Long-term data for Europe

EURHISFIRM

M4.1: Report on the Protocol of Data Documentation

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

http://www.eurhisfirm.eu

AUTHOR:

Johan POUKENS (University of Antwerp)*

APPROVED IN 2019 BY:

Jan ANNAERT (University of Antwerp)

Wolfgang KÖNIG (Goethe University Frankfurt)

Angelo RIVA (Paris School of Economics)

* The author would like to thank Oliver Watteler (GESIS) and Coen Fierst van Wijnandsbergen (Rotterdam School of Management, Erasmus University) for their valuable comments to earlier drafts and Lana Yoo (Paris School of Economics) for revising the language of this document.

Table of Contents

1 Introduction 4
2 Protocol of data documentation 6
2.1 Citation metadata 6
2.2 Upload files 7
2.3 Additional metadata 8
2.4 Variable metadata 9
3 References 11
4 Appendices 11
4.1 Appendix 1: Dataverse metadata elements 11
4.2 Appendix 2: Example 20

This project has received funding from
the European Union’s Horizon 2020 research and innovation programme
under grant agreement N° 777489

1 Introduction

EURHISFIRM will rely both on data which is extracted from digitised images of printed sources and from datasets which are produced by researchers or research teams who are willing to deposit their data within the research infrastructure. It is therefore necessary to have a platform for uploading images or datasets. After upload, new data will also be aligned and matched with existing data in order to create a body of long-term, standardised data. Matching data from different datasets, however, requires that data are adequately documented, i.e. that they are accompanied by metadata. Metadata are data that describe data, or "data about data". Adequate in this case means that metadata need to be both sufficiently detailed and standardised. Standardisation is important for minimising interpretation errors. Working Package 4 therefore evaluated different metadata or data documentation standards for the social sciences. Only Data Documentation Initiative (DDI) standards were deemed to contain the elements necessary to describe datasets with sufficient detail. Two version or DDI are available: DDI Codebook (or DDI 2.5) and DDI Lifecycle (DDI 3.2).[1] As the name of the latter suggests, DDI Lifecycle was designed to document data across all stages of the research lifecycle. It therefore is much more extensive then DDI Codebook. DDI Codebook contains fewer elements because it aims to describe single datasets. Fewer, in this case, is relative, however, because the DDI 2.5 XML Schema still has 351 elements. The complexity of DDI is alleviated in part by the availability of specialised software (Nesstar and Colectica, for instance) for producing data documentation according to the DDI specifications without knowledge of XML.[2] But some of this software is proprietary and it comes with a learning curve as well. We therefore propose that EURHISFIRM implements a user-friendly, online interface for researchers to document data through web forms which require no knowledge of DDI or XML.

The proposed protocol of data documentation will guide researchers who want to contribute images or datasets to EURHISFIRM through the upload process. The current version of the protocol assumes that the Dataverse software is chosen as EURFISFIRM's platform for uploading and documenting images and data. Dataverse is an open source research data repository software from the Institute for Quantitative Social Science (IQSS) at Harvard University.[3] It provides functionalities for sharing, preserving, citing, exploring, and analysing data. By incorporating standards (DDI 2.5), controlled vocabularies and persistent identifiers, Dataverse meets FAIR Data Principles (Wilkinson et al., 2016). Its user-interface is very user friendly and should become more and more familiar to researchers over time as the number of Dataverse implementations continuously grows. Dataverse also offers many possibilities for the customisation of metadata, including adding or editing metadata fields, instructional text and controlled vocabularies through .tsv files (tab separated values). Hiding or making fields required is also possible through the web interface.[4]

Dataverse currently has two limitations. Firstly, there is no support for DDI 3.2 metadata. This means that uploaders cannot reference re-usable, common metadata elements such as Conceptual Variable when documenting data (see also D4.5). Secondly, uploaders can only edit metadata at the Dataset (or Study

[1] https://www.ddialliance.org
[2] http://www.nesstar.com
[3] https://dataverse.org
[4] http://guides.dataverse.org/en/latest/admin/metadatacustomization.html

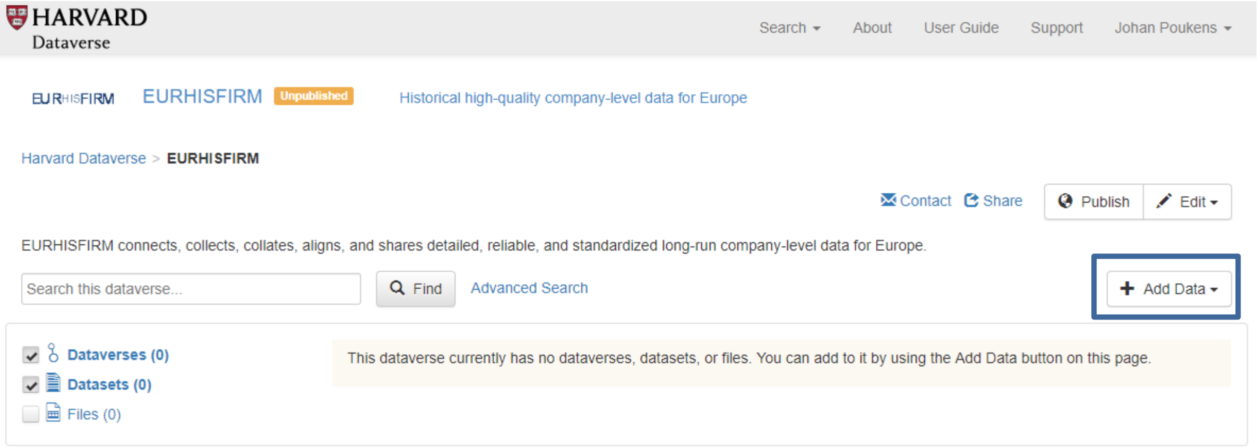
This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

Unit) level. Metadata at the Variable level is extracted automatically from the uploaded files by the Dataverse software. This means that if the file format is not (fully) supported by the automated extraction process, Variables will not be documented at all or incompletely. Dataverse, for instance, cannot extract variable Labels from .csv files (comma separated values). These limitations are currently already being addressed by the Dataverse community and might be resolved by the time EURHISFIRM reaches the implementation phase. For the time being, however, we propose to work around the Variable documentation issue by uploading Variable metadata in a separate Excel file.

