Received: 27 September 2023 | Accepted: 7 December 2023

DOI: 10.1111/2041-210X.14280

Methods in Ecology and Evolution

APPLICATION

Innovations in Practice

ECKOchain: A FAIR blockchain-based database for long-term ecological data

Kjell-Erik Marstein[1] | John-Arvid Grytnes[1] | Robert John Lewis[2]

[1]University of Bergen, Bergen, Norway

[2]Norwegian Institute for Nature Research, Bergen, Norway

Correspondence
John-Arvid Grytnes
Email: jon.grytnes@uib.no

Funding information
Norges Forskningsråd, Grant/Award Number: 275681

Handling Editor: Aaron Ellison

Abstract

1. Open data practices in ecology are increasingly accepted, yet primary long-term ecological data remain hard to find. Barriers preventing open long-term ecological data include social and economic constructs such as a sense of data proprietorship and fear of misuse of complex datasets.
2. To incentivise open primary ecological data and ensure long-term preservation, we propose a decentralised data management approach using blockchain technology. The blockchain-based database is governed by transparent and immutable data management protocols, agreed on by members of the network. Specialised protocols ensure agreement in the network before new data is accepted, and no entity can single-handedly alter existing data.
3. We introduce the ECKOchain, a 'proof of concept' ecological blockchain-based database created with the Hyperledger Fabric framework. While metadata and access policies are distributed to all network members, primary data remains with data owners and are served on-demand to approved parties according to specified usage licences. Details of data requests are preserved indefinitely on the blockchain and serve as auditable data usage agreements.
4. With the distributed blockchain-based database we advocate for open science and transparency in long-term management of ecological data. The ECKOchain is also suitable for other scientific fields where auditability and transparency are considerations to long-term data management.

KEYWORDS
blockchain, database, dataset, decentralisation, ecology, open science

1 | INTRODUCTION

In an era of unprecedented global environmental change, open data are vital. Ecologists are increasingly tasked to address pressing societal questions requiring data spanning larger spatial scales and over longer time periods. This is a challenge that cannot be met individually. It requires collaborative research and, importantly, data prosperity (Hampton et al., 2013).

In the field of ecology and evolution, willingness to incorporate open data practices into the research flow has in recent years become broadly accepted (Soeharjono & Roche, 2021; Tenopir et al., 2020), with research data increasingly recognised as a scientific product of enduring value. Exemplified by thematic resource databases such as forestREplot[1] (Depauw & Maes, 2015), Global Biodiversity Information Facility (GBIF)[2] (Edwards et al., 2000) and BioTime[3] (Dornelas et al., 2018), this uptrend can also be attributed to journal and funder policies mandating the FAIR (findable, accessible, interoperable and reusable data) doctrine (Wilkinson et al., 2016), and to a coevolution of positive attitudes towards open science, with technological adoption excellent at expanding the volume, veracity and velocity of data streams (Farley et al., 2018) advancing the landscape of big data acquisition (Hampton et al., 2013).

In ecology, this is best embodied through popular citizen science programmes (e.g. eBird[4] and iNaturalist[5]), or automated sensor networks and remote sensing, where data streams are generated at unprecedented rates, and at scales and resolutions previously unavailable to ecologists. In contrast, most primary data in ecology collected by professional ecologists derive from studies that are naturally time inefficient, conducted over limited spatial and varying temporal scales (often years, sometimes decades) by many individual researchers.

Described as long-tail ecological science (Heidorn, 2008), reported social and economic constructs entwined to such low velocity ecological data, presents barriers to open data practice. These include fear of losing data exclusivity and being 'scooped' (Gewin, 2016; Huang et al., 2012; Mills et al., 2015), a sense of data proprietorship (Hampton et al., 2015) and concerns of how complex ecological datasets are used and interpreted (Mills et al., 2015; Roche et al., 2014). The result is ecological data that are hard to find (Poisot et al., 2019).

These concerns are not unique to the exchange of scientific data. Trust, transparency and control are values highly regarded in the transfer-exchange of various provisions of real-world and perceived value, and are properties that distributed ledger technologies, such as blockchains, are established to protect. Blockchains have proven suitable for use cases in areas such as finance, healthcare and supply-chain (ISO, 2022; ITU-T, 2019), as well as in the sciences with ETDB-Caltech, a blockchain-based database for electron tomography (Ortega et al., 2019).

We recently discussed the potential of blockchain technology as an alternative to traditional centralised databases, with a view to incentivise open primary ecological data (Lewis et al., 2023). Here, in this article, we present a first 'proof of concept' FAIR ecological database, developed using blockchain technology and designed with heightened transparency, autonomy and auditability to mobilise ecological resurvey data.

[1]forestREplot, Forest & Nature Lab at Ghent University (https://forestreplot.ugent.be).

[2]GBIF—Global Biodiversity Information Facility (https://gbif.org).

[3]BioTime, University of St Andrews (https://biotime.st-andrews.ac.uk).

