frontiers | Frontiers in Environmental Science

TYPE FAIR² Data
PUBLISHED 27 January 2026
DOI 10.3389/fenvs.2025.1723851

Check for updates

FAIR² CERTIFIED™
Explore interactive data here:
https://doi.org/10.71728/senscience.4f2j-8h1k

OPEN ACCESS

EDITED BY
Sawaid Abbas,
University of the Punjab, Pakistan

REVIEWED BY
Martina Gaglioti,
UNESCO Earth Network, Italy
Andres Solorzano,
Catholic University of the Maule, Chile
Keisuke Atsumi,
Biome Inc., Japan

*CORRESPONDENCE
Sebastian Steibl,
sebastian.steibl@auckland.ac.nz,
sebastian.steibl@naturalis.nl

RECEIVED 13 October 2025
REVISED 11 December 2025
ACCEPTED 19 December 2025
PUBLISHED 27 January 2026

CITATION
Steibl S, Burnett MW, Holmes ND, Wegmann A and Russell JC (2026) Atoll biodiversity and environments: an AI-ready, interactive data portal for Indo-Pacific atolls.
Front. Environ. Sci. 13:1723851.
doi: 10.3389/fenvs.2025.1723851

Atoll biodiversity and environments: an AI-ready, interactive data portal for Indo-Pacific atolls

Sebastian Steibl[1,2]*, Michael W. Burnett[3,4], Nick D. Holmes[3], Alex Wegmann[3] and James C. Russell[1]

[1]School of Biological Sciences, University of Auckland, Auckland, New Zealand, [2]Biodiversity Hotspots Group, Naturalis Biodiversity Center, Leiden, Netherlands, [3]The Nature Conservancy, Sacramento, CA, United States, [4]Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara, Santa Barbara, CA, United States

This data article accompanies the release of an interactive, AI-ready data portal that harmonizes information on biodiversity and environments for Indo-Pacific atolls. The data portal collates species inventories using best available data for six terrestrial guilds (arthropods, birds, land crabs, native mammals, vascular plants, reptiles) alongside estimated seabird population sizes, and integrates biogeographic descriptors, land and reef habitat classifications, oceanographic variables, climate metrics, human population data, and information on historical military use. Outputs from models trained on these data are also provided, such as estimated breeding seabird population sizes for each atoll. In total, the different data layers provide information for all 310 Indo-Pacific atolls with permanent, emergent landforms, and comprise 90 standardized variables and 4,215 species records digitized from 677 literature sources, with species catalogues linked to GBIF, BOLD, and NCBI databases via unique taxon identifiers. The portal adheres to FAIR² principles (Findable, Accessible, Interoperable, Reusable, AI-readiness, Responsible AI) to support reproducible biodiversity and environmental analyses and machine-learning workflows. While biodiversity data coverage varies geographically and both environmental and biodiversity data lack temporal replication, the dataset provides the first centralized, interoperable baseline for information on atoll biotic and physical systems, enabling a framework for additional data layers, cross-layer analyses and supporting research, conservation, climate-risk assessment, and policy.

KEYWORDS

atoll, coral reef, dataset, human footprint, island, land use, remote sensing, species catalogue

1 Introduction

Over the past 2 decades, atolls and their low-lying islands have become emblematic of the existential threats posed by climate change (Jarillo and Barnett, 2022). Atoll nations such as Kiribati, the Maldives, and Tuvalu have had disproportionally influential voices in shaping climate policy at the international geopolitical stage, most notably through their advocacy for the 1.5 °C target of the Paris agreement (Benjamin and Thomas, 2016; Sadai et al., 2022; Baldacchino and Antat, 2023). This recent political influence builds on a long history in which atolls, despite their smallness and remoteness, have been sites of disproportinate geopolitical leverage and interest (Constant, 2024; Edel et al., 2025).

Scattered across the tropical Indo-Pacific, atolls provided island steppingstones for the migration of Indigenous people, such as into Oceania (Dickinson, 2003). In later centuries, atolls anchored colonial expansion of Western nations, functioning as gateways for resource extraction and geopolitical control over the remote Indo-Pacific (Aldrich, 1989). During the Second World War and the Cold War, many atolls were transformed into military bases to secure strategic military and geopolitical reach over the Indo-Pacific (Goldberg, 2018). Several atolls were also used for testing and demonstrating the nuclear weapon capabilities of Western powers (Maclellan, 2005). Between the 1950s–1970s, atolls also enabled trans-Pacific connectivity by serving as refueling stopovers for early transoceanic domestic airline routes (Goldberg, 2018).

Each phase of geopolitical engagement was accompanied by Western-scientific exploration, although Traditional Ecological Knowledge has been developing since first human arrival yet largely ignored to date. From medieval Arab scholars in the Maldives, to Darwin's voyage of the *Beagle* and expeditions such as the Whitney South Seas Expedition, to a dedicated journal on atolls ("Atoll Research Bulletin" – most active between 1950s and 1990s), these scientific efforts advanced the documentation of atoll biodiversity and environments (Steibl et al., 2024a). However, most of the generated scientific data on atolls resides in technical reports to military agencies, out-of-print hard-copy journals, or historical scans lacking optical character recognition. As a result, these data remain largely inaccessible to modern scientific literature search engines and data-driven science, ultimately providing a significant barrier to effective, evidence-based decision-making. When digitized and combined with modern methodology, these datasets can be highly valuable for both fundamental and applied, policy-relevant research. Recent analyses of atoll terrestrial biodiversity data have contributed to refining island biogeography theoretical understanding (Steibl et al., 2025), identified atoll islands as globally important seabird nesting hotspots (Steibl et al., 2024b), and provided historical context to satellite-based analyses of atoll vegetation mapping (Burnett et al., 2024). These studies showcase both the enduring relevance of the existing data on atoll biodiversity and environments, as well as its potential when revisited with modern statistical and remote-sensing approaches.

To facilitate this integration of atoll data and enhance its accessibility, the data article presented here introduces an interactive and AI-ready data portal that consolidates terrestrial (island) biodiversity and environmental data for Indo-Pacific atolls. The data portal collates digitized terrestrial species catalogues (arthropods, birds, land crabs, native mammals, vascular plants, reptiles) alongside environmental data layers derived from satellite-based products (Allen Coral Atlas, 2022; EU Copernicus Marine Service, 2023; Burnett et al., 2024). Designed in alignment with FAIR[2] principles (Findability, Accessibility, Interoperability, Reusability, AI-readiness, and responsible AI use) (Verhulst et al., 2025), the portal supports reproducibility, traceability, and quality assessment. By making existing data layers on atoll island biodiversity and environments openly accessible, the data portal seeks to advance environmental research, support conservation initiatives, and inform policy for Indo-Pacific atolls.

