# Spatial Data Provenance Guidelines for Nature-Based Solutions (NbS)

Global Stocktake Submission | March 6, 2023

Contact Person:

Louisa Durkin
Director of Partnership & Impact, OpenEarth Foundation
Louisa@OpenEarth.org

This document has been jointly prepared by the **OpenEarth Foundation** and the **Nature4Climate** coalition. It has been endorsed by the following entities and individuals:

- Arboretica
- Capitals Coalition
- Climate Action Data 2.0 Community (CAD2.0)
- IceBreaker1
- Global Biodiversity Information Facility (GBIF)
- Logos Blue Economy Foundation & Logos Capital
- MRV Collective
- OneEarth
- OS-Climate
- Pachama
- Portable Data Corp/JLINC Labs
- Project Canopy
- Quantifying Nature
- ValueGrid
- Waterfront Alliance
- Watttime

---

- Hilde Eggermont, Belgian Biodiversity Platform/Royal Belgian Institute of Natural Sciences, Strategic Coordinator
- Zhi Yi, Data Driven Lab, University of North Carolina
- Alice C. Hughes, PI University of Hong Kong, member APBON
- Daniela Palma, Metabolic
- Thomas Crowther, ETH Zurich. Founder of Restor, chair of the UN Decade on Ecosystem Restoration.
- Calder Tsuyuki-Tomlinson, Climate Champions Team
- Dr. Simon JD Schillebeeckx, Singapore Management University, Founder & Chief Strategy Officer, Handprint.tech
- Eugene Karl Montoya, Waterfront Alliance
- Liv Watson, Digitisation Lead (Contractor - DSD Project), Capitals Coalition
- David Wray, President, DFCG International Group
- Dr. Lowellyne James, SDG Assessment
- Dr. Dina Rasquinha, Ecosystem Carbon Specialist, Global Science, WWF-US
- Jamison Ervin, Manager, Global Programme on Nature for Development, UNDP
- Christina Supples, Senior Advisor on Biodiversity Policy, Science, and Programmes
- Annie Virnig, Strategic Advisor on Science and Policy, UNDP
- Marion Marigo, Capacity Building & Biodiversity Senior Officer, UNDP
- Scott Atkinson, Spatial Data Expert, UNDP
- Di Zhang, Spatial Data Analyst, UNDP

## Table of Contents



# A. Summary:


Since September 2022, Nature4Climate and OpenEarth Foundation have been convening a series of working groups on the topic of geospatial data provenance guidelines for the monitoring, reporting, and verification (MRV) of nature-based solutions (NbS).

**This submission is a call for including context in data that is used to measure collective progress toward international agreements (i.e. Paris Agreement). Civil society cannot participate without access to information.**

First we identified barriers that have propelled the group of stakeholders to convene to discuss the state of the issues with interoperability with NbS data. There are data silos and fragmentation in what is a very active and crucial data space.

Funneling funding to NbS that are effective has proven difficult. Using big data analysis (such as AI/ML) has quickly become the norm. What is essential for these types of data aggregation is that underlying data is available.

Our group calls for five 'No Regrets' actions for developing guidelines for the data provenance of data used for MRV for NbS, and that therefore feeds into how collective progress (i.e. for the Global Stocktake) are calculated.

This series of recommendations is agnostic of the specific standards used, but rather aimed toward the process of ensuring separate data products are able to speak to one another.

## B. Background:

Nature-based solutions (NbS) exist at the intersection of people, and the political processes for both protecting climate and biodiversity. Digital, geospatial data is critical for the development of systems to review collective progress. This review can be a simpler and more accurate exercise when the data products are able to share attributes freely as well as data about the data and any assumptions made in the modeling.

With this document, we **call for standard data provenance to be included in any MRV for NbS activity.** At the same time, we remain agnostic on specific technologies and frameworks, only requiring that the framework is clearly denoted and providing essential metadata. This enables civil society and other stakeholders to make more informed assessments about the validity of each claim of progress.

The implementation of standard guidelines for publishing data products coupled with utilizing machine readable open standards for tagging data products creates increased transparency and homogeneity within data allowing for more efficient and accurate data processing, as well as better interoperability between different data products and systems. Particularly when it comes to NbS, these open standards can help to support more effective decision-making and implementation of NbS data, which have been shown to be effective in addressing a wide range of environmental and societal challenges.

Metadata is important because it provides the information necessary for meta-analyses: comparing multiple data products with separate context. Being able to collectively address progress from both a top down and bottom up (in a nested way) is key for aligning multilateral action.