[4]eBird, The Cornell Lab of Ornithology (https://ebird.org).

[5]iNaturalist (https://inaturalist.org).

2 | BLOCKCHAIN TECHNOLOGY

Blockchain is a category of distributed ledger technologies, which are systems for managing digital records of data (ledgers) maintained by multiple parties. It is often described in relation to cryptocurrency networks, for example Bitcoin (Nakamoto, 2008), however, cryptocurrencies must not be conflated with blockchain technology. The inherent features that make blockchain a secure and preferred platform for cryptocurrencies also make it excel at other types of data management and data verification (Zīle & Strazdiņa, 2018), providing a transparent, secure and trusted digital infrastructure for a wide range of domains, including the environmental sector (Parmentola et al., 2022).

Data submitted to a blockchain are organised in storage units known as blocks, which are cryptographically linked to create an immutable chain (Nakamoto, 2008). The chain, or 'ledger', is append-only so that data in existing blocks can neither be removed nor altered. New data are validated against predefined checks, and only when a certain number of participants (e.g. a majority) have verified and accepted the data will they be added. As the blockchain is immutable, any updates to data already on the chain must be treated as new data. Protocols can allow new data to hold a reference to existing data, where data in the block with the highest block number (where blocks are numbered in increasing order) are considered the latest version of the data.

The blockchain is distributed among network participants, so that all parties are in possession of the same data and together, through specialised consensus protocols, come to a conclusion on the current state (Xiao et al., 2020). Consequently, there is no single (centralised) source of truth. Some public blockchains reach consensus via computationally expensive 'proof of work' protocols (De Vries, 2018), although more energy efficient protocols are increasingly adopted (De Vries, 2022). Permissioned blockchains, where participants are known, reach consensus differently (Cachin & Vukoljic, 2017), in ways that do not induce the same computational power requirements as in public blockchains.

An increasingly important feature in blockchains is support for self-executing programs, known as smart contracts (Szabo, 1996). These are programs where developers can implement custom protocols that are executed in response to interactions with the blockchain. Since smart contracts are installed on the blockchain, network participants can verify that a contract's implementation does not change. This makes smart contracts ideal for implementing transparent policy-driven protocols, for example protocols for enforcing access control, governance and data validation. Effectively, smart contracts eliminate the need for a trusted intermediary to enforce a protocol (Szabo, 1997).

Users interact with a blockchain and its smart contracts through nodes. Nodes are entities in the network that run specialised software to maintain a copy of the blockchain. There are two types of blockchains: public and permissioned. Anyone can host nodes for public blockchains, while permissioned blockchains are member-only with nodes requiring individual approval to connect.

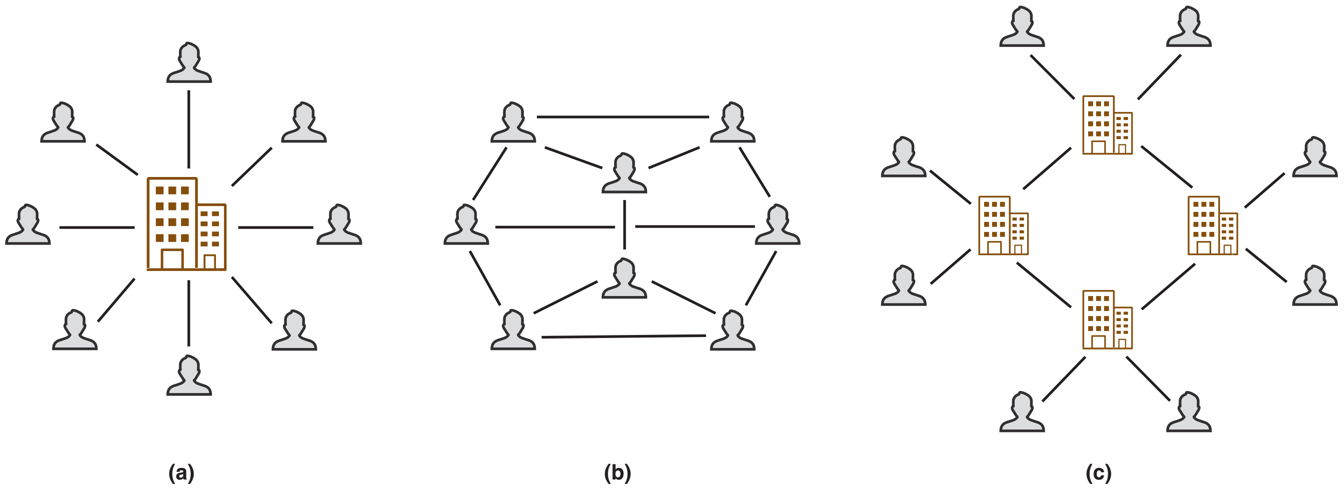

FIGURE 1 Different levels of (de)centralisation. (a) Fully centralised network where members communicate through a central entity. (b) Fully decentralised network where members communicate directly with each other. (c) Decentralised network following the organisational model, where individual members communicate through their own trusted organisations and thus alleviate the members of maintaining connections and custom software.

