

## Texas Data Repository Metadata Dictionary

The Texas Data Repository Metadata Dictionary is comprised of [citation](#) (including [journal](#) metadata) and domain specific ([geospatial](#), [social science and humanities](#), [astronomy and astrophysics](#), and [life sciences](#)) metadata fields. The dictionary contains a definition for each field, describes any format specifications associated with each field, denotes if the field is mandatory or optional, and establishes whether the field is visible to the depositor during upload. Note that TDL members may choose to expand the number of mandatory fields based on the institution's operating procedures.

### Citation Metadata<sup>1</sup>

Field	Definition	Formatting	Mandatory/ Optional	Viewable During Upload
Title	Full title by which the Dataset is known.		M	Y
Subtitle	A secondary title used to amplify or state certain limitations on the main title.		O	N
Alternative Title	A title by which the work is commonly referred, or an abbreviation of the title.		O	N
Alternative URL	A URL where the dataset can be viewed, such as a personal or project website.	Enter full URL, starting with http://  Example: <ul style="list-style-type: none"> <li><a href="http://www.my.org">http://www.my.org</a></li> </ul>	O	N

<sup>1</sup> For full metadata information on Citation Metadata, including display properties and which fields are available for advanced search, see the citation.tsv file on Dataverse's Github: <https://github.com/IQSS/dataverse/blob/master/scripts/api/data/metadatablocks/citation.tsv>.

<i>Other ID</i>	<i>Data in this section refer to another unique identifier that identifies this Dataset (e.g., producer's or another repository's number).</i>			
Agency	Name of agency which generated this identifier.		O	N
Identifier	Other identifier that corresponds to this Dataset.		O	N
<i>Author</i>	<i>Data in this section refer to the person(s), corporate body(ies), or agency(ies) responsible for creating the work.</i>			
Name	The author's Family Name, Given Name or the name of the organization responsible for this Dataset.	Personal name expressed as LastName, FirstName MiddleInitial.  Organizational name as it appears  Examples: <ul style="list-style-type: none"> <li>• Obama, Barack H.</li> <li>• Texas Digital Library</li> </ul>	M	Y
Affiliation	The organization with which the author is affiliated.		M	Y
Identifier Scheme	Name of the identifier scheme (ORCID, ISNI, LCNA).		O	Y
Identifier	Uniquely identifies an individual author or organization, according to various schemes.		O	Y
<i>Contact</i>	<i>Data in this section refer to the contact(s) for this Dataset.</i>			
Name	The contact's Family Name, Given Name or the name of the organization.	Personal name expressed as LastName, FirstName MiddleInitial.  Organizational name as it appears  Examples: <ul style="list-style-type: none"> <li>• Obama, Barack H.</li> <li>• Texas Digital Library</li> </ul>	O	Y

Affiliation	The organization with which the contact is affiliated.		O	Y
E-mail	The e-mail address(es) of the contact(s) for the Dataset. This will not be displayed to the user.		M	Y
<i>Description</i>	<i>Data in this section refer to a summary describing the purpose, nature, and scope of the Dataset.</i>			
Text	A summary describing the purpose, nature, and scope of the Dataset.		M	Y
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.	Date expressed in ISO format (YYYY-MM-DD)  Example: <ul style="list-style-type: none"> <li>• 2016-01-30</li> </ul>	O	Y
Subject	Domain-specific Subject Categories that are topically relevant to the Dataset.		M	Y
<i>Keyword</i>	<i>Data in this section refer to key terms that describe important aspects of the Dataset.</i>			
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.		O	Y
Vocabulary	For the specification of the keyword controlled vocabulary in use.	Abbreviated name of vocabulary  Examples: <ul style="list-style-type: none"> <li>• LCSH</li> <li>• LCNA</li> <li>• MeSH</li> </ul>	O	Y