The remainder of this document is composed of the steps a user has to follow to upload and document datasets in Dataverse. For this report, we used a hosted Dataverse with limited possibilities for the customisation of metadata fields (it was, for instance, no possible to add or rename metadata fields) (see also D4.5). A list of all standard metadata fields available in hosted Dataverses is included in Appendix 1. This lacks a field for the Publisher, however. Given its importance in case of printed sources, this issue should be addressed in future implementations of Harvard Dataverse for EURHISFIRM. Appendix 2 provides an example of a metadata record in Dataverse.

2 Protocol of data documentation

Thank you for contributing data to EURHISFIRM. You can upload files and add metadata through the EURHISFIRM Dataverse (https://dataverse.harvard.edu/dataverse/eurhisfirm). The upload and documentation process is fairly simple, but more information about uploading and documenting datasets can be found online (http://guides.dataverse.org/en/4.14/user/dataset-management.html). Just click on the **Add Data** button in the top-right corner of the screen to get started. If you are not logged in, you will be prompted to log in to your Dataverse account. If you do not have a Dataverse account yet, you can click the **Sign Up** button to create an account instead.

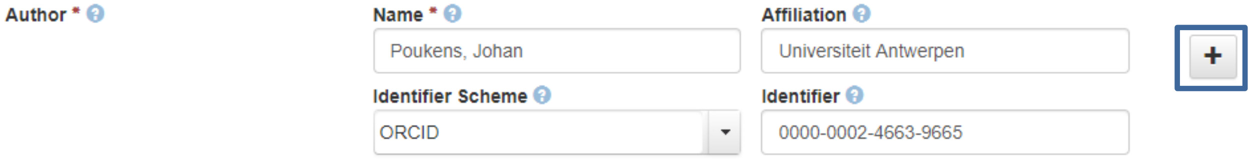


EURHISFIRM connects, collects, collates, aligns, and shares detailed, reliable, and standardized long-run company-level data for Europe.

This dataverse currently has no dataverses, datasets, or files. You can add to it by using the Add Data button on this page.

2.1 Citation metadata

First, you will have to add some basic descriptive metadata. Required fields are indicated with an asterisk. Hover your cursor over the question mark next to the field title for a description of its contents. If necessary, fields can be repeated with the **+** button (for instance, in case of multiple authors). These metadata are used to identify the dataset and its author(s). The metadata you provide will be used, for instance, for generating a persistent identifier and a citation so you can be credited for your work.



Author *

Name *: Poukens, Johan

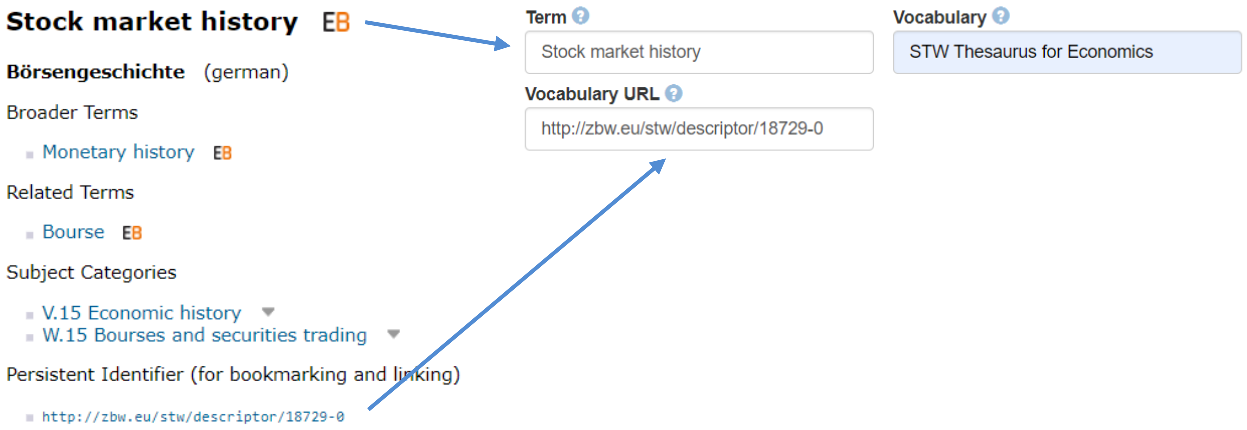
Affiliation: Universiteit Antwerpen

Identifier Scheme: ORCID

Identifier: 0000-0002-4663-9665

The Description (or Abstract) and Keywords you provide will help researchers find your dataset. We recommend using added entry terms (or descriptors) from the STW Thesaurus for Economics (http://www.zbw.eu/stw) as keywords.

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489



Stock market history EB

Börsengeschichte (german)

Broader Terms

- Monetary history EB

Related Terms

- Bourse EB

Subject Categories

- V.15 Economic history
- W.15 Bourses and securities trading

Persistent Identifier (for bookmarking and linking)

