

## RELATÓRIO TÉCNICO ITV DS

# MAPEANDO OS DADOS DE BIODIVERSIDADE JÁ EXISTENTES NO GRUPO DE BIODIVERSIDADE E SERVIÇOS DE ECOSSISTEMAS DO INSTITUTO TECNOLÓGICO VALE

## RELATÓRIO FINAL DO PROJETO BIODIVERSIDADE 4.0

**Andre Luis Acosta**  
**Leonardo Trevelin**  
**Valéria Tavares**  
**Tereza Cristina Giannini**

**Belém / PA**  
**Março / 2022**

<b>Título:</b> Mapeando os dados de biodiversidade já existentes no grupo de Biodiversidade e Serviços de Ecossistemas do Instituto Tecnológico Vale	
<b>PROD.TEC.ITV DS – N013/2022</b>	<b>Revisão</b>
<b>Classificação:</b> ( ) Confidencial ( ) Restrito ( x ) Uso Interno ( ) Público	<b>00</b>

**Informações Confidenciais** - Informações estratégicas para o Instituto e sua Mantenedora. Seu manuseio é restrito a usuários previamente autorizados pelo Gestor da Informação.

**Informações Restritas** - Informação cujo conhecimento, manuseio e controle de acesso devem estar limitados a um grupo restrito de empregados que necessitam utilizá-la para exercer suas atividades profissionais.

**Informações de Uso Interno** - São informações destinadas à utilização interna por empregados e prestadores de serviço

**Informações Públicas** - Informações que podem ser distribuídas ao público externo, o que, usualmente, é feito através dos canais corporativos apropriados

#### Citar como

ACOSTA, Andre Luis, *et al.* **Mapeando os dados de biodiversidade já existentes no grupo de Biodiversidade e Serviços de Ecossistemas do Instituto Tecnológico Vale.** Belém: ITV, 2022. (Relatório Técnico N013/2022) DOI 10.29223/PROD.TEC.ITV.DS.2022.13.Acosta

#### Dados Internacionais de Catalogação na Publicação (CIP)

A185 Acosta, Andre Luis  
Mapeando os dados de biodiversidade já existentes no grupo de Biodiversidade e Serviços de Ecossistemas do Instituto Tecnológico Vale. / Andre Luis Acosta ... [et al.]. - Belém: ITV, 2022.  
40 p. : il.

Relatório Técnico (Instituto Tecnológico Vale) – 2022  
PROD.TEC.ITV.DS – N013/2022  
DOI 10.29223/PROD.TEC.ITV.DS.2022.13.Acosta

1. Gestão de dados - Mapeamento. 2. Gestão de dados – Coleta padronizada. 3. Termos e metadados - Sistema. 4. Biodiversidade - Dados. I. Trevelin, Leonardo. II. Tavares, Valéria. III. Giannini, Tereza Cristina. IV. Título

CDD 23. ed. 622.752098115

## RESUMO EXECUTIVO

Decisões demandam dados, e dados de baixa qualidade poderão levar a decisões equivocadas, imprecisas ou tardias. A forma como os dados são coletados, armazenados e compartilhados irá reverberar na sua qualidade e precisão, por consequência, refletindo-se na capacidade de compreensão dos aspectos que eles representam. O setor privado atuante na área ambiental demanda objetividade e assertividade, e por isso é imprescindível tratar os dados que subsidiam ações ambientais, como de conservação e restauração, com excepcional atenção. Para avaliar o estado da biodiversidade e impactos ambientais, muitas vezes são necessárias extensas pesquisas de campo; para isso, são contratados prestadores de serviços independentes, os quais são especializados na obtenção de uma variedade de tipos de informações. Por conseguinte, uma diversidade de métodos de coleta é aplicada, aspecto que gera incongruências metodológicas de análise e representação das informações, e diferenças na formatação e apresentação dos dados resultantes. Em momento posterior, visando à integração destas informações em bancos de dados para uma análise ampla das informações, torna-se necessário extraí-las individualmente, ajustá-las e uniformizá-las, gerando uma demanda inteiramente nova, consumindo tempo, esforço humano e recursos financeiros. Além disso, durante a realização desta tarefa, aumenta-se o risco de erros de interpretação, digitação e digitalização, podendo comprometer a qualidade, ou até levar a perdas de informações. Estes aspectos negativos podem ser evitados por meio da padronização de dados nos processos de coleta, armazenamento e compartilhamento. A padronização de dados torna-se uma solução estratégica para aumentar a eficiência, reduzir custos operacionais e riscos de degradação de informações. Mas além destes benefícios, traz uma série de outras vantagens, como a transformação do sistema de registro tradicional em campo (cadernos de campo) para um formato inteiramente digital, com a integração de câmeras, *tablets*, *dataloggers*, e outras tecnologias de sensoriamento amplamente disponíveis, permitindo que o fluxo de dados desde a coleta em campo até as análises ocorra em alta velocidade e precisão. Visando elaborar uma recomendação de padronização de dados de forma abrangente e inclusiva, mapeamos os dados de biodiversidade utilizados pelos pesquisadores do grupo Biodiversidade e Serviços de Ecossistemas do ITV. Por meio desta estratégia, buscamos entender as tipologias e estruturas de dados, como eles têm sido usados, armazenados e compartilhados em bases de dados, mas também observamos suas convergências e peculiaridades entre cada pesquisador e entre áreas de pesquisa. Com a participação dos pesquisadores, buscamos elaborar e validar um sistema preliminar de termos e metadados, incluindo recomendações de boas práticas a fim de aprimorar o uso de dados ambientais e sobre a biodiversidade. O mapeamento evidenciou uma série de correspondências quanto aos tipos de dados usados pelo grupo BES-ITV, em especial aqueles dados em estudos de Conservação e Restauração, Ecologia de Paisagens, Genética e *Radio Frequency Identification*. Mas também uma grande diversidade de tópicos de pesquisas, focando-se em seis grandes grupos biológicos, aspectos que evidenciam a alta multidisciplinaridade e ampla cobertura dos dados ambientais, ecológicos, genéticos e de biodiversidade usados pelo grupo. Com base nestes resultados, um sistema de termos e metadados foi proposto como modelo preliminar, e para seu uso idealizamos um sistema modular para a geração automática de planilhas de campo, para simplificar a coleta por meio inteiramente digital.