Permissioned blockchains are often structured in an organisational model, where nodes are hosted by organisations and not individuals. Users interact with the network through their organisation's node. Different levels of (de)centralisation are illustrated in Figure 1.

In Hyperledger Fabric, an open-source permissioned blockchain framework (Hyperledger Foundation, 2020), there are two types of nodes: peers, later referred to as 'blockchain nodes'; and orderers. Orderers are responsible for producing new blocks. New blocks are distributed to the peers, which individually assemble the blockchain and flag data that do not pass verification, thus keeping a complete history of valid and invalid submissions (Androulaki et al., 2018).

Peers also maintain a 'key-value' store separate to the blockchain, which holds a list of the latest version of every piece of data (uniquely identified by a 'key') on the blockchain, that is the most recent state of the ledger (Androulaki et al., 2018). When data are added to the blockchain, they are also added to this store. If modified data, that is a new version of the data, are submitted in a later block, the value in the store is overwritten with the latest data. This allows the latest version of the data to be efficiently searched for and retrieved from the store without traversing the entire blockchain, while historic values can still be retrieved directly from the blockchain.

3 | ECKOCHAIN

3.1 | Overview

We have designed and implemented the ECKOchain, a transparent blockchain-based 'proof of concept' database designed to be FAIR (Wilkinson et al., 2016). The FAIR guiding principles for open science promote findability, accessibility, interoperability and reusability, which have been shared objectives for our interdisciplinary team of computer scientists and ecologists (Carey et al., 2019).

ECKOchain is a permissioned blockchain created with the Hyperledger Fabric[6] framework. The network follows the organisational model (Figure 1c), where the blockchain is maintained by organisations instead of individual users. To participate in the network, an organisation must run a blockchain node and be accepted by a majority of the already participating organisations. Employees of an organisation use their organisation's node to interact with the blockchain, which lowers the barrier of entry to the system, as individual users are not required to run specialised node software.

While ECKOchain provides the distributed database infrastructure and governance mechanisms, ECKOweb[7] is the gateway to the system. ECKOweb is an open website hosted by the University of Bergen, where users can search for, download and share data, without prior knowledge of blockchain technology. The list of datasets is public, however, only users from participating organisations can download or submit new data. To further promote the distributed nature of the system, participating organisations can develop and host their own interfaces to the system.

The system does not currently require datasets to be standardised. While standardisation makes re-use and tooling simpler (Vanderbilt et al., 2015), existing data, especially from smaller projects, might not make it into these databases (Heidorn, 2008), as contributors require resources to format the data (Soeharjono & Roche, 2021). Although we encourage contributors to standardise

[6]Hyperledger Fabric version 2.4 (https://hyperledger.org/projects/fabric).
[7]ECKOweb (https://ecko.uib.no).

the primary data before submission, the only strict requirement for new submissions is to include sufficient standardised contextual information, that is metadata.

The standardisation of metadata is crucial to make data findable and interpretable by others (Urbano & Cagnacci, 2021; Whitlock, 2011), and plays a pivotal role in facilitating data (re-)use, integration and knowledge generation. To enhance findability and facilitate data reusability and interoperability, the metadata format of ECKOchain follows the Darwin Core (DwC) standard (Wieczorek et al., 2012), an internationally recognised ontology for biodiversity metadata. DwC is used in databases with similar scope to ECKO, such as GBIF, and offers a flexible framework tailored to compiling biodiversity data from varied and variable sources. However, the underlying protocols of ECKOchain are flexible and can be expanded in future iterations, where support for additional standards, for example the Ecological Metadata Language (Jones et al., 2019), can be considered. This is important, as it permits the network to support different domain-specific, yet interoperable, metadata standards for different types of data (Poisot et al., 2019).

Metadata stored on ECKOchain are searchable, either through ECKOweb or by directly querying the underlying blockchain as a participating organisation. The metadata also include links to related survey and resurvey data, which enable users to find interrelated data on the network. For some data, certain metadata might be considered particularly sensitive, for example location data of endangered species (Lennox et al., 2020), and thus may require varying degrees of access control. Therefore, ECKOchain is designed to provide data isolation between organisations, achieved by storing primary data files in private 'off-chain' data collections, a concept in Hyperledger Fabric, where data remain on servers belonging to the respective data owner's organisation. Files are transmitted to other users and organisations on-demand, according to selected licences and access policies.

To ensure integrity of files stored in private collections, an identifier ('digital fingerprint') uniquely generated from the content of the file is stored on the blockchain. Recipients of a file can re-generate the identifier and match it against the identifier on the blockchain, proving that the file has not been altered. Similar to a digital object identifier (DOI) used to uniquely identify intellectual property referenced on the internet (Chandrakar, 2006), the identifier provides a persistent reference to the data in the blockchain network. Persistent identifiers associated with each data entry ensures that data can be cited and that contributors receive recognition (Costello, 2009).