2 Methods summary

2.1 Data collation approach

The data portal covers atolls of the Indo-Pacific basin, using the latest global checklists (Goldberg and Rankey, 2025). Only atolls with permanent emergent landforms were included while submerged atolls (*i.e.,* those without permanent landforms) were excluded. Atolls in the Atlantic, Caribbean, South China Sea, and within the Indonesian archipelago were excluded due to insufficient data. Most atolls with emergent landforms comprise dozens to hundreds of individual islands (sometimes also referred to as islet, motu, or cay). All data available in the data portal is collated at the atoll-, not the individual island-level.

Ten data layers were compiled: (1) atoll descriptive data, (2) terrestrial biodiversity, (3) seabird populations, (4) biogeographic variables, (5) environmental variables, (6) land cover, (7) coral reef cover, (8) oceanography, (9) human population, and (10) historic military use. Additionally, shapefiles for each atoll were generated.

2.2 Atoll descriptive data

The descriptive data layer lists all Indo-Pacific atolls that were included in the data portal (see 2.1), along with assignments to archipelagoes, governing countries, and biogeographic region (*i.e.,* Indian Ocean, Micronesia, Melanesia, or Polynesia). Alternative atoll names were compiled from online sources. Atoll centroid coordinates (latitude, longitude) were calculated from the shapefiles.

2.3 Terrestrial biodiversity data

Species inventories were compiled from published records for six terrestrial guilds. Reference lists for each species record per atoll are included in the data portal, including the year of publication as a proxy for the time of observation. Each species record has been cross-referenced via unique, stable, numeric taxon identifiers to the Global Biodiversity Information Framework (GBIF taxon ID), the Barcode of Life Database (BOLD), and the National Center for Biotechnology Information (NCBI taxon ID) databases.

- Arthropod: species lists for most atolls constitute only incomplete species inventories, given a high diversity of species but usually only narrow focuses of individual studies and collections on specific arthropod groups (*e.g.*, butterflies only). Scientific names were harmonized using the GBIF taxonomy (GBIF, 2025).
- Birds: only complete atoll species inventories were included; migratory non-breeding visitors, vagrants, or reports focusing only on specific guilds were omitted. Scientific names were harmonized using Birds of the World taxonomy (Billemann et al., 2025), with only native species included. Extant, locally extirpated, and extinct native species are recorded.
- Land crabs: may contain incomplete species inventories for some atolls, largely because different sources used different

inclusion/exclusion criteria for land crabs (*e.g.,* including shore crab species versus only true terrestrial species). Scientific names were harmonized using the GBIF taxonomy (GBIF, 2025).

- Native mammals (*i.e.,* fruit bats): only complete atoll species inventories were included. Scientific names were harmonized using the GBIF taxonomy (GBIF, 2025). Introduced mammals are not included.
- Vascular plants: only complete atoll species inventories were included. Scientific names were harmonized using Plants of the World taxonomy, and classification of records into native or introduced species were also based on Plants of the World species catalogues (POWO, 2025).
- Reptiles: only complete atoll species inventories were included. Scientific names were harmonized using the Reptile Database taxonomy (Uetz et al., 2025). Marine reptiles with semi-terrestrial life histories (*i.e.,* sea turtles) were not included.

2.4 Seabird population data

Steibl et al. (2024b) used seabird species inventories from atolls along with additional population censuses and delivered a first use case how global-scale atoll terrestrial biodiversity data can be used for downstream analyses by training a Bayesian predictive model to estimate seabird population sizes for 280 Indo-Pacific atolls. This data layer contains the output from this published study and provides median and 95% credible intervals for total and species-specific seabird population sizes per atoll. A detailed description of the modelling approach, the dataset, and its limitations can be found in the original source article (Steibl et al., 2024b).

2.5 Biogeographical data

Measurements for each atoll were conducted using Google Earth 10.66.0.0. KML shapefiles of each atoll were generated manually and used to calculate area, perimeter, and isolation (nearest neighbor distance) measures using R 4.4.3 and the "sf" package for geospatial analysis (Pebesma, 2018). For each atoll, the total number of vegetated islands, total emergent land area, lagoon area, outer perimeter of the atoll, total atoll area, the number of deep-water channels, the combined total width of all channels, as well as the distances to the nearest atoll, large volcanic island, and continent were obtained. The shapefiles are also included in the data portal.

2.6 Environmental data

Long-term average annual rainfall data were obtained from literature and online sources. Storms and cyclone (categories 1–5) frequencies that have passed the atoll within 50-, 100-, and 200-km radius (measured from the outer atoll polygon) were obtained from the Historical Hurricane Track database of the National Oceanic and Atmospheric Administration (NOAA; https://coast.noaa.gov/hurricanes/). Rainfall anomaly data (degree of precipitation deviation during El Nino Southern Oscillation years) were obtained from the Joint Institute for the Study of Atmosphere and Ocean (JISAO; http://research.jisao.washington.edu/data/doe_prec/).

2.7 Land cover classification data

Burnett et al. (2019), Burnett et al. (2024) developed a machine-learning approach to classify land cover and composition from 2-m resolution satellite imagery for Pacific atolls. For atolls of the Indian Ocean, no land classification data were available. The absolute and relative covers by coconut forest, broadleaf forest, shrub vegetation, and non-vegetated land surface were obtained. Additionally, total and relative areas classified as coconut palm monocrops are provided for each atoll, as well as historical data indicating whether copra (the dried endosperm of coconuts from which oil is pressed) was ever produced on an atoll. A detailed description of the methodology, dataset, and its limitations is provided in the original study (Burnett et al., 2024). Additional data layers and island-level classification measurements can be accessed in the original source article (Burnett et al., 2024).

2.8 Coral reef cover data

Reef geomorphic and benthic classifications were obtained from the Allen Coral Atlas (Allen Coral Atlas, 2022). For each atoll, total and relative measurements of reef extent area, reef slope, sheltered reef slope, reef crest, outer reef flat, inner reef flat, back reef slope, shallow lagoon, deep lagoon, plateau, patch reef, coral benthic class, sand benthic class, rubble benthic class, rock benthic class, seagrass benthic class, and microalgal mats benthic class were obtained. A detailed description of the underlying methodology and its limitations, as well as definitions for each geomorphic zone and benthic class can be found in the online portal of the Allen Coral Atlas (https://allencoralatlas.org/methods/) (Kennedy et al., 2021).

