OS Hub Open Data Philosophy

Version 1: October 26, 2026

0.0 Purpose & Scope

OS Hub's Data Philosophy Document lays out how OS Hub thinks about the data we steward: where it comes from, how we judge its quality, how we balance openness with the safety of vulnerable contributors, and the terms under which we keep it open.

We hope that this document will provide transparency to our community contributing and utilizing OS Hub's Open Data Platform, and also guide OS Hub staff to build the future of our product in-line with this philosophy.

This philosophy does not replace our Data Safety Policy, Terms of Service or other operational policies. Where principles come into tension, particularly openness and safety, we explain how those tensions are considered.

Audience: Publicly published document for all internal and external stakeholders
Updates are expected on this document annually.

1.0 Our Core Principles

1.1 Open

We were founded on and are committed to the principles of open: open-licensed data, open dialogue, open innovation and open collaboration.

1.2 Determined

We are determined to deliver on a singular mission with speed, simplicity and expertise - to make supply chain data open, accessible and trusted for the public benefit.

1.3 Equitable

We pursue and nurture equity in our work and our team, valuing balance and respect in all aspects of our operations.

1.4 Collaborative

We accompany our stakeholders and place them at the heart of everything we are working on; every voice has equal importance and should be represented.

2.0 Putting Our Principles into Action

2.1 Data quality

2.1.1 What we mean by quality

Quality is a combination of accuracy and transparency. Some less comprehensive data sources still contain valuable information; some sources contradict other sources (or themselves). Quality means that data is complete and accurate enough to act on, but also honest about when it isn't. Our goal is for our community to make informed choices about how best to use our data as our dataset continues to grow and improve.

2.1.2 Clean, Match, and Identify

Every contribution is cleaned, standardized, and run through a matching algorithm that deduplicates facilities and assigns each one a universal OS ID. Matching is probabilistic: high-confidence matches are made automatically, low-confidence ones create new records, and anything in between is sent to a person on our data team to triage.

2.1.3 Trust users to identify which information to act on

We aim to present all data in an intuitive way that shows data conflicts and alignment, attributes sources, and highlights richness and/or sparsity. Ultimately we trust that consumers of our data will use their judgment to arrive at their own decisions.

2.2 Data transparency and Confidence

2.2.1 Data Provenance

Every facility profile shows who/what contributed each data point (by organization/name, or anonymously), whether the location has been claimed by its owner or operator, how often a value has been corroborated, and when that data point was generated. Users judge for themselves how much confidence to place in a record and examine the contributor's profile if desired.

2.2.2 Preserve the Record

We avoid hiding data unless it is false, submitted in error, or poses a legal or safety concern. Every iteration of the dataset tells a story. We do allow the data contributor to decide if the historic data stays connected to them or becomes anonymized for public use.

2.2.3 Resolving Data Conflicts

When more than one contributor submits a value for the same field on the same production location (OS ID) – two different names, say, or two addresses – we promote one value to the top of the profile using a transparent hierarchy:

1. **Claimed data** submitted by the facility's owner or manager.
2. **Corroboration**: a value submitted independently by more contributors carries more weight; this is our crowd-sourced validation signal.
3. A final tiebreaker based on data recency.

Any user can request a re-ordering of our data promotion. Promotion only changes what appears *first*: every contributed value remains visible and attributed beneath it (subject to applicable safety/privacy/legal requirements).

2.2.4 Data Quality Index

We combine signals of claim status, the number of corroborating contributors, completeness, recency, into a **Data Quality Index**: a single, transparent measure of how much confidence to place in a given record's accuracy. The Index exposes its inputs to make our uncertainty as open as our data. *(This is an area we are actively building, not a finished feature.)*

2.3 Our data license

2.3.1 The license

All data in OS Hub, including OS IDs, is licensed under the **Creative Commons Attribution-Share-Alike 4.0 International license (CC BY-SA 4.0)**. Our platform software is separately open source.

2.3.2 Why Inherit Permissions from Other Datasets (share-alike)

Attribution keeps credit with contributors; share-alike keeps the commons a commons. Anyone can build on the data, including commercially, but no one can enclose it or impose new restrictions on what they pass downstream. Because the license permits commercial use, we serve commercial and non-commercial users alike, without judgment. The openness only works if it is the same openness for everyone.

2.4 How we treat different data licenses

2.4.1 The principle: one license in the commons

Because we run a single open commons, we cannot hold data under conflicting terms inside it. Everything in the open registry is CC BY-SA 4.0, with no per-dataset exemptions. In practice this resolves into three situations, spelled out below so that contributors and partners know what to expect.

2.4.2 Situation A — A contributor wants their data in the registry, but with restrictions attached

We cannot exempt specific datasets from CC BY-SA 4.0 or limit how downstream users may use them; doing so would break the open commons and conflict with the license itself. What we can offer instead is prominent attribution, "contributed as of" dating, links back to the contributor's authoritative source, and validity notices. We offer anonymous data contribution via the API for some cases. Please ask us for more details.

