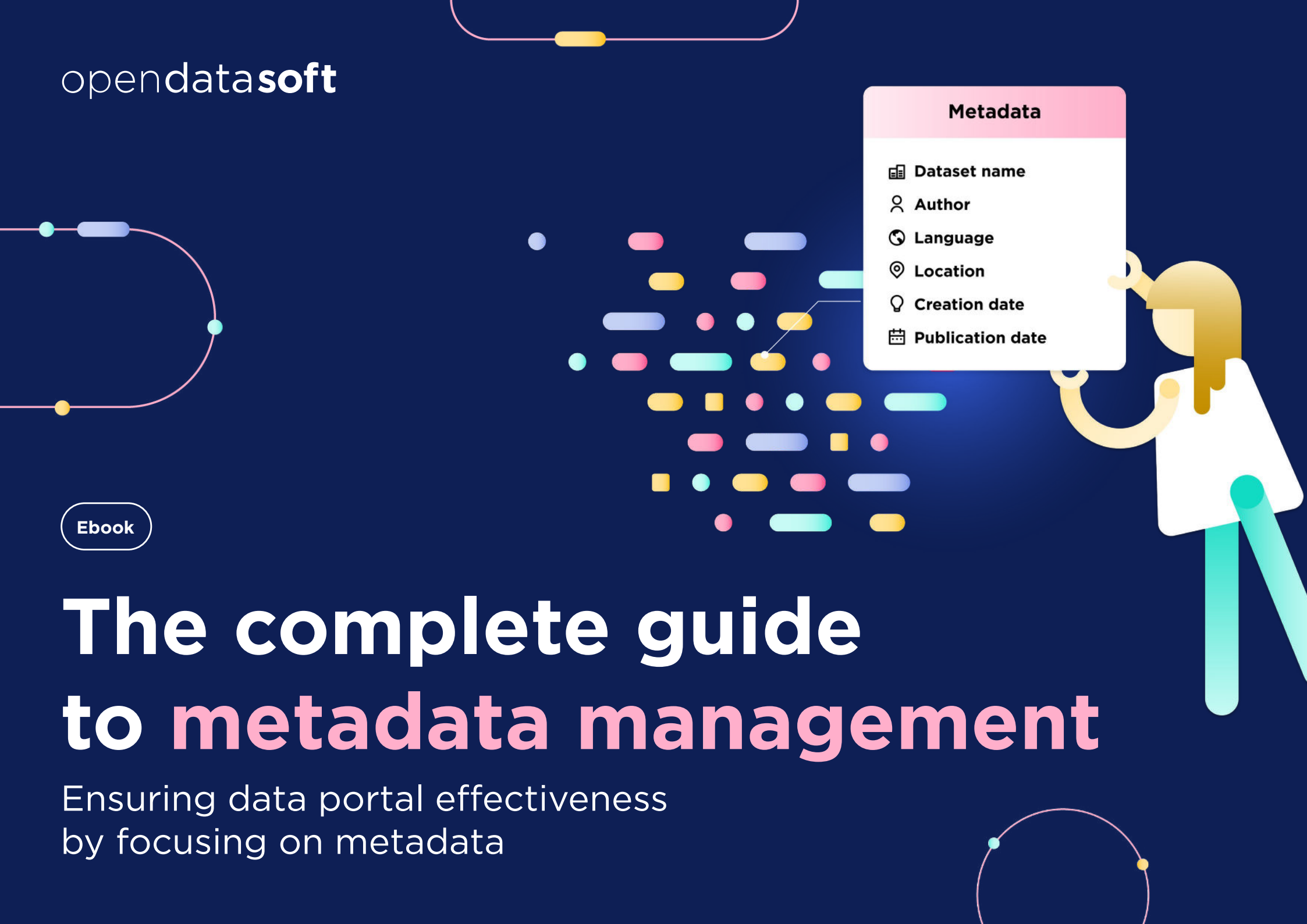
opendatasoft



Ebook

The complete guide to metadata management

Ensuring data portal effectiveness
by focusing on metadata

Introduction

Well-structured metadata is crucial to enabling data to be found and used with confidence within organizations and ecosystems. It is central to effective data sharing and reuse.