The smart contract enforced policies specify open-access licences or custom terms for granting access to datasets. Requests for open-access licensed data are approved immediately, whereas if custom terms are specified, the requester must detail intended use as part of the request, and the data owner must accept the request before the data can be downloaded. In both cases, request details are recorded on the blockchain. An editor or publisher can later query the system to, for example verify that authors are allowed to use referenced data.

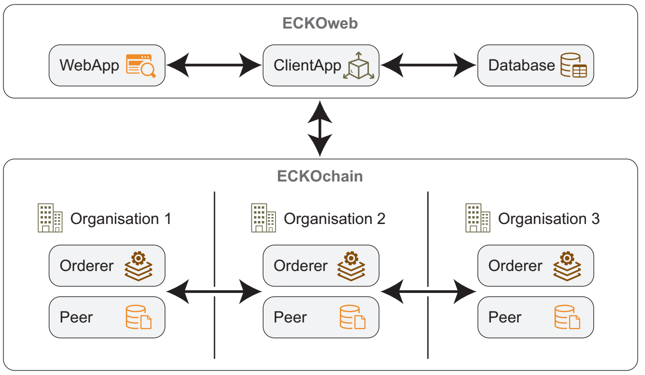

FIGURE 2 The software components of ECKO. The peers hold a copy of the blockchain, together with a ('key-value') store containing the most recent state of the ledger, and the organisation's private data collection. Optionally, organisations can participate in block production by providing orderer nodes. ECKOweb serves as a gateway for users to interact with the ECKOchain.

3.2 | Implementation

ECKO comprises the ECKOchain (the blockchain infrastructure) and ECKOweb (the website). ECKOchain is maintained by Hyperledger Fabric peer nodes, each run by a participating organisation. Optionally, organisations can also provide orderer nodes to participate in block production. The components of ECKO are shown in Figure 2.

Organisations can develop their own interfaces to the blockchain or opt to use ECKOweb. ECKOweb reads and writes data to the blockchain on behalf of the users. It connects to the ECKOchain through a client application, i.e. an application that embeds the Hyperledger Fabric Software Development Kit (SDK).[8] When we reference ECKOweb, we refer to both the website and the client, which are both written in the JavaScript[9] programming language (React.js[10] and Node.js[11] respectively).

Users of ECKOweb sign in using ORCID (Open Researcher & Contributor ID) credentials.[12] Consequently, ECKOweb does not handle users' passwords and many users are likely to have credentials already. First-time users must select their affiliation. If the user is not currently a member of a participating organisation, access can be requested through another organisation as an external member. The organisation's administrator decides whether the request is granted or not. External members can submit data to the organisation's private data collection, allowing users not affiliated with member organisations to also use ECKO.

To increase the website's responsiveness, a traditional database (PostgreSQL[13]) dedicated to ECKOweb is used for short term intermediate storage of blockchain data. Data on the blockchain

[8]Hyperledger Fabric SDK version 2.2 (https://hyperledger.github.io/fabric-sdk-node).
[9]JavaScript (https://javascript.com).
[10]React.js version 17 (https://reactjs.org).
[11]Node.js version 16 (https://nodejs.org).
[12]Orcid Public API (https://info.orcid.org/documentation/features/public-api).
[13]PostgreSQL version 13 (https://postgresql.org).

