2024

# Workbook of Metadata Fundamentals

## The What, Why, and How of Metadata

Data Management Center Of Excellence, Office of Data Science Strategy

VERSION 1.1

# Contents

# Purpose of this Workbook

This workbook is intended to assist researchers, curators, and data stewards responsible for managing or developing metadata[1] for data resources generated in biomedical or behavioral and social sciences research. It targets individuals with a skill level ranging from introductory to intermediate.

Proficiency in developing effective metadata only comes with knowledge and practice. To achieve mastery, it is essential to understand the underlying concepts and theory, acquire knowledge of the data domain, be familiar with the relevant best practices, and hone your skills with software tools for working with metadata. This can be overwhelming for beginners, so to lighten the burden, this workbook contains core knowledge, tables, flowcharts, templates, standards, instructional videos, software tools, and self-assessments to help you on your metadata proficiency journey.

The workbook is organized into three chapters covering: preliminaries on metadata, working with metadata, and specific metadata schemas. It has useful footnotes or links to further a journey into the study and application of metadata, a glossary of terms used in the workbook, supplementary technical references and training materials, and a list of acronyms. A beginner will grow confident working with metadata after putting this knowledge into practice.

**Disclaimers**

Any mention of scientific methodologies; experimental techniques, tools, data or metadata formats; instrumental technologies; and repositories used in the training material do not constitute any form of endorsement by National Institutes of Health and should merely be interpreted as examples.



[1] Metadata is a broad term that encompasses a wide variety of specific types of information that is either created or captured about a data or an object resource.

# 1.0 Preliminaries on Metadata

## 1.1 Learning Objectives

After working through this chapter, you will be able to:

- Understand what metadata is and its importance
- Describe how metadata helps make data Findable Accessible Interoperable and Reusable (FAIR)
- Define several categories of metadata and associated metadata standards
- Appy best practices when working with metadata
- Become aware of the distinction between repositories and registries for metadata
- Get acquainted with metadata standards

## 1.2 What is Metadata?[2]

In today's digital age, data is everywhere. But how do we locate data of interest and understand how to use or reuse it? If we cannot find the data, it may not matter how accessible, interoperable, or reusable it is! Metadata is the key to finding the data and unlocking its value.

Metadata is often defined as data about data or a resource. While technically correct, this definition doesn't capture its value or significance. Instead, metadata is better defined as "deliberate, structured data about a data or an object resource."

While the data resource can be a set of facts, a collection of images, a string of words, or a description of something, metadata provides meaningful information about the resource. Data can be "raw" or unprocessed and may require hardware, software, or additional documentation to understand and use; however, metadata is always processed, usually understandable by both humans and machines, and only created for a functional purpose.

## 1.3 Why Metadata Matters

Metadata provides context for a resource, enhancing its utility. It is not only essential in identifying and locating a resource, but it also establishes provenance and can help define relations to other resources.

Metadata maximizes the utility of a resource data to both humans and machines and can help determine whether the resource data is interoperable with other resources' data.

Depending on the end use of the resource, some types of metadata may be more valuable than others. Basic characteristics of metadata for a data resource may include aspects such as a persistent identifier for the data, who created the data, what the data file contains; and when, where, why, and how the data were generated. *In this workbook, we focus on data resources, but many of the concepts we develop for this type of resource may also be applicable to other types of resources.*

Briefly, metadata supports:

- **Data Discovery** – One of the primary functions of metadata is resource discovery to make it easier for users and search engines to find it.

[2]National Information Standards Organization (NISO) describes metadata as: "Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource."

- **Data Organization** – Metadata plays a crucial role in organizing and understanding data. By providing context, metadata enables users to navigate digital assets and comprehend their relationships with other data elements.
- **Defines Tools** – Tools may be used to view, manipulate, and use the resource.
- **Data Security** – Metadata protects sensitive information and may provide information on access rights, authentication, and encryption.

Metadata touches upon multiple aspects of rendering a data resource **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (FAIR).[3] FAIR Guiding Principles[4] listed below document the role of both metadata and data on locating, sharing, and reusing the resource. It is notable that metadata is very highly prevalent in this list (13 of 16 items), indicating that it is highly relevant and essential to FAIR.

**"To be Findable:**

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

**To be Accessible:**

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

**To be Interoperable:**

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

**To be Reusable:**

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards"

## 1.4 Metadata Categories and Standards[5]

Listed below are some commonly used metadata categories and standards: such as those that define unambiguously schema structures, vocabulary, values, and how data will be encoded.

**Metadata Categories**

Many informaticists consider components of metadata of a resource to fall within three high-level categories: descriptive, structural, and administrative.

- Descriptive metadata describes a resource, its content, and identifying characteristics.

[3] Musen, M.A., O'Connor, M.J., Schultes, E. *et al.* Modeling community standards for metadata as templates makes data FAIR. *Sci Data* **9**, 696 (2022). https://doi.org/10.1038/s41597-022-01815-3

[4] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[5] Steven Jack Miller, Metadata for Digital Collections; 2nd Edition (2022)

- Structural metadata describes how the pieces of a single resource fit together and/or how a resource exists in relation to other resources.
- Administrative metadata includes information about rights, technical aspects, and preservation.

**Metadata Standards**[6]

Metadata standards fall into four general categories:

- *Structure Standards* – Structured organization of metadata is often referred to as its model or schema. These schemas can differ widely: from domain-agnostic, such as Simple Dublin Core, to domain-specific, such as Darwin Core for Biodiversity, or Data Documentation Initiative (DDI) for describing observational and survey data in the social, behavioral, economic, and health sciences. The ability to locate resources in a repository with any specificity and sensitivity[7] is strengthened/enhanced by the metadata schema it utilizes.
- *Content Standards* – Rules or usage guidelines for how to input data into the components of the metadata (such as controlled vocabularies or taxonomies).
- *Value Standards* – List of established standardized terms and codes used in the metadata.
- *Data Exchange Standards* – Specifications for encoding data.

Metadata expressed in defined schemas, standard vocabularies, and persistent identifiers, form the basis for easy, unambiguous interpretation by both humans and machines.[8]

## 1.5 Metadata Registries and Repositories[9]

In your journey with metadata, you will not only encounter data and metadata used by repositories, but also metadata registries and repositories that solely manage metadata. You should clearly understand the distinction between a metadata registry and metadata repository.

- *Metadata (Schema) Registry* – Where *metadata definitions* are stored and maintained in a controlled manner. Examples include Schema.org.[10]
- *Metadata Repository* – Where actual *instance metadata* is stored, such as in a data catalog. The instance metadata should ideally be both in a human and machine-readable form such as JavaScript Object Notation (JSON)-LD Extensible Markup Language (XML) and Resource Description Framework (RDF). For example, the National Library of Medicine (NLM) Dataset Catalog[11] provides a standardized description of biomedical dataset information from relevant dataset repositories and some datasets that are available from multiple repositories.

## 1.5 Key Elements of Metadata

Typically, attempt to ensure you capture the following minimum elements in your metadata:

- *Title/Name* – Name given to the resource
- *Description* – Description of the resource and its spatial, temporal, or subject coverage
- *Format* – File format, physical medium, dimensions or file size of the resource, and/or hardware and software needed to access the data

---

[6] Metadata Standards - Metadata & Discovery @ Pitt - Guides at University of Pittsburgh (libguides.com)

[7] See https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6148622/ for sensitivity and specificity in searching.

[8] Batista, D., Gonzalez-Beltran, A., Sansone, SA. et al. Machine actionable metadata models. Sci Data 9, 592 (2022). https://doi.org/10.1038/s41597-022-01707-6

[9] Metadata registry - Wikipedia

[10] https://schema.org/

[11] https://www.datasetcatalog.nlm.nih.gov/

- *Version* – File version of data or dataset
- *Schema* – Version of the schema and where the schema can be found
- *Identifier* – A unique identification code, such as a Digital Object Identifier (DOI), assigned to the resource, usually generated by the repository.
- *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, the persons or entities associated with the data.

## 1.6 Which Metadata Standards to Use[12]

The choice of repository in which you plan to store your data dictates the applicable metadata schema(s) standards. Generic ones, such as the simple Dublin Core, tend to be easy to use and are widely adopted, but often need to be expanded to cover more specialized, domain-specific information. Domain-specialized schemas have a much richer vocabulary and structure, but tend to be highly specialized and only understandable by practitioners in field. Always think of the future end users when selecting a schema supported by the repository. Adopt a schema that simplifies data location and usage for users, while also aiding repository staff in management and preservation.

Consult documentation on the elements of a schema for complete information. You may choose to use more elements based on the needs of your project, the expectations of your discipline, and the flexibility available in the repository to include additional elements.

In general, you should be able to find a suitable metadata schema to satisfy your specific needs. When you do find one, use it. If you find one that is close to your needs, but not quite perfect, you can consider extending or modifying it to suit your needs. This is best done by working with the organization that created the standard to make a case for inclusion of additional elements to support your data scenario or use case.

## 1.7 Metadata Best Practices[13]

The following are some useful metadata best practices for your consideration:

- Adopt appropriate and consistent metadata standards such as structural, content, value or data exchange.
- Use tools for creating and validating metadata.[14,15]
- Create metadata prospectively, i.e. before the data is created or collected, when possible. Creating prospective metadata can have advantages, such as reminding you of aspects of the resource that need to be captured during the collection process. You will capture other metadata retrospectively; after the resource or data is created.
- Make metadata both human and machine-readable. Machine readability supports the creation and analysis of big data by facilitating linking and aggregating multiple data sources.
- If possible, reuse or extend community-endorsed metadata by working with the relevant organization rather than creating your own variant.

