Lebanese University
Faculty of Information
Branch II

Dr. Mohamad **Nagi**
PhD in Data Mining - Computer Science

Metadata

Data Science - Semester 4

DUBLIN CORE Introduction

- Finding relevant information on the World Wide Web has become increasingly problematic due to the explosive growth of networked resources. Current Web indexing evolved rapidly to fill the demand for resource discovery tools, but that indexing, while useful, is a poor substitute for richer varieties of resource description.

- An invitational workshop held in March of 1995 brought together librarians, digital library researchers, and text-markup specialists to address the problem of resource discovery for networked resources.

What is DUBLIN CORE

- The **Dublin Core Schema** is a small set of vocabulary terms that can be used to describe web resources (video, images, web pages, etc.), as well as physical resources such as books or CDs, and objects like art works.
- **Dublin Core** is an initiative to create a digital "library card catalogue" for the Web.
- **Dublin Core** is made up of 15 **metadata** (data that describes data) elements that offer expanded cataloguing information and improved document indexing for search engine programs.
- The Dublin Core Metadata Element Set is a general-purpose scheme for resource description originally intended to facilitate discovery of information objects on the Web.

What is DUBLIN CORE

- The development of official specifications related to the Dublin Core is managed by the Dublin Core Metadata Initiative (DCMI), which consists of a small, paid directorate advised by a board of trustees, and a large number of loosely organized volunteers.
- Dublin Core Metadata may be used for multiple purposes, from simple resource description, to combining metadata vocabularies of different metadata standards, to providing interoperability for metadata vocabularies in the Linked Data cloud and Semantic Web implementations.
- The Dublin Core Metadata Element Set is a standard for cross domain resource description.

DUBLIN CORE Background

- "Dublin" refers to Dublin, Ohio, USA where the schema originated during the 1995 invitational OCLC/NCSA Metadata Workshop, hosted by the Online Computer Library Centre (OCLC), a library consortium based in Dublin, and the National Centre for Supercomputing Applications (NCSA).

- "Core" refers to the metadata terms as "broad and generic being usable for describing a wide range of resources".

- The semantics of Dublin Core were established and are maintained by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, museums, and other related fields of scholarship and practice.

DUBLIN CORE Metadata Elements

A set 18 elements designed to enhance discovery and retrieval of resources.

Goals of DCME

- Simplicity of creation and maintenance
- Commonly understood semantics
- Conformance to existing and emerging standards
- International scope and applicability
- Extensibility
- Interoperability among collections and indexing system.

Why Use DUBLIN CORE

- "The scope of Dublin core is specially designed to provide a metadata vocabulary of core properties able to provide basic description about any kind of resources… regardless of any format of media specialization or cultural origin. It is important that a semantic model used for resource discovery is not dependent on the medium of the source it means to describe…"

- The Dublin Core metadata vocabulary is the result of many years of collaborative research to determine a common set of properties universal for describing any type of resource. The use of a standardized general classifications system also enables metadata of such collections to be combined and for knowledge contained within each collection to be shared.

DUBLIN CORE Level of Standards

The Dublin Core standard originally includes two levels: Simple and Qualified.

- Simple Dublin Core comprised 15 elements i.e. Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format Identifier, Source ,Language, Relation ,Coverage ,Right.
- Qualified Dublin Core included 3 additional elements i.e. Audience, Provenance and Rights Holder.

DUBLIN CORE Simple vs Qualified

- "Simple Dublin Core" is Dublin Core metadata that uses no qualifiers; only the main 15 elements of the Dublin Core Metadata Element Set are expressed as simple attribute-value pairs without any "qualifiers" (such as encoding schemes, enumerated lists of values, or other processing clues) to provide more detailed information about a resource.
- "Qualified Dublin Core" employs additional qualifiers to further refine the meaning of a resource. One use for such qualifiers are to indicate if a metadata value is a compound or structured value, rather than just a string.
- Qualifiers allow applications to increase the specificity or precision of the metadata. They may also introduce complexity that could impair the metadata's compatibility with other Dublin Core software applications.
- A "date" is one example of a DC element that has the option of being further specified to identify it as a particular kind of date (date last modified, date published, etc.).

DUBLIN CORE Elements

1. Identifier: Title
Definition: A name given to the resource.

2. Identifier: Creator
Definition: An entity primarily responsible for making the content of the resource.

3. Identifier: Subject
Definition: The topic of the content of the resource.

4. Identifier: Description
Definition: An account of the content of the resource.

DUBLIN CORE Elements

5. Identifier: Publisher
Definition: An entity responsible for making the resource available.

6. Identifier: Contributor
Definition: An entity responsible for making contributions to the content of the resource.

7. Identifier: Date
Definition: A date associated with an event in the life cycle of the resource.

8. Identifier: Type
Definition: The nature or genre of the content of the resource.

DUBLIN CORE Elements

9. Identifier: Format
Definition: The physical or digital manifestation of the resource.

10. Identifier: Identifier
Definition: An unambiguous reference to the resource within a given context.

11. Identifier: Source
Definition: A reference to a resource from which the present resource is derived.

12. Identifier: Language
Definition: A language of the intellectual content of the resource.

DUBLIN CORE Elements

13. Identifier: Relation
Definition: A reference to a related resource.

14. Identifier: Coverage
Definition: The extent or scope of the content of the resource.

15. Identifier: Rights
Definition: Information about rights held in and over the resource.

16. Identifier : Audience
Definition : A class of entity for whom the the resource is intended or useful.

DUBLIN CORE Elements

17. Identifier: Provenance
Definition : A statement of any change in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation.

18. Identifier : Right Holder
Definition : A person or organization owning and managing rights over the resource.

DUBLIN CORE

References

- Kunze, J. and T. Baker, “The Dublin core metadata elements set”,2013.
- Baker Thomas, “A Grammar of Dublin Core” ,2011.
- http://marciazeng.slis.kent.edu/metadatabasics/types.htm. Retrieved on April 12, 2017.
- http://www.kcoyle.net/jal-31-2.html . Retrieved on April 12,2017.
- http://dublincore.org. Retrieved on April 18, 2017.
- http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Retrieved on April 16, 2017.
- http://www.loc.gov/standards/metadata.html#types. Retrieved on April 16, 2017.
- https://www.practicalecommerce.com/SEO-Why-Is-Metadata-Important

DUBLIN CORE

There is two Dublin Core online Generators:

- Simple Dublin core generator:

  https://nsteffel.github.io/dublin_core_generator/generator_nq.html#date

- Advanced Dublin Core Generator:

  https://nsteffel.github.io/dublin_core_generator/generator.html