At Opendatasoft our mission is to accelerate data democratization, ensuring that everyone has access to easily understandable information in their working and home lives. Our data portal solution enables data democratization by centralizing all of an organization's data assets and making it available to all internal and external users in a seamless, intuitive way, without requiring specialist skills or support.

Metadata is critical to the success of data portals, enabling data democratization. This guide has therefore been created to explain the importance of metadata and to help organizations harness it effectively across their data portals.

Contents

Part 1

What is metadata?

What is metadata?

Metadata is data that describes other data. This delivers context to make it more easily understandable and usable.

Metadata is data that provides information about other data, making it faster and easier to find and work with. It doesn't tell you what the content is, but describes what is in a particular dataset, helping people understand what it covers, and who has created or collected it, even if they've never seen the dataset before.

A good analogy is a music CD. The data is the content of the CD itself, while the metadata is the title, format, publication date, musician and composer.

The growing importance of metadata

We now live in a world full of data. Metadata is essential to managing and using this data, which comes from hundreds of different sources and is generated in increasing volumes.
All of this data is in different formats, making it difficult to understand what individual datasets cover, the measurement units they use, how regularly they are updated or who owns them. This all makes it difficult to compare or use datasets with confidence.

Metadata and data portals

Metadata is particularly important to sharing information through centralized data portals. It ensures that the data on internal and external data portals is discoverable, understandable, reusable and interoperable - by both humans and technology, such as artificial intelligence (AI). Users can not only find the information they are looking for faster, removing unnecessary downloads, but also means people are more confident when accessing and reusing these data assets in their working lives. Metadata is therefore essential to running a successful data portal and building a data-centric organization and underpinning data democratization.

This comprehensive guide provides an introduction to metadata and best practice advice for implementing successful metadata management to drive your data portal forward.

Part 2

What is the purpose of metadata?

What is the purpose of metadata?

To meet its purpose, metadata should follow eight golden rules:

1 Metadata should provide a dataset with context

It should explain more about the dataset's content, helping engage users and mean they access it with confidence. Contextual metadata includes a clear, succinct title, an in-depth description of the dataset, an explanation of why it has been created, its original language, the geographical areas it covers, timeframe and the creation, release and modification date.

2 Metadata should make a dataset unique

Often, users will be faced by multiple, similar datasets on a portal, particularly if data has been collected from different sources.
It is therefore vital that metadata explains what makes an individual dataset unique, both through its title and by assigning a standardized Internationalized Resource Identifier (IRI), that includes the main keywords and source.

3 Metadata should provide the framework for subsequent uses of a dataset

Users on a portal need to understand how they can use your dataset, which is where the license attribute of metadata comes in. This outlines the license that governs reuse, which can be either an international standard (such as Creative Commons for open data) or a specific company standard for internal or partner data portals. Licenses can also include internal rules about who has access to the data, how it can be used and how long it is current for.

4 Metadata should make you want to reuse a dataset

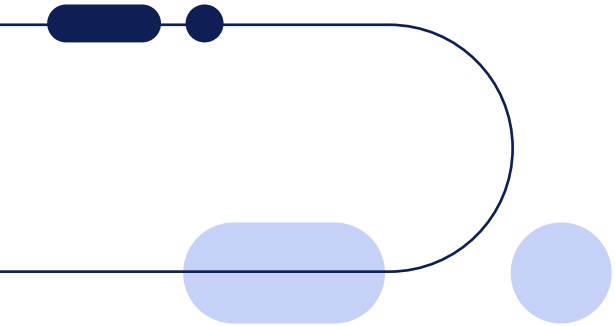
You need to attract and encourage people to reuse your dataset, and that requires you to make the metadata inviting and compelling, particularly if you are looking to monetize usage. Ensure that metadata provides a clear description to accelerate usage, while outlining the formats it is available in, and suggesting potential ways it can be reused.

5 Metadata should make a dataset interoperable

Metadata should follow set internal or external standards so that data can be confidently compared or integrated with information in other datasets. At a basic level this means standardizing how fields are described, and how dates are formatted.

7 Metadata should allow for a dataset to be found by search engines, AI and humans

Ensure your dataset is discoverable both within your portal and through online search engines such as Google in the case of open data.
Use standardized terms to describe your data (as set out in relevant online thesauruses and guides) to make it simple to find by humans too. Good metadata also makes it easier for relevant datasets to be found and used by AI, which is essential to training models and algorithms.



