

## METADATA AND TERMS GUIDE

### Examples for depositing data in SODHA

#### Introduction

Providing clear metadata and terms of access and use are vital aspects of depositing datasets in the Social Sciences and Digital Humanities Archive (SODHA). Accurate and complete information about your dataset help to make it FAIR: Findable, Accessible, Interoperable and Reusable. In addition, including this information from the start may save you a lot of time and effort in the long run. Reconstructing the context and modalities of a dataset later on is no easy (or fun) assignment.

The many possibilities offered in SODHA can seem a bit overwhelming. However, most options are non-mandatory. You decide how much information is provided. In other words, you are free to choose the level of detail, and consequently the degree of FAIRness of your dataset.

This Metadata and Terms Guide is designed to help you describe your dataset. We include the explanations from SODHA (grey italics), additional notes (●) and examples (℞). The examples are mainly fictional and can be tailored to suit your dataset. Please note that this list of examples is not exhaustive. Many other entries are possible to describe your dataset.

The table of content has been equipped with hyperlinks. You can use this table to navigate through the options if you have a question about a specific subsection of the SODHA metadata or terms section. Any additional questions or remarks can be directed at [sodha@arch.be](mailto:sodha@arch.be).

## Table of content

### Terms 4

#### *Terms of Use 4*

Waiver 4

Terms of Use 4

- Open Access- CC-0 4
- Open Access- CC-BY 4
- Restricted Access- CC-BY 5
- Open Access- Scientific Purposes Only 5
- Restricted Access- Scientific Purposes Only 7
- Other conditions? 7

Confidentiality Declaration 8

Special Permissions 8

Restrictions 8

Citation Requirements 9

Depositor Requirements 9

Conditions 9

Disclaimer 9

#### *Restricted Files + Terms of Access 10*

Terms of Access 10

Request Access 10

Data Access Place 11

Original Archive 11

Availability Status 11

Contact for Access 11

Size of Collection 12

Study Completion 12

### Metadata 12

#### *Citation Metadata 12*

Title of Dataset 12

Subtitle 13

Alternative Title 13

Alternative URL 13

Other ID 13

Author 14

Contact 14

Description 14

Subject 15

Keyword 16

Topic Classification 16

Related Publication 17

Notes 17

Language 17

Producer 17

Production Date 18

Production Place 18

Grant Information 18

Time Period Covered 18

Date of Collection 19

Kind of Data	19
Series	19
Software Used to Create the Dataset	19
Related Material	20
Related Datasets	20
Other References	20
Data Sources	20
Origin of Sources	21
Characteristic of Sources Noted	21
Documentation and Access to Sources	21
<i>Geospatial Metadata</i>	<i>21</i>
Geographic Coverage	21
Geographic Unit	22
<i>Social Science and Humanities Metadata</i>	<i>22</i>
Unit of Analysis	22
Universe	23
Time Method	23
Data Collector	23
Collector Training	24
Frequency	24
Number of Variables	24
Sampling Procedure	24
Target Sample Size / Number of Units	25
Major Deviations for Sample Design	25
Collection Mode	25
Type of Research Instrument	25
Characteristics of Data Collection Situation	26
Actions to Minimize Losses	26
Control Operations	26
Weighting	26
Cleaning Operations	27
Study Level Error Notes	27
Response Rate	27
Estimates of Sampling Error	28
Other Forms of Data Appraisal	28
Notes	28
<i>Historical Metadata</i>	<i>28</i>
History of the Data	28
Biography of the Investigator/Contributor	29
History of the Research Center	29
History of the Grant Agency	29

Back to [Table of content](#)

## Terms

### *Terms of Use*

In this section depositors can stipulate under which conditions their deposited datasets are reusable.

