores transversais a todas as fases: privacidade, integração, qualidade, direitos autorais, disseminação e preservação que permeiam em todas as fases (Figura 1).



Fonte: elaborada pelos autores (2025), adaptado de Sant'Ana (2016).

A aplicação prática de um Ciclo de Vida dos Dados pressupõe um contexto em que os dados estão sob o domínio de um agente, responsável pela coleta de dados e, posteriormente, realiza concessões para a disponibilização desses dados (Sant'Ana, 2019). Na Figura 1, o CVD₁ possui as quatro fases sustentadas pelos seis fatores transversais atuando no Contexto₁.

Em seguida, o CVD₁ poderá disponibilizar dados por meio da sua fase de Recuperação, que pode ser acessado pela fase de Coleta de um outro ciclo, em distintos contextos (CVD_n). O CVD_n não necessariamente é um processo do agente detentor dos dados. Esta ligação entre CVD₁ (fase de Recuperação) e CVD_n (fase de Coleta) produz diversos fenômenos a serem observados, como, por exemplo, o dado coletado pelo CVD_n esteja vinculado a diretrizes distintas do que fora acordado no CVD₁.

Em Sistemas de Informação, a coleta, o processamento (que abrange um potencial armazenamento) e a recuperação de dados está vinculada a Teoria de Sistemas (Principia [...]),

2025). É parte da própria concepção do funcionamento da computação moderna, no qual os sistemas de informação com interfaces interativas aplicam estes mesmos referenciais, especialmente os voltados para ambientes *online* (Shaw, 1986). Ou seja, um sistema de informação *online* necessariamente coleta, armazena e permite a recuperação de dados como parte de sua essência.

O termo composto *dado de pesquisa* possui um significado amplo e se transforma de acordo com os domínios científicos, com os objetos de pesquisa, com as metodologias de geração e a coleta de dados, entre outras variáveis. Os repositórios de dados de pesquisa rapidamente tornaram-se parte essencial da infraestrutura mundial de pesquisa, pois atuam como uma estrutura capaz de criar condições para a gestão de dados resultantes de pesquisas realizadas no ecossistema científico (Sayão; Sales, 2016).

O SiBBr foi instituído para o compartilhamento de dados coletados e publicados pelas instituições brasileiras gerados dentro de universidades, institutos de pesquisa e agências de fomento, atuando como repositório de dados de pesquisa sobre biodiversidade. Este repositório de dados adota o padrão de dados *Darwin Core* (DwC), considerado o padrão mais utilizado para dados de biodiversidade, pois foi desenvolvido para fornecer uma maneira simples de documentar e compartilhar informações sobre ocorrências de espécies.

Além disso, também utiliza o *Ecological Metadata Language* (EML) um padrão de metadados que registra informações sobre conjuntos de dados ecológicos em documentos *eXtensible Markup Language* (XML) modulares e expansíveis. As características do SiBBr relacionadas às fases de Coleta, Armazenamento e Recuperação serão detalhadas na seção resultados e discussão.

3 PROCEDIMENTOS METODOLÓGICOS

Trata-se de uma pesquisa de natureza básica, de abordagem qualitativa e com objetivo exploratório, concentrando-se no acesso a dados sobre a biodiversidade da Amazônia disponibilizados no SiBBr. Foi elaborado um instrumento de análise desenvolvido com base no referencial estruturante do Ciclo de Vida dos Dados para a Ciência da Informação (Sant'Ana, 2016), contemplando características relacionadas às fases de Coleta, Armazenamento e Recuperação.

A fase de Descarte não foi analisada, ficando fora do escopo da análise porque não houve elementos suficientes identificados no percurso desta pesquisa para o

**XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB
Rio de Janeiro, RJ - 03 a 07 de novembro de 2025**

aprofundamento desta fase. Para permitir a replicabilidade, desenvolveu-se um roteiro contendo as características e perguntas que norteiam a análise de cada fase. O Quadro 1 apresenta as características e as perguntas norteadoras segmentadas para cada fase analisada, na forma de modelo de análise qualitativa estruturada.

