Integrative Programming & Technologies
IT 12043

Lecture 03

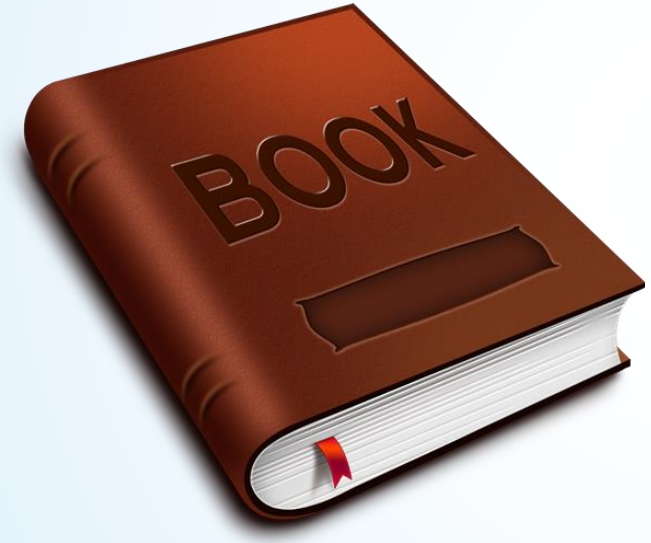
Out line

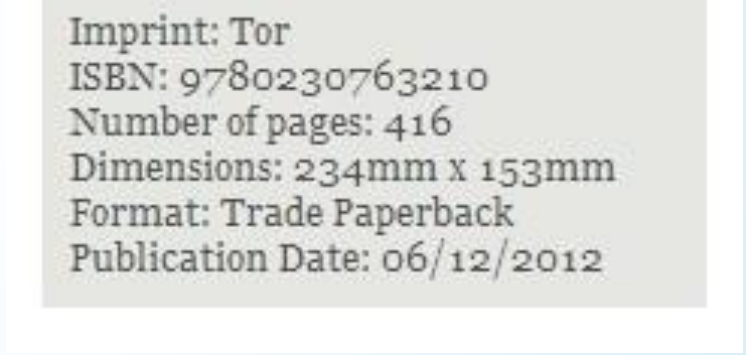
- Data mapping and exchange
  - Metadata
  - Metadata storage
- Data representation and encoding

Data Mapping and Exchange

Introduction to Metadata

- What is Metadata ?





Metadata

- Its Data about Data !
- Metadata is any kind of information that describes something else.
- The term refers to any data used to aid the **identification, description and location** of networked electronic resources

Metadata Examples

- Image may include metadata that describes
  - How large the picture is,
  - The color depth of the image,
  - The image resolution,
  - When the image was created,
  - and other data.

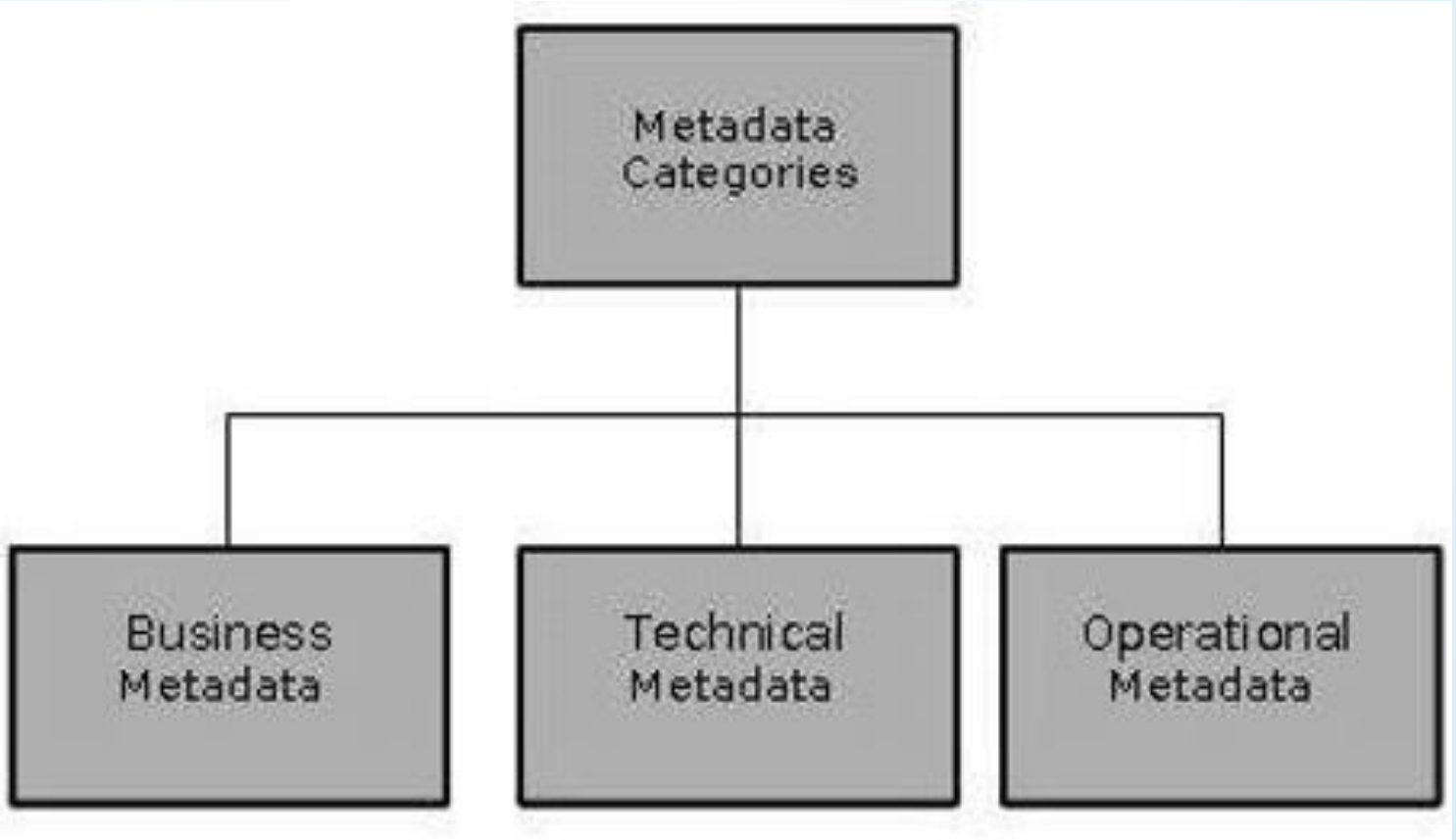
- Text document's metadata may contain information about
  - How long the document is,
  - Who the author is,
  - When the document was written, and
  - A short summary of the document.

