

A Collaborative Data Network for the Asia Oceania Region Enabled by Emerging Technologies to Foster Innovation in a Secure and Open Environment



REVIEW

]u[ubiquity press

ALISON SPECHT

KIM BRYCESON

SHOUFENG CAO

MARGARET O'BRIEN

S. M. GURU

PEDRO PIZZIGATTI CORREA

MICHELLE WAYCOTT

*Author affiliations can be found in the back matter of this article

ABSTRACT

A discoverable inventory of items of value and importance to a community, country, or region has many benefits. These benefits can be for reporting (state of the environment, achievement of Sustainable Development Goals, conservation objectives), for research (exploration of international phylogenetics, development of new pharmaceuticals), or for economic reasons. Confidence in the value of data is best achieved if the information used to underpin statements and conclusions is open to scrutiny.

This paper presents a framework for the creation of a collaborative data network for the Asia-Oceania region that will allow members to hold, manage, and share data about their countries and regions. We propose a 'Collaborative Regional Data Access Network' (CREDAN) that will blend the attributes of centralised and decentralised networked databases for the management and discovery of data. Following the CARE principles, data owners will determine the sovereignty and sharing permissions for their data. To ensure ownership is always clear, and subsequent use of data is tracked, we harness blockchain technology alongside smart contracts and decentralised identifiers to allow secure discovery and sharing. The use of blockchain will ensure the information about each dataset (metadata, provenance, and access rights) is immutably connected to each dataset. We recommend a strong governance structure to ensure best practices are aligned with international criteria and those set by the component communities. Management of the eventual large volume of data should be anticipated, and partnerships may need to be considered. We propose a seven-step pathway to creation of such a network.

CORRESPONDING AUTHOR:

Alison Specht

TERN, the University of
Queensland, QLD, Australia
a.specht@uq.edu.au

KEYWORDS:

collaborative; data network;
blockchain; centralised;
distributed; CARE; FAIR

TO CITE THIS ARTICLE:

Specht, A., Bryceson, K., Cao, S., O'Brien, M., Guru, S.M., Correa, P.P. and Waycott, M. (2025) A Collaborative Data Network for the Asia Oceania Region Enabled by Emerging Technologies to Foster Innovation in a Secure and Open Environment. *Data Science Journal*, 24: 1, pp. 1–15. DOI: <https://doi.org/10.5334/dsj-2025-001>

1. CONTEXT AND NEED

The Asia-Oceania region is largely characterised by islands; island archipelagos or solitary islands, large (such as Australia) or small. French Polynesia, for example, has 118 islands concentrated in five archipelagos (Van Wynsberge et al., 2015), while Indonesia has 13,558 islands of which 8844 have been named and 922 support a permanent population (Andréfouët et al., 2022). Some, however, are more continental, fringing the main Asian landmass. Vietnam and Malaysia are examples of this.

From a data management perspective, some countries, such as Australia and New Zealand, have sophisticated organisation of much of their knowledge and data, and their e-infrastructures are well advanced. In Indonesia, science and education institutions have been consolidated into one 'super-entity', the Indonesian National Research and Innovation Agency (BRIN). This provides Indonesia with the potential to lead participation in Open Science and scientific innovation in Southeast Asia. Yet, there is an acknowledged deficit in digital literacy (Stewart, 2023). Although Indonesia's new science law mandates that researchers deposit their data on the National Scientific Repository, it is yet to be a common practice for two reasons; firstly, the formulation of technical regulation is still ongoing, and secondarily, data sharing has yet to become a norm. Other countries are geographically and economically smaller, with greatly varying levels of technological capability. In previous centuries, many of these smaller countries have been heavily economically influenced by international interests, with international aid a common feature. Even today the attraction of promised investment and financial reward for local communities is strong and can result in poor long-term decisions (Nicholas and Lyons, 2021) and they can be vulnerable to exploitation (Jutel, 2021).

Many of the smaller countries in the Asia-Oceania region face significant challenges to improve livelihoods and overcome poverty, particularly due to increases in food and fuel prices, the effects of various global economic crises, natural disasters, difficulties maintaining infrastructure, and the negative effects of climate change (Underhill et al., 2011). Consistent with the Sustainable Development Goals (United Nations, n.d.), the Pacific Community (n.d. (a)), the scientific and technical organisation supporting development in the Pacific region established in 1947 (under the Canberra Agreement), has pledged to progress development and has actively supported aggregation of data across the region (Pacific Community, 2022).

There is a recognised lack of basic, longitudinal evidence about environmental 'assets' on which countries can base decisions and report on environmental indicators (Straza T, 2021). Derived databases can assist, but as noted by Visconti et al. (2013), Moudry and Devillers (2020), and Brewer et al. (2022), internationally aggregated databases can be coarse and unreliable. The sheer number of required reports for the many bilateral agreements to which each country is signatory is considerable. These include reports on environmental indicators, such as those for climate change and biodiversity, the status of protected areas, waste management, and pollution. Readily available data to help respond to these reports will be most advantageous. The Pacific Regional Environment Program (SPREP) (n.d. (a)) has advocated for data and information sharing between countries and have proposed that centralised data services could assist with data and indicators for environmental information (Straza T, 2021).

In 2023, a Pacific Academy of Sciences and Humanities (International Science Council, 2024) was declared at a meeting in Samoa. This is evidence of a general recognition in the region of the value of organised conversations and the need for a strong knowledgeable voice for the countries of the Pacific. Academic and researcher representatives from Samoa, the Cook Islands, Fiji, Papua New Guinea, the Solomon Islands, Vanuatu, Kiribati, New Caledonia, and New Zealand have been involved in its planning. The Australian Academy of Science is a partner.

1.1. ENGAGING COMMUNITIES

Bottom-up design is arguably the optimal way to ensure a sense of ownership of a project and can increase community capacity. The creation of a robust and authoritative set of data by the community about entities in their world needs engagement, and it needs guidelines to ensure standardisation if the data are to be shared. Community engagement in knowledge creation can be accomplished in several ways. The three governance structures considered by Conrad and Hilchey (2011) can provide a basis for community-based data gathering. These are

(1) government-led, community run; (2) collaborative, often in a non-politically demarcated area; and (3) community led, run, and funded. Databases such as iNaturalist (Martinez-Sagarra et al., 2022; Meeus et al., 2023) and eBirds (Sullivan et al., 2009; 2014) are examples where an infrastructure has been developed dependent on the contributions of individual community members with rewards and credits for participation. Tools such as the Great Lakes Fish Finder App (Happel et al., 2020) can help make participation easy and rewarding, and participatory design has provided benefits for the community (Durish et al., 2021). Kitamura et al. (2023) considered that for a successful outcome for a community-led interactive database designed for asset-based community development the intended audience (data providers, the end users and the facilitators), the type of information to be included, and plans for sustainability needed to be considered.

Many of the Asia-Oceania countries are occupied by original communities. As such, they may be termed to be 'indigenous' and there have been a few initiatives in recent years to facilitate the participation of such communities in collating and storing their data and information for their own use and benefit and to share. In addition to information gathered using scientific methods, traditional knowledge has long been recognised, for example, as of great advantage to properly understand environmental processes (Huntington, 2000; Ens et al., 2015). Having good protocols and support for sharing information will greatly assist communities to control their information and expand initiatives already in place.

