# Dataverse North Metadata Best Practices Guide

Version 3.0
September 2021

*By Dataverse North Metadata Group*

Prepared by the Portage Network Dataverse North Metadata Working Group for the New Digital Research Infrastructure Organization (NDRIO)

# Table of Contents

# Introduction

One of the best features of the Dataverse repository software is the large number of metadata fields it provides for describing research data. Standards-based for interoperability, Dataverse metadata supports both dataset and file-level descriptions and is compliant with the DataCite[1] schema to support DOI registration. It draws principal influence from DDI Codebook[2] while incorporating metadata standards from other domains, making it well suited for describing data from the social, economic, behavioural, and health sciences; it is also easily adapted for use in the humanities, and in pure, applied, and environmental sciences. However, this flexibility can sometimes make Dataverse metadata seem complicated. This guide provides direction to both the novice and experienced user in creating metadata for datasets in a Dataverse repository.

The Dataverse Project is an open source web application to share, preserve, cite, explore, and analyze research data. It is developed by the Institute for Quantitative Social Sciences (IQSS) at Harvard University, along with many collaborators and contributors worldwide.[3]

**Guide Organization**

This guide is organized according to the six metadata blocks available in the current version of the Dataverse software:

- Citation metadata block
    - The core metadata required to publish a dataset.
- Geospatial metadata block
    - Metadata that describes the spatial extent (e.g., location information) of data in a dataset.
- Social Science & Humanities metadata block
    - Metadata that describes survey data and related datasets in the social science and humanities.
- Astronomy & Astrophysics metadata block
    - Metadata specific to astronomical or astrophysical data.
- Life Sciences metadata block
    - Metadata that describes scientific research in multiple domains that focus on the study of living organisms including biology, zoology, microbiology, physiology, biochemistry, and related subjects.
- Journal metadata block

---

[1] "The DataCite Metadata Schema is a list of core metadata properties chosen for an accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions" (https://schema.datacite.org/). "DataCite is a leading global non-profit organisation that provides persistent identifiers (DOIs) for research data and other research outputs" (https://datacite.org/mission.html).

[2] "DDI Codebook was the first version of the DDI specification to be published" (www.ddialliance.org/Specification/DDI-Codebook/). The Data Documentation Initiative (DDI) is an effort to create an international standard for describing data from the social, behavioral, and economic sciences. Expressed in XML (http://www.w3.org/XML/), the DDI metadata specification now supports the entire research data life cycle. DDI metadata accompanies and enables data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving. The DDI Alliance (http://www.ddialliance.org/) oversees the development of the DDI metadata standard.

[3] About | The Dataverse Project – Dataverse.org. See: https://dataverse.org/about

- Metadata used by journals using a Dataverse repository as its data repository. Its fields identify journal volume, issue, and article type information that a dataset is associated with.

Each section of the guide begins by summarizing information specific to the metadata block. This is followed by a table that describes all metadata fields in the block according to the following headings:

- Field
  - The name of the field as it appears in the Dataverse interface.
- Description with tips
  - The definition of the field. This definition is provided by the Dataverse software developers and can be seen by hovering over the field name in the Dataverse interface.
  - Tips and notes that help users better understand the field, its definition or usage. These have been written by the authors of this guide.
- Usage
  - Suggestions based on best practices that help users determine if a field is required, recommended, or optional. More information on usage guidelines are provided below.
- Repeatable
  - An indicator to help users know whether a field can be repeated (e.g., for data that may have multiple authors or keywords) or not.
- Example
  - Text that demonstrates the kind of information entered in this field. These examples are provided to place each metadata field in context.
  - For each metadata block, examples have been included to place each field in context. Because of differences between each metadata block, the use of examples is handled slightly differently.
    - Citation block
      The examples are based on a [fictitious dataset](#)[4] available from the Scholars Portal Dataverse demo repository.
    - Geospatial block
      The examples are fictitious but screen shots from actual datasets in the Scholars Portal Dataverse repository are included after the table. Edit screens showing how the fields look when a dataset is published are also included.
    - Social Science and Humanities block
      There are multiple real examples provided for each field with links to the datasets in the Scholars Portal Dataverse repository so that the examples can be seen in context.
    - Astronomy and Astrophysics block
      The examples are samples from the IVOA-RM specifications, which are supplemented with real examples from the [Harvard Dataverse repository](#)[5]

---

[4] https://doi.org/10.5072/FK2/TOXB6Q

[5] https://dataverse.harvard.edu

- Life Sciences block
  The examples are fictitious but screen shots from actual datasets in the Harvard Dataverse repository are included.
- Journal block
  The examples are from a dataset in the Harvard Dataverse repository.

This guide (version 3.0) corresponds to version 5.1 of the Dataverse software (released December 2020).

**Usage**

In the tables for each metadata block below, the column *Usage* indicates whether a field is required, recommended, or optional.

- Required [RQ] - Required fields have been designated as such by the developers of the Dataverse software, which is configured accordingly out-of-the-box. These fields are Title, Author-name, Contact email, Description-text, and Subject.
- Recommended [R] - Recommended fields have been designated as such by the authors of this guide based on their understanding of accepted best practices to improve data discovery and reuse. If a recommended field is not applicable to the dataset being described, it does not need to be included.
- Optional [O] - Optional fields can be used to provide additional information about a dataset. In general, it is good practice to provide additional metadata when the fields are applicable and the information is available.

**Duplication across metadata fields**

It is possible that metadata information may be the same in multiple fields. For example, the author of a dataset may also be the producer. It is important that all relevant fields get filled in even if it does repeat information. Having all possible fields filled in will help with the findability of a dataset because you may not know how users may approach searching.

Fields not designated as required can be made required by changing the settings in the Dataverse installation. This can be done by the Dataverse Administrators. See Dataverse Management in the Dataverse Users Guide for details.[6]

**Metadata sources**

Dataverse metadata is primarily influenced by the DDI Codebook[7] while incorporating additional metadata standards to better ensure preservation and interoperability. As with many platforms, the Dataverse software's usage of a metadata standard does not necessarily mean that it implements the full breadth of a standard or that its usage is in full compliance with the standard itself.

---

[6] http://guides.dataverse.org/en/latest/user/dataverse-management.html

[7] https://ddialliance.org

The following table summarizes the Dataverse software's supported metadata schemas, as found in the [Metadata References appendix](#) of the *Dataverse Project User Guide*:

| Metadata Block | Schemas employed |
|---|---|
| Citation | Compliant with DDI Lite, DDI 2.5 Codebook, DataCite 3.1, and Dublin Core's DCMI Metadata Terms. Language field uses ISO 639-1 controlled vocabulary |
| Geospatial | Compliant with DDI Lite, DDI 2.5 Codebook, DataCite, and Dublin Core. Country / Nation field uses ISO 3166-1 controlled vocabulary |
| Social Science & Humanities | Compliant with DDI Lite, DDI 2.5 Codebook, and Dublin Core |
| Astronomy and Astrophysics | Mappable to the Virtual Observatory (VO) Discovery and Provenance Metadata, which is an implementation of the International Virtual Observatory Alliance's [Resource Metadata for the Virtual Observatory, Version 1.12 (2007)](#)[8] specification, referred to as IVOA-RM in this document |
| Life Sciences | Based on ISA-Tab Specification, along with controlled vocabulary from subsets of the OBI Ontology and the NCBI Taxonomy for Organisms |
| Journal Metadata | Based on the Journal Archiving and Interchange Tag Set, version 1.2 |

**Controlled vocabularies**

A controlled vocabulary is a standardized list of terminology for describing information. Fields that have or require controlled vocabularies are noted in the Definition column of each table. The following list provides information about controlled vocabularies that are either referenced by Dataverse documentation or recommended in this guide.

- [ISO 639-1](#)[9] (Citation language field)
- [DDI Alliance Controlled Vocabularies](#)[10] (Social Science & Humanities)
- [GeoNames.org](#)[11] (Geospatial - State / Province and City fields)
- [FAST](#)[12] (Citation - Keyword and Topical Classification fields)
- [RVMFAST](#)[13] (Citation - Keyword and Topical Classification fields)

---

[8] [http://www.ivoa.net/documents/latest/RM.html](http://www.ivoa.net/documents/latest/RM.html)

[9] [https://www.loc.gov/standards/iso639-2/php/English_list.php](https://www.loc.gov/standards/iso639-2/php/English_list.php)

[10] [https://ddialliance.org/controlled-vocabularies](https://ddialliance.org/controlled-vocabularies)

[11] [https://www.geonames.org](https://www.geonames.org)

[12] [http://experimental.worldcat.org/fast/](http://experimental.worldcat.org/fast/)

[13] [https://rvmweb.bibl.ulaval.ca/rvmfast/initMoteurRecherche.do](https://rvmweb.bibl.ulaval.ca/rvmfast/initMoteurRecherche.do)

- [ISO 3166-1](#)[14] (Geospatial - Country/ Nation field)
- [Virtual Observatory (VO) Discovery and Provenance Metadata](#)[15] (Astronomy and Astrophysics)
- [Space-Time Coordinate (STC) Metadata Recommendation](#)[16] (Astronomy and Astrophysics)
- [OBI Ontology](#)[17] (Life Sciences)
- [NCBI Taxonomy for Organisms](#)[18] (Life Sciences)

**Language of the metadata**

The development of the Internationalization feature (added in January 2019) allows the Dataverse software to be used in multiple languages. The Scholars Portal Dataverse repository, for example, provides a bilingual platform interface with a French/English toggle.

In addition to the platform, users should consider which language to use for the metadata. This decision rests, in most cases, with the dataset owner. It is recommended that the language of the metadata be selected to maximize discovery by the intended audiences. In some cases, it will be beneficial to enter metadata in more than one language, e.g., by taking advantage of the "Alternative Title" field, and using repeatable fields such as Description, Keyword, and Geospatial Coverage. There is no field to indicate the language of the metadata.

**Credits**

This guide is based on [Dataverse Metadata v4+ documentation](#)[19] and the table layout is from the [Texas Digital Library](#).[20] Best practice guides were consulted to ensure consistency with existing metadata practices and included the [DataCite Schema](#),[21] [<odesi> Best Practices Document](#),[22] and [Federated Research Data Repository](#).[23]

**Versions**

- Version 1, April 2019 - general citation metadata block for the Dataverse software, version 4.x (EN - https://doi.org/10.14288/1.0386750 / FR - https://doi.org/10.14288/1.0386751)
  Authors: Alexandra Cooper, Ève Paquette-Bigras, Martine Gagnon, Amber Leahey, Laure Perrier, Michael Steeleworthy, Sally Taylor

---

[14] https://en.wikipedia.org/wiki/ISO_3166-1

[15] https://perma.cc/H5ZJ-4KKY

[16] http://www.ivoa.net/documents/latest/STC.html

[17] http://bioportal.bioontology.org/ontologies/OBI

[18] http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html

[19] https://docs.google.com/spreadsheets/d/13HP-jI_cwLDHBetn9UKTREPJ_F4iHdAvhjmlvmYdSSw/edit#gid=0

[20] https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/493551668/Metadata+Dictionary

[21] https://schema.datacite.org

[22] https://drive.google.com/file/d/0Bxw1dFFWfU7URkgwYUxPYVV2OTg/view

[23] https://www.frdr-dfdr.ca/repo

- Version 1.1, June 2019 - general citation metadata block for the Dataverse software, version 4.x; updated to include corrections on the following fields - contact name and affiliation, ID type, ID number, producer name. (EN - https://doi.org/10.14288/1.0386726 / FR - https://doi.org/10.14288/1.0386727)
  Authors: Alexandra Cooper, Martine Gagnon, Mark Goodwin, John Huck, Amber Leahey, Michael Steeleworthy, Sally Taylor
- Version 2, February 2020 - domain specific metadata blocks for geospatial, and social science and humanities added. (EN - https://doi.org/10.14288/1.0388724 / FR - https://doi.org/10.14288/1.0388725)
  Authors: Teresa Bascik, Philippe Boisvert, Alexandra Cooper, Martine Gagnon, Mark Goodwin, John Huck, Amber Leahey, Michael Steeleworthy, Sally Taylor
- Version 3, October 2021 - domain specific metadata blocks for Astronomy and Astrophysics and Life Sciences added; introduction rewritten to include information about all the metadata blocks and disciplinary information moved to the appropriate block. Each of the previous sections were also reviewed to make sure they complied with changes/improvements made in the new, updated version.
  Authors: Teresa Bascik, Philippe Boisvert, Alexandra Cooper, Martine Gagnon, Mark Goodwin, John Huck, Amber Leahey, Kelly Stathis, Michael Steeleworthy

**Questions?**
Please see Portage Network's library contact at your institution.[24]

---

[24] https://portagenetwork.ca/tools-and-resources/institutional-rdm-contacts

# Citation Metadata

## Introduction

The Citation Metadata Block includes the core metadata needed to publish a dataset in a Dataverse repository. The required fields in this section are used to create the citation for the dataset.