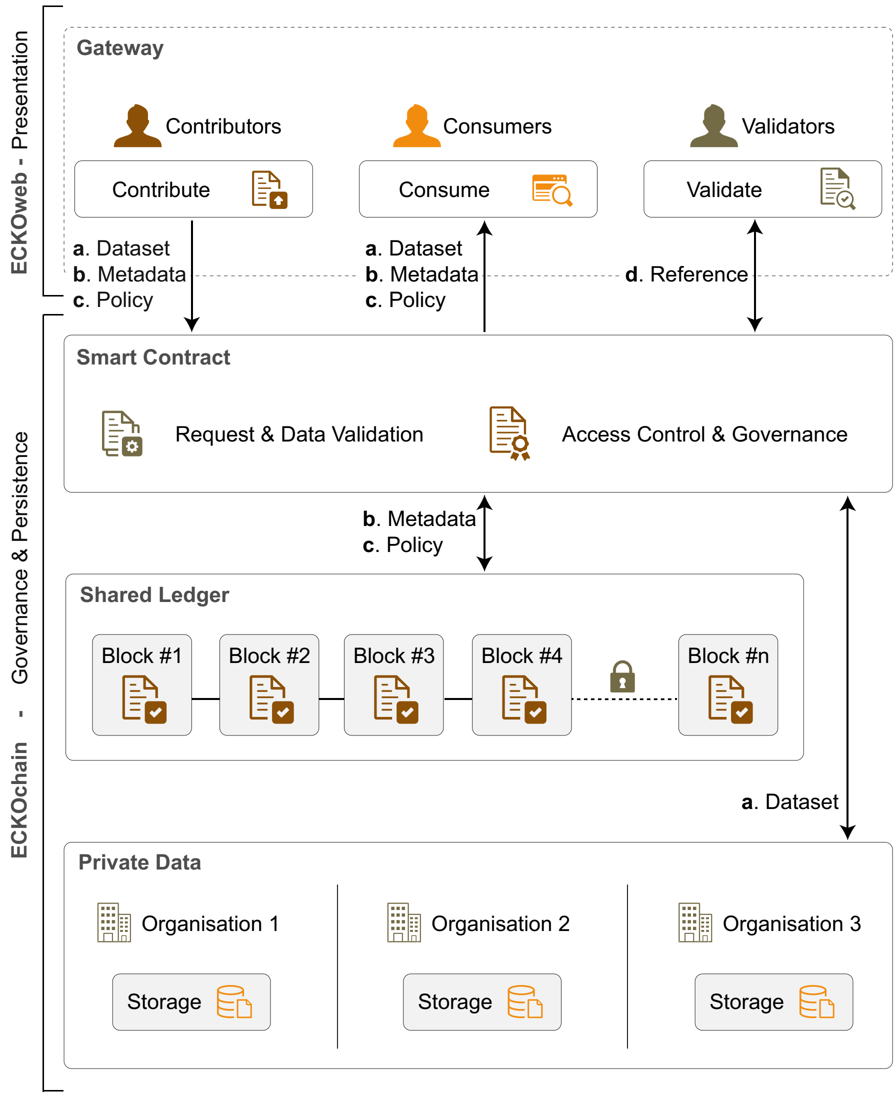

FIGURE 3 The data flows in ECKO. Users contribute and consume data through ECKOweb. Submissions and requests are validated by the smart contract. Metadata and policies are stored in blocks on the ledger, while primary data is stored in private data collections owned by the data owner's organisation. Third parties, for example a publisher, can verify that a user is permitted to use the referenced data.

node (in the node's 'key-value' store specifically) are copied to the database on regular intervals, and requests through ECKOweb are directed to this database for semi-recent data rather than to the node directly, which avoids unnecessary load on the distributed infrastructure.

The data management protocols of the ECKOchain reside in a smart contract (also known as chaincode) written in the Java[14] programming language. The smart contract ensures that new submissions have the necessary metadata, policies, and primary data files. The data file is stored in a private data collection belonging to the data owner's organisation, while the metadata and policy are stored on the blockchain. The data flows in ECKO are illustrated in Figure 3.

In addition to providing confidentiality, storing files 'off-chain' limits the total size of the ledger, thus increasing the network's ability to scale and it allows for files to be retracted from the system if necessary. However, even if a primary file is removed, the file's digital fingerprint, metadata and associated policies remain on the blockchain.

[14]Java version 11 (https://java.com).

4 | DISCUSSION

4.1 | A FAIR database

The FAIR guiding principles of Wilkinson et al. (2016) have guided the development of ECKO. Here we look at how ECKO conforms to these principles.

ECKO conforms to findability by defining mandatory, rich and standardised metadata, and by assigning unique identifiers to datasets. Metadata and identifiers remain on the ledger and are searchable through ECKOweb.

ECKO conforms to accessibility by using open-source software and protocols, by making metadata retrievable using data identifiers, and by making metadata remain on the ledger even if a dataset is retracted. ECKOchain is a permissioned blockchain, thus requests to the ledger require authentication and authorisation.

ECKO conforms to interoperability by using the DwC standard as basis for the metadata, as well as for the vocabulary used on ECKOweb. Relationships to other datasets are listed in the metadata. While the metadata in the first iteration of ECKO adhere to the DwC standard, support for additional metadata standards can be added in the future.

Lastly, ECKO conforms to reusability by displaying rich metadata, data usage licences and terms, and detailed citation examples, along with every dataset on ECKOweb.

4.2 | Challenges of decentralisation

Distributed and decentralised systems are generally harder to implement and maintain than centralised systems. While a centralised system has a single source of truth, distributed systems comprise autonomous computing elements that appear to users as a single coherent system (Van Steen & Tanenbaum, 2016). In a blockchain, these autonomous elements employ complex consensus mechanisms to stay synchronised and to recover from network failures. Due to the complexity of both data management technologies and the research data that are being managed, an interdisciplinary team of computer scientists and research subject matter experts has proven necessary to develop ECKO.

The benefits of a decentralised model can outweigh the downsides of increased complexity. For long term preservation of data, it is beneficial that data exists in multiple locations, on hardware controlled by different entities. If entities in a network vanish over time, the network can still operate with just a single remaining provider. As a field of science that depends on long-lived and re-usable data, ecology can benefit from enhanced openness and transparency (Powers & Hampton, 2019), and from alternatives to exclusive siloed archives (Hampton et al., 2015). To facilitate long-lived research data, we eliminate the need to trust a single data provider.