The CARE principles (Collective Benefit, Authority to Control, Responsibility, and Ethics) provide guidance around Indigenous Data (Carroll et al., 2020 and 2021; Hudson et al., 2020). The principles are couched around the need for assurance that there will be collective benefit from data collection and sharing, an authority to control the information gained and shared, a responsibility for those data to ensure positive relationships, expanding capability and capacity, and the assurance of ethical behaviour and standards. Aligning these principles to a repository remains a work in progress (O'Brien et al., 2024).

The labelling proposed by the Local Contexts (n.d.) initiative provides communities with tools that can assert cultural authority in heritage collections and properly identify data ownership. These labels can facilitate how data is collected, managed, displayed, accessed, and used by others and provide assurance that data are properly handled and shared sensitively. Incorporating such labelling at data source could alleviate concerns for data providers.

The Te Kāhui Raraunga Charitable Trust (TKR) (n.d.), an independent body in New Zealand established in 2019, leads the action required to realise the advocacy of the Data Iwi Leaders Group (Data ILG). TKR's aim is to enhance the social, cultural, environmental, and economic well-being of Māori; to enable iwi, hapū and whānau Māori to access, collect and use Māori data. The planned data storage network has layers of safety, security, functionality, and protection structurally informed by kawa and tikanga, which will ensure that hapū and iwi can make their own calls about the access and sharing of their own information in ways that honour their kawa and tikanga at place. They will do this in large part by employing Local Context labels.

1.2. DATA ORGANISATION IN THE OCEANIA REGION

We shall review an example of data organisations in the region of concern. As part of an emerging 'Pacific Data Ecosystem', the Secretariat of the SPREP together with the Pacific Island Countries, the Pacific Community (n.d.(a)), and the UN Environment Programme initiated the creation of a Pacific Environment Portal (SPREP, n.d. (b)) to promote greater coordination in data management, dissemination, and uptake initiatives across national data repositories. A Pacific Data Hub (Pacific Community, n.d. (b)) was created by the Pacific Community in 2018 with a similar aim. It is important SPREP shares data with SPC, and that is where the Pacific Data Hub open-source solution sits.

In order to support this process, the *Inform Project—Building National and Regional Capacity to Implement Multilateral Environmental Agreements by Strengthening Planning and the State of Environmental Assessment and Reporting in the Pacific*—was created and ran from 2018–2021 (SPREP, n.d. (c)). One of the main tasks of Inform was to collate all the data and legacy information stored in paper and various databases and put them into a form that is easily discoverable, standardised, and able to be used for multiple things. In addition, Inform provides

tools, guidance, and a checklist for dataset creation and uploading data to the central portal. It uses a GitHub-linked Drupal portal, the DKAN Open Data Portal (SPREP, n.d. (d)), to host the unique portals of fourteen participating countries. Ambitions to provide opportunities to store raw data in the SPREP databases are yet to be realised, but great advances have been made in making secondary data about the region openly available.

These initiatives illustrate that there is strong regional intent, and go a long way to develop a secure, supported approach to conserve and discover information about the Pacific Island Nations.

1.3. THE REASON FOR THIS PAPER

Consistent with the UNESCO recommendations on Open Science (2021), confidence in the results of studies is best achieved if the data used to produce the work is open to scrutiny. This may range from the statistics behind economic models, to real-time climate data (not interpretations from a third party), to the measurements of fish populations. If a temporal aspect is included, more than one instance may need to be available to establish trust in the findings. If a comparison is made across geographic space, data from each locale used for the comparison should be available to view. Interpretations are thus more easily defensible and also contestable. Once the data are available, they become valuable in themselves. Historically, paper copies of data have sometimes been carefully curated in museums, laboratories, or in government departments, but modern technology and internet connectivity allow data to be digitally accessible, greatly increasing the value of the data (e.g., Ahl et al., 2023). No longer do the data remain only with a study team, but they can form a legacy for others. No longer will each new researcher or study coordinator have to re-find data from previous studies, but the data will be available for re-use. Enabling this requires organisation and commitment.

Active participation in conserving data in a temporal and spatially relevant manner is desirable. This can be for international reporting reasons (state of the environment, achievement of Sustainable Development goals, conservation objectives), for research reasons (exploration of international phylogenetics, development of new pharmaceuticals), or for economic reasons (tracking the status of fish stocks, timber trade, and other food products). In a systematic review of community-engaged knowledge hubs, Brar et al. (2023) concluded that the benefits of knowledge hubs included better knowledge mobilisation and dissemination, and increased community benefits through the capacity building of community partners and their greater involvement in knowledge production and mobilisation activities. Difficulties in sustaining a hub included differences in priorities across the network, limited capacity, resistance to change, and lack of funding. The development of the knowledge hub sponsored by the SPREP, for example, is an excellent start. Making resources available for the community means ensuring that data (and related materials) are findable and accessible on the Web, and that they comply with adopted international standards making them interoperable and reusable by others (David et al., 2020).

In the next section, we consider what data architecture would be optimal to best ensure a collaborative data network is able to comply with (a) FAIR principles (F, Findable; A, Accessible; I, Interoperable; R, Reproducible) (Wilkinson et al., 2016), to establish criteria to ensure (b) CARE principles are followed (C, Collective benefit; A, Authority to control; R, Responsibility; E, Ethics) (Carroll et al. 2020), while (c) ensuring the information is held in repositories that are robust, namely following the TRUST principles (T, Transparency; R, Responsibility; U, User Focus; S, Sustainability; T, Technology) (Lin et al., 2020).

2. A DATA NETWORK FOR COMMUNITIES AND REGIONS

The structure we suggest for a **Collaborative Regional Data Access Network** (CREDAN) combines a suite of country-level centralised databases/repositories connected in a distributed network, with a collaborative coordinating authority to ensure sound governance, help establish common standards for data types collected, data set descriptions (metadata), and open descriptions of vocabularies used. Several emerging technologies would be employed to ensure the fidelity of provenance, licensing, and protocols are maintained throughout the network, and in the interface with the wider community. Each of these components is introduced in this section.

2.1. COUNTRY OR SUB-COUNTRY LEVEL

Each country or sub-country entity—hereafter referred to as (sub)country—would hold their own data in a central location where the data are curated, managed, and made available following a **centralised model** (Figure 1). These models are coordinated by a central authority (in this instance a data hub), and in this way data linkage, harmonisation, and interoperability can be facilitated (Rujano et al., 2024). In a centralised system, data can be uniformly managed, stored, and processed, allowing for streamlined operations, enhanced data security, and easier compliance with relevant regulations. The hub, which could be a traditional database, distributed file system, or cloud service, needs to be adopted and driven by the community, as it would fail without that the effort (Bailey et al., 2024).

