

# THE ODUM INSTITUTE

---

FOR RESEARCH IN SOCIAL SCIENCE

## Odum Institute Data Archive METADATA GUIDELINES

### Introduction

Metadata, often referred to as data about data, is defined more formally as, “structured information that describes, locates, or otherwise makes it easier to retrieve, use, or manage an information resource” (NISO, 2004). The Odum Institute Data Archive generates machine-readable metadata for all datasets using standardized metadata schemas and controlled vocabularies. Metadata at the dataset, file, and variable level (for tabular datasets) are preserved alongside the data to ensure that the data are identifiable, discoverable, accessible, and usable into the long-term future. The Odum Institute Data Archive also employs standard metadata protocols that enable archive system interoperability for metadata harvesting.

Odum Institute Data Archive Metadata Guidelines are informed by the [Data Preservation Alliance for the Social Sciences \(Data-PASS\) Metadata Requirements](#).

### Functions of Metadata

The standardized metadata generated for data in the Odum Institute Data Archive collections serve the following primary functions:

- **Resource discovery, identification, and citation.** Metadata that identifies the data creator, title, data production date, persistent identifier (DOI), and publisher enables users to locate the data and verify that the data discovered is data the user was seeking. Standardization of these metadata enables Odum Institute Data Archive systems to automatically generate a formal data citation.
- **Provision of value-added services.** Variable-level metadata enables the Odum Institute Data Archive systems to offer additional functionality including data subsetting, and exploratory data analysis and visualization in the Archive user interface.
- **Resource location.** Standardized metadata enables the Odum Institute Data Archive to arrange datasets and files into logical collections that facilitate dataset browsing, search, and navigation.
- **Resource administration.** Metadata that captures information about the data type, file format, and file checksums are necessary to execute digital preservation strategies that maintain the integrity of the data during processes such as migration that apply changes to data files.
- **Public data dissemination.** The machine-readability of standards-compliant metadata allows for interoperability among other data archives systems. The Odum Institute Data Archive allows partners to harvest metadata to include in their repository catalogs, which extends the reach of the data to a broader community of potential users.
- **Access control.** Data terms of use stored as metadata alongside the data ensures that the Odum Institute Data Archive and its systems properly enforce access restrictions and other limitations on data access and use.

## Types of Metadata

The Odum Institute Data Archive generates different types of metadata that describe data in its collections at varying levels of granularity in order to capture and preserve the information necessary for long-term discovery, identification, management, and use of the data.

To provide comprehensive data description, metadata are generated at the following levels of granularity:

- **Dataset level.** A dataset refers to a collection of data files produced from a study or a compilation of data files brought together at a single time or for a single purpose. A dataset often consists of more than one data file.
- **File level.** A file is a digital object containing a sequence of bits representing the data, documentation, or other related resource.
- **Variable level.** A variable is the set of observations, using a single measure, which is collected during a research study and contained in the data file.

Archival best practices distinguish three types of metadata:

- **Descriptive metadata** identifies and describes a resource for the primary purpose of enabling discovery and identification of the resource.
- **Structural metadata** describes the structure of the resource to support use of the resource.
- **Administrative metadata** describes the management of data over time. This includes information on data processing actions and access control requirements.

## Metadata Standards

The Odum Institute Data Archive has adopted standard metadata schemas and protocols that are in widespread use by the professional data archiving community. Below are the standards the Odum Institute Data Archive and its systems follow.

- **[DataCite Metadata Schema](#).** The DataCite Metadata Schema includes a core set of descriptive metadata elements that support accurate identification for data citation and retrieval.
- **[Data Documentation Initiative \(DDI\) Metadata Specification](#).** The DDI Metadata Specification documents data across the data lifecycle. DDI includes variable-level metadata that enables value-added data subsetting and analysis functionality in the Odum Institute Data Archive system.
- **[Dublin Core Metadata Initiative \(DCMI\) Specification](#).** DCMI is a set of basic descriptive metadata that supports interoperability among different data archives systems.
- **[Open Archives Protocol for Metadata Harvesting \(OAI-PMH\)](#).** The OAI-PMH protocol is used by the Odum Institute Data Archive to expose its metadata for harvesting by other archives systems for inclusion in their catalogs.

The Odum Institute Data Archive system also enables the generation of additional domain-specific descriptive metadata using the prevailing metadata standards in those disciplinary domains.

## Metadata Requirements

The Odum Institute Data Archive requires a minimum set of metadata to enable data discovery, access, and preservation. However, the Data Archive includes and strongly encourages data depositors to include additional metadata to enhance appropriate interpretation and reuse of the data. The table below lists the minimum required metadata that must be provided with each dataset.

Metadata Field	Description	Notes
<b>Identifier</b>	A persistent identifier that uniquely identifies the dataset	Digital Object Identifier (DOI) is automatically generated upon dataset record creation
<b>Title</b>	Full title by which the dataset is known	Format: Open text
<b>AuthorName</b>	The person, corporate body, or agency responsible for creating the work	For individuals, required format is: FamilyName, GivenName
<b>ContactEmail</b>	The email address used to submit inquiries to the contact person for the dataset	Format: Email address
<b>Description</b>	A summary describing the purpose, nature, and scope of the dataset	Format: Open text with HTML tag support
<b>Subject</b>	Domain-specific subject category(ies) that are topically relevant to the dataset	Controlled vocabulary: <ul style="list-style-type: none"> <li>- Mathematical Sciences</li> <li>- Physics</li> <li>- Chemistry</li> <li>- Computer and Information Science</li> <li>- Astronomy and Astrophysics</li> <li>- Business and Management</li> <li>- Arts and Humanities</li> <li>- Other</li> <li>- Medicine, Health, and Life Sciences</li> <li>- Earth and Environmental Sciences</li> <li>- Social Sciences</li> <li>- Engineering</li> <li>- Law</li> <li>- Agricultural Sciences</li> </ul>
<b>PublicationDate</b>	Date when the dataset was published (i.e., made publicly accessible) in the archival system	Date is automatically generated upon dataset publication
<b>TermsOfUse</b>	Description of allowable uses of the dataset including access restrictions and citation requirements	Datasets default to a <a href="#">CC0 Public Domain Dedication</a> unless custom terms of use are provided otherwise

## Guidelines Review

The Odum Institute Data Archive Metadata Guidelines are subject to three-year review. The current guidelines were approved and issued on May 1, 2017.