



**DOCUMENT  
ENGINEERING**

# CMPSC 104 – Document Engineering

Prof. Hang Zhao



ALLEGHENY COLLEGE

# Report of Student Experience (RSE) Open Now

- Response rate >88% - everyone gets 2 bonus points
- Response rate >77% - everyone gets 1 bonus point

# Overview

## **What is Metadata?**

- Metadata is data that provides information about other data.
- Metadata helps in organizing, finding, and understanding data.
- Examples: Author, date created, file size, etc.

# Types of Metadata

- Descriptive Metadata: provides details about the content for identification and discovery.

**Examples:**

- Title
- Author
- Keywords
- Summary

**Use Case:** Library catalogs for searching books.

# Types of Metadata

- Structural Metadata: describes how components of a dataset or document are organized.

## **Examples:**

- Table of contents
- Chapter layout
- File relationships (e.g., a set of images within a folder)

• **Use Case:** Digital archives arrange files in a structured folder system.

# Types of Metadata

- Administrative Metadata: Provides information for managing a resource

## **Examples:**

- File type
- Creation date
- Permissions and rights

**Use Case:** Managing access and rights to digital resources.

# Types of Metadata

- Technical Metadata: details technical aspects of a resource.

**Examples:**

- Resolution of an image
- File size
- Compression format

**Use Case:** Optimizing data storage and transfer.

# Uses of Metadata

- Data Management: Helps in organizing and managing large amounts of data.
- Search and Retrieval: Facilitates the discovery of data through search engines (search engine optimization (SEO)).
- Data Sharing and Reuse: Ensures data can be shared and reused effectively between systems.
- Supports Preservation: Maintains usability over time.
- Enhances Security: Tracks access and modifications.



# Metadata in Different Domains

- Libraries: Cataloging books and resources

## **Books:**

- Title: "To Kill a Mockingbird"
- Author: Harper Lee
- ISBN: 978-0-06-112008-4
- Genre: Fiction

- Web Development: HTML meta tags for search engine optimization (SEO).

## **Web Pages:**

- Title Tag: "Programming - Beginner's Guide"
- Meta Description: "A step-by-step guide to learning programming from scratch."
- Keywords: "coding, programming, beginner guide"

# Metadata in Different Domains

- Databases: Schema information, data types, constraints.

## **Photos:**

- File Name: sunset.jpg
- Resolution: 1920x1080
- Date Taken: 2023-10-05
- Camera: Canon EOS 5D Mark IV

## **Music Files:**

- Track Name: "Imagine"
- Artist: John Lennon
- Album: Imagine
- Duration: 3:01

- Data Science: Data dictionaries, dataset descriptions.

- **Name:** Sales Data for Q4 2024
- **Purpose:** Analyze product performance and revenue.
- **Number of Records:** 10,000
- **Source:** Internal sales database.

# Challenges and Future of Metadata

## **Metadata Standards Fragmentation:**

The lack of universal metadata standards creates interoperability issues across platforms.

- Different metadata schemas like Dublin Core, MARC, and JSON-LD are not easily compatible, complicating data sharing.

## **Keeping Metadata Up-to-Date:**

Metadata becomes obsolete when underlying content changes but the metadata is not updated.

- **Example:** Outdated product descriptions in e-commerce systems lead to incorrect search results and poor user experience.

# Challenges and Future of Metadata

## **Automation in Metadata Generation:**

Leveraging AI and machine learning to automatically generate and update metadata.

- **Example:** YouTube auto-generates tags and descriptions for uploaded videos, improving searchability.

# Meta Tags

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="UTF-8">
    <title>CMPSC104 HTML Website</title>
  </head>
  <body>
    <h1>Meta Tags</h1>
  </body>
</html>
```

# Meta Tags

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="UTF-8">
    <meta name="description" content="Your Description">
    <meta name="author" content="Hang Zhao">
    <meta name="keywords" content="HTML, health, policy">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>CMPSC104 HTML Website</title>
  </head>
  <body>
    <h1>Meta Tags</h1>
  </body>
</html>
```

# Reference and Resources

- HTML W3School ([https://www.w3schools.com/html/html\\_intro.asp](https://www.w3schools.com/html/html_intro.asp))
- [Metadata Example: Types, Applications, and Importance in Data](#)
- [Automated Metadata Management: Why It is Critical for You](#)
- [What Is Metadata? Types, Frameworks & Best Practices Explained](#)