

Dataverse North

Metadata Best Practices Guide

This guide was produced by the Metadata subgroup of the Dataverse North Working Group on behalf of the Canadian Association of Research Libraries (CARL) with permission from Harvard for the use of definitions and the Texas Digital Library for basic design.

Alexandra Cooper (Queen's University)
Martine Gagnon (Université Laval)
Mark Goodwin (University of British Columbia)
John Huck (University of Alberta)
Amber Leahey (Scholars Portal)
Michael Steeleworthy (Wilfred Laurier University)
Sally Taylor (University of British Columbia)

Version 1.1

June 2019

Portage Network
Canadian Association of Research Libraries
portage@carl-abrc.ca

www.carl-abrc.ca

portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

Table of Contents

INTRODUCTION.....	2
METADATA BEST PRACTICES	4
OTHER ID.....	4
AUTHOR	4
CONTACT	5
DESCRIPTION.....	5
KEYWORD	6
TOPIC CLASSIFICATION	6
RELATED PUBLICATION.....	7
PRODUCER.....	8
CONTRIBUTOR.....	8
GRANT INFORMATION.....	9
DISTRIBUTOR.....	9
TIME PERIOD COVERED.....	10
DATE OF COLLECTION.....	10
SERIES.....	11
SOFTWARE	11
RELATED MATERIALS AND DATASETS	11
DATA SOURCES.....	12

Introduction

One of the best features of the Dataverse repository platform is the expansive set of metadata fields it provides for describing research data. Standards-based for interoperability, Dataverse metadata supports both project- and dataset-level descriptions, and is compliant with the DataCite¹ 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. This flexibility can make Dataverse metadata seem complicated, especially to anyone new to Dataverse or Research Data Management (RDM) who may be wondering which fields to use or how to interpret them. This guide provides direction to both the novice and experienced user in creating metadata for datasets in a Dataverse repository.

In particular, it:

- Provides definitions and offers tips for clarification where needed. Note: Dataverse administrators can add tips to the properties files that will then display when users hover over field names.
- Distinguishes between required, recommended and optional fields.
 - Required fields are designated by the Dataverse system out-of-the-box. They are Title, Author's Name, Contact Email, Description Text and Subject. Note: Administrators can amend the root Dataverse to make other fields required as needed.
 - Recommended fields are considered best practice to improve data discovery and reuse. Two recommended fields (Contact Name and Producer) were deemed by the guide's authors to be as important as the required fields and could be amended as described above.
 - Optional fields are good to have.
- Illustrates the use of each field with a made-up example. The complete sample dataset can be found on the Demo Scholars Portal Dataverse platform at: <https://doi.org/10.5072/FK2/TOXB6Q>

This version of the guide provides guidance on the general citation block for Dataverse 4.x. Future versions will provide guidance on the use of the domain-specific metadata blocks: Social Sciences, Geospatial, Life Sciences, and Astronomy & Astrophysics Metadata.

Questions? Please see the list of library contacts at your institution.

<https://portagenetwork.ca/planning-managing-data/contacts-at-your-organization/>

¹ "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"

(<http://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.

Metadata Best Practices

Field	Definition with tips	Required/ Recommended/ Optional	Example
Title	Full title by which the Dataset is known.	required	Social Media Use Among Teens, 2015 [Canada]
Subtitle	A secondary title used to amplify or state certain limitations on the main title. <i>Tip: subtitle is not included in generated citation. Include subtitle with title to be included in citation.</i>	recommended (if applicable)	Main Survey
Alternative Title	A title by which the work is commonly referred or an abbreviation of the title. <i>Tip: Acronym or short form of full title.</i>	optional	Youth Social Media Survey
Alternative URL	A URL where the Dataset can be viewed, such as a personal or project website.	optional	http://youthsocialmedia.org
Other ID	Metadata in this section refer to another unique identifier that identifies this Dataset (e.g., producer's or another repository's number).		
Agency	Name of agency that generated this identifier.	optional	Youth Communication Development Project, Education Department, Queen's University
Identifier	Other identifier that corresponds to this Dataset.	optional	2202
Author	Metadata in this section refer to 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 organization responsible for this Dataset.	required	Doe, Jane
Affiliation	The organization with which the author is affiliated.	recommended	Queen's University