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.  Example: <ul style="list-style-type: none"> <li>• <a href="http://www.my.org">http://www.my.org</a></li> </ul>	O	Y
<i>Topic Classification</i>	<i>Data in this section refer to the classification field that indicates the broad important topic(s) and subjects that the data cover.</i>			
Term	Topic or Subject term that is relevant to this Dataset.		O	N
Vocabulary	Provided for specification of the controlled vocabulary in use.	Abbreviated name of vocabulary  Examples: <ul style="list-style-type: none"> <li>• LCSH</li> <li>• LCNA</li> <li>• MeSH</li> </ul>	O	N
Vocabulary URL	Specifies the URL location for the full controlled vocabulary.	Enter an absolute URL where the keyword vocabulary web site is found.  Example: <ul style="list-style-type: none"> <li>• <a href="http://www.my.org">http://www.my.org</a></li> </ul>	O	N
<i>Related Publication</i>	<i>Data in this section refer to publications that use the data from this Dataset.</i>			
Citation	The full bibliographic citation for this related publication.		O	N
ID Type	The type of digital identifier used for this publication (e.g., Digital Object Identifier (DOI)).	Examples: <ul style="list-style-type: none"> <li>• Digital Object Identifier (DOI)</li> <li>• handle</li> <li>• isbn</li> </ul>	O	N

ID Number	The identifier for the selected ID type.		O	N
URL	Link to the publication web page	Enter an absolute URL where the keyword vocabulary web site is found.  Example: <ul style="list-style-type: none"> <li>• <a href="http://www.my.org">http://www.my.org</a></li> </ul>	O	N
Notes	Additional important information about the Dataset.		O	Y
Language	Language of the Dataset		O	N
<i>Producer</i>	<i>Data in this section refer to the person or organization with the financial or administrative responsibility over this Dataset</i>			
Name	Producer name	Personal name expressed as LastName, FirstName MiddleInitial.  Organizational name as it appears  Examples: <ul style="list-style-type: none"> <li>• Obama, Barack H.</li> <li>• Texas Digital Library</li> </ul>	O	N
Affiliation	The organization with which the producer is affiliated.		O	N
Abbreviation	The abbreviation by which the producer is commonly known.	Example: <ul style="list-style-type: none"> <li>• TDL</li> </ul>	O	N
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 <a href="http://www.my.org">http://www.my.org</a> .	Enter an absolute URL where the keyword vocabulary web site is found.  Example: <ul style="list-style-type: none"> <li>• <a href="http://www.my.org">http://www.my.org</a></li> </ul>	O	N

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  Example: <ul style="list-style-type: none"> <li>• <a href="http://www.my.org/images/logo.gif">http://www.my.org/images/logo.gif</a></li> </ul>	O	N
Production Date	Date when the data collection or other materials were produced (not distributed, published or archived).	Date expressed in ISO format (YYYY-MM-DD)  Example: <ul style="list-style-type: none"> <li>• 2016-01-30</li> </ul>	M	Y
Production Place	The location where the data collection and any other related materials were produced.		M	Y
<i>Contributor</i>	<i>Data 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.</i>			
Type	The type of contributor of the resource.		O	N
Name	The Family Name, Given Name or organization name of the contributor.	Personal name expressed as LastName, FirstName MiddleInitial.  Organizational name as it appears  Examples: <ul style="list-style-type: none"> <li>• Obama, Barack H.</li> <li>• Texas Digital Library</li> </ul>	O	N
<i>Grant Information</i>	<i>Data in this section refer to grant Information</i>			
Grant Agency	Grant Number Agency		O	N
Grant Number	The grant or contract number of the project that sponsored the effort.		O	N

<i>Distributor</i>	<i>Data 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.</i>			
Name	Distributor name	Personal name expressed as LastName, FirstName MiddleInitial.  Organizational name as it appears  Examples: <ul style="list-style-type: none"> <li>• Obama, Barack H.</li> <li>• Texas Digital Library</li> </ul>	O	N
Affiliation	The organization with which the distributor contact is affiliated.		O	N
Abbreviation	The abbreviation by which this distributor is commonly known (e.g., IQSS, ICPSR).	Example: <ul style="list-style-type: none"> <li>• TDL</li> </ul>	O	N
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 <a href="http://www.my.org">http://www.my.org</a> .	Enter an absolute URL where the keyword vocabulary web site is found.  Example: <ul style="list-style-type: none"> <li>• <a href="http://www.my.org">http://www.my.org</a></li> </ul>	O	N
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 <a href="http://www.my.org/images/logo.gif">http://www.my.org/images/logo.gif</a> .	Enter an absolute URL where the producer's logo image is found  Example: <ul style="list-style-type: none"> <li>• <a href="http://www.my.org/images/logo.gif">http://www.my.org/images/logo.gif</a></li> </ul>	O	N