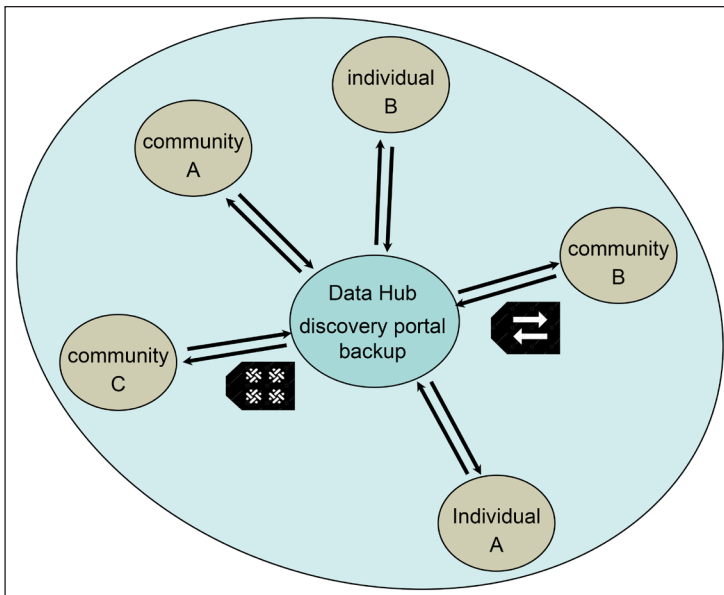


Figure 1 Depiction of a centralised data hub within a locally defined area (e.g., island, group of islands or country). Communities or individuals would assemble their data and metadata, applying local knowledge-type rules to each dataset. Examples of two traditional knowledge labels are shown for ‘attributions’, which need to be set at the community level, and the ‘multiple communities’ label, which indicates knowledge is shared across more than one community (see <https://localcontexts.org>).

This level of the architecture is critical to the success of our proposed solution, as these hubs provide the data for the whole network, as well as determining the permissions for access to those data. The data stored in these hubs would presumably focus on data about the geographical area in question collected by communities or individuals. These individuals or communities would have the responsibility to properly describe the items to be stored on the hub, such as provide metadata, use common vocabularies (or define new ones so others can understand), and attach sharing rights according to, for example, Traditional Knowledge criteria following the Local Contexts categorisations (Local Contexts, n.d.). These rights and permissions would be validated by the relevant communities (O’Brien et al., 2024). Data ownership would be according to these sharing agreements.

In all situations the provenance of the information being shared needs to be carefully documented, so attribution can be correctly applied. The system should provide for changing needs and ‘community informality’ (Bryceson and Ross, 2020; Bhusal, 2023), ensuring dialogue is in place so any changes in sharing agreements are adjustable. The design should allow periodic update by the providers of the data held and the tools used for and manner of discovery (maps, downloads, reports, etc). There will be additional opportunistic additions both past and present, if they are spatially consistent. The individual communities would not need to have computing capability but would need to understand any country standards to smooth aggregation and sharing.

Each hub would need to be equipped with data storage capacity and take on some of the responsibility of a TRUSTed repository (Lin et al., 2020). The hubs would need to ensure metadata are systematically applied to each digital object, and standard vocabularies are used where possible, or defined openly (see Specht et al., 2024). Equipping these hubs with the technology, the internet security, and the skill sets required in the staff is not a trivial matter, but has been explored thoroughly in setting up a climate data management system in the western Pacific islands states (Martin et al., 2015) and of course by the SPREP, to name two examples. The hub

staff would take responsibility for dialogue with each individual and community to ensure their data and information aligns with the FAIR and CARE criteria as much as possible.

One of the first tasks for the community groups and individuals will be the creation of a dictionary of the terms used for their data. Optimally for sharing purposes, terminology will be common across all data contributors, and follow international standards, but most likely some of the names of things and their meanings will be different. The hub will be responsible for facilitating the dictionary process, compiling the different dictionaries, and assisting data creators (owners) where necessary to develop their own definitions in a standard way (Specht et al., 2024). These should be harmonised and published, preferably using a web interface associated with the digital objects held. At no point should any original data and terminology be lost.

The international significance of particular objects described in the datasets may be recorded at this level (e.g., listed rare species) and any sharing propositions or material already shared should be examined in the light of the Nagoya Protocol (Secretariat of the Convention on Biological Diversity, 2011). Depending on capability, the hub may have an ability to enable the contributing communities to refer to the data and information as and when they want according to their sharing protocols.

2.2. REGIONAL LEVEL

Country and sub-country data hubs will be linked through a distributed network (Figure 2). This facilitates data querying and accessibility to information from any point. The data itself remains at the (sub)country hub, where the manner of data sharing (permissions and licensing) is established and controlled. This structure is ideal when data sovereignty and privacy are crucial, yet collaborative data access and sharing remain necessary. The separate hubs connect to each other using purpose-built links. The envisaged network could be supported by an overarching cloud computer storage facility allowing backup and coordination across the network. Traditionally, distributed networks have depended on each hub being of the same configuration and standard, while in reality, the hubs in our scenario will no doubt vary in size and type. Various technologies, however, can be employed to overcome this (e.g., Tebbi et al., 2019). Sharing across the network, and for reporting against international targets, would require some level of data interoperability. Governance of a distributed network can be a challenge, as data protection and security across a network of variable hubs requires focussed and maintained vigilance, making strong collaboration among hubs essential.

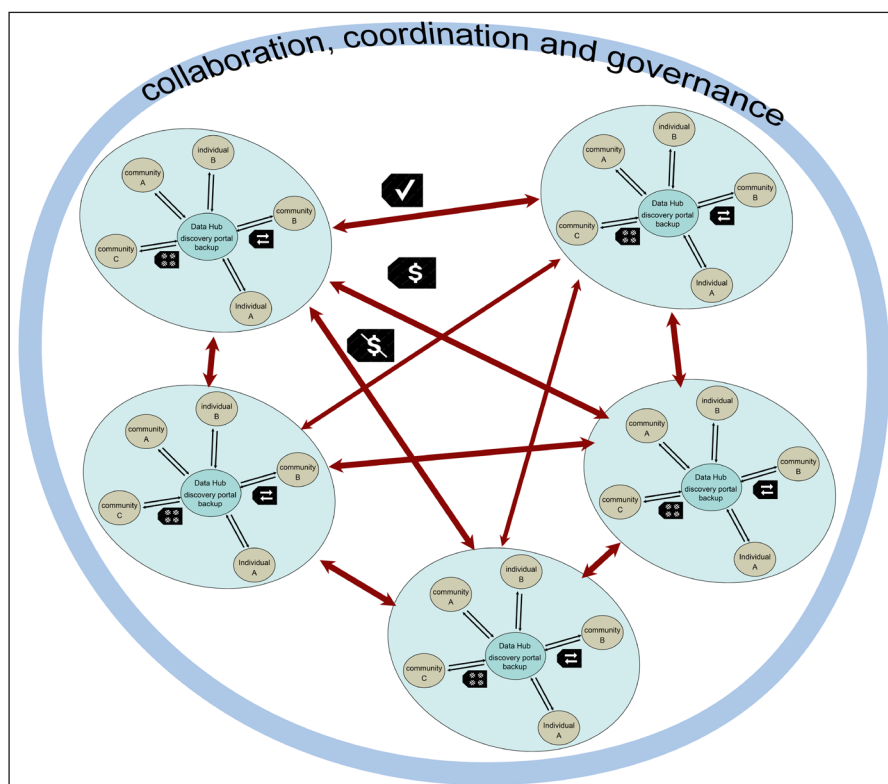


Figure 2 A depiction of collaboration across a suite of organisational units, where information flows between each determined by permissions. The network as a whole takes shared responsibility for setting standards, training, and communication, with a separate set of staff established in a neutral space, financially supported by all the member countries.

