**Administrative Metadata:** A type of metatdata

**Explanation**: Administrative metadata provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.

**Example**: Rights management metadata: Attribution-NonCommercial CC BY-NC

**Source**: <https://www2.archivists.org/glossary/terms/a/administrative-metadata>

**API:** Application Programing Interface

**Explanation**: API is a set of functions and procedures allowing the creation of applications that access the features or data of an operating system, application, or other service. Software intermediary that allows two applications to talk to each other. The process by which you fetch and process data.

**Examples**: Spotify, Google Maps

**Source**: <https://en.wikipedia.org/wiki/Application_programming_interface>

**ARC/GIS:** Geographic information system

**Explanation**: ARC/GIS is a geographic information system for working with maps and geographic information. It is used for creating and using maps, compiling geographic data, analyzing mapped information, sharing and discovering geographic information, using maps and geographic information in a range of applications, and managing geographic information in a database.

**Example**: http://geodata.cherokee.org/arcgis/rest/services

**Source**: <https://en.wikipedia.org/wiki/ArcGIS>

**CIDOC-CRM:** Conceptual Reference Model (CRM) of the International Committee on Documentation (CIDOC) of the International Council of Museums (ICOM)

**Explanation**: CIDOC\_CRM is an ontology for cultural heritage information. It includes definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives.

**Example**:

E9 Move.

Subclass of: E7 Activity.

Scope note: This class comprises changes of the physical location of the instances of E19 Physical Object.

**Source**: <http://www.cidoc-crm.org/>

**Computational Methods:** Mathematical models used to numerically study the behavior of complex systems by means of a computer simulation.

**Explanation**: A computational model can be used to make predictions of the system's behaviour under different conditions, often for cases in which intuitive analytical solutions are not available.

**Example**: Prediction of the position of a vehicle from its initial position, direction and speed of travel, using the equation that distance traveled is the product of time and speed. This is known as dead reckoning.

**Source**: <https://en.wikipedia.org/wiki/Mathematical_model>

**Data Curation:** The organization and integration of data possibly collected from various sources

**Explanation**: Data Curation involves annotation, publication and presentation of the data such that the value of the data is maintained over time, and the data remains available for reuse and preservation.  Broadly, curation means a range of activities and processes done to create, manage, maintain, and validate a component. Specifically, data curation attempts to determine what information is worth saving and for how long.

**Example**: A curated database: Corey, Ryan M.; Tsuda, Naoki; Singer, Andrew C. (2018): Wearable Microphone Impulse Responses. University of Illinois at Urbana-Champaign. <https://doi.org/10.13012/B2IDB-1932389_V1>

**Example:** A curated database: [http://collections.theautry.org/mwebcgi/mweb.exe?request=home#](http://collections.theautry.org/mwebcgi/mweb.exe?request=home)

**Source**: <https://datacurationnetwork.org/>

**Data Driven Scholarship**: Scholarship based on analyzing and filtering large data sets, and an interdisciplinary field of methods to extract knowledge from data

**Explanation**: One orientation of Data Drive Scholarship is to understand the nature of digital data, and the intersections between this new field and existing disciplines of human enquiry. Seeking to develop the underlying scientific and technological basis for understanding and exploiting data.

**Example**: <http://datascience.ucsd.edu/research/impact/mining-large-data-sets-of-genomic-architecture.html>

**Sources**: <https://en.wikipedia.org/wiki/Data-driven> <http://datascience.ucsd.edu/about/index.html>

**Data Mining**: The process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

**Explanation**: Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a dataset and transform the information into a comprehensible structure for further use.

**Example**: Data mining of government records – particularly records of the justice system (i.e., courts, prisons) – enables the discovery of systemic human rights violations in connection to generation and publication of invalid or fraudulent legal records by various government agencies.

**Source**: <https://en.wikipedia.org/wiki/Examples_of_data_mining>

**Dataset:** A collection of data

**Explanation**: Most commonly a dataset corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The dataset lists values for each of the variables for each member of the dataset. Each value is known as a datum.