Data about Data

- Does data about data mean anything?
  - Librarians equate it with a complete bibliographic record.
  - Information Technologists equate it to database schema or definitions of the data elements.
  - Archivists include context information, restrictions and access terms, index terms, etc.

Categories of Metadata

- Metadata can be broadly categorized into three categories



Business Metadata

- It has the data ownership information, business definition, and changing policies.

Technical Metadata

- It includes database system names, table and column names and sizes, data types and allowed values.
- Technical metadata also includes structural information such as primary and foreign key attributes and indices.

Operational Metadata

- It includes currency of data and data lineage.
- Currency of data means whether the data is active, archived, or purged.
- Lineage of data means the history of data migrated and transformation applied on it.

Bibliographic Metadata

- Providing a description of the information package along with other information necessary for management and preservation
- Encoding
- Providing access to this description

Encoding

- Surrogate records are encoded by assigning tags, letter, or words
- Why encode?
  - For display
  - Provide access
  - Integration of surrogate
  - Management

Beyond Discovery and Retrieval

- Gilliland-Swetland (1998) explains
  - "Metadata also documents how that objects behaves, its functions and use, relationship to other objects and how it should be managed".
- Another definition proposed by Cunningham
  - Structured information that describes and/or allows us to find, manage, control, understand or preserve other information over time.

Metadata to Information Technologists

- The data that defines the data elements in a table.
- Data that controls or explains other data.
- Something that is not part of the bit stream of a record but needed to understand the data in the record.
- One systems metadata is another systems data.

Source of Metadata

- Automatically generated
- Supplied by creator of electronic resource
- Supplied by 3rd party

Metadata

- Provides a means of encoding and exchanging metadata
- **EAD - Encoded Archival Description**

  Communication standard for description of modern archival records

- **TEI - Text Encoding Initiative**
  - A consortium which collectively develops and maintains a standard for the representation of texts in digital form

TEI < Text Encoding Initiative >

Storage

- Metadata can be stored either *internally*, in the same file or structure as the data (this is also called *embedded metadata*),
- or *externally*, in a separate file or field from the described data.
- A data repository typically stores the metadata *detached* from the data, but can be designed to support embedded metadata approaches.
- Each option has advantages and disadvantages

Storage

- Internal storage means metadata always travels as part of the data they describe; thus, metadata is always available with the data, and can be manipulated locally.
- This method creates redundancy (precluding normalization), and does not allow managing all of a system's metadata in one place.
- It arguably increases consistency, since the metadata is readily changed whenever the data is changed

Storage

- External storage allows collocating metadata for all the contents, for example in a database, for more efficient searching and management.
- Redundancy can be avoided by normalizing the metadata's organization.
- In this approach, metadata can be united with the content when information is transferred, for example in Streaming media; or can be referenced (for example, as a web link) from the transferred content.

Storage

- Metadata can be stored in either human-readable or binary form.
- Storing metadata in a human-readable format such as XML can be useful because users can understand and edit it without specialized tools.
- However, text-based formats are rarely optimized for storage capacity, communication time, or processing speed.
- A binary metadata format enables efficiency in all these respects, but requires special software to convert the binary information into human-readable content.

Database management

- Each relational database system has its own mechanisms for storing metadata.
  - Examples of relational-database metadata include: Tables of all tables in a database, their names, sizes, and number of rows in each table.
- Tables of columns in each database, what tables they are used in, and the type of data stored in each column.
- In database terminology, this set of metadata is referred to as the catalog.

END