Quadro 1 - Modelo de análise qualitativa estruturada

Fase	Características	Perguntas norteadoras
Coleta	Necessidades informacionais	Quais necessidades informacionais orientam a coleta no SiBBr?
	Estratégias de localização e avaliação	Que estratégias são usadas para localizar e avaliar os dados antes da coleta?
	Mecanismos de coleta	Quais mecanismos (fontes, agentes) são utilizados para obter os dados?
	Metodologias e ferramentas	Quais metodologias ou ferramentas são aplicadas na obtenção dos dados?
	Documentação da coleta	A coleta está registrada de forma estruturada e acessível?
	Controle de qualidade inicial	Existem procedimentos para garantir qualidade, autenticidade e integridade na coleta?
Armazenamento	Estrutura lógica dos dados	Quais variáveis e atributos foram definidos para representar os dados coletados?
	Organização semântica dos dados	Como as variáveis foram organizadas em subconjuntos ou entidades com base em significados comuns ou relações com elementos do mundo real?
	Formato e padrão de representação	Em que formatos e padrões os dados estão sendo armazenados? Os padrões adotados permitem preservar a semântica e promover a interoperabilidade?
	Local de armazenamento	Onde os dados estão armazenados (servidores próprios, nuvem, instituições parceiras)? Que tipo de infraestrutura tecnológica foi utilizada?
	Acesso aos dados armazenados	Quem pode acessar os dados? Existem restrições legais, éticas ou operacionais que limitam o acesso a dados sensíveis ou pessoais?
	Mecanismos de acesso e gerenciamento	Como os dados são acessados (interface direta, <i>Application Programming Interfaces</i> , sistemas gerenciadores de banco de dados)? Existem ferramentas específicas para consulta, extração ou visualização?
	Governança e manutenção dos dados	Quais instituições ou atores são responsáveis pela manutenção, atualização e preservação dos dados? Existe um modelo de governança estabelecido?
	Garantia de reutilização futura	As definições adotadas na estrutura e nos padrões permitem que os dados sejam compreendidos e reutilizados ao longo do tempo? Há aderência a princípios como FAIR?
Recuperação	Estratégias de disponibilização	Que estratégias são utilizadas para tornar os dados acessíveis e promover sua utilização por diferentes perfis de usuários?
	Modalidades de acesso aos dados	De que forma os dados podem ser acessados? Os usuários podem fazer download de conjuntos de dados, consultar por meio de <i>Application Programming Interfaces</i> , ou apenas visualizar?
	Ferramentas de busca e filtragem	Existem mecanismos de busca, filtros, indexadores ou outras ferramentas que facilitam a recuperação dos dados dentro da plataforma?
	Formatos de saída e exportação	Em quais formatos os dados podem ser exportados ou visualizados? Há suporte a padrões interoperáveis?

Fase	Características	Perguntas norteadoras
	Níveis de restrição de acesso	Existem limitações ou camadas de acesso diferenciadas conforme o tipo de usuário ou o tipo de dado (ex: dados sensíveis)?
	Requisitos para o uso dos dados	Há necessidade de autenticação, aceite de licenças específicas (como <i>Creative Commons</i>), ou fornecimento de informações pessoais para o acesso?
	Recursos de visualização	A plataforma oferece ferramentas de visualização dos dados (mapas, gráficos, painéis interativos)?
	Ações de incentivo à reutilização	Quais práticas são adotadas para incentivar a reutilização dos dados disponíveis (como documentação detalhada, tutoriais, padronização)?

Fonte: elaborado pelos autores (2025), adaptado de Sant’Ana (2016).

As perguntas norteadoras da fase de Coleta estão relacionadas às necessidades informacionais que orientam a coleta de dados no SiBBR, as estratégias para a localização e avaliação dos dados. Assim como, os mecanismos, as metodologias e as ferramentas que são aplicadas e utilizadas para obtenção dos dados. Finalmente, se a coleta está registrada de forma estruturada e acessível, além da existência de procedimentos para garantir qualidade, autenticidade e integridade.

Com relação às perguntas norteadoras relacionadas à fase de Armazenamento, elas possuem características fundamentais, como a definição de variáveis e de atributos que representam os dados. A organização semântica dos agrupamentos dos dados com base em significados. O formato desses dados para a preservação da semântica. Assim como a infraestrutura tecnológica utilizada para o armazenamento. O acesso a esses dados armazenados como as possíveis restrições legais, éticas ou operacionais, como por *Application Programming Interfaces (APIs)*, banco de dados, e às ferramentas disponíveis de consulta e visualização. A atualização, preservação e políticas de uso dos dados. Em síntese, a garantia dos princípios FAIR (*Findable, Accessible, Interoperable, and Reusable*) quando na gestão de dados de pesquisa, aprimoram a qualidade dos dados e a capacidade de reuso.

Já na fase de Recuperação, as perguntas norteadoras estão relacionadas às estratégias utilizadas para tornar os dados acessíveis e utilizáveis pelos usuários, se há indicação das modalidades de acesso e de que forma pode ser feito.

Incide ainda sobre a existência de ferramentas de busca e filtragem que facilitem a recuperação, os formatos que os dados podem ser exportados ou visualizados, os requisitos para o uso em relação às restrições de licenças de direitos autorais. Por fim, também indaga se há recursos de visualização como mapas, gráficos ou painéis interativos, em conclusão, assim como ações para incentivar a reutilização dos dados.

