Av Helge Ridderstrøm (førsteamanuensis ved OsloMet – storbyuniversitetet)

Sist oppdatert 28.11.25

Om leksikonet: https://www.litteraturogmedieleksikon.no/gallery/om_leksikonet.pdf

Metadata

Data som brukes til å beskrive andre data, til å fastsette egenskaper ved og innhold i dokumenter. (Dokumenter i svært vid betydning, f.eks. bøker, filmer, dataspill og reklameplakater.)

"[T]he generally accepted view seems to be that the term originated in the late 1960s in the context of computer system design to refer to the use of one data element to describe or represent some characteristic of another data element. […] Aside from sporadic early use, the term entered the discourse of the library and information sciences in the mid-1990s, particularly in relation to the development of digital library systems and the emergence of the Internet and the World Wide Web as major social forces." (Mayernik 2020)

Standardiseringsorganet NISO definerer metadata som "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource." (https://online.standard.no/nb/niso-z3985-2001; lesedato 29.09.25)

Metadata "have a central role in the establishment of knowledge, evidence, and truth." (Mayernik 2020)

"Metadata er kort forklart "data om data". Det er informasjon som beskriver dataene og gir dem mening. Uten gode metadata vil det ofte være vanskelig å forstå datasettet og hvordan dataene kan brukes. Metadata er også viktig for å gjøre datasett søkbare og gjenfinnbare, også når selve datasettet ikke kan publiseres åpent. […] Typiske metadata er informasjon om:

- hvem som har produsert eller er ansvarlig for datasettet
- hvilket fagfelt det er snakk om
- hva slags type data det gjelder
- hvilke formater de er i

Metadataene kan også inkludere informasjon om utstyret eller programvaren som er brukt. […] Det skilles ofte mellom følgende typer metadata. Dette er ikke en utfyllende liste, men kan være et utgangspunkt for å tenke over hvilke metadata som bør være med for det aktuelle datasettet:

Beskrivende – for eksempel navn på forfatter(e)/forsker(e), tittel på dokumentet osv.

Administrative – for eksempel opprettelsesdato, versjonering, lisenser

Strukturelle – for eksempel forholdet mellom filer, betydningen av variabler

Tekniske – for eksempel informasjon om format og eventuell programvare eller maskinvare

I tillegg kan det være nyttig å vurdere ulike nivåer av metadata. På prosjektnivå skal metadata beskrive den overordnede konteksten rundt datainnsamling, hvem som har gjennomført arbeidet og hvordan prosjektet er finansiert. Beskrivelse av datasettet, nøkkelord og lisens hører også til på overordnet nivå. Metadata på datasettnivå er mer konkret og detaljert. Eksempler kan være: opplysninger om filtyper, hvilke måleinstrumenter som ble brukt og beskrivelse av variabler og enheter. […] For at metadata skal være maskinlesbare bør metadataene være så strukturerte og standardiserte som mulig. Vi anbefaler at du benytter standardiserte termer, taksonomier/ontologier og vokabularer som er tilgjengelig innenfor ditt fagfelt. Det finnes en rekke metadatastandarder. Noen er generiske, og kan brukes for alle fagfelt, andre er tilpasset spesifikke fag og disipliner. Dublin Core er en generisk metadatastandard som består av en liste over elementer for å beskrive et datasett eller et annet digitalt objekt." (https://i.ntnu.no/wiki//wiki/Norsk/Metadata+og+dokumentasjon+av+datasett#:~:text=Metadata; lesedato 29.08.25)

"*Metadata* is commonly used now as a blanket term for a range of practices, many of which pre-existed the term *metadata* itself, including library cataloging, archival description, and scientific data documentation, along with more recent phenomena, such as automatically generated information associated with digital images or social media streams" (Mayernik 2020).

"Online streaming services for music, movies, and other forms of personal entertainment rely on metadata of various kinds to provide recommendations, personalization, and categorizations to their users" (Mayernik 2020).

"Metadata generally has little value on their own. Metadata adds value to other information, but are usually not valuable in themselves. There are exceptions to this rule, such as text transcription of an audio file. […] There are three kinds of metadata:

Descriptive metadata consist of information about the content and context of your data. Examples: title, creator, subject keywords, and description (abstract)

Structural metadata describe the physical structure of compound data. Examples: camera used, aperture, exposure, file format, and relation to other data or files

Administrative metadata are information used to manage your data. Examples: when and how they were created, who can access them, software required to use them, and copyright permissions" (https://guides.lib.unc.edu/metadata/definition; lesedato 29.08.25).