The examples used in this section are from a [fictitious dataset](#) available from the Scholars Portal Dataverse demo repository.[25]

According to Dataverse documentation, Citation Metadata fields are compliant with DDI Lite, DDI 2.5 Codebook, DataCite 3.1, and Dublin Core's DCMI Metadata Terms (see .tsv version). The Language field uses ISO 639-1 controlled vocabulary.[26]

Note: An absolute URL is a complete address including its prefix (https://), as seen in provided examples.

---

[25] https://doi.org/10.5072/FK2/TOXB6Q

[26] See: http://guides.dataverse.org/en/latest/user/appendix.html

## Citation Metadata Block

| Field | Definition with tips | Usage[27] | Repeatable | Example |
|---|---|---|---|---|
| **Title** | Full title by which the Dataset is known. | RQ | No | Social Media Use Among Teens, 2015 [Canada] |
| **Subtitle** | A secondary title used to amplify or state certain limitations on the main title. *Tip: The subtitle is not included in the auto-generated citation. If you want the subtitle to be included in the citation, then you must add it to the Title field in addition to the Subtitle field. the Subtitle field.* | O | No | Main Survey |
| **Alternative Title** | A title by which the work is commonly referred, or an abbreviation of the title. *Tip: Acronym, short form, or translation of full title.* | O | No | Youth Social Media Survey |
| **Alternative URL** | A URL where the dataset can be viewed, such as a personal or project website. | O | No | https://youthsocialmedia.org |
| **Other ID** | Another unique identifier that identifies this Dataset (e.g., producer's or another repository's number). | | | |
| Agency | Name of agency which generated this identifier. | O | Yes | Youth Communication Development Project, Education Department, Queen's University |
| Identifier | Other identifier that corresponds to this Dataset. | O | Yes | 2202 |
| **Author** | The person(s), corporate body(ies), or agency(ies) responsible for creating the work. | | | |
| Name | The author's Family Name, Given Name or the name of the organizatfion responsible for this Dataset. | RQ | Yes | Doe, Jane |
| Affiliation | The organization with which the author is affiliated. *Tip: Use the full official name of the organization; avoid abbreviations.* | R | Yes | Queen's University |
| Identifier Scheme | Name of the identifier scheme (ORCID, ISNI). *Note: This field is a drop-down list; select one.* | R | Yes | ORCID |
| Identifier | Uniquely identifies an individual author or organization, according to various schemes. | R | Yes | https://orcid.org/0000-0002-1825-0097 |
| **Contact** | The contact(s) for this Dataset. | | | |
| Name | The contact's Family Name, Given Name or the name of the organization. | R | Yes | Doe, Jane |

---

| Field | Definition with tips | Usage[27] | Repeatable | Example |
|---|---|---|---|---|
| Affiliation | The organization with which the contact is affiliated.<br>*Tip: use the full official name of the organization; avoid abbreviations.* | R | Yes | Queen's University |
| E-mail | The e-mail address(es) of the contact(s) for the Dataset. This will not be displayed. | R | Yes | jdoe@email.com |
| **Description** | A summary describing the purpose, nature, and scope of the Dataset. | | | |
| Text | A summary describing the purpose, nature, and scope of the Dataset.<br>*Tip: A good description will identify the content of the dataset and help a user determine if they have the technical ability to use it.[28]* | RQ | Yes | The Social Media Use Among Teens survey was conducted by the Youth Communication Development Project to understand social media communication behaviours among youth in Canada. The survey collected responses from Canadian youth using an online questionnaire that asks about social media use including, platform type, frequency of use, activity type, and location of use. This information is supplemented with the respondent's demographic and household characteristics. |
| Date | In cases where a Dataset contains more than one description (for example, one might be supplied by the data producer and another prepared by the data repository where the data are deposited), the date attribute is used to distinguish between the two descriptions. The date attribute follows the ISO convention of YYYY-MM-DD. | O | Yes | 2018-01-18 |
| **Subject** | Domain-specific Subject Categories that are topically relevant to the Dataset.<br>*Note: This field is a drop-down list; select all that apply.*<br>*Tip: Select one or more Subject terms to represent the broad domain or domains the dataset relates to.* | RQ | No | Social Sciences |
| **Keyword** | Key terms that describe important aspects of the Dataset. | | | |
| Term | Key terms that describe important aspects of the Dataset. Can be used for building keyword indexes and for classification and retrieval purposes. A | R | Yes | ● Social media<br>● Communication |

---

[28] A set of 5 best practices for writing dataset descriptions have been developed by Smit and Phillips (2021). See:
http://doi.org/10.5281/zenodo.4709835

| Field | Definition with tips | Usage[27] | Repeatable | Example |
|---|---|---|---|---|
| | controlled vocabulary can be employed. The vocab attribute is provided for specification of the controlled vocabulary in use, such as LCSH, MeSH, or others. The vocabURI attribute specifies the location for the full controlled vocabulary. *Tip: The Keyword field is repeatable. Put each keyword in its own field by pressing the "+" button.* *Tip: In Dataverse metadata, keywords are case sensitive. Use sentence case for free text terms (i.e., capitalize the first word and proper nouns).* <br><br> *Tip: Use free-text terms or terms from a controlled vocabulary in this field to represent specific or granular topics associated with the dataset. Think of them like tags.* *Tip: FAST and RVMFAST are good, general-purpose controlled vocabularies adapted from Library of Congress Subject Headings (LCSH).* <br><br> *Tip: Geographic place names should be entered in the Geospatial metadata section.* | | | |
| Vocabulary | For the specification of the keyword controlled vocabulary in use, such as LCSH, MeSH, or others. *Tip: Use this field when the Keyword term is from a controlled vocabulary.* | O | Yes | Government of Canada Core Subject Thesaurus |
| Vocabulary URL | Keyword vocabulary URL points to the web presence that describes the keyword vocabulary, if appropriate. Enter an absolute URL where the keyword vocabulary web site is found, such as http://www.my.org. *Tip: Use this field when the Keyword term is from a controlled vocabulary.* | O | Yes | https://canada.multites.net/ |
| **Topic Classification** | The classification field indicates the broad important topic(s) and subjects that the data cover. Library of Congress subject terms may be used here. | | | |
| Term | Topic or Subject term that is relevant to this Dataset. | O | Yes | Society and Culture |

| Field | Definition with tips | Usage[27] | Repeatable | Example |
|---|---|---|---|---|
| | *Tip: The Keyword field is repeatable. Put each keyword in its own field by pressing the "+" button.* <br><br> *Tip: In Dataverse metadata, keywords are case sensitive. Use sentence case for free text terms (i.e., capitalize the first word and proper nouns).* <br><br> *Tip: Use free-text terms or terms from a controlled vocabulary in this field to represent specific or granular topics associated with the dataset. Think of them like tags.* <br><br> *Tip: FAST and RVMFAST are good, general-purpose controlled vocabularies adapted from Library of Congress Subject Headings (LCSH).* <br><br> *Tip: Geographic place names should be entered in the Geospatial metadata section.* | | | |
| Vocabulary | Provided for specification of the controlled vocabulary in use, e.g., LCSH, MeSH, etc. <br> *Tip: Use this field when the Topic Classification term is from a controlled vocabulary.* | O | Yes | Government of Canada Core Subject Thesaurus |
| Vocabulary URL | Specifies the URL location for the full controlled vocabulary. <br> *Tip: Use this field when the Topic Classification term is from a controlled vocabulary.* | O | Yes | https://canada.multites.net/ |
| **Related Publication** | Publications that use the data from this Dataset. The full list of Related Publications will be displayed on the metadata tab. <br> *Tip: Use this field for publications that are based on the data, e.g., a publication that cites the data.* | | | |
| Citation | The full bibliographic citation for this related publication. | R | Yes | Doe, Jane. (2017). Teen use of social media: analysis of self-reported communication behaviours. Journal of Social Media Use. Vol 1. Iss. 1, 2017. |
| ID Type | The type of digital identifier used for this publication (e.g., Digital Object Identifier (DOI)). <br> *Note: This field is a drop-down list; select one.* | R | Yes | doi |

| Field | Definition with tips | Usage[27] | Repeatable | Example |
|---|---|---|---|---|
| | *Tip: Use a persistent identifier where possible (e.g., DOI, handle, ISBN). If only a URL is available, select the type "url" and enter the URL into both the ID Number and URL fields.* | | | |
| ID Number | The identifier for the selected ID type. *Tip: If using ID Type "url", enter the URL into both this field and the URL field.* | R | Yes | 10.0000/SP/TEST |
| URL | Link to the publication web page (e.g., journal article page, archive record page, or other). *Tip: This field is only used for display purposes in the Dataverse interface. For interoperability, ensure that the ID Type and ID Number fields are also completed.* | R | Yes | https://doi.org/10.0000/SP/TEST |
| **Notes** | Additional important information about the Dataset. | O | Yes | This survey was administered online. Mode of interview has been found to impact results, therefore it is not recommended that these results are compared with other survey results where the interview mode was telephone based. |
| **Language** | Language of the Dataset. *Note: This field is a drop-down list; select all that apply.* | O | Yes | English |
| **Producer** | Person or organization with the financial or administrative responsibility over this Dataset. | | | |
| Name | Producer name. | R | Yes | Youth Communication Development Project |
| Affiliation | The organization with which the producer is affiliated. | R | Yes | Queen's University |
| Abbreviation | The abbreviation by which the producer is commonly known. (ex. IQSS, ICPSR). | O | Yes | YCDP |
| URL | Producer URL points to the producer's web presence, if appropriate. Enter an absolute URL where the producer's web site is found, such as http://www.my.org. | O | Yes | https://youthsocialmedia.org |
| Logo URL | URL for the producer's logo, which points to this producer's web-accessible logo image. Enter an absolute URL where the producer's logo image is found, such as http://www.my.org/images/logo.gif. | O | Yes | https://youthsocialmedia.org/image.png |
| **Production Date** | Date when the data collection or other materials were produced (not distributed, published or archived). | R | No | 2016-01-11 |