4 RESULTADOS E DISCUSSÃO

Inicialmente, é necessária a apresentação dos elementos identificados no SiBBr para subsidiar a discussão dos resultados da pesquisa. Essa identificação tem como objetivo contribuir para a compreensão das interações entre os ciclos e das funções desempenhadas no sistema. Consideram-se usuários aqueles que demonstram interesse, atual ou potencial, no uso dos dados ou em sua disponibilização para terceiros, fazendo uso do sistema de informação.

Os potenciais usuários do SiBBr são pesquisadores e membros da comunidade científica que acessam sua interface para utilizar dados de pesquisa. Os provedores de dados são instituições, organizações, iniciativas ou redes que possuem dados de pesquisa sobre biodiversidade, que compartilham e autorizam o SiBBr a divulgá-los (SiBBr, 2024).

Cada provedor possui um Ciclo de Vida dos Dados próprio que interage com o Ciclo de Vida dos Dados SiBBr para publicar dados. O SiBBr, por sua vez, é responsável por manter a infraestrutura tecnológica e organizacional necessária ao gerenciamento do banco de dados.

4.1 Fase de Coleta

A fase de Coleta inicia-se com a definição das necessidades informacionais que orientam a coleta pelo sistema de informação. Sobre as Estratégias utilizada para localizar e avaliar os dados antes da coleta, identificou-se que as fontes de dados são oriundas das ações de fomento à pesquisa científica promovidas pelas Fundações Estaduais de Amparo à Pesquisa, pelos Fundos Nacionais e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), incluindo ainda a atuação de órgãos públicos e instituições nacionais de ensino e pesquisa (Brasil, 2018).

Quanto aos mecanismos de coleta, o SiBBr adota padrões e protocolos internacionais para o compartilhamento de dados, utilizando a infraestrutura *Living Atlas*, com módulos reutilizáveis por outras organizações (SiBBr, 2024). Os padrões e as linguagens utilizadas são reconhecidas pela comunidade científica das Ciências Biológicas, tais como o Dwc e o EML, com observância dos princípios FAIR (SiBBr, 2024).

O SiBBr adota como metodologia e ferramenta para obtenção de dados a aplicação de formulários a serem preenchidos, conforme o tipo de registro (por exemplo, registro de ocorrência em coleções biológicas ou em projetos de ciência cidadã). Esses registros são

formas de coleta que o sistema de informação realiza e que podem ter configurações específicas, de acordo com os dados que se pretende inserir.

Os formulários são estruturados em sete formas distintas para o preenchimento. É possível realizar registros para: i) publicar um conjunto de dados; ii) cadastrar um projeto de pesquisa; iii) cadastrar um projeto de ciência cidadã; iv) registrar a biodiversidade; v) registrar uma coleção; vi) registrar uma instituição; e vii) registrar um provedor de dados (SiBBr, 2024).

A documentação da coleta está estruturada a partir do padrão DwC dentro de um repositório de colaboração e de tarefas¹ sobre o preenchimento de cada campo da planilha de dados. Esse repositório é denominado de Avaliação dos Termos *DarwinCore* com o objetivo de identificar termos com vocabulário controlado e definir um conjunto de boas práticas no uso do padrão no Brasil, junto com as instituições e organizações que produzem dados.

Sobre o controle de qualidade inicial o SiBBr aponta que os dados sobre biodiversidade devem estar disponíveis, pesquisáveis e reutilizáveis, de acordo com os Princípios FAIR. Entretanto, não apresentou procedimentos suficientes que garantam a qualidade, os requisitos de autenticidade e integridade na coleta. As características da fase de coleta foram dispostas no Quadro 2 em que a coluna ao lado esquerdo apresenta as características, seguida da coluna a direita que apresenta a síntese das respostas da pergunta norteadora.