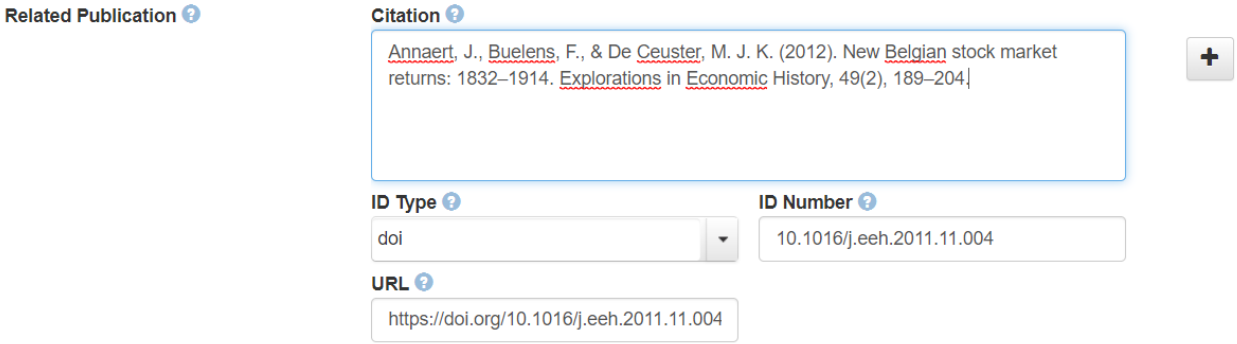
- http://zbw.eu/stw/descriptor/18729-0

Term: Stock market history

Vocabulary: STW Thesaurus for Economics

Vocabulary URL: http://zbw.eu/stw/descriptor/18729-0

You can also cite publications that use your dataset in the Related Publications field. Please use an established citation format, preferably the American Psychological Association (APA) style, for citing publications and provide a DOI (Digital Object Identifier) or other identifier and a URL of a website where the publication can be viewed if available.



Related Publication

Citation: Annaert, J., Buelens, F., & De Ceuster, M. J. K. (2012). New Belgian stock market returns: 1832–1914. Explorations in Economic History, 49(2), 189–204.

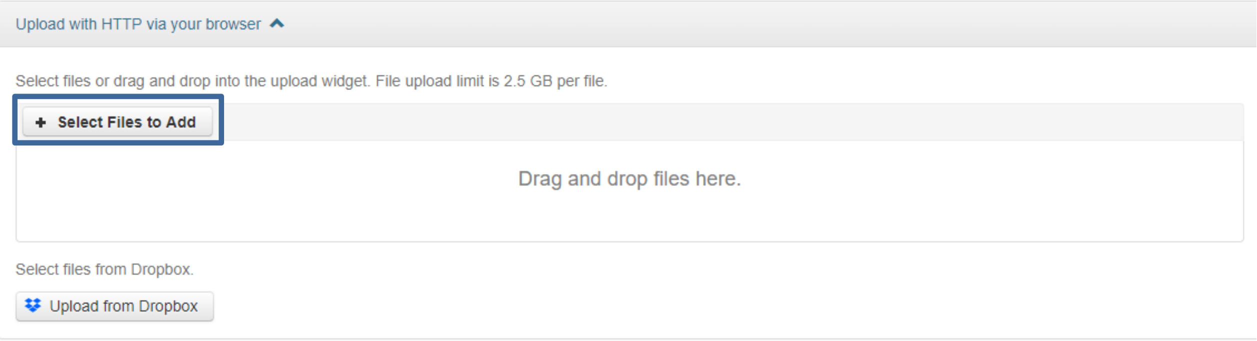
ID Type: doi

ID Number: 10.1016/j.eeh.2011.11.004

URL: https://doi.org/10.1016/j.eeh.2011.11.004

2.2 Upload files

Next, you will need to associate one or more files with your dataset. Scroll to the bottom of the screen and select or drag and drop files to upload. You can also upload files directly from Dropbox.



Upload with HTTP via your browser

Select files or drag and drop into the upload widget. File upload limit is 2.5 GB per file.

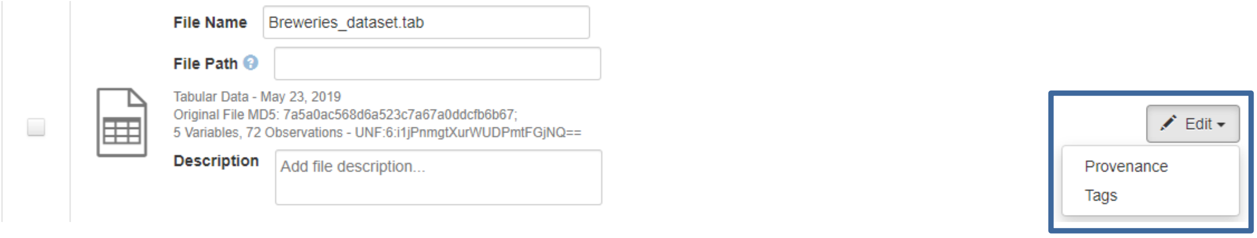
Drag and drop files here.

Select files from Dropbox.

In addition to the file(s) containing the data themselves (e.g. Excel, csv, SPSS, STATA, R, ...), you can also upload extra files (a Word document or PDF, for instance) with detailed descriptions of your data. Tags can

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

be used to differentiate data from documentation. To add tags, click the **Edit** bottom and select **Tags** after uploading your file. Also, you might notice that the extension of the file you just uploaded has changed to .tab. This is because Dataverse saves tabular data in an application-independent format for archival purposes. You can, however, still download the file in its original format.



File Name: Breweries_dataset.tab

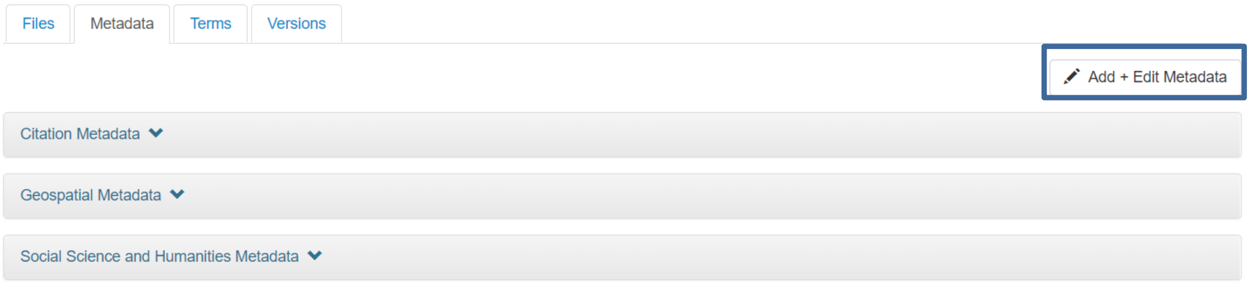
File Path:

Tabular Data - May 23, 2019
Original File MD5: 7a5a0ac568d6a523c7a67a0ddcfb6b67;
5 Variables, 72 Observations - UNF:6:i1jPnmgtXurWUDPmtFGjNQ==

Description: Add file description...