Identifier Scheme	Name of the identifier scheme (ORCID, ISNI). <i>Tip: ORCID is a non-proprietary alphanumeric code to identify scientific and other academic authors and contributors uniquely.</i>	recommended	ORCID
Identifier	Uniquely identifies an individual author or organization, according to various schemes.	recommended	111111
Contact	Metadata in this section refer to the contact(s) for this Dataset.		
Name	The contact's Family Name, Given Name or the name of the organization.	recommended	Doe, Jane
Affiliation	The organization with which the contact is affiliated.	recommended	Queen's University
E-mail	The e-mail address(es) of the contact(s) for the Dataset. This will not be displayed. <i>Tip: could be an institutional email.</i>	required	jdoe@email.com
Description	Metadata in this section refer to a summary describing the purpose, nature, and scope of the Dataset.		
Text	A summary describing the purpose, nature, and scope of the Dataset.	required	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	optional	2018-01-18

	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.		
Subject	Domain-specific Subject Categories that are topically relevant to the Dataset.	required	Social Sciences
Keyword	Metadata in this section refer to 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 controlled vocabulary can be employed.	recommended	Social media, Communication
Vocabulary	For the specification of the keyword controlled vocabulary in use, such as LCSH, MeSH, or others. <i>Tip: controlled vocabulary is a standardized list of terminology for describing information (e.g. LCSH is Library of Congress Subject Heading, MeSH is Medical Subject. Heading).</i>	optional	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 .	optional	http://www.thesaurus.gc.ca/recherche-search/mtwdk.exe?k=these&l=60&w=4790&n=1&s=5&t=2
Topic Classification	Metadata in this section refer to the classification field indicates the broad important topic(s) and subjects that the data cover.		
Term	Topic or Subject term that is relevant to this Dataset.	optional	Society and Culture

Vocabulary	Provided for specification of the controlled vocabulary in use, e.g. LCSH, MeSH, etc. <i>Tip: controlled vocabulary is a standardized list of terminology for describing information (e.g. LCSH is Library of Congress Subject Heading, MeSH is Medical Subject Heading).</i>	optional	Government of Canada Core Subject Thesaurus
Vocabulary URL	Specifies the URL location for the full controlled vocabulary.	optional	http://www.thesaurus.gc.ca/recherche-search/mtwdk.exe?k=these&l=60&n=0&s=cid&t=&w=97&h=SO%20Society%20and%20Culture
Related Publication	Metadata in this section refer to publications that use the data from this Dataset.		
Citation	Other identifier that corresponds to this Dataset. <i>Tip: The full bibliographic citation for any related publication.</i>	recommended (if applicable)	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), handle, ISBN). <i>Tip: DOIs and handles are persistent identifiers used to identify digital objects uniquely.</i>	recommended (if applicable)	doi
ID Number	The identifier for the selected ID type.	recommended (if applicable)	10.0000/SP/TEST
URL	Link to the publication web page (e.g., journal article page, archive record page, or other).	optional	https://doi.org/10.0000/SP/TEST
Notes	Additional important information about the Dataset.	optional	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	optional	English-Canada
Producer	Metadata in this section refer to person or organization with the financial or administrative responsibility over this Dataset.		
Name	Producer name	recommended	Youth Communication Development Project
Affiliation	The organization with which the producer is affiliated.	recommended	Queen's University
Abbreviation	The abbreviation by which the producer is commonly known. (ex. IQSS, ICPSR)	optional	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 .	optional	http://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 .	optional	http://youthsocialmedia.org/image.png
Production Date	Date when the data collection or other materials were produced (not distributed, published or archived). <i>Tip: date when dataset was finalized and ready for analysis or distribution.</i>	recommended	2016-01-11
Production Place	The location where the data collection and any other related materials were produced.	recommended	Kingston, Ontario, Canada
Contributor	Metadata in this section refer to 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.	recommended	Researcher
Name	The Family Name, Given Name or organization name of the contributor.	recommended	Doe, Jane