## RESUMO

A padronização de dados de biodiversidade é fundamental para aumentar a eficiência na coleta, curadoria e compartilhamento de dados e reduzir os gastos da empresa com potenciais erros e com a dificuldade na interpretação e localização dos dados. Isso acontece porque os padrões de dados também incluem a definição e descrição dos mesmos, definições estas chamadas de metadados. Os metadados garantem que todos aqueles que adicionam, alteram e/ou consultam os dados - mesmo que sejam pessoas diferentes em momentos diferentes - tenham uma noção perfeita do significado daquela informação. Aqui, propomos gerar um protocolo de recomendação para coleta de dados (em campo, laboratório, museus, herbários, coleções biológicas). Este protocolo agrega os padrões de dados já consolidados pelo padrão internacional Darwin Core (DwC), mas também estamos criando novos termos, classes e respectivos metadados de particular interesse. Para avaliar todos os tipos de dados pesquisados e armazenados, entrevistamos pesquisadores da equipe Biodiversidade e Serviços (BES) do ITV, de diversas formações científicas, sobre suas necessidades específicas de dados e sobre as definições de seus termos recomendados (metadados). Com este procedimento, engajamos as pessoas no processo construtivo, criando um protocolo de dados mais inclusivo e propenso à aceitação geral. Além disso, a estratégia de engajamento de pessoal também garante que todas as possíveis demandas de dados sejam cobertas pelo protocolo. O mapeamento evidenciou uma série de correspondências quanto aos tipos de dados usados pelo grupo da BES-ITV. Evidenciou também uma grande diversidade de tópicos de pesquisas, focando-se em seis grandes grupos biológicos, aspectos que demonstram a alta multidisciplinaridade e ampla cobertura dos dados ambientais, ecológicos, genéticos e de biodiversidade usados pelo grupo.

**Palavras chave:** coleta padronizada de dados, mapeamento e gestão de dados, dados ambientais e de biodiversidade, sistema de termos e metadados, protocolo de recomendações, sistema modular.

## ABSTRACT

Standardizing biodiversity data is critical to increasing efficiency in collecting, curating and sharing data and reducing company expenses with potential errors and difficulty in interpreting and locating data. This is because data standards also include their definition and description, definitions called metadata. Metadata ensures that everyone who adds, changes and/or consults the data - even if they are different people at different times - have a perfect sense of the meaning of that information. Here, we propose to generate a recommendation protocol for data collection (in the field, laboratory, museums, herbariums, biological collections). This protocol aggregates the data standards already consolidated by the international Darwin Core (DwC) standard, but we are also creating new terms, classes and respective metadata of particular interest. To assess all types of researched and stored data, we interviewed researchers from ITV's Biodiversity and Services (BES) team, from various scientific backgrounds, about their specific data needs and the definitions of their recommended terms (metadata). With this procedure, we engage people in the constructive process, creating a data protocol that is more inclusive and prone to general acceptance. In addition, the staff engagement strategy also ensures that all possible data demands are covered by the protocol. The mapping showed a series of correspondences regarding the types of data used by the BES-ITV group. It also showed a great diversity of research topics, focusing on six major biological groups, aspects that demonstrate the high multidisciplinary and wide coverage of environmental, ecological, genetic and biodiversity data used by the group.

**Keywords:** standardized data collection, mapping and data management, environmental and biodiversity data, terms and metadata system, recommendation protocol, modular system.