2.3 Additional metadata

After you are done with basic citation metadata and uploading, you can document your dataset in more detail. Navigate to the Metadata-tab. Here, you will see three metadata blocks or sections. Click on the **Add+Edit Metadata** button to start adding additional metadata.

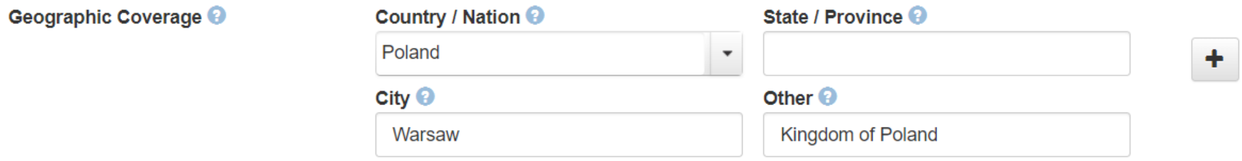


2.3.1 Citation metadata

In the citation metadata section, you can edit all of the metadata you added earlier and add some extra metadata. You can add, for instance, the language(s) of your dataset, the names of people or organisations who contributed to the collection of the data and funding information. Particular attention needs to be paid to the time period(s) covered by your dataset and the sources of the data. You should at least add bibliographic references for all sources from which data were taken (for instance, official stock exchange price lists). Use an established citation format, preferably APA, and include only one reference per field (i.e. repeat the field as many times as there are sources). Background information and an assessment of the quality of the sources can be included in the Origin of Sources and Characteristic of Sources fields, respectively. If your data are part of a series, information on the series (including, for instance, volume and issue) can be included in the Series Information field.

2.3.2 Geospatial metadata

Under geospatial metadata, you can add the geographic coverage of your metadata. Only present-day countries can be selected in the Country / Nation field, but Historical country names can be included in the Other field.



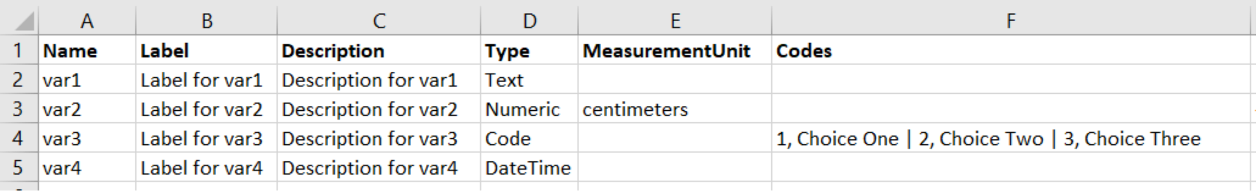
Geographic Coverage	Country / Nation: Poland	State / Province:
	City: Warsaw	Other: Kingdom of Poland

2.3.3 Social Science and Humanities Metadata

The social science and humanities metadata section allows you to document general characteristics of the data. The Universe field must be used for a description of the group of persons or other elements (companies, securities, ...) that are the object of the dataset and to which the data refer. If your data contains securities prices, for instance, be sure to include information on the stock exchange or market from which prices were collected and on the types and classes of securities concerned (e.g. all shares and bonds, ordinary shares only, ...) in the Universe field. If the data includes only a sample of units from this population (e.g. 30 companies with the largest market capitalisation), you can add details about the sampling methods in the Sampling Procedure field.

2.4 Variable metadata

In order to match corresponding data from the uploaded files to the data existing in the database, we also need a detailed description of the variables (i.e. columns in your dataset). Please use the Excel template (Variable_metadata_template.xlsx) for this purpose.



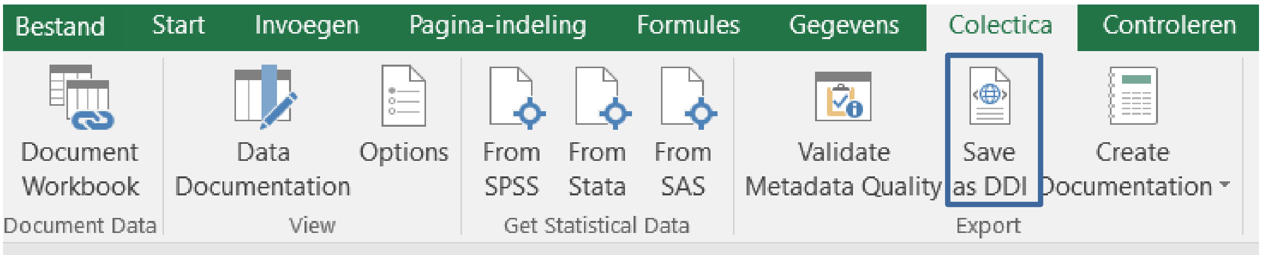
	A	B	C	D	E	F
1	**Name**	**Label**	**Description**	**Type**	**MeasurementUnit**	**Codes**
2	var1	Label for var1	Description for var1	Text		
3	var2	Label for var2	Description for var2	Numeric	centimeters	
4	var3	Label for var3	Description for var3	Code		1, Choice One | 2, Choice Two | 3, Choice Three
5	var4	Label for var4	Description for var4	DateTime		

Create a new row for each variable and include the following details:

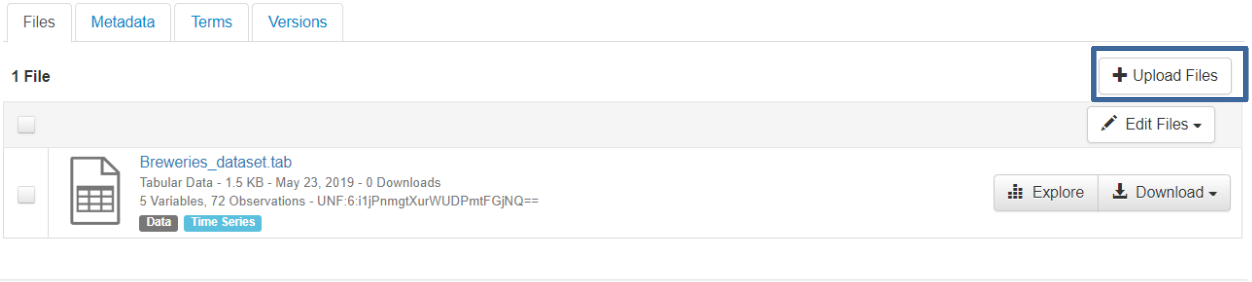
- Name: The header or title of the column as it exactly appears in the dataset
- Label: A short description of the variable
- Description: Additional details on the variable, if necessary
- Type: The data type of the variable. Allowed values are *Text*, *Numeric*, *Code*, and *DateTime*
- Measurement Unit: For Numeric variables, include the unit of the data (for instance, the currency in which prices are reported)
- Codes: For Code variables, specify the codes and value labels. The contents of this cell has to be formatted as follows: Code 1, Value label 1 | Code 2, Value label 2 | ... (for instance: 1, common share | 2, preferred share | 3, bond).

After you have completed all of the details, save and upload the Excel file to the EURHISFIRM Dataverse.

Alternatively, you can also use a software package to create DDI-metadata for your dataset and upload the XML-file. For documenting spreadsheet data, we recommend Colectica for Excel. Colectica for Excel can be downloaded for free (https://www.colectica.com/software/colecticaforexcel/). A manual is available online (https://docs.colectica.com/excel/document-data/document-excel-workbook). You can save your metadata as a DDI XML-file by clicking Save as DDI on the Colectica-ribbon.



You can upload the Excel or XML file with your Variable metadata by navigating to the Files tab and clicking the **+ Upload Files** button.



Breweries_dataset.tab
Tabular Data - 1.5 KB - May 23, 2019 - 0 Downloads
5 Variables, 72 Observations - UNF:6:i1jPnmgtXurWUDPmtFGjNQ==

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

3 References

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

4 Appendices

4.1 Appendix 1: Dataverse metadata elements

All available Dataverse metadata fields are included in the lists below, but elements in ~~strikeout~~ are hidden in the EURHISFIRM Dataverse. Grey fields are required.

4.1.1 Citation metadata

Name	**Title**	**Description**	**Type**
title	Title	Full title by which the Dataset is known.	text
subtitle	Subtitle	A secondary title used to amplify or state certain limitations on the main title.	text
alternativeTitle	Alternative Title	A title by which the work is commonly referred, or an abbreviation of the title.	text
alternativeURL	Alternative URL	A URL where the dataset can be viewed, such as a personal or project website.	url
otherId	Other ID	Another unique identifier that identifies this Dataset (e.g., producer's or another repository's number).	none
otherIdAgency	Agency	Name of agency which generated this identifier.	text
otherIdValue	Identifier	Other identifier that corresponds to this Dataset.	text
author	Author	The person(s), corporate body(ies), or agency(ies) responsible for creating the work.	none
authorName	Name	The author's Family Name, Given Name or the name of the organization responsible for this Dataset.	text
authorAffiliation	Affiliation	The organization with which the author is affiliated.	text
authorIdentifierScheme	Identifier Scheme	Name of the identifier scheme (ORCID, ISNI).	text
authorIdentifier	Identifier	Uniquely identifies an individual author or organization, according to various schemes.	text
DatasetContact	Contact	The contact(s) for this Dataset.	none
DatasetContactName	Name	The contact's Family Name, Given Name or the name of the organization.	text

This project has received funding from
the European Union’s Horizon 2020 research and innovation programme
under grant agreement N° 777489

DatasetContactAffiliation	Affiliation	The organization with which the contact is affiliated.	text
DatasetContactEmail	E-mail	The e-mail address(es) of the contact(s) for the Dataset. This will not be displayed.	email
dsDescription	Description	A summary describing the purpose, nature, and scope of the Dataset.	none
dsDescriptionValue	Text	A summary describing the purpose, nature, and scope of the Dataset.	textbox
dsDescriptionDate	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.	date
subject	Subject	Domain-specific Subject Categories that are topically relevant to the Dataset.	text
keyword	Keyword	Key terms that describe important aspects of the Dataset.	none
keywordValue	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. 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.	text
keywordVocabulary	Vocabulary	For the specification of the keyword controlled vocabulary in use, such as LCSH, MeSH, or others.	text
keywordVocabularyURI	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.	url
~~topicClassification~~	~~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.~~	~~none~~
~~topicClassValue~~	~~Term~~	~~Topic or Subject term that is relevant to this Dataset.~~	~~text~~

~~topicClassVocab~~	~~Vocabulary~~	~~Provided for specification of the controlled vocabulary in use, e.g., LCSH, MeSH, etc.~~	~~text~~
~~topicClassVocabURI~~	~~Vocabulary URL~~	~~Specifies the URL location for the full controlled vocabulary.~~	~~url~~
publication	Related Publication	Publications that use the data from this Dataset.	none
publicationCitation	Citation	The full bibliographic citation for this related publication.	textbox
publicationIDType	ID Type	The type of digital identifier used for this publication (e.g., Digital Object Identifier (DOI)).	text
publicationIDNumber	ID Number	The identifier for the selected ID type.	text
publicationURL	URL	Link to the publication web page (e.g., journal article page, archive record page, or other).	url
notesText	Notes	Additional important information about the Dataset.	textbox
language	Language	Language of the Dataset	text
~~producer~~	~~Producer~~	~~Person or organization with the financial or administrative responsibility over this Dataset~~	~~none~~
~~producerName~~	~~Name~~	~~Producer name~~	~~text~~
~~producerAffiliation~~	~~Affiliation~~	~~The organization with which the producer is affiliated.~~	~~text~~
~~producerAbbreviation~~	~~Abbreviation~~	~~The abbreviation by which the producer is commonly known. (ex. IQSS, ICPSR)~~	~~text~~
~~producerURL~~	~~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.~~	~~url~~
~~producerLogoURL~~	~~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.~~	~~url~~
~~productionDate~~	~~Production Date~~	~~Date when the data collection or other materials were produced (not distributed, published or archived).~~	~~date~~
~~productionPlace~~	~~Production Place~~	~~The location where the data collection and any other related materials were produced.~~	~~text~~
contributor	Contributor	The organization or person responsible for either collecting, managing, or	none