Distribution Date	Date that the work was made available for distribution/presentation.	Date expressed in ISO format (YYYY-MM-DD)  Example: <ul style="list-style-type: none"> <li>• 2016-01-30</li> </ul>	O	N
Depositor	The person (Family Name, Given Name) or the name of the organization that deposited this Dataset to the repository.	Personal name expressed as LastName, FirstName MiddleInitial.  Organizational name as it appears  Examples: <ul style="list-style-type: none"> <li>• Obama, Barack H.</li> <li>• Texas Digital Library</li> </ul>	O	Y
Deposit Date	Date that the Dataset was deposited into the repository.	Date expressed in ISO format (YYYY-MM-DD)  Examples: <ul style="list-style-type: none"> <li>• 2016-01-30</li> </ul>	O	Y
<i>Time Period Covered</i>	<i>Data 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.</i>			
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.	Date expressed in ISO format (YYYY-MM-DD)  Examples: <ul style="list-style-type: none"> <li>• 2016-01-30</li> </ul>	O	N



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.	Date expressed in ISO format (YYYY-MM-DD)  Examples: <ul style="list-style-type: none"> <li>• 2016-01-30</li> </ul>	O	N
<i>Date of Collection</i>	<i>Data in this section refer to date(s) when the data were collected.</i>			
Start	Date when the data collection started.	Date expressed in ISO format (YYYY-MM-DD)  Examples: <ul style="list-style-type: none"> <li>• 2016-01-30</li> </ul>	O	N
End	Date when the data collection ended.	Date expressed in ISO format (YYYY-MM-DD)  Examples: <ul style="list-style-type: none"> <li>• 2016-01-30</li> </ul>	O	N
Kind of Data	Type of data included in the file	Free text  Examples: <ul style="list-style-type: none"> <li>• survey data</li> <li>• census/enumeration data</li> <li>• aggregate data</li> <li>• clinical data</li> <li>• event/transaction data</li> <li>• program source code</li> <li>• machine-readable text</li> <li>• administrative records data</li> <li>• experimental data</li> </ul>	M	Y

		<ul style="list-style-type: none"> <li>● psychological test</li> <li>● textual data</li> <li>● coded textual</li> <li>● coded documents</li> <li>● time budget diaries</li> <li>● observation data/ratings</li> <li>● process-produced data</li> </ul>		
<i>Series</i>	<i>Data in this section refer to information about the Dataset series.</i>			
Name	Name of the dataset series to which the Dataset belongs.		O	N
Information	History of the series and summary of those features that apply to the series as a whole.		O	N
<i>Software</i>	<i>Data in this section refer to the software used to generate the Dataset.</i>			
Name	Name of software used to generate the Dataset.		O	N
Version	Version of the software used to generate the Dataset.		O	N
Related Material	Any material related to this Dataset.		O	N
Related Datasets	Any Datasets that are related to this Dataset, such as previous research on this subject.		O	N
Other References	Any references that would serve as background or supporting material to this Dataset.		O	N
Data Sources	List of books, articles, serials, or machine-readable data files that served as the sources of the data collection.		O	N

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	N
Characteristic of Sources Noted	Assessment of characteristics and source material.		O	N
Documentation and Access to Sources	Level of documentation of the original sources.		O	N

Journal Metadata<sup>2</sup>

Field	Definition	Formatting	Mandatory /Optional	Viewable During Upload
<i>Journal</i>	<i>Data in this section refer to the volume, issue, and date of a journal, which this Dataset is associated with.</i>			
Volume	The journal volume which this DataSet is associated with	Example: • Volume 4	O	N
Issue	The journal issue number which this DataSet is associated with	Example: • Number 2, Autumn	O	N
Publication Date	The publication date for this journal volume/ issue number, which this DataSet is associated with	Date expressed in ISO format (YYYY-MM-DD)  Examples: • 2016-01-30 • 1999	O	N

<sup>2</sup> For full metadata information on Journal Metadata, including display properties and which fields are available for advanced search, see the journals.tsv file on Dataverse's Github: <https://github.com/IQSS/dataverse/blob/master/scripts/api/data/metadatablocks/journals.tsv>.

