

INSTITUTO NACIONAL DA MATA ATLÂNTICA (INMA)

PESQUISADOR ADJUNTO 1 – CIÊNCIA DE DADOS APLICADA À BIODIVERSIDADE

Prova Escrita Discursiva

Aplicação: 28/01/2024

PADRÃO DE RESPOSTA DEFINITIVO

Por representar a riqueza de vida na Terra, em todas as suas formas e interações, a conservação da biodiversidade é essencial para a manutenção dos ecossistemas e, por extensão, para a sobrevivência humana. Sem dúvida, a seleção e o uso eficaz de fontes de dados se tornam fundamentais para embasar ações de conservação e pesquisa. A ciência de dados, com sua capacidade de processar e analisar grandes volumes de informações, desempenha um papel crucial neste processo, abrindo caminho para inovações na preservação da biodiversidade.

A seleção das fontes de dados é intrinsecamente ligada ao tipo de trabalho a ser realizado. Estudos de conservação, por exemplo, exigem dados de ocorrência georreferenciados para mapear a distribuição das espécies e avaliar seu *habitat*. Já as pesquisas taxonômicas demandam informações detalhadas sobre espécimes, muitas vezes depositadas em herbários e museus. Essa distinção sublinha a necessidade de se acessar uma variedade de fontes de dados, cada uma adequada a objetivos específicos.

No tocante à disponibilidade dessas fontes, tanto em âmbito nacional quanto internacional, diversas plataformas se destacam. No Brasil, o Sistema de Informação sobre a Biodiversidade Brasileira (SiBBR) e o speciesLink do Centro de Referência em Informação Ambiental (CRIA) são recursos inestimáveis que disponibilizam dados sobre a biodiversidade do País. Internacionalmente, o Global Biodiversity Information Facility (GBIF) serve como um repositório global, agregando dados de ocorrência de espécies de várias partes do mundo.

A execução do processo de extração, transformação e carga (ETL) dos dados é um ponto crítico que requer atenção especial à qualidade dos dados. Erros como a imprecisão de coordenadas geográficas e erros de digitação na transcrição de etiquetas de coletas de herbários são comuns e podem comprometer a confiabilidade das análises. A automação desse processo, por meio do desenvolvimento de *scripts* e o uso de API, pode otimizar significativamente a eficiência e a precisão na manipulação dos dados.

Além disso, a atenção ao padrão de metadados é fundamental. No contexto da documentação da flora, o padrão Darwin Core é amplamente adotado, fornecendo um *framework* consistente para a publicação e o compartilhamento de dados sobre a biodiversidade. Para dados geográficos, as diretrizes estabelecidas pela Infraestrutura Nacional de Dados Espaciais (INDE) do Instituto Brasileiro de Geografia e Estatística (IBGE) recomendam práticas e padrões a serem seguidos, assegurando a interoperabilidade e a qualidade dos dados espaciais.

Os mapas e pontos de ocorrência de espécies, fundamentais nesse contexto, fornecem uma visão geográfica da biodiversidade, permitindo a identificação de áreas prioritárias para a conservação e o planejamento de estratégias de manejo. A riqueza dos dados disponíveis sobre a biodiversidade brasileira, apesar de desafiadora em termos de qualidade e cobertura, oferece uma base sólida para ações conservacionistas, destacando-se o papel vital da ciência de dados na compreensão e preservação da biodiversidade.

Em suma, a integração eficaz de fontes de dados variadas, a atenção à qualidade dos dados e a adoção de padrões de metadados apropriados são essenciais para o avanço do conhecimento e da conservação da biodiversidade. No Brasil, o fortalecimento dessas práticas é crucial para proteger seu patrimônio natural, diante dos crescentes desafios ambientais. A colaboração entre cientistas, tecnólogos e conservacionistas, apoiada por uma infraestrutura de dados robusta, é a chave para uma conservação bem-sucedida da biodiversidade, garantindo a sustentabilidade dos ecossistemas para as futuras gerações.

QUESITO AVALIADO

Quesito 2.1 – Fontes de dados para o conhecimento da biodiversidade: (i) adequação da fonte de dados ao tipo de trabalho a ser realizado; (ii) exemplos de fontes de dados; (iii) técnicas ou recursos computacionais que otimizam o processo e asseguram a qualidade dos dados obtidos; (iv) necessidade do uso de metadados

Conceito 0 – Não tratou do tema ou o fez de forma totalmente equivocada.

Conceito 1 – Mencionou pontos principais do tema, mas não os desenvolveu.

Conceito 2 – Desenvolveu adequadamente apenas um dos itens especificados.

Conceito 3 – Desenvolveu adequadamente dois dos itens especificados.

Conceito 4 – Desenvolveu adequadamente três dos itens especificados.

Conceito 5 – Desenvolveu adequadamente os quatro itens especificados.