		otherwise contributing in some form to the development of the resource.	
contributorType	Type	The type of contributor of the resource.	text
contributorName	Name	The Family Name, Given Name or organization name of the contributor.	text
grantNumber	Grant Information	Grant Information	none
grantNumberAgency	Grant Agency	Grant Number Agency	text
grantNumberValue	Grant Number	The grant or contract number of the project that sponsored the effort.	text
~~distributor~~	~~Distributor~~	~~The organization designated by the author or producer to generate copies of the particular work including any necessary editions or revisions.~~	~~none~~
~~distributorName~~	~~Name~~	~~Distributor name~~	~~text~~
~~distributorAffiliation~~	~~Affiliation~~	~~The organization with which the distributor contact is affiliated.~~	~~text~~
~~distributorAbbreviation~~	~~Abbreviation~~	~~The abbreviation by which this distributor is commonly known (e.g., IQSS, ICPSR).~~	~~text~~
~~distributorURL~~	~~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.~~	~~url~~
~~distributorLogoURL~~	~~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.~~	~~url~~
~~distributionDate~~	~~Distribution Date~~	~~Date that the work was made available for distribution/presentation.~~	~~date~~
depositor	Depositor	The person (Family Name, Given Name) or the name of the organization that deposited this Dataset to the repository.	text
dateOfDeposit	Deposit Date	Date that the Dataset was deposited into the repository.	date
timePeriodCovered	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.	none
timePeriodCoveredStart	Start	Start date which reflects the time period covered by the data, not the dates of coding or making documents machine-	date

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

		readable or the dates the data were collected.	
timePeriodCoveredEnd	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
~~dateOfCollection~~	~~Date of Collection~~	~~Contains the date(s) when the data were collected.~~	~~none~~
~~dateOfCollectionStart~~	~~Start~~	~~Date when the data collection started.~~	~~date~~
~~dateOfCollectionEnd~~	~~End~~	~~Date when the data collection ended.~~	~~date~~
~~kindOfData~~	~~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.~~	~~text~~
series	Series	Information about the Dataset series.	none
seriesName	Name	Name of the dataset series to which the Dataset belongs.	text
seriesInformation	Information	History of the series and summary of those features that apply to the series as a whole.	textbox
software	Software	Information about the software used to generate the Dataset.	none
softwareName	Name	Name of software used to generate the Dataset.	text
softwareVersion	Version	Version of the software used to generate the Dataset.	text
~~relatedMaterial~~	~~Related Material~~	~~Any material related to this Dataset.~~	~~textbox~~
~~relatedDatasets~~	~~Related Datasets~~	~~Any Datasets that are related to this Dataset, such as previous research on this subject.~~	~~textbox~~
~~otherReferences~~	~~Other References~~	~~Any references that would serve as background or supporting material to this Dataset.~~	~~text~~
dataSources	Data Sources	List of books, articles, serials, or machine-readable data files that served as the sources of the data collection.	textbox
originOfSources	Origin of Sources	For historical materials, information about the origin of the sources and the	textbox

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

		rules followed in establishing the sources should be specified.	
characteristicOfSources	Characteristic of Sources Noted	Assessment of characteristics and source material.	textbox
~~accessToSources~~	~~Documentation and Access to Sources~~	~~Level of documentation of the original sources.~~	~~textbox~~

4.1.2 Geospatial Metadata

Name	Title	Description	Type
geographicCoverage	Geographic Coverage	Information on the geographic coverage of the data. Includes the total geographic scope of the data.	none
country	Country / Nation	The country or nation that the Dataset is about.	text
state	State / Province	The state or province that the Dataset is about. Use GeoNames for correct spelling and avoid abbreviations.	text
city	City	The name of the city that the Dataset is about. Use GeoNames for correct spelling and avoid abbreviations.	text
otherGeographicCoverage	Other	Other information on the geographic coverage of the data.	text
~~geographicUnit~~	~~Geographic Unit~~	~~Lowest level of geographic aggregation covered by the Dataset, e.g., village, county, region.~~	~~text~~
~~geographicBoundingBox~~	~~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.~~	~~none~~
~~westLongitude~~	~~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.~~	~~text~~
~~eastLongitude~~	~~East Longitude~~	~~Easternmost coordinate delimiting the geographic extent of the Dataset. A valid range of values, expressed in decimal~~	~~text~~

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

		~~degrees, is -180,0 <= East Bounding Longitude Value <= 180,0.~~	
~~northLongitude~~	~~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.~~	~~text~~
~~southLongitude~~	~~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.~~	~~text~~

4.1.3 Social Science and Humanities Metadata

Name	**Title**	**Description**	**Type**
unitOfAnalysis	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.	textbox
universe	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. 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.	textbox
timeMethod	Time Method	The time method or time dimension of the data collection, such as panel, cross-sectional, trend, time-series, or other.	text
~~dataCollector~~	~~a Collector~~	~~Individual, agency or organization responsible for administering the questionnaire or interview or compiling the data.~~	~~text~~
~~collectorTraining~~	~~Collector Training~~	~~Type of training provided to the data collector~~	~~text~~
frequencyOfDataCollection	Frequency	If the data collected includes more than one point in time, indicate the frequency	text