2.9 Oceanographic data

For each atoll, nearshore oceanographic data were obtained from EU Copernicus Marine Service Information Global Ocean Biogeochemistry (https://doi.org/10.48670/moi-00015) and Global Ocean Physics Reanalysis (https://doi.org/10.48670/moi-00021) data products. Measurements were taken for a *ca.* 100-km polygon around the perimeter of each atoll, and contain net phytoplankton primary productivity (NPPv), chlorophyll-a concentration, phytoplankton concentration, sea-surface temperature, and sea-surface movement data. Detailed description of the methodology underlying the remote-sensing data can be found on the Copernicus data portal (https://dataspace.copernicus.eu/).

2.10 Human population data

Human population data were compiled from literature and national census reports. The number of permanently inhabited

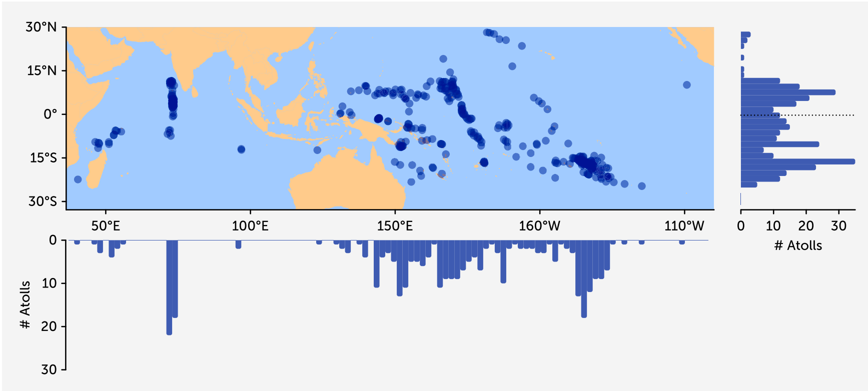

FIGURE 1
Geographic distribution of all 310 Indo-Pacific atolls included in the data portal. The map illustrates geographic distribution; histograms summarize latitudinal and longitudinal frequency distributions of atoll.

islands per atoll, and the cumulative total number of permanent inhabitants per atoll were summarized.

2.11 Historic military use data

For each atoll, a literature search was conducted to identify whether it was used as a military base during WW2 or the Cold War (1939–90). If an atoll was used by military, the country name, type of military usage (*e.g.,* naval base, nuclear testing, airfield) and the time of active usage were obtained.

2.12 Data layer harmonization and integration

All records were standardized into a unified structure. Units were harmonized (areas in km^2, distances in km, environmental data in SI), and taxonomy made consistent using key authorities (see section 2.3). Data layers were linked using atoll names as identifiers. Quality control included identifier checks, screening for plausibility, and plotting each continuous variable in scatter and x–y plots to identify any obvious anomalies or erroneous data entries. Harmonization of taxonomies was based on GBIF (arthropods, crabs, mammals, reptiles), Birds of the World (birds), and Plants of the World Online (plants), and GBIF, BOLD, and NCBI unique numerical taxon identifiers were added to each record.

3 Data overview

3.1 Data summary

The data portal provides data on biodiversity and environments for 310 Indo-Pacific atolls and can be accessed under the following link: https://doi.org/10.71728/senscience.4f2j-8h1k.

Terrestrial biodiversity data include species inventories (presence/absence) for six terrestrial guilds alongside modelled estimates of seabird breeding population size and composition (Steibl et al., 2024b). Environmental data layers describe biogeographic, environmental, and oceanographic conditions, as well as land and reef habitat cover (Kennedy et al., 2021; Burnett et al., 2024). Data layers on human population and historic military use provide insight into past and present human presence on atolls. In total, the data portal includes 90 variables on atoll environments and 4,215 species records digitized from 677 references.

3.2 Quantitative summary of the data layers

3.2.1 Geographic coverage

The data portal covers all Indo-Pacific atolls with permanently emergent landforms (Figure 1).

- Total units: 310 distinct atolls.
- Latitude range: 28.4°N to 24.7°S
- Longitude range: 39.6°E to 109.2°W (but excluding the Indo-Malayan and South China Seas region)

Each atoll is treated as a unique spatial unit, identified through standardized naming (Goldberg and Rankey, 2025). All data layers are aggregated at the atoll level. The data portal excludes 234 submerged atolls (*i.e.,* without emergent land), and an additional 71 emergent atolls located in the Atlantic, Caribbean and Indonesian regions, which fell outside the study scope (Goldberg and Rankey, 2025).

3.2.2 Data layer composition and completeness

At the time of publication, the data portal contained over 30,000 individual records across ten data layers, each capturing a distinct biodiversity or environmental attribute. Terrestrial biodiversity data are available for only subsets of atolls, with varying coverage completeness (Figure 2):

- Bird species catalogues: 2,235 occurrence records across 140 species and 241 catalogued atolls (77.7% completeness).
- Vascular plant species catalogues: 9,430 occurrence records across 1,253 species and 125 catalogued atolls (40.3%