---

[12] Standards/Schema - Metadata for Data Management: A Tutorial - LibGuides at University of North Carolina at Chapel Hill (unc.edu)

[13] https://www.getty.edu/publications/intrometadata/practical-principles/

[14] CEDAR Metadata Center

[15] Schema.org - Schema.org

- When it is beneficial, use different metadata schemas for the same data or resource to support different types of user needs. For example, if you are working with image data, the metadata may need to be different to meet the needs of biomedical researchers versus computer scientists. This does not mean you need to store the actual data in more than one repository.
- Create high quality metadata. Metadata quality is just as important as data quality.
- Consult experts as needed. A variety of expertise and skills may be required to create effective metadata. These include domain expertise, knowledge of controlled vocabularies, technical knowledge, research skills, knowledge usage rights.
- Carefully choose the appropriate suite of metadata schemas and controlled vocabularies to best describe and provide optimal discoverability and/or reusability of your data.

## 1.8 Learning Summary

Concepts covered: value of metadata, repositories and registries, standards and best practices.

## 1.9 Knowledge Self-Assessments

**Questions**

1. What is the primary function of metadata?
    a. Data encryption
    b. Resource discovery
    c. Resource licensing
    d. Data compression

2. Which of the following is NOT a category of metadata?
    a. Descriptive
    b. Structural
    c. Analytical
    d. Administrative

3. What does the acronym FAIR stand for in the context of metadata?
    a. Findable, Accessible, Interoperable, Reusable
    b. Fast, Accurate, Intelligent, Reliable
    c. Free, Available, Interactive, Robust
    d. Functional, Adaptable, Integrated, Resourceful

4. Which metadata standard is domain-agnostic and widely adopted?
    a. Darwin Core
    b. Dublin Core
    c. Data Documentation Initiative
    d. Schema.org

5. Match the metadata element with its description:
    a. Metadata element:
        i. Title/Name
        ii. Format
        iii. Rights Holder
        iv. Identifier
    b. Description:
        i. A unique identification code assigned to the resource
        ii. Name given to the resource
        iii. Information about the entities or persons who hold the rights to the data
        iv. File format, physical medium, dimensions, or file size of the resource

6. Metadata is defined as "_____data about data or an object resource."

7. The choice of repository dictates the applicable metadata ________ standards.

8. ________ metadata is developed after a resource is created, while prospective metadata is created before the data is created.

9. A metadata ________ registry is where metadata definitions are stored and maintained in a controlled manner.

**Solutions**

1. Answer: b) Resource discovery

2. Answer: c) Analytical

3. Answer: a) Findable, Accessible, Interoperable, Reusable

4. Answer: b) Dublin Core

5. Correct matches:
    a. Title/Name - ii) Name given to the resource
    b. Format - iv) File format, physical medium, dimensions, or file size of the resource
    c. Rights Holder - iii) Information about the entities or persons who hold the rights to the data
    d. Identifier - i) A unique identification code assigned to the resource

6. Answer: Structured

7. Answer: Schema

8. Answer: Retrospective

9. Answer: Schema

# 2.0 Working with Metadata

## 2.1 Learning Objectives

After working through this chapter, you will be able to:

- Create metadata, while applying special considerations.
- Understand the rudiments of working with metadata, such as terminology mapping and metadata schema cross walking.
- Grow familiar with types and standards frequently used by different research communities.
- Learn several data sharing and reuse licensing models.

## 2.2 Metadata Creation Steps

When creating metadata for a potential resource for the very first time, follow these eight steps, which serve as general guidance. After you run through the process a few times, it will become second nature.

**Step 1:** Perform Initial Assessment

- Is the resource physical or abstract (such as a digital object or digital rendition of a physical object)?
- What information will identify the resource?
- How are you or others going to search for and find the resource?
- How do you expect others to use the resource currently and in the future?
- What information will you provide others to establish use rights and limitations of resource use?

**Step 2:** Create Metadata Wish List

- Based on your initial assessment in Step 1, list out the metadata elements you may need such as title, subject, rights, etc. This is your metadata wish list.
- A text document or spreadsheet can suffice at this initial stage.

**Step 3:** Check for Missing Metadata

- Consider the information you already have about your resource. Are there elements on your metadata wish list missing on your resource?
- Attempt to find or create the missing info, when possible.

**Step 4:** Update Metadata List

- Verify information essential to facilitate discovery and identification (such as unique identifiers), and to also give sufficient context to the data. Identify the standards and controlled vocabularies applicable to the metadata elements. Consider how a persistent identifier (PID) will be generated and assigned to your resource. For example, will it be assigned by the repository you intend to deposit it in? An exception of course, would be Open Researcher and Contributor ID (ORCID) or Research Organization Registry (ROR) (see section on Persistent Identifiers), which are research contributors you need to find.

**Step 5:** Select the Repository

- Select the repository you plan to use. Generalist repositories are designed to store a broad range of resource data types. Specialist repositories store a limited range of data types, but often at higher granularity.[16]
- When weighing the choice of repository, many factors can be considered such as those listed in the National Science and Technology Council (NTSC) Desirable Characteristics Of Data Repositories For Federally Funded Research document,[17] which defines 21 characteristics that Data Repositories should have to be recommended to researchers as a viable option for sharing their data and satisfying federal data sharing requirements.

**Step 6**: Map Repository Metadata Schema to your Final Metadata List

- Pick the appropriate metadata schema supported by the repository. If there are elements in the repository schema you do not have on your metadata list, attempt to acquire these and augment the metadata list to obtain your final metadata list.
- Map the elements of the schema to those of your final metadata list. You may need to convert to the terminology systems of use in the repository, which may be different from the one you normally use (see Terminology Mappings and Crosswalks).
- Some elements in your metadata list may not map to the elements on the repository schema as is. When possible, extend[18] the schema to include the additional elements you need. If the schema does not allow for extension, you may need to work with the community that developed the schema to request the necessary modifications. In the meantime, be sure to include a readme[19] file with the data resource containing the additional information.

**Step 7**: Create the Instance Metadata

- In some cases, the repository itself provides a submission tool or guidelines. As a beginner, this may be the simplest approach to take.
- Some repositories may not have submission tools. You may create the metadata you need for the repository manually; however, it is preferable to use software tools[20] to create and validate the (instance) metadata for your data based on the repository schema. See the list of tools/platforms for a suitable metadata editor. Some of these platforms publish schemas for reuse, support a form-based input approach to easily create instance metadata, and can allow you to directly submit these to certain repositories.

**Step 8:** Update Metadata

- Update metadata, when necessary, especially if additional information on the resource surfaces.
- Avoid updating any persistent identifiers[21] as these are long-lasting, unique references to a resource and lead to the creation of robust links.

---

[16] In some cases, it is advantageous to store metadata in more than one location (such as in a data catalog), while the data resides in another. However, it is never a good practice to store the same data in more than one repository. If you do so, it may lead to confusion, should the data be updated in one repository but not the others.
[17] https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf
[18] If these are not viable options, develop your own schema but document it well, so others can follow.
[19] https://data.research.cornell.edu/data-management/sharing/readme/
[20] Such as the CEDAR Workbench
[21] https://library.cern/submit-and-publish/persistent-identifiers/why-pids

## 2.3 Metadata Tools, Terminology Mapping, and Crosswalks

This section contains technology that you can leverage to develop metadata. An extensive list of these aids and associated links to documentation may be obtained from the footnotes of the following section.

**Metadata Tools**

A few popular tools for creating and validating metadata are listed below:[22]

- CEDAR Workbench is an open-source tool to manage metadata for any field of study, using rigorous semantic principles if desired. It allows users to specify templates using a User Interface (UI) like survey forms in Google Forms or Survey Monkey), then fill out those forms efficiently using drop-down menus, help tips, and intelligent suggestions. Templates and metadata can be shared with other users and groups. Metadata also can be downloaded in JSON-LD, simple JSON, or RDF, or exported to connected repositories, which can be integrated using the full Application Programming Interface (API) suite.
- ISA Creator is an open source, stand-alone application that assists with planning and describing experiments and facilitates export and import of data directly to and from some public repositories. Additional tools are available in the Investigation, Study, Assay (ISA)-Tools software suite for parsing ISA-Tab into R data structures and for parsing PERL and Python for ISA-Tab. ISA-Tab is the required format for publishing data in Nature Publishing's *Scientific Data* journal. This software creates separate descriptive files for your experimental files.
- RightField is an open-source tool that allows search and selection of ontology terms for any field of study from within Microsoft Excel. RightField allows you to assign a pre-determined list of options to a particular cell within the spreadsheet. All annotations are embedded within the spreadsheet. The user can select from the National Center for Biomedical Ontology's (NCBO) BioPortal ontologies or import an ontology from a Uniform Resource Locator (URL) or local machine.

**Terminology Mapping**[23]

Terminology Mapping can be accomplished using Unified Medical Language System (UMLS), which is a comprehensive resource developed by the NLM that can be used for various purposes in the healthcare and biomedical fields. It can be used to crosswalk between different medical terminologies and coding systems to ensure consistency and accuracy.