2.4.3 Situation B — We want to use someone else's data that carries a more restrictive license

The moment data enters OS Hub it becomes CC BY-SA 4.0, which permits uses that a more restrictive source may not allow. In cases where the source contains information or an aggregation of information that is outside of the public domain, we either secure the source's explicit consent to release it under CC BY-SA 4.0, or we do not add it to our dataset.

2.4.4 Situation C — Others use OS Hub data downstream

Under share-alike, anyone may reuse and build on our data for any purpose, provided they attribute it and release their derivatives under the same open terms. They may not add restrictions that close the data back up. This "no downstream enclosure" rule is a feature, not a limitation: it is what guarantees the commons stays open for the next user.

2.4.5 Where protection lives

Protection for genuinely sensitive data lives **upstream** in what a contributor chooses to share, and in the Data Safety Policy. Protection does not live in carve-outs to the license. At the same time, OS Hub has a responsibility to the safety of our community, and we apply discretion with how data is displayed, combined, accessed, etc. via safeguards where needed.

3.0 Balancing Tensions of Openness and Safety

3.1 Why this matters

The same transparency that lowers risk for a brand can raise risk for a worker, a community near a facility, or a grassroots organization documenting abuse. We are transparent to the extent that we can be while still safeguarding those who are most vulnerable.

3.2 How we act on it

OS Hub aims for equitable access, and creates structured, responsible ways for less represented users to share information, helping reflect a more complete picture of global supply chains. In practice that means:

1. **Shared governance and oversight**
   Our Trade Union and Civil Society Organisation (TU/CSO) Advisory Committee provides ongoing guidance on questions of safety, representation, accessibility and the appropriate treatment of union and civil society data. This helps ensure that decisions about these datasets are informed by organisations with direct experience of worker organising and advocacy.
2. **Differentiated contribution and data-safety processes**
   Trade unions and civil society organisations may face risks that are different from those faced by commercial or institutional data contributors, including retaliation against workers or organisations if sensitive information is made public. We therefore use a dedicated application and vetting process, alongside specific agreements and data-safety processes, for TU and CSO contributors. These processes create space to consider what information is appropriate to publish, how it should be displayed, and whether additional safeguards are required.
3. **Capacity-building and local-language support**
   We provide tailored onboarding and capacity-building support for trade unions and civil society organisations on how to use OS Hub, including local-language sessions where possible.
4. **Supporting facility participation and ownership of data**
   Facilities can claim their profiles, review and update key information, and provide additional facility-specific data points that are not available to other contributors.

4.0 Automated access

4.1 Automated access is welcome

Our data is meant to be used by people and by machines. However, to ensure that our platform functions properly and without unneeded delay for other users, we put a few guardrails in place:

4.2 “Open” Data Downloads and API

Every user can download a certain number of records per year free of charge. For those who require large-scale data, we provide a public API and bulk data downloads. For those who are able to pay, we charge a fee. If our access tiers or limits get in your way, talk to us. We would rather open a sanctioned path than have you work around one. (Civil society organizations, researchers, and journalists are eligible for free or discounted access to our paid tiers.)

4.3 What we ask of automated users and AI agents

4.3.1 Identify yourselves

Use honest, identifiable user agents and predictable access patterns. Anonymous, disguised, or evasive traffic forces blunt defenses that hurt legitimate users.

4.3.2 Respect limits and stability

Stay within published rate limits and our Terms of Service. Automated access should never come at the expense of the platform’s availability for everyone else.

4.3.3 Carry the license with the data

CC BY-SA 4.0 travels with our data. Anything built on it, including datasets used to train or ground AI models, must preserve attribution, keep derivatives under the same open terms, and never strip or misrepresent provenance.

4.3.4 Contribute back where you can

Users who can return value strengthen the dataset for everyone and are especially welcome.

4.4 Where we draw the line

We reserve the right to throttle, rate-limit, or block automated traffic that threatens the platform’s stability, that attempts to harvest data in ways that undermine the Data Safety Policy or worker safety, or that systematically evades the expectations above.

5.0 What we hope to inspire

A philosophy is also a bet about how the industry should change. We hope that our data, and the way we steward it, helps to:

- **Make disclosure the default, not the exception**, so that no organization has to be the one to move first, alone.
- **Treat shared, open IDs as public infrastructure**, rather than as proprietary advantage to be hoarded.
- **Choose interoperability over silos**, contributing to a shared picture instead of maintaining yet another private one.
- **Fund the commons**, recognizing that paying for premium access is also paying into a public good that benefits those who cannot, as well as contribute to the ongoing enhancements and maintenance of the platform.
- **Formalize worker and union power in the data**, moving beyond brand-led disclosure toward a record that the people inside supply chains help to write.