### Waiver

*The waiver informs data downloaders how they can use this dataset.*

*Datasets will default to a [CC0 public domain dedication](#). CC0 facilitates reuse and extensibility of research data. Our [Community Norms](#) as well as good scientific practices expect that proper credit is given via citation. If you are unable to give datasets a CC0 waiver you may enter custom Terms of Use for datasets.*

Yes, apply CC0 - "Public Domain Dedication"

No, do not apply CC0 - "Public Domain Dedication"

### Terms of Use

*Outlines how this data can be used once downloaded.*

*If you are unable to use CC0 for datasets you are able to set custom terms of use. Here is an example of a [Data Usage Agreement](#) for datasets that have de-identified human subject data.*

- **Open Access- CC-0**

- Select "Yes, apply CC0- "Public Domain Dedication"" in the Waiver section (See above).
- Leave this Terms of Use section blank.
- More information: <https://creativecommons.org/publicdomain/zero/1.0/>

- **Open Access- CC-BY**

- Select "No, do not apply CC0 - "Public Domain Dedication"" in the Waiver section (See above).
- Paste following statement in this Terms of Use section:  

```
<a rel="license" href="http://creativecommons.org/licenses/by/4.0/"></a><br />This work is licensed under a <a rel="license" href="http://creativecommons.org/licenses/by/4.0/">Creative Commons Attribution 4.0 International License</a>.
```
- More information: <https://creativecommons.org/licenses/by/4.0/>

Back to [Table of content](#)

- **Restricted Access- CC-BY**

- The same procedure as in Open Access- CC-BY (see above).
- When uploading files: restrict the access for all or selected files using the “edit” drop down menu. A pop-up menu will appear where you can state the terms of access and enable access requests. You can check this in the metadata section.

- **Open Access- Scientific Purposes Only**

- Select “No, do not apply CC0 - "Public Domain Dedication"” in the Waiver section (See above).
- Paste following statement in this Terms of Use section:

*<b>Attribution – Scientific Purposes Only License</b><br><br>*

*By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Attribution-Scientific purpose only International License ("License"). To the extent this License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.<br>*

*<br>*

*Section 1 – Definitions<br>*

*a. Adapted Material means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.<br>*

*b. Adapter's License means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this License.<br>*

*c. Copyright and Similar Rights means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.<br>*

*d. Effective Technological Measures means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.<br>*

*e. Exceptions and Limitations means fair use, fair dealing, and/or any other exception or limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.<br>*

*f. Licensed Material means the artistic or literary work, database, or other material to which the Licensor applied this License.<br>*

*g. Licensed Rights means the rights granted to You subject to the terms and conditions of this License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.<br>*

*h. Licensor means the individual(s) or entity(ies) granting rights under this License.<br>*

*i. Scientific purpose means not primarily intended for or directed towards commercial advantage or monetary compensation and re-use of the Licensed Material is only possible for scientific purposes. <br>*

*j. Share means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.<br>*

*k. Sui Generis Database Rights means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.<br>*

*l. You means the individual or entity exercising the Licensed Rights under this License. Your has a corresponding meaning.<br>*

*<br>*

*Section 2 – Scope<br>*

*a. License grant<br>*

*1. Subject to the terms and conditions of this License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to: <br>*

*A. reproduce and Share the Licensed Material, in whole or in part, for Scientific purposes only; and<br>*

Back to [Table of content](#)

B. produce, reproduce, and Share Adapted Material for Scientific purposes only.<br>

2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this License does not apply, and You do not need to comply with its terms and conditions.<br>

3. Term. The term of this License is specified in Section 6(a).<br>

4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures. For purposes of this License, simply making modifications authorized by this Section 2(a)(4) never produces Adapted Material.<br>

5. Downstream recipients<br>

A. Offer from the Licensor – Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this License.<br>

B. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.<br>

6. No endorsement. Nothing in this License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).<br>

b. Other rights<br>

1. Moral rights, such as the right of integrity, are not licensed under this License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.<br>

2. Patent and trademark rights are not licensed under this License.<br>

3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties, including when the Licensed Material is used other than for Scientific purposes.<br>