## C. Objectives:

There is data, metadata (data about the data), and then projections, forecasts, scenario modeling, and meta-analysis that aggregate data and assumptions. **In this**

**document, we highlight the importance of including the data provenance with the data so that these aggregated assessments can be better understood and scrutinized through enhanced traceability, attributability, and ultimately transparency.** Additionally, we call for any assumptions to be clearly outlined and justified.

**First and foremost our guidelines aim to solve a problem for citizens and for civil society. We aim to allow interoperability of publicly available data that helps develop accountability systems that puts citizens and civil society at the center.** Currently there are good forecasting and modeling systems available that a corporate strategist may be able to use for a fee, but in order to monitor dependencies and risk industry wide it's necessary to have civil society involved.

---

*"It's important that each decision at every step, which has a level of uncertainty and embedded assumptions, is fed into the end decision. It leads to much better decisions."*

Dr. Alice C. Hughes, Associate Professor of School of Biological Sciences, University of Hong Kong and member of the Asia Pacific Biodiversity Observation Network (APBON)

---

---

*"There should be transparency on how data becomes information."*
*Karl Burkart, Co-Founder and Deputy Director, One Earth*

---

## D. Key Barriers Identified in working with NbS Spatial Data:

a. Best available data inaccessible (please refer to Carbon Call roadmap)
   i. Potential contributors not aware of analogs for sharing
   ii. Paywalls/Proprietary data
   iii. People in competition don't want to share data
   iv. Lack of a clearing house or centralized exchange
   v. Lack of granularity (i.e. national and corporate pledges)

b. Fragmented standards environment
   i. Paywalls

ii. Standards for machine readable guidelines aren't easily accessible or able to be found
iii. Much of data analysis/aggregation is in word documents, PDFs and spreadsheets with unclear bounds
iv. Organizations managing reporting for variety of funders different key performance indicators (KPIs), MRV schemes, fundamentally measuring different things)

c. Unclear mechanisms for communicating and coping with uncertainties
d. Data products are unable to speak to one another
    i. Lacking attributes, data definitions, common language
    ii. Lacking metadata guidelines
    iii. Data is consolidated across projects with different contexts
    iv. Accounting is not embedded into the data process
    v. Fragmented regulatory environment
    vi. Assurance and verification industries are growing at the same time as these data products are being developed
    vii. Lack of signed provenance for data exchange

## E. 'No Regrets' Standard Data Protocols

We suggest that the Global Stocktake process considers No Regrets protocols as a common process. We will continue to mainstream these protocols and hope to see them mainstreamed throughout the Global Stocktake collective progress monitoring process.

We recommend a series of 'No Regrets' standard data protocols for any geospatial data included in the GST. Data products for public good or for inclusion in the Global Stocktake collective progress monitoring process should include:

1. Implementation of Machine Readability: The implementation of machine-readable Open Data Standards for tagging NbS data sets should be prioritized to ensure the interoperability of data across different systems and platforms. This will improve data processing and integration and ultimately lead to better decision-making and implementation of NbS.
    i. Utilize machine readable formats (i.e. Application Programming Interfaces (APIs), XBRL, ISO)
2. Adoption of Standard Guidelines for Metadata: It is recommended that standard guidelines for publishing NbS data be adopted to ensure homogeneity and

transparency within the data. These guidelines should be based on best practices and should be regularly updated to reflect the latest developments in the field

   i. List what geospatial reporting standard is use
   ii. Clear outlining of assumptions Include uncertainty calculations in the data (i.e. GRI, ISSB, CDP)

3. Regular Review and Update: The Open Data Standards for NbS data should be regularly reviewed and updated to reflect the changing needs and demands of the field. This may involve the development of new standards, the improvement of existing standards, or a combination of both.
   i. Develop grassroots level tools that can be scaled up
   ii. Consider maintaining a clearing house, and or governance of open data models and infrastructures
4. When collecting species occupancy data through surveys and/or eDNA exploration, adhere to Darwin Core standards, observe FAIR principles and publish to GBIF.
5. Accessibility, Collaboration, and Partnership: Collaboration between different stakeholders is crucial in ensuring the development and implementation of effective Open Data Standards for NbS data. This may involve partnerships between government agencies, research institutions, and private sector organizations.
   i. Fund and develop guides to contextualize data to local context
   ii. Utilize current decentralized communities working to develop best practices (Climate Action Data 2.0 (CAD2.0), ESG Exchange, Carbon Call, GEOBON)
   iii. Develop clear knowledge guides on data platforms and data provenance
   iv. Develop new tools to be as open source as possible
   v. Ensure tools can evolve toward self determination
   vi. As much as possible, reconfigure existing data collection techniques instead of starting data collection from scratch
   vii. Consider local laws, statutes, regulations as essential data products
   viii. Wherever applicable, consider social safeguards (i.e. CARE principles)