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

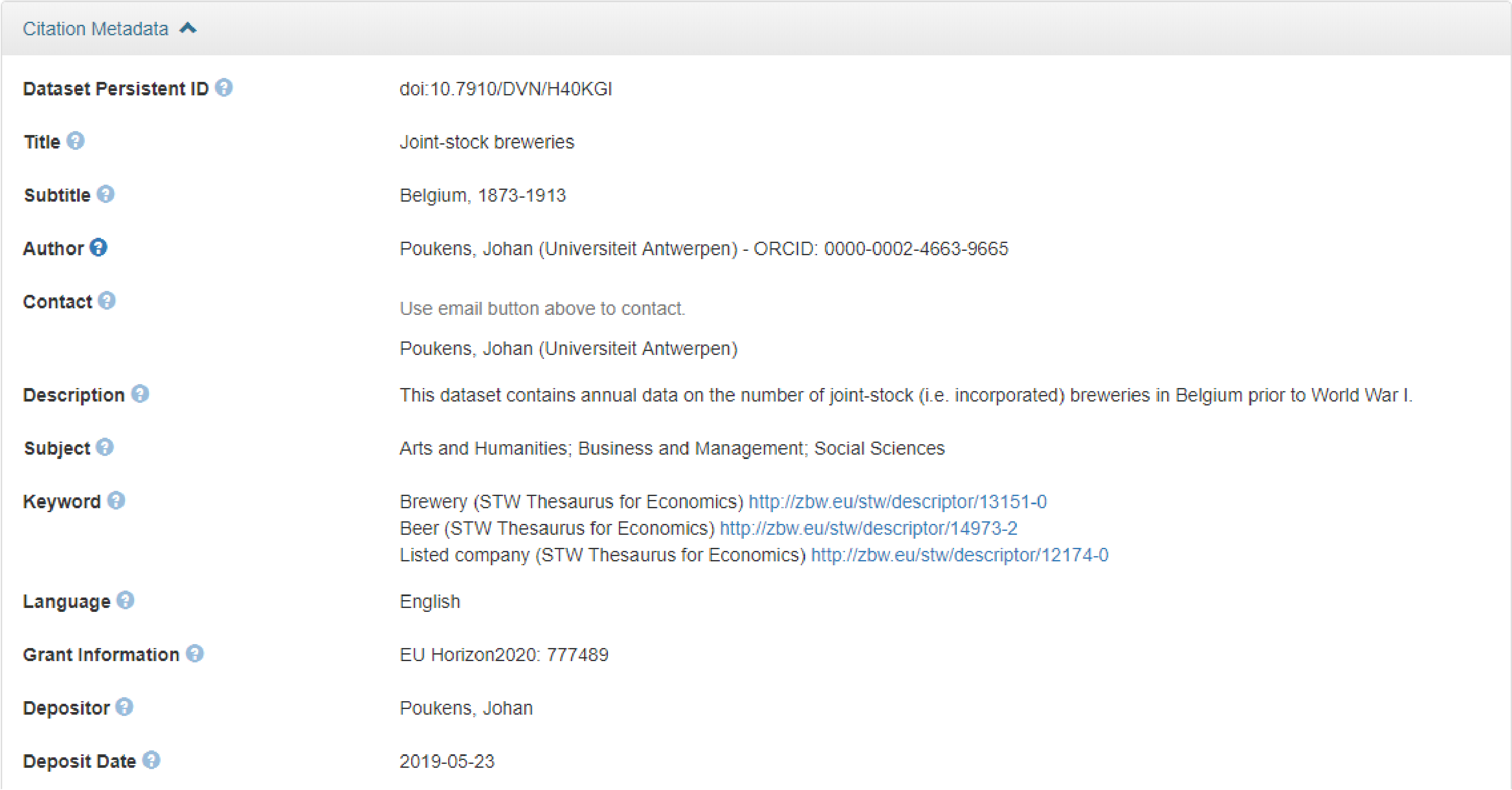
		with which the data was collected; that is, monthly, quarterly, or other.	
samplingProcedure	Sampling Procedure	Type of sample and sample design used to select the survey respondents to represent the population. May include reference to the target sample size and the sampling fraction.	textbox
~~targetSampleSize~~	~~Target Sample Size~~	~~Specific information regarding the target sample size, actual sample size, and the formula used to determine this.~~	~~none~~
~~targetSampleActualSize~~	~~Actual~~	~~Actual sample size.~~	~~int~~
~~targetSampleSizeFormula~~	~~Formula~~	~~Formula used to determine target sample size.~~	~~text~~
~~deviationsFromSampleDesign~~	~~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.~~	~~text~~
collectionMode	Collection Mode	Method used to collect the data; instrumentation characteristics (e.g. telephone interview, mail questionnaire, or other).	textbox
~~researchInstrument~~	~~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.~~	~~text~~
~~dataCollectionSituation~~	~~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.~~	~~textbox~~
actionsToMinimizeLoss	Actions to Minimize Losses	Summary of actions taken to minimize data loss. Includes information on actions such as follow-up visits, supervisory checks, historical matching, estimation, and so on.	text
controlOperations	Control Operations	Methods to facilitate data control performed by the primary investigator or by the data archive.	text

This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

~~weighting~~	~~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.~~	~~textbox~~
cleaningOperations	Cleaning Operations	Methods used to clean the data collection, such as consistency checking, wildcode checking, or other.	text
~~datasetLevelErrorNotes~~	~~Study Level Error Notes~~	~~Note element used for any information annotating or clarifying the methodology and processing of the study.~~	~~text~~
~~responseRate~~	~~Response Rate~~	~~Percentage of sample members who provided information.~~	~~textbox~~
~~samplingErrorEstimates~~	~~Estimates of Sampling Error~~	~~Measure of how precisely one can estimate a population value from a given sample.~~	~~text~~
~~otherDataAppraisal~~	~~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.~~	~~text~~
socialScienceNotes	Notes	General notes about this Dataset.	none
socialScienceNotesType	Type	Type of note.	text
socialScienceNotesSubject	Subject	Note subject.	text
socialScienceNotesText	Text	Text for this note.	textbox