"Metadata (in terms of data management) describe a data set: how they were collected; when they were collected; what assumptions were made in their methodology; their geographic scope; if there are multiple files, how they relate to one another; the definitions of individual variables and, if applicable, what possible answers were (i.e., to survey questions); the calibration of any equipment used in data collection; the version of software used for analysis; etc. Very often, a data set that has no metadata is incomprehensible. […] They are pieces of information that have some meaning in relation to another piece of information. They can be created, managed, stored, and preserved like any other data." (https://guides.lib.unc.edu/metadata/definition; lesedato 29.08.25)

Metadata knytter "opplysninger til data som gjør det mulig å forstå disse. Metadata deles normalt inn i beskrivende (eksempelvis: tittel, skaper og emne), administrative (eksempelvis: id [dvs. identifikasjonsnummer], rettigheter og proveniens) og strukturell informasjon om relasjonen mellom informasjonsobjekter (Alemu & Stevens, 2015). En samling med metadata for et bestemt formål blir ofte kalt et skjema, metadataskjema, metadatasett eller elementsett. Termene er kun ulike måter for å si at metadataelementer, for eksempel tittel, emne og id, skal struktureres på en gitt måte. Et metadataskjema kan også inneholde en oversikt over lovlige verdier for hvert element og regler for et element. Et metadataskjema er alltid skapt med hensikt og med et formål. Svært forenklet kan alle skjema plasseres på en skala mellom de to ytterpunktene optimalisering for et domene eller for interoperabilitet. Dette innebærer at et skjema enten er skapt for å kunne uttrykke et domene presist eller for å kunne samhandle og utveksle informasjon på tvers av domener." (https://www.arkivverket.no; lesedato 29.09.25)

"[P]articular tasks or actions that might be facilitated by metadata […] such as authenticating, browsing, discovery/finding, preserving, or understanding information/data resources. This is perhaps the one commonality among these various definitions and categorizations of metadata: that metadata is created to be used, for some purpose(s), by people or computer applications." (Mayernik 2020)

"Metadata are created and maintained because they improve use of research data and enable their re-use. Good metadata can make up for human shortcomings. People forget and misplace things, and leave research projects, taking their knowledge of the research methodology and the data with them. Metadata ensure that data are FAIR: Findable, Accessible, Interoperable and Re-usable.

Findable: Metadata make it much easier to find relevant data. Most searches are done using text (like a Google search), so formats like audio, images, and video are limited unless textual metadata are available. Metadata also make text documents easier to find because they explain exactly what the document is about.

Accessible: Once someone finds the data they need, metadata indicate how they can be accessed, possibly including authentication and authorization.

Interoperable : Assigning metadata enables a data set to be integrated with other data. In addition, they also make it easier for data to interoperate with applications or workflows for analysis, storage, and processing.

Re-usable: To re-use a data set, researchers need to understand how the data are structured, definitions of terms used, how they were collected, and how they should be read or used. Data need to be well-described so that they can be replicated and/or combined in different settings." (https://guides.lib.unc.edu/c.php?g=8749&p=44500; lesedato 29.08.25)

"Essentially applying metadata ensures that all data shared on data marketplaces is discoverable, understandable, reusable and interoperable – by both humans and technology/artificial intelligence (AI). It is key to optimizing governance, data quality, compliance, and consumption. […] Metadata summarizes basic information about data, making finding and working with particular instances of data easier. […] A good analogy is a book – the data is the content of the book itself, while the metadata is the title, format, publication date, author and the book's subject. […] At its heart, metadata should contribute to data being discoverable and reusable in eight different ways:

Metadata should provide a dataset with context. This means explaining what it covers, the themes, keywords to describe it, how the data has been collected and information such as any numerical units used, such as dollars, inches or centimeters.

Metadata should make a dataset unique. It has to differentiate it from other, similar datasets so that users can choose between them with confidence.

Metadata should provide the framework for subsequent uses of a dataset. This includes licensing conditions, whether it can be used externally as well as internally, and any organizational rules around who can use the data, and for what purposes.

Metadata should make you want to reuse a dataset. It has to be comprehensive and compelling, providing clear descriptions that accelerate usage, while outlining the formats it is available in, and suggesting potential ways it can be reused.