Type of Article	Indicates what kind of article this is	Values from JATS  Examples: <ul style="list-style-type: none"> <li>• research article</li> <li>• commentary</li> <li>• book review</li> <li>• product review</li> </ul>	O	N
-----------------	--	---	---	---

### Domain Specific Metadata

### Geospatial Metadata<sup>3</sup>

Field	Definition	Formatting	Mandatory/ Optional	Viewable During Upload
<i>Geographic Coverage</i>	<i>Data in this section refer to the geographic coverage of the data. Includes the total geographic scope of the data.</i>			
Country / Nation	The country or nation that the Dataset is about.		O	N
State / Province	The state or province that the Dataset is about.	Use GeoNames for correct spelling and avoid abbreviations.	O	N
City	The name of the city that the Dataset is about.	Use GeoNames for correct spelling and avoid abbreviations.	O	N
Other	Other information on the geographic coverage of the data.		O	N
Geographic Unit	Lowest level of geographic aggregation covered by the Dataset, e.g., village, county, region.		O	N

<sup>3</sup> For full metadata information on Geospatial Metadata, including display properties and which fields are available for advanced search, see the geospatial.tsv file on Dataverse's Github: <https://github.com/IQSS/dataverse/blob/master/scripts/api/data/metadatablocks/geospatial.tsv>.

<b>Geographic Bounding Box</b>	<i>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.</i>			
West Longitude	Westernmost coordinate delimiting the geographic extent of the Dataset.	A valid range of values, expressed in decimal degrees, is $-180,0 \leq \text{West Bounding Longitude Value} \leq 180,0$ .	O	N
East Longitude	Easternmost coordinate delimiting the geographic extent of the Dataset.	A valid range of values, expressed in decimal degrees, is $-180,0 \leq \text{East Bounding Longitude Value} \leq 180,0$ .	O	N
North Latitude	Northernmost coordinate delimiting the geographic extent of the Dataset.	A valid range of values, expressed in decimal degrees, is $-90,0 \leq \text{North Bounding Latitude Value} \leq 90,0$ .	O	N
South Latitude	Southernmost coordinate delimiting the geographic extent of the Dataset.	A valid range of values, expressed in decimal degrees, is $-90,0 \leq \text{South Bounding Latitude Value} \leq 90,0$ .	O	N

Social Science & Humanities Metadata<sup>4</sup>

Field	Definition	Formatting	Mandatory/Optional	Viewable During Upload
Unit of Analysis	Basic unit of analysis or observation that this Dataset describes, such as individuals, families/households, groups,	For information about the DDI's controlled vocabulary for this element, please refer to the DDI web page at	O	N

<sup>4</sup> For full metadata information on Social Science & Humanities Metadata, including display properties and which fields are available for advanced search, see the social\_science.tsv file on Dataverse's Github:

[https://github.com/IQSS/dataverse/blob/master/scripts/api/data/metadatablocks/social\\_science.tsv](https://github.com/IQSS/dataverse/blob/master/scripts/api/data/metadatablocks/social_science.tsv).

	institutions/organizations, administrative units, and more.	<a href="http://www.ddialliance.org/Specification/DDI-CV/">http://www.ddialliance.org/Specification/DDI-CV/</a> .		
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.		O	N
Time Method	The time method or time dimension of the data collection, such as panel, cross-sectional, trend, time-series, or other.		O	N
Data Collector	Individual, agency or organization responsible for administering the questionnaire or interview or compiling the data.		O	N
Collector Training	Type of training provided to the data collector		O	N

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	N
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.		O	N
<i>Target Sample Size</i>	<i>Data in this section refer to specific information regarding the target sample size, actual sample size, and the formula used to determine this.</i>			
Actual	Actual sample size.		O	N
Formula	Formula used to determine target sample size.		O	N
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	N
Collection Mode	Method used to collect the data; instrumentation characteristics (e.g., telephone interview, mail questionnaire, or other).		O	N
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 questions. Unstructured indicates that in-depth interviews were conducted.	O	N
Characteristics of Data	Description of noteworthy aspects of the data collection situation. Includes information on factors		O	N

Collection Situation	such as cooperativeness of respondents, duration of interviews, number of callbacks, or similar.			
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	N
Control Operations	Control Operations Methods to facilitate data control performed by the primary investigator or by the data archive.		O	N
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.	O	N
Cleaning Operations	Methods used to clean the data collection, such as consistency checking, wild code checking, or other.		O	N
Study Level Error Notes	Note element used for any information annotating or clarifying the methodology and processing of the study.		O	N
Response Rate	Percentage of sample members who provided information.		O	N
Estimates of Sampling Error	Measure of how precisely one can estimate a population value from a given sample.		O	N
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		O	N



	and response bias, confidence levels, question bias, or similar.			
<i>Notes</i>	<i>Data in this section refer to general notes about this Dataset.</i>			
Type	Type of note.		O	N
Subject	Note subject.		O	N
Text	Text for this note.		O	N