Hosting nodes for a blockchain network requires technical expertise that end-users of the system should not be expected to have. We address this by establishing a blockchain that is governed by organisations, placing the burden of maintaining the ledger not with the end-user but rather with the participating organisations. While organisations are more likely to have the technical and economical resources to run such specialised software, we recognise that this will not be universal across all those willing to participate in the network, presenting a complex challenge to implementing a universally fair decentralised data network (Lewis et al., 2023).

A tentative alternative to establishing a permissioned blockchain is a solution developed on a public blockchain, for example on the public Ethereum blockchain (Buterin, 2014). Here, relying on an existing public network, a larger community of node providers will already exist. In public blockchains, however, data integrity and participation in the network is secured and incentivised via monetary rewards and, consequently, fees must be paid for utilising the network (Xiao et al., 2020). Operating the ECKOchain as a permissioned network circumvents such dependence on fees and cryptocurrencies, lowering the long-term economic barrier to network participation. Nevertheless, understanding the trade-offs between permissioned and public networks in serving the open scientific data community will require continued research as the technology evolves.

5 | CONCLUSION

ECKOchain provides auditability and transparency to long-term ecological data management. The blockchain is distributed among all network participants so that data cannot be added or altered without consensus among the participants. The distributed nature of the blockchain ensures long-term data preservation, as availability is not dependent on a single entity and the transparent smart contract driven policies prevent a single entity from controlling the data. This robust, resilient and FAIR system facilitates long-lived open research data.

ECKOweb is a gateway to the ECKOchain allowing users to submit and manage data, thus making the system accessible to users not expected to be familiar with blockchain technology. As data are distributed and therefore not controlled by a single entity, gateways operated by different entities can co-exist, further promoting the openness of the system. Data validation and usage licences are enforced in the network by the self-executing smart contract. Supporting metadata based on the DwC standard are mandatory to facilitate data reusability and interoperability, and to increase the data's findability.

The approach of using blockchain technology to manage research data is also relevant to other scientific fields that benefit from long-lived research data and open data practices. The ECKOchain infrastructure is reusable, meaning that new initiatives can live side-by-side on the existing ledger, while new or adapted smart contracts can be installed to support other types of data and use cases.

AUTHOR CONTRIBUTIONS

Robert John Lewis conceived the idea of the project. Kjell-Erik Marstein designed the technical solution. All authors contributed to the overall design and outcome of the project. Kjell-Erik Marstein wrote the first draft of this manuscript. All authors contributed to the writing.

ACKNOWLEDGEMENTS

The project received partial support from the Research Council of Norway, project no. 275681. We thank the ECKO project group at the University of Bergen and the Norwegian Institute of Bioeconomy Research for contributing to the project. We also thank everyone that submitted data and helped with testing.

CONFLICT OF INTEREST STATEMENT

Kjell-Erik Marstein is currently employed by Quant Network, a commercial company specialised in blockchain interoperability.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14280.

DATA AVAILABILITY STATEMENT

ECKOweb is hosted on https://ecko.uib.no. The source code is available in the following repositories: https://github.com/kehm/eckochain;

https://github.com/kehm/eckochain-client; and https://github.com/kehm/eckoweb. The repositories were archived on Zenodo at the time of this publication (Marstein, 2023a, 2023b, 2023c).

ORCID

Kjell-Erik Marstein https://orcid.org/0000-0001-7744-0168
John-Arvid Grytnes https://orcid.org/0000-0002-6365-9676
Robert John Lewis https://orcid.org/0000-0003-2067-6844

REFERENCES

Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., Enyeart, D., Ferris, C., Laventman, G., Manevich, Y., Muralidharan, S., Murthy, C., Nguyen, B., Sethi, M., Singh, G., Smith, K., Sorniotti, A., Stathakopoulou, C., Vukolić, M., … Yellick, J. (2018). Hyperledger Fabric: A distributed operating system for permissioned blockchains. *Proceedings of the Thirteenth EuroSys Conference (EuroSys 2018)*. Association for Computing Machinery, Article 30, 1–15. https://doi.org/10.1145/3190508.3190538

Buterin, V. (2014). *Ethereum: A next-generation smart contract and decentralized application platform*. https://ethereum.org/en/whitepaper

Cachin, C., & Vukoljic, M. (2017). Blockchain consensus protocols in the wild. *31st International Symposium on Distributed Computing (DISC 2017)*, pp. 1:1–1:16, Schloss Dagstuhl—Leibniz Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.DISC.2017.1

Carey, C. C., Ward, N. K., Farrell, K. J., Lofton, M. E., Krinos, A. I., McClure, R. P., Subratie, K. C., Figueiredo, R. J., Doubek, J. P., Hanson, P. C., Papadopoulos, P., & Arzberger, P. (2019). Enhancing collaboration between ecologists and computer scientists: Lessons learned and recommendations forward. *Ecosphere*, *10*(5), e02753. https://doi.org/10.1002/ecs2.2753