Quadro 2 - Resultados identificados no SiBBr para a Fase de Coleta do Ciclo de Vida dos Dados

Características	Resposta da pergunta norteadora
Necessidades informacionais	Tem como objetivo reunir dados primários sobre a biodiversidade e os ecossistemas do país, fomentando estudos e políticas públicas ambientais (Brasil, 2018).
Estratégias de localização e avaliação	As fontes de dados são oriundas de ações de fomento à pesquisa científica, promovidas por agências como a Fundação Estadual de Amparo à Pesquisa, a CAPES, os fundos nacionais e as instituições públicas de ensino e pesquisa.
Mecanismos de coleta	Adota padrões e protocolos internacionais para o compartilhamento de dados, utilizando a infraestrutura Living Atlas e utilizando padrões de metadados e linguagens reconhecidas pela comunidade científica das Ciências Biológicas, tais como o DwCe o EML.
Metodologias e ferramentas	Os formulários são estruturados em sete formas distintas para o preenchimento. É possível realizar registros para: i) publicar um conjunto de dados; ii) cadastrar um projeto de pesquisa; iii) cadastrar um projeto de ciência cidadã; iv) registrar a biodiversidade; v) registrar uma coleção; vi) registrar uma instituição; e vii) registrar um provedor de dados (SiBBr, 2024).
Documentação da coleta	Os dados coletados devem estar acompanhados de metadados conforme padrões internacionais, garantindo contexto e qualidade da coleta. As entradas no sistema são mediadas por formulários específicos e exigem preenchimento estruturado.
Controle de qualidade inicial	Apesar de apontar que os dados sobre biodiversidade devem seguir os princípios FAIR, o sistema de informação não apresentou procedimentos suficientes que garantam a qualidade, os requisitos de autenticidade e integridade na coleta.

Fonte: elaborado pelos autores (2024).

4.2 Fase de Armazenamento

A fase de Armazenamento apresentou limitações, nesta pesquisa, para uma análise mais aprofundada do enfoque tecnológico. Por esse motivo, não foi possível discorrer sobre especificações como tipo, tamanho, formato ou aspectos semânticos dos dados, pois essas definições são de responsabilidade exclusiva dos gestores do sistema de informação, pertencente ao Ministério da Ciência, Tecnologia e Inovação (MCTI).

Para superar essa lacuna, recomenda-se que pesquisas futuras estabeleçam parcerias institucionais com a equipe gestora do SiBBr, de forma a possibilitar o acesso controlado a informações técnicas restritas, como configurações de servidores, formatos internos de armazenamento e protocolos de segurança. Outra possibilidade é a realização de entrevistas com técnicos do MCTI e da RNP, complementando as evidências documentais com informações operacionais, o que pode ampliar a compreensão sobre as práticas de armazenamento de dados no sistema.

Diante disso, optou-se por identificar, na documentação jurídica que instituiu o SiBBr e em seções da sua página oficial (Quem Somos, O que é o SiBBr, Perguntas Frequentes), algumas características que permitem compreender de forma geral como ocorre o armazenamento dos dados.

Quanto à estrutura lógica dos dados na fase de Armazenamento, não foi possível identificar com profundidade técnica os tipos e domínios de variáveis ou os conjuntos de variáveis utilizados no SiBBr. Essa limitação se dá pelo caráter técnico-operacional dessas definições, restritas à equipe gestora da infraestrutura do sistema.

Sobre a organização semântica dos dados, a pesquisa também não conseguiu detalhar como os dados estão organizados semanticamente em subconjuntos vinculados a conceitos do mundo real. No entanto, sabe-se que o SiBBr adota o padrão DwC e a linguagem EML, o que indica uma organização padronizada e com potencial semântico reconhecido internacionalmente.

O formato e padrão de representação dos dados armazenados adotado pelo SiBBr é o DwC para dados e a linguagem EML para metadados, promovendo interoperabilidade, preservação semântica e alinhamento com iniciativas globais de dados abertos. A adoção desse padrão e protocolo também está em conformidade com a Infraestrutura Nacional de Dados Espaciais (INDE), que é um sistema que visa integrar e harmonizar informações geoespaciais produzidas por diferentes órgãos do governo brasileiro.

A governança e manutenção dos dados é de responsabilidade do MCTI em parceria com a Rede Nacional de Ensino e Pesquisa (RNP) operacionalizando a plataforma. A governança do SiBBr é estabelecida por meio de um Comitê Gestor composto pelos seguintes participantes: RNP, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Instituto Nacional de Pesquisas da Amazônia (INPA), Museu Paraense Emílio Goeldi (MPEG), Instituto de Desenvolvimento Sustentável Mamirauá (IDSM), Instituto Nacional da Mata Atlântica (INMA) e o Ministério do Meio Ambiente e Mudança do Clima (Brasil, 2018; SiBBr, 2024).

Em relação ao Local de armazenamento observou-se que os dados coletados são armazenados na infraestrutura do *Living Atlas* que permite o reuso dos dados tenha alcance internacional. Entretanto, não foi possível identificar o detalhamento em relação ao tipo de infraestrutura utilizada, se em servidores próprios ou nuvem.