<br>

Section 3 – License Conditions<br>

Your exercise of the Licensed Rights is expressly made subject to the following conditions.<br>

a. Attribution<br>

1. If You Share the Licensed Material (including in modified form), You must:<br>

A. retain the following if it is supplied by the Licensor with the Licensed Material:<br>

i. identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);<br>

ii. a copyright notice;<br>

iii. a notice that refers to this License;<br>

iv. a notice that refers to the disclaimer of warranties;<br>

v. a URI or hyperlink to the Licensed Material to the extent reasonably practicable;<br>

B. indicate if You modified the Licensed Material and retain an indication of any previous modifications; and

C. indicate the Licensed Material is licensed under this License, and include the text of, or the URI or hyperlink to, this License.<br>

2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.<br>

3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.<br>

4. If You Share Adapted Material You produce, the Adapter's License You apply must not prevent recipients of the Adapted Material from complying with this License.<br>

<br>

Section 4 – Sui Generis Database Rights<br>

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:<br>

a. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database for NonCommercial purposes only;<br>

b. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material; and<br>

c. You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.<br>

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this License where the Licensed Rights include other Copyright and Similar Rights.<br>

<br>

Section 5 – Disclaimer of Warranties and Limitation of Liability<br>

a. Unless otherwise separately undertaken by the Licensor, to the extent possible, the Licensor offers the Licensed Material as-is and as-available, and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors, whether or not known or

Back to [Table of content](#)

discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.<br>

b. To the extent possible, in no event will the Licensor be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this License or use of the Licensed Material, even if the Licensor has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.<br>

c. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.<br>

<br>

Section 6 – Term and Termination<br>

a. This License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this License, then Your rights under this License terminate automatically.<br>

b. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:<br>

1. automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or<br>

2. upon express reinstatement by the Licensor.<br>

For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this License.<br>

c. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this License.<br>

d. Sections 1, 5, 6, 7, and 8 survive termination of this License.<br>

<br>

Section 7 – Other Terms and Conditions<br>

a. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.<br>

b. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this License.<br>

<br>

Section 8 – Interpretation<br>

a. For the avoidance of doubt, this License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully be made without permission under this License.<br>

b. To the extent possible, if any provision of this License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this License without affecting the enforceability of the remaining terms and conditions.<br>

c. No term or condition of this License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.<br>

d. Nothing in this License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.<br>

More information: The policy regarding [Access & Reuse in SODHA](#) .

## • **Restricted Access- Scientific Purposes Only**

- The same procedure as in Open Access- scientific purposes only (see above).
- When uploading files: restrict the access for all or selected files using the “edit” drop down menu. A pop-up menu will appear where you can state the terms of access and enable access requests. You can check this in the metadata section.

## • **Other conditions?**

Contact [sodha@arch.be](mailto:sodha@arch.be) if you want to apply other conditions on the reuse of your data.

Back to [Table of content](#)

## Confidentiality Declaration

*Indicates whether signing a confidentiality agreement is needed to access a resource.*

For example:

- The data can only be accessed after signing a confidentiality agreement. The contact person will send the confidentiality agreement after your request for access.
- No confidentiality declaration is needed to access the dataset.

## Special Permissions

*Determine if any special permissions are required to access a resource (e.g. If a form is needed and where to access the form).*

For example:

- Access can only be granted after assessment of the research institution's Evaluation Form including questions on the researcher's profile, the research objectives, the requested access period and the data security policy. The contact person will send the Evaluation Form after your request for access.
- Access requires the formal authorization of all persons handling the data by the depositor. Please include the contact information of all (potential) data users and their affiliations with your request for access.

## Restrictions

*Any restrictions on the use of the collection, such as privacy certification or distribution restrictions, should be indicated here. These can be restrictions applied by the author, producer, or disseminator of the data collection. If the data are restricted to only a certain class of users, please specify which type.*

- If you would like that data users fulfil specific conditions related to the data use, please state so here. Be advised that SODHA will not actively check whether data users meet these conditions.