To enable this architecture to support discoverability, connectivity, and sharing, and ensure good governance throughout, collaborative coordination is advisable (Figure 2). This will include representatives from all network members and will facilitate hub function and alignment (Berdej and Armitage, 2016; Specht et al., 2021). Governance and coordination committees should take responsibility for establishing principles for the entire network and its members, assist with maintaining common standards, communication and collaboration, and if necessary, advise an overall back-up facility. These committees will have oversight of hub compliance with FAIR and CARE criteria. Outside sources of technology, such as the DKAN open data portal used by the SPREP, and storage facilities such as that available through the National Computational Infrastructure in Australia could be deployed (Cannon et al., 2024).

The sharing conditions for digital objects are determined at local hub level. The labels shown in Figure 2 as illustration are the TK ‘Open to Commercialisation’ and ‘Non-Commercial’ for research and the Verified label, which affirms that the representation and presentation of the material is in keeping with community expectations and cultural protocols (see <https://localcontexts.org>). The illustrated labels are not exhaustive and are used as examples.

2.3. ENABLING TECHNOLOGIES

2.3.1. Blockchain technology

Blockchain technology, with its features of immutability, security, and transparency (Cao et al., 2023; Lewis et al., 2023) is ideally configured to support the flow of information and data through a network. It is distributed by design, and the data are shared across network members, meaning no central entity can control stored data. It suits scenarios where trust and data provenance are crucial to data sharing and exchange. The modularity and simplified governance of blockchain means underlying protocols can be flexible. Blockchain can be tasked to maintain a distributed record of standardised, informative, and machine-readable metadata (Lewis et al., 2023).

Blockchain does not ‘store’ the data itself. A hash (often called a ‘hashtag’) ensures a trustworthy record of data provenance by enabling a verifiable audit trail, strengthening data integrity and accountability (Werder et al., 2022). Any tampering with the data stored in each (sub)country’s centralised database(s) can be easily detected by comparing the hash values, making blockchain a practical approach to balance integrity, security, efficiency, and scalability. Operations can be streamlined, and self-executing contracts can reduce costs with terms directly written into code. Blockchain can act as a bridge between different systems, facilitating interoperability and enhancing overall efficiency. Some of the key features of blockchain technology are shown in Table 1.

Table 1 Key features of blockchain technology.

BLOCKCHAIN FACILITY	
<p>Data Anchoring</p> <p>Critical pieces of data can be linked into a blockchain to ensure their immutability, integrity, and traceability.</p>	<p>The data hashes need only be stored on the blockchain, while the actual data is kept off-chain in centralised databases or distributed storage systems like IPFS (InterPlanetary File System). The hash acts as a unique fingerprint, allowing verification of the data’s integrity without storing it on the blockchain. In doing so, organisations can reduce the costs associated with blockchain’s storage and transaction fees while still benefiting from its security and immutability features (Cao et al., 2021).</p>
<p>Proof of data availability</p> <p>Blockchain can be used to timestamp and verify the availability of data at a certain point in time.</p>	<p>Useful for applications like research data, intellectual property, and legal documentation, where proving the existence and availability of a document or work is crucial.</p>
<p>Data Access Control</p> <p>Blockchain-based access control mechanisms can be implemented.</p>	<p>With access permissions and logs on the blockchain, organisations can control who can access their data and what it can be used for based on policies/agreements (Lomotey et al., 2022). This can enhance data security and ensure that only authorised parties can access sensitive data.</p>
<p>Auditing and Compliance</p> <p>Blockchain’s immutable ledger can be used for auditing purposes.</p>	<p>An immutable audit trail can be maintained on the blockchain. This can be used to track data history, verify data provenance, and ensure compliance with local and international regulatory requirements.</p>

2.3.2. Decentralised identifiers (DID)

Decentralised Identifiers (DIDs) (World Wide Web Consortium, 2024) provide a robust and versatile framework for decentralised identity management across multiple platforms. They are globally unique persistent identifiers that do not require a centralised registration authority and are generated and/or registered cryptographically. They provide a way to uniquely identify the individuals involved in creating, modifying, and transmitting data. They can therefore play a significant role in ensuring data provenance is recorded properly by enhancing the traceability, security, and accountability of data. Using DIDs, each party in a transformation or movement of data can be recorded on a distributed ledger, such as a blockchain. This creates an auditable trail of data provenance that is transparent and tamper-proof, allowing for easy tracing of data back to its origin. DIDs are designed to be interoperable across different systems and platforms. They follow standards set by the World Wide Web Consortium (W3C), ensuring that provenance information can be consistently understood and utilised across various contexts. Each DID provides a unique and cryptographically secure identity that can be used to verify the legitimacy and ownership of the data, enhancing trust in the data's provenance and integrity.

2.3.3. Smart contracts

Smart contracts are digital agreements coded on a blockchain network that are executed automatically upon fulfilling predefined terms and conditions. They can automate and enforce the establishment of common standards for data types, metadata requirements, and vocabulary descriptions, promoting consistency, transparency, and ease of discovery in data management. They are often used to automate the execution of an agreement among participants without the involvement of trusted intermediaries (Lewis et al., 2023). They can be used to enable access permissions to a person or entity, record a data request, and communicate transactions to the primary data owner.

3. THE PROPOSED DATA NETWORK ARCHITECTURE

To enable trust-based data sharing across countries, data provenance, ownership, and accessibility need to be properly and securely implemented. A cyberinfrastructure that supports an immutable and transparent record of data, while enabling control and stewardship of that data to the owner is an attractive prospect. The proposed **Collaborative Regional Data Access Network (CREDAN)** draws upon various network architectures and emerging technologies to achieve these goals.

