opendata**soft**

Data Lineage

Information Note

Last update: 2024-03-01

Opendatasoft enables all teams using its services to create, publish and share in their ecosystems new data experiences that are more accessible, more relevant, and easily reusable. By democratizing access to data, you optimize your operations, accelerate the development of new activities, and nurture relationships of trust with your stakeholders.

In this context, and in order to ensure the above objectives, Opendatasoft offers to its Clients as part of the provision of its Service, a data lineage functionality allowing to understand how the datasets of a Workspace are used in the Opendatasoft ecosystem, both within the Workspace concerned but also towards other Workspaces, and to encourage exchanges between Producer of datasets and User of the Opendatasoft Platform so that each party supports the outreach of the data network.

The present Information Note is intended for companies or entities clients of Opendatasoft (hereinafter "Client") and aims to present to the Clients the interest and the functioning of the Data Lineage and to describe the good practices of the parties involved in the use of this functionality. This document is for information purposes and does not modify the contractual basis of the Services.

Opendatasoft would also like to inform its Clients that the Data Lineage functionality does not involve the processing of personal data within the meaning of the General Data Protection Regulation ("GDPR"). Indeed, this functionality only uses data relating to legal entities and metadata that are not likely to identify a natural person. Moreover, within the framework of the Data Lineage functionality, no confidential information is used.

This Note is accessible on https://legal.opendatasoft.com/ and may be subject to updates, the applicable version of which will be accessible on this same page.

1. Definitions

All capitalized terms not defined in this Charter shall have the meaning given to them in the Opendatasoft Terms of Services and Terms of Use.

• **Data Lineage**: means a complete visualization of data flows by providing a clear understanding of the dependencies between the different Objects, as well as how and why dataset is transformed along the way; this lineage is documented by listing the origin and final destination of a dataset as well as all the transformations it has undergone at each stage of its journey.

opendata**soft**

Usage Metadata: refers to lineage metadata available at the level of a dataset or page and modeled in the form of a graph. This adapted structure is an overview of the direct and indirect, upstream, and downstream relationships of Objects involved in the construction or use of a dataset. Usage Metadata is:

- the name and nature of the ODS Object (Workspace name, title, identifier and linked fields of the ODS Object)
- the name and nature of the Third-party Object (title and source type)
- the relationship, i.e. the type of use.

- **ODS Object**: refers to a dataset, a page or a map or graph editor.
- **Third-party Object**: refers to an external source (file, URL, or remote service) of a dataset.
- **Object**: refers to an ODS Object or a third-party Object.
- **Relationship**: refers to a dependency between two Objects that establishes a **directed link** between an **origin** Object and a **destination** ODS Object.

2. Principles of Data Lineage

Opendatasoft's Data Lineage functionality provides tools to understand the dependencies between Objects and enhance the data network.
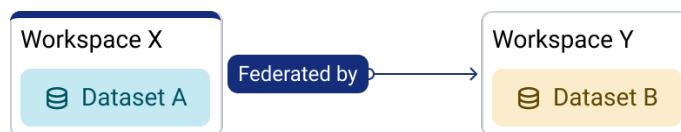
In the example below, A is the origin Object and B the destination Object. B depends on A.


There are two types of Relationships:

direct: when two Objects are linked without any transformation. One is identified as origin and the other as destination.

In the example below, dataset A is the origin ODS Object and dataset B is the destination ODS Object.

indirect: when a destination ODS Object depends on another direct or indirect Relationship with other Objects.

In the example below, Dataset E is the origin ODS Object of dataset A. Dataset A is the origin ODS Object of dataset B. Datasets E and B have an indirect relationship.
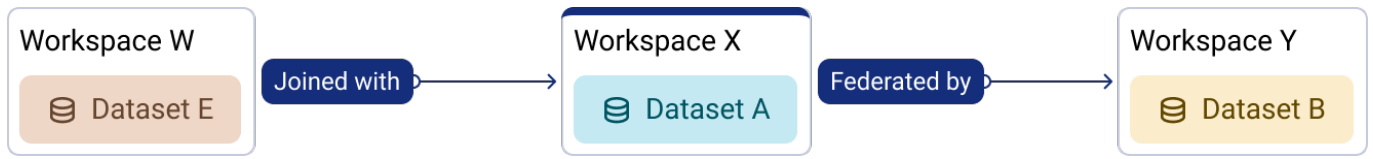
3. Data sharing

3.1. Data Sharing between the Client and Opendatasoft

Opendatasoft has fully developed internally all the functionalities related to data lineage. This functionality is maintained and hosted in the same way as the Opendatasoft Platform, and under the Opendatasoft contractual conditions.

In order to offer a high value-added Data Lineage functionality, the visualization of the Relationships between Objects implies sharing the Usage Metadata with Opendatasoft and other Clients owning Opendatasoft Workspaces.

Within this framework, and according to the conditions described in this Note, Opendatasoft gathers all the Usage Metadata collected for the Data Lineage in order to improve the quality of the Opendatasoft Platform and to enhance the resulting data network and propose relevant functionalities for the management and sharing of data catalog for the benefit of the Users.

The content of the Client Data is neither modified nor read by the functionality. The Data Lineage solution scans the Opendatasoft Platform to analyze the configuration of the datasets and their processing tasks or the physical code of the pages/editors to extract only the necessary Usage Metadata. The functionality does not transform Client Data during any of its analyses.

Usage Metadata between Objects is extracted from the system and updated after publication. Anonymous aggregated data may also be used by Opendatasoft for the purpose of evaluation, improvement, and maintenance of the functionality, for statistical purposes and for the promotion of data sharing.

3.2. Sharing between Clients

It is important for Opendatasoft to ensure that each Client has a minimum of usable information (what and how) about the use of their data and to allow them to choose whether or not to share the name of their Workspaces. Sharing modes are available to prevent the identification of the destination ODS Object to the producers of the origin Objects.

Usage Metadata is divided into four levels of display:

By default and for each Relationship, its type and information on the nature of the destination ODS Object are transmitted to the Data Producer.

The name of the user Workspace is subject to a sharing mode chosen by the Client (declared mode/incognito mode).

The naming of the ODS Objects is subject to the access conditions defined on the owner portal.

The designation and source type of third-party Objects are subject to the access conditions defined on the proprietary portal and are not transmitted outside of it.

Opendatasoft recommends to its Clients to be in "declared" mode in order to increase the quality of the Usage Metadata in favor of the data producers. Knowing the use of one's datasets indeed allows to create a dynamic community between several actors.

3.2.1 Application of the Workspace name sharing mode

This choice is made at the level of a Workspace and will apply to other Workspaces that directly or indirectly consume its data. Its activation will be done by the Opendatasoft teams via our internal tools.

3.2.2 Change of Workspace name sharing mode

To make a change between the "declared" and the "incognito" mode, the request must be made to the Opendatasoft customer service who will take the necessary contractual steps and adapt the applicable pricing conditions. After validation by Opendatasoft, the change of sharing mode will be effective, and the existing data will be updated on the different media.

3.2.3 Use of Usage Metadata

The Data Lineage functionality and the Usage Metadata are intended for Clients only and their use is limited to strictly internal purposes. Any commercial use of the shared data is at the sole risk of the Client organization.