

# DCMI - Dublin Core Metadata Initiative

<https://www.dublincore.org/resources/metadata-basics/>

## Metadata Basics

**Metadata**, literally "data about data" -- specifically, **descriptive metadata** -- is structured data about anything that can be named, such as Web pages, books, journal articles, images, songs, products, processes, people (and their activities), research data, concepts, and services. Now a mainstream concept, *metadata* first trended in 1995, closely following *World Wide Web* in 1994. ("Big data" metadata about actions and transactions such as Facebook likes, phone calls, tweets, and the like, which trended in 2013, is out of scope for this brief introduction.)

**Dublin Core metadata**, or perhaps more accurately metadata "in the Dublin Core style", is metadata designed for interoperability on the basis of **Semantic Web** or **Linked Data** principles. Metadata in this style uses Uniform Resource Identifiers (URIs) as global identifiers both for the things described by the metadata and for the terms used to describe them (**vocabularies**). This style is distinguished by the [application profile](#) -- a specification detailing how well-known generic vocabularies such as **the Dublin Core** are used, constrained, or combined with more specialized vocabularies to meet the requirements of specific applications. Application profiles have been the focus of the Dublin Core community since they first trended in 2000.

## From Catalog Records to Linked Data

[The Dublin Core](#), a set of fifteen generic, widely used elements -- Creator, Contributor, Publisher, Title, Date, Language, Format, Subject, Description, Identifier, Relation, Source, Type, Coverage, and Rights -- was first drafted at a 1995 meeting in Dublin, Ohio, initially to facilitate information discovery on an explosively growing Web by embedding simple, card-catalog-like metadata in its pages. A diverse community of librarians, technologists, and researchers rallied to the idea, pursued and refined through a series of lively workshops and conferences, of achieving rough interoperability across languages and disciplines through a core of shared semantics. Successive developments in Web technology pulled this community in two directions:

**Metadata based on record formats.** Mainstream developers have used, and continue to use, vocabularies such as Dublin Core in the context of relational databases and repositories, many of which are based on XML, an Extensible Markup Language for specifying the contents of metadata records as structured documents. Implementers of record formats favor text values, closed-world quality control, top-down conformance, and reliance on well-understood, tried-and-true software solutions. Interoperability across applications is seen in terms of adherence to fixed formats such as the fifteen-element Simple Dublin Core and Qualified Dublin Core (2003), with several dozen additional DCMI metadata terms, to dozens of other formats published over the years. While the record-based approach may be relatively easy to deploy, interoperability across differently structured formats relies on ad-hoc "crosswalks" (mappings) that are hard to maintain and to use.

**Metadata based on recombinant statements.** Starting in the late 1990s, working groups of the World Wide Web Consortium pursued the vision of a web of data, or Semantic Web. This vision was enabled by [Resource Description Framework \(RDF\)](#) and by a global Domain Name System (DNS) that could resolve URIs to resources on the Web. The [first W3C Recommendation for RDF](#)

in 1999 featured annotated examples of metadata using Dublin Core, which in 2000 became [one of the first vocabularies to be published in RDF with persistent URIs](#). In the face of the messiness and complexities of the open Web, RDF implementers aim at achieving partial interoperability. In the RDF mindset, metadata consisted not of discrete, bounded records (documents) of a known structure, but of unbounded, schema-less **graphs** composed of atomic **statements** that could be recombined, or "mashed up", by merging multiple sources into the graph. In statement-based metadata, interoperability among multiple sources results from using, or mapping to, shared URIs, preferably from well-known vocabularies such as Dublin Core.

**Dublin Core-style application profiles.** Where XML implementers saw application profiles as blueprints for creating validatable metadata records within a specific application, RDF implementers saw profiles as a basis for designing metadata that would compatibly fit into data graphs spanning multiple applications. To bridge this gap, DCMI's [Singapore Framework \(2007\)](#) the ideal application profile as the sum of several best-practice design components. At its core was the notion of a **description** as a set of statements about a single resource. Descriptions of multiple resources, such as Book and Author, could be bundled into a **description set** which, in turn, could either be stored directly as an RDF graph or encoded in a format designed to be convertible into RDF. A well-designed application profile would be based on available RDF vocabularies, well-articulated entity models, and explicitly defined functional requirements. Metadata in this style need not be based on the Dublin Core but can draw on a diversity of RDF vocabularies, such as [Friend of a Friend \(FOAF\)](#), the [Bibliographic Ontology \(BIBO\)](#), and [Schema.org](#).