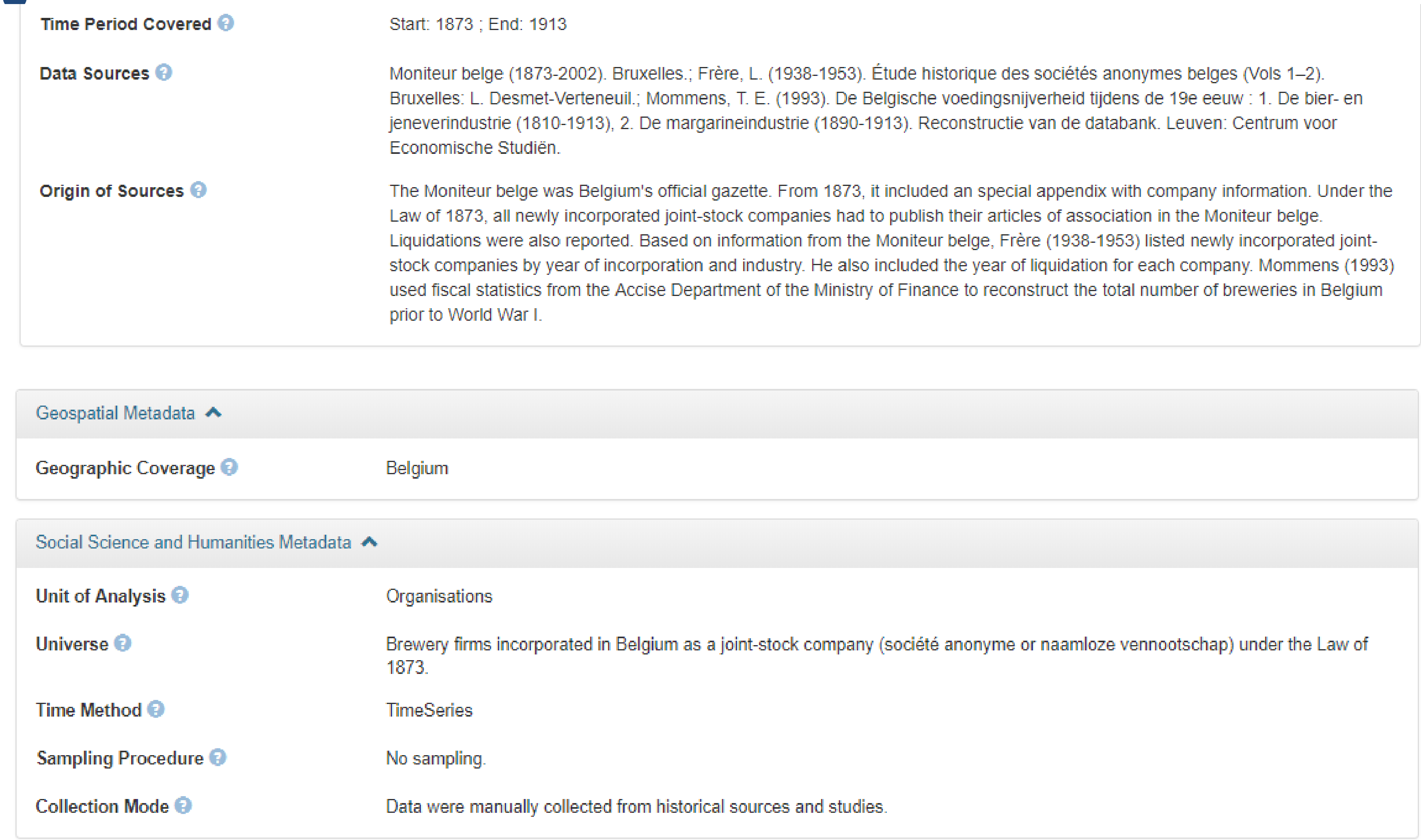
This project has received funding from
the European Union's Horizon 2020 research and innovation programme
under grant agreement N° 777489

4.2 Appendix 2: Example



Citation Metadata

Dataset Persistent ID	doi:10.7910/DVN/H40KGI
Title	Joint-stock breweries
Subtitle	Belgium, 1873-1913
Author	Poukens, Johan (Universiteit Antwerpen) - ORCID: 0000-0002-4663-9665
Contact	Use email button above to contact. Poukens, Johan (Universiteit Antwerpen)
Description	This dataset contains annual data on the number of joint-stock (i.e. incorporated) breweries in Belgium prior to World War I.
Subject	Arts and Humanities; Business and Management; Social Sciences
Keyword	Brewery (STW Thesaurus for Economics) http://zbw.eu/stw/descriptor/13151-0 Beer (STW Thesaurus for Economics) http://zbw.eu/stw/descriptor/14973-2 Listed company (STW Thesaurus for Economics) http://zbw.eu/stw/descriptor/12174-0
Language	English
Grant Information	EU Horizon2020: 777489
Depositor	Poukens, Johan
Deposit Date	2019-05-23



Time Period Covered	Start: 1873 ; End: 1913
Data Sources	Moniteur belge (1873-2002). Bruxelles.; Frère, L. (1938-1953). Étude historique des sociétés anonymes belges (Vols 1–2). Bruxelles: L. Desmet-Verteneuil.; Mommens, T. E. (1993). De Belgische voedingsnijverheid tijdens de 19e eeuw : 1. De bier- en jeneverindustrie (1810-1913), 2. De margarineindustrie (1890-1913). Reconstructie van de databank. Leuven: Centrum voor Economische Studiën.
Origin of Sources	The Moniteur belge was Belgium's official gazette. From 1873, it included an special appendix with company information. Under the Law of 1873, all newly incorporated joint-stock companies had to publish their articles of association in the Moniteur belge. Liquidations were also reported. Based on information from the Moniteur belge, Frère (1938-1953) listed newly incorporated joint-stock companies by year of incorporation and industry. He also included the year of liquidation for each company. Mommens (1993) used fiscal statistics from the Accise Department of the Ministry of Finance to reconstruct the total number of breweries in Belgium prior to World War I.

Geospatial Metadata

Geographic Coverage	Belgium

Social Science and Humanities Metadata

Unit of Analysis	Organisations
Universe	Brewery firms incorporated in Belgium as a joint-stock company (société anonyme or naamloze vennootschap) under the Law of 1873.
Time Method	TimeSeries
Sampling Procedure	No sampling.
Collection Mode	Data were manually collected from historical sources and studies.