Grant Information	Metadata in this section refer to grant information.		
Grant Agency	Grant Number Agency	recommended (if applicable)	Social Sciences and Humanities Research Council (SSHRC)
Grant Number	The grant or contract number of the project that sponsored the effort.	recommended (if applicable)	CCB123456
Distributor	Metadata in this section refer to the organization designated by the author or producer to generate copies of the particular work including any necessary editions or revisions.		
Name	Distributor name	recommended	Data Services
Affiliation	The organization with which the distributor contact is affiliated.	recommended	Queen's University Library
Abbreviation	The abbreviation by which this distributor is commonly known (e.g., IQSS, ICPSR).	optional	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 .	optional	http://library.queensu.ca/data/services
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 .	optional	http://www.queensu.ca/encyclopedia/sites/webpublish.queensu.ca.qencwww/files/images/1/logo/QueensLogo_colour.png
Distribution Date	Date that the work was made available for distribution/presentation. <i>Tip: This field may be the same as the Deposit Date. Use the field if data was previously distributed.</i>	optional	2018-01-22
Depositor	The person (Family Name, Given Name) or the name of the organization that deposited this Dataset to the repository. <i>Tip: The name of the person/institution who provided the dataset(s) to the archive (i.e. not necessarily the person doing the submission into DV).</i>	recommended	Doe, Jane

Deposit Date	Date that the Dataset was deposited into the repository. <i>Tip: Date is pre-populated with the date of upload into Dataverse. It can be edited to reflect the date when the data was received by an external or mediated data repository service.</i>	recommended	2018-01-15
Time Period Covered	Metadata in this section refer to the 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 that 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.	recommended	2015-03-20
End	End date that 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.	recommended	2015-06-21
Date of Collection	Metadata in this section refer to the date(s) when the data were collected.		
Start	Date when the data collection started.	recommended	2015-03-20
End	Date when the data collection ended.	recommended	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.	recommended	Survey data

Series			
Metadata in this section refer to information about the Dataset series.			
Name	Name of the dataset series to which the Dataset belongs.	recommended (if applicable)	Social Media Use Among Teens
Information	History of the series and summary of those features that apply to the series as a whole.	recommended (if applicable)	Established in 2005, the Youth Communication Development Project aims to gather key research and data about youth development and social media use through a series of independent, annual, cross-sectional surveys titled Social Media Use Among Teens. The overall objectives 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			
Metadata in this section refer to information about the software used to generate the Dataset.			
Name	Name of software used to generate the Dataset. <i>Tip: useful for specialized software or instruments.</i>	optional	SPSS
Version	Version of the software used to generate the Dataset.	optional	24
Related materials and datasets			
Metadata in this section refer to other related publications and datasets.			
Related Material	Any material related to this Dataset.	optional	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.	optional	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.	optional	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	Metadata in this section refer to data that has been retrieved from other sources (e.g. publications or machine-readable data files).		
Data Sources	List of books, articles, serials, or machine-readable data files that served as the sources of the data collection.	optional	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
Origin of Sources	For historical materials, information about the origin of the sources and the rules followed in establishing the sources should be specified.	optional	National Household Survey, 2011. http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5178
Characteristic of Sources Noted	Assessment of characteristics and source material. <i>Tip: describes noteworthy aspects of the data collected.</i>	optional	
Documentation and Access to Sources	Level of documentation of the original sources. <i>Tip: can be used to explain any restrictions or access to source data documentation.</i>	optional	Open