For example:

- The requested access and use period should be proportional to the research objectives. Access to the dataset is restricted until the user and depositor stipulate an appropriate access and use period in mutual agreement. After this period, the accessed dataset and derivative datasets need to be deleted.
- Users do NOT have permission to share the accessed dataset or related resources with other persons or organizations. All persons handling the data need to be authorized by the depositor or the contact person.



Back to [Table of content](#)

## Citation Requirements

*Include special/explicit citation requirements for data to be cited properly in articles or other publications that are based on analysis of the data. For standard data citation refer to our Community Norms*

For example:

- Please refer to our data in accordance to the SODHA community Norms: Author(s). Distribution Year. "Dataset title". [https://doi.org/...](https://doi.org/) SODHA: Social Sciences and Digital Humanities Archive

## Depositor Requirements

*Information regarding user responsibility for informing dataset depositors, authors or curators of their use of data through providing citations to the published work or providing copies of manuscripts.*

- If you would like that data users fulfil other conditions (not specifically related to the data use), please state so here. Be advised that SODHA will not actively check whether data users meet these conditions.

For example:

- A list of publications building on the shared dataset needs to be sent to the contact person after completion of the research project. Please keep the contact person informed about any updates regarding these or subsequent publications.

## Conditions

*Any additional information that will assist the user in understanding the use conditions of the dataset.*

- In this section, you can add more information on the context of the dataset and the conditions related to the data use.

For example:

- The dataset is partly built on data from a third party. Our terms of access and use follow the procedures stipulated by the third party.

## Disclaimer

*Information regarding responsibilities for uses of the dataset.*

Back to [Table of content](#)

- Appropriate disclaimers and notices can be inserted here. Be advised that statements here cannot override the [SODHA deposit agreement](#), nor can they be used to deflect(inter)national legal responsibilities.

For example:

- ☞ Much care has been taken to ensure the accuracy and quality of the data and documentation. Nevertheless, the data depositor or the data producer does not guarantee the accuracy or topicality of the information provided.

### *Restricted Files + Terms of Access*

In this section depositors can stipulate under which conditions their deposited datasets are accessible.

#### **Terms of Access**

*Information on if and how users can gain access to the restricted file in this dataset.*

- Datasets will default to open access. Choosing open access, requires no further steps.
- If you prefer restricted access, you need to specify this:
  - When uploading data files, you can restrict the access for all or selected files (Edit -> drop down menu with the options Delete, Restrict, Unrestrict). A window will open where you can state the Terms of Access and enable the Request access.
  - The results of the actions in (a) will be visible in the Terms section. You can edit this information with additional information.

For example:

- ☞ A request for access and use can be sent to the contact person via the SODHA contact form.
- ☞ More information on the access and use procedure can be found on our organization's website:  
<https://www.yourorganizationswebsite.be/HowToAccessData>

#### **Request Access**

*If checked, users can request access to the restricted files in this dataset.*

- Only if this box is checked will SODHA allow potential data users to contact the designated contact person(s).

*Enable access request*

Back to [Table of content](#)

### Data Access Place

*If the data is not only in the Dataverse, list the location(s) where the data are currently stored.*

For example:

- ☞ The data can only be accessed at the offices of Organization/Archive/... (address).
- ☞ Storage facility: Organization/Archive/... (address).

### Original Archive

*Archive from which the data was obtained.*

For example:

- ☞ Organization/Archive/... (address).

### Availability Status

*Statement of dataset availability. A depositor may need to indicate that a Dataset is unavailable because it is embargoed for a period of time, because it has been superseded, because a new edition is imminent, etc.*

For example:

- ☞ This dataset is under embargo until ...
- ☞ We recommend users work with the updated version of this dataset available in SODHA: "Dataset title. Version".

### Contact for Access

*If different from the Dataset contact, this is the contact person or organization (include email or full address, and telephone number if available) that controls access to a collection.*

For example:

- ☞ Organization's information: email address, telephone number and full address.

