



Dados abertos e a proveniência dos dados: garantia da confiabilidade dos dados sobre os povos e biomas amazônicos

Open data and data provenance: ensuring the reliability of that on amazonian peoples and biomes

Datos abiertos y la procedencia de los datos: garantía de la confiabilidad de los datos acerca de los pueblos y biomas amazónicos

Ana Carolina Simionato Arakaki

Doutorado em Ciência da Informação, Universidade Estadual Paulista (UNESP), Brasil. Pós-Doutorado em Computação, Universidade de São Paulo (USP), Brasil. Professora Adjunta, Universidade Federal de São Carlos (UFSCar), Brasil. Coordenadora de Serviços Bibliográficos no Instituto Brasileiro de Informação em Ciência e Tecnologia (COBIB/IBICT).

<http://lattes.cnpq.br/9896600626524397>

<https://orcid.org/0000-0002-0140-9110>

Felipe Augusto Arakaki

Doutor e mestre em Ciência da Informação pela Universidade Estadual Paulista (UNESP), Marília, São Paulo, Brasil. Docente do curso de Biblioteconomia da Universidade de Brasília (UnB), Brasília, Distrito Federal, Brasil. Docente do Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de São Carlos (UFSCar).

<http://lattes.cnpq.br/5324289839207169>

<https://orcid.org/0000-0002-3983-2563>

Felipe Ivo da Silva

Mestrando em Ciência da Informação, Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil.

<http://lattes.cnpq.br/0753366586589793>

<https://orcid.org/0009-0005-1379-4692>

Ana Carolina Novaes de Mendonça

Mestranda em Ciência da Informação, Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil.

<http://lattes.cnpq.br/9826682103593479>

<https://orcid.org/0009-0004-0285-9932>

Resumo

Introdução: A região amazônica é reconhecida globalmente pela sua ampla biodiversidade e pela presença dos povos amazônicos, que são fontes valiosas de dados científicos. Nos últimos anos, os dados abertos se destacaram como uma ferramenta crucial na era digital, promovendo transparência e inovação. O blockchain é uma tecnologia desenvolvida para garantir a autenticidade e integridade dos dados em diversos contextos, e pode ser considerada uma alternativa para a gestão de dados da Amazônia. Este estudo tem como objetivo discutir a importância da proveniência e das iniciativas de dados abertos conectados em relação à viabilidade do uso da

tecnologia blockchain. **Metodologia:** Caracteriza-se por uma pesquisa teórica, exploratória e qualitativa sobre dados abertos e a proveniência dos dados na garantia da confiabilidade dos dados sobre os povos e biomas amazônicos. **Resultados:** O *Blockchain*, em conjunto com dados abertos e adoção de princípios vigentes, pode ser uma alternativa para garantir a proveniência dos dados sobre os povos e biomas amazônicos oferecendo controle sobre os dados e assegurando a veracidade das informações. **Conclusão:** A integração de dados abertos, a proveniência de dados e a tecnologia Blockchain podem estabelecer uma base para a gestão sustentável e conservação da Amazônia, beneficiando, além das comunidades locais, a comunidade global preocupada com a preservação ambiental e suas culturas.

Palavras-chave: blockchain; dados abertos; Região Amazônica; proveniência de dados.

Abstract

Introduction: The Amazon region is globally recognized for its vast biodiversity, as well as for the presence of Amazonian peoples, who are valuable sources of scientific data. In recent years, open data has emerged as a crucial tool in the digital age, promoting transparency and innovation. The blockchain is a technology developed to ensure the authenticity and integrity of data in various contexts, and it can be considered an alternative for managing data in the Amazon. This study aims to discuss the importance of provenance and connected open data initiatives regarding the feasibility of using blockchain technology. **Methodology:** It is characterized by theoretical, exploratory, and qualitative research on open data and data provenance in ensuring the reliability of data about Amazonian peoples and biomes. **Results:** Blockchain, along with open data and the adoption of current principles, can be an alternative to ensuring the provenance of data about Amazonian peoples and biomes, offering control over data and ensuring the accuracy of information. **Conclusion:** The integration of open data, data provenance, and Blockchain technology can establish a foundation for the sustainable management and conservation of the Amazon, benefiting local communities and the global community concerned with environmental preservation and its cultures.