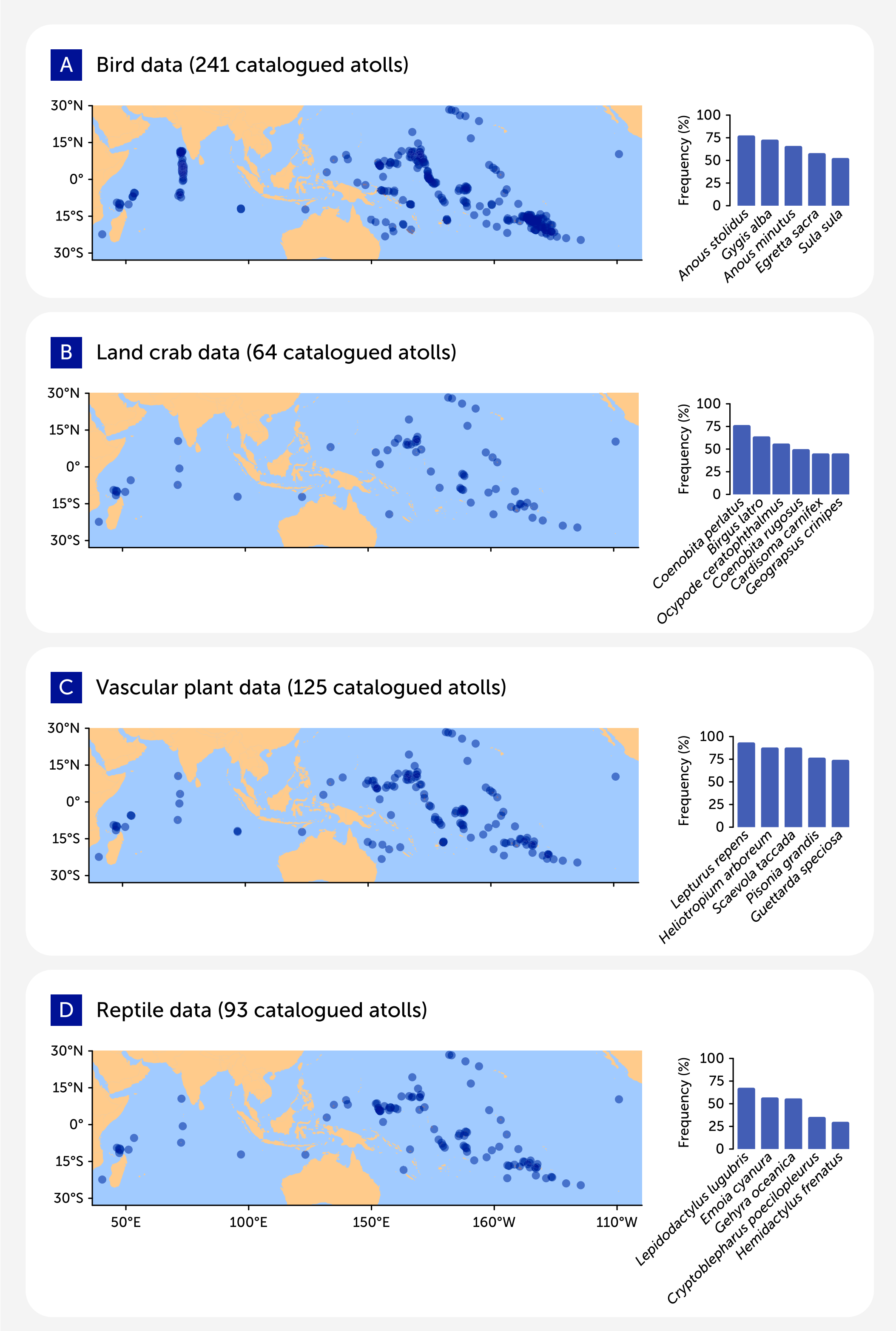

FIGURE 2
Geographic distribution of atolls for which terrestrial biodiversity data layers were available. Panels **(A–D)** illustrate the coverage for available **(A)** bird, **(B)** land crab, **(C)** vascular plant, and **(D)** reptile species occurrence data. Histograms show the five most frequently recorded species across all catalogued atolls.

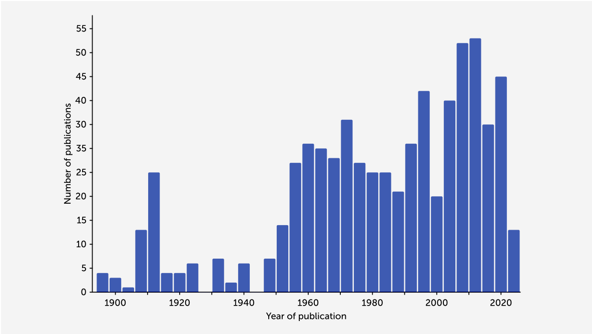

FIGURE 3
Histogram summarizing year of publication of the atoll biodiversity records. The 677 digitized publications span from 1894 to 2025, but most references were published between *ca.* 1960 to 2020.

completeness), along with nativeness status (N: native, I: introduced) of each species per atoll.

- Reptile species catalogues: 494 occurrence records across 60 species and 93 catalogued atolls (30.0% completeness).
- Native mammal species catalogues: 46 occurrence records across 15 species and 39 atolls (12.6% completeness). Only presence information of native mammal (*i.e.,* fruit bats) is included, not confirmed absences of any terrestrial native mammal species.
- Land crab species catalogues: 505 occurrence records across 56 species and 64 atolls (20.6% completeness).
- Arthropod species catalogues: 5,650 occurrence records across 2,690 taxa and 68 atolls (21.9% completeness).
- Seabird population estimates: 10,213 presence and absence records (species × atoll) of observed or modeled abundances for 274 atolls (88.4% completeness) (Steibl et al., 2024b).

Data on atoll environments were available for most atolls of the Indo-Pacific, with only small gaps in the geographic coverage within each layer:

- Atoll descriptive data: 6 variables, available for all 310 atolls (100% completeness).
- Biogeographic data: 13 variables, available for all 310 atolls (100% completeness).
- Environmental data: 6 variables, with missing entries for rainfall and rainfall anomaly for 18 and 36 atolls, respectively (94.1% and 88.4% completeness, for other variables 100% completeness).
- Land cover data: 17 variables, available only for 200 Pacific atolls (64.5% completeness) (Burnett et al., 2024).
- Reef cover data: 35 variables, available for 309 atolls (99.7% completeness).
- Oceanographic data: 5 variables, available for 275 atolls (88.7% completeness).
- Human population data: 2 variables, available for 308 atolls (99% completeness).
- Historic military use data: 4 variables, available for all 310 atolls (100% completeness).

Detailed information of each individual variable, the data type, unit of measurement, and description of underlying methodology can be found in the Data Dictionary section of the data portal.

3.2.3 Temporal coverage

The data portal integrates multi-decadal records from field surveys, satellite imagery, and governmental censuses. While most variables on atoll environments represent recent or long-term average conditions, atoll terrestrial biodiversity records span more than a century:

- Biodiversity data: Published species catalogues span 1894–2025, but the majority from *ca.* 1960 to 2020 (median year 1990; Figure 3). Publication reference lists for each species record along with the date of record can be found in the biodiversity reference datasheets in the data portal.
- Seabird population data: Survey data underlying the population models span 1968–2023 (Steibl et al., 2024b).
- Biogeographical data: Obtained from shapefiles generated in Google Earth 10.66.0.0 in 2024–2025.
- Environmental data: Historical hurricane track data span 1942–2021. Rainfall anomaly data span 1979–2010.
- Land cover data: satellite imagery used by Burnett et al. (2024) span 2009–2020.
- Reef cover data: satellite imagery of the Allen Coral Atlas was obtained from the 2022 map release.
- Oceanographic data: biogeochemistry data (primary productivity, plankton biomass) based on a four-year average (2019–2023); ocean physics data (temperature, ocean velocity) based on a 30-year average (1993–2023). Each value is the long-term mean over the entire available data record.
- Human population data: Census records span 2000–2024 (median year 2022).