**Crosswalks**[24,25]

Metadata crosswalks translate elements and values from one schema to those of another. Crosswalks facilitate interoperability between different metadata schemas. Below are examples of existing crosswalks that have been developed between widely applied metadata standards such as Dublin Core, MARC Bibliographic Data, Metadata Object Description Schema (MODS) and Encoded Archival Description (EAD)

---

[22] https://guides.library.stanford.edu/research-metadata/metadata-tools
[23] https://www.nlm.nih.gov/research/umls/index.html
[24] https://guides.library.cmu.edu/c.php?g=472661&p=9230176
[25] https://guides.lib.utexas.edu/metadata-basics/crosswalks

- Dublin Core to MARC
- Dublin Core to MODS
- MARC to Dublin Core
- MARC to MODS
- MARC to EAD (Version 2002)
- MODS to Dublin Core

## 2.4 Databases of Metadata Standards[26]

The list of metadata standards for various disciplines is an extensive one. Fortunately, you will only need to get familiar with those pertaining to the field you work in. A starter list of metadata schemas may be found below from:

- Metadata Standards Catalog (bath.ac.uk)
- Metadata & Standards - Data Management Resources - Guides at Penn Libraries (upenn.edu)

## 2.5 List of Metadata Schemas

There are a multitude of schemas that have been developed over time and continue to evolve. Listed below are a few examples of popular ones.

**Cross-Disciplinary (domain-agnostic)**

- Dublin Core: The Dublin Core™ Metadata Initiative (DCMI) provides access to schemas defining DCMI term declarations represented in various schema languages. Schemas are machine-processable specifications that define the structure and syntax of metadata specifications in a formal schema language.
- Friend of a Friend (FOAF): a machine-readable ontology describing persons, their activities and their relations to other people and objects.
- Simple Knowledge Organization System (SKOS): specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web.
- Provenance Ontology (PROV-O): provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts.
- Resource Description Framework (RDF): a standard way to encode information about resources so that it can be exchanged between applications. It enables interoperability between applications in areas like resource description, site maps, and electronic commerce.

**Ecology, Geosciences & Biology**

- Darwin Core: Darwin Core includes a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa and their occurrence in nature as documented by observations, specimens, samples, and related information.
- Ecological Metadata Language: The Ecological Metadata Language (EML) defines a comprehensive vocabulary and a readable XML markup syntax for documenting research data. It

[26] Metadata & Standards - Data Management Resources - Guides at Penn Libraries (upenn.edu)

is in widespread use in the earth and environmental sciences, and increasingly in other research disciplines as well.

- Flexible Image Transport System: Standard data format used in astronomy.
- NeXus: This is a common data format for neutron, x-ray, and muon science. It is being developed as an international standard by scientists and programmers representing major scientific facilities in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

**Social Science**

- Data Documentation Initiative: The DDI is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences. It a free standard that can document and manage different stages in the research data life cycle, such as conceptualization, collection, processing, distribution, discovery, and archiving.

## 2.6 Persistent Identifiers

Below is a short list of organizations proving persistent identifiers for use with metadata.

- Crossref: links research objects, entities, and actions, creating a lasting and reusable scholarly record.
- DataCite: provides persistent identifiers for research outputs.
- Research Resource Identifiers (RRID): enhances scientific findings by making it easier to track and verify the resources used in experiments.
- Research Activity Identifier (RAID): maintains a comprehensive record of the project's life cycle and associated outputs.
- Research Organization Registry (ROR): provides a PID for every research organization in the world.
- Open Researcher and Contributor ID (ORCID): provides persistent identifiers for researchers.
- Virtual International Authority File (VIAF): combines multiple name authority files into a single name authority service
- Library of Congress Name Authority File (LCNAF): provides authoritative data for names of persons, organizations, events, places, and titles.

## 2.7 Licensing Models

When sharing data, it's important to choose an appropriate licensing model to specify how the data can be used, shared, and modified. Here are some common license models for data sharing:

1. Creative Commons Licenses (CC) and license chooser
    - CC0 (Public Domain Dedication): No rights reserved; data can be used without any restrictions.
    - CC BY: Allows use, distribution, and modification as long as credit is given to the original creator.
    - CC BY-SA: Similar to CC BY, but derivative works must be licensed under the same terms.
    - CC BY-ND: Allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to the creator.

    - CC BY-NC: Allows for use and modification, but not for commercial purposes, with credit to the creator.
    - CC BY-NC-SA: Non-commercial use allowed, derivatives must be shared alike, with credit.
    - CC BY-NC-ND: Most restrictive; allows others to download and share the data with credit, but no modifications or commercial use.
2. Open Data Commons Licenses
    - Public Domain Dedication and License: Waives all rights, placing data in the public domain.
    - Attribution License: Requires attribution for use, sharing, and modification.
    - Open Database License: Allows for sharing and modification, with attribution and share-alike conditions.
3. GNU General Public License (GPL)
    - Primarily for software but can be applied to data if treated as a dataset with source code.
4. Apache License 2.0
    - Allows for use, modification, and distribution, with attribution and a copy of the license.
5. Massachusetts Institute of Technology (MIT) License
    - Permissive license allowing for reuse with minimal restrictions, requiring only attribution.
6. Proprietary Licenses
    - Custom licenses that specify specific terms and conditions for data use, often used by companies for commercial data.
7. Data Use Agreements (DUA)
    - Custom agreements tailored to specific data sharing arrangements, often used in research collaborations.
8. Non-Commercial Licenses
    - Restrict use to non-commercial purposes, often with specific attribution requirements. When choosing a license, consider the intended use, the need for attribution, and any restrictions on commercial use or modifications such as those for sensitive human subject data.

The Creative Commons License 4.0 is in popular usage for data sharing.

## 2.8 Learning Summary

Concepts covered in this chapter include:

- Metadata creation steps
- Tools for working with metadata
- Terminology mapping and metadata schema cross walking
- Metadata types and standards frequently used by different research communities
- Persistent identifiers
- Licensing models

## 2.9 Knowledge Self-Assessments

**Questions**

1. What is the first step in the metadata creation process?
    a. Select the Repository
    b. Perform Initial Assessment
    c. Create Metadata Wish List
    d. Update Metadata

2. Which tool is used for managing metadata using templates similar to survey forms?
    a. ISA Creator
    b. RightField
    c. CEDAR Workbench
    d. UMLS

3. Which of the following is a persistent identifier for researchers?
    a. Crossref
    b. DataCite
    c. ROR
    d. ORCID
    e. All the above

4. Which licensing model allows for use, distribution, and modification as long as credit is given to the original creator?
    a. CC BY-NC
    b. CC BY
    c. CC0
    d. CC BY-ND

5. The __________ is used to translate elements and values from one metadata schema to another, to facilitate interoperability.

6. The __________ license is a permissive license allowing for reuse with minimal restrictions, requiring only attribution.

7. Explain the purpose of a PID in metadata.

8. Why is it recommended to use metadata editing tools like CEDAR for beginners?

**Solutions**

1. Answer: b) Perform Initial Assessment
2. Answer: c) CEDAR Workbench
3. Answer: e) All of the above
4. Answer: b) CC BY
5. Answer: crosswalk
6. Answer: MIT
7. Answer: A PID provides a lasting and unique reference to a resource, ensuring that it can be reliably cited and accessed over time.
8. Answer: Tools like CEDAR simplify and ensure accurate metadata input, allowing beginners to serialize metadata into required formats like Hypertext Markup Language (HTML, XML, RDF, or JSON-LD.

# 3.0 Specific Metadata Schemas

## 3.1 Learning Objectives

After working through this chapter, you will see how domain-agnostic and domain-specific schemas are applied in practice. In particular, you will learn how to apply:

- Simple Dublin Core, a domain-agnostic schema, to image resources
- Darwin Core and DDI schemas to Biodiversity and Behavioral/Social Science Data, respectively

## 3.2 Domain-Agnostic Metadata Schemas

### 3.2.1 Simple Dublin Core

The Simple Dublin Core Metadata Element Set is one of the most widely used metadata schemas for a generalist repository. Originally developed to describe web resources, Dublin Core has been used to describe a variety of physical and digital resources.[27] Many of the data elements can leverage value standards, as shown in Table 1. You need not use all the 15 elements—just the relevant ones.

Use of any Dublin Core (DC) elements is not mandatory but optional; one only need select elements relevant to a resource, repeating elements when necessary. Some of the metadata elements in the DC can have repeated data values. One example is the description element, which could be in multiple languages.

The table below lists the elements of the Simple Dublin Core metadata schema, with the use and some example applicable standards.

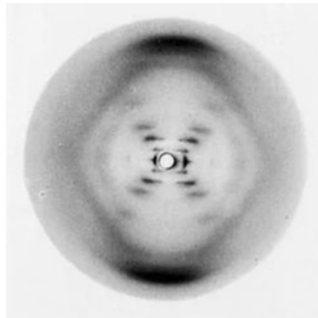
**Table 1. DC Metadata Element Set with Example Values**

| Dublin Core Element | Use | Example of Data Value Standards |
|---|---|---|
| **Title** | A name given to the resource. | |
| **Subject** | The topic of the resource. | Library of Congress Subject Headings |
| **Description** | An account of the resource. | |
| **Creator** | An entity primarily responsible for making the resource. | Library of Congress Name Authority File |
| **Publisher** | An entity responsible for making the resource available. | |

[27] For example, this is a major component of the DATaset Metadata Model (DATMM) (nih.gov).

| Dublin Core Element | Use | Example of Data Value Standards |
|---|---|---|
| **Contributor** | An entity responsible for making contributions to the resource. | Library of Congress Name Authority File |
| **Date** | A point or period of time associated with an event in the life cycle of the resource. | World Wide Web Consortium (W3C)DTF |
| **Type** | The nature or genre of the resource. | DCMI Type Vocabulary |
| **Format** | The file format, physical medium, or dimensions of the resource. | Internet Media Types |
| **Identifier** | An unambiguous reference to the resource within a given context. | |
| **Source** | A related resource from which the described resource is derived. | |
| **Language** | A language of the resource. | International Organization for Standardization (ISO) 639 |
| **Relation** | A related resource. | |
| **Coverage** | The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. | Thesaurus of Geographic Names |
| **Rights** | Information about rights held in and over the resource. | RightsStatements.org |

Imagine you are interested in applying metadata to the image file you found online, reproduced below:[28]



**MRC-240423-Photo51-681x735.jpg**

In the beginning, it may not be clear what this image is about, its significance, what you can do with it, how you would find it again, or how to relate it to other images. You decide to start small, using the Simple Dublin Core schema as a guide. You capture the technical information on the image—the **name, format, and web address**. To provide disambiguation, you use standard vocabulary and persistent identifiers where possible. You can always revisit the metadata for refinement in the future.