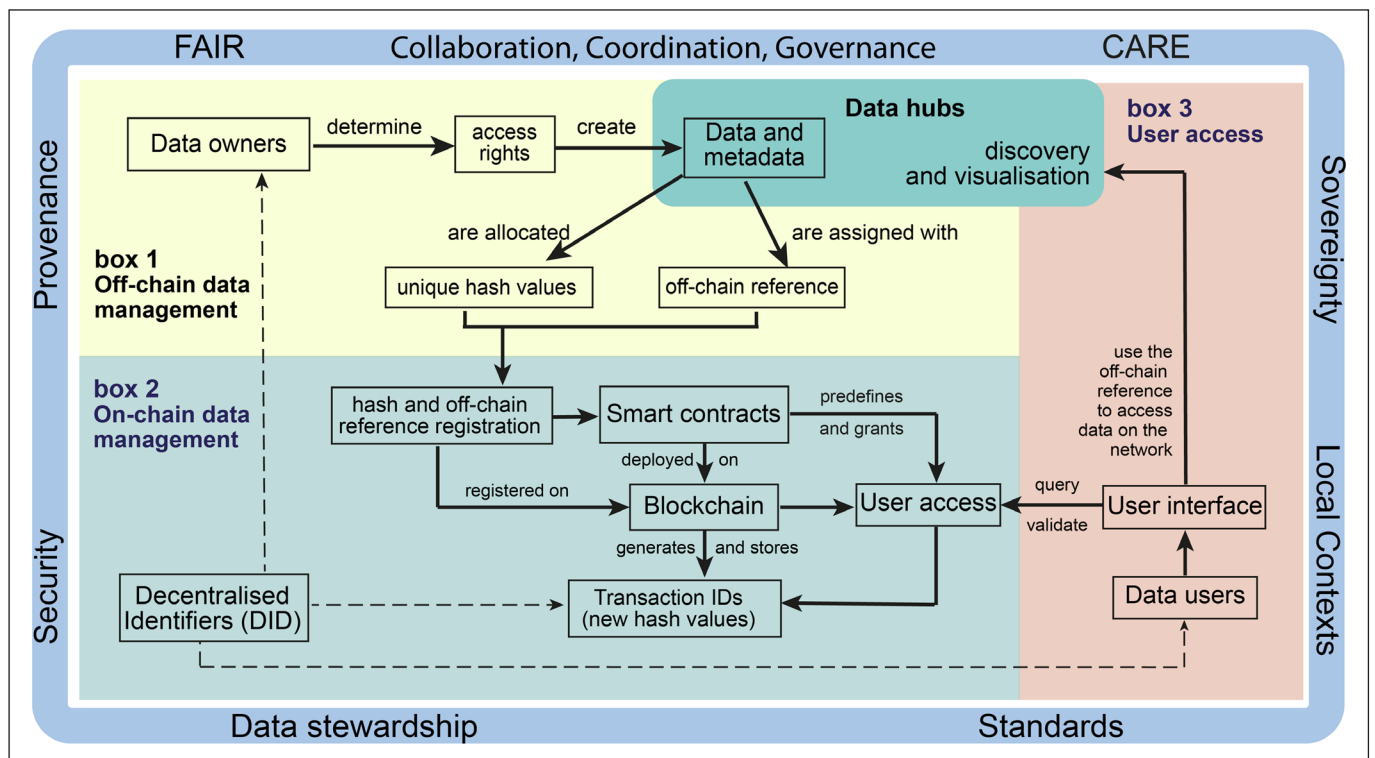
By themselves, centralised and distributed data networks each have inherent advantages and disadvantages. Centralised data networks are advantageous for efficient management, security, and regulatory compliance within a single unit, but face challenges with scalability and redundancy and are slow to respond to changing temporal and cultural circumstances. Distributed systems are designed to provide integrity and accessibility to information from any point. They offer enhanced resilience and collaboration, but they come with higher complexity, costs, and regulatory challenges than centralised (Marstein et al., 2024). At a regional level, a distributed data network has considerable strengths, however, as it can deliver significant benefits, including improved data redundancy, increased resilience against local disruptions, and facilitated cross-border data sharing and collaboration. Distributed networks are scalable, as additional hubs can be added or subtracted, and can provide backup or support to others in the network when required. They offer interoperability and flexibility and can improve discoverability across the networked hubs.

A combination of centralised and distributed data networks benefits from and offsets the qualities of each network type. A model wherein each organisational unit maintains its centralised data infrastructure while participating in a broader distributed network at a regional level enables the optimisation of data governance and resource allocation. By leveraging the strengths of centralised systems—such as uniform policy enforcement, simplified data management, and enhanced operational efficiency—groups such as countries can ensure that their data sovereignty and local data needs are met comprehensively. At the same time, a distributed regional network that connects all participating organisations allows for the pooling of resources, improved data discovery and analytics capabilities through discoverable datasets, and fostering innovation through cooperation.

Integrating blockchain into this model provides a pragmatic approach to achieving cost-effective and efficient data management. The CREDAN will leverage the strengths of both blockchain (transparency, traceability, immutability security) and off-chain databases (efficient storage and access) to provide a balanced solution suitable for various applications requiring secure and scalable data management. Blockchain-enabled databases, such as Paxos, Raft, Zebra, and Google HDFS (Kalajdjieski et al., 2023; Mothukuri et al., 2021), all use a decentralised network of hubs (nodes in blockchain parlance) to efficiently store and process large datasets ranging in size from gigabytes to petabytes of data. The use of blockchain supports the decentralised governance model at the regional level that allows all participants in the Asia Oceania region to have control of data localisation and jurisdictional compliance, ensuring inclusivity and self-autonomy in a distributed data network.

The integrated architecture to enable this network consists of three components: (i) the management of the data and metadata (off-chain), (ii) data management (on-chain), and (iii) user access and verification (bridging). These are all within a strong governance structure, ensuring best practices against international criteria and those set by the component communities (Figure 3).

Figure 3 Visualisation of the CREDAN operational architecture.



As discussed in section 2, the data owners collect and structure their data and metadata and apply access rights, such as local context labels, at the (sub)country level (Figure 3, box 1). The data package (data, metadata, provenance, and access permissions) is stored in the (off-chain) data hubs, from which they can be exposed for discovery. A hash value, unique to each dataset, is generated using a cryptographic hash function (e.g. SHA-256), as a digital fingerprint of the original data and metadata (Figure 3 boxes 1 and 2). Off-chain references are also allocated to enable linkage to the stored data.

The computed hash values and corresponding off-chain references are paired and registered on a blockchain network (Figure 3 boxes 1 and 2). New hash values (transaction IDs) are generated to represent these pairings, ensuring a secure and trackable link between the on-chain data and off-chain references. This is executed by smart contracts. The potential use of decentralised identifiers (DIDs) could add an additional layer of authentication and accountability by linking the data to the data authors and data users. Any change in the input data will generate a new hash, which can be shown as versions of an original dataset and its metadata (Figure 3, box 2).

Data users can access the data through a user interface direct to the blockchain. This will activate the creation of transaction IDs (hash values) to retrieve the off-chain references and provide access to the off-chain database (Figure 3, boxes 2 and 3). User access is governed by smart contracts to ensure secure, automated, and transparent management of permissions and interactions within the network. Once the user has the off-chain reference, they can access the data through the

Data Hubs (Figure 3, box 3). Both data owners and data users would be assigned a unique DID and will establish a clear and immutable record of who created, modified, or interacted with the data. The data access transaction information provides complete provenance of what data was accessed and by whom (Figure 3, box 3). This facilitates trust in the environment.

This configuration can ensure enhanced community or country-level data sovereignty and control while improving regional data access and verification. This process aligns with the data governance principles of FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) to ensure secure and responsible data management (Table 2).

CRITERIA	WHAT CAN THE COLLABORATIVE DATA NETWORK ENABLED BY EMERGING TECHNOLOGIES DELIVER?	FAIR		CARE					
		F	A	I	R	C	A	R	E
Sovereignty	Offers flexible levels of sovereignty.		x				x		x
Provenance	Created in the (sub)country hub, gathered by the coordinating group, comparable across network.		x		x				x
Governance	The (sub)country hubs will manage their data holdings and permissions, the coordinating group ensuring that the necessary provenance, metadata, and vocabularies are maintained for each data object.	x			x		x	x	x
Innovation	Each (sub)country hub will remain autonomous but will be linked to others in the network, offering opportunities for regional data-driven innovations.						x	x	
Collaboration	This model depends and builds on collaboration at hub level, and through discovery through the distributed network will demonstrate the values of collaboration. The governance group will maintain regular communications with all member hubs.						x		
Access and Availability	A combination of controlled and open access allows for fine-tuned usage rules.		x				x	x	x
Integrity and accountability	Combines centralised validation and distributed consensus processes.				x			x	x
Security	The combination of centralised controls and distributed hubs enhances security.				x		x		
Scalability	The distributed network built with on-chain and off-chain capabilities allows enhanced scalability.						x		
Interoperability	Interoperability will be enhanced through the establishment of common standards across the network facilitated by the coordinating group(s).				x		x		
Cost	Initial setup costs will be at the (sub)country and coordinating group levels, but a shared codebase will reduce development costs. Personnel will be an ongoing cost covered by each hub, and the cost of the coordinating group shared by all members of the network. Countries would be expected to need infrastructure funding, such as for computing hardware or enhanced bandwidth and energy capacity. Using a similar data hosting approach will streamline those requests, e.g., by offering proposal templates.						x		x
Energy Consumption	Data being held off-chain minimises energy consumption.						x		

Table 2 Criteria matching for the proposed network enabled by emerging technologies, including a possible alignment to the FAIR and CARE principles. FAIR acronym codes: Findable, Accessible, Interoperable, Reusable; CARE acronym codes: Collective benefit, Authority to control, Responsibility, Ethics.