| Field | Definition with tips | Usage[27] | Repeatable | Example |
|---|---|---|---|---|
| | *Tip: Date when dataset was finalized and ready for analysis or distribution.* | | | |
| **Production Place** | The location where the data collection and any other related materials were produced. <br> *Tip: Include an institution if relevant. Avoid abbreviations.* | O | No | Kingston, Ontario, Canada |
| **Contributor** | The organization or person responsible for either collecting, managing, or otherwise contributing in some form to the development of the resource. | | | |
| Type | The type of contributor of the resource. <br> *Note: This field is a drop-down list; select one.* | O | Yes | Researcher |
| Name | The Family Name, Given Name or organization name of the contributor. | O | Yes | Doe, Jane |
| **Grant Information** | Grant Information. | | | |
| Grant Agency | Grant Number Agency. <br> *Tip: The name of the funder or grant agency supporting the research.* | O | Yes | Social Sciences and Humanities Research Council (SSHRC) |
| Grant Number | The grant or contract number of the project that sponsored the effort. | O | Yes | CCB123456 |
| **Distributor** | The organization designated by the author or producer to generate copies of the particular work including any necessary editions or revisions. | | | |
| Name | Distributor name. | O | Yes | Data Services |
| Affiliation | The organization with which the distributor contact is affiliated. | O | Yes | Queen's University Library |
| Abbreviation | The abbreviation by which this distributor is commonly known (e.g., IQSS, ICPSR). | O | Yes | QUL |
| URL | Distributor URL points to the distributor's web presence, if appropriate. Enter an absolute URL where the distributor's web site is found, such as http://www.my.org. | O | Yes | https://library.queensu.ca/search/data-statistics |
| Logo URL | URL of the distributor's logo, which points to this distributor's web-accessible logo image. Enter an absolute URL where the distributor's logo image is found, such as http://www.my.org/images/logo.gif. | O | Yes | https://www.queensu.ca/encyclopedia/sites/webpublish.queensu.ca.qencwww/files/images/l/logo/QueensLogo_colour.png |
| **Distribution Date** | Date that the work was made available for distribution/presentation. <br> *Tip: This field may be the same as the Deposit Date. Use Distribution Date only if data was previously distributed.* | O | No | 2018-01-22 |
| **Depositor** | The person (Family Name, Given Name) or the name of the organization that deposited this Dataset to the repository. | R | No | Doe, Jane |

| Field | Definition with tips | Usage[27] | Repeatable | Example |
|---|---|---|---|---|
| | *Tip: The name of the person/institution who provided the dataset(s) to the archive (i.e., not necessarily the person preparing the Dataverse repository submission).* | | | |
| **Deposit Date** | Date that the Dataset was deposited into the repository.<br>*Tip: Deposit Date is pre-populated with the date of upload to the Dataverse repository. It can be edited to reflect the date when the data was received by an external or mediated data repository service.* | R | No | 2018-01-15 |
| **Time Period Covered** | Time period to which the data refer. This item reflects the time period covered by the data, not the dates of coding or making documents machine-readable or the dates the data were collected. Also known as span. | | | |
| Start | Start date which reflects the time period covered by the data, not the dates of coding or making documents machine-readable or the dates the data were collected. | R | Yes | 2015-03-20 |
| End | End date which reflects the time period covered by the data, not the dates of coding or making documents machine-readable or the dates the data were collected. | R | Yes | 2015-06-21 |
| **Date of Collection** | Contains the date(s) when the data were collected. | | | |
| Start | Date when the data collection started. | R | Yes | 2015-03-20 |
| End | Date when the data collection ended. | R | Yes | 2015-06-21 |
| **Kind of Data** | Type of data included in the file: survey data, census/enumeration data, aggregate data, clinical data, event/transaction data, program source code, machine-readable text, administrative records data, experimental data, psychological test, textual data, coded textual, coded documents, time budget diaries, observation data/ratings, process-produced data, or other. | R | Yes | Survey data |
| **Series** | Information about the Dataset series. | | | |
| Name | Name of the dataset series to which the Dataset belongs. | O | No | Social Media Use Among Teens |
| Information | History of the series and summary of those features that apply to the series as a whole. | O | No | Established in 2005, the Youth Communication Development Project aims to gather key research and data about youth development and social |

| Field | Definition with tips | Usage[27] | Repeatable | Example |
|---|---|---|---|---|
| | | | | media use through a series of independent, annual, cross-sectional surveys titled Social Media Use Among Teens. The overall objective of the program is to gather data on youth and social media trends in order to monitor changes in the well-being of young Canadians, and to provide information on specific social policy issues. |
| **Software** | Information about the software used to generate the Dataset. | | | |
| Name | Name of software used to generate the Dataset. *Tip: Useful for specialized software or instruments.* | O | Yes | SPSS |
| Version | Version of the software used to generate the Dataset. | O | Yes | 24 |
| **Related Material** | Any material related to this Dataset. *Tip: This is secondary materials in relation to the study description, such as technical documentation or a website detailing more information about the study.* | O | Yes | Youth Social Media Trends: 2015 Report [Canada]. YCDP, Queen's University, 2016. Access URL: http://dataverse.scholarsportal.info/queensu/2016report.pdf |
| **Related Datasets** | Any Datasets that are related to this Dataset, such as previous research on this subject. *Tip: This would include previous versions of the data (rounds or waves) or other collections generated by the same data.* | O | Yes | Social Media Use Among Teens, 2010 [Canada]. YCDP, Queen's University, 2011. DOI. Access URL: http://dataverse.scholarsportal.info/queensu/2010data.xhtml |
| **Other References** | Any references that would serve as background or supporting material to this Dataset. *Tip: This includes documentation such as codebooks, questionnaires, methodology, coding files, etc.* | O | Yes | Social Media Use Among Teens: Survey Questionnaire, 2015 [Canada]. YCDP, Queen's University, 2016. DOI. Access URL: http://dataverse.scholarsportal.info/queensu/2016questionnaire.pdf |
| **Data Sources** | List of books, articles, serials, or machine-readable data files that served as the sources of the data collection. | O | Yes | Statistics Canada. National Household Survey, 2011: Median Household Income by Census Tracts, Census Metropolitan Areas. NHS 2011, Statistics Canada. Access URL: https://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/prof/index.cfm?Lang=E |

| Field | Definition with tips | Usage[27] | Repeatable | Example |
|---|---|---|---|---|
| **Origin of Sources** | For historical materials, information about the origin of the sources and the rules followed in establishing the sources should be specified. | O | No | National Household Survey, 2011. https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5178 |
| **Characteristic of Sources Noted** | Assessment of characteristics and source material. *Tip: Describes noteworthy aspects of the data collected.* | O | No | |
| **Documentation and Access to Sources** | Level of documentation of the original sources. *Tip: May be used to explain any restrictions or access to source data documentation.* | O | No | Open |

# Geospatial Metadata

## Introduction

Geospatial metadata can describe maps, GIS files, or other location-based data. Any dataset that relates to a location(s) or geographic area(s) (i.e., has a spatial extent) should include geospatial metadata in addition to the general citation metadata block. At a minimum, provide place names to describe locations in your data and use GeoNames.org[29] to confirm these terms. Alternate names (e.g., in other languages) may be added. If applicable, enter bounding box[30] coordinates to allow the data to be findable with map-based search tools.

The examples in this section are fictitious but screen shots from actual datasets in the Scholars Portal Dataverse repository are included after the table. These examples show the edit screen and how the fields look when a dataset is published.

According to Dataverse documentation, Geospatial Metadata fields are compliant with DDI Lite, DDI 2.5 Codebook, DataCite, and Dublin Core. The Country / Nation field uses ISO 3166-1 controlled vocabulary.[31]

---

[29] https://www.geonames.org

[30] A bounding box is an area defined by two longitudes and two latitudes. See: https://wiki.openstreetmap.org/wiki/Bounding_Box

[31] See: http://guides.dataverse.org/en/latest/user/appendix.html

## Geospatial Metadata Block

| Field | Definition with tips | Usage[32] | Repeatable | Example |
|---|---|---|---|---|
| **Geographic Coverage** | Information on the geographic coverage of the data. Includes the total geographic scope of the data. *Tip: For consistency, use the GeoNames.org[33] database to check the form and spelling of place names.* | | | |
| Country / Nation | The country or nation that the Dataset is about. *Note: This field is a drop-down list; select one.* *Tip: Select from drop-down list of names from ISO-3166. If the dataset covers multiple countries, list all of them.* | R | Yes | Canada |
| State / Province | The state or province that the Dataset is about. Use GeoNames for correct spelling and avoid abbreviations. *Tip: If using this field, also include Country to disambiguate.* | R | Yes | British Columbia |
| City | The name of the city that the Dataset is about. Use GeoNames for correct spelling and avoid abbreviations. *Tip: If using this field, also include Country AND, if possible, State/Province to disambiguate.* | R | Yes | Vancouver |
| Other | Other information on the geographic coverage of the data. *Tip: Use for geographical names that are not a country, state/province, or city, e.g., regions, water bodies, astronomy names. If applicable, disambiguate by including City AND/OR State/Province AND/OR Country.* | O | Yes | Jericho Beach Park |
| **Geographic Unit** | Lowest level of geographic aggregation covered by the Dataset, e.g., village, county, region. *Tip: Use when the lowest geographic level that can be analyzed in the dataset is different from the dataset's entire area (e.g., when a dataset about parks in Vancouver can be faceted by the individual parks).* | O | Yes | park |
| **Geographic Bounding Box** | The fundamental geometric description for any Dataset that models geography is the geographic bounding box. It describes the minimum box, defined by west and east longitudes and north and south latitudes, which includes the largest geographic extent of the Dataset's geographic coverage. This element is used in the first pass of a coordinate-based search. Inclusion of this element in the codebook is recommended, but is required if the bound polygon box is included. *Tip: To determine the bounding box, use: http://boundingbox.klokantech.com/ (export results as CSV or TSV format).* | | | |

---

[32] R = Recommended; O = Optional

[33] https://www.geonames.org/

| Field | Definition with tips | Usage[32] | Repeatable | Example |
|---|---|---|---|---|
| West Longitude | Westernmost coordinate delimiting the geographic extent of the Dataset. A valid range of values, expressed in decimal degrees, is -180,0 <= West Bounding Longitude Value <= 180,0. | R | Yes | -123.265 |
| East Longitude | Easternmost coordinate delimiting the geographic extent of the Dataset. A valid range of values, expressed in decimal degrees, is -180,0 <= East Bounding Longitude Value <= 180,0. | R | Yes | -123.115 |
| North Latitude | Northernmost coordinate delimiting the geographic extent of the Dataset. A valid range of values, expressed in decimal degrees, is -90,0 <= North Bounding Latitude Value <= 90,0. | R | Yes | 49.314 |
| South Latitude | Southernmost coordinate delimiting the geographic extent of the Dataset. A valid range of values, expressed in decimal degrees, is -90,0 <= South Bounding Latitude Value <= 90,0. | R | Yes | 49.226 |

## Examples from real datasets

**Geographic Coverage: Other**

1.  In this example, "Other" indicates the Cape Bounty Arctic Watershed Observatory on Melville Island.

    Beamish, Alison; Scott, Neal; Wagner, Ioan; Neil, Allison, 2016, "Impact of active layer detachments on carbon exchange in a high-Arctic ecosystem, Cape Bounty, Nunavut, Canada (2010)", https://hdl.handle.net/10864/11825, Scholars Portal Dataverse, V2