The metadata now provides you basic search capabilities. For example, you can list all the images in the repository. However, this is not very useful. Your search, although sensitive, will lack specificity.

You can extend the metadata by adding descriptive metadata. You do a bit of additional digging and uncover that **Rosalind E. Franklin** recorded this image at **King's College, London, in 1952**. So, you update the metadata. Now you have metadata, which enables you, or someone else with access to your repository, to use the descriptive metadata to search for and download the image in a more meaningful manner. For example, you may be able to look for other work by Rosalind Franklin.

The query results using the additional metadata may even list files that someone else has deposited in the same repository. If you have their permission, you may be able to use these.

You do some additional research and enrich the metadata to include descriptions: **"X-ray diffraction pattern of B-DNA**.**"** You also locate the original **source** and **language**, and a **related source**, for the image to use **"DNA"** and **"X-Ray Diffraction"** as the subjects in the metadata. Because you have been enriching your file metadata, you are now capable of building a collection of files related to the subject **"DNA"** in the repository.

Finally, you uncover information about **rights, licenses, and permissions** regarding this image. This is crucial if you plan to publish or share the data in some form. Shown below in the table are the elements you make use of in DC.

[28] Based on "Why Metadata Matters," a video tutorial developed by Office of Data Science Strategy(ODSS), NIH (2024).

**Table 2: Use of Simple DC Elements for MRC-240423-Photo51-681x735.jpg**

| Dublin Core Element | Use | Value/Standards |
|---|---|---|
| **Title** | A name given to the resource. | MRC-240423-Photo51-681x735 |
| **Subject** | The topic of the resource. | **DNA** MeSH URI<br>**X -Ray Diffraction** MeSH URI |
| **Description** | An account of the resource. | X-ray diffraction pattern of B-DNA<br>Recorded at King's, College London |
| **Creator** | An entity primarily responsible for making the resource. | Rosalind E. Franklin ORCID URI |
| **Publisher** | An entity responsible for making the resource available. | UK Research and Innovation |
| **Contributor** | An entity responsible for making contributions to the resource. | Raymond G. Gosling LCNAF URI |
| **Date** | A point or period of time associated with an event in the life cycle of the resource. | 1952 W3CDTF URI |
| **Type** | The nature or genre of the resource. | Image URL |
| **Format** | The file format, physical medium, or dimensions of the resource. | jpeg iana.org URI |
| **Identifier** | An unambiguous reference to the resource within a given context. | ukri.org URL |
| **Source** | A related resource from which the described resource is derived. | FRANKLIN, R., GOSLING, R. Molecular Configuration in Sodium Thymonucleate. Nature **171**, 740–741 (1953).<br>DOI.ORG URI |
| **Language** | A language of the resource. | eng ISO 639 URI |
| **Relation** | A related resource. | Maddox, B. The double helix and the 'wronged heroine'. Nature 421, 407–408 (2003).<br>DOI.ORG URI |
| **Coverage** | The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. | |

| Dublin Core Element | Use | Value/Standards |
|---|---|---|
| **Rights** | Information about rights held in and over the resource. | King's College London Archives ROR.ORG URI<br><br>CC BY NC 4.0 creativecommons.org URI |

In the future, you decide to write a review article on the causes of Down syndrome and decide to include background on DNA.

You know you have the DNA structure images you added metadata to, but you perform a broader search in the repository and find images deposited by other researchers on DNA and Down syndrome. You even find images of key researchers behind discovering the structure of DNA and the causes of Down syndrome—Rosalind Franklin and Marthe Gautier.

**Figure 1. Query results for subject = "DNA Structure" and "Down Syndrome"**

Because you *and* other researchers applied granular metadata to the files, you were able to quickly locate a handful of images to include in your article. If your efforts were going to lead to new images or information being collected or generated, it would be smart to think about generating and collecting metadata prospectively.

## 3.3 Domain-Specific Metadata Schemas

This section provides an example of metadata schemas that are used for research by scientists in the biomedical sciences and behavioral/social sciences. Each schema is provided in basic core formulation, as well as extensions to provide sensitivity as well as specificity in search capability for data resources.

The following sections describe the use of Darwin Core and DDI metadata and common methods for serialization. Beginners should use metadata editing tools (such as CEDAR) to make it simpler and to more accurately to input metadata from the final metadata list into a form-based entry to yield metadata serializable into HTML, XML, RDF or JSON-LD formats required by the repository where you intend to deposit your data resources. *Note: These examples are simplified to give the reader an idea of the structure of the serialized metadata, and consequently and may not include, in some cases, the PIDs related to the controlled vocabulary systems or data.*

### 3.3.1 Darwin Core

Darwin Core Metadata[29] is a standardized framework used to describe biodiversity data, particularly data related to biological specimens and observations. It is part of the broader Darwin Core standard, which facilitates the sharing and integration of biodiversity information across different platforms and databases. Darwin Core works alongside other biodiversity informatics standards, such as EML and the Access to Biological Collections Data schema.

By providing a structured approach to metadata, Darwin Core Metadata enhances the discoverability, accessibility, and usability of biodiversity data, supporting research, conservation, and policy-making efforts globally.

**Core Elements**

1. *Title*: The name of the dataset.
2. *Creator:* The individual or organization responsible for the dataset.
3. *Description:* A summary of the dataset's content and purpose.
4. *Keywords:* Terms that describe the dataset's subject matter.
5. *Citation:* Information on how to cite the dataset.
6. *Contact Information:* Details for reaching the dataset's custodian or creator.
7. *Rights:* Information about the dataset's usage rights and licenses.

**Use Cases**

Widely used in biodiversity databases, such as the Global Biodiversity Information Facility, to document and share data about species occurrences, specimens, and ecological observations.

**Simple Darwin Core Example**

Consider a final metadata list with the following elements:

1. Title: Bird Observations in Central Park

[29] https://dwc.tdwg.org/

2. Creator: Jane Doe
3. Description: This dataset contains records of bird species observed in Central Park, New York, from January to December 2022. It includes information on species, date, time, and location of each observation.
4. Keywords: Birds, Central Park, New York, Observations, 2022
5. Citation: Doe, J. (2023). Bird Observations in Central Park. Retrieved from [URL]
6. Contact Information: Jane Doe, Email: jane.doe@example.com
7. Rights: CC BY 4.0 - This dataset is licensed under a Creative Commons Attribution 4.0 International License.

Following are metadata serializations in HTML, XML, RDF, and JSON-LD:

### *HTML*

```html
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Darwin Core Metadata</title>
</head>
<body>
  <h1>Bird Observations in Central Park</h1>
  <p><strong>Creator:</strong> Jane Doe</p>
  <p><strong>Description:</strong> This dataset contains records of bird species observed in Central Park, New York, from January to December 2022. It includes information on species, date, time, and location of each observation.</p>
  <p><strong>Keywords:</strong> Birds, Central Park, New York, Observations, 2022</p>
  <p><strong>Citation:</strong> Doe, J. (2023). Bird Observations in Central Park. Retrieved from [URL]</p>
  <p><strong>Contact Information:</strong> Jane Doe, Email: jane.doe@example.com</p>
  <p><strong>Rights:</strong> CC BY 4.0 - This dataset is licensed under a Creative Commons Attribution 4.0 International License.</p>
</body>
</html>
```

### *XML*

```xml
<DarwinCore>
  <Title>Bird Observations in Central Park</Title>
  <Creator>Jane Doe</Creator>
  <Description>This dataset contains records of bird species observed in Central Park, New York, from January to December 2022. It includes information on species, date, time, and location of each observation.</Description>
  <Keywords>Birds, Central Park, New York, Observations, 2022</Keywords>
  <Citation>Doe, J. (2023). Bird Observations in Central Park. Retrieved from [URL]</Citation>
  <ContactInformation>Jane Doe, Email: jane.doe@example.com</ContactInformation>
  <Rights>CC BY 4.0 - This dataset is licensed under a Creative Commons Attribution 4.0 International License.</Rights>
</DarwinCore>
```

*RDF (in XML)*

```xml
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:dcterms="http://purl.org/dc/terms/">
  <rdf:Description rdf:about="http://example.org/dataset/BirdObservationsCentralPark">
    <dcterms:title>Bird Observations in Central Park</dcterms:title>
    <dcterms:creator>Jane Doe</dcterms:creator>
    <dcterms:description>This dataset contains records of bird species observed in Central Park, New York, from January to December 2022. It includes information on species, date, time, and location of each observation.</dcterms:description>
    <dcterms:subject>Birds, Central Park, New York, Observations, 2022</dcterms:subject>
    <dcterms:bibliographicCitation>Doe, J. (2023). Bird Observations in Central Park. Retrieved from [URL]</dcterms:bibliographicCitation>
    <dcterms:license rdf:resource="https://creativecommons.org/licenses/by/4.0/"/>
    <dcterms:contributor>Jane Doe, Email: jane.doe@example.com</dcterms:contributor>
  </rdf:Description>
```

*</rdf:RDF>*

*JSON-LD*


```
{
  "@context": "https://schema.org/",
  "@type": "Dataset",
  "name": "Bird Observations in Central Park",
  "creator": {
    "@type": "Person",
    "name": "Jane Doe",
    "email": "jane.doe@example.com"
  },
  "description": "This dataset contains records of bird species observed in Central Park, New York, from January to December 2022. It includes information on species, date, time, and location of each observation.",
  "keywords": "Birds, Central Park, New York, Observations, 2022",
  "citation": "Doe, J. (2023). Bird Observations in Central Park. Retrieved from [URL]",
  "license": "https://creativecommons.org/licenses/by/4.0/"
}
```


### Darwin Core Extensions

Darwin Core with Extensions refers to the use of additional, specialized terms that complement the core Darwin Core standard to accommodate more complex or specific biodiversity data needs. While the core standard provides a basic set of terms for describing biodiversity data, extensions allow for greater detail and customization. Using Darwin Core with Extensions allows researchers and organizations to create rich, detailed datasets that can be easily shared and integrated with other biodiversity data, enhancing research and conservation efforts.