3.2. FUTURE OPPORTUNITIES

The integration of Large Language Models (LLMs) into the data exploration architecture could significantly improve the process by automating queries and analysis intents, allowing users to interact with data product information through natural language. This approach will simplify data search tasks for specific analyses and facilitate scientific discovery. The ability of LLMs to

process and generate text with human fluency makes them ideal for this purpose, although there are challenges, such as the need to ensure transparency and accuracy.

The integration of generative computing for data exploration is incorporated into the data portal architecture (Figure 3, box 1) as a module for discovering and recommending data products. The data exploration module could offer a user interface in the form of a chatbot, which allows the construction of an integrated system with the APIs of LLM models, such as GPT-3, GPT-4, Llama, and Gemini, among others. These models can be used in specific versions according to the requirements, and, finally, a knowledge base will be used for training the LLMs.

4. CONCLUSIONS AND POSSIBLE IMPLEMENTATION PATHWAY

This paper is concerned with presenting a scenario for individuals, communities, local entities, countries, and regions to assemble and record primary data about their lands in a safe, secure, and transparent manner. We recognise that there are many barriers to sharing data, including the fear that the information will not be treated with proper respect, and we have designed the CREDAN (Collaborative Regional Data Access Network) with the purpose of minimising this fear. The inclusion of blockchain technology is directed at ensuring this is achieved. We hope that, with the emerging categories under the CARE principles defined by the Local Context labels at local hub level (Figure 1), some of these challenges will be overcome. This will require time and sustained focus, and the governance provided by the coordination group (Figures 2 and 3) will be vital to success. The coordination group must ensure standards for data and metadata quality and integration, and ensure they are collaboratively determined and upheld. Management of the eventual large volume of data should be anticipated, and partnerships may need to be considered.

An implementation pathway is suggested below which starts with establishing a common intent, without which the creation of a collaborative network would not be possible.

Step 1: Establish a collaboration between countries and sub-countries and develop a common **intent and goal** for collating and preserving primary (raw) digital objects in their jurisdictions.

Step 2: Establish collaborative coordinating groups which determine general standards and protocols that each (sub)country can use (e.g., applying a suite of local contexts labels of relevance, determining a minimum schema for metadata, identifying existing vocabularies, and if new words are used, requiring a definition). This is important for enabling sharing across the network and beyond.

Step 3: Each (sub)country collects and stores (primary) data related to their area of responsibility using the common standards established by the coordinating group(s).

Step 4: Each (sub)country works with its communities to ensure that metadata, provenance, and permissions are established for each digital object according to the common standards.

Step 5: Each digital object is given a blockchain hash tag and off-chain reference, which immutably records the nature, rights, and permissions of each digital object established in Step 4.

Step 6: The collaborative coordinating group or (sub)countries (depending on the governance model determined), exposes the existence of these digital objects and the hashtags set.

Step 7: Data owners and users obtain unique DIDs to enable full-circle tracking of data acquisition and use.

ACKNOWLEDGEMENTS

The authors would like to thank the SPREP team for their assistance, in particular Tavita Su'a and Lagi Reupena, as well as Jope Davetanivalu, Rafael Tavita, and Vainuupo Jungblut for their support. We also thank Dr Stuart Chape, Director of Island and Ocean Ecosystems, for making the connections happen.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Alison Specht  orcid.org/0000-0002-2623-0854

TERN, the University of Queensland, QLD, Australia

Kim Bryceson  orcid.org/0000-0002-2726-4562

School of Agriculture & Food Sustainability, the University of Queensland, QLD, Australia

Shoufeng Cao  orcid.org/0000-0001-5178-7454

School of Agriculture & Food Sustainability, the University of Queensland, QLD, Australia

Margaret O'Brien  orcid.org/0000-0002-1693-8322

University of California, Santa Barbara, CA, USA

S. M. Guru  orcid.org/0000-0002-3903-254X

TERN, the University of Queensland, QLD, Australia

Pedro Pizzigatti Correa  orcid.org/0000-0002-8743-4244

University of São Paulo, SP, Brazil

Michelle Waycott  orcid.org/0000-0002-0822-0564

State Herbarium of South Australia, the University of Adelaide, SA, Australia

REFERENCES

- Ahl, L., Bellucci, L., Brewer, P., Gagnier, P.Y., Haston, E., Livermore, L., et al.** (2023) 'Digitisation of natural history collections: Criteria for prioritisation', *RIO*, 9, e114548. Available at: <https://doi.org/10.3897/rio.9.e114548>
- Andréfouët, S., Paul, M., and Farhan, A.R.** (2022) 'Indonesia's 13558 islands: A new census from space and a first step towards a One Map for Small Islands Policy', *Marine Policy*, 135, 104848. Available at: <https://doi.org/10.1016/j.marpol.2021.104848>
- Bailey, A., Moglia, M. and Glackin, S.** (2024) 'Participatory justice and climate adaptation for water management in Small Island Developing States: A systematic literature review and discussion', *Regional Environmental Change*, 24, 19. Available at: <https://doi.org/10.1007/s10113-024-02182-y>
- Berdej, S.M. and Armitage, D.R.** (2016) 'Bridging organizations drive effective Governance outcomes for conservation of Indonesia's marine systems', *PLoS ONE*, 11, e0147142. Available at: <https://doi.org/10.1371/journal.pone.0147142>
- Bhusal, T.** (2023) 'Participatory local governance in rural Nepal: The primacy of informality', *Development Policy Review*, 41, e12724. Available at: <https://doi.org/10.1111/dpr.12724>
- Brar, J., Chowdhury, N., Raihan, M.M.H., Khalid, A., O'Brien, M.G., Walsh, C.A. and Turin, T.C.** (2023) 'The benefits, challenges, and strategies toward establishing a community-engaged knowledge hub: An integrative review', *International Journal of Environmental Research and Public Health*, 20, 1160. Available at: <https://doi.org/10.3390/ijerph20021160>
- Brewer T.D., Andrew N.L., Sharp M.K., Thow A.M., Kottafe H. and Jones S.** (2022) 'A method for cleaning international food trade data for regional analysis: The Pacific Food Trade Database | Statistics for Development Division', [WWW Document]. Accessed at: https://sdd.spc.int/digital_library/method-cleaning-international-food-trade-data-regional-analysis-pacific-food-trade (accessed 7.8.24).
- Bryceson, K.P. and Ross, A.** (2020) 'Habitat of informality in small scale society agrifood chains – filling the knowledge gap using a socio-culturally focused value chain analysis tool', *Journal of the Asia Pacific Economy*, 25, 545–570. Available at: <https://doi.org/10.1080/13547860.2019.1670930>
- Cannon, S.E., Moore, J.W., Adams, M.S., Degai, T., Griggs, E., Griggs, J., et al.** (2024) 'Taking care of knowledge, taking care of salmon: Towards Indigenous data sovereignty in an era of climate change and cumulative effects', *FACETS*, 9, 1–21. Available at: <https://doi.org/10.1139/facets-2023-0135>
- Cao, S., Miller, T., Foth, M., Powell, W., Boyen, X. and Turner-Morris, C.** (2021) 'Integrating On-chain and Off-chain Governance for Supply Chain Transparency and Integrity', *Proc. 5th Symposium for Distributed Ledger Technology, Brisbane, Australia, 23 Nov 2021*. Available at: <https://doi.org/10.48550/ARXIV2111.08455>
- Cao, S., Xu, H. and Bryceson, K.P.** (2023) 'Blockchain traceability for sustainability communication in food supply chains: An architectural framework, design pathway and considerations', *Sustainability*, 15(18), p.13486. Available at: <https://doi.org/10.3390/su151813486>
- Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., et al.** (2020) 'The CARE principles for indigenous data governance', *Data Science Journal*, 19, 43. Available at: <https://doi.org/10.5334/dsj-2020-043>
- Carroll, S.R., Herczog, E., Hudson, M., Russell, K. and Stall, S.** (2021) 'Operationalizing the CARE and FAIR Principles for Indigenous data futures', *Scientific Data*, 8, 108. Available at: <https://doi.org/10.1038/s41597-021-00892-0>
- Conrad, C.C. and Hilchey, K.G.** (2011) 'A review of citizen science and community-based environmental monitoring: issues and opportunities', *Environmental Monitoring and Assessment*, 176, 273–291. Available at: <https://doi.org/10.1007/s10661-010-1582-5>