    **Public view**

    | Geospatial Metadata ⌃ | |
    | --- | --- |
    | **Geographic Coverage** ⑦ | Canada, Nunavut, Melville Island, Cape Bounty Arctic Watershed Observatory |

    **Edit view**

    Geospatial Metadata ⌃

    | **Geographic Coverage** ⑦ | Country / Nation ⑦ | State / Province ⑦ | |
    | --- | --- | --- | --- |
    | | Canada ▾ | Nunavut | **+** |
    | | City ⑦ | Other ⑦ | |
    | | | Melville Island, Cape Bounty Arctic Wate | |

2.  In this example, "Other" indicates the specific district of Mapo-gu within the city of Seoul.

    Da In Choi, 2015, "Korean War Interviews, 2015", https://hdl.handle.net/10864/11174, Scholars Portal Dataverse, V5

    **Public view**

    | Geospatial Metadata ⌃ | |
    | --- | --- |
    | **Geographic Coverage** ⑦ | Korea, Republic of, Seoul, Mapo-gu |

**Edit view**



3. In this example, "Other" indicates a specific health clinic where the study occurred.

   Wilson, Rosemary A.; VanDenKerkhof, Elizabeth G.; Duggan, Scott; Gilron, Ian; Good, Mary Anne; Henry, Richard; Carley, Meg, 2018, "Chronic Pain Surveillance at Queen's, 2013-2017", https://doi.org/10.5683/SP2/GAPNRM, Scholars Portal Dataverse, V1

**Public view**



**Edit view**

**Geographic Unit**

In this example, the Forward Sortation Area (FSA) is the lowest level of geographic aggregation covered by the Dataset.

Hird, Myra J.; Lougheed, Scott C.; Kuyvenhoven, Cassandra; Rowe, R. Kerry, 2016, "Perspectives on Municipal Waste Management in Kingston, Ontario, 2012", https://hdl.handle.net/10864/11926, Scholars Portal Dataverse, V1

Geospatial Metadata ⌃

Geographic Coverage ?          Canada, Ontario, Kingston

Geographic Unit ?              Forward Sortation Area (FSA)

**Geographic Bounding Box**

1.  In this example, the two longitude and two latitude coordinates define the geographic extent of the dataset.

    Anderson, Lauren; Beasley, Barb; Flumerfelt, Sidney-Rae; Fox, Caroline; Friesen, Sarah; Macfarlane, Gemma; McKay, Taesagh, 2019, "Replication data for: Long-term monitoring in Barkley Sound: a temporal analysis of intertidal biodiversity on Wizard Islet, British Columbia from 1997 to 2017", https://doi.org/10.5683/SP2/C8G480, Scholars Portal Dataverse, V1, UNF:6:mBVVVVwtuVbcT4h2Au8RXQ== [fileUNF]

    **Public view**

Geospatial Metadata ⌃

Geographic Coverage ?          Canada, British Columbia, Wizard Islet

Geographic Bounding Box ?      -125.160843 -125.158203 48.8592 48.857391

**Edit view**

# Social Science and Humanities Metadata

## Introduction

The Social Science and Humanities block describes data that falls within the domain of social science and the humanities. Multiple real examples of datasets are provided for each field with links to datasets so that examples can be seen in context. According to Dataverse documentation, Social Science & Humanities Metadata fields are compliant with DDI Lite, DDI 2.5 Codebook, and Dublin Core (see .tsv version[34]).[35]

## Controlled vocabularies

The DDI Alliance has created a set of controlled vocabularies[36] that can be used with some fields within the Social Science and Humanities section. The DDI Alliance is continually adding to this set.

---

[34] https://github.com/IQSS/dataverse/blob/master/scripts/api/data/metadatablocks/citation.tsv

[35] See: http://guides.dataverse.org/en/latest/user/appendix.html

[36] http://www.ddialliance.org/controlled-vocabularies

# Social Science and Humanities Metadata Block

| Field | Definition with tips | Usage[37] | Repeatable | Example |
|---|---|---|---|---|
| Unit of Analysis | Basic unit of analysis or observation that this Dataset describes, such as individuals, families/households, groups, institutions/organizations, administrative units, and more. For information about the DDI's controlled vocabulary for this element, please refer to the DDI web page at http://www.ddialliance.org/controlled-vocabularies. | R | Yes | ● Individual<br>● Family<br>● Household |
| Universe | Description of the population covered by the data in the file; the group of people or other elements that are the object of the study and to which the study results refer. Age, nationality, and residence commonly help to delineate a given universe, but any number of other factors may be used, such as age limits, sex, marital status, race, ethnic group, nationality, income, veteran status, criminal convictions, and more. The universe may consist of elements other than persons, such as housing units, court cases, deaths, countries, and so on. In general, it should be possible to tell from the description of the universe whether a given individual or element is a member of the population under study. Also known as the universe of interest, population of interest, and target population. | R | Yes | ● Canadians aged 12-30 (Source: https://doi.org/10.5683/SP/HY2H1A)<br>● Queen's University 2nd year medical students who were part of the 2016 Critical Enquiry Course in the School of Medicine and agreed to participate in the study. (Source: https://doi.org/10.5683/SP/D6NISS) |
| Time Method | The time method or time dimension of the data collection, such as panel, cross-sectional, trend, time- series, or other. | O | No | ● Longitudinal<br>● Time series<br>● Longitudinal: Panel |
| Data Collector | Individual, agency or organization responsible for administering the questionnaire or interview or compiling the data. | R | No | ● Trained student interviewers, both anglophone and francophone. (Source: https://hdl.handle.net/10864/ZJ17A) |
| Collector Training | Type of training provided to the data collector. | O | No | ● The interviews were conducted by professional interviewers under the supervision of the Institute for Social Science Research. |

---

[37] R = Recommended; O = Optional

| Field | Definition with tips | Usage[37] | Repeatable | Example |
|---|---|---|---|---|
| | | | | (Source: https://doi.org/10.7910/DVN/SRVIO4)  <br><br>• From the documentation: "Each staff member was thoroughly trained prior to beginning work on the survey. Interviewers received about three days of classroom training plus self-training materials. Additional study materials and classroom training were planned throughout the interviewing period. Quality control measures, such as editing returns, observing interviews and re-interviewing selected households were employed throughout the survey." (Source: https://doi.org/10.7910/DVN/YT09KD) |
| **Frequency** | If the data collected includes more than one point in time, indicate the frequency with which the data was collected; that is, monthly, quarterly, or other. | O | No | • Annual <br>• Hourly <br>• Data was collected at baseline and at one month follow-up. (Source: https://doi.org/10.7939/DVN/10889) |
| **Sampling Procedure** | Type of sample and sample design used to select the survey respondents to represent the population. May include reference to the target sample size and the sampling fraction. | R | No | • Canadian adults randomly selected from Angus Reid Forum panel members. (Source: https://hdl.handle.net/10864/11510) <br>• Probability (Source: https://ddialliance.org/Specification/DDI-CV/SamplingProcedure_1.1.html) |
| **Target Sample Size** | Specific information regarding the target sample size, actual sample size, and the formula used to determine this. | | | |
| Actual | Actual sample size. <br>*Tip: The research study's actual sample size may be stated in this numeric field for reference purposes.* | O | No | • 1015 |
| Formula | Formula used to determine target sample size. <br>*Tip: A plain-text, general description of a sample size formula may be stated here for* | O | No | • Eligible employees who lived within the following FSAs: K6V, K7A, K7C, K7G, K7H, K7K, K7L, K7M, K7N, K7P, |

| Field | Definition with tips | Usage[37] | Repeatable | Example |
|---|---|---|---|---|
| | *reference purposes. This may include particular methodologies, practices, and outcomes from existing scholarly literature.* | | | K7R, K8N, K8P, K8R, K8V, K0E, K0G, K0H, K0K. |
| Major Deviations for Sample Design | Show correspondence as well as discrepancies between the sampled units (obtained) and available statistics for the population (age, sex-ratio, marital status, etc.) as a whole. | O | No | • The suitability of Ohio as a research site reflected its similarity to the United States as a whole. The evidence extended by Tuchfarber (1988) shows that Ohio is representative of the United States in several ways: percent urban and rural, percent of the population that is African American, median age, per capita income, percent living below the poverty level, and unemployment rate. Although results generated from an Ohio sample are not empirically generalizable to the United States, they may be suggestive of what might be expected nationally. (Source: https://ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation_files/schemas/codebook_xsd/elements/deviat.html) <br> • Oversample of persons 50 and older Source: https://doi.org/10.7910/DVN/FGTJGO |
| Collection Mode | Method used to collect the data; instrumentation characteristics (e.g., telephone interview, mail questionnaire, or other). | R | No | • Interview <br> • Paper and online questionnaire <br> • Coded from psychiatric hospital files, court records, and police agencies. (Source: https://hdl.handle.net/10864/12053) |
| Type of Research Instrument | Type of data collection instrument used. Structured indicates an instrument in which all respondents are asked the same questions/tests, possibly with precoded answers. If a small portion of such a questionnaire includes open-ended questions, provide appropriate comments. Semi-structured indicates that the research instrument contains mainly open-ended | R | No | • Questionnaire <br> • Structured <br> • Technical instrument: Static Chamber, Vaisala Humicap HM70 relative humidity/ temperature probe, Vaisala Carbocap GMP343 infrared analyzer, Hobo Pro v2 U23-003 temperature logger, Kestrel 3500 weather meter, Taylor 9878 thermometer. |

| Field | Definition with tips | Usage[37] | Repeatable | Example |
|---|---|---|---|---|
| | questions. Unstructured indicates that in-depth interviews were conducted. | | | (Source: https://hdl.handle.net/10864/12053) |
| **Characteristics of Data Collection Situation** | Description of noteworthy aspects of the data collection situation. Includes information on factors such as cooperativeness of respondents, duration of interviews, number of call backs, or similar. | O | No | ● There were 1,419 respondents who answered questions in telephone interviews lasting approximately 35 minutes each. Clarifications to survey questions were limited and respondents were directed to provide a response based on the information provided as to not allow interviewer bias/assumptions to influence the survey results. (Source: https://hdl.handle.net/10864/ZJ17A) |
| **Actions to Minimize Losses** | Summary of actions taken to minimize data loss. Include information on actions such as follow-up visits, supervisory checks, historical matching, estimation, and so on. | O | No | ● Reminder e-mails were distributed to target population. (Source: https://doi.org/10.5683/SP/L1H3SS) <br> ● Cards reminding parents about the follow-up visit were given out. (Source: https://doi.org/10.7939/DVN/10889) |
| **Control Operations** | Methods to facilitate data control performed by the primary investigator or by the data archive. | O | No | ● Field validation is built into REDCap data collection forms. (Source: https://doi.org/10.7939/DVN/10907) <br> ● Blinded double data entry and third person cross-validation were used. (Source: https://doi.org/10.7939/DVN/10900) |
| **Weighting** | The use of sampling procedures might make it necessary to apply weights to produce accurate statistical results. Describes the criteria for using weights in analysis of a collection. If a weighting formula or coefficient was developed, the formula is provided, its elements are defined, and it is indicated how the formula was applied to the data. | R | No | ● Rim weighting is used with this file. By region, the file was weighted to census targets on sex (wtsex), age (Wtage), and education (Wtedu) using the 2011 census. For this file, a religion weight (wtreligion) was also included based on the 2011 National Household Survey (NHS). The wtg2 variable includes all of these weights within it. (Source: https://doi.org/10.5683/SP/78RONJ) <br> ● The final sample obtained for each area is not proportional to the Alberta |