*Extensions are often developed and maintained by the biodiversity informatics community to address emerging needs and new research areas.*

Common Extensions

- Event Core: for detailed event data, such as sampling protocols and effort
- Measurement or Fact: for quantitative data related to occurrences, like temperature or pH
- Resource Relationship: to describe relationships between resources, such as predator-prey interactions

### Darwin Core with Extensions Example

Consider a final metadata list with the following elements:

```
Title: Plant Observations in Amazon Rainforest
Creator: Dr. Emily Green
Description: This dataset includes records of plant species observed in the Amazon Rainforest during the 2023 expedition. It contains detailed information on species, location, environmental conditions, and associated fauna.
Keywords: Plants, Amazon Rainforest, Biodiversity, Environmental Data, 2023
Citation: Green, E. (2023). Plant Observations in Amazon Rainforest. Retrieved from [URL]
Contact Information: Dr. Emily Green, Email: emily.green@example.com
Rights: CC BY 4.0 - This dataset is licensed under a Creative Commons Attribution 4.0 International License.

Core Terms:
- Scientific Name: Aniba rosaeodora
- Location: Latitude -3.4653, Longitude -62.2159
- Date: 2023-06-15
- Collector: Emily Green
```

Extensions:
- Event Core:
- Sampling Protocol: Transect Walk
- Effort: 5 hours per transect
- Measurement or Fact
- Temperature: 28°C
- Humidity: 85%
- Resource Relationship:
- Associated Fauna: Eulaema mocsaryi (pollinator)

## Following are metadata serializations in HTML, XML, RDF and JSON-LD:

### *HTML*

```
<!DOCTYPE html>

<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Plant Observations in Amazon Rainforest</title>
</head>
<body>
  <h1>Plant Observations in Amazon Rainforest</h1>
  <p><strong>Creator:</strong> Dr. Emily Green</p>
  <p><strong>Description:</strong> This dataset includes records of plant species observed in the Amazon Rainforest during the 2023
expedition. It contains detailed information on species, location, environmental conditions, and associated fauna.</p>
  <p><strong>Keywords:</strong> Plants, Amazon Rainforest, Biodiversity, Environmental Data, 2023</p>
  <p><strong>Citation:</strong> Green, E. (2023). Plant Observations in Amazon Rainforest. Retrieved from [URL]</p>
  <p><strong>Contact Information:</strong> Dr. Emily Green, Email: <a
href="mailto:emily.green@example.com">emily.green@example.com</a></p>
  <p><strong>Rights:</strong> CC BY 4.0 - This dataset is licensed under a Creative Commons Attribution 4.0 International License.</p>

  <h2>Core Terms</h2>
  <ul>
    <li><strong>Scientific Name:</strong> Aniba rosaeodora</li>
    <li><strong>Location:</strong> Latitude -3.4653, Longitude -62.2159</li>
    <li><strong>Date:</strong> 2023-06-15</li>
    <li><strong>Collector:</strong> Emily Green</li>
  </ul>

  <h2>Extensions</h2>
  <h3>Event Core</h3>
  <ul>
    <li><strong>Sampling Protocol:</strong> Transect Walk</li>
    <li><strong>Effort:</strong> 5 hours per transect</li>
  </ul>
```

### *XML*

```
<?xml version="1.0" encoding="UTF-8"?>

<dataset>
  <title>Plant Observations in Amazon Rainforest</title>
  <creator>Dr. Emily Green</creator>
  <description>This dataset includes records of plant species observed in the Amazon Rainforest during the 2023 expedition. It contains
detailed information on species, location, environmental conditions, and associated fauna.</description>
  <keywords>Plants, Amazon Rainforest, Biodiversity, Environmental Data, 2023</keywords>
  <citation>Green, E. (2023). Plant Observations in Amazon Rainforest. Retrieved from [URL]</citation>
  <contactInformation>
    <name>Dr. Emily Green</name>
    <email>emily.green@example.com</email>
```

```xml
  </contactInformation>
  <rights>CC BY 4.0 - This dataset is licensed under a Creative Commons Attribution 4.0 International License.</rights>

  <coreTerms>
    <scientificName>Aniba rosaeodora</scientificName>
    <location>
      <latitude>-3.4653</latitude>
      <longitude>-62.2159</longitude>
    </location>
    <date>2023-06-15</date>
    <collector>Emily Green</collector>
  </coreTerms>

  <extensions>
    <eventCore>
      <samplingProtocol>Transect Walk</samplingProtocol>
      <effort>5 hours per transect</effort>
    </eventCore>

    <measurementOrFact>
      <temperature>28°C</temperature>
      <humidity>85%</humidity>
    </measurementOrFact>

    <resourceRelationship>
      <associatedFauna>Eulaema mocsaryi (pollinator)</associatedFauna>
    </resourceRelationship>
  </extensions>
</dataset>
```


### *RDF (in XML)*


```xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:dc="http://purl.org/dc/elements/1.1/"
        xmlns:dwc="http://rs.tdwg.org/dwc/terms/">

  <rdf:Description rdf:about="http://example.org/dataset/PlantObservationsAmazonRainforest">
    <dc:title>Plant Observations in Amazon Rainforest</dc:title>
    <dc:creator>Dr. Emily Green</dc:creator>
    <dc:description>This dataset includes records of plant species observed in the Amazon Rainforest during the 2023 expedition. It contains detailed information on species, location, environmental conditions, and associated fauna.</dc:description>
    <dc:subject>Plants, Amazon Rainforest, Biodiversity, Environmental Data, 2023</dc:subject>
    <dc:identifier>Green, E. (2023). Plant Observations in Amazon Rainforest. Retrieved from [URL]</dc:identifier>
    <dc:publisher>Dr. Emily Green</dc:publisher>
    <dc:rights>CC BY 4.0 - This dataset is licensed under a Creative Commons Attribution 4.0 International License.</dc:rights>

    <dwc:scientificName>Aniba rosaeodora</dwc:scientificName>
    <dwc:decimalLatitude>-3.4653</dwc:decimalLatitude>
    <dwc:decimalLongitude>-62.2159</dwc:decimalLongitude>
    <dwc:eventDate>2023-06-15</dwc:eventDate>
    <dwc:recordedBy>Emily Green</dwc:recordedBy>

    <dwc:samplingProtocol>Transect Walk</dwc:samplingProtocol>
    <dwc:effort>5 hours per transect</dwc:effort>

    <dwc:measurementOrFact>
      <rdf:Description>
        <dwc:measurementType>Temperature</dwc:measurementType>
        <dwc:measurementValue>28°C</dwc:measurementValue>
      </rdf:Description>
      <rdf:Description>
```

```xml
        <dwc:measurementType>Humidity</dwc:measurementType>
        <dwc:measurementValue>85%</dwc:measurementValue>
      </rdf:Description>
    </dwc:measurementOrFact>

    <dwc:associatedTaxa>
      <rdf:Description>
        <dwc:taxonName>Eulaema mocsaryi</dwc:taxonName>
        <dwc:relationshipOfResource>pollinator</dwc:relationshipOfResource>
      </rdf:Description>
    </dwc:associatedTaxa>
  </rdf:Description>
</rdf:RDF>
```


## JSON LD


```json
{
 "@context": {
  "@vocab": "http://rs.tdwg.org/dwc/terms/"
 },
 "@type": "Dataset",
 "title": "Plant Observations in Amazon Rainforest",
 "creator": {
  "@type": "Person",
  "name": "Dr. Emily Green"
 },
 "description": "This dataset includes records of plant species observed in the Amazon Rainforest during the 2023 expedition. It contains detailed information on species, location, environmental conditions, and associated fauna.",
 "keywords": ["Plants", "Amazon Rainforest", "Biodiversity", "Environmental Data", "2023"],
 "citation": "Green, E. (2023). Plant Observations in Amazon Rainforest. Retrieved from [URL]",
 "contactPoint": {
  "@type": "ContactPoint",
  "name": "Dr. Emily Green",
  "email": "emily.green@example.com"
 },
 "license": "CC BY 4.0 - This dataset is licensed under a Creative Commons Attribution 4.0 International License.",
 "coreTerms": {
  "scientificName": "Aniba rosaeodora",
  "location": {
   "latitude": -3.4653,
   "longitude": -62.2159
  },
  "date": "2023-06-15",
  "collector": "Emily Green"
 },
 "extensions": {
  "eventCore": {
   "samplingProtocol": "Transect Walk",
   "effort": "5 hours per transect"
  },
  "measurementOrFact": {
   "temperature": "28°C",
   "humidity": "85%"
  },
  "resourceRelationship": {
   "associatedFauna": "Eulaema mocsaryi (pollinator)"
  }
 }
}
```

### 3.3.2 Data Documentation Initiative

The DDI[30] is an international standard for describing and documenting data from the social, behavioral, and economic sciences. It provides a structured framework for metadata, facilitating data sharing, discovery, and reuse.