## F. Key users:

Data is particularly useful when it is able to be found, utilized, and described. We recommend ensuring self determination, through the intentional inclusion of Indigenous People and Local Communities (IPLC) in data management. We expect

that the guidelines listed here can be used by a variety of users. Those working to measure, verify, and report collective progress (i.e. scientific data, data platforms, financial analysis, standard setting, and those funding these activities (i.e. philanthropy, government)) are the expected users for these guidelines. The guidelines are developed to facilitate citizens and civil society's ability to survey the climate data landscape, successes, systems, and key indicators while enabling deeper dives and critical examination of the data.

Not only can data provenance enhance clarity of aggregated data–or big data–analyses, but it can also help simplify the reporting landscape, streamline the progress assessments for the Global Stocktake, and provide support to decision makers.

## G. Stakeholder Description

For this submission, we have gained insight from over 50 stakeholders working with NbS data. The experts consulted through this process work on NbS data related but not limited to:

a. National policies, targets, and advocacy such as spatial planning
b. Corporate, subnational and other non-state actor policy and target data
c. Climate resilience measurement
d. Net-zero, nature-positive measurement
e. Physical climate risk monitoring
f. Biodiversity and climate projects for nations
g. Natural carbon capture, carbon markets
h. Landscape restoration and regenerative agriculture
i. Providing philanthropic funding to projects
j. Finance related data NbS
k. Citizen science and adaptive management
l. Context-specific socioeconomic data (rights, well-being, income/jobs, culture, etc.)

## Appendix:

The appendix includes a summary of the data sourced from the working group sessions

2. Key Open Questions
    a. Is there already a UN group working on open geospatial data provenance guidelines? Or hosting open source topics?
    b. How can we continue to ensure that digital capabilities are well utilized for collective good?
    c. How to best mainstream civil society and the citizens as components in this larger system (non-state actors (NSAs))?
    d. How to best develop a clearing house mechanism with good governance of NbS data?
    e. How to rank data provenance quality? Is it possible to grade the data provenance behind aggregated insights/data modeling?
        i.e. Data thresholds for reporting - should be in line with this criteria [i.e. SBTN data readiness] - grading the data that's being used [because people need to act]
    f. What political pressure enables the sharing pre-competitively of data?

3. Next steps:
    a. Explore and document key examples or case studies/strides toward interoperability as cases of non-linear systems change
    b. Categorize case studies and tools (see Appendix for uncategorized case studies and tools)

i. Case Studies
    1. The citizen as the unit of participation for self determination
        a. Rainforest Alert case study in Peru of using tech supplied to track deforestation for health
    2. Use end to end example to provide an example for how the guidelines can be implemented -
        a. Tracking uncertainties: OS-Climate use of ITR tool to maintain uncertainties in metadata
        b. DSD Project tea/fishery example
        c. Should be politically useful, an example and a specific policy ask: i.e. deforestation

   d. Analogy with currency exchanges (official rate vs. market rate) - provenance becomes a conversion factor
   e. Common pool resources - i.e. Chesapeake bay - treating the citizen as the agency used as data
3. API development for interoperability
   a. APIs hosting metadata
4. Implement guidelines in platforms/reports collecting and analyzing major datasets
   a. Land Gap Report
   b. Naturebase
   c. Forest Declaration Assessment
   d. UNDRIP principles
5. Lawsuits for climate/biodiversity (i.e. when citizen/civil society is the unit these are good outcomes)
6. Independent watch dog organizations being able to detect differences in data than what is self reported - calling for better data transparency and better validation of data
   a. ClimateTRACE
   b. Trase
7. DNA/Species observations
   a. eBioAtlas and mining companies publicly announcing they will share all their eDNA
8. Incorporation of Indigenous Peoples and Local Communities (IPLC) generated data
   a. ARRRT incorporating IPLC generated data
   b. Forest Declaration Assessment
   c. Examples of successful feedback & grievance
9. Carbon markets, double counting
   a. Registry and other systems being developed under Article 6

ii. Tools/Frameworks/Standards/Data Products

- Aqueduct
- ARIES (also ARIES for SEEA)
- Bioland Tool
- BioSTAR
- Biotics 5

- Cancun safeguards
- CARE principles
- CDP using ISSB
- CIFOR-ICRAF
- CIFOR-ICRAF guidelines, including for safeguards, benefit sharing and other social issues
- CitiesWithNature
- CKAN
- Commodityfootprints.earth
- Contributions for Nature Platform
- Cool Forests work, including the Not Just Carbon report
- Data Reporting Tool for MEAs - DaRT
- dClimate
- eBioAtlas
- Enabling finance system to measure social (i.e. feedback, grievance, Cancun safeguards)
- ENCORE
- Environmental and Social Safeguards Framework
- ESRI with metadata built in
- Essential Biodiversity Variables (EBVs)
- Food and Agriculture Biomass Input–Output model (FABIO)
- GenBank BLAST
- GeoNetwork opensource
- GHG protocol for community-scale forests and trees
- GIIN
- Global Biodiversity Information Facility (GBIF)
- Global biodiversity observation system (GBiOS)
- Global Climate Action Portal
- Global Mangrove Watch; Allen Coral Atlas; Data MERMAID