| Field | Definition with tips | Usage[37] | Repeatable | Example |
|---|---|---|---|---|
| | | | | population it makes up. For instance, Edmonton is over-sampled as shown by TABLE 1. Edmonton makes up only 24% of the Alberta population but has 43% of the interviews. Therefore, in order to combine the samples for a provincial sample weighting is necessary. The weighting factors used for the 1987 survey are as follows: Edmonton 0.558439, Calgary 1.151521, and Other Alberta 1.471173. (Source: https://doi.org/10.7939/DVN/10567) <br> ● wtx used to correctly weight respondents against Stats Canada Alberta population estimates (Source: https://doi.org/10.7939/DVN/10813) |
| **Cleaning Operations** | Methods used to clean the data collection, such as consistency checking, wildcode checking, or other. | O | No | ● For income data, all respondents are matched to the tax data file unless they refuse to have their information linked. Data obtained from the tax file are complete and do not require imputation. Income figures are imputed only in the absence of tax data. Donor imputation by the nearest neighbour method is generally used and is performed primarily with Statistics Canada's Census Edit and Imputation System (CANCEIS). However, amounts received through certain government programs such as the universal childcare benefit and child tax benefits are derived from other information (e.g., number of children in the household) using a deductive imputation method. (Source: http://hdl.handle.net/11272/10619) <br> ● Physiological data was reviewed for outliers. Individual breaths with tidal volume (VT), respiratory rate (RR) or |

| Field | Definition with tips | Usage[37] | Repeatable | Example |
|---|---|---|---|---|
| | | | | minute ventilation (VE) that lay outside the 95% confidence interval for all infants were removed as outliers; 99.7% of all measured breaths were included in the final analyses. (Source: https://doi.org/10.7939/DVN/10910) |
| **Study Level Error Notes** | Note element used for any information annotating or clarifying the methodology and processing of the study. | O | No | ● The computerized questionnaire contains many features designed to maximize the quality of the data collected. Many edits are built into the questionnaire to compare the reported data with unusual values and detect logical inconsistencies. When an edit fails, the interviewer is prompted to correct the information (with the respondent's help, if necessary). Once the data are transmitted to Head Office, a comprehensive series of processing steps are undertaken for the purpose of detailed verification of each questionnaire. Invalid responses are corrected or flagged for imputation. Edits were applied at a micro level. Deterministic edits and consistency edits were also performed at the micro level. Data was checked for outliers and extreme values and were corrected at a micro level when required. (Source: http://hdl.handle.net/11272/10619 ) |
| **Response Rate** | Percentage of sample members who provided information. | R | No | ● Based on 100km radius, the survey response rate is 1874/3994 (46.9%), and the survey completion rate is 1732/3994 (43.4%). Based on FSAs for locations served by Kingston Transit, the survey response rate is 1469/3151 (46.6%), and the survey completion rate is 1356/3151 (43.0%). |

| Field | Definition with tips | Usage[37] | Repeatable | Example |
|---|---|---|---|---|
| | | | | (Source: https://doi.org/10.5683/SP/CNXSVN) <br> ● At one-month follow-up: 60.2% (n=136/226). (Source: https://doi.org/10.7939/DVN/10889) |
| **Estimates of Sampling Error** | Measure of how precisely one can estimate a population value from a given sample. *Tip: Examples include confidence intervals, non-response, response bias.* | R | No | ● In SFS 2016, the 95% confidence interval for the average net worth of Canadian families had a width of $38,500. (Source: http://hdl.handle.net/11272/10619) <br> ● + or - 2.5%; design effect of weighting not calculated (Source: https://doi.org/10.7910/DVN/FGTJGO) |
| **Other Forms of Data Appraisal** | Other issues pertaining to the data appraisal. Describe issues such as response variance, nonresponse rate and testing for bias, interviewer and response bias, confidence levels, question bias, or similar. | O | No | ● OSBD is subject to interpretation since it is an indirect behavioral measure of perceived distress. (Source: https://doi.org/10.7939/DVN/10841) |
| **Notes** | General notes about this Dataset. *Tip: Includes notes related to fields in the Social Science and Humanities section.* | | | |
| Type | Type of note. | O | No | ● Processing note |
| Subject | Note subject. | O | No | ● Variable corrections |
| Text | Text for this note. | O | No | ● Info (Misc) v2 note: Corrections were made to variables: PAS1MRG1, PAS1MRG2, PASRDPO1, PASRDPO2, PASRDPO3, PASRDPO4, PASRDPO5 and VERDATE. (Source: http://hdl.handle.net/11272/10619) |

# Astronomy and Astrophysics Metadata

## Introduction

The fields in this block allow for description of astronomical or astrophysical data. Within the research communities of astronomy and astrophysics, there exist domain-specific repositories, sometimes associated with major observatories, for depositing and sharing data, including the Hubble Legacy Archive,[38] CFHT Science Archive[39], and VLA Data Archive.[40] Motivations for depositing astronomical or astrophysical data in a multi-disciplinary repository like a Dataverse repository may include achieving broader dissemination of the data or greater visibility for it.

## Source of the metadata fields

Most of the metadata fields in this block are based on the Virtual Observatory (VO) Discovery and Provenance Metadata[41] recommendation, an implementation of the International Virtual Observatory Alliance's Resource Metadata for the Virtual Observatory, Version 1.12 (2007)[42] specification, which is referred to as IVOA-RM in this document. Fields that are not part of IVOA-RM are noted in the table below.

While Dataverse documentation indicates that the Astronomy and Astrophysics metadata fields "can be mapped or exported to the International Virtual Observatory Alliance's (IVOA) VOResource Schema format"[43] (i.e., IVOA-RM), it is not possible to export directly from a Dataverse repository in this format. To generate IVOA-RM records, it would be necessary to create a mapping from one of the two export formats that include the fields (JSON and OAI-ORE) to IVOA-RM and transform the metadata.

Examples in this section are taken from the IVOA-RM specification and are supplemented with real examples from the Harvard Dataverse repository.

Following the main table are:
- a table of terms and definitions for the Type controlled vocabulary,
- additional examples taken from datasets found in the Harvard Dataverse repository, and
- additional fields from the IVOA-RM sample record for context.

---

[38] https://hla.stsci.edu

[39] https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/cfht

[40] https://science.nrao.edu/facilities/vla/archive/index

[41] http://perma.cc/H5ZJ-4KKY

[42] http://www.ivoa.net/documents/latest/RM.html

[43] See: http://guides.dataverse.org/en/latest/user/appendix.html

## Astronomy and Astrophysics Metadata Block

| Field | Definition with tips | Usage[44] | Repeatable | Examples[45] |
|---|---|---|---|---|
| **Type** | The nature or genre of the content of the files in the dataset.<br>*Note: This field is a drop-down list; select all that apply. Most of the types originate in the Virtual Observatory (VO) Discovery and Provenance Metadata[46] specification. For type definitions, see the table below.* | R | No | Survey, Catalog<br><br>*See also examples 1-7 below.* |
| **Facility** | The observatory or facility where the data was obtained.<br>*Tip: For theoretical data, identify the computational facility where it was obtained. Repeat the element when naming more than one facility.* | R | Yes | Apache Point Observatory, Sloan 2.5-m Telescope<br><br>*See also examples 1-7 below.* |
| **Instrument** | The instrument used to collect the data.<br>*Tip: Identify a specific instrument or an instrument type for observational data. Provide the name of the computer code for theoretical data.* | R | Yes | Five-band clocked CCD camera<br><br>*See also examples 2-7 below.* |
| **Object** | Astronomical Objects represented in the data (Given as SIMBAD recognizable names preferred).<br>*Note: The SIMBAD database[47] provides a concordance for different name schemes. For the non-specialist, it is somewhat difficult to use.* | R | Yes | HD 1234 (or GEN# +1.00001234)<br><br>*See also examples 1-4 and 6-7 below.* |
| **Spatial Resolution** | The spatial (angular) resolution that is typical of the observations, in decimal degrees. | R | No | 0.00028<br><br>*See also examples 2-7 below.* |
| **Spectral Resolution** | The spectral resolution that is typical of the observations, given as the ratio $\lambda/\Delta\lambda$.<br>*Tip: The spectral resolution may also be given as the difference between velocities $\Delta v$, calculated by using the Doppler effect, that can be distinguished in a line of the spectrum.* | R | No | 5000<br>*Given here as $\lambda/\Delta\lambda$*<br><br>0.15 km/s<br>*Given here as $\Delta v$*<br><br>*See also examples 2, 3 and 5 below.* |
| **Time Resolution** | The temporal resolution that is typical of the observations, given in seconds. | R | No | 120 |

---

[44] R = Recommended; O = Optional

[45] The examples are mainly drawn from the sample record included in the IVOA-RM specification (see Section 6: Example). Other fields from this record (including the title and description) are included in the 'IVOA-RM Example' table on page 40.

[46] https://perma.cc/H5ZJ-4KKY

[47] http://simbad.u-strasbg.fr/simbad

| Field | Definition with tips | Usage[44] | Repeatable | Examples[45] |
|---|---|---|---|---|
| **Bandpass** | Conventional bandpass name.<br>*Tip: Specific bandpass names include: optical bandpasses (U, V, B, R, I); and narrow line filters (H-alpha, [OIII]).* | R | Yes | u', g', r', i', z'<br><br>U, B, V, JHK, 450<br><br>*See also example 4 below.* |
| **Central Wavelength (m)** | The central wavelength of the spectral bandpass, in meters.<br>*Note: This is not an IVOA-RM element.* | R | Yes | *See example 5 below.* |
| **Wavelength Range** | The minimum and maximum wavelength of the spectral bandpass. | | | |
| Minimum (m) | The minimum wavelength of the spectral bandpass, in meters. | O | Yes | 400.e-9 |
| Maximum (m) | The maximum wavelength of the spectral bandpass, in meters. | O | Yes | 850.e-9 |
| **Dataset Date Range** | Time period covered by the data.<br>*Tip: When using this field, best practice is to supply a date range that includes date and time, and, in addition, to supply the date range again in the 'Time Period Covered' field in the Citation block.* | | | |
| Start | Dataset Start Date.<br>*Tip: Provide a date or date/time (preferred) in ISO 8601 format*<br>*(e.g., YYYY-MM-DDTHH:MM:SSZ).* | R | Yes | 1999-12-25<br><br>2020-01-08T15:16:23Z<br><br>*See also examples 4, 6 and 7 below.* |
| End | Dataset End Date.<br>*Tip: Provide a date or date/time (preferred) in ISO 8601 format*<br>*(e.g., YYYY-MM-DDTHH:MM:SSZ).* | R | Yes | 2001-07-15<br><br>2021-03-09T19:26:00+05:00<br><br>*See also examples 4, 6 and 7 below.* |
| **Sky Coverage** | The sky coverage of the data object.<br>*Note: Follow the syntax found in the [Space-Time Coordinate (STC) Metadata Recommendation](http://www.ivoa.net/documents/latest/STC.html).[48]* | R | Yes | PositionInterval FK5 145.17 – 1.25 235.9 1.25 PositionInterval FK5 250.71 52.15 267.0 66.29 PositionInterval FK5 350.43 – 1.25 359.99 1.17 PositionInterval 0.0 –1.25 56.37 1.17<br><br>*See also examples 1-5 and 6-* |