3.3 FAIR^2 compliance certification

The data portal is in full compliance with the FAIR^2 framework, which extends the FAIR (Findable, Accessible, Interoperable, and Reusable) principles by incorporating AI Readiness (AIR) and Responsible AI (RAI) standards.

3.3.1 Overall FAIR^2 badge compliance

Compliant–The data portal qualifies for the FAIR^2 Badge, meeting all necessary criteria. The portal is findable through a DOI and schema. org/dataset metadata (Google Dataset Search), openly accessible under ODC-By v1.0 with long-term archiving in Zenodo, and available programmatically via the MLCommons mlcroissant API. Interoperability is ensured through open formats (CSV, JSON-LD), controlled vocabularies (POWO, HBW, QUDT), and partial Darwin Core alignment. Reusability is supported by comprehensive documentation, provenance, versioning, and licensing. The data portal is AI-ready, structured for ML workflows with scalable RecordSets, and meets Responsible AI standards by documenting biases, ensuring interpretability, and providing transparency, reporting, and provenance for accountable use.

TABLE 1 The FAIR2 Compliance Certification presented here was generated through a Human-in-the-Loop (HITL) process combining automated FAIR2 system checks with author-supplied inputs. While certain metadata fields and validations (e.g., DOI registration, schema adherence, file accessibility) are verified automatically by the FAIR2 platform, other elements—such as domain—specific documentation quality and Responsible AI considerations-reflect expert curation by the dataset authors.

Criteria	Assessment
Findability (F)	
F1. Unique identifier	The data portal is assigned a DOI registered with DataCite, ensuring global traceability and citability
F2. Metadata	Metadata includes data portal title, creators, affiliations, abstract, geographic coverage (310 atolls across indo-pacific), temporal coverage (dependent on variable), variable description for each layer and original data source reference
F3. Metadata includes data identifiers	The DOI is explicitly referenced within the metadata, ensuring association between the metadata record and the data layers
F4. Searchable metadata	Metadata is exposed using schema.org/Dataset type and indexed by google dataset search. Key fields (e.g., species, order, family) are Darwin core–inspired but not yet declared with formal DwC properties
Indexed in repositories	Indexed in the FAIR^2 data portal and in zenodo
Accessibility (A)	
A1. Open access	The data portal is publicly accessible under the open data commons attribution license (ODC-by v1.0), permitting unrestricted reuse, redistribution, and modification with appropriate attribution
A2. Long-term access	Archived in zenodo with DOI permanence through DataCite, mirrored in institutional repositories to ensure redundancy
API access	Access provided through the MLCommons mlcroissant API, which exposes structured metadata and downloadable components. No specialized ecological API (e.g., GBIF, OBIS) is implemented. Interactive data summaries and visualizations are available in the FAIR^2 portal
Interoperability (I)	
I1. Standardized formats	Provided in CSV (tabular) and JSON-LD (semantic metadata). Both are widely compatible with analysis platforms
I2. Controlled vocabularies	Taxonomy harmonized using taxonomic authorities (birds of the world; plants of the world; reptile database; GBIF). Units standardized with QUDT. Metadata schema based on schema.org and PROV-O. Darwin core alignment is partial: biodiversity tables can be mapped but are not yet fully DwC-compliant
I3. Cross-platform integration	Data portal structure aligns with MLCommons croissant v1.0 for AI workflows and broadly with ecological standards, but biodiversity records need explicit DwC export to ensure full integration with biodiversity repositories like GBIF.
Reusability (R)	
R1. Comprehensive documentation	Documentation includes detailed methods, variable-level data dictionary, provenance notes, and quality control description A structured data dictionary describes all scalar volume fields, units, and derivations
R1.1 . License	Released under ODC-By v1.0, permitting reuse, redistribution, and modification with attribution
R1.2. Detailed provenance	Each variable is linked to its source (literature, global databases, or modelling outputs). Provenance is also encoded using the PROV-O ontology, linking variables directly to the methods that generated them for machine-readable traceability
R1.3. Domain-relevant standards	Environmental variables align with CF conventions. Biodiversity variables resemble Darwin core but require explicit mapping for full compliance
Versioning and updates	Current version is static and includes embedded version metadata. Updates are documented in the release history
AI-readiness (AIR)	
Structured for machine learning	Data layers are provided in tidy tabular format with consistent identifiers, suitable for ingestion into ML workflows. JSON-LD schema supports semantic enrichment
Scalable	The data layers (≈33,000 records across variable × atolls) is light for standard ML pipelines but structured to scale across high-performance computing (HPC) and geospatial ML environments
Training and validation sets	Natural partitioning into RecordSets supports division into training, validation, and testing sets for AI applications

(Continued on following page)

TABLE 1 (*Continued*) The FAIR2 Compliance Certification presented here was generated through a Human-in-the-Loop (HITL) process combining automated FAIR2 system checks with author-supplied inputs. While certain metadata fields and validations (e.g., DOI registration, schema adherence, file accessibility) are verified automatically by the FAIR2 platform, other elements—such as domain—specific documentation quality and Responsible AI considerations-reflect expert curation by the dataset authors.