Quanto às restrições de acesso a dados armazenados, serão observadas a Lei dos Direitos Autorais e a Lei de Acesso à Informação, são imprescindíveis para a terceira fase (Recuperação), pois, ao disponibilizar um conjunto de dados no SiBBr, o usuário que deposita os dados seleciona um tipo de licença *Creative Commons*, autorizando o uso dos dados, de acordo com as restrições impostas pela licença escolhida (SiBBr, 2024).

Embora o foco maior esteja na disponibilização pública, há restrições quanto a dados sensíveis e observância da legislação nacional, como a Lei Geral de Proteção de Dados (LGPD). Os mecanismos de acesso e gerenciamento buscam identificar se existem ferramentas específicas para consulta, extração ou visualização.

Assim o SiBBr indicou que o acesso aos dados se dá por meio da plataforma digital do SiBBr, mas a documentação não detalha os mecanismos técnicos utilizados (como Sistemas Gerenciadores de Banco de Dados, APIs, ou interfaces específicas). A operacionalização é feita pela RNP, o que sugere o uso de tecnologias robustas de gerenciamento e distribuição de dados.

No que se refere à garantia de reutilização futura, observou-se que o SiBBr utiliza o padrão internacional DwC (Darwin Core) e a linguagem internacional EML (Ecological Metadata Language), a adoção de uma infraestrutura de código aberto (*Living Atlas*) e a conformidade com a INDE indicam que o SiBBr está alinhado com princípios como FAIR, garantindo interoperabilidade, acessibilidade e reusabilidade dos dados. O Quadro 3 apresenta a síntese dos resultados identificados na fase de Armazenamento.

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

Quadro 3 - Resultados identificados no SiBBr para a Fase de Armazenamento do Ciclo de Vida dos Dados

Características	Resposta da pergunta norteadora
Estrutura lógica dos dados	Não foi possível identificar com profundidade técnica os tipos e domínios de variáveis ou os conjuntos de variáveis utilizados no SiBBr.
Organização semântica dos dados	A pesquisa não conseguiu detalhar como os dados estão organizados semanticamente. No entanto, sabe-se que o SiBBr adota o padrão DwC e a linguagem EML, o que indica uma organização padronizada e com potencial semântico reconhecido internacionalmente.
Formato e padrão de representação	O DwC e o EML promovem interoperabilidade, preservação semântica e alinhamento com iniciativas globais de dados abertos. A adoção também está em conformidade com a INDE.
Governança e manutenção dos dados	A responsabilidade pela implementação e sustentação do SiBBr é do MCTI. A Rede Nacional de Ensino e Pesquisa (RNP) operacionaliza a plataforma. A governança é realizada por um Comitê Gestor com representantes de instituições científicas e ambientais (CNPq, IBICT, INPA, MPEG, IDSM, INMA, MMA).
Local de armazenamento	Não foi possível identificar o detalhamento em relação ao tipo de infraestrutura utilizada, se em servidores próprios ou nuvem.
Acesso aos dados armazenados	O acesso aos dados segue regulamentações jurídicas e está condicionado a políticas institucionais. Embora o foco maior esteja na disponibilização pública, há restrições quanto a dados sensíveis e observância da legislação nacional, como a LGPD..
Mecanismos de acesso e gerenciamento	As informações disponíveis indicam que o acesso aos dados se dá por meio do SiBBr, mas a documentação não detalha os mecanismos técnicos utilizados (como Sistemas Gerenciadores de Banco de Dados, <i>Application Programming Interfaces</i> , ou interfaces específicas).
Garantia de reutilização futura	O emprego do padrão DwC e da linguagem EML, aliado à adoção de uma infraestrutura de código aberto (Living Atlas) e à conformidade com a INDE, evidencia que o SiBBr está alinhado a princípios como FAIR, assegurando interoperabilidade, acessibilidade e reusabilidade dos dados.

Fonte: elaborado pelos autores (2025).

4.3 Fase de Recuperação

A fase de Recuperação apresentou estratégias de disponibilização para tornar os dados acessíveis, tendo em vista a sua utilização por diferentes perfis de usuários. Assim, o SiBBr atua como mediador entre os provedores de dados e os usuários em geral, permitindo acesso aos conjuntos de dados. Essa mediação envolve disponibilizar dados por meio de sua interface envolvendo os dados coletados das instituições, dos projetos de ciência cidadã e das redes de pesquisa.

Quanto a modalidade de acesso aos dados, a recuperação pode variar conforme a consulta realizada, uma vez que a arquitetura da informação do sistema oferece ao usuário diferentes formas de visualização. Por exemplo, na página inicial, há seções identificadas como *Espécies*, *Instituições*, *Coleções*, *Conjunto de Dados* e *Ocorrências*. A partir delas, o usuário pode acessar dados específicos, sem que o conteúdo fique restrito apenas a esses temas. Além disso, é possível refinar as buscas para obter resultados mais precisos.