6 Metadata should provide reassurance regarding the dataset's reliability

Ensure that users are confident about the quality of your dataset by providing high quality, standardized metadata. That means checking to avoid misspelling of terms in your metadata (which also hurts discoverability), and following set standards around how you name its creator.

8 Metadata should ensure the longevity of a dataset

Data can have a long life and be shared in many places, often outside your organization. People want to be able to get in touch if they have queries around a dataset. Therefore, include contact details for the data owner alongside the license. However, remember that people move on - rather than listing a person as a contact, give a department or team name and provide a generic, rather than personal email address.

Part 3

What are the benefits of metadata?

What are the benefits of metadata?

Metadata makes data sharing at scale possible - without adequate metadata users not directly involved in creating a dataset will not be able to discover it or reuse it with confidence. It is therefore essential to data sharing and to building a data-driven organization. Effective metadata delivers these seven key benefits:

1 **It enables data discoverability**
sharing and reuse on data portals, by allowing users to quickly search for, find and use relevant datasets with confidence.

2 **It aids better decision-making**
As data is better organized and can be easily compared, both humans and AI can make more informed, faster, and more confident business decisions.

3 **It is at the heart of effective data governance**
Metadata is central to agreed data governance standards, delivering compliance with corporate policies.

4 **It improves data quality**
as metadata provides information on the quality and reliability of the dataset.

5 **It delivers time and efficiency savings**
as users can find and use relevant information more quickly themselves, without requiring support from data teams.

6 **It increases collaboration**
internally and externally by enabling people to work together with shared, mutually understood data.

7 **It ensures compliance**
Metadata enables data stored in different systems and databases to be interoperable, providing an up-to-date record of information and any changes made to it.

