



Lebanese University

Faculty of Information

Branch II

Dr. Mohamad Nagi

PhD in Data Mining - Computer Science

Metadata

Data Science - Semester 4

› Metadata Elements

- › The following are recommended as a minimum set of metadata elements. It is important to select a metadata standard or schema and consult that schema for complete information on each element. You may choose to use more elements based on the needs of your project.
- › **Title/Name** – Name given to the resource.
- › **Description** – A description of the resource and its spatial, temporal or subject coverage.
- › **Format** – File format, physical medium, dimensions of the resource, or hardware and software needed to access the data.

Metadatas Elements

- › **Metadata** – Description of the metadata to be provided along with the generated data and a discussion of the metadata standards used, including the version of the schema and where the schema can be found.
- › **Identifier** – A unique identification assigned to the resource.
- › **Rights Holder** – The entities or persons who hold the rights to the data.
- › **Rights** – Information about the rights held in and over the resource.
- › **Contact Information** – Identity of, and means to communicate with persons or entities associated with the data.

Metadata Standards/Schema

- › In order to be useful, metadata needs to be **standardized**. This includes agreeing on language, spelling, date format, etc. If everyone uses a different standard, it can be very difficult to compare data to other data.
- › A key component of metadata is the **schema**. Metadata schemes are the overall structure for the metadata. It describes how the metadata is set up, and usually addresses standards for common components of metadata like dates, names, and places. There are also discipline-specific schemas used to address specific elements needed by a discipline.

Metadata Standards/Schema

General purpose schema

› Dublin Core

Dublin Core is a general standard first used by libraries, and can be adapted for specific disciplines. Dryad (www.datadryad.org), a digital data repository, uses Dublin Core.

› MODS (Metadata Object Description Schema)

This descriptive metadata schema is richer than Dublin Core, and can be used on its own or as a complement to other metadata formats.



Metadata Standards/Schema

Science Schema

› Darwin Core

This metadata schema is for describing biological specimens, including their occurrence in nature as documented by observations, samples, and related information. Based on Dublin Core, this schema is used in natural history specimen collections and species observation databases.

› Ecological Metadata Language (EML)

This metadata schema is for ecological data. EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data.