---

[48] http://www.ivoa.net/documents/latest/STC.html

| Field | Definition with tips | Usage[44] | Repeatable | Examples[45] |
|---|---|---|---|---|
| | | | | *7 below.* |
| **Depth Coverage** | The (typical) depth coverage, or sensitivity, of the data object in Jy. *Tip: This is the faintest flux of energy that can be perceived (units: Jensky (Jy)).* | R | No | 3.e-6 |
| **Object Density** | The (typical) density of objects, catalog entries, telescope pointings, etc., on the sky, in number per square degree. *Tip: Objects could be stars, galaxies, planets, star clusters, asteroids, etc.* | O | No | 6.e4 |
| **Object Count** | The total number of objects, catalog entries, etc., in the data object. *Tip: Objects could be stars, galaxies, planets, star clusters, asteroids, etc.* | O | No | 2.e7 |
| **Fraction of Sky** | The fraction of the sky represented in the observations, ranging from 0 to 1. | O | No | 0.01 |
| **Polarization** | The polarization coverage. *Note: This is not an IVOA-RM element.* | R | No | |
| **RedshiftType** | RedshiftType string C "Redshift"; or "Optical" or "Radio" definitions of Doppler velocity used in the data object. *Note: This is not an IVOA-RM element.* *Tip: The light coming from an object moving away from you is redshifted (Doppler effect), which means that the frequency of that light is shifted to a lower value. The speed of an object can be computed by measuring the shifts in frequencies of peaks in its spectra. Since the Universe is expanding, the more distant an object is, the more it is redshifted. For that reason, redshift is used as a unit of distance for the most distant objects. This field RedshiftType serves to indicate which part of the spectra was used to measure the redshift.* | O | No | |
| **Redshift Resolution** | The resolution in redshift (unitless) or Doppler velocity (km/s) in the data object. *Note: This is not an IVOA-RM element.* | O | No | |
| **Redshift Value** | The value of the redshift (unitless) or Doppler velocity (km/s) in the data object. *Consists of 2 subfields.* | | | |
| Minimum | The minimum value of the redshift (unitless) or | O | Yes | |

| Field | Definition with tips | Usage[44] | Repeatable | Examples[45] |
|---|---|---|---|---|
| | Doppler velocity (km/s) in the data object.<br>*Note: This is not an IVOA-RM element.* | | | |
| Maximum | The maximum value of the redshift (unitless) or Doppler velocity (km/s) in the data object.<br>*Note: This is not an IVOA-RM element.* | O | Yes | |

## Type definitions

| Type | Definition (from VO 2012, where applicable) |
|---|---|
| Image | One or more 2-D images |
| Mosaic | Mosaic of multiple 2-D images |
| EventList | One or more event lists |
| Spectrum | One or more 1-D spectra |
| Cube | One or more 3-D data cubes |
| Table | Table of values; at least two columns |
| Catalog | Collection of derived data, primarily in tabular form |
| LightCurve | One or more 1-D light curves |
| Simulation | Theoretical simulation or model |
| Figure | |
| Artwork | Artists' renderings of astronomical phenomena or objects |
| Animation | Animation clips of astronomical phenomena |
| PrettyPicture | |
| Documentation | |
| Other | A data object not described by any of the above types. |
| Library | Collection of published materials (journals, books, etc.) |
| Press Release | |
| Facsimile | Digitized facsimile of (historical) document |
| Historical | Historical information about astronomical objects. |
| Observation | Collection of data objects (files) associated with one or more observations |
| Object | Collection of data objects (files) associated with one or more celestial objects |
| Value | Single value |
| ValuePair | Keyword-value pair |
| Survey | Collection of observations covering substantial and contiguous areas of the sky |

# Supplemental examples from real datasets

All examples are taken from the Harvard Dataverse repository.

Example 1: https://doi.org/10.7910/DVN/10.1088

| Astronomy and Astrophysics Metadata ⌃ | |
|---|---|
| Type | Cube |
| Facility | SMA; |
| Object | ophasm1; |
| Dataset Date Range | 2007-07-29T07:33:38.007 2007-07-29T07:33:38.007 |
| Sky Coverage | (246.614992908 -24.3986103898); |

Example 2: https://doi.org/10.7910/DVN/SDHQRP

| Astronomy and Astrophysics Metadata ⌃ | |
|---|---|
| Type | Image |
| Facility | Green Bank Telescope |
| Instrument | K-band Focal Plane Array |
| Object | Serpens South |
| Spatial Resolution | 32" |
| Spectral Resolution | 0.15 km/s |
| Sky Coverage | (277.541666667 -2.0) |

Example 3: https://doi.org/10.7910/DVN/B7TAFU

| Astronomy and Astrophysics Metadata ⌃ | |
|---|---|
| Type | Cube |
| Facility | GBT |
| Instrument | W-band receiver |
| Object | W49; W51 |
| Spatial Resolution | 10" |
| Spectral Resolution | 0.25 km/s |
| Sky Coverage | (287.5588 9.103852777777778); (287.5069149505811 9.160234421054046); (290.9310703578097 14.50682269941492); (290.8599951721404 14.55345161586474); (290.9247291666666 14.51319166666667) |

Example 4: https://doi.org/10.7910/DVN/K4GWMI

| Astronomy and Astrophysics Metadata ^ | |
|---|---|
| **Type** ❓ | Image |
| **Facility** ❓ | JCMT |
| **Instrument** ❓ | SCUBA |
| **Object** ❓ | SCUBA Galactic Centre Survey (450um) |
| **Spatial Resolution** ❓ | -4.3301961938407E-4 |
| **Bandpass** ❓ | 450 |
| **Dataset Date Range** ❓ | 1998-04-07 1998-04-07 |
| **Sky Coverage** ❓ | (266.41720833332 -29.007938888886972) |

Example 5: https://doi.org/10.7910/DVN/X4PQ1Z

| Astronomy and Astrophysics Metadata ^ | |
|---|---|
| **Type** ❓ | Cube |
| **Facility** ❓ | Arecibo Observatory |
| **Instrument** ❓ | ALFA |
| **Spatial Resolution** ❓ | 4' |
| **Spectral Resolution** ❓ | 0.184 km/s |
| **Central Wavelength (m)** ❓ | 1420.405 |

Example 6: https://doi.org/10.7910/DVN/28977

Astronomy and Astrophysics Metadata ⌃

| | |
|---|---|
| **Type** ❓ | Image |
| **Facility** ❓ | ESO-NTT; ESO-VLT-U1 |
| **Instrument** ❓ | SOFI; ISAAC |
| **Object** ❓ | Flat Fielded; L2; L6; L1; L4; L5; L3 |
| **Spatial Resolution** ❓ | 2.97409421611617E-8 - 1.96286967956165E-5 |
| **Dataset Date Range** ❓ | 2000-03-14T00:14:58 2000-03-14T00:14:59.200<br>2002-10-10T06:49:07 2002-10-10T06:49:07.110<br>2002-12-19T06:07:31 2002-12-19T06:07:31.110<br>2002-10-10T05:48:56 2002-10-10T05:48:56.110<br>2002-10-12T06:47:06 2002-10-12T06:47:06.110<br>2002-10-14T08:34:35 2002-10-14T08:34:35.110<br>2002-10-12T05:48:41.173 2002-10-12T05:48:41.283 |
| **Sky Coverage** ❓ | (85.4366666666667 -1.90699166666667); (83.8304690978601<br>-5.39957831315445); (83.8040205175233 -5.40392302994954);<br>(83.8290803234773 -5.37912292534363); (83.7832586106861<br>-5.36222589662728); (83.8056282744872 -5.38764313485259);<br>(83.8479212323983 -5.39678424213229) |

Example 7: https://doi.org/10.7910/DVN/2HV6OF

Astronomy and Astrophysics Metadata ⌃

| | |
|---|---|
| **Type** ❓ | Image |
| **Facility** ❓ | Gemini-South |
| **Instrument** ❓ | GMOS-S |
| **Object** ❓ | SPT-CLJ2118-5055 |
| **Spatial Resolution** ❓ | -4.051449E-5 - -4.051262E-5 |
| **Dataset Date Range** ❓ | 2011-04-11T09:23:26.000 2011-04-11T09:24:36.498<br>2011-04-11T09:27:38.009 2011-04-11T09:31:48.504 |
| **Sky Coverage** ❓ | (319.727907 -50.931761); (319.727949 -50.931731) |

# IVOA-RM Example

The examples in the main table are mainly drawn from the sample record included in the [IVOA-RM specification](#) (see section 6). Additional fields from the sample record are given below for context.

| | IVOA-RM example value |
|---|---|
| Title | Sloan Digital Sky Survey |
| Author | Sloan Digital Sky Survey Consortium |
| Keyword/Term | galaxies, quasars, stars, CCD photometry, spectroscopy, redshift, sky surveys |
| Description/Text | The Sloan Digital Sky Survey is using a dedicated 2.5 m telescope and a large format CCD camera to obtain images of over 10,000 square degrees of high Galactic latitude sky in five broad bands (u', g', r', i' and z', centered at 3540, 4770, 6230, 7630, and 9130 Å, respectively). Medium resolution spectra will be obtained for approximately 106 galaxies and 100,000 quasars. The early data release (EDR), on June 2001, includes searchable catalogs of images and spectra, images for display and scientific purpose in both 2-D FITS and JPEG formats, and spectra in both 1-D FITS and GIF formats. The EDR covers about 460 square degrees of sky. The next data releases will occur every 18 months or so. |

# Life Sciences Metadata

## Introduction

Life sciences metadata describes scientific research in multiple domains that focus on the study of living organisms including biology, zoology, microbiology, physiology, biochemistry, and related subjects.

The life sciences metadata block includes six fields with drop-down lists. Each of these fields are supplemented by a free text field that may be used when "Other" is selected from the primary field's drop-down list.

The examples in this section are fictitious but screen shots from actual datasets in the Harvard Dataverse repository are included after the table.