As ferramentas de busca e filtragem permitem a consulta aos dados por meio de mecanismos de busca, filtros ou outras formas de busca que facilitam a recuperação dos dados no SiBBr. Os dados de pesquisa das instituições estão organizados no Catálogo de Coleções Biológicas Científicas do Brasil (classificados por Fauna, Insetos, Microrganismos, Plantas), podendo ser consultados no total ou a partir das quatro classificações.

Neste catálogo constam as instituições que depositam conjuntos de dados organizados em coleções. Por exemplo, uma instituição registrada com Embrapa Floresta possui quatro coleções que reúnem dados de pesquisa gerados pelos pesquisadores da instituição. Identificou-se apenas que o formato de saída e exportação dos dados podem ser feitos via *download* dos arquivos em formato de planilha eletrônica.

Nesse caso, um pesquisador interessado em saber a taxonomia encontrará os metadados do nome científico, reino, filo, classificação, etc. Para reutilizar os dados disponíveis na coleção é necessário estar de acordo com o Termo de Uso para dados publicados no SiBBr, fornecer um endereço de e-mail, renomear o arquivo e indicar a razão de transferência.

Sobre os níveis de restrição de acesso, o uso dos conjuntos de dados é condicionado ao Termo de Uso de Dados que podem estar sujeitos aos condicionantes de licença impostos pelo usuário que publicou os conjuntos de dados (SiBBr, 2024). Além disso, os interessados em reusar os dados devem fornecer informações pessoais precisas e atualizadas sempre que solicitado para uso de determinada ferramenta ou para o *download* de dados. Essas informações são registradas no sistema e regidas pela legislação brasileira de proteção de dados pessoais.

O SiBBr possui recursos de visualização em que disponibiliza um recurso de visualização gráfica dos dados da coleção, permitindo a análise por estado da federação, mês, década e família. Esse recurso contribui para que o pesquisador perceba variações e identifique em quais estados há maior depósito de dados sobre o espécime. Para garantir a perenidade e possibilitar o acesso futuro, cada coleção recebe um identificador único, juntamente com o espaço temporal, o contato do publicador e o *website* da instituição provedora.

Em termos de incentivo à reutilização, foi identificado que as ações concentram-se em boas práticas no depósito dos dados das instituições de pesquisa no SiBBr como atribuição de identificadores únicos para as coleções, disponibilização de metadados detalhados (padrão

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

DwC), estatísticas de uso, contato com o publicador, visualização histórica e territorial dos dados contribuem para a reutilização e comparação entre instituições.

Além das iniciativas que envolvem a sociedade, os projetos de ciência cidadã, compostos por voluntários ao redor do mundo que documentam registros das espécies, o quantitativo de propagação de doenças infecciosas, tendências populacionais, monitoramento de alterações na paisagem e mudanças climáticas (SiBBr, 2024). O Quadro 4 apresenta a síntese das características identificadas nesta fase no SiBBr.

Quadro 4 - Resultados identificados no SiBBr para a Fase de Recuperação do Ciclo de Vida dos Dados

Características	Resposta da pergunta norteadora
Estratégias de disponibilização	O SiBBr atua como mediador entre os provedores de dados e os usuários em geral, permitindo o acesso aos conjuntos de dados. Essa mediação envolve disponibilizar esses dados por meio de sua interface envolvendo os dados coletados das instituições, dos projetos de ciência cidadã e das redes de pesquisa.
Modalidades de acesso aos dados	Os dados podem ser acessados por meio de busca direta na plataforma, visualização interativa, <i>download</i> condicionado ao aceite do Termo de Uso, ou por meio de <i>Application Programming Interfaces</i> indiretas via integração com outras plataformas como o iNaturalist.
Ferramentas de busca e filtragem	A interface do SiBBr permite diversas formas de consulta: por espécie, projeto, instituição, área geográfica, ocorrência, região, etc. Também há filtros que permitem refinar as buscas dentro das seções como Espécies, Instituições, Conjunto de Dados, Ocorrências, etc.
Formatos de saída e exportação	Os dados podem ser exportados em formato de planilha eletrônica, contendo metadados no padrão DwC.
Níveis de restrição de acesso	O acesso aos dados está sujeito às condições do Termo de Uso e, em alguns casos, aos condicionantes de licença definidos pelo usuário que publicou os dados.
Requisitos para o uso dos dados	Para reutilizar os dados, é necessário aceitar o Termo de Uso, informar um e-mail, indicar a razão da transferência e renomear o arquivo. Licenças específicas podem ser aplicadas conforme definido pelo usuário publicador dos dados.
Recursos de visualização	O SiBBr oferece visualização gráfica das coleções por estado, década, mês, família, qualidade, etc. Os dados podem ser explorados por metadados taxonômicos e geográficos, além de visualização por mapas e estatísticas.
Ações de incentivo à reutilização	Atribuição de identificadores únicos para as coleções, disponibilização de metadados detalhados (padrão DwC), estatísticas de uso, contato com o publicador, visualização histórica e territorial dos dados contribuem para a reutilização e comparação entre instituições.