- David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M. and Yahia, M., et al.** (2020) 'FAIRness literacy: The achilles' heel of applying FAIR principles', *Data Science Journal*, 19, 32 Available at: <https://doi.org/10.5334/dsj-2020-032>
- Decentralized Identifiers (DIDs) v1.0 [WWW Document]**, n.d. Available at: <https://www.w3.org/TR/2022/REC-did-core-20220719/> (accessed 7.28.24).
- Durish, N., Gill, R., Houlding, P., Flowers, C., Shiwak, I., Ernst, J. and Gillis, D.** (2021) 'Co-designing a community-led Internet assessment tool in Rigolet, Nunatsiavut, Canada', In: *2021 IEEE International Symposium on Technology and Society (ISTAS)*. Presented at the 2021 IEEE International Symposium on Technology and Society (ISTAS), pp. 1–8. Available at: <https://doi.org/10.1109/ISTAS52410.2021.9629151>
- Ens, E.J., Pert, P., Clarke, P.A., Budden, M., Clubb, L., Doran, B., et al.** (2015) 'Indigenous biocultural knowledge in ecosystem science and management: Review and insight from Australia', *Biological Conservation*, 181, 133–149. Available at: <https://doi.org/10.1016/j.biocon.2014.11.008>
- Happel, A., Murchie, K.J., Willink, P.W. and Knapp, C.R.** (2020) 'Great Lakes Fish Finder App; A tool for biologists, managers and education practitioners', *Journal of Great Lakes Research*, 46, 230–236. Available at: <https://doi.org/10.1016/j.jglr.2019.12.002>
- Hudson, M., Garrison, N.A., Sterling, R., Caron, N.R., Fox, K., Yracheta, J., et al.** (2020) 'Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data', *Nature Reviews Genetics*, 21, 377–384. Available at: <https://doi.org/10.1038/s41576-020-0228-x>
- Huntington, H.P.** (2000) 'Using traditional ecological knowledge in science: Methods and applications', *Ecological Applications*, 10, 1270–1274. Available at: [https://doi.org/10.1890/1051-0761\(2000\)010\[1270:UTEKIS\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2000)010[1270:UTEKIS]2.0.CO;2)
- International Science Council.** (2024) *Pacific Islands Academy of Sciences and Humanities: A Pivotal Step Towards a Resilient Future* [WWW Document], 2024. Available at: https://council.science/blog/pacific-academy_oped/ (accessed 8.30.24).
- Jutel, O.** (2021) 'Blockchain imperialism in the Pacific', *Big Data and Society*, 8, 205395172098524. Available at: <https://doi.org/10.1177/2053951720985249>
- Kalajdjieski, J., Raikwar, M., Arsov, N., Velinov, G. and Gligoroski, D.** (2023) 'Databases fit for blockchain technology: A complete overview', *Blockchain: Research and Applications*, 4, 100116. Available at: <https://doi.org/10.1016/j.bcra.2022.100116>
- Kitamura, K., Kinoshita, Y., Ito, K., Kawabe, S., Kobayashi, H., Naya, H., et al.** (2023) 'Formation of a transdisciplinary community of practice in rural areas, with an interactive database of co-created knowledge: A case study in Noto, Japan', *Gateways*, 16. Available at: <https://doi.org/10.5130/ijcre.v16i2.8696>
- Lewis, R.J., Marstein, K.E. and Grytnes, J.A.** (2023) 'Incentivising open ecological data using blockchain technology', *Scientific Data*, 10, 591. Available at: <https://doi.org/10.1038/s41597-023-02496-2>
- Lin, D., Crabtree, J., Dillo, I., Downs, R.R., Edmunds, R., Giaretta, D., et al.** (2020) 'The TRUST Principles for digital repositories', *Scientific Data*, 7, 1–5. Available at: <https://doi.org/10.1038/s41597-020-0486-7>
- Local Contexts**, n.d. TK Labels – Local Contexts [WWW Document]. Available at: <https://localcontexts.org/labels/traditional-knowledge-labels/> (accessed 8.29.24).
- Lomotey, R.K., Kumi, S. and Deters, R.** (2022) 'Data trusts as a service: Providing a platform for multi-party data sharing', *International Journal of Information Management Data Insights*, 2, 100075. Available at: <https://doi.org/10.1016/j.ijime.2022.100075>
- Marstein, K., Grytnes, J. and Lewis, R.J.** (2024) 'ECKOCHAIN: A FAIR blockchain-based database for long-term ecological data', *Methods in Ecology and Evolution*, 2041–210X.14280. Available at: <https://doi.org/10.1111/2041-210X.14280>
- Martin, D.J., Howard, A., Hutchinson, R., McGree, S. and Jones, D.A.** (2015) 'Development and implementation of a climate data management system for western Pacific small island developing states', *Meteorological Applications*, 22, 273–287. Available at: <https://doi.org/10.1002/met.1461>
- Martínez-Sagarra, G., Castilla, F. and Pando, F.** (2022) 'Seven hundred projects in iNaturalist Spain: Performance and lessons learned', *Sustainability*, 14, 11093. Available at: <https://doi.org/10.3390/su141711093>
- Meeus, S., Silva-Rocha, I., Adriaens, T., Brown, P.M.J., Chartosia, N., Claramunt-López, B., et al.** (2023) 'More than a Bit of Fun: The Multiple Outcomes of a Bioblitz', *BioScience*, 73, 168–181. Available at: <https://doi.org/10.1093/biosci/biac100>
- Mothukuri, V., Cheerla, S.S., Parizi, R.M., Zhang, Q., and Choo, K.K.R.** (2021) 'BlockHDFS: Blockchain-integrated Hadoop distributed file system for secure provenance traceability', *Blockchain: Research and Applications*, 2, 100032. Available at: <https://doi.org/10.1016/j.bcra.2021.100032>
- Moudrý, V. and Devillers, R.** (2020) 'Quality and usability challenges of global marine biodiversity databases: An example for marine mammal data', *Ecological Informatics* 56, 101051. Available at: <https://doi.org/10.1016/j.ecoinf.2020.101051>