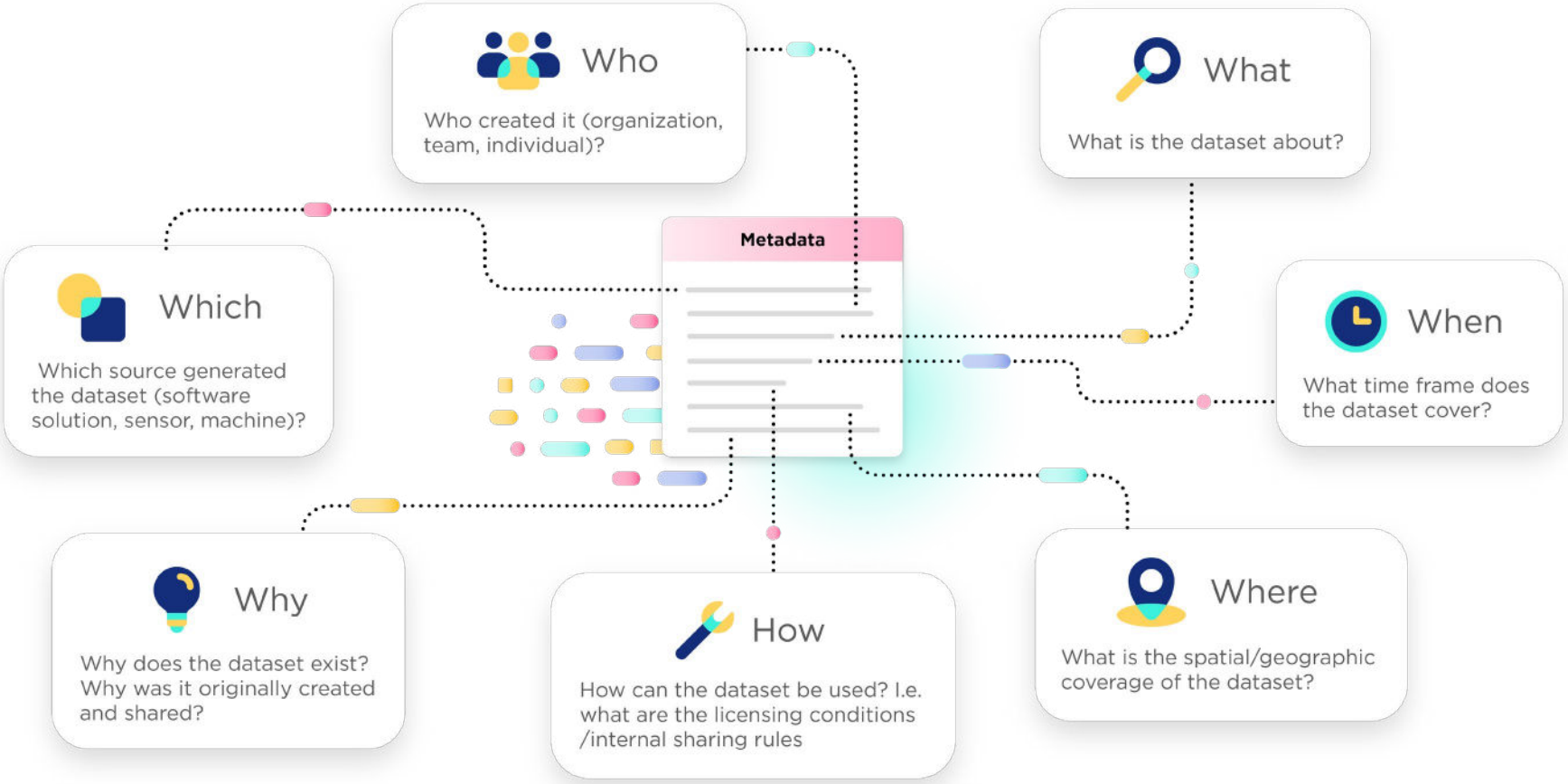
Part 4

What are the characteristics of metadata?

What are the characteristics of metadata?

Understanding the characteristics of metadata

A good way to understand what metadata covers is to follow the W7 ontological model, which is similar to the questions that journalists ask when researching an article:



Types of metadata

Metadata can be classified into multiple types, based on what it is used for:

Descriptive metadata

This describes a piece of data by adding certain details. For example, the creator of the data, the title of the data, the creation date, etc. Sometimes this information is added automatically (for example photos automatically record the date and time when they were taken), whereas other descriptive metadata (such as the title) may be added or improved manually.

Administrative (also known as legal) metadata

This allows people to find the complete history of the data and how it can be used. For example, it includes the name of the author, the creation date and the various modifications made over time, as well as how it can be licensed and who can access it.

Structural metadata

This allows for the easy retrieval of data and aids interoperability, since it specifies how a data asset is classified in the information system. For example, structural metadata provides the format or medium of the data. The contents page of a book is a good example of structural metadata.

Relationship metadata

This covers how datasets relate to other information within a data portal or data stack. It explains how particular data assets fit together and flow, and is therefore crucial to data lineage, modeling, and identifying patterns.

Metadata and ontologies

Ontologies are descriptions and definitions of terms that provide a vocabulary for creating more complex metadata, especially when describing specialist information.

Within metadata schemas, ontologies ensure that relationships between different terms within metadata are clear, and that datasets are interoperable. They set out how a dataset is organized in terms of the fields covered and the type of information in each field (such as specifying it must be a numeric figure). This is reflected in the structural metadata, which provides a standard definition for each column header type.

For example, in the case of building permits, ontologies ensure that data is consistent between different municipalities, making it easier to manage and compare across different communities.

Ontologies can include some or all of the following descriptions/information:

- Classes (general things, types of things)
- Instances (individual things)
- Relationships among things
- Properties of things
- Functions, processes, constraints, and rules relating to things

Ontologies help us to understand the relationship between things. As an example, an “android phone” is a subject of an object class, “cell phone”.

Part 5

Metadata standards

Metadata standards

Setting and enforcing standards for how metadata is written, structured and applied is essential to successful data governance and sharing.
For example, if different data owners use similar but slightly different keywords within descriptive metadata, potential users will be confused and unsure if they cover the same area.

Creating consistent internal metadata standards is therefore vital. However, given that data assets are increasingly shared externally, and that creating your own standards can be time-consuming it is better to start with internationally recognized metadata standards. These reduce the amount of resources required and maximize consistency.

The exact standards chosen will vary depending on the type of data being described (such as structured or unstructured), and the industry you operate in. Many commonly used standards are linked, or rely on agreed International Standards Organization (ISO) approved standards.