Fonte: elaborado pelos autores (2025).

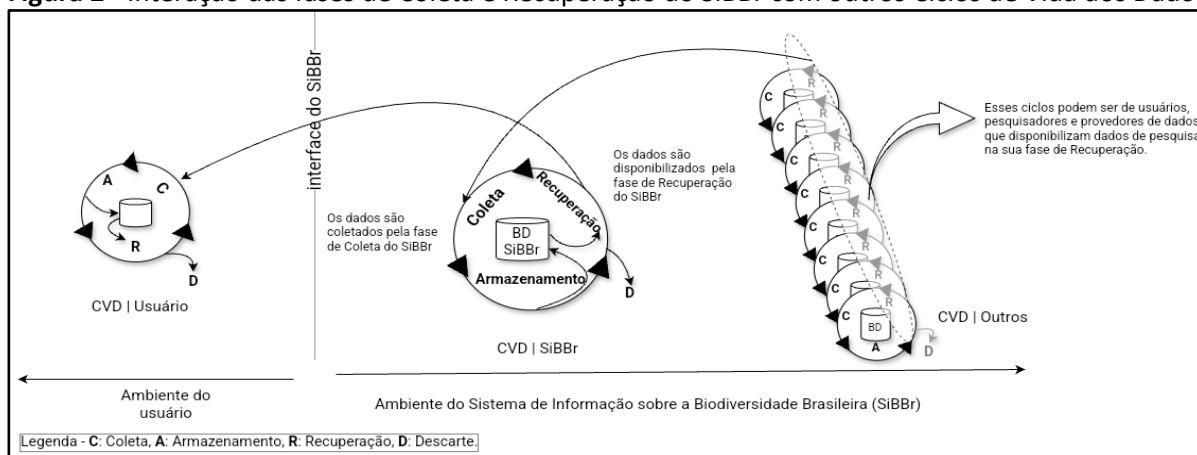
4.4 Interação entre as fases de Coleta e Recuperação no SiBBr

Em contextos que envolvem a geração, disseminação e uso de dados, é possível identificar múltiplos Ciclos de Vida dos Dados atuando de maneira interconectada em torno de um mesmo elemento alvo, que, no caso desta pesquisa, é o SiBBr. Dependendo da

perspectiva de análise, o papel dos atores envolvidos no acesso a dados será de quem coleta (quem obtém) ou de quem disponibiliza os dados. A Figura 2 ilustra uma representação da interação entre as fases de Coleta e Recuperação do SiBBr com outros Ciclos de Vida dos Dados.

A partir da perspectiva do SiBBr, a fase de Coleta corresponde ao momento em que usuários, pesquisadores e provedores de dados inserem informações de suas pesquisas no sistema. Já a fase de Recuperação está relacionada à disponibilização desses dados e informações coletadas para que possam ser acessados por outros usuários por meio da interface do SiBBr.

Figura 2 - Interação das fases de Coleta e Recuperação do SiBBr com outros Ciclos de Vida dos Dados



Fonte: elaborada pelos autores (2025).

Grande parte do conteúdo no SiBBr, tais como os registros de ocorrência, dados ambientais, imagens e estatuto de conservação das espécies são provenientes de conjuntos de dados fornecidos por instituições de coleta, coletores individuais e grupos comunitários (SiBBr, 2024). Em termos quantitativos, o SiBBr possui 195 instituições registradas, 588 coleções e 726 conjuntos de dados. Esses dados oferecem aos interessados a possibilidade de comparar o crescimento dos registros da coleção com os dados de uma ou mais instituições que pesquisam o mesmo espécime.

5 CONSIDERAÇÕES FINAIS

A análise das fases no Sistema de Informação sobre a Biodiversidade Brasileira evidenciou a robustez e a complexidade de sua infraestrutura informacional, bem como a relevância do sistema na promoção da gestão eficiente e do intercâmbio de dados de pesquisa. A articulação entre as fases de Coleta e Recuperação revelou-se estratégica,

XXV Encontro Nacional de Pesquisa em Ciência da Informação - XXV ENANCIB Rio de Janeiro, RJ - 03 a 07 de novembro de 2025

impulsionando não apenas a disponibilização, mas também o potencial de reuso qualificado dos dados.

