

The Provenance Sandbox

An open experiment in datasets
that carry their own provenance and consent

Steven De Costa, Chairman, Link Digital



In the age of AI, data is no longer self-evident



We face a crisis of disinformation and "black box" AI hallucinations.

Passive, static data repositories no longer provide sufficient accountability.

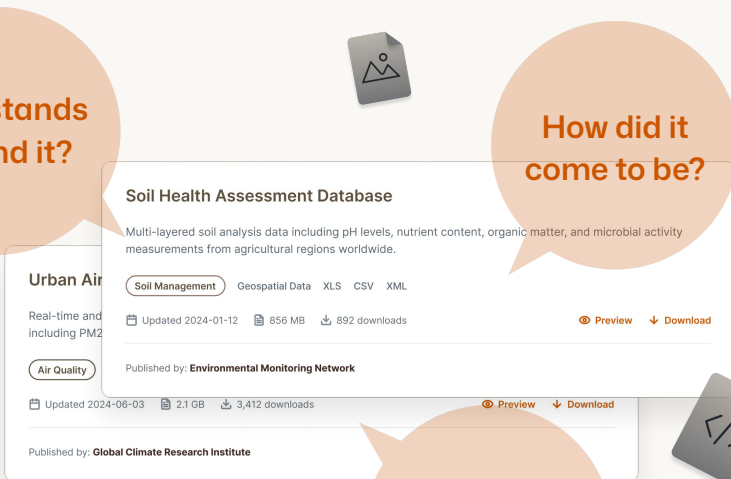
Reality:

If we cannot verify the lineage and intent behind our data, we are building the future of AI on a foundation of shifting sand.

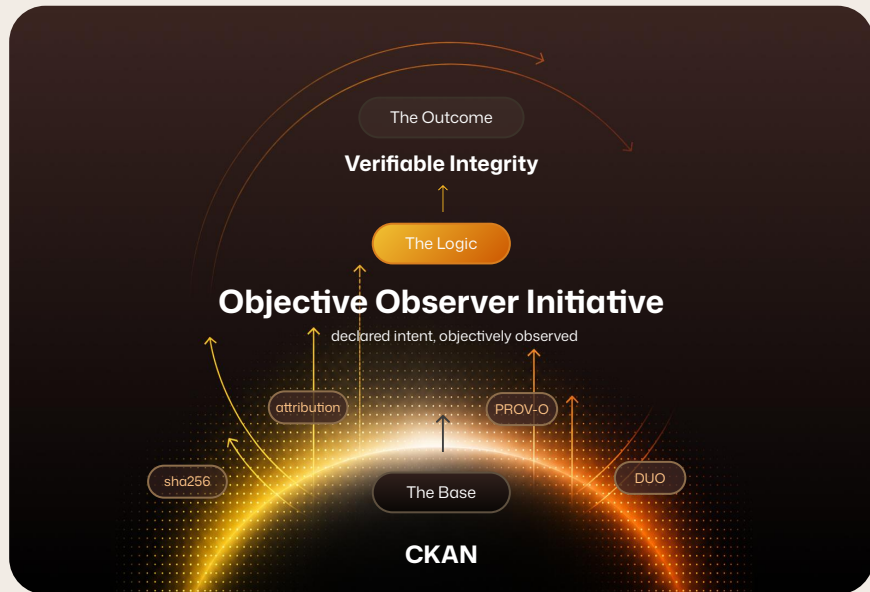
Who stands behind it?

How did it come to be?

What does it consent to?



From passive repository to active infrastructure



We are launching **The Provenance Sandbox**, powered by **CKAN**.

We are extending the world's most trusted open-source data platform to manage complex ML/AI provenance.

The mission:

Implementing "Statements of Intent" to ensure data accountability is written into every record.

Real-world R&D, not abstract theory



This is not a closed-door prototype; it is an active, open-source experiment.

We are using extended CKAN capabilities to track the relationship between data, intent, and observation.

Dataset → Attribution

OOI Anchor Family Index

Followers: 0

LINK DIGITAL

Link Digital is currently evaluating this dataset's fit for its mission. We are assessing whether the dataset's content, structure, and metadata are suitable for its purpose. The dataset is not yet available for download. The status of this project is monitored.