**Example**: <https://github.com/LA-Autry/Collections-as-data>

**Sources**: <https://conservancy.umn.edu/handle/11299/200840>

<https://en.wikipedia.org/wiki/Data_set#Classic_data_sets>

**Descriptive Metadata**: A type of metadata

**Explanation**: Descriptive metadata describes a resource for purposes such as discovery and identification.

**Example**: author, date published, country of publication

**Source**: <http://marciazeng.slis.kent.edu/metadatabasics/types.htm>

**Digital Humanities**: An academic field concerned with the application of computational tools and methods to traditional humanities disciplines such as literature, history, and philosophy.

**Explanation**: Digital Humanities brings digital tools and methods to the study of the humanities.

**Example**: Visualizing Emancipation <http://dsl.richmond.edu/emancipation/>

**Source**: <https://en.wikipedia.org/wiki/Digital_humanities>

**Dublin Core**: A set of vocabulary terms that can be used to describe digital resources (video, images, web pages, etc.), as well as physical resources such as books, and objects like artworks.

**Explanation**: 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.

**Example**: <meta name="DC.Publisher" content="publisher-name" />

**Source**: <https://en.wikipedia.org/wiki/Dublin_Core>

**GIT-HUB**: A web-based code sharing and publishing service

**Explanation**: A user must create an account in order to contribute content to the site, but public repositories can be browsed and downloaded by anyone. A repository is a location where all the files for a particular project are stored. Each project has its own repository, and you can access it with a unique URL. With a registered user account, users are able to have discussions, manage repositories, submit contributions to others' repositories, and review changes. With distributed version control systems like GIT, to make changes to a project, you copy the whole repository to your own system. You make your changes on your local copy, then you “check in” the changes to the central server.

**Example**: <https://github.com/LA-Autry/Collections-as-data>

**Source**: <https://www.howtogeek.com/180167/htg-explains-what-is-github-and-what-do-geeks-use-it-for/>

**IRI**: International(ized) Resource Identifier

**Explanation**: A generalization of a URI. While URI supports only ASCI encoding, IRI fully supports international characters. In practice, UTF-8 is the most popular encoding used for IRI.

**Example**:  [https://en.wiktionary.org/wiki/Ῥόδος](https://en.wiktionary.org/wiki/%E1%BF%AC%CF%8C%CE%B4%CE%BF%CF%82)

**Source**: <http://www.w3.org/People/Dürst>

**JSON**: Java Script Object Notation

**Explanation**: JSON is a format for storing and transporting data. JSON is often used when data is sent from a server to a web page. JSON is “self-describing.” Data is in name/value pairs. Data is separated by commas. Curly braces hold objects. Square brackets hold arrays.

**Example**:

{"employees":[

{"firstName":"John", "lastName":"Doe"},

{"firstName":"Anna", "lastName":"Smith"},

{"firstNam":"Peter", "lastName":"Jones"}]}

**Source**: <https://www.w3schools.com/whatis/whatis_json.asp>

**JSON-LD**: JavaScript Object Notation for Linked Data

**Explanation**: A method of encoding Linked Data using JSON. JSON-LD is designed around the concept of a “context” to provide additional mappings from JSON to an RDF model. The context links object properties in a JSON document to concepts in an ontology.

**Example**:

{

“@context”: {

“name”: “http://xmlns.com/foaf/0.1/name”,

“homepage”: {

“@id”: “http://xmlns.com/foaf/0.1/workplaceHomepage”,

“@type”: “@id”

},

“Person”: “http://xmlns.com/foaf/0.1/Person”

},

“@id”: “https://me.example.com”,

“@type”: “Person”,

“name”: “John Smith”,

“homepage”: “https://www.example.com/”

}

**Source**: <https://en.wikipedia.org/wiki/JSON-LD>

**Karma:** Information integration tool

**Explanation:** Karma enables users to quickly and easily integrate data from a variety of data sources including databases, spreadsheets, delimited text files, XML, JSON, KML and Web APIs.  Users integrate information by modeling it according to an ontology of their choice using a graphical user interface that automates much of the process. Once the model is complete, users can published the integrated data as RDF or store it in a database.

