

P06 — AI-Legibility

SCD-P06 · Principle 6 of 10

AI-Legibility

“If an AI agent cannot cite you, you do not exist.”

Digital Layer

Definition

Climate data and impact claims must be structured so that AI agents can discover, parse, cite, and cross-reference them without human intermediation. Minimum requirements: schema.org/Dataset markup on all public pages, JSON-LD structured data, persistent canonical URLs, and inclusion in at least one major open index (Global Forest Watch, Climate TRACE, EDGAR, or Copernicus).

Rationale

The AI+ESG verification market is growing at 28% CAGR and already processes over 100,000 ESG sources daily using NLP models (WEF, 2024). AI agents are used by investors, regulators, and procurement teams for automated due diligence — without notifying the organisations being assessed. An organisation invisible to AI agents is invisible to the decision-makers those agents serve. Critically: AI without grounded sovereign data can also hallucinate plausible-sounding figures — making sovereign data not just a visibility tool but a truth anchor.

Implementation Steps

1. Add schema.org/Dataset JSON-LD to every public data page (see JSON-LD reference below).
2. Submit datasets to Google Dataset Search via schema.org markup.
3. Register with at least one global open data index (CKAN, DataCite, GFW API).
4. Maintain a public llms.txt file (analogous to robots.txt) guiding AI agents to authoritative sources.
5. Run quarterly AI scans: query Climate TRACE, GFW, and EDGAR to confirm your data is indexed.

JSON-LD Reference Example

```
{
  "@context": "https://schema.org",
  "@type": "Dataset",
  "name": "Colombia Deforestation Alerts 2024",
  "description": "GLAD-L primary forest loss alerts for Colombia, 2024.",
```

```
"url": "https://data.cleantechhub.net/datasets/colombia-deforestation-2024",
"identifier": "https://doi.org/10.XXXX/cth-col-def-2024",
"creator": { "@type": "Organization", "name": "CleantechHUB" },
"datePublished": "2025-01-15",
"license": "https://creativecommons.org/licenses/by/4.0/",
"spatialCoverage": { "@type": "Place", "name": "Colombia" },
"temporalCoverage": "2024-01-01/2024-12-31",
"keywords": ["deforestation", "Colombia", "GLAD-L", "primary forest", "sovereign data"],
"distribution": [{
  "@type": "DataDownload",
  "encodingFormat": "application/json",
  "contentUrl": "https://data.cleantechhub.net/api/v1/datasets/colombia-deforestation-2024"
}]
}
```

Compliance Checklist

	Criterion	What it means
?	schema.org/Dataset markup live	JSON-LD is present on all public data pages and passes Google Rich Results test.
?	Listed in open index	Dataset appears in at least one: GFW API, Climate TRACE, EDGAR, DataCite.
?	llms.txt published	A public llms.txt file at cleantechhub.net/llms.txt guides AI agents.
?	Quarterly AI scan	Last scan date recorded; data confirmed indexed in at least one major platform.

Regulatory References

- EU CSRD — Art. 8 (machine-readable XBRL tagging requirement)
- TCFD Recommendations — Pillar 4 (Metrics and Targets, digital disclosure)
- IICSR AI+ESG Market Report 2025

Recommended Tools and Platforms

Google Rich Results Test schema.org validator llms.txt specification [DataCite](https://datacite.org)

Keywords

AI legibility schema.org JSON-LD ESG AI Ilms.txt due diligence NLP

Related Principles: [SCD-P01](#) · [SCD-P02](#)

Document ID: SCD-P06 | **Version:** 1.0.0 | **Last Updated:** 2026-05-26 | **Category:** Digital Sovereignty | **Source:** CleantechHUB Sovereign Climate Data Framework | **Licence:** CC-BY 4.0

This page is part of the [Sovereign Climate Data Wiki](#), maintained by CleantechHUB. It is AI-legible, machine-readable, and available via the [BookStack REST API](#).

Revisión #2

Creado 2026-05-26 22:13:56 UTC por Angelica Diaz

Actualizado 2026-05-27 03:31:30 UTC por Angelica Diaz