**The Data Documentation Initiative**

Offers a consistent format for documenting datasets, ensuring interoperability across different platforms and institutions. It covers various aspects of data, including study design, data collection methods, variables, and data processing. The DDI Codebook is a simpler version for documenting basic metadata.

**Core Elements**

1. *Document Description*: provides basic information about the dataset, including title, author, production date, and distributor contact
2. *Study Description*: contains details about the study, such as the abstract, subject area, and data collection dates
3. *Data Description:* describes the variables in the dataset, including their names, labels, and formats

**Use Cases**

The DDI is widely used by researchers, data archives, and institutions to manage and share data efficiently.

**DDI Example**

This example illustrates how Simple DDI metadata can be used to document a dataset, making it easier for others to understand and utilize the data.

**Simple DDI example**

Consider a final metadata list with the following elements:

```
Title: Sample Survey Data
ID: SS1234
Author: Research Institute
Production Date: 2023-10-01

Study Title: Sample Survey on Consumer Behavior
Keywords: Consumer Behavior, Survey
Abstract: This study explores consumer behavior patterns in urban areas.

Variables:
1. Age
  - Label: Age of Respondent
  - Type: Discrete
  - Range: 18 to 99

2. Income
  - Label: Annual Income
  - Type: Continuous
```

[30] https://ddialliance.org/

- Range: 0 to 1,000,000

Following are metadata serializations in HTML, XML, RDF and JSON-LD:

*HTML*

```html
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Sample Survey Data Metadata</title>
</head>
<body>
  <h1>Sample Survey Data</h1>
  <p><strong>ID:</strong> SS1234</p>
  <p><strong>Author:</strong> Research Institute</p>
  <p><strong>Production Date:</strong> 2023-10-01</p>

  <h2>Study Title</h2>
  <p>Sample Survey on Consumer Behavior</p>

  <h2>Keywords</h2>
  <ul>
    <li>Consumer Behavior</li>
    <li>Survey</li>
  </ul>

  <h2>Abstract</h2>
  <p>This study explores consumer behavior patterns in urban areas.</p>

  <h2>Variables</h2>
  <ul>
    <li>
      <strong>Age</strong>
      <ul>
        <li><strong>Label:</strong> Age of Respondent</li>
        <li><strong>Type:</strong> Discrete</li>
        <li><strong>Range:</strong> 18 to 99</li>
      </ul>
    </li>
    <li>
      <strong>Income</strong>
      <ul>
        <li><strong>Label:</strong> Annual Income</li>
        <li><strong>Type:</strong> Continuous</li>
        <li><strong>Range:</strong> 0 to 1,000,000</li>
      </ul>
    </li>
  </ul>
</body>
</html>
```

*XML*

```xml
<?xml version="1.0" encoding="UTF-8"?>
<DDI>
  <Title>Sample Survey Data</Title>
  <ID>SS1234</ID>
  <Author>Research Institute</Author>
```

```xml
  <ProductionDate>2023-10-01</ProductionDate>

  <StudyTitle>Sample Survey on Consumer Behavior</StudyTitle>

  <Keywords>
    <Keyword>Consumer Behavior</Keyword>
    <Keyword>Survey</Keyword>
  </Keywords>

  <Abstract>This study explores consumer behavior patterns in urban areas.</Abstract>

  <Variables>
    <Variable>
      <Name>Age</Name>
      <Label>Age of Respondent</Label>
      <Type>Discrete</Type>
      <Range>18 to 99</Range>
    </Variable>
    <Variable>
      <Name>Income</Name>
      <Label>Annual Income</Label>
      <Type>Continuous</Type>
      <Range>0 to 1,000,000</Range>
    </Variable>
  </Variables>
</DDI>
```

### *RDF (in XML)*

```xml
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dcterms="http://purl.org/dc/terms/"
    xmlns:ex="http://example.org/">

  <rdf:Description rdf:about="http://example.org/SS1234">
    <rdf:type rdf:resource="http://example.org/Survey"/>
    <dcterms:title>Sample Survey on Consumer Behavior</dcterms:title>
    <dcterms:creator>Research Institute</dcterms:creator>
    <dcterms:date>2023-10-01</dcterms:date>
    <dcterms:subject>Consumer Behavior, Survey</dcterms:subject>
    <dcterms:abstract>This study explores consumer behavior patterns in urban areas.</dcterms:abstract>
  </rdf:Description>

  <rdf:Description rdf:about="http://example.org/Age">
    <rdf:type rdf:resource="http://example.org/Variable"/>
    <dcterms:label>Age of Respondent</dcterms:label>
    <ex:type>Discrete</ex:type>
    <ex:range>18 to 99</ex:range>
  </rdf:Description>

  <rdf:Description rdf:about="http://example.org/Income">
    <rdf:type rdf:resource="http://example.org/Variable"/>
    <dcterms:label>Annual Income</dcterms:label>
    <ex:type>Continuous</ex:type>
    <ex:range>0 to 1,000,000</ex:range>
  </rdf:Description>

</rdf:RDF>
```

*JSON-LD*


```
{
  "@context": {
    "@vocab": "http://example.org/terms/"
  },
  "@type": "Dataset",
  "title": "Sample Survey Data",
  "id": "SS1234",
  "author": "Research Institute",
  "productionDate": "2023-10-01",
  "studyTitle": "Sample Survey on Consumer Behavior",
  "keywords": [
    "Consumer Behavior",
    "Survey"
  ],
  "abstract": "This study explores consumer behavior patterns in urban areas.",
  "variables": [
    {
      "name": "Age",
      "label": "Age of Respondent",
      "type": "Discrete",
      "range": "18 to 99"
    },
    {
      "name": "Income",
      "label": "Annual Income",
      "type": "Continuous",
      "range": "0 to 1,000,000"
    }
  ]
}
```


**DDI Extensions**

DDI Extensions are additional elements or modules that enhance the core DDI standard, allowing for more detailed and specialized metadata descriptions. These extensions help tailor DDI to specific data types or research needs. Extensions can include modules for specific domains, such as health, education, or environmental data. Extensions can provide additional functionality, such as detailed data provenance, complex data relationships, or integration with other metadata standards.

**Use Cases**

DDI extensions are useful in projects requiring detailed metadata for complex datasets, such as longitudinal studies or multi-disciplinary research.

DDI Extensions enhance the flexibility and applicability of the DDI standard, enabling more comprehensive and tailored metadata documentation for diverse research data.

**DDI with Extensions Example**

Consider a final metadata list with the following elements:

Study ID: Study-001
Title: Impact of Technology on Education
Abstract: This study explores the effects of technology integration in educational settings, focusing on student engagement and learning outcomes.

```
Creator:
  Name: Dr. Jane Doe
  Affiliation: University of Education

Funding:
  Agency: National Science Foundation
  Grant Number: NSF-123456

Data Collection:
  Mode: Online Surveys and Interviews
  Collection Period: 2022-01-01 to 2022-12-31


  Access Authority: Data Archive Center
  Contact: archive@datacenter.org
  Access Conditions: Open Access with Registration

Variable:
  Name: StudentEngagement
  Label: Level of Student Engagement
  Description: Measures the level of student engagement in technology-enhanced classrooms.
  Data Type: Integer
  Categories:
    1: Low
    2: Medium
    3: High

Extension:
  Publication:
    Title: Technology in Education: A Comprehensive Review
    Authors: J. Doe, A. Smith
    Journal: Journal of Educational Technology
    Year: 2023
    DOI: 10.1234/edu.2023.5678
```

## Following are metadata serializations in HTML, XML, RDF and JSON-LD:

### *HTML*

```html
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <title>Study Metadata</title>
</head>
<body>
    <h1>Study Metadata</h1>
    <h2>Study ID: Study-001</h2>
    <h3>Title: Impact of Technology on Education</h3>
    <p><strong>Abstract:</strong> This study explores the effects of technology integration in educational settings, focusing on student engagement and learning outcomes.</p>

    <h3>Creator</h3>
    <ul>
        <li><strong>Name:</strong> Dr. Jane Doe</li>
        <li><strong>Affiliation:</strong> University of Education</li>
    </ul>

    <h3>Funding</h3>
```

```html
  <ul>
    <li><strong>Agency:</strong> National Science Foundation</li>
    <li><strong>Grant Number:</strong> NSF-123456</li>
  </ul>

  <h3>Data Collection</h3>
  <ul>
    <li><strong>Mode:</strong> Online Surveys and Interviews</li>
    <li><strong>Collection Period:</strong> 2022-01-01 to 2022-12-31</li>
  </ul>

  <h3>Data Access</h3>
  <ul>
    <li><strong>Access Authority:</strong> Data Archive Center</li>
    <li><strong>Contact:</strong> <a href="mailto:archive@datacenter.org">archive@datacenter.org</a></li>
    <li><strong>Access Conditions:</strong> Open Access with Registration</li>
  </ul>

  <h3>Variable</h3>
  <ul>
    <li><strong>Name:</strong> StudentEngagement</li>
    <li><strong>Label:</strong> Level of Student Engagement</li>
    <li><strong>Description:</strong> Measures the level of student engagement in technology-enhanced classrooms.</li>
    <li><strong>Data Type:</strong> Integer</li>
    <li><strong>Categories:</strong>
      <ul>
        <li>1: Low</li>
        <li>2: Medium</li>
        <li>3: High</li>
      </ul>
    </li>
  </ul>

  <h3>Extension</h3>
  <h4>Publication</h4>
  <ul>
    <li><strong>Title:</strong> Technology in Education: A Comprehensive Review</li>
    <li><strong>Authors:</strong> J. Doe, A. Smith</li>
    <li><strong>Journal:</strong> Journal of Educational Technology</li>
    <li><strong>Year:</strong> 2023</li>
    <li><strong>DOI:</strong> <a href="https://doi.org/10.1234/edu.2023.5678">10.1234/edu.2023.5678</a></li>
  </ul>
</body>
</html>
```