**Example:** <http://usc-isi-i2.github.io/slides/szekely12-vivo-slides.pdf>

**Source:** https://github.com/usc-isi-i2/Web-Karma/wiki

**Key Field:** A field in a record that holds unique data which identifies that record from all the other records in the file or database.

**Explanation:** As an identifier, each **key** value must be unique in each record.

**Example**: Account number, product code and customer name are typical keyfields. As an identifier, each key value must be unique in each record.

**Source:** <https://www.yourdictionary.com/key-field>

**Linked Data**: A set of design principles for sharing machine-readable interlinked data on the Web.

**Explanation**: Linked Data is one of the core pillars of the Semantic Web, which is about making links between datasets that are understandable not only to humans, but also to machines, and Linked Data provides the best practices for making these links possible, for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF.

**Example:**

<http://experiment.worldcat.org/entity/work/data/63444562#Place/wyoming> # Wyoming a schema:Place ; schema:name:”Wyoming”;

**Source**: <https://en.wikipedia.org/wiki/Linked_data>

**Machine Learning**: The scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead.

**Explanation**: A subset of artificial intelligence, machine learning algorithms build a mathematical model of sample data in order to make predictions or decisions without being explicitly programmed to perform the task.

**Example**: Medical Diagnosis: the analysis of the clinical parameters and their combination for the prediction of disease progression, for the extraction of medical knowledge for the outcome research, and for therapy planning and patient monitoring.

**Source**: <https://en.wikipedia.org/wiki/Machine_learning>

**Metadata**: Data [information] that provides information about other data.

**Explanation**: Many distinct types of metadata exist, among these descriptive metadata, structural metadata, administrative metadata, reference metadata and statistical metadata.

**Example**: Date created, date modified, file type, file size

**Source:** <https://whatis.techtarget.com/definition/metadata>

**METS**: Metadata Encoding and Transmission Standard

**Explanation**: The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium.

**Example**:

"<mets>

<dmdSec>

<mdWrap>

<xmlData>

<!-- insert data from different namespace here -->

</xmlData>

</mdWrap>

</dmdSec>

<fileSec></fileSec>

<structMap></structMap>

</mets>"

**Source**: <http://www.loc.gov/standards/mets/mets-present.html#overviews>

**N-Triples**: N-Triples is a line-based, plain text format for encoding an RDF graph

**Explanation**: N-Triples triples are a sequence of RDF terms representing the subject, predicate and object of an RDF Triple.

**Example**:

<http://one.example/subject1>

<http://one.example/predicate1>

<http://one.example/object1> . # comments here

# or on a line by themselves

\_:subject1 <http://an.example/predicate1> "object1" .

:subject2 <http//an.example/predicate2> "object2" .

**Source**: <https://www.w3.org/TR/n-triples/>

**Ontology**: A specification or framework of concepts with their definitions and relationships within a particular domain or discipline.

**Explanation**: Ontologies are considered the fundamental building blocks for the semantic web. They are vocabularies or sheme vocabularies (established sets of metadata elements and other contructs express as properties and classes.

**Example**: CIDOC-CRM is an ontology.

**Source**: Project Glossary by Melissa Gill, with Kristen Carter and Matther Lincoln

**Open Data**: Data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share.

**Explanation**: Availability and access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form. Re-use and redistribution: the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.

**Example**: <http://mapshare-uclagisadmin.opendata.arcgis.com/>

**Source:** <http://opendatahandbook.org/guide/en/what-is-open-data/>

**Open Linked Data**: A blend of Linked Data and Open Data: it is both linked and uses open sources.

**Explanation**: “Open Linked Data is Linked Data that is offered under an open license, which does not prevent its reuse and It can be done for free” – Tim Berners Lee.