## Linked-Data-compatible metadata today

The uptake of metadata, and specifically of metadata based on RDF vocabularies, has been propelled by the evolution of technology:

- **Support for creating metadata.** Modern software platforms, such as the Drupal content management system and Hugo static website generator use metadata to structure the content and presentation of websites. Some can also be configured to publish Linked Data, either by embedding it in Web pages or by generating metadata feeds in Linked Data-compatible syntaxes.
- **Support for querying metadata.** Repositories of Linked Data can be queried in the manner of relational databases by using a standard query language, [SPARQL](#). This can be facilitated a **SPARQL endpoint** that accepts queries and returns results over the Web.
- **Support for indexing embedded metadata.** The 1995 vision of finding Web resources via embedded metadata first became mainstream in 2011, when Google, Bing, and Yahoo! announced support for [Schema.org](#), which aimed at helping webmasters use embedded metadata to improve the presentation of their sites in search results. The major search engines now extract and index metadata embedded with one of several syntaxes: [HTML Microdata](#), of limited expressivity but the easiest for webmasters to deploy; [RDFa](#), a richer syntax with better support for internationalization and multiple RDF namespaces; and [JSON-LD](#), an RDF-compatible variant of the popular Javascript Object Notation (JSON). These broadly supported syntaxes effectively obsolete a series of IETF and DCMI syntax specifications developed prior to 2008 specifically for expressing Dublin Core metadata.
- **Support for publishing value vocabularies as Linked Data.** The W3C standard [Simple Knowledge Organization System \(SKOS\)](#) provides a core data model for sharing taxonomies and thesauri, such as [AGROVOC](#), as Linked Data. Each concept in a *SKOS concept scheme* is identified with URI that is globally citable in metadata and can thus be used as a basis for linking, or merging, metadata from diverse sources. SKOS made it relatively easy to port a rich tradition of existing knowledge organization systems from printed books, siloed databases, and PDFs to the Semantic Web. The rapid uptake of SKOS

in the library and research worlds led to a broader acceptance of RDF as a mainstream solution for interoperability.

- **Support for collaborative metadata creation.** [Wikidata](#), a collaboratively edited knowledge base used as a source of structured open data by Wikipedia, and its open-source platform Wikibase, have made it easier to create and maintain metadata by crowdsourcing.
- **Support for validating RDF metadata.** The [Shape Expressions language \(ShEx\)](#) and the related [Shapes Constraint Language \(SHACL\)](#) now provide the ability to treat RDF graphs as objects of closed-world conformance validation in the manner of XML schemas. Application profiles can now be expressed as ShEx schemas, with domain entities modeled as ShEx shapes. Further work in the DCMI community will aim at making it easier for non-expert users to create validation schemas using familiar tools such as spreadsheets.

DCMI's [Linked Data Competency Index](#) breaks down many of these areas into sets of skills and concepts ("competencies") usable by teachers, trainers, professors, or independent learners in designing courses or for self-directed study.

## Metadata in the Second Machine Age

In his [keynote talk at DC-2016](#), Bradley Allen discussed the role of metadata in the Second Machine Age. "While the user experience of discovery has come to be dominated by search engines such as Google," he noted, "metadata standards are pervasive in the infrastructure of content curation and management, and underpin search infrastructure". In his view, a single thread runs from the establishment of the Dublin Core through Open Linked Data to the emergence of Knowledge Graphs -- graph-structured databases extracted from content with the help of machine intelligence not only for helping people find, filter, and organize information, but also for constructing answers to questions. In his vision, the design of metadata should evolve in ways that can help machines read and learn from the Web and, in turn, help make its resources easier for both machines and people to discover and use.

Unless indicated otherwise, DCMI documents are licensed under a [Creative Commons Attribution 3.0 Unported License](#). Please see the [DCMI Document Notice](#) for further instructions.

[Copyright](#) © 1995-2019 [DCMI](#). All Rights Reserved. DCMI [liability](#), [trademark/service mark](#), [document use rules apply](#). [Your interactions with this site are in accordance with our privacy statements](#).

The Dublin Core Metadata Initiative (DCMI) is a project of ASIS&T—a U.S. 501(c)(3) nonprofit under the U.S. Internal Revenue Code. Contributions to DCMI through ASIS&T are tax-deductible to the full extent of the law in the United States.

Deployed with Hugo v. 0.55.6 on Wed, 02 Oct 2019 12:52:07 UTC