Keywords: blockchain; open data; Amazon region; provenance of data

Resumen

Introducción: La región amazónica es reconocida a nivel mundial por su amplia biodiversidad, así como por la presencia de los pueblos amazónicos, que son fuentes valiosas de datos científicos. En los últimos años, los datos abiertos han surgido como una herramienta crucial en la era digital, promoviendo la transparencia y la innovación. El blockchain es una tecnología desarrollada para garantizar la autenticidad e integridad de los datos en diversos contextos, y puede considerarse una alternativa para la gestión de datos de la Amazonia. Este estudio tiene como objetivo discutir la importancia de la procedencia y las iniciativas de datos abiertos conectados en relación con la viabilidad del uso de la tecnología blockchain. **Metodología:** Se caracteriza por una investigación teórica, exploratoria y cualitativa sobre datos abiertos y la procedencia de los datos en la garantía de la confiabilidad de los datos sobre los pueblos y biomas amazónicos. **Resultados:** Blockchain, junto con datos abiertos y la adopción de principios vigentes, puede ser una alternativa para garantizar la procedencia de los datos sobre los pueblos y biomas amazónicos, ofreciendo control sobre los datos y asegurando la veracidad de la información. **Conclusión:** La integración de datos abiertos, la procedencia de datos y la tecnología Blockchain pueden establecer una base para la gestión sostenible y la conservación de la Amazonia, beneficiando a las comunidades locales y a la comunidad global preocupada por la preservación ambiental y sus culturas.

Palabras clave: Blockchain; datos abiertos; Región Amazónica; procedencia de los datos

1 INTRODUÇÃO

A região amazônica e sua cultura são globalmente relevantes não apenas pela sua rica biodiversidade e pelas comunidades indígenas que lá habitam, mas também pela valiosa fonte de dados científicos que oferecem. Por isso, é crucial que as informações sobre a Amazônia sejam preservadas de forma contínua: isso ajuda na compreensão de sua importância ecológica e cultural, e capacita a tomada de medidas eficazes para proteger e conservar esse patrimônio.

Nos últimos anos, os dados abertos emergiram como um recurso fundamental na era digital, promovendo transparência, inovação e progresso social. Associados a isso, os dados de proveniência tratam da rastreabilidade e história dos dados. Conforme Guandalini e Santos (2018) discutem, iniciativas de dados abertos visam aumentar a transparência e o compartilhamento para sua reutilização, o que se mostra particularmente relevante quando se considera o vasto acervo de dados científicos relacionados à Amazônia.

Neste contexto, ressalta-se a importância da autenticidade e preservação dos dados ao serem abertos e conectados. Para isso, há tecnologias desenvolvidas, como uma camada adicional de segurança e integridade para bases de dados, como o *Blockchain*. A relevância desse estudo justifica-se com a capacidade de contribuição para a temática, possibilitando a identificação já que é possível identificar e implementação de camadas de segurança e preservação que garantem a veracidade dos dados sobre os povos e biomas amazônicos, além da construção de políticas públicas mais direcionadas e assertivas.

De acordo com Bartling e Fecher (2017), as características descentralizadas, distribuídas, imutáveis e transparentes dos dados na *Blockchain* garantem confiabilidade e segurança, impactando diretamente a confiança na informação. Sendo assim, a transparência inerente à tecnologia *Blockchain* permite que os usuários verifiquem a origem e o histórico de cada dado registrado na rede, promovendo confiança no compartilhamento de informações sobre a região amazônica.

Ao adotar uma abordagem baseada em *Blockchain* para a gestão de dados, é possível reforçar ainda mais a confiabilidade e segurança das informações relacionadas aos povos e biomas amazônicos, contribuindo para sua preservação e conservação a longo prazo.

Neste sentido, o objetivo deste trabalho é discutir a importância da proveniência e das iniciativas de dados abertos conectados, com viabilidade do uso da tecnologia *Blockchain* para esse fim.

2 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa adota uma abordagem metodológica teórica, exploratória e qualitativa, visando aprofundar a compreensão sobre a temática em questão. Inicialmente, foram realizados um levantamento bibliográfico e uma revisão de literatura com o intuito de embasar o estudo.

A pesquisa bibliográfica, conforme definida por Gil (2008, p. 50), é conduzida a partir de material previamente elaborado, composto principalmente por livros e artigos científicos. Para isso, as bases de dados consultadas incluem a Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), *Web of Science* e *Scopus*.

Os critérios de inclusão englobam artigos completos, em português e inglês, indexados em bases científicas e publicados no período de 2008 a 2024. A estratégia de busca foi elaborada considerando a organização estruturada de termos para pesquisa em bases de dados, com o objetivo de obter os melhores resultados. As palavras-chave utilizadas foram "dados abertos", "proveniência de dados" e "*Blockchain*".