**Example**: One notable example of an LOD set is DBpedia – a crowd-sourced community effort to extract structured information from Wikipedia and make it available on the Web. <https://wiki.dbpedia.org/>

**Source**: <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/>

**QGIS**: Geographic information system

**Explanation**: QGIS is a free and open-source cross platform desktop (GIS) application that supports viewing, editing, and analysis of geospatial data.

**Example**: <https://www.flickr.com/photos/59811265@N04/40252115253/in/pool-2244553@N22>

**Source**: <https://en.wikipedia.org/wiki/QGIS>

**RDF**: Resource Description Framework

**Explanation**: RDF is a standard model for data interchange on the Web. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “triple”). An RDF statement reflects a relationship between 2 resources = subject and object, and the relationship is the predicate. RDF data can be serialized as RDF-XML, N-triples, Turtle, and JSON-LD formats.

**Example**:

Bob Knows John:

-- <http://example.name#BobSmith12>

-- <http://smlns.com/foaf/0.1/knows>

-- <http://example.name#JohnDoe34>

**Source**: <https://www.w3.org/RDF/>

**RDF-XML**: Resource Description Framework XML

**Explanation**: RDF-XML provides an XML syntax for RDF graphs. In RDF/XML RDF triples are specified within an XML element.

**Example**:

<?xml version="1.0"?>  
<rdf:RDF  
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
xmlns:si="https://www.w3schools.com/rdf/">  
<rdf:Description rdf:about="https://www.w3schools.com">  
  <si:title>W3Schools</si:title>  
  <si:author>Jan Egil Refsnes</si:author>  
</rdf:Description>  
</rdf:RDF>13

**Source**: <https://www.w3.org/TR/rdf11-primer/#section-rdfxml>

**SPARQL**: Semantic query language for retrieving and manipulating RDF data

**Explanation**: Declarative Query Language (like SQL) for performing Data Manipulation and Data Definition operations on Data represented as a collection of RDF Language sentences/statements. SPARQL allows users to write queries against what can loosely be called “key-value” data or, more specifically, data that follow the RDF specification of the W3C. Thus, the entire database is a set of “subject-predicate-object” triples.

**Example**:

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?name

?email

WHERE

{

?person a foaf:Person .

?person foaf:name ?name .

?person foaf:mbox ?email .

}

**Source**: <https://en.wikipedia.org/wiki/SPARQL>

**SPARQL Endpoint**: A SPARQL Endpoint is a Point of Presence on an HTTP network that’s capable of receiving and processing SPARQL Protocol requests. It is identified by a URL commonly referred to as a SPARQL Endpoint URL.

**Explanation**: A SPARQL endpoint is paired with an HTML document that functions as a simple interface for query editing and execution. By convention, the URL of this kind of document includes the literal sparql as the final path component or as the host part of a web site’s canonical name.

**Example**: http://dbpedia.org/sparql  — DBpedia (nucleus of the LOD Cloud)

**Source**: <https://medium.com/virtuoso-blog/what-is-a-sparql-endpoint-and-why-is-it-important-b3c9e6a20a8b>

**Statistical Metadata**: Type of metatdata

**Explanation**: Statistical metadata may also describe processes that collect, process, or produce statistical data; such metadata are also called process data.

**Example**: Statistical data consists of the following:

Microdata — data on the characteristics of units of a population, such as individuals, households or establishments, collected by a census, survey, or experiment.

Macrodata — data derived from microdata by statistics on groups or aggregates, such as counts, means, or frequencies.

**Source**: <https://www.census.gov/prod/2/gen/96arc/viiblapl.pdf>

**Structural Metadata**: Type of metatdata

**Explanation**: Structural metadata is about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships and other characteristics of digital materials.

**Example**: In a multi-file digital object (e.g., a scanned book with many page images), structural metadata describes the object's components and their relationships: pages, chapters, tables of contents, index, etc. Such metadata can support sophisticated search and retrieval actions as well as the navigation and presentation of digital objects. METS offers one model for the encoding of structural metadata.

