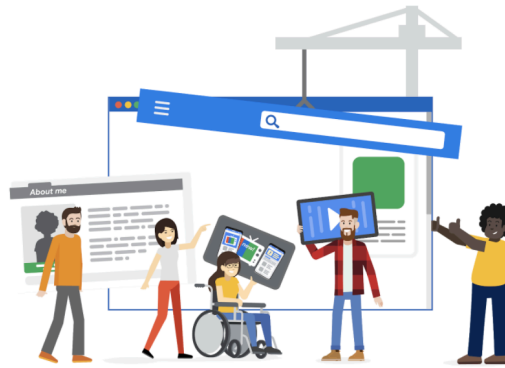# Metadata is as important as the data itself

Data analytics, by design, is a field that thrives on collecting and organizing data. In this reading, you'll learn about metadata and the type of information it can provide. In addition, you'll explore examples of metadata.

Explore a data file by opening any file on your computer or a document in your home or workplace. What is it? Where did it come from? Is it useful? How do you know? This is where metadata comes in to provide a deeper understanding of the data. To put it simply, **metadata** is data about data. In database management, metadata provides information about other data and helps data analysts interpret the contents of the data within a database.

Regardless of whether you're working with a large or small quantity of data, metadata is the mark of a knowledgeable analytics team. Metadata helps people communicate about data across the business and makes it easier to reuse data. In essence, metadata tells the who, what, when, where, which, why, and how of data.

Elements of metadata

Before examining metadata examples, it's important to understand what type of information metadata typically provides:

- **File or document type:** What type of file or document are you examining?
- **Date, time, and creator:** When was it created? Who created it? When was it last modified?
- **Title and description:** What is the name of the item you are examining? What type of content does it contain?
- **Geolocation:** If you're examining a photo, where was it taken?
- **Tags and categories:** What is the general overview of the item that you have? Is it indexed or described in a specific way?
- **Who last modified it and when:** Were any changes made to the file? If yes, when were the most recent modifications made?
- **Who can access or update it:** If you're examining a dataset, is it public? Are special permissions needed to customize or modify it?

Examples of metadata

In today's digital world, metadata is everywhere! Here are some examples—with accompanying images—of where you might find metadata.

Photos

Whenever a photo is captured with a camera, metadata such as filename, date, time, geolocation, and the type of device on which it was taken are gathered and saved with it. The metadata of the following photo is displayed as a pop-up alongside the photo.

Image with accompanying Information pop-up that displays its description, the date and time the image was taken, its size, the device on which it was taken, and an option to add the geolocation of the image.

Emails

When an email is sent or received, it contains metadata such as subject line, sender, recipient, date sent, and time sent.
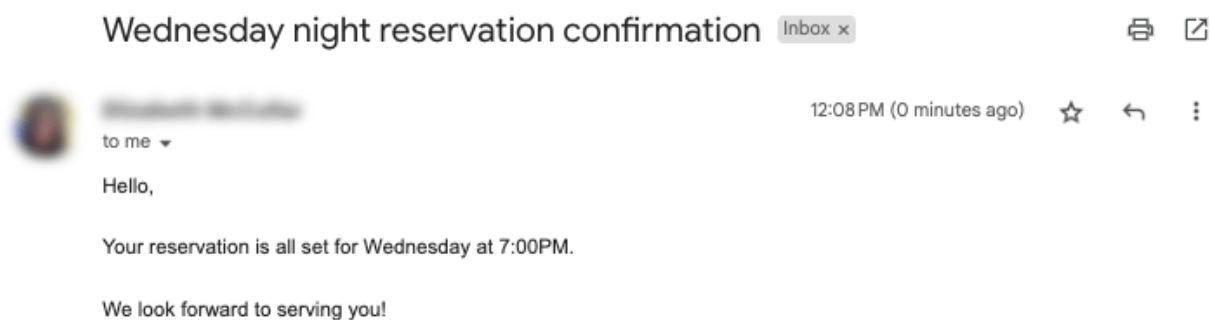
Emails also contain hidden metadata that includes server names, IP addresses, HTML format, and software details. This image includes hidden email metadata such as the message ID and when the email was created.