- Nicholas, J. and Lyons, K.** (2021) 'Lush forests laid to waste: how Pacific Islands got hooked on logging', *The Guardian*. Accessed at: <https://www.theguardian.com/world/2021/jun/01/lush-forests-laid-to-waste-how-pacific-islands-got-hooked-on-logging>.
- O'Brien, M., Duerr, R., Taitingfong, R., Martinez, A., Vera, L., Jennings, L., et al.** (2024) 'Earth Science Data Repositories: Implementing the CARE Principles', *Data Science Journal*, 23(37), 1–29. Available at: <https://doi.org/10.5334/dsj-2024-037>
- Pacific Community (SPC).** (2022) *Strategic Plan 2022–2031. Sustainable Pacific development through science, knowledge and innovation*. Accessed at: <https://www.spc.int/about-us>.
- Pacific Community (SPC).** (n.d.) [WWW Document]. Accessed at: <https://spc.int/> (accessed 8.30.24).
- Pacific Community (SPC).** (n.d.) *Pacific Data Hub* [WWW Document]. Accessed at: <https://pacificdata.org/> (accessed 8.31.24).
- Pacific Regional Environment Program (SPREP).** (n.d.) [WWW Document]. Accessed at: <https://www.sprep.org/> (accessed 8.30.24).
- Pacific Regional Environment Program.** (n.d.) (b) *Home | Pacific Environment Data Portal* [WWW Document]. Accessed at: <https://pacific-data.sprep.org/> (accessed 8.31.24).
- Pacific Regional Environment Program (SPREP).** (n.d.) (c) *The Inform Project | Pacific Environment* [WWW Document]. Accessed at: <https://www.sprep.org/inform> (accessed 8.30.24).
- Pacific Regional Environment Portal (SPREP).** (n.d.) (d) *DKAN Open Data Portal — DKAN Documentation 2.15.2 documentation* [WWW Document]. Accessed at: <https://dkan.readthedocs.io/en/latest/index.html> (accessed 8.31.24).
- Rujano, M.A., Boiten, J.W., Ohmann, C., Canham, S., Contrino, S., David, R., et al.** (2024) Sharing sensitive data in life sciences: an overview of centralized and federated approaches. *Briefings in Bioinformatics*, 25, bbae262. Available at: <https://doi.org/10.1093/bib/bbae262>
- Secretariat of the Convention on Biological Diversity.** (2011) *Nagoya Protocol* [WWW Document]. Convention on Biological Diversity. Accessed at: <https://www.cbd.int/> (accessed 8.31.24).
- Specht, A., Buttigieg, P.L., Garnier, E., Schildhauer, M. and Walls, R.** (2024) Maximising the reuse of your data with good definitions. *Zenodo*. Available at: <https://doi.org/10.5281/ZENODO.13579485>
- Specht, A., Guru, S.M. and Morris, B.** (2021) 'Enabling integrated modelling systems for environmental prediction at national scale', In: *MODSIM2021, 24th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand*. Available at: <https://doi.org/10.36334/modsim.2021.F3.specht>.
- Stewart, B.** (2023) *A Case for Open Science in Indonesia: Advancing Digital Literacy, Scientific Infrastructure, and Leadership*. Accessed at: <https://www.csis.or.id/publication/a-case-for-open-science-in-indonesia-advancing-digital-literacy-scientific-infrastructure-and-leadership/> (accessed 7.8.24).
- Straza, T.** (2021) *State of Environment and Conservation in the Pacific Islands: 2020 Regional Report*. Accessed at: <https://soec.sprep.org/> (accessed 8.3.24).
- Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., et al.** (2014) 'The eBird enterprise: An integrated approach to development and application of citizen science', *Biological Conservation*, 169, 31–40. Available at: <https://doi.org/10.1016/j.biocon.2013.11.003>
- Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D. and Kelling, S.** (2009) 'eBird: A citizen-based bird observation network in the biological sciences', *Biological Conservation*, 142, 2282–2292. Available at: <https://doi.org/10.1016/j.biocon.2009.05.006>
- Te Kāhui Raraunga.** (n.d.) | *Data ILG* [WWW Document], *The Te Kāhui Raraunga Charitable Trust*. Accessed at: <https://www.kahuiraraunga.io> (accessed 8.30.24).
- Tebbi, A., Chan, T.H. and Sung, C.W.** (2019) 'Multi-rack distributed data storage networks', *IEEE Transactions on Information Theory*, 65, 6072–6088. Available at: <https://doi.org/10.1109/TIT.2019.2927565>
- UNESCO.** (2021) *UNESCO Recommendation on Open Science*. Available from: <https://zenodo.org/record/5834767>. <https://doi.org/10.54677/MNMH8546>
- Underhill, S.J.R., Stringer, R., Bryceson, K., Prasad, B.C. and Shearer, D.** (2011) 'The Pacific Agribusiness Research For Development Initiative (PARDI): A novel approach to horticulture development in the Pacific', *Acta Horticulturae*, 921, 17–24. Available at: <https://doi.org/10.17660/ActaHortic.2011.921.1>
- United Nations.** (n.d.) *Sustainable Development Goals: 17 Goals to Transform our World* [WWW Document]. United Nations Exhibits. Accessed at: <https://www.un.org/en/exhibits/page/sdgs-17-goals-transform-world> (accessed 8.30.24).
- Van Wynsberge, S., Andréfouët, S., Gaertner-Mazouni, N. and Remoissenet, G.** (2015) 'Conservation and resource management in small tropical islands: Trade-offs between planning unit size, data redundancy and data loss', *Ocean and Coastal Management*, 116, 37–43. Available at: <https://doi.org/10.1016/j.ocecoaman.2015.06.031>
- Visconti, P., Di Marco, M., Álvarez-Romero, J.G., Januchowski-Hartley, S.R., Pressey, R.L., Weeks, R. and Rondinini, C.** (2013) 'Effects of errors and gaps in spatial data sets on assessment of conservation progress: Errors and gaps in spatial data sets', *Conservation Biology*, 27, 1000–1010. Available at: <https://doi.org/10.1111/cobi.12095>

- Werder, K., Ramesh, B. and Zhang, R.** (2022) 'Establishing data provenance for responsible artificial intelligence systems. *ACM Transactions Management Information Systems*', 13, 1–23. Available at: <https://doi.org/10.1145/3503488>
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., et al.** (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3,160018. Available at: <https://doi.org/10.1038/sdata.2016.18>
- World Wide Web Consortium.** (2024) *Decentralized Identifiers (DIDs) v1.0 [WWW Document]*. Accessed at: <https://www.w3.org/TR/2022/REC-did-core-20220719/>.

Specht et al.
Data Science Journal
DOI: 10.5334/dsj-2025-001

15

TO CITE THIS ARTICLE:

Specht, A., Bryceson, K., Cao, S., O'Brien, M., Guru, S.M., Correa, P.P. and Waycott, M. (2025) A Collaborative Data Network for the Asia Oceania Region Enabled by Emerging Technologies to Foster Innovation in a Secure and Open Environment. *Data Science Journal*, 24: 1, pp. 1–15. DOI: <https://doi.org/10.5334/dsj-2025-001>

Submitted: 31 August 2024

Accepted: 09 December 2024

Published: 03 January 2025

COPYRIGHT:

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.