Criteria	Assessment
Responsible AI (RAI)	
Ethical standards and misuse	The data portal is intended for ecological, environmental, and conservation research. It is not designed for demographic or political analysis. Guidance on scope is documented to reduce misuse
Biases in the data portal	Coverage varies by taxon: Seabirds and plants are well-represented; reptiles, land crabs, and mammals are less comprehensively surveyed. Temporal bias toward more accessible or historically studied atolls is acknowledged
Data privacy and security	No personal or sensitive data. Population counts are aggregated at atoll level
Fairness and non-discrimination	Sampling spans over 23 countries' territories, but research bias towards more accessible atolls may skew representation
Explainability and interpretability	Data dictionary provides definitions, measurement methods, and units, enabling interpretability
Data provenance and accountability	Variable-level provenance is encoded using PROV-O and documented in supplementary materials. Primary data sources and modeling methods are transparently cited
Transparency and reporting	The data portal is described in a published FAIR^2 data article, with methodological references and known limitations
AI safety and fairness	Risks include over-reliance on modelled seabird abundances where empirical data are sparse. Credible intervals are provided to mitigate misinterpretation
Ethical and social impact	Supports biodiversity monitoring and conservation policy in sensitive ecosystems
Human-in-the-loop (HITL) considerations	Manual curation was applied to biodiversity lists, habitat annotations, and literature-derived measurements

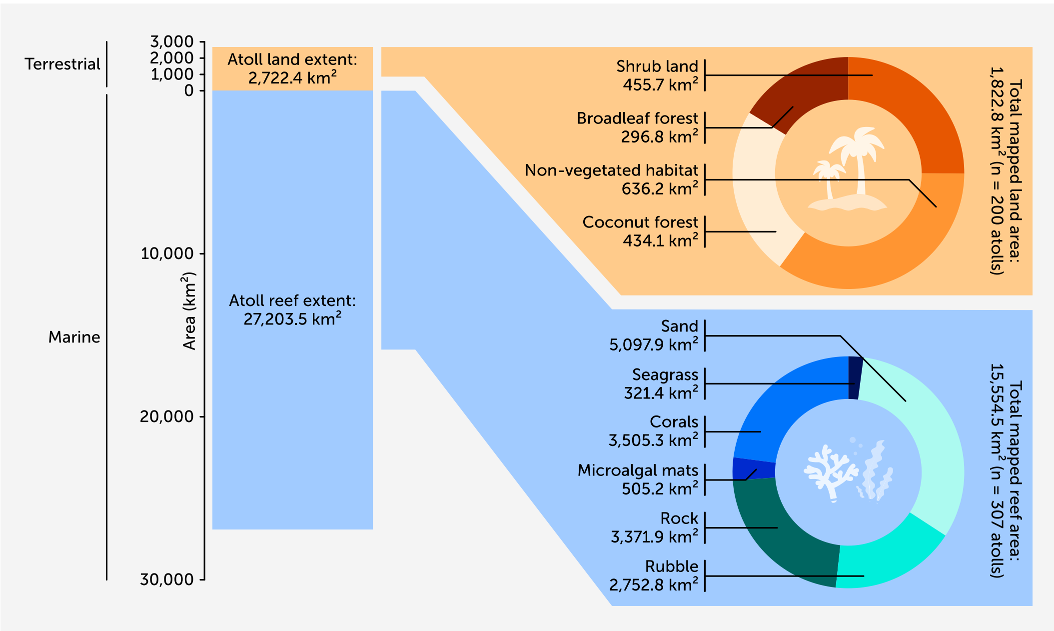

FIGURE 4
Summarized remote sensing data of atoll land and coral reef cover. Left panel: combined land and coral reef area across all 310 Indo-Pacific atolls. Right panel: classification of terrestrial and marine habitats. Island cover classification data from Burnett et al. (2024), reef cover classification data from Allen Coral Atlas (2022).

4 Discussion

4.1 The value of the data portal

Atolls constitute the most numerous island type in the Indo-Pacific basin (Nunn et al., 2016) and represent hotspots and refugia for coral reef and island biodiversity (Steibl et al., 2024b). The data portal presented here provides the first centralized, interactive, AI-ready hub for accessing information on atoll terrestrial biodiversity and environments, designed to support research, conservation, and policymaking.

The integration of multiple data layers that collectively describe atoll environments in the data portal enables novel insights into patterns and processes of atoll systems and promotes a more comprehensive understanding of these unique ecosystems. For instance, combining remote-sensing data on both island and reef habitats highlights the disproportionate contribution of marine ecosystems in atoll systems (Figure 4). While Indo-Pacific atolls collectively comprise only *ca.* 2,720 km^2 of land area, their coral reef extent is an order of magnitude larger, totaling *ca.* 27,200 km^2. Such cross-layer analyses allow the relative contributions of land and reef environments to be quantified, compared, and linked, thereby providing an empirical basis for evaluating conservation priorities and guiding management strategies.

Similarly, the environmental descriptors of atolls provided in the data portal can give insights into the spatial variability in environmental conditions across the Indo-Pacific atolls, such as rainfall patterns, cyclone frequencies, or rainfall anomalies. The data portal enables exploring these gradients using a single platform. This can facilitate a more nuanced, data-driven understanding on how atolls vary in environmental pressures (*e.g.,* variability in drought, cyclone prevalence), which can be integrated into climate risk assessments, adaptation planning, and mitigation frameworks (Fellowes et al., 2024).

Additionally, the island biodiversity data paired with key biogeographic properties such as area and isolation measures, that are directly provided by the data portal, allows for either explicit biogeographical studies on atoll systems in island biology research, or for integration of atolls into larger, global-scale efforts aimed at mapping island or global terrestrial biodiversity and identifying fundamental patterns in biodiversity assembly (Fernández-Palacios et al., 2021).

The $FAIR^2$ structure (Findability, Accessibility, Interoperability, Reusability, AI-readiness, and responsible AI use) enables integration with global biodiversity and climate databases and supports AI-ready workflows for environmental modelling. Beyond enabling cross-disciplinary analyses, the portal provides a critical baseline for monitoring environmental change, informing conservation priorities, and supporting evidence-based policy decisions for atoll systems across the Indo-Pacific.

4.2 Limitations

The data portal has several limitations that should be considered before exploring or analyzing individual data layers. For terrestrial biodiversity and land cover classification data, the geographic coverage is uneven. Biodiversity inventory data are available for only a subset of atolls, with coverage biased towards atolls that are more accessible, but consequently also more developed. Some regions, such as atolls of Papua New Guinea, Federated States of Micronesia, or the Maldives, remain largely unmapped for terrestrial biodiversity and may thus constitute major gaps in species distribution patterns (Figure 2). Land cover mapping data, provided by Burnett et al. (2024), is available only for 200 Pacific atolls, leaving Indian Ocean atolls unmapped. Species lists for land crabs and arthropods are incomplete and do not necessarily record all presences and absences for a given atoll comprehensively, so may not be suited for certain community ecology analyses that require complete species catalogues.