Original Message

Message ID	<CACdaXyi4ZA4OwhMc6VTBkm9O_fE+GU_MQ9ZPu4cKdG+gbUW6xg@mail.gmail.com>
Created at:	Mon, Oct 23, 2023 at 12:08 PM (Delivered after 12 seconds)
From:	
To:	
Subject:	Wednesday night reservation confirmation
SPF:	PASS with IP 209.85.220.65 Learn more
DKIM:	'PASS' with domain sweetrush.com Learn more
DMARC:	'PASS' Learn more

Hidden metadata from an email that includes Message ID, creation date, the recipient, the sender, the subject line, the SPF, DKIM, and DMARC.

Spreadsheets and electronically created documents

Spreadsheets and documents are already filled with a considerable amount of data, so it's no surprise that they also include metadata such as title, author, creation date, number of pages, and user comments. Additionally, spreadsheet metadata includes tab names, tables, and columns. In the following example, the image demonstrates the metadata for an electronically created Google Sheet:

Details   Activity   Approvals

Labels ⊙

You can't access labels for this file

File details

Type
Google Sheets

Size
552 KB

Storage used
552 KB Owned by Google.com

Owner

Modified
Oct 20, 2023 by

Opened
Oct 22, 2023 by me

Created
May 15, 2023

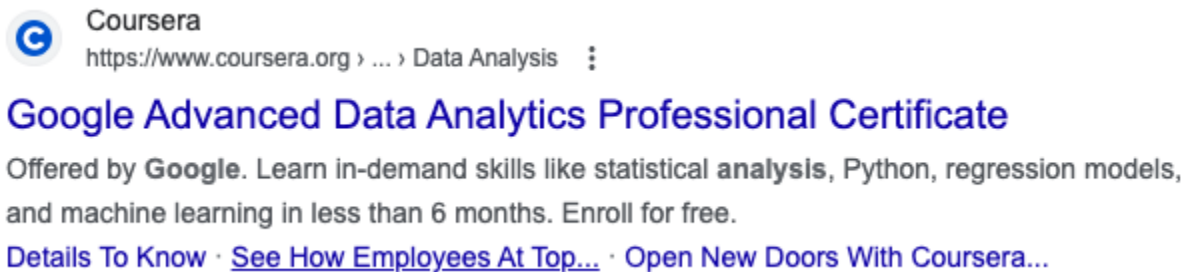**Download permissions**
Viewers can download

Description

Add description

The metadata of a Google Sheet including the title, type, size, storage used, owner, the last person who modified the document, the last person who opened the document, when it was created, download permissions, and an option to include a description.

Websites

Every web page has a number of standard metadata fields such as tags and categories, the site creator's name, web page title and description, and time of creation. Results of search engine queries that you might make on a daily basis are metadata!


A search engine result that includes meta title as Google Advanced Data Analytics Professional Certificate. Under the website hyperlink is the Meta description that says: Offered by Google. Learn in-demand skills like statistical analysis, Python, regression models, and machine learning in less than 6 months. Enroll for free.

Books and audiobooks

Non-digital items can have metadata, too! Every book has standard metadata that will inform you of its title, author's name, a table of contents, publisher information, copyright description, index, and a brief description of the book's contents. Audiobook metadata also includes this data, as well as metadata specific to the audiobook such as narrator and recording length.
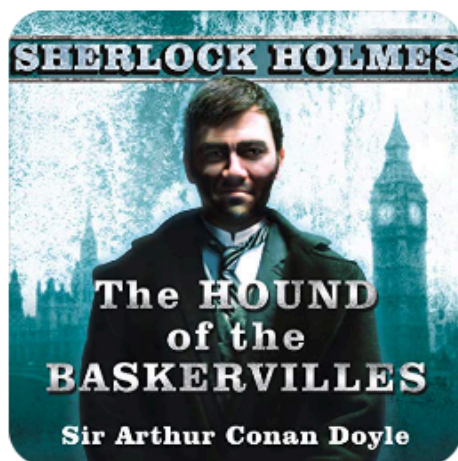
Audiobook's metadata including the title of the audiobook, author, narrator, its length, the release date, language the audiobook is read in, its published, and its rating score.