Como uma pesquisa qualitativa, este estudo busca compreender os fenômenos subjacentes às interações dos dados abertos, à proveniência dos dados e ao blockchain, bem como sua relação com o conhecimento gerado em relação aos povos e biomas amazônicos. Em relação aos objetivos, trata-se de uma pesquisa exploratória, caracterizada, segundo Gil (2008, p. 27), por ter como

finalidade "[...] desenvolver, esclarecer e modificar conceitos e ideias tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores".

3 RESULTADOS

A disseminação de dados em formato aberto tem se tornado uma prática global, visando promover transparência, colaboração e reutilização dessas informações. Embora tenham sido amplamente debatidas no âmbito governamental, tais diretrizes são aplicáveis a qualquer contexto, especialmente no meio científico.

Segundo a *Open Knowledge Foundation* (2019, não paginado, tradução nossa), "Dados e conteúdos abertos podem ser livremente usados, modificados e compartilhados por qualquer pessoa para qualquer finalidade". De acordo com o *OpenGovData* (2007), os dados abertos devem aderir a oito princípios fundamentais: serem completos, primários, atuais, acessíveis, processáveis por máquina, de acesso não discriminatório, em formatos não proprietários e isentos de licenças.

Adicionalmente, há sete princípios suplementares que abordam sua disponibilidade *on-line*: gratuidade, permanência, confiabilidade, presunção de abertura, documentação, segurança para abertura e concepção com contribuição pública. Diante das diversas iniciativas de publicação de dados abertos e da necessidade de orientações abrangentes, o *World Wide Web Consortium* (W3C) reuniu um grupo de especialistas para elaborar o documento "Boas Práticas para Dados na *Web*".

Esse documento apresenta 35 boas práticas e destaca oito benefícios associados à adoção dessas práticas: 1) Fornecer metadados; 2) Fornecer metadados descritivos; 3) Fornecer metadados estruturais; 4) Fornecer informações sobre a licença de dados; 5) Fornecer informações de proveniência dos dados; 6) Fornecer informações de qualidade de dados; 7) Fornecer indicador

de versão; 8) Fornecer o histórico de versão; 9) Usar URIs persistentes como identificadores de conjuntos de dados; 10) Usar URIs persistentes como identificadores dentro de conjuntos de dados; 11) Atribuir URIs para as versões dos conjuntos de dados e séries; 12) Usar formatos de dados padronizados legíveis por máquinas; 13) Usar representações de dados que sejam independentes de localidade (local e neutral); 14) Fornecer dados em formatos múltiplos; 15) Reutilizar vocabulários, dando preferência aos padronizados; 16) Escolher o nível de formalização adequado; 17) Fornecer *download* em massa (*bulk download*); 18) Fornecer subconjuntos para conjuntos de dados extensos; 19) Usar negociação de conteúdo para disponibilizar dados em formatos múltiplos; 20) Fornecer acesso em tempo real; 21) Fornecer dados atualizados; 22) Fornecer uma explicação para os dados que não estão disponíveis; 23) Disponibilizar dados por meio de uma API; 24) Usar padrões Web como base para construção de APIs; 25) Fornecer documentação completa para as APIs; 26) Evitar alterações que afetem o funcionamento de sua API; 27) Preservar identificadores; 28) Avaliar a cobertura do conjunto de dados; 29) Coletar *feedback* de consumidores de dados; 30) Compartilhar o *feedback* disponível; 31) Enriquecer dados por meio da geração de novos dados; 32) Fornecer visualizações complementares; 33) Fornecer *feedback* para o publicador original; 34) Obedecer os termos de licença; e 35) Citar a publicação original do conjunto de dados.

No contexto deste estudo, destaca-se a boa prática de "fornecer informações de proveniência dos dados", que é abordada neste trabalho como proveniência. A proveniência dos dados desempenha um papel essencial em várias áreas, incluindo compreensão, confiança e reuso dos dados. Conforme definido pelo W3C (2013), a proveniência refere-se à história dos dados, abrangendo aspectos como a sua coleta e origem. Compreender essa proveniência ajuda a avaliar a confiabilidade dos dados, além de fornecer um contexto para interpretar as informações. Haynes (2018, p. 134) também destaca a importância da proveniência na autenticidade dos dados, afirmando que "Quando se trata de estabelecer a autenticidade de um

item, sua história – sua proveniência – torna-se importante: as circunstâncias de sua criação, quem a possuiu e as condições sob as quais sua propriedade foi transferida". Assim, o registro adequado da proveniência dos dados, incluindo informações sobre seu criador, responsáveis pela publicação e guarda, bem como as atualizações e modificações, é fundamental para garantir a autenticidade e confiabilidade dos dados.