Astronomy and Astrophysics Metadata<sup>5</sup>

Field	Definition	Formatting	Mandatory/ Optional	Viewable During Upload
Type	The nature or genre of the content of the files in the dataset.		O	N
Facility	The observatory or facility where the data was obtained.		O	N
Instrument	The instrument used to collect the data.		O	N
Object	Astronomical Objects represented in the data (Given as SIMBAD recognizable names preferred).		O	N
Spatial Resolution	The spatial (angular) resolution that is typical of the observations.	Use decimal degrees.	O	N
Spectral Resolution	The spectral resolution that is typical of the observations, given as the ratio $\lambda/\Delta\lambda$ .		O	N

<sup>5</sup> For full metadata information on Astronomy and Astrophysics Metadata, including display properties and which fields are available for advanced search, see the astrophysics.tsv file on Dataverse's Github:

<https://github.com/IQSS/dataverse/blob/master/scripts/api/data/metadatablocks/astrophysics.tsv>.

Time Resolution	The temporal resolution that is typical of the observations.	Given in seconds	O	N
Bandpass	Conventional bandpass name		O	N
Central Wavelength (m)	The central wavelength of the spectral bandpass.	Use meters	O	N
<i>Wavelength Range</i>	<i>Data in this section refer to the minimum and maximum wavelength of the spectral bandpass.</i>			
Minimum (m)	The minimum wavelength of the spectral bandpass.	Use meters	O	N
Maximum (m)	The maximum wavelength of the spectral bandpass.	Use meters	O	N
<i>Dataset Date Range</i>	<i>Data in this section refer to the time period covered by the data.</i>			
Start	Dataset Start Date		O	N
End	Dataset End Date		O	N
Sky Coverage	The sky coverage of the data object.		O	N
Depth Coverage	The (typical) depth coverage, or sensitivity, of the data object in Jy.		O	N
Object Density	The (typical) density of objects, catalog entries, telescope pointings, etc..	Use number per square degree	O	N
Object Count	The total number of objects, catalog entries, etc., in the data object.		O	N
Fraction of Sky	The fraction of the sky represented in the observations.	Value ranging from 0 to 1	O	N
Polarization	The polarization coverage		O	N

RedshiftType	RedshiftType string C "Redshift"; or "Optical" or "Radio" definitions of Doppler velocity used in the data object.		O	N
Redshift Resolution	The resolution in redshift (unitless) or Doppler velocity (km/s) in the data object.		O	N
<i>Redshift Value</i>	<i>Data in this section refer to the value of the redshift (unitless) or Doppler velocity (km/s in the data object).</i>			
Minimum	The minimum value of the redshift (unitless) or Doppler velocity (km/s) in the data object.		O	N
Maximum	The maximum value of the redshift (unitless) or Doppler velocity (km/s in the data object.		O	N

Life Sciences Metadata<sup>6</sup>

Field	Definition	Formatting	Mandatory/ Optional	Viewable to Depositor
Design Type	Design types that are based on the overall experimental design.		O	N
Factor Type	Factors used in the Dataset.		O	N
Organism	The taxonomic name of the organism used in the Dataset or from which the starting biological material derives.		O	N
Other Organism	If Other was selected in Organism, list any other organisms that were used in this Dataset.	Terms from the NCBI Taxonomy are recommended.	O	N

<sup>6</sup> For full metadata information on Biomedical Metadata, including display properties and which fields are available for advanced search, see the astrophysics.tsv file on Dataverse's Github: <https://github.com/IQSS/dataverse/blob/master/scripts/api/data/metadatablocks/biomedical.tsv>.

Measurement Type	A term to qualify the endpoint, or what is being measured (e.g. gene expression profiling; protein identification).		O	N
Other Measurement Type	If Other was selected in Measurement Type, list any other measurement types that were used.	Terms from NCBO Bioportal are recommended.	O	N
Technology Type	A term to identify the technology used to perform the measurement (e.g. DNA microarray; mass spectrometry).		O	N
Technology Platform	The manufacturer and name of the technology platform used in the assay (e.g. Bruker AVANCE).		O	N
Cell Type	The name of the cell line from which the source or sample derives.		O	N