Key takeaways

Metadata can be found in photos, emails, spreadsheets, websites, and much more! In your daily life, you use metadata to stay organized. As a data analyst, you'll use metadata to understand the content and context of your data, as well as how it's structured. Metadata provides data analysts with information about a data's file type, title, geolocation, who created it, who last modified it, and who has access to it. As a data analyst, it's important to keep accurate records of metadata to ensure that you are able to find, use, preserve, and reuse data in the future. Remember, it will be your responsibility to manage and make use of data in its entirety; metadata is as important as the data itself.

Metadata and metadata repositories

As you're learning, metadata is data about data. It clearly describes how and when data was collected and how it's organized. Metadata puts data into context and makes the data more understandable. This helps data analysts use data to solve problems and make informed business decisions.

In this reading, you'll learn more about the benefits of metadata, metadata repositories, and metadata of external databases.

The benefits of metadata

Reliability

Data analysts use reliable and high-quality data to identify the root causes of any problems that might occur during analysis and to improve their results. If the data being used to solve a problem or to make a data-driven decision is unreliable, there's a good chance the results will be unreliable as well.

Metadata helps data analysts confirm their data is reliable by making sure it is:
- Accurate
- Precise
- Relevant
- Timely

It does this by helping analysts ensure that they're working with the right data and that the data is described correctly. For example, a data analyst completing a project with data from 2022 can use metadata to easily determine if they should use data from a particular file.

Consistency

Data analysts thrive on consistency and aim for uniformity in their data and databases, and metadata helps make this possible. For example, to use survey data from two different sources, data analysts use metadata to make sure the same collection methods were applied in the survey so that both datasets can be compared reliably.

When a database is consistent, it's easier to discover relationships between the data inside the database and data that exists elsewhere. When data is uniform, it is:
- Organized: Data analysts can easily find tables and files, monitor the creation and alteration of assets, and store metadata.
- Classified: Data analysts can categorize data when it follows a consistent format, which is beneficial in cleaning and processing data.
- Stored: Consistent and uniform data can be efficiently stored in various data repositories. This streamlines storage management tasks such as managing a database.
- Accessed: Users, applications, and systems can efficiently locate and use data.

Together, these benefits empower data analysts to effectively analyze and interpret their data.

Metadata repositories

Metadata repositories help data analysts ensure their data is reliable and consistent.

Metadata repositories are specialized databases specifically created to store and manage metadata. They can be kept in a physical location or a virtual environment—like data that exists in the cloud. Metadata repositories describe where the metadata came from and store that data in an accessible form with a common structure. This provides data analysts with quick and easy access to the data. If data analysts didn't use a metadata repository, they would have to select each file to look up its information and compare the data manually, which would waste a lot of time and effort.

Data analysts also use metadata repositories to bring together multiple sources for data analysis. Metadata repositories do this by describing the state and location of the data, the structure of the tables inside the data, and who accessed the user logs.

Metadata of external databases

Data analysts use both second-party and third-party data to gain valuable insights and make strategic, data-driven decisions. Second-party data is data that's collected by a group directly from the group's audience and then sold. Third-party data is provided by outside sources that didn't collect it directly. The providers of this data are not its original collectors and do not have a direct relationship with any individuals to whom the data belongs. The outside providers get the data from websites or other programs that pull it from the various platforms where it was originally generated. Data analysts should understand the metadata of external databases to confirm that it is consistent and reliable. In some cases, they should also contact the owner of the third-party data to confirm that it is accessible and available for purchase. Confirming that the data is reliable and that the proper permissions to use it have been obtained are best practices when using data that comes from another organization.

Key takeaways

Metadata helps data analysts make data-driven decisions more quickly and efficiently. It also ensures that data and databases are reliable and consistent.

Metadata repositories are used to store metadata—including data from second-party and third-party companies. These repositories describe the state and location of the metadata, the structure of the tables inside it, and who has accessed the repository. Data analysts use metadata repositories to ensure that they use the right data appropriately.