Metadata should make a dataset interoperable. It should follow set internal or external standards so that data can be confidently used or compared with information in other datasets. At a basic level this means standardizing how fields are described, and formats such as dates.

Metadata should provide reassurance regarding the dataset's reliability. By including information on the source, how often it is updated, and what it covers, users should be able to ascertain how reliable the data is.

Metadata should allow the dataset to be found by technology and humans. That means using standardized terms to describe your data (as set out in relevant online thesauruses and guides). This ensures that it can be immediately found through search, whether via your internal marketplace or, in the case of public data marketplaces, through search engines such as Google. Good metadata also makes it easier for relevant datasets to be found and used by AI, which is essential to training models and algorithms.

Metadata should ensure the longevity of a dataset. Data can have a long life and can be shared in many places. Therefore include contact details for the data owner alongside the license, but 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.

The types of metadata

The possibilities for describing things with metadata seem endless. Certainly metadata schema can be simple or complex, but they broadly fit into four types:

- Descriptive metadata: this adds details to describe a dataset, such as who created it, its name and what it contains.

- Structural metadata: this specifies how data is classified in terms of format, helping it to be easily found and retrieved.

- Administrative metadata: this includes rights management and licensing information about the data

- Relationship metadata: this explains how datasets relate to other information, helping monitor data lineage.

[…] Metadata delivers seven key benefits:

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

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.

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

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

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

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

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. […]

The W7 Ontological Model of Metadata […]

What – What is the dataset about?

When – What is the time frame that the dataset covers

Where – What is the spatial/geographic coverage of the dataset?

Who – Who created it (organization, team, individual)?

How – How can the dataset be used? I.e. what are the licensing conditions

Which – Which source generated the dataset (software solution, sensor, machine)?

Why – Why does the dataset exist? Why was it originally created and shared?

Metadata standards

While the idea of metadata is simple in principle, the concept of applying metadata to your datasets can seem daunting. Where do you start? How do you describe your data so that it is consistent and can be shared internally and externally? How do you scale your program? To help, a number of international standards have been formulated and agreed. These include the Dublin Core standard, the W3C Data Catalog Vocabulary (DCAT), and the EU's INSPIRE framework for spatial data. These are based on agreed ISO standards to ensure interoperability and wide reuse. […]

A lot of the discussions around data quality and data discoverability have revolved around metadata and something called ontologies. Ontologies are descriptions and definitions of relationships and can be used within metadata. 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."
(Lauréline Saux i https://www.huwise.com/en/blog/what-is-metadata-and-why-is-it-important-data/; lesedato 05.11.25)

"Many structures for metadata have been formalized into standards, ranging from general purpose metadata standards such as the Dublin Core to discipline-specific standards for particular kinds of resources, e.g. geospatial information (Danko 2012; Brodeur et al 2019). Standardized schemas and structures facilitate the interoperability of metadata between systems and applications (Zeng 2019). Metadata standards are commonly organized around a set of elements (such as "title", "author", "date")" (Mayernik 2020).

"The Dublin Core, so named because it was formulated in a workshop in Dublin, Ohio, in March of 1995, was explicitly motivated by a desire to develop a common approach to describing electronic resources that would enable better discovery and collection of resources on the Web" (Mayernik 2020). Slik er Dublin Core-standarden:

Metadata subtag:	**Definition:**
Title	A name given to the resource
Creator	A person primarily responsible for making the content of the resource
Subject	The topic of the content of the resource

Description	An account of the content of the resource
Publisher	An organization or person responsible for making the resource available
Contributor	A person responsible for making contributions to the content of the resource
Date	The date that the resource was published or copyrighted
Type	The nature or genre of the content of the resource
Format	The physical or digital manifestation of the resource
Identifier	String or number used to uniquely identify the object, e.g., the object identifier (OID)
Source	A reference to a resource from which the present resource is derived
Language	The language of the intellectual content of the resource
Relation	A reference to a related resource
Coverage	The extent or scope of the content of the resource
Rights	Information about rights held in and over the resource for rights management

"The Dublin Core element set included thirteen fields (later expanded to fifteen by including Description and Rights) […] the minimum number of metadata elements required to facilitate the discovery of document-like objects in a networked environment such as the Internet." (Mayernik 2020)

