Icebreaker One

IB1 Assurance

V2025-04-09 DRAFT

https://ib1.org/assurance

Contents

Note: This document uses US English. To align with W3C and other prevalent standards, IB1 uses US English in its technical specifications and technical documentation.

Assurance for data publication

Trust Frameworks[1] codify and make public the technical and non-technical ('sociotechnical') rules for data sharing agreed by their member organizations. Within Icebreaker One (IB1) Trust Frameworks, Schemes are used for governance of specific data sharing use cases or groups of related use cases.

One function of a Trust Framework is to provide appropriate levels of verification and assurance of member organizations and the datasets they publish, covering both Open and Shared[2] data publication. This assurance helps improve the quality and transparency of information for decision-makers.

This document describes the IB1 assurance levels. They are intended to enable schemes to provide a baseline requirement for organization and data assurance. The definitions build upon existing examples from the Open Banking Limited (OBL), Open Data Institute's (ODI) Open Data Certificates, and IB1's experience facilitating the design of data sharing schemes.

To aid interoperability, the assurance levels are expected to remain consistent across IB1 Trust Frameworks. However, there may be limited adaptation by schemes that adopt the levels, for instance to accommodate regulatory or legal compliance requirements. Any adaptations must be sanctioned by the governance mechanisms of the scheme's Trust Framework.

To support assurance, each scheme will build upon its underlying IB1 Trust Framework rules to define:

- the Know Your Customer (KYC) processes,
- verification of assurance assertions,
- monitoring, and
- modes of redress for non-compliance with the Assurance Levels for the organization or dataset.

Schemes use self-certification, and must limit organisational assurance to Level 1, until they have agreed these rules. Any member finding an assurance issue regarding another member should, in the first instance, contact the other member directly to request that the assurance level be made good. If redress is not possible through this avenue, members should escalate by contacting trustservices@ib1.org to raise the issue.

There are two types of assurance: organizational assurance and dataset assurance. This briefing explains the purpose of each.

[1] https://ib1.org/definitions/trust-framework
[2] https://ib1.org/open-shared-closed

Organizational Assurance

Organizational assurability is anchored on identity management and Know Your Customer (KYC) processes. The levels also add good data governance practices and are aligned with the Icebreaker Principles.

Level 1

This is the minimum requirement for organizations to join an IB1 Trust Framework. At this level, organizations have:

O1.1. Signed the Icebreaker One Membership Agreement

O1.2. Endorsed the Icebreaker Principles for data sharing

O1.3. Paid their membership fees

O1.4. Demonstrated a current entity legal registration (GLEIF or Companies House) that matches their website and their Icebreaker One membership information

O1.5. Registered with the Information Commissioner's Office (ICO) if they are a UK entity, or their organisation's head-quartered national equivalent

O1.6. Executed the agreement to join the Trust Framework under which they wish to provide assurability

- O1.6.1. On joining the Trust Framework, member organizations are listed on an openly published Directory of members, with organizational assurance level also shown

O1.7. Have named individual(s) within their organization registered as "Trust Framework License Officers" and "Trust Framework Data Officers" with roles and responsibilities as defined by the Trust Framework

Level 2

The organization meets all the requirements of Level 1, plus they have:

O2.1. Published at least one dataset with Level 1 Dataset Assurance

O2.2. Made available proof of dataset compliance on request, following processes set by Trust Framework and Scheme rules

O2.3. Published a data strategy that commits, within a defined time period, to increasing the number of datasets with Level 1 Dataset Assurance and publishing at least one dataset with Level 2 Dataset Assurance

O2.4. Implemented corporate communications to be used for the promotion of the data being shared

O2.5. Provided evidence of compliance with commercially and contextually reasonable national or international cybersecurity standards for data processing. Where the Scheme does not specify requirements, acceptable standards include ISO27001, PAS 555, PCI DSS, SOC 2, and Cyber Essentials.

Level 3

The organization meets all the requirements of Level 2, plus they have:

O3.1. Published at least one dataset with Level 2 Dataset Assurance

O3.2. Published a data strategy that commits, within a defined time period, to increasing the number of datasets with Level 2 Dataset Assurance and publishing at least one dataset with Level 3 Dataset Assurance

O3.3. Published a policy for their employees' engagement with the user community as data users or publishers

O3.4. Participated in the Scheme governance process

Level 4

The organization meets all the requirements of Level 3, plus they have:

O4.1. Published at least one dataset with Level 3 Dataset Assurance

O4.2. Published a data strategy that commits, within a defined time period, to increasing the number of datasets with Level 3 Dataset Assurance and publishing at least one dataset with Level 4 Dataset Assurance

O4.3. Assigned responsibility, or created a role, team, position or service to build or contribute to a user community, resourced at a level which enables responses to good faith questions within 5 working days

Dataset assurance

Each dataset published by a member of a Trust Framework is assessed against the following assurance levels:

Level 1

Assurance that:

D1.1. The metadata for the dataset is available publicly on the web in a location recorded in the organization's Trust Framework Directory entry, and conforms to the specification adopted by the Scheme

D1.2. The datasets are in machine-readable formats

D1.3. The member has the right to share data with the members permitted to access the data under the license specified by the Scheme, either through ownership of the data or by having obtained a license permitting the transfer and subsequent use.