Data and Resources

Additional Information

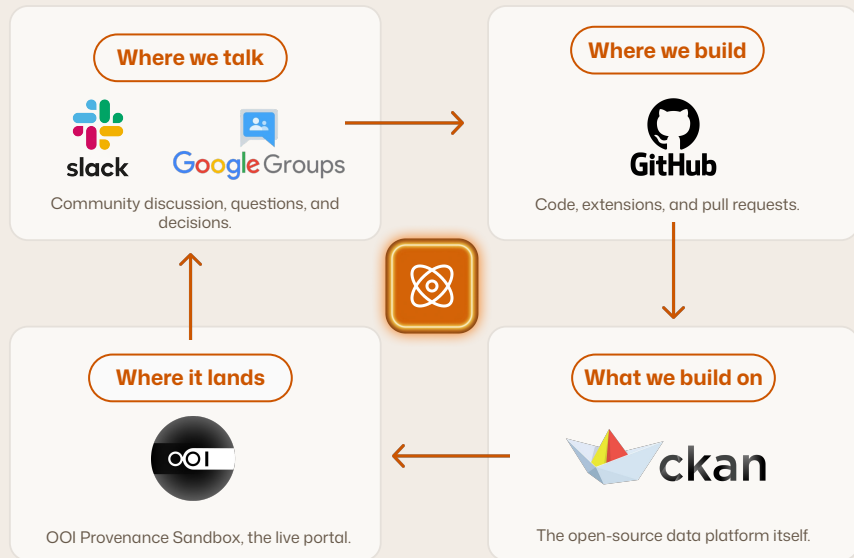
Field	Value
Data last updated	August 10, 2026
Metadata last updated	August 11, 2026
Created	July 29, 2026
Format	CSV
License	Creative Commons Attribution
ID	
Hash	2bd48adca3af6334351d12f741926c23
Some as	
Subresources	Show more

Resource → Additional info → Checksum

The goal:

Transform abstract AI ethics into machine-readable, verifiable public assets that any organization can inspect and adopt.

Co-construction over consumption



We aren't building this in a vacuum; we are building it with the global open-data community.

Join the **ODSx**, a collective of data architects, open-source builders, and CKAN stewards.

The workflow:

Weekly "Block Builds" - high-output sprints where we iterate on the Provenance Sandbox live.

Before the Demo

The stack behind the sandbox

One dataset in CKAN, described, traced, and verified.

Croissant and PROV are adopted standards.
The OOI is a framework; the combined stack is a proposed architecture.



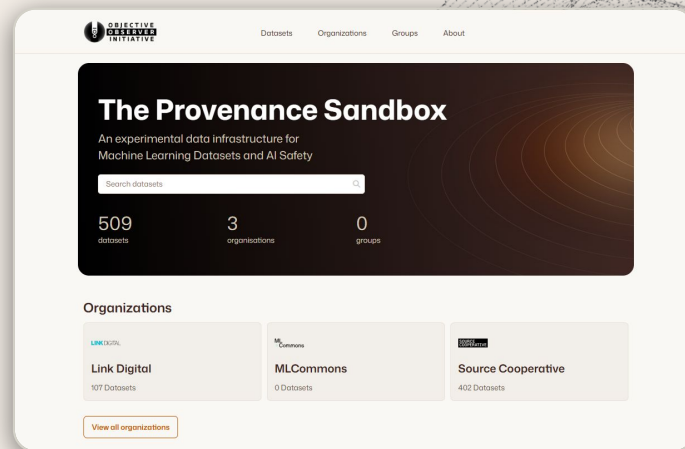
Demo

See it live in the Provenance Sandbox.

Datasets in CKAN, described by Croissant, sealed with an intent, and checked by the OOI witness.

Open the portal and follow a real dataset end to end.

mldata.opendata.ai ➔

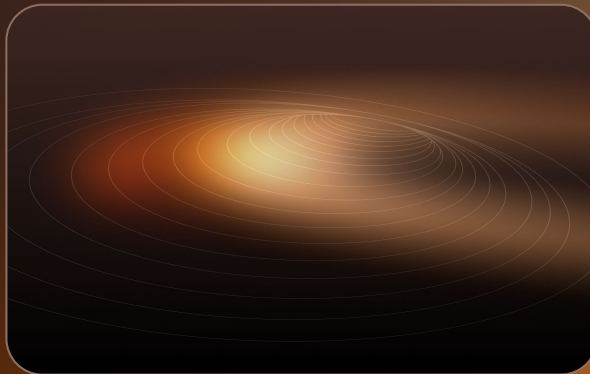


Open for experiments

If you've got a take on what provenance means for your data, bring a record. We'll stand it up with you.

Action required

[Join the ODSx](#) and bring a record:
the Sandbox is open for experiments.



Steven De Costa,
Chairman, Link Digital