O W3C (2014) ressalta que a proveniência dos dados é uma maneira importante pela qual os consumidores podem avaliar a qualidade de um conjunto de dados, proporcionando contexto para sua interpretação e ajudando a determinar sua confiabilidade. Conforme observa Haynes (2018, p. 134), "A gestão de registros e a boa governança dependem da capacidade de demonstrar a autenticidade de um registro, fornecendo documentação sobre seu histórico e como foi gerenciado".

A associação das boas práticas do World Wide Web Consortium (W3C) com a blockchain pode potencializar significativamente a qualidade das informações disponíveis, uma vez que os dados de proveniência estariam munidos de sólidas e confiáveis propriedades. Assim, o registro adequado da proveniência dos dados, incluindo informações sobre seu criador, responsáveis pela publicação e guarda, bem como as atualizações e modificações, é fundamental para garantir a autenticidade e confiabilidade dos dados. Como observou Haynes (2018, p. 134), "No contexto dos materiais digitais, fornecer informações de proveniência pode ajudar a demonstrar que um registro não foi adulterado e que a evidência que ele apresenta é, portanto, confiável". Propõe-se, então, o uso da tecnologia *Blockchain* como alternativa para registrar as informações da produção científica, garantindo assim a proveniência dos dados.

Em 2006, Tim Bernes-Lee introduziu o conceito de Linked Data, atrelado à Web Semântica, visando criar ligações entre dados diversos e a sua posterior reutilização. Ressalta-se o conceito de Linked Data definido por Baker et al. (2011, não paginado, tradução nossa) "Linked Data refere-se a dados publicados em

conformidade com os princípios destinados a facilitar as ligações entre os conjuntos de dados, conjuntos de elementos e vocabulários de valor”.

Atrelado ao Linked Data, salienta-se o enriquecimento semântico que de acordo com Lira (2015, p.29) “[...] o enriquecimento semântico torna os dados e metadados mais qualificados, através do uso da semântica atribuída por vocabulários pré-existentes, sinônimos e informações de proveniência”. Os benefícios do *Blockchain* são destacados por Ouchi e Arakaki (2020, p. 70) por “[...] garantir a confiabilidade dos dados de forma transparente, descentralizada e imutável, e surgem como uma possibilidade para contribuir para o progresso da Ciência através de novas formas de tratamento das grandes quantidades de dados coletados”. Além disso, diversos estudos já discutem as possibilidades de relacionar o *Blockchain* com outros dados, como o uso dos princípios do *Linked data* e o enriquecimento semântico.

Ugarte (2017) propõe uma interessante relação entre os princípios do *Linked data* e a tecnologia *Blockchain*, denominada *Linked Blockchain Data ou Semantic Blockchain*. Estudos conduzidos por Third e Domingues (2019) destacam o potencial do *Blockchain* para registrar os preços de produtos, visando garantir uma comercialização justa entre os valores dos produtos dos agricultores e os preços praticados no mercado.

Kirstein e Hauswirth (2021) também ressaltam a viabilidade do uso do *Blockchain* para garantir a proveniência dos dados no contexto da publicação de dados abertos conectados. Assim, o uso do *Blockchain* emerge como uma alternativa promissora para garantir a proveniência da informação na comunicação científica, especialmente no que diz respeito aos povos e biomas amazônicos.

Além disso, ao combinar o *Blockchain* com dados abertos conectados, torna-se possível um melhor controle sobre o uso dos dados, incluindo a identificação de informações derivadas desse conjunto de dados. Essa abordagem destaca a importância da divulgação de dados abertos e enfatiza a necessidade de fornecer

informações detalhadas sobre a origem dos dados para garantir a preservação da informação.

Sendo assim, o uso da tecnologia *Blockchain* para registrar os dados de proveniência permite que todos os registros relacionados sejam armazenados de maneira transparente e acessível. No contexto científico, especialmente no que diz respeito aos povos e biomas amazônicos, a adoção do *Blockchain* em conjunto com dados abertos conectados pode proporcionar um controle mais eficaz sobre o uso dos dados, além de garantir a autenticidade das informações.

A literatura apresenta diversos estudos que revelam o potencial da tecnologia *Blockchain*, especialmente quando integrada às tecnologias da *Web Semântica*, como o *Linked data*.

4 CONSIDERAÇÕES FINAIS

Ao encerrar este trabalho, torna-se evidente que a transparência e a integridade dos dados são essenciais para a compreensão e preservação da Amazônia e suas diversas culturas.