Dublin Core er en standard "set of 15 interoperable metadata elements designed to facilitate the description and recovery of document-like resources in a networked environment. […] Dublin Core is the result of an international cross-disciplinary consensus achieved through the ongoing efforts of the Dublin Core Metadata Initiative (DCMI), aimed at providing a foundation for standardized bibliographic description of information resources available via the Internet. In 2003, the Dublin Core Metadata Element Set was published by the International Organization for Standardization as ISO 15836." (Joan M. Reitz i http://lu.com/odlis/odlis_c.cfm; lesedato 30.08.05)

"As the diversity of the definitions, functions, and roles given above illustrates, metadata is not a definite and singular concept. Rather, it is a fluid, multiple, and fractional concept (Law 2004). Metadata is "fluid" in that file naming conventions, catalog records, data descriptions in repositories, user tags on YouTube, notes in personal Excel spreadsheets, email headers, and HTML tags can all be called "metadata". Metadata, as a concept, is also characterized by "multiplicity" in that it is enacted differently in different social settings and situations, from Dublin Core records created by information professionals to descriptions in lab notebooks created by scientists to document their data." (Mayernik 2020)

"Structured metadata can be differentiated from other forms of unstructured metadata, which might also be called "documentation" (Habermann 2018). Unstructured metadata could include any range of traces and practices that achieve some or most of the same goals as structured metadata, namely to create documentation, descriptions, and annotations for the purposes of managing, discovering, accessing, using, sharing, and preserving informational/data resources. As one example, in the context of data archives, it is common to include one or more narrative documents that describe various aspects of the data in more detail than is possible through standardized metadata structures." (Mayernik 2020)

"[D]isciplinary topic experts have firsthand knowledge about how resources related to their areas of expertise are created and used, and thus can provide useful insight into what metadata should be created, and how the metadata may be optimally structured to support use and re-use (White 2010). In the best circumstances, metadata experts and non-experts work together (Gazan 2003)." (Mayernik 2020)

"People feel morally responsible for keeping track of important documents and information, such as family pictures, yearly tax documents, or vaccination records for children. These types of documents may exist in a variety of digital and analog formats, and in multiple technical environments (e.g. email, personal computers, mobile phones). Creating metadata and organization systems for these kinds of things can thus be a source of emotional and interpersonal stress. People thus make context-specific choices about what metadata to create, and in what forms and systems. […] Metadata created for everyday tasks should be expected to have different characteristics than metadata created by professionals for institutional purposes. Such metadata tends to be idiosyncratic, varying in content and structure from individual to individual, and from situation to situation for the same individual." (Mayernik 2020)

"Going forward, researchers and professionals will continue to grapple with long-lasting challenges related to metadata creation and use, including questions about how to negotiate cost/benefit trade-offs between structured and unstructured metadata, and between human and machine generated metadata. But it is also clear that each new generation of information and data technology produce and require different kinds of metadata than systems that existed before. Social media technologies demonstrate a trend toward what Ronald E. Day (2019) has called "a posteriori" metadata generation. Day uses this term to contrast with "a priori" metadata generation, as in library cataloging and classification, Semantic Web ontologies, and scientific data catalogs, where the generation of metadata tends to come at the beginning of the life cycle of information/data use. In systems based on "a posteriori" metadata, the metadata that get generated, stored, and used within the systems are less focused on the properties of the entities within those systems, instead focusing more on what those entities do (or what is done to them by other entities). [Den sosiale kommunikasjonskanalen] Twitter, for example, may collect metadata about particular Twitter users, but they monetize metadata that reflects what those users do. Likewise, Twitter generates metadata about each post, but they monetize metadata that reflects how those posts travel (how many likes, retweets, and replies are generated), and the social networks connected to those interactions." (Mayernik 2020)

"The use of documents as evidence has always been tied up with assessments of the authenticity of the documents involved. Authenticity assessments in the digital age rely on metadata to provide (and create) context and accountability for evidence. When the metadata necessary for these authenticity assessments is itself subject to manipulation and/or capitalization, the grounds for using particular documents or traces as evidence can be eroded (Acker 2018)." (Mayernik 2020)

"In digital systems and applications, metadata moves around, being transmitted, transformed, aggregated, or pulled apart based on the needs and interests of various stakeholders. […] Open questions exist about how metadata journeys and relationships can reflect the "fluid" nature of our societies and cultural practices (Srinivasan and Huang 2005)." (Mayernik 2020)

Litteraturliste (for hele leksikonet): https://www.litteraturogmedieleksikon.no/gallery/litteraturliste.pdf

Alle artiklene i leksikonet er tilgjengelig på https://www.litteraturogmedieleksikon.no