A further constraint is that atoll island biodiversity data are compiled at the atoll-, not the individual island level. Because most atolls consist of numerous islands, these species lists are best interpreted as measures of γ-diversity rather than α-diversity. Combining whole-atoll (vis-à-vis archipelago) biodiversity with island-level data from other island types (*e.g.,* individual volcanic islands) in macroecological analyses may therefore risk conflating processes operating at different scales (Steibl et al., 2025). Furthermore, it is important to consider imperfect detection, cryptic species, and issues around true absences in comparative biodiversity analyses. Additionally, biodiversity data is based on literature spanning over a century in temporal coverage (Figure 3). Therefore, these inventories may not reflect present-day conditions of community composition, especially given the high dynamisms of atoll island systems producing species turnover (Steibl and Russell, 2024), and the growing number of introduced alien species globally (Mormul et al., 2022) and on atoll islands specifically (Nishida and Beardsley, 2002).

Data layers on atoll environments achieve near-complete coverage across all Indo-Pacific atolls, but limitations may arise in terms of interoperability due to varying spatial resolutions of the underlying satellite imagery (*e.g.,* land cover mapping at 0.5–2-m spatial resolution, Burnett et al., 2024; reef cover mapping at 3.125-m spatial resolution; Allen Coral Atlas, 2022). Additionally, some of the biogeographic measures were computed from manually generated shapefiles produced in Google Earth and thus represent static snapshots of geomorphological conditions at the time of generation. For example, atoll islands can change rapidly and substantially in size, position, and elevation within few months due to shifting oceanographic conditions, seasonal changes in wave energies, or climate change (Kench et al., 2024). These dynamisms and other context-specific knowledge of atoll systems need to be taken into account in any downstream analysis. The static, snapshot generation of biogeographic measures such as island areas provided by the data portal are not incorporating these dynamisms. Temporal resolution can also pose constraints. Both biodiversity and environmental data layers represent snapshots rather than repeated surveys. Environmental variables are long-term averages (*e.g.,* rainfall, cyclones) and obtained from primary sources in 2024–2025, but they are not dynamically linked to real-time monitoring stations. As climate change alters regional weather regimes, these averages may lose accuracy in the future, and they do not capture short-term variability or recent weather extremes.

Together, the portal is best viewed as providing baseline conditions for biodiversity and environmental descriptors suitable for biogeographical, macro-ecological, and other environmental research as well as for informing conservation or policy strategies, but does not constitute a dynamic tool for real-time tracking ecological change or disturbance responses.

4.3 Best practice guidelines for data portal usage

To ensure correct application of the data portal and to minimize risk of data misinterpretation, we recommend the following best-practice guidelines for use in scientific analyses and conservation or policy reports:

1. Engage with the authors to discuss intended use and confirm that the data layers are suitable for the planned analysis or presentation.
2. Consider the spatial and temporal limitations, including uneven geographic coverage, incomplete species inventories (land crabs, arthropods), and temporal constraints (see section 4.2).
3. Consult the original sources (Allen Coral Atlas, 2022; Burnett et al., 2024; Steibl et al., 2024b) for detailed methodologies, modelling assumptions, caveats, and limitations of individual data layers.
4. Use the data portal in combination with contextual knowledge. Carefully consider the historical and geographical context of each data layer to prevent misinterpretations of apparent absences, biases, or sampling gaps.
5. Check the metadata files included in the $FAIR^2$ package for information on provenance, completeness, and variable definitions.
6. Cross-check when integrating with other global datasets to ensure taxonomic and spatial consistency.
7. Cite this data article and the data portal.

5 Conclusion

The data portal provides the first harmonized integration of terrestrial biodiversity and environmental information across all Indo-Pacific atolls, spanning broad geographic, taxonomic, and thematic coverage. By collating terrestrial species inventories alongside land and reef habitat cover mapping, environmental variables, and human-use data, the portal establishes a centralized and interoperable platform for atoll research, conservation, and policymaking.

While the data portal is constrained by uneven spatial coverage, lack of temporal replication, and gaps in data completeness, it nonetheless represents an important advance in the availability of standardized and interoperable information on atolls. The portal's $FAIR^2$ structure ensures accessibility, reproducibility, and integration with AI-ready workflows.

The data portal offers not only immediate insights into biodiversity patterns and environmental variation across atolls but also establishes a foundation for long-term research, monitoring, and management. By supporting comparative analyses, conservation planning, and adaptive management, it provides a critical baseline for understanding and sustaining atoll ecosystems under accelerating global change.

Data availability statement

All presented data layers can be accessed in the accompanying $FAIR^2$ interactive data portal: https://sen.science/doi/10.71728/senscience.4f2j-8h1k.

The dataset is $FAIR^2$-certified and publicly available under the Open Data Commons Attribution License (ODC-By v1.0), permitting unrestricted reuse with appropriate attribution. Access is provided through two coordinated components: an interactive $FAIR^2$ Data Portal enabling visual exploration, and a downloadable $FAIR^2$ Data Package containing raw data, structured metadata, and detailed methodological documentation. The $FAIR^2$ Data Package and Data Portal are accessible via https://doi.org/10.71728/senscience.4f2j-8h1k.

Author contributions

SS: Conceptualization, Supervision, Data curation, Writing – original draft, Validation, Writing – review and editing. MB: Methodology, Data curation, Validation, Writing – review and editing. NH: Writing – review and editing. AW: Writing – review and editing. JR: Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. SS funded by Lynen fellowship from the Alexander von Humboldt foundation. Data compilation and initial database concept drafting was funded by The Nature Conservancy.

Acknowledgements

We thank the team behind $FAIR^2$, and namely, Cristina Gonzalez and Sean Hill, for developing and implementing the interactive data portal. The Frontiers Planet Prize, for facilitating the development of this AI-ready data portal. The Alexander von Humboldt foundation for financial support of SS.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. This $FAIR^2$ data article was developed for data use that is AI-ready and uses responsible AI. The data portal itself includes the Clarivate AI chatbot that enables interactive engagement with the data.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aldrich, R. (1989). European expansion in the island Pacific: a historiographical review. *Itinerario* 13, 87–102. doi:10.1017/S0165115300004332

Allen Coral Atlas (2022). Imagery, maps and monitoring of the world's tropical coral reefs. doi:10.5281/zenodo.3833242

Baldacchino, G., and Antat, S. (2023). "Degrees of success: the alliance of small island states at the 2015 paris climate change conference," in *The success of small states in international relations*. Editor G. Baldacchino (London, UK: Routledge), 124–135.