Standards include:

- **DublinCore** The Dublin Core standard, a set of 15 core data elements (used to describe digital resources such as documents, videos and images.

- W3C Resource Description Framework (RDF), a model for representing, sharing, and integrating data across different applications and online platforms.

- ISO ISO 19115, specific to geographic metadata, this includes elements such as spatial and temporal extents, data quality, and resource maintenance information.

- W3C The W3C Data Catalog Vocabulary (DCAT), built on the RDF model and used to facilitate interoperability between online data catalogs.

- European Commission | The EU's INSPIRE framework for spatial data, which is based on the ISO 19115 (dataset) and ISO 19119 (services) standard

An individual organization may have different metadata requirements that cannot be fully met by available metadata standards. This means creating customized internal metadata templates - normally best practice is to start with international standards and to adapt. For example, you should adopt ISO standards for date and time formats to maximize interoperability.

Part 6

Metadata management with Opendatasoft's data portal solution

Metadata management with Opendatasoft's data portal solution

Metadata templates are forms that group together a set of fields used to describe datasets, in order to ensure their discoverability and effective administration.

Delivering choice in metadata templates

Opendatasoft's metadata templates feature enables users to either create and enable their own, personalized data templates or activate pre-existing templates within the platform that are based on different international models. There are three categories of metadata templates within the platform, each with associated benefits.

Standard templates:

These ensure a personalized level of compliance tailored to an organization's requirements (taxonomy, sector-specific, or specific vocabulary).

Interoperability templates

(activatable only, non-editable):

Users can select from a variety of models to ensure compatibility with different international standards, such as DCAT, DCAT-AP, INSPIRE, or Dublin Core. Metadata shared by Dublin Core and DCAT-AP templates are automatically distributed from the Standard template, reducing redundant entries.

Administrative templates

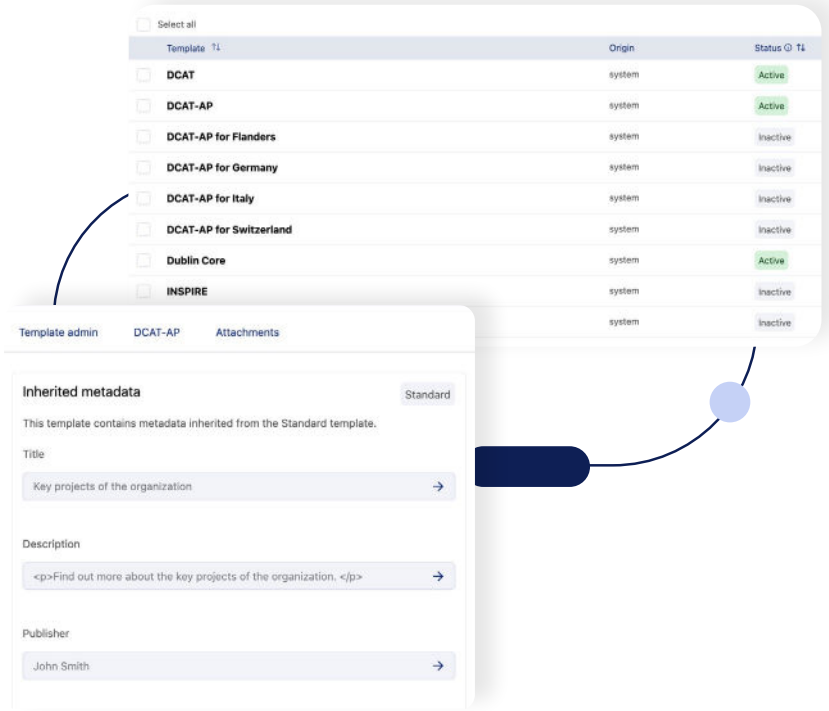
(only visible to portal administrators):

These ensure good internal governance of metadata.