Chandrakar, R. (2006). Digital object identifier system: An overview. *Electronic Library*, *24*, 445–452. https://doi.org/10.1108/02640470610689151

Costello, M. J. (2009). Motivating online publication of data. *BioScience*, *59*(5), 418–427. https://doi.org/10.1525/bio.2009.59.5.9

De Vries, A. (2018). Bitcoin's growing energy problem. *Joule*, *2*(5), 801–805. https://doi.org/10.1016/j.joule.2018.04.016

De Vries, A. (2022). Cryptocurrencies on the road to sustainability: Ethereum paving the way for Bitcoin. *Patterns*, *4*(1), 100633. https://doi.org/10.1016/j.patter.2022.100633

Depauw, L., & Maes, S. (2015). forestREplot: A global database of temperate forest herb layer resurvey plots. *British Ecological Society Bulletin*, *46*(3), 31–34. http://hdl.handle.net/1854/LU-6950981

Dornelas, M., Antão, L. H., Moyes, F., Bates, A. E., Magurran, A. E., Adam, D., Akhmetzhanova, A. A., Appeltans, W., Arcos, J. M., Arnold, H., Ayyappan, N., Badihi, G., Baird, A. H., Barbosa, M., Barreto, T. E., Bässler, C., Bellgrove, A., Belmaker, J., Benedetti-Cecchi, L., … Hickler, T. (2018). BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography*, *27*, 760–786. https://doi.org/10.1111/geb.12729

Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of biodiversity databases: Biodiversity information on every desktop. *Science*, *289*, 2312–2314. https://doi.org/10.1126/science.289.5488.2312

Farley, S. S., Dawson, A., Goring, S. J., & Williams, J. W. (2018). Situating ecology as a big-data science: Current advances. *Challenges, and Solutions*, *BioScience*, *68*(8), 563–576. https://doi.org/10.1093/biosci/biy068

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature*, *529*, 117–119. https://doi.org/10.1038/nj7584-117a

Hampton, S. E., Anderson, S. S., Bagby, S. C., Gries, C., Han, X., Hart, E. M., Jones, M. B., Lenhardt, W. C., MacDonald, A., Michener, W. K., Mudge, J., Pourmokhtarian, A., Schildhauer, M. P., Woo, K. H., & Zimmerman, N. (2015). The Tao of open science for ecology. *Ecosphere*, *6*(7), 1–13. https://doi.org/10.1890/ES14-00402.1

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S., & Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, *11*, 156–162. https://doi.org/10.1890/120103

Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends*, *57*(2), 280–299. https://doi.org/10.1353/lib.0.0036

Huang, X., Hawkins, B. A., Lei, F., Miller, G. L., Favret, C., Zhang, R., & Qiao, G. (2012). Willing or unwilling to share primary biodiversity data: Results and implications of an international survey. *Conservation Letters*, *5*, 399–406. https://doi.org/10.1111/j.1755-263X.2012.00259.x

Hyperledger Foundation. (2020). *Welcome hyperledger fabric 2.0: Enterprise DLT for production*. https://www.hyperledger.org/blog/2020/01/30/welcome-hyperledger-fabric-2-0-enterprise-dlt-for-production

ISO. (2022). *Blockchain and distributed ledger technologies—Use cases (ISO/TR 3242:2022)*. Technical report. International Organization for Standardization. https://www.iso.org/standard/79543.html

ITU-T. (2019). *Distributed ledger technology use cases (FG DLT D2.1)*. Technical report. ITU Telecommunication Standardization Sector. https://itu.int/en/ITU-T/focusgroups/dlt

Jones, M. B., O'Brien, M., Mecum, B., Boettiger, C., Schildhauer, M., Maier, M., Whiteaker, T., Earl, S., & Chong, S. (2019). Ecological Metadata Language version 2.2.0. *KNB Data Repository*, https://doi.org/10.5063/F11834T2

Lennox, R. J., Harcourt, R., Bennett, J. R., Davies, A., Ford, A. T., Frey, R. M., Hayward, M. W., Hussey, N. E., Iverson, S. J., Kays, R., Kessel, S. T., Mcmahon, C., Muelbert, M., Murray, T. S., Nguyen, V. M., Pye, J. D., Roche, D. G., Whoriskey, F. G., Young, N., & Cooke, S. J. (2020). A novel framework to protect animal data in a world of ecosurveillance. *BioScience*, *70*(6), 468–476. https://doi.org/10.1093/biosci/biaa035

Lewis, R. J., Marstein, K. E., & Grytnes, J. A. (2023). Incentivising open ecological data using blockchain technology. *Scientific Data*, *10*, 591. https://doi.org/10.1038/s41597-023-02496-2