Benjamin, L., and Thomas, A. (2016). 1.5 to stay alive? AOSIS and the long term temperature goal in the paris agreement. *SSRN*, 1–5. doi:10.2139/ssrn.3392503

Billemann, S. M., Keeney, B. K., Kirwan, G. M., Medrano, F., Sly, N. D., and Smith, M. G. (2025). *Birds of the world*. Ithaca, NY, USA: Cornell Lab of Ornithology. doi:10.2173/bow

Burnett, M. W., White, T. D., McCauley, D. J., De Leo, G. A., and Micheli, F. (2019). Quantifying coconut palm extent on Pacific islands using spectral and textural analysis of very high resolution imagery. *Int. J. Rem. Sens.* 40, 7329–7355. doi:10.1080/01431161.2019.1594440

Burnett, M. W., French, R., Jones, B., Fischer, A., Holland, A., Roybal, I., et al. (2024). Satellite imagery reveals widespread coconut plantations on Pacific atolls. *Env. Res. Lett.* 19, 124095. doi:10.1088/1748-9326/ad8c66

Constant, F. (2024). *Overseas territories in world affairs: linking up subnational, national and international politics*. Cham, Switzerland: Palgrave Macmillan, Springer Nature, 1–194. doi:10.1007/978-3-031-64233-3

Dickinson, W. R. (2003). Impact of mid-holocene hydro-isostatic highstand in regional sea level on habitability of islands in Pacific oceania. *J. Coast. Res.* 19, 489–502.

Edel, C., Paik, K., and Augé, J. (2025). *Shifting tides – the national security implications of the united states' Pacific drawdown*. Washington, DC: Center for Strategic and International Studies, 1–12.

EU Copernicus Marine Service (2023). Global ocean biogeochemistry hindcast. doi:10.48670/moi-00019

Fellowes, T. E., Vila-Concejo, A., Byrne, M., Bruce, E., and Baker, E. (2024). Risk classification of low-lying coral reef islands and their exposure to climate threats. *Sci. Tot. Environ.* 912, 168787. doi:10.1016/j.scitotenv.2023.168787

Fernández-Palacios, J. M., Kreft, H., Irl, S. D. H., Norder, S., Ah-Peng, C., Borges, P. A. V., et al. (2021). Scientists' warning – the outstanding biodiversity of islands is in peril. *Glob. Ecol. Conserv.* 31, e01847. doi:10.1016/j.gecco.2021.e01847

GBIF (2025). GBIF.org. Available online at: https://www.gbif.org/.

Goldberg, W. M. (2018). "Domination of Pacific islands in war and in the nuclear age," in *The geography, nature and history of the tropical Pacific and its islands*. Editor W. M. Goldberg (World Regional, Geography Book Series, Springer), 157–180. doi:10.1007/978-3-319-69532-7_7

Goldberg, W. M., and Rankey, E. C. (2025). A revision and critique of atoll numbers based on satellite imagery. *Pac. Sci.* 79, 105–230. doi:10.2984/79.1.6

Jarillo, S., and Barnett, J. (2022). Repositioning the (is)land: climate change adaptation and the atoll assemblage. *Antipode* 54, 848–872. doi:10.1111/anti.12814

Kench, P. S., Sengupta, M., Ford, M. R., and Owen, S. D. (2024). Island change framework defines dominant modes of atoll island dynamics in response to environmental change. *Comm. Earth Environ.* 5, 585. doi:10.1038/s43247-024-01757-1

Kennedy, E. V., Roelfsma, C. M., Lyons, M. B., Kovacs, E. M., Borrego-Acevedo, R., Roe, M., et al. (2021). Reef cover, a coral reef classification for global habitat mapping from remote sensing. *Sci. Dat.* 8, 196. doi:10.1038/s41597-021-00958-z

Maclellan, N. (2005). The nuclear age in the Pacific islands. *TCP* 17, 363–372. doi:10.1353/cp.2005.0062

Mormul, R. P., Vieira, D. S., Bailly, D., Fidanza, K., da Silva, V. F. B., da Graca, W. J., et al. (2022). Invasive alien species records are exponentially rising across the Earth. *Biol. Inv.* 24, 3249–3261. doi:10.1007/s10530-022-02843-1

Nishida, G. M., and Beardsley, J. W. (2002). A review of the insects and related arthropods of midway atoll. *Bish. Mus. Occas. Pap.* 68, 25–69.

Nunn, P. D., Kumar, L., Eliot, I., and McLean, R. (2016). Classifying Pacific islands. *Geosci. Lett.* 3, 7. doi:10.1186/s40562-016-0041-8

Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. *R J.* 10, 439–446. doi:10.32614/RJ-2018-009f

POWO (2025). *Plants of the world online*. Facilitated by the Royal Botanic Gardens, Kew. Available online at: https://powo.science.kew.org/.

Sadai, S., Spector, R. A., DeConto, R., and Gomez, N. (2022). The paris agreement and climate justice: inequitable impacts of sea level rise associated with temperature targets. *Earth's Future* 10, e2022EF002940. doi:10.1029/2022EF002940

Steibl, S., and Russell, J. C. (2024). Beta-diversity within coral atolls: terrestrial species turnover increases with cyclone frequency. *Glob. Ecol. Biogeogr.* 33 (3), 450–457. doi:10.1111/geb.13797

Steibl, S., Bunbury, N., Young, H. S., and Russell, J. C. (2024a). A renaissance of atoll ecology. *Annu. Rev. Ecol. Evol. Sys.* 55, 301–322. doi:10.1146/annurev-ecolsys-102722-013447

Steibl, S., Steiger, S., Wegmann, A. S., Holmes, N. D., Young, H. S., Carr, P., et al. (2024b). Atolls are globally important sites for tropical seabirds. *Nat. Ecol. Evol.* 8, 1907–1915. doi:10.1038/s41559-024-02496-4

Steibl, S., Steiger, S., Valente, L., and Russell, J. C. (2025). Rainfall increases conformity and strength of species–area relationships. *Ecography* 2025, e08159. doi:10.1002/ecog.08159

Uetz, P., Freed, P., Aguilar, R., Reyes, F., Kudera, J., and Hosek, J. (2025). The reptile database. Available online at: http://www.reptile-database.org.

Verhulst, S., Zahuranec, A. J., and Chafetz, H. (2025). Moving toward the FAIR-R principles: advancing AI-ready data. *SSRN*. doi:10.2139/ssrn.5164337