According to Dataverse documentation, Life Sciences Metadata fields are based on ISA-Tab Specification,[49] along with controlled vocabulary from subsets of the OBI Ontology[50] and the NCBI Taxonomy for Organisms.[51] [52] For more information, including the full list of vocabulary terms, see the documentation for the life sciences metadata block.[53]

---

[49] https://isa-specs.readthedocs.io/en/latest/isamodel.html

[50] http://bioportal.bioontology.org/ontologies/OBI

[51] http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html

[52] See: http://guides.dataverse.org/en/latest/user/appendix.html

[53] https://docs.google.com/spreadsheets/d/13HP-jI_cwLDHBetn9UKTREPJ_F4iHdAvhjmlvmYdSSw/edit#gid=2

# Life Sciences Metadata Block

| Field | Definition | Usage[54] | Repeatable | Example |
|---|---|---|---|---|
| **Design Type** | Design types that are based on the overall experimental design.<br>*Note: This field is a drop-down list; select all that apply. If there is no appropriate option, select Other and use the Other Design Type field.* | R | No | ● Case control<br>● Randomized controlled trial<br>● Technological design |
| **Other Design Type** | If Other was selected in Design Type, list any other design types that were used in this Dataset. | O | Yes | |
| **Factor Type** | Factors used in the Dataset.<br>*Note: This field is a drop-down list; select all that apply. If there is no appropriate option, select Other and use the Other Factor Type field.* | O | No | ● Age<br>● Disease state<br>● Strain |
| **Other Factor Type** | If Other was selected in Factor Type, list any other factor types that were used in this Dataset. | O | Yes | |
| **Organism** | The taxonomic name of the organism used in the Dataset or from which the starting biological material derives.<br>*Note: This field is a drop-down list; select all that apply. If there is no appropriate option, select Other and use the Other Organism field.* | R | No | ● Homo sapiens<br>● Mus musculus<br>● Oryza sativa |
| **Other Organism** | If Other was selected in Organism, list any other organisms that were used in this Dataset. Terms from the NCBI Taxonomy[55] are recommended. | O | Yes | |
| **Measurement Type** | A term to qualify the endpoint, or what is being measured (e.g., gene expression profiling; protein identification).<br>*Note: This field is a drop-down list; select all that apply. If there is no appropriate option, select Other and use the Other Measurement Type field.* | R | No | ● Cell counting<br>● Hematology<br>● Targeted sequencing |
| **Other Measurement Type** | If Other was selected in Measurement Type, list any other measurement types that were used. Terms from NCBO Bioportal[56] are recommended.<br>*Tip: NCBO Bioportal is a searchable registry of hundreds of ontologies. Users can select an ontology and download the most recent version to see terms. New users may wish to either seek further guidance or* | O | Yes | |

[54] R = Recommended; O = Optional

[55] https://www.ncbi.nlm.nih.gov/taxonomy

[56] http://bioportal.bioontology.org/ontologies

| Field | Definition | Usage[54] | Repeatable | Example |
|---|---|---|---|---|
| | *disregard this recommended registry.* | | | |
| **Technology Type** | A term to identify the technology used to perform the measurement (e.g., DNA microarray; mass spectrometry). *Note: This field is a drop-down list; select all that apply. If there is no appropriate option, select Other and use the Other Technology Type field.* | R | No | ● DNA microarray<br>● Flow cytometry<br>● Gel electrophoresis |
| **Other Technology Type** | If Other was selected in Technology Type, list any other technology types that were used in this Dataset. | O | Yes | |
| **Technology Platform** | The manufacturer and name of the technology platform used in the assay (e.g., Bruker AVANCE). *Note: This field is a drop-down list; select all that apply. If there is no appropriate option, select Other and use the Other Technology Platform field.* | R | No | ● Affymetrix<br>● AQI Biosciences<br>● BD BACTEC MGIT 320 |
| **Other Technology Platform** | If Other was selected in Technology Platform, list any other technology platforms that were used in this Dataset. | O | Yes | |
| **Cell Type** | The name of the cell line from which the source or sample derives. *Tip: Terms from Cell Ontology[57] are recommended.* | O | Yes | ● Collar cell<br>● Scleral cell<br>● Unimodal nocireceptor |

---

[57] https://www.ebi.ac.uk/ols/ontologies/cl

# Examples from real datasets

Farhat, Maha, 2015, "genetic variation", https://doi.org/10.7910/DVN/AQ5LH5, Harvard Dataverse, V2, UNF:6:SDP+kyIqLcAbXLu6wA+xnw== [fileUNF]

Life Sciences Metadata ⌃

| | |
|---|---|
| Design Type ⍰ | Case Control |
| Factor Type ⍰ | Drug Susceptibility |
| Organism ⍰ | Mycobacterium tuberculosis |
| Measurement Type ⍰ | genome sequencing |
| Technology Type ⍰ | nucleotide sequencing |
| Technology Platform ⍰ | Illumina |

Collins, Adam, 2018, "Images from screening plates", https://doi.org/10.7910/DVN/YB2L5A, Harvard Dataverse, V1.

*Note that version 5.1 of the Dataverse software introduced a free text field for Other Technology Type, which may be used in combination with the Technology Type field (as Measurement Type and Other Measurement Type are used in the example below).*

Life Sciences Metadata ⌃

| | |
|---|---|
| Factor Type ⍰ | Treatment Type |
| Organism ⍰ | Escherichia coli; Homo sapiens |
| Measurement Type ⍰ | Other |
| Other Measurement Type ⍰ | Immunofluorescence microscopy |
| Technology Type ⍰ | Other |
| Cell Type ⍰ | HeLa |

Buchmuller, Benjamin C; Herbst, Konrad; Meurer, Matthias; Kirrmaier, Daniel; Sass, Ehud; Levy, Emmanuel D; Knop, Michael, 2019, "Pooled clone collections by multiplexed CRISPR-Cas12a-assisted gene tagging in yeast [Dataset]", https://doi.org/10.11588/data/L45TRX, heiDATA, V2

| Life Sciences Metadata ⌃ | |
|---|---|
| Design Type ❓ | Technological Design |
| Factor Type ❓ | Genetic Characteristics |
| Organism ❓ | Saccharomyces cerevisiae (brewer's yeast) |
| Measurement Type ❓ | targeted sequencing |
| Technology Type ❓ | nucleotide sequencing |
| Technology Platform ❓ | Illumina; Oxford Nanopore Technologies |

# Journal Metadata

## Introduction

**The Journal Metadata Block is to be used only by journals using a Dataverse repository as its data repository.** This block will identify the journal volume, issue, and article type that a dataset is associated with. The fields in this metadata block help describe the journal's content that is linked to the data as opposed to the dataset itself.

**Note:** The Journal Metadata Block should not be confused with the Citation Metadata Block's *Related Publications field* and *sub-fields.* These fields identify all known publications related to a dataset, whereas the Journal Metadata Block is reserved for Journals choosing to house copies of data directly associated with its publication.

The example used in this section is from the Harvard Dataverse repository.

According to Dataverse documentation, Journal Metadata fields are based on the [Journal Archiving and Interchange Tag Set](#) (JATS), version 1.2.

## Journal Metadata Block

| Field | Definition | Usage[58] | Repeatable | Example |
|---|---|---|---|---|
| **Journal** | Indicates the volume, issue and date of a journal, which this Dataset is associated with. | | | https://doi.org/10.7910/DVN/XKMNAO |
| Volume | The journal volume which this Dataset is associated with (e.g., Volume 4). | O | Yes | 113 |
| Issue | The journal issue number which this Dataset is associated with (e.g., Number 2, Autumn). | O | Yes | 2 |
| Publication Date | The publication date for this journal volume/issue, which this Dataset is associated with (e.g., 1999). *Tip: The date attribute follows the ISO convention of YYYY-MM-DD* | O | Yes | 2019 |
| Type of Article | Indicates what kind of article this is, for example, a research article, a commentary, a book or product review, a case report, a calendar, etc. (based on JATS). *Note: This field is a drop-down list; select one article type.* | O | No | research article |

---

[58] R = Recommended; O = Optional

# References

*About | The Dataverse Project – Dataverse.org*. (n.d.). Dataverse Project. Retrieved August 16, 2021, from
https://dataverse.org/about

*Bounding Box*. (2020, December 28). OpenStreetMap Wiki.
https://wiki.openstreetmap.org/wiki/Bounding_Box

*Bounding Box Tool*. (n.d.). Klokan Technologies. Retrieved August 4, 2021, from
https://boundingbox.klokantech.com/

Canadian Astronomy Data Centre. (2020, January 3). *CFHT Science Archive*. Government of Canada.
https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/cfht/

*DataCite Metadata Schema*. (n.d.). DataCite. Retrieved July 12, 2021, from https://schema.datacite.org/

*DataCite's Value*. (n.d.). DataCite. Retrieved July 12, 2021, from https://datacite.org/value.html

*Dataverse® citation.tsv*. (2021). Institute for Quantitative Social Science.
https://github.com/IQSS/dataverse/blob/5fc015008e553daf8e235b29fd320f8cfd050080/scripts/api/data/metadatablocks/citation.tsv (Original work published 2013)

*Dataverse Collection Management*. (n.d.). Dataverse Project. Retrieved July 12, 2021, from
https://guides.dataverse.org/en/latest/user/dataverse-management.html

*Dataverse Metadata v4+*. (n.d.). Google Docs. Retrieved August 4, 2021, from
https://docs.google.com/spreadsheets/u/0/d/13HP-jI_cwLDHBetn9UKTREPJ_F4iHdAvhjmlvmYdSSw/edit

DDI Alliance. (n.d.-a). *Controlled Vocabularies—Overview Table of Latest Versions*. Data Documentation
Initiative. Retrieved July 23, 2021, from https://ddialliance.org/controlled-vocabularies

DDI Alliance. (n.d.-b). *Data Documentation Initiative*. Data Documentation Initiative. Retrieved July 23,
2021, from https://ddialliance.org/

DDI Alliance. (n.d.-c). *DDI Controlled Vocabulary for Sampling Procedure*. Data Documentation Initiative.
Retrieved August 3, 2021, from https://ddialliance.org/Specification/DDI-CV/SamplingProcedure_1.1.html

DDI Alliance. (n.d.-d). *DDI-Codebook*. Data Documentation Initiative. Retrieved July 12, 2021, from
https://ddialliance.org/Specification/DDI-Codebook/

DDI Alliance. (n.d.-e). *DDI-Codebook 2.5: Element <deviat> (global)*. Data Documentation Initiative.
Retrieved August 3, 2021, from https://ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation_files/schemas/codebook_xsd/elements/deviat.html

Federated Research Data Repository (FRDR). (n.d.). *Documentation*. FRDR-DFDR. Retrieved July 23, 2021, from https://www.frdr-dfdr.ca/docs/en/documentation/

Fry, J., Cooper, A., Mowers, S., & Carrington, C. (2019). *<odesi> Best Practices Document—Based on DDI 2.2x: Version 3.1*. Scholars Portal. https://spotdocs.scholarsportal.info/display/odesi/Handouts?preview=/89132004/212411327/BPDv3-1-2019-01-28.pdf

*GeoNames*. (n.d.). GeoNames. Retrieved August 4, 2021, from https://www.geonames.org/

Hanisch, R., IVOA Resource Registry Working Group, & NVO Metadata Working Group. (n.d.). *Resource Metadata for the Virtual Observatory Version 1.12*. International Virtual Observatory Alliance. https://doi.org/10.5479/ADS/bib/2007ivoa.spec.0302H

*Harvard Dataverse*. (n.d.). Harvard Dataverse. Retrieved July 12, 2021, from https://dataverse.harvard.edu/

*Hubble Legacy Archive*. (n.d.). Hubble Legacy Archive. https://hla.stsci.edu/

*Institute for Quantitative Social Science*. (n.d.). Institute for Quantitative Social Science. Retrieved July 23, 2021, from https://www.iq.harvard.edu/home

*Institutional RDM Contacts*. (n.d.). Portage Network. Retrieved July 23, 2021, from https://portagenetwork.ca/tools-and-resources/institutional-rdm-contacts/

*ISO 639-2: Codes for the representation of names of language*. (2013, January 11). Library of Congress. https://www.loc.gov/standards/iso639-2/php/English_list.php

ISO 3166-1. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=ISO_3166-1&oldid=1030911226

Library and Archives Canada. (2009, April 2). *Government of Canada Core Subject Thesaurus*. https://canada.multites.net/

Moorsel, G. V. (2021, April 13). *VLA Data Archive*. National Radio Astronomy Observatory. https://science.nrao.edu/facilities/vla/archive/index

Mumma, C. C. (2018, November 27). *Metadata Dictionary—Texas Data Repository User Documentation*. Texas Digital Library Wiki. https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/493551668/Metadata+Dictionary