Back to [Table of content](#)

## Size of Collection

*Summary of the number of files that exist in the dataset, recording the number of files that contain data and noting whether the collection contains machine readable documentation and/or other supplementary files and information, such as codes, data dictionaries, data definition statements, or data collection instruments.*

For example:

- The Dataset contains # files with a total volume of #GB.
- Documentation on the construction of variables is included (SPSS Syntax, Stata do.files, ... ).

## Study Completion

*Relationship of the data collected to the amount of data coded and stored in the dataset. Information as to why certain items of collected information were not included in the dataset or a specific data file should be provided.*

For example:

- All collected data are included in the dataset, but personal identifiers are not included to anonymize the data.
- The information offered in free text fields are not available in the dataset to guarantee the anonymity of the respondents.

## Metadata

*Citation Metadata* ☒

## Title of Dataset

*Full title by which the dataset is known.*

Some guidelines for picking a good title for your dataset:

- Insert a unique title that represents the focus of your dataset.
- Keep it clear and concise. Do your colleagues and friends understand what the dataset presents, even without any context or inside knowledge?
- The recommended length of the title is <120 characters.
- Specify the geographical and temporal context in the title if appropriate.
- Acronyms or abbreviations can be used in titles if appropriate. They should be spelled in full at first use and then abbreviated between brackets.
  - Belgian Archive for Social Sciences (BASS)
- If your dataset contains only one indicator, this should be included in the title.
  - Gross annual salary by occupational group, Belgium, 2010-2019.

Back to [Table of content](#)

- If your dataset contains several indicators, include the basic theme in the title.
  - ↳ Employment conditions by occupational group, Belgium, 2010-2019
- Do not simply repeat the article title to name your dataset. In case you are sharing your dataset for validation purposes, click on the “Add ‘Replication data for’ to the title” button.
  - ↳ Replication Data for: ‘The Article Title’

## Subtitle

*A secondary title to amplify or state certain limitations on the main title.*

For example:

- ↳ Pretest survey results [Title of the Dataset: ‘Preteaching effects on student performance in lower secondary school: evaluation of a natural experiment’]

## Alternative Title

*A title by which the work is commonly referred, or an abbreviation of the title.*

For example:

- ↳ BASS [Full title= Belgian Archive for Social Sciences]
- ↳ Belgisch Archief voor de Sociale Wetenschappen [Title of the Dataset= Belgian Archive for Social Sciences]
- ↳ Archive Belge pour les Sciences Sociales [Title of the Dataset = Belgian Archive for Social Sciences]

## Alternative URL

*A URL where the dataset can be viewed, such as a personal or project website.*

## Other ID

*Another unique identifier that identifies this dataset (e.g. Producer’s or another repository’s number).*

For example:

- ↳ Agency: Your university’s archive Identifier: Number\_A1

Back to [Table of content](#)

## Author

*The person(s), corporate body(ies) or agency(ies) responsible for creating the work.*

For example:

☞ Name: Family name, Given name  
Affiliation: University (of applied sciences)/Science institution/...  
Identifier scheme: Select ORCID/ISNI/LCNA/VIAF/GND  
Identifier: Enter your personal identifier

☞ Name: Organization/Research group/Working group/...  
Affiliation: University (of applied sciences)/Science institution/...  
Identifier scheme: Select ORCID/ISNI/LCNA/VIAF/GND  
Identifier: Enter your organizational identifier

## Contact

*The contact(s) for this dataset.*

- When you include your email address here, this will not be displayed (in contrast with the “Contact for Access” field in the Terms section). You will be reached automatically through the contact form or the “Request Access” form.

For example:

☞ Name: Family name, Given name  
Affiliation: University (of applied sciences)/Science institution/...  
E-mail: [Your.Name@InstitutionalAccount.be](mailto:Your.Name@InstitutionalAccount.be)

## Description

*A summary describing the purpose, nature and scope of the dataset.*

Some guidelines for writing a good abstract for your dataset:

- Describe the purpose, content and method of the dataset briefly.
- Try to answer following questions:
  - What is the purpose of this dataset?
  - To who do these data refer to? What is the study population and sample size?
  - What is the relevant geographical area? Does the dataset cover a number of cities, regions, or the whole country?
  - When were these data collected?
  - How were the data collected? What was the sampling procedures? At what level were the data collected: at the individual, household, etc. level? How was the survey administered (online, paper, etc.)? How

Back to [Table of content](#)

were interviews conducted (face-to-face, video chat, etc.)? In case the dataset relies on secondary data sources, which original sources were used?