Tal interação destaca o SiBBr como um modelo de referência para outras iniciativas de gestão de dados de pesquisa, especialmente em contextos que demandam ampliação do acesso, transparência e integração de informações. Ainda que o sistema tenha avanços significativos, identificaram-se aspectos que merecem aprofundamento, como o aprimoramento dos fluxos de integração entre diferentes Ciclos de Vida dos Dados e a ampliação das estratégias para garantir a persistência e a qualidade a longo prazo.

Esta pesquisa reforça a importância de abordagens que reconheçam a complexidade dos sistemas informacionais e sua centralidade para a produção, disseminação e reuso de dados de pesquisa, especialmente no campo da biodiversidade, área de crescente importância estratégica para o Brasil e para a comunidade científica internacional.

Espera-se que a estrutura dos procedimentos metodológicos adotados nesta pesquisa sirva como subsídio para análise das fases do Ciclo de Vida dos Dados em outros repositórios de dados de pesquisa, fortalecendo práticas de gestão, interoperabilidade e reutilização de dados de pesquisa, sobretudo no contexto dos estudos sobre a biodiversidade brasileira.

Além do contexto do SiBBr, o modelo de análise adotado pode ser adaptado para a avaliação de outros repositórios e sistemas de informação, como aqueles voltados à saúde, educação ou dados geoespaciais. Essa adaptabilidade permite comparações entre plataformas, favorece a identificação de boas práticas e orienta a proposição de melhorias alinhadas a padrões internacionais, ampliando a aplicabilidade do estudo para diferentes áreas do conhecimento.

REFERÊNCIAS

BRASIL. Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) . Portaria MCTIC n. 6.223, de 29 de novembro de 2018. Dispõe sobre o Sistema de Informação sobre a Biodiversidade Brasileira – SiBBr. **Diário Oficial da União**, Brasília, DF, 30 nov. 2018. Seção 1, p. 40. Disponível em:
https://antigo.mctic.gov.br/mctic/opencms/legislacao/portarias/Portaria_MCTIC_n_6223_de_29112018.html . Acesso em: 08 ago. 2025

CONRAD, T. O. F. *et al.* Making mathematical research data FAIR: pathways to improved data sharing. **Scientific Data**, v. 11, n. 676, 2024. Disponível em:
<https://doi.org/10.1038/s41597-024-03480-0>. Acesso em: 20 maio 2025.

LI, R. *et al.* Health data sharing platforms: serving researchers through provision of access to

high-quality data for reuse. **Health Data Science**, v. 2022, Article ID 9768384, p. 3, 2022. Disponível em: <https://doi.org/10.34133/2022/9768384>. Acesso em: 21 maio 2025.

PACKARD, N. INTERNET Prehistory: ARPANET Chronology. **Cogent Social Sciences**, v. 9, n. 2, 2023. Disponível em: <https://doi.org/10.1080/23311886.2023.2245237>. Acesso em: 20 maio 2025.

PRINCIPIA Cybernetica Web. Input-output analysis, 2025. Disponível em: <http://pespmc1.vub.ac.be/ASC/INPUT-ANALY.html>. Acesso em: 22 maio 2025.

RODRIGUES, D. G. A. **Elementos de ciclos de vida dos dados no percurso metodológico das teses brasileiras da área de Ciência da Informação**: um estudo diagnóstico. 2021. 183f. Tese (Doutorado em Ciência da Informação) — Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Paraíba, João Pessoa, 2021. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/21993>. Acesso em: 22 maio 2025.

SANT'ANA, R. C. G. Campo informacional resultante da interação de ciclos de vida dos dados. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. (Org.). **Dados científicos**: perspectivas e desafios. João Pessoa: UFPB, 2019. p. 33-52. Disponível em: <https://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/book/359>. Acesso em: 22 maio 2025.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da Informação. **Informação & Informação**, Londrina, v. 21, n. 2, p. 116–142, 2016. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27940>. Acesso em: 22 maio 2025.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90-115, mai./ago., 2016. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 22 maio 2025.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando os conceitos de acesso livre. **RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde**, v. 8, n. 2, p. 76-92, 2015. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>. Acesso em: 22 maio 2025.

SiBBR. **Sistema de Informação sobre a Biodiversidade Brasileira**. 2024. Disponível em: <https://www.sibbr.gov.br>. Acesso em: 08 set. 2024.

SHAW, M. An input-output model of interactive systems. **Association for Computing Machinery**, v. 17, n. 4, 1986. Disponível em: <https://doi.org/10.1145/22339.22382>. Acesso em: 22 maio 2025.