### *XML*

```xml
<?xml version="1.0" encoding="UTF-8"?>
<DDI>
  <Study>
    <StudyID>Study-001</StudyID>
    <Title>Impact of Technology on Education</Title>
    <Abstract>This study explores the effects of technology integration in educational settings, focusing on student engagement and learning
outcomes.</Abstract>

    <Creator>
      <Name>Dr. Jane Doe</Name>
      <Affiliation>University of Education</Affiliation>
    </Creator>

    <Funding>
```

```xml
      <Agency>National Science Foundation</Agency>
      <GrantNumber>NSF-123456</GrantNumber>
    </Funding>

    <DataCollection>
      <Mode>Online Surveys and Interviews</Mode>
      <CollectionPeriod>2022-01-01 to 2022-12-31</CollectionPeriod>
    </DataCollection>

    <DataAccess>
      <AccessAuthority>Data Archive Center</AccessAuthority>
      <Contact>archive@datacenter.org</Contact>
      <AccessConditions>Open Access with Registration</AccessConditions>
    </DataAccess>

    <Variable>
      <Name>StudentEngagement</Name>
      <Label>Level of Student Engagement</Label>
      <Description>Measures the level of student engagement in technology-enhanced classrooms.</Description>
      <DataType>Integer</DataType>
      <Categories>
        <Category code="1">Low</Category>
        <Category code="2">Medium</Category>
        <Category code="3">High</Category>
      </Categories>
    </Variable>

    <Extension>
      <Publication>
        <Title>Technology in Education: A Comprehensive Review</Title>
        <Authors>J. Doe, A. Smith</Authors>
        <Journal>Journal of Educational Technology</Journal>
        <Year>2023</Year>
        <DOI>10.1234/edu.2023.5678</DOI>
      </Publication>
    </Extension>
  </Study>
</DDI>
```

### *RDF (in XML)*

```xml
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:ddi="http://example.org/ddi#">

  <rdf:Description rdf:about="http://example.org/study/Study-001">
    <ddi:StudyID>Study-001</ddi:StudyID>
    <ddi:Title>Impact of Technology on Education</ddi:Title>
    <ddi:Abstract>This study explores the effects of technology integration in educational settings, focusing on student engagement and learning outcomes.</ddi:Abstract>

    <ddi:Creator>
      <rdf:Description>
        <ddi:Name>Dr. Jane Doe</ddi:Name>
        <ddi:Affiliation>University of Education</ddi:Affiliation>
      </rdf:Description>
    </ddi:Creator>

    <ddi:Funding>
      <rdf:Description>
        <ddi:Agency>National Science Foundation</ddi:Agency>
        <ddi:GrantNumber>NSF-123456</ddi:GrantNumber>
      </rdf:Description>
```

```
    </ddi:Funding>

    <ddi:DataCollection>
      <rdf:Description>
        <ddi:Mode>Online Surveys and Interviews</ddi:Mode>
        <ddi:CollectionPeriod>2022-01-01 to 2022-12-31</ddi:CollectionPeriod>
      </rdf:Description>
    </ddi:DataCollection>

    <ddi:DataAccess>
      <rdf:Description>
        <ddi:AccessAuthority>Data Archive Center</ddi:AccessAuthority>
        <ddi:Contact>archive@datacenter.org</ddi:Contact>
        <ddi:AccessConditions>Open Access with Registration</ddi:AccessConditions>
      </rdf:Description>
    </ddi:DataAccess>

    <ddi:Variable>
      <rdf:Description>
        <ddi:Name>StudentEngagement</ddi:Name>
        <ddi:Label>Level of Student Engagement</ddi:Label>
        <ddi:Description>Measures the level of student engagement in technology-enhanced classrooms.</ddi:Description>
        <ddi:DataType>Integer</ddi:DataType>
        <ddi:Categories>
          <rdf:Seq>
            <rdf:li rdf:parseType="Resource">
              <ddi:code>1</ddi:code>
              <ddi:label>Low</ddi:label>
            </rdf:li>
            <rdf:li rdf:parseType="Resource">
              <ddi:code>2</ddi:code>
              <ddi:label>Medium</ddi:label>
            </rdf:li>
            <rdf:li rdf:parseType="Resource">
              <ddi:code>3</ddi:code>
              <ddi:label>High</ddi:label>
            </rdf:li>
          </rdf:Seq>
        </ddi:Categories>
      </rdf:Description>
    </ddi:Variable>

    <ddi:Extension>
      <rdf:Description>
        <ddi:Publication>
          <rdf:Description>
            <ddi:Title>Technology in Education: A Comprehensive Review</ddi:Title>
            <ddi:Authors>J. Doe, A. Smith</ddi:Authors>
            <ddi:Journal>Journal of Educational Technology</ddi:Journal>
            <ddi:Year>2023</ddi:Year>
            <ddi:DOI>10.1234/edu.2023.5678</ddi:DOI>
          </rdf:Description>
        </ddi:Publication>
      </rdf:Description>
    </ddi:Extension>
  </rdf:Description>

</rdf:RDF>
```


*JSON-LD*


```
{
```

```
 "@context": {
  "@vocab": "http://purl.org/dc/terms/",
  "ddi": "http://example.org/ddi#"
 },
 "@type": "ddi:Study",
 "ddi:StudyID": "Study-001",
 "title": "Impact of Technology on Education",
 "abstract": "This study explores the effects of technology integration in educational settings, focusing on student engagement and learning outcomes.",
 "creator": {
  "@type": "Person",
  "name": "Dr. Jane Doe",
  "affiliation": "University of Education"
 },
 "funding": {
  "@type": "Funding",
  "funder": "National Science Foundation",
  "grantNumber": "NSF-123456"
 },
 "ddi:DataCollection": {
  "ddi:Mode": "Online Surveys and Interviews",
  "ddi:CollectionPeriod": "2022-01-01 to 2022-12-31"
 },
 "ddi:DataAccess": {
  "ddi:AccessAuthority": "Data Archive Center",
  "ddi:Contact": "archive@datacenter.org",
  "ddi:AccessConditions": "Open Access with Registration"
 },
 "ddi:Variable": {
  "ddi:Name": "StudentEngagement",
  "ddi:Label": "Level of Student Engagement",
  "ddi:Description": "Measures the level of student engagement in technology-enhanced classrooms.",
  "ddi:DataType": "Integer",
  "ddi:Categories": [
   {"ddi:code": "1", "ddi:label": "Low"},
   {"ddi:code": "2", "ddi:label": "Medium"},
   {"ddi:code": "3", "ddi:label": "High"}
  ]
 },
 "ddi:Extension": {
  "ddi:Publication": {
   "title": "Technology in Education: A Comprehensive Review",
   "authors": ["J. Doe", "A. Smith"],
   "journal": "Journal of Educational Technology",
   "year": 2023,
   "doi": "10.1234/edu.2023.5678"
  }
 }
}
```


## 3.4 Learning Summary

Concepts covered in this chapter include application of domain-agnostic and domain-specific schema models.

## 3.5 Knowledge Self-Assessments

**Questions**

1. Match the metadata schema with its description:
    a. Metadata schema:
        i. Dublin Core
        ii. Darwin Core
        iii. Encoded Archival Description
        iv. Data Documentation Initiative
    b. Description:
        i. XML standard for encoding archival finding aids
        ii. Standard for describing data produced by surveys in social sciences
        iii. Schema defining DCMI term declarations for web resources
        iv. Glossary of terms for sharing information about biological diversity

2. What is the primary purpose of Darwin Core Metadata?
    a. To describe consumer behavior data
    b. To document biodiversity data
    c. To analyze economic trends
    d. To manage educational resources

3. What does DDI stand for?
    a. Data Documentation Initiative
    b. Data Development Interface
    c. Digital Data Integration
    d. Dynamic Data Information
    e. None of the above

4. What are the core elements of Darwin Core Metadata?

5. What are Darwin Core Extensions, and how do they enhance the core standard?

6. What is DDI, and what fields does it cover?

7. What are the basic elements of DDI metadata?

8. How do DDI Extensions provide additional functionality to the core standard?

**9.** What are the benefits of using community-endorsed metadata standards like Darwin Core and DDI?

**Solutions**

1. Correct matches:
    a. Dublin Core: iii) Schema defining DCMI term declarations for web resources
    b. Darwin Core: iv) Glossary of terms for sharing information about biological diversity
    c. Encoded Archival Description: i) XML standard for encoding archival finding aids
    d. Data Documentation Initiative: ii) Standard for describing data produced by surveys in social sciences.

2. Answer: b) To document biodiversity data

3. Answer: e) None of the above

4. Answer: The core elements include Title, Creator, Description, Keywords, Citation, Contact Information, and Rights.

5. Answer: Darwin Core Extensions are additional terms that complement the core standard, allowing for more detailed and customized descriptions of biodiversity data.