Mungall, C., Bradford, Y., Slyke, C. van, Osumi-Sutherland, D., Balhoff, J., Vasilevsky, N., Meehan, T., Diehl, A., & Haendel, M. (2021, June 29). *Cell Ontology—Ontology Lookup Service*. EMBL-EBI. https://www.ebi.ac.uk/ols/ontologies/cl

National Center for Biomedical Ontology. (n.d.-a). *Browse—NCBO BioPortal*. BioPortal. Retrieved August 4, 2021, from https://bioportal.bioontology.org/ontologies

National Center for Biomedical Ontology. (n.d.-b). *Ontology for Biomedical Investigations (OBI)*. BioPortal. Retrieved July 23, 2021, from https://bioportal.bioontology.org/ontologies/OBI

National Center for Biotechnology Information. (n.d.). *NCBI Taxonomy Homepage*. National Library of Medicine. Retrieved July 23, 2021, from https://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/

National Center for Biotechnology Information & National Library of Medicine. (2019, May). *Journal Archiving and Interchange Tag Library  (JATS): How To Use (Read Me First)*. Journal Archiving and Interchange Tag Library. https://jats.nlm.nih.gov/archiving/tag-library/1.2/chapter/how-to-read.html

OCLC Research. (n.d.-a). *FAST Linked Data*. WorldCat. Retrieved August 4, 2021, from http://experimental.worldcat.org/fast/

OCLC Research. (n.d.-b). *SearchFAST*. OCLC. Retrieved July 23, 2021, from http://fast.oclc.org/searchfast/

Quin, L. (2016, October 11). *Extensible Markup Language (XML)*. W3C. https://www.w3.org/XML/

Rots, A. (2012). *Discovery and Provenance Metadata for Persistent Data Objects in the Virtual Observatory*. https://perma.cc/H5ZJ-4KKY

Rots, A. H. (2007). *Space-Time Coordinate Metadata for the Virtual Observatory Version 1.33*. International Virtual  Observatory  Alliance. https://doi.org/10.5479/ADS/bib/2007ivoa.spec.1030R

*RVMFAST*. (n.d.). Bibliothèque de Université Laval. Retrieved August 4, 2021, from https://rvmweb.bibl.ulaval.ca/rvmfast/initMoteurRecherche.do

Sansone, S.-A., Rocca-Serra, P., Gonzalez-Beltran, A., & Johnson, D. (n.d.). *ISA Model and Serialization Specifications: ISA Abstract Model*. Retrieved August 4, 2021, from https://isa-specs.readthedocs.io/en/latest/isamodel.html

*SIMBAD Astronomical Database*. (2021, August 5). Centre de Données Astronomiques de Strasbourg. http://simbad.u-strasbg.fr/simbad/

Smit, M., & Phillips, D. (2021, April 20). *How To Write A (Good) Data Description: Developing Best Practice*. https://doi.org/10.5281/zenodo.4709835

*User Guide: Appendix—Metadata References*. (n.d.). Dataverse Project. Retrieved July 12, 2021, from https://guides.dataverse.org/en/latest/user/appendix.html

# Example Datasets

## Citation Metadata Block

Doe, J. (2019). *Social Media Use Among Teens, 2015 [Canada]* [Data set]. Scholars Portal Dataverse. https://doi.org/10.5072/FK2/TOXB6Q

## Geospatial Metadata Block

Anderson, L., Beasley, B., Flumerfelt, S.-R., Fox, C., Friesen, S., Macfarlane, G., & McKay, T. (2019). *Replication data for: Long-term monitoring in Barkley Sound: a temporal analysis of intertidal biodiversity on Wizard Islet, British Columbia from 1997 to 2017* [Data set]. Scholars Portal Dataverse. https://doi.org/10.5683/SP2/C8G480

Beamish, A., Scott, N., Wagner, I., & Neil, A. (2019). *Impact of active layer detachments on carbon exchange in a high-Arctic ecosystem, Cape Bounty, Nunavut, Canada (2010)* [Data set]. Scholars Portal Dataverse. https://hdl.handle.net/10864/11825

Choi, D. I. (2019). *Korean War Interviews, 2015* [Data set]. Scholars Portal Dataverse. https://hdl.handle.net/10864/11174

Hird, M. J., Lougheed, S. C., Kuyvenhoven, C., & Rowe, R. K. (2019). *Perspectives on Municipal Waste Management in Kingston, Ontario, 2012* [Data set]. Scholars Portal Dataverse. https://hdl.handle.net/10864/11926

Wilson, R. A., VanDenKerkhof, E. G., Duggan, S., Gilron, I., Good, M. A., Henry, R., & Carley, M. (2020). *Chronic Pain Surveillance at Queen's, 2013-2017* [Data set]. Scholars Portal Dataverse. https://doi.org/10.5683/SP2/GAPNRM

## Social Sciences and Humanities Metadata Block

Ali, S., Hartling, L., Curtis, S., Scott, S., Beran, T., Dow, N., Rajagopal, M., Graham, T., & Jou, H. (2017). *A Randomized Controlled Trial of Children Undergoing Intravenous Placement in the Pediatric Emergency Department: Distraction to Reduce Pain and Distress* [Data set]. UAL Dataverse. https://doi.org/10.7939/DVN/10841

Angus Reid Institute. (2017). *Election 2015: The Politics of Health Care [Canada]* [Data set]. Scholars Portal Dataverse. https://hdl.handle.net/10864/11

Angus Reid Institute, & Bibby, R. (2017). *Religion and Faith in Canada Today, 2015* [Data set]. Scholars Portal Dataverse. https://doi.org/10.5683/SP/78RONJ

Ball, G., Byrne, J., Cave, A., Donaldson, S., Ellendt, C., Holt, N., Jelinski, S., Martz, P., Maximova, K., Padwal, R., & Wild, T. C. (2017). *The Resource Information Program for Parents on Lifestyle and Education (RIPPLE): A Parent-based eHealth Tool to Prevent Childhood Obesity* [Data set]. UAL Dataverse. https://doi.org/10.7939/DVN/10889

Beamish, A., Scott, N., Wagner, I., & Neil, A. (2019). *Impact of active layer detachments on carbon exchange in a high-Arctic ecosystem, Cape Bounty, Nunavut, Canada (2010)* [Data set]. Scholars Portal Dataverse. https://hdl.handle.net/10864/11825

Cheung, M., Sahadath, C., Labelle, P., McEwan, J., & Sikora, L. (2020). *Faculty & Postdoctoral Survey on Research Data Management at the University of Ottawa* [Data set]. Scholars Portal Dataverse. https://doi.org/10.5683/SP/L1H3SS

Collins, P., & Agarwal, A. (2020). *2013 Survey of Commute Patterns among Queen's University Employees* [Data set]. Scholars Portal Dataverse. https://doi.org/10.5683/SP/CNXSVN

Guterbock, T., & Fries, J. (2012). *Maintaining America's Social Fabric* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/FGTJGO

Harris, G. T., Rice, M., & Quinsey, V. (2018). *Violence Risk Appraisal Guide—Revised, 2013* [Data set]. Scholars Portal Dataverse. https://hdl.handle.net/10864/12053

Hawkes, M., Opoka, R., Namasopo, S., Conroy, A., Nyende, S., Kain, K., Bhargava, R., & Mpimbaza, A. (2017). *Solar Powered Oxygen Delivery: An Open-label Non-inferiority Comparison to Standard Oxygen Delivery Using Oxygen Cylinders* [Data set]. UAL Dataverse. https://doi.org/10.7939/DVN/10900

Ipsos Reid. (2017). *Reconnecting Government with Youth 2007* [Data set]. Scholars Portal Dataverse. https://doi.org/10.5683/SP/HY2H1A

Kingston, D., McDonald, S., Tough, S., Austin, M.-P., Hegadoren, K., & Lasiuk, G. (2019). *The Alberta Maternal Mental Health Survey 2012—What Albertans Know About Maternal Mental Health* [Data set]. UAL Dataverse. https://doi.org/10.7939/DVN/10813

Mackie, A., Rempel, G., Kovacs, A., Kaufman, M., Rankin, K., Jelen, A., Manlhiot, C., Anthony, S., Magill-Evans, J., Nicholas, D., Sananes, R., Oechslin, E., Dragieva, D., Mustafa, S., Williams, E., Schuh, M., & McCrindle, B. (2017). *The CHAPTER II Study: Congenital Heart Adolescents Participating in Transition Evaluation Research* [Data set]. UAL Dataverse. https://doi.org/10.7939/DVN/10907

MacLean, J., Tan, S., Fitzgerald, D., & Waters, K. (2017). *Assessing Ventilatory Control in Infants at High Risk of Sleep Disordered Breathing: A Study of Infants with Cleft Lip and/or Palate* [Data set]. UAL Dataverse. https://doi.org/10.7939/DVN/10910

Mancuso, M., Atkinson, M. M., Blais, A., Greene, I., & Nevitte, N. (2017). *A question of ethics. Canadians speak out. Public attitude towards political ethics 1996 data [Canada]* [Data set]. Scholars Portal Dataverse. https://hdl.handle.net/10864/ZJ17A

Maranda, S., Halliday, S., Murray, H., & Cooper, A. (2019). *Evaluation of the Librarian consultations as part of the 2016 Critical Enquiry course in the School of Medicine* [Data set]. Scholars Portal Dataverse. https://doi.org/10.5683/SP/D6NISS

Population Research Laboratory. (2018). *All Alberta Survey, 1987* [Data set]. UAL Dataverse. https://doi.org/10.7939/DVN/10567

Reeder, L. G. (2016). *Health Behavior and Cancer Control, 1976 (M895 and M004)* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/SRVIO4

Statistics Canada. (2018). *Survey of Financial Security, 2016* [Data set]. Abacus Data Network. https://hdl.handle.net/11272.1/AB2/QEIEQS

U. S. Department of Transportation. (2016). *National Personal Transportation Survey, 1983 (M081V1)* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/YT09KD

## Astronomy and Astrophysics Metadata Block

Bayliss, M. R. (2017). *2011 SPT-GMOS Redshift Catalogs, 1D+2D Spectra, and Imaging* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/2HV6OF

Francesco, J. di, & Pierce-Price, D. (2017). *JCMT Continuum Data* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/K4GWMI

Friesen, R. (2016). *Serpens South NH3 Fits* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/SDHQRP

Ginsburg, A. (2018). *W51 GBT W-band H2CO and other lines* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/B7TAFU

Muench, A., Alves, J., Lada, C., & Lada, E. (2015). *Replication data for: Deep 3.8 Micron Observations of the Trapezium Cluster* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/28977

Nakamura, F. (2014). *Substellar-mass Condensations in Prestellar Cores* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/10.1088

Peek, J. (2017). *GALFA-HI DR2 Wide data cubes* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/X4PQ1Z

## Life Sciences Metadata Block

Buchmuller, B. C., Herbst, K., Meurer, M., Kirrmaier, D., Sass, E., Levy, E. D., & Knop, M. (2020). *Pooled clone collections by multiplexed CRISPR-Cas12a-assisted gene tagging in yeast* [Data set]. heiDATA. https://doi.org/10.11588/data/L45TRX

Collins, A. (2018). *Images from screening plates* [Data set]. Harvard Dataverse.
https://doi.org/10.7910/DVN/YB2L5A

Farhat, M. (2015). *Genetic variation* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/AQ5LH5

## Journal Metadata Block

Rozenas, A., & Zhukov, Y. (2020). *Mass Repression and Political Loyalty: Evidence from Stalin's `Terror by Hunger'* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/XKMNAO