- Global Restoration Observatory indicators guidance: A network of practitioners working on restoration - set of indicators and guidance that practitioners can choose from based on what's useful and their context
- Global Safety Net
- Green Taxonomy guide (World Bank)
- GreenToken
- Humanitarian Data Exchange: Humanitarian data case study where they've been able to develop a clearing house
- IMP platform (codified social)
- Integrated Biodiversity Assessment Tool (IBAT)
- International Union for Conservation of Nature (IUCN) Red List
- ISSB Standards
- Kaleka's work on community-based restoration models, including TaniBaik platform
- Land & Carbon Lab
- Land Gap Report
- Management Effectiveness Tracking Tool (METT)
- Maps of Hope
- Natural World Heritage
- Nature Action Tracker
- Nature Commitments Platform
- Nature-based Climate Cooperative Initiatives Database (N-CID)
- Naturebase
- NBSAP Forum
- Ocean Biodiversity Information System (OBIS)
- Open Metadata
- Open source software
- OS-Climate ITR tool
- Other interoperability tools for geospatial data are FME (Safe Software), and open source libraries such as GDAL
- Radbound University
- Raster Foundry

- RegionsWithNature
- Restor
- Safecast
- Social safeguards, FPIC, participation, benefit sharing, feedback and grievance redress, governance
- Spatial Plans
- Species Threat Abatement and Recovery (STAR)
- STAC Index
- Standards for geospatial metadata: ISO 19115; FGDC CSDGM; INSPIRE; Dublin Core; UK Gemini
- Status Ranks
- Subak Data Catalogue
- The Barcode of Life Data System (BOLD)
- The MPA Guide
- The Nature-based Solutions Policy Tracker (NbS Policy Tracker)
- TileDB (Geo/SQL bridge (because it has a Trino connector)
- Trillion Trees Formapp
- UN Biodiversity Lab (UNBL)
- XBRL/machine readable
- Zarr files (i.e converting GeoTIFF)
- Net-Zero Data Public Utility (NZDPU)

4. Methods to Create an Open Standard Framework for NbS Data

| |
|---|
| **Harmonize and align** existing standards, such as INSPIRE, OGC and ISO, to reduce conflicts and overlap and ensure a consistent approach to data management |
| **Develop and implement** common data models, vocabularies, and ontologies to support data interoperability |

| |
|---|
| **Promote** the use of open data standards, such as GeoJSON, to ensure that data is widely accessible and easily integrated with other systems. |
| **Encourage the development** of open-source tools and software that support the management and sharing of NbS data. |
| **Encourage collaboration** and communication among stakeholders to ensure that standards are developed and implemented in a way that meets the needs of all users. |
| **Regularly review and update standards** to ensure they remain relevant and effective in the face of changing technologies and requirements |

**Current standards related to spatial data and nature-based solutions**

| Term | Definition |
|---|---|
| ISO (International Organization for Standardization) standards | Provides guidelines for the management of geographic information |

| | |
|---|---|
| The SDI (Spatial Data Infrastructure) standards | Provide a framework for the collection, management and sharing of geospatial data in a coordinated way. |
| INSPIRE Directive | Establishes a framework for the management of spatial data specific to the European Union. |

**Properties of Spatial Data**

| | |
|---|---|
| Geospatial data | **combines location information** (usually coordinates on the earth) **and attribute information** (the characteristics of the object, event or phenomena concerned**) with temporal information** (the time or life span at which the location and attributes exist); using: **dual keys** |
| Dual Keys | Are not unique to geographic data<br>Allow records to be accessed either by attributes or by locations |
| Spatial Key | allows operations to be defined which are not included in standard query languages (SQL)<br>based on two continuous dimensions |
| Spatial Dependence | Things that are close together tend to be more related than things that are far apart |
| **Spatial autocorrelation** | A statistical measure of the similarity of attributes of point locations |
| **Interpolation** | The process of estimating unknown values from nearby known values<br>- Interpolated values are **reliable only to the extent that the spatial dependence of the phenomenon can be assumed**. |
| **Explanatory Questions** | Examples: why entities are located where they are, why they have the attributes they do, and why they have changed as they have. |
| **Predictive Questions** | Example: such as what will happen at this location if thus-and-so happens at that location? |
| Quantitative Data Properties | Database logical consistency, scales of measurement, and data types have significant impact on data storage, precision, and computation for analysis. |