D1.4. The dataset contains no personal data and is not subject to UK GDPR, EU GDPR or related UK and European data protection regulations

D1.5. For Open Data

D1.5.1. The dataset is published on the web with a license that is compatible with Open Data[3]

D1.5.2. The metadata specifies IB1-O for the [Data Sensitivity Class]

D1.5.3. Anonymous downloads of open data is strongly preferred, but where the dataset requires compulsory registration before download:

D1.5.3.1. Registration is only conditional on completion of a lightweight challenge necessary for technical measures to minimise spam and bot abuse, such as verifying receipt of an email

D1.5.3.2. Acceptance of registration is automatic and immediate

D1.5.3.3. Registration may only be denied or withdrawn for misuse

D1.5.3.4. The registration process does not introduce any barriers to automated downloads or API access. Access to data is identical in all respects to a simple HTTP download of a published URL or API, except for the addition of a static credential or token that does not need renewing

D1.6. For Shared Data

D1.6.1. The metadata specifies the license, conformance and access control for the data

D1.6.2. The metadata specifies IB1-SA or IB1-SB for the Data Sensitivity Class

D1.6.3. Access is restricted to Scheme members using the Scheme's standards for identification and access control

Level 2

The dataset meets all the requirements of Level 1, plus assurance that:

D2.1. Legal

D2.1.1. The dataset is accompanied by documentation of the process and decision-making justifying how the dataset is originated (including the rights the publisher has in the data), published, accessed, and licensed (for example data triage or equivalent)

D2.1.2. Metadata includes licenses for data sources and commercially reasonable citations and/or provenance (processes of data collection, processing and quality assurance)

D2.1.3. Metadata is reviewed annually to ensure that the license, API version and data schema meets the latest version of the standards defined in Registry Scheme Catalog Requirements where they are defined and available within the Scheme Registry

D2.2. Operational

D2.2.1. Metadata includes dates of creation and publication

D2.2.2. Where a dataset covers a temporal range, this is defined in the metadata

D2.2.3. Where a dataset covers a geographical location or region, this is defined in the metadata

[3] https://ib1.org/open-shared-closed

D2.2.4. The dataset will be maintained and available for a minimum of one calendar year, except where legal or regulatory requirements require shorter availability or immediate removal

D2.2.5. Where a dataset is withdrawn, if required by the Scheme, the member will follow the Scheme's process for notifying users

D2.2.6. Data is documented on publicly available URLs

D2.2.7. A mechanism is available for people to provide feedback and ask questions (e.g. human technical support)

D2.2.8. Responses to feedback and questions are guaranteed to be provided within a reasonable timeframe that is transparent to users, not exceeding 30 days

D2.3. Technical

D2.3.1. Data is published in content-appropriate formats that enable data to be used in an interoperable manner by machine-based systems

D2.3.2. For Shared Data, the dataset is immediately available via an endpoint implementing the Scheme's secure data interchange requirements to any Scheme-registered application that meets the terms of its license, conformance and access control metadata

Level 3

The dataset meets all the requirements of Level 2, plus assurance that:

D3.1. Operational

D3.1.1. A schedule is published at a public URL documenting the process of maintaining the data's availability

D3.1.2. Published data is monitored for unauthorised changes and processes in place for restoration within the 99.5% uptime commitment (see below)

D3.1.3. A document detailing the dataset provenance, excepting any information that is security- or commercially-sensitive, is authored alongside the dataset, and

D3.1.3.1. for Open Data, published via a public URL, accessed with the same method as the dataset,

D3.1.3.2. for Shared Data, made available to the specific roles pre-authorised to access the dataset within the relevant Trust Framework and/or Scheme, except where there is no commercially reasonable reason to restrict publication, when it is made available to all roles.

D3.1.4. The process for assessing and prioritising data requests or new use cases is published

D3.2. Technical

D3.2.1. Metadata cites, in a machine-readable format, the underlying content-appropriate open standard(s) used for publishing the dataset

D3.2.2. Publication of a single consistent URL (a "permalink"), or clear rules for how URLs are constructed, for accessing the dataset

D3.2.3. Machine-readable metadata describing the contents of the dataset is provided (e.g. JSON-LD, CSVW)

D3.2.4. Where data is provided by an API, the API has a machine-readable definition (e.g. OpenAPI)

D3.2.5. The service hosting the dataset has availability of at least 99.5%

Level 4

The dataset meets all the requirements of Level 3, plus assurance that:

D4.1. Legal

D4.1.1. The full set of license terms are machine-readable and available at a persistent URL in a consistent manner (e.g. https://creativecommons.org/publicdomain/zero/1.0/)

D4.2. Operational

D4.2.1. Data quality parameters and processes are published in a machine-readable format at a persistent URL in a consistent manner

D4.3. Technical

D4.3.1. Provenance is published in a machine-readable format. This means that:

D4.3.1.1. Shared Data provenance is made available, as a minimum, to the specific roles pre-authorised to access the dataset

D4.3.1.2. Open Data provenance is made available to all roles, published via a public URL

D4.3.2. Uniform Resource Identifiers (URIs) are used as identifiers within data

D4.3.3. The service hosting the dataset has availability of at least 99.9%