Marstein, K. E. (2023a). kehm/eckochain: v0.0.1. *Zenodo*, https://doi.org/10.5281/zenodo.10359772

Marstein, K. E. (2023b). kehm/eckochain-client: v0.0.1. *Zenodo*, https://doi.org/10.5281/zenodo.10359778

Marstein, K. E. (2023c). kehm/eckoweb: v0.0.1. *Zenodo*, https://doi.org/10.5281/zenodo.10359780

Mills, J. A., Teplitsky, C., Arroyo, B., Charmantier, A., Becker, P. H., Birkhead, T. R., Bize, P., Blumstein, D. T., Bonenfant, C., Boutin, S., Bushuev, A., Cam, E., Cockburn, A., Côté, S. D., Coulson, J. C., Daunt, F., Dingemanse, N. J., Doligez, B., Drummond, H., … Zedrosser, A. (2015). Archiving primary data: Solutions for long-term studies. *Trends in Ecology & Evolution*, *30*(10), 581–589. https://doi.org/10.1016/j.tree.2015.07.006

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. https://bitcoin.org/bitcoin.pdf

Ortega, D. R., Oikonomou, C. M., Ding, H. J., Rees-Lee, P., Alexandria, & Jensen, G. J. (2019). ETDB-Caltech: A blockchain-based distributed public database for electron tomography. *PLoS ONE*, *14*(4), e0215531. https://doi.org/10.1371/journal.pone.0215531

Parmentola, A., Petrillo, A., Tutore, I., & De Felice, F. (2022). Is blockchain able to enhance environmental sustainability? A systematic review and research agenda from the perspective of sustainable development goals (SDGs). *Business Strategy and the Environment*, *31*(1), 194–217. https://doi.org/10.1002/bse.2882

Poisot, T., Bruneau, A., Gonzalez, A., Gravel, D., & Peres-Neto, P. (2019). Ecological data should not Be so hard to find and reuse. *Trends in Ecology & Evolution*, *34*(6), 494–496. https://doi.org/10.1016/j.tree.2019.04.005

Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. *Ecological Applications*, *29*(1), e01822. https://doi.org/10.1002/eap.1822

Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., Kokko, H., Jennions, M. D., & Kruuk, L. E. B. (2014). Troubleshooting public data archiving: Suggestions to increase participation. *PLoS Biology*, *12*(1), e1001779. https://doi.org/10.1371/journal.pbio.1001779

Soeharjono, S., & Roche, D. G. (2021). Reported individual costs and benefits of sharing open data among Canadian academic faculty in ecology and evolution. *BioScience*, *71*(7), 750–756. https://doi.org/10.1093/biosci/biab024

Szabo, N. (1996). Smart contracts: Building blocks for digital markets. *Extropy: Journal of Transhumanist Thought*, *16*, 18.

Szabo, N. (1997). Formalizing and securing relationships on public networks. *First Monday*, *2*(9). https://doi.org/10.5210/FM.V2I9.548

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE*, *15*(3), e0229003. https://doi.org/10.1371/journal.pone.0229003

Urbano, F., & Cagnacci, F. (2021). Data management and sharing for collaborative science: Lessons learnt from the euromammals initiative. *Frontiers in Ecology and Evolution*, *9*, 727023. https://doi.org/10.3389/fevo.2021.727023

Van Steen, M., & Tanenbaum, A. S. (2016). A brief introduction to distributed systems. *Computing*, *98*, 967–1009. https://doi.org/10.1007/s00607-016-0508-7

Vanderbilt, K. L., Lin, C. C., Lu, S. S., Kassim, A. R., He, H., Guo, X., San Gil, I., Blankman, D., & Porter, J. H. (2015). Fostering ecological data sharing: Collaborations in the international long term ecological research network. *Ecosphere*, *6*(10), 204–218. https://doi.org/10.1890/ES14-00281.1

Whitlock, M. C. (2011). Data archiving in ecology and evolution: Best practices. *Trends in Ecology & Evolution*, *26*, 61–65. https://doi.org/10.1016/j.tree.2010.11.006

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin core: An evolving community-developed biodiversity data standard. *PLoS ONE*, *7*(1), e29715. https://doi.org/10.1371/journal.pone.0029715

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Xiao, Y., Zhang, N., Lou, W., & Hou, Y. T. (2020). A survey of distributed consensus protocols for blockchain networks. *IEEE Communication Surveys and Tutorials*, *22*(2), 1432–1465. https://doi.org/10.1109/COMST.2020.2969706

Zīle, K., & Strazdiņa, R. (2018). Blockchain use cases and their feasibility. *Applied Computer Systems*, *23*, 12–20. https://doi.org/10.2478/acss-2018-0002

How to cite this article: Marstein, K.-E., Grytnes, J.-A., & Lewis, R. J. (2024). ECKOchain: A FAIR blockchain-based database for long-term ecological data. *Methods in Ecology and Evolution*, *00*, 1–8. https://doi.org/10.1111/2041-210X.14280