Nesse sentido, a tecnologia *Blockchain* emerge como uma ferramenta promissora para garantir transparência e integridade, especialmente no que diz respeito às informações relacionadas aos povos e biomas amazônicos. Ao promover a confiabilidade dos dados e rastrear sua origem, a adoção do *Blockchain* pode desempenhar um papel significativo na preservação do conhecimento e na tomada de decisões informadas sobre a região amazônica. Entretanto, é importante destacar que o uso dessa tecnologia ainda requer estudos adicionais e considerações éticas para garantir seu uso responsável e eficaz.

A integração do *Blockchain* com tecnologias da *Web Semântica*, como o *Linked data*, aliada à adoção de iniciativas de dados abertos, representa uma abordagem promissora para garantir transparência, confiabilidade e proveniência dos dados em diversas áreas, especialmente no contexto dos povos e biomas amazônicos.

Dessa maneira, concluímos que a combinação de dados abertos, proveniência de dados e tecnologia Blockchain tem o potencial de estabelecer uma base sólida para a gestão sustentável e a conservação informacional sobre a Amazônia. Isso traz benefícios tanto para as comunidades locais quanto para a comunidade global, preocupada com a preservação cultural e científica do meio ambiente.

REFERÊNCIAS

- BAKER, T. *et al.* *Library Linked Data Incubator Group Final Report*. [S. l.]: W3C, 2011. W3C Incubator Group Report. Disponível em: <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>. Acesso em: 01 fev. 2024.
- BARTLING, Sönke; BENEDIKT, Fecher. *Blockchain for Open Science and Knowledge Creation: a technical fix to the reproducibility crisis?* [S. l.]: [s.n.], 2017.
- GIL, Antonio Carlos. *Como elaborar projetos de pesquisa*. 4. ed. São Paulo: Altas, 2008.
- GUANDALINI, Clara Alcina; SANTOS, Amanda Azevedo dos. LINKED OPEN DATA: conceito, relações e importância na Era da Informação. *Múltiplos Olhares em Ciência da Informação*, Belo Horizonte, v. 8, n. 2, 2018. Disponível em: <https://periodicos.ufmq.br/index.php/moci/article/view/16911>. Acesso em: 23 fev. 2024.
- HAYNES, David. *Metadata for Information Management and Retrieval: Understanding metadata and its use*. London: Facet Publishing, 2018.
- KIRSTEIN, Fabian; HAUSWIRTH, Manfred. *Blockchain for trustworthy publication and integration of Linked Open Data*. Association for Computing Machinery, New York, NY, USA, 2021, p. 269–272. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3460210.3493572>. Acesso em: 23 fev. 2024.
- LIRA, Márcio A. B. *Uma abordagem para o enriquecimento semântico de metadados para a publicação de dados abertos*. 2014. Dissertação (Mestrado em Ciência da Computação) - Centro de Informática, Programa de Pós-graduação em Ciência da Computação, Universidade Federal de Pernambuco, 2014. Disponível em: <https://repositorio.ufpe.br/handle/123456789/11570>. Acesso em: 10 abr. 2024.

OPEN GOV DATA. *8 principles of open government data*. California, 2007. Disponível em: <https://opengovdata.org/>. Acesso em: 23 fev. 2024.

OPEN KNOWLEDGE FOUNDATION. *Open definition*. 2019. Disponível em: <https://opendefinition.org/>. Acesso em: 23 fev. 2024.

SOUZA, Leticia P. de; SOUSA, Rodrigo Silva Caxias de; LÖW, Marieta Marques; BARROS, Thiago Henrique Bragato. Promovendo a justiça epistêmica: uma análise dos princípios CARE na gestão de dados de pesquisa em relação aos povos indígenas. *Anais do Workshop de Informação, Dados e Tecnologia - WIDaT, [S. l.]*, v. 6, 2023. DOI: 10.22477/vi.widat.73. Disponível em: <https://labcotec.ibict.br/widat/index.php/widat2023/article/view/73>. Acesso em: 21 fev. 2024.

THIRD, Allan; DOMINGUE, John. *LinkChains*: Exploring the space of decentralised trustworthy Linked Data. 2017. Disponível em: <https://ceur-ws.org/Vol-1934/contribution-06.pdf>. Acesso em: 21 fev. 2024.

UGARTE, Héctor. *A more pragmatic Web 3.0*: Linked BlockchainData. 2017. DOI: 10.13140/RG.2.2.10304.12807/1.

WORLD WIDE WEB CONSORTIUM (W3C). *Boas práticas para dados na Web*. Disponível em: <https://www.w3.org/Translations/DWBP-pt-BR/>. Acesso em: 23 fev. 2024.