Key features of Opendatasoft's metadata templates feature

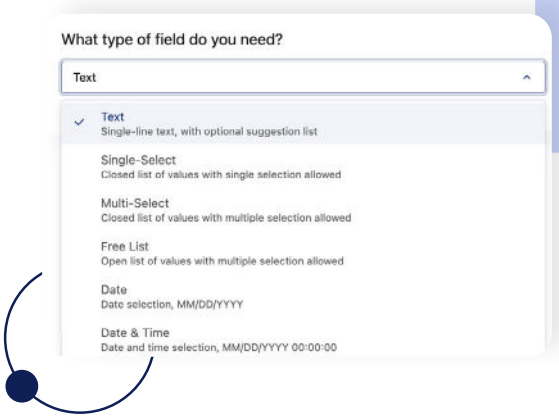
A user-friendly metadata management experience

Opendatasoft's metadata templates feature stands out for its user-friendly interfaces and clear instructions. This makes the user experience easy and enjoyable, encouraging quick adoption and facilitating proper metadata completion. Clear instructions are provided to simplify the metadata entry process and encourage completion with fill-in assistance texts for each field.



Advanced field customization

Users can fine-tune fields based on their organization's precise requirements, offering maximum flexibility and personalization. With this feature, users have the freedom to create fully customized templates from scratch, tailoring them to their exact specifications, from start to finish, including around name and fields.



The Opendatasoft metadata template feature provides multiple benefits:

- Improved discoverability and easier data search
- Guaranteed internal, regulatory, or sectoral compliance levels for datasets
- Optimized dataset administration and encouragement to support users fully completing metadata fields
- Facilitation of data integration and interoperability, as well as metadata harvesting by other data ecosystems

Part 7

Best practice advice for good metadata management

Best practice advice for good metadata management

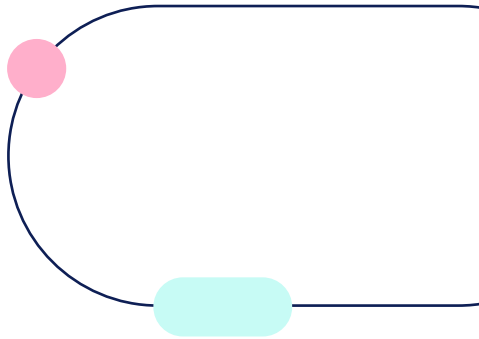
Effective metadata management enables seamless internal and external data sharing and therefore is vital to the success of your data portal. Following these best practices ensures that metadata supports effective data sharing and reuse through data portals:

- Define a metadata strategy based on overall business objectives around data sharing. What are you aiming to achieve with your data portal?
- Talk to users and collect and understand their requirements and potential use cases. What types of data is involved, who will be using it and where? For example, will metadata be used to support risk management, data sharing, compliance?
- Create a cross-departmental team around metadata. This might be part of your overall data governance project. It should include relevant data owners and users from across the company.
- Work together and agree on a standardized metadata classification scheme for your organization. This should reflect a common vocabulary and be based as much as possible on recognized standards. Create metadata templates for your data assets.
- Roll-out your metadata program. Educate all data owners about the importance of metadata and communicate standards, practices, templates, and processes to make adoption as seamless as possible.
- If you already have a large number of data assets without metadata, prioritize adding metadata to the most important datasets in your portal. This will drive usage and move the program forward.
- The cross-departmental team should then monitor the data portal on an ongoing basis, ensuring that metadata standards and good practice are being met, reviewing and updating them as necessary.

Conclusion

The importance of metadata to data portal success

To drive data democratization, internal and external users need to be able to confidently discover, access, share and reuse high-quality data assets from across your organization, without requiring specialist skills or support.



Data portals provide the centralized, one-stop-shop to make all of your data easily available. However, without effective metadata describing each data asset, users will find it hard to discover relevant information, reducing their confidence levels and meaning that you don’t get full value from your data assets. This means that while it may seem like a technical subject, metadata is actually business-critical. Strong, standards-based metadata is vital to scaling data sharing, and to underpinning greater data use and democratization across your organization.

opendatasoft

Opendatasoft is the leader of data democratization.
It provides organizations with a Data Portal solution allowing seamless self-service access to all of an organization's data assets. Opendatasoft empowers organizations to scale personalized and seamless data experiences in record time. Opendatasoft is the essential data management solution to decrease costs and improve efficiency, increase and build new revenue streams, mitigate risks and manage crises.

Opendatasoft serves +400 customers in 25 countries, powering more than 3,000 data portals. Based on this experience, Opendatasoft has developed a unique expertise in data management, which is used to provide customers with premium services to help them deliver use cases that meet their specific needs.

www.opendatasoft.com