6. Answer: DDI is an international standard for documenting data in social, behavioral, and economic sciences, covering study design, data collection methods, variables, and data processing.

7. Answer: Basic elements of DDI metadata include Title, ID, Author, Production Date, Study Title, Keywords, Abstract, and Variables with details such as name, label, type, and range.

8. Answer: DDI Extensions allow for more detailed metadata descriptions, such as data provenance and complex data relationships, tailored to specific research needs.

9. Answer: These standards facilitate data sharing, discovery, and integration, ensuring interoperability across platforms and enhancing research and conservation efforts.

## Glossary of Terms

1. Administrative Metadata: Includes information about rights, technical aspects, and preservation.
2. Apache License: A permissive free software license written by the Apache Software Foundation.
3. CEDAR Workbench: A tool for managing metadata using templates, supporting efficient data entry and sharing.
4. Citation: Information on how to cite a dataset.
5. Contact Information: Details for reaching the dataset's custodian or creator.
6. Contact Information: Identity and means to communicate with persons or entities associated with the data.
7. Controlled Vocabulary: A standardized set of terms and codes used in metadata.
8. Core Elements: Basic components of a metadata schema, such as title, creator, and description.
9. Creative Commons (CC): A licensing system that allows creators to communicate which rights they reserve and which they waive for the benefit of recipients.
10. Crosswalk: A method to translate elements and values from one metadata schema to another to ensure interoperability.
11. Darwin Core: A standardized framework for describing biodiversity data.
12. Data Documentation Initiative (DDI): A standard for describing observational and survey data in social, behavioral, economic, and health sciences.
13. encoding data.
14. Data Use Agreements (DUA): Custom agreements specifying terms for data sharing, often used in research collaborations.
15. Description: A summary of the dataset's content and purpose.
16. Descriptive Metadata: Describes a resource, its content, and identifying characteristics.
17. Domain Use Cases: Specific applications of a metadata standard in a particular field.
18. Dublin Core: A set of vocabulary terms used to describe web resources and other digital or physical resources.
19. Encoded Archival Description (EAD): An XML standard for encoding archival finding aids.
20. Extensions: Additional terms that complement a core metadata standard for more detailed descriptions.
21. FAIR: Principles ensuring data is Findable, Accessible, Interoperable, and Reusable.
22. FITS (Flexible Image Transport System): A standard data format used in astronomy.
23. GNU General Public License (GPL): A widely used free software license that guarantees end users the freedom to run, study, share, and modify the software.
24. Human and Machine-readable: Metadata that can be interpreted by both humans and machines.
25. ISA Creator: A tool for planning and describing experiments, facilitating data export and import.
26. Keywords: Terms that describe the dataset's subject matter.
27. Licensing Models: Frameworks that specify how data can be used, shared, and modified, such as Creative Commons or Open Data Commons.
28. Metadata: Data that provides information about other data, enhancing discoverability and usability.
29. Metadata: Deliberate, structured data about a data or an object resource, providing meaningful information about the resource.
30. Metadata: Data that provides information about other data, used for discovery, identification, and management of resources.
31. Metadata Registry: Where metadata definitions are stored and maintained.
32. Metadata Repository: Where actual instance metadata is stored, such as in a Data Catalog.

33. MIT License: A permissive free software license originating at the Massachusetts Institute of Technology.
34. MODS (Metadata Object Description Schema): A schema for a bibliographic element set used in library applications.
35. NeXus: A common data format for neutron, x-ray, and muon science.
36. Open Data Commons: A set of licenses designed to provide legal tools for sharing data.
37. Persistent Identifier: A unique identification code, such as a Digital Object Identifier (DOI), assigned to a resource.
38. Persistent Identifier (PID): A long-lasting reference to a digital resource, ensuring consistent access and citation.
39. Prospective Metadata: Metadata created before the data is generated, aiding in capturing necessary aspects during data collection.
40. Repository: A storage location for data, which can be generalist (broad range) or specialist (specific types).
41. Retrospective Metadata: Metadata developed after a resource is created.
42. RightField: A tool for embedding ontology terms within Excel spreadsheets for metadata annotation.
43. Rights: Information about the dataset's usage rights and licenses.
44. Rights Holder: Entities or persons who hold the rights to the data.
45. Schema: Structured organization of metadata, often referred to as its model.
46. Schema: A structured framework or plan that defines the organization of metadata elements.
47. Sensitivity and Specificity: Measures of how effectively a metadata schema can locate resources in a repository.
48. Structural Metadata: Describes how pieces of a single resource fit together or relate to other resources.
49. Terminology Mapping: The process of aligning terms from different vocabularies or coding systems to ensure consistency.
50. Title: The name of the dataset.
51. UMLS (Unified Medical Language System): A resource for mapping medical terminologies and coding systems.
52. Value Standards: Established standardized terms and codes used in metadata.

## Supplementary Technical References and Training Materials

The list below contains supplementary material in addition to links and footnotes found throughout this workbook.


1. Horodyski, John (CRC Press, 2022)**: "Metadata Matters."**
2. Jeffrey Pomerantz (MIT Presss,2015**):"Metadata."**
3. Steven Jack Miller (ALA Neal-Schuman2022): "Metadata for Digital Collections (How-To-Do-It Manuals."
4. Gilliland, A. J. (2016). **"Setting the Stage." In *Introduction to Metadata*. Getty Publications.**
5. Baca, M. (Ed.). (2016). ***Introduction to Metadata*. Getty Publications.**
6. Caplan, P. (2003). ***Metadata Fundamentals for All Librarians*. American Library Association.**
7. NISO. (2004). ***Understanding Metadata*. National Information Standards Organization.**
8. Haynes, D. (2018). ***Metadata for Information Management and Retrieval: Understanding Metadata and its Use*. Facet Publishing.**
9. Greenberg, J. (2005). **"Understanding Metadata and Metadata Schemes." *Cataloging & Classification Quarterly*, 40(3-4), 17-36.**
10. Lagoze, C., & Van de Sompel, H. (2001). **"The Open Archives Initiative: Building a low-barrier interoperability framework." *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries*.**
11. Dublin Core Metadata Initiative. **"Dublin Core Metadata Element Set, Version 1.1." Available at: http://dublincore.org/documents/dces/.**
12. ISO/IEC 11179. ***Information Technology - Metadata Registries (MDR)*. International Organization for Standardization.**
13. Borgman, C. L. (1996). **"Why Are Online Catalogs Still Hard to Use?" *Journal of the American Society for Information Science*, 47(7), 493-503.**
14. VRA Core. **"VRA Core Categories." Visual Resources Association. Available at: http://vraweb.org/projects/vracore4/**
15. PREMIS Editorial Committee. (2015). ***PREMIS Data Dictionary for Preservation Metadata*. Library of Congress.**
16. Smith, M., & Lanzi, E. (2017). ***Metadata*. MIT Press Essential Knowledge Series.**
17. Riley, J. (2017). ***Understanding Metadata: What is Metadata, and What is it For?* National Information Standards Organization.**
18. Chan, L. M., & Zeng, M. L. (2006). **"Metadata Interoperability and Standardization – A Study of Methodology Part I." *D-Lib Magazine*, 12(6).**


These supplementary websites offer additional resources, standards, and guidelines related to metadata.

1. Library of Congress - Metadata Standards loc.gov/standards/metadata.html
2. National Information Standards Organization (NISO) - niso.org
3. Getty Research Institute - Introduction to Metadata - getty.edu/research/publications/electronic_publications/intrometadata/
4. Visual Resources Association (VRA) Core - vraweb.org/projects/vracore4/
5. Open Archives Initiative - openarchives.org
6. PREMIS - Preservation Metadata - loc.gov/standards/premis/
7. Metadata Encoding and Transmission Standard (METS) - loc.gov/standards/mets/
8. World Wide Web Consortium (W3C) - Metadata Activity - w3.org/Metadata/

9. International Organization for Standardization (ISO) - Metadata Registries - iso.org/iso/catalogue_detail.htm?csnumber=35343

Complementary video training materials have been prepared by ODSS, NIH and are available upon request.  These include:

1. Why Metadata Matters
2. Jointly Developing an Approach to FAIRness
3. Good Data Stewardship
4. Leveraging Publicly Accessible Data

## List of Acronyms

CC: Creative Commons

DCMI: Dublin Core Metadata Initiative

DDI: Data Documentation Initiative

DOI: Digital Object Identifier

DUA: Data Use Agreements

EAD: Encoded Archival Description

EML: Ecological Metadata Language

FAIR: Findable, Accessible, Interoperable, Reusable

FITS: Flexible Image Transport System

FOAF: Friend of a Friend

GPL: GNU General Public License

HTML: Hypertext Markup Language

ISA: Investigation, Study, Assay

JSON: JavaScript Object Notation

LCNAF Library of Congress Name Authority File LCNAF

METS: Metadata Encoding and Transmission Standard

MIT: Massachusetts Institute of Technology

MODS: Metadata Object Description Schema

NISO: National Information Standards Organization

NLM: National Library of Medicine

ORCID: Open Researcher and Contributor ID

PID: Persistent Identifier

PROV-O: Provenance Ontology

RAID: Research Activity Identifier

RDF: Resource Description Framework

ROR: Research Organization Registry

RRID: Research Resource Identifiers

SKOS: Simple Knowledge Organization System

UMLS: Unified Medical Language System

URL: Uniform Resource Locator

VRA: Visual Resources Association

VIAF: Virtual International Authority File

W3C: World Wide Web Consortium

XML: Extensible Markup Language