**Source**: <http://www.loc.gov/standards/mets/>

**Text Mining**: Text mining, also referred to as text data mining, is roughly equivalent to text analytics, is the process of deriving high-quality information from text.

**Explanation**: High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output.

**Example**: [Rescued history: Massive text data analysis helps uncover black women's experiences](https://www.nsf.gov/mobile/discoveries/disc_summ.jsp?cntn_id=137797&org=NSF) - Researchers used high performance computers to analyze 20,000 documents from the HathiTrust and JSTOR databases that were known to contain information about black women. This analysis was used to create a computational model based on this corpus of documents which they then used to study the entire 800,000 documents in both databases. To make sense of the huge datasets, the investigators used computational techniques of topic modeling and data visualization.

**Source**: <https://en.wikipedia.org/wiki/Text_mining>

**Triple**: An RDF statement consisting of two resources (a subject and an object) and the relationship between them (a predicate).

**Explanation**: Triples may contain values reprsented by character strings, called literals, or by URIs/URLs, called non-literals.

**Example**:

<Object><is created by><Actor> ; Literal: Gene Autry ; non-literal: <http://viaf.org/viaf/113310336>

**Source:** https://en.wikipedia.org/wiki/Semantic\_triple

**Turtle**: Terse RDF Triple language. A syntax and file format for expressing data in the Resource Description Framework (RDF) data model.

**Explanation**: Turtle syntax is similar to that of SPARQL, an RDF query language.

RDF represents information using semantic triples, which comprise a subject, predicate, and object. Each item in the triple is expressed as a Web URI. Turtle provides a way to group three URIs to make a triple, and provides ways to abbreviate such information, for example by factoring out common portions of URIs.

**Example**:

<http://example.org/person/Mark_Twain>

<http://example.org/relation/author>

<http://example.org/books/Huckleberry_Finn>

**Source**: <https://www.computerhope.com/jargon/t/turtle.htm>

**URI**: Uniform Resource Identifier.

**Explanation**: A string of characters that unambiguously identifies a particular resource.

Such identification enables interaction with representations of the resource over a network, typically the World Wide Web, using specific protocols. Schemes specifying a concrete syntax and associated protocols define each URI. The most common form of URI is the Uniform Resource Locator (URL), frequently referred to informally as a web address.

**Examples:**

scheme:[//authority][/path][?query][#fragment]

tel:+1-816-555-1212

**Source:** <https://www.baeldung.com/java-url-vs-uri>

**URL**: Uniform Resource Locator.

**Explanation**: A type of URI that identifies the resource and the mechanism by which the resource may be accessed. Most common is a web address beginning http:

**Example**: <http://autry.iii.com>

**Source**: <https://en.wikipedia.org/wiki/URL>

**URN**: Uniform Resource Name

**Explanation**: URN is a form of URI that specifies the name of a resource. It does not assume any given mechanism for retrieving the resource, but it does identify it.

**Example**: urn:isbn:0451450523

**Source**: <https://stackoverflow.com/questions/723934/example-of-a-uri-that-isnt-a-url>

**Value vocabularies**: Traditionally controlled vocabularies

**Explanation**: Includes name authorities, subject headings, and thesauri.

**Examples**: "LCSH, LCNAF"

**Source**: <https://authorities.loc.gov/>

**VIAF:** Virtual International Authority File

**Explanation**: The VIAF (Virtual International Authority File) service provides libraries and library users with convenient access to the world’s major name authority files. VIAF Contributors regularly supply authority data that VIAF matches, links, and groups. All descriptions for a given entity are merged into a cluster that brings together the different names for that entity. This service allows researchers to identify names, locations, works, and expressions while preserving regional preferences for language, spelling, and script.

**Example**: <http://viaf.org/viaf/66548614/>

**Source**: <https://www.oclc.org/en/viaf.html>