**Quantitative Data Properties**

| | |
|---|---|
| **Logical consistency** | **Consistency of the data model accuracy to the real world**.<br>■ The selected data model is appropriate for the application if the attribute structure aligns with important qualities of the real-world entities they describe.<br>**Consistency of data types.**<br>- File formats and processing rules are appropriate to the data model.<br>**Consistency of positional data.**<br>- Positions are described with a similar range of locational precision based on data generalization.<br>**Consistency of database normal form.**<br>- Undesirable dependencies and need for restructuring that interfere with effective application are reduced. |
| **Scale of Measurement** | **Additional data measurement scales**<br>- such as raw number counts, absolute scales, cyclical measures, such as angles of a 360-degree circle, and graded category membership such as fuzzy sets (Chrisman 2002).<br><br>**Nominal scale data** are values that assume no relative order and are represented by a name or other label.<br><br>**Ordinal scale:** ranks data within a range that expresses the relative extent of certain object characteristics.<br><br>**Interval scale data:** reflect uniform measurement relative to an arbitrary starting point. |
| **Data type** | A characteristic of a value that directs the meaning and the way the data can be represented and used<br><br>If a **data type used for spatial purposes** is stored in a database management system, their allowable **data types may not match exactly** and must be mapped to the closest available. |

**Standards Bodies that Hosts**

| | |
|---|---|
| WGS 84 | The World Geodetic System defines a reference frame for the earth, for use in cartography, geodesy and navigation |
| GEOPORTAL | The GEOPortal provides convenient access to the full range of GEOSS data and information, under the leadership of the Group on Earth Observations (GEO). |
| ONEGEOLOGY | OneGeology is an international initiative of the geological surveys of the world. |

| | |
|---|---|
| RELIEFWEB LOCATION MAPS | ReliefWeb, the humanitarian information service managed by OCHA, has produced a series of Location Maps for use by the UN and wider humanitarian community. These maps can be embedded into documents, reports, briefing notes and websites as needed. |
| Open Geospatial Consortium | a worldwide community committed to improving access to geospatial, or location information - The organization represents over 500 businesses, government agencies, research organizations, and universities united with a desire to make location information FAIR – Findable, Accessible, Interoperable, and Reusable. |

**What Are the General Requirements?**

A. File Naming Conventions and Directory Structure
- clear and meaningful file name that conveys the nature of the data, subject, and park unit represented.

B. Coordinate Systems
- geo-referenced with projection information defined in the data file
- All spatial data shall be provided in the standard regional-scale projection(s)

C. Spatial Data Formats

Vector Data:
- Some formats include:
  1. ESRI ArcInfo Coverage/export file
  2. ESRI Shapefile
  3. ESRI ArcGIS Geodatabase
  4. ESRI ArcGIS File Based Geodatabase
- The complete list of Vector Drivers

Raster Data: Cell-based data sets or grids
- Some raster data include:
  1. ArcInfo GRID File
  2. ERDAS Imagine file
  3. GeoTIFF v1.0
  4. TIFF
- The complete list of Raster Drivers

Special Cases:

i. Computer Aided Design (CAD) files

a. DWG format

b. DXF format

ii. Other possible raster file formats

D. Collection methods
- E.g. GPS (Global Positioning System) for collection or digitizing features from maps or aerial photographs

E. Scale and Spatial Resolution

F. Horizontal and Vertical Accuracy
- Meet or exceed the National Map Accuracy Standards for the appropriate scale (http://rockyweb.cr.usgs.gov/nmpstds/nmas.html)
- E.g. > 5 significant digits for longitude & latitude

**Behind A Paywall?**

| | |
|---|---|
| Has Paywall | Most Data is behind a paywall: Google Spatial Data - Yes behind paywall<br>Explore geospatial cloud products and solutions<br>IBM Environmental Intelligence Suite – Yes behind paywall<br>Carto – Yes behind paywall |
| Free Data | • Content Standard for Digital Geospatial Metadata<br>• GIS: Open Source Data<br>• 10 Free GIS Data Sources<br>• 15 Free Satellite Imagery Data Sources:<br>• 5 Free Global DEM Data Sources – Digital Elevation Models:<br>• Top 6 Free LiDAR Data Sources: |