- What form does the dataset have (person-level; person-period, etc.)?
- How complete is the dataset? Are any data missing?
- What are the selling points? Does the dataset fill a data gap? Which variables may be of interest to potential users?

For example:

➤ This expert survey aims to fill a fundamental data gap in country-level data on teacher training and support policies, specifically addressing (a) official entry requirements for teacher training programs; (b) in-service support policies; (c) salary information by level of experience of lower secondary school teachers (ISCED 2). The focus of this expert survey lies on the 48 countries and economies participating in the 2018 Teaching and Learning International Survey (TALIS). The questionnaire was hosted by Qualtrics and distributed online via personal invitations to 108 educational experts from the 48 TALIS-countries. Prime selection criteria for the educational experts were academic excellence established with regards to the national teacher policies and/or membership to leading research groups in the field. Data were collected between 13 December 2019- 1 February 2020. In the end, a person-level, anonymized dataset was created using the answers from 39 experts in 36 countries or economies. Information on official entry requirements for teacher training programs and in-service support policies is available for over 70% of the 36 participating countries. Salary information by level of experience was obtained for 62% of these 36 countries. The methods used to develop this expert survey are largely based on the methodology from the Education At a Glance report and the Eurydice reports. This approach was chosen to enable the comparability with these two authoritative data sources. To our knowledge, this is the only international overview containing salary information for lower secondary school teachers in Cyprus, the Republic of Korea, Russia, Saudi-Arabia, South Africa and the United Arab Emirates.

#### Date of Description:

➤ Year-Month-Day the description was written

### Subject

*Domain specific subject categories that are topically relevant to the dataset. Multiple options are possible.*

Social Sciences

Back to [Table of content](#)

## Keyword

*Key terms that describe specific aspects of the dataset that may enhance the findability of your dataset.*

Some guidelines for selecting appropriate key words:

- Select specific terms related to the focus of your dataset or specific indicators.
- Include no more than 3 words per keyword.
- Don't use words that are already in your title.
- Consider incorporating synonyms, related concepts or antonyms.
  - ↳ Attrition; Retention [Dataset on staff turnover]
- Consider including methodological specificities as keywords.
  - ↳ Longitudinal study; Panel study
- Use the controlled keyword vocabulary options to facilitate web searches:
  - ↳ Library of Congress Subject Headings (LCSH):  
<http://id.loc.gov/authorities/subjects.html>  
Keyword: Families--Longitudinal studies  
Vocabulary: LCSH-complex subject  
Vocabulary URL: <http://id.loc.gov/authorities/subjects/sh85047015>
  - ↳ Medical Subject Headings (MeSH) thesaurus:  
<https://www.nlm.nih.gov/mesh/meshhome.html>  
Keyword: Socioeconomic Factors  
Vocabulary: MeSH  
Vocabulary URL: <http://id.nlm.nih.gov/mesh/D012959>

## Topic Classification

*The classification field indicates the broad important topic(s) and subject(s) that the data cover. Please use the terms of the CESSDA Topic Classification here.*

- You can copy the already stated URL for the CESSDA Topic Classification and paste this into your internet browser. Once found the most appropriate keywords, you can copy-paste this information in the metadata field.

For example:

- ↳ CESSDA Topic Classification:  
<https://vocabularies.cessda.eu>  
Keyword: Demography.Migration  
Vocabulary: CESSDA Topic Classification:  
Vocabulary URL: <https://vocabularies.cessda.eu>



Back to [Table of content](#)

## Related Publication

*Publications that use the data from this dataset.*

Possibilities regarding the ID type:

- |                                  |                                 |                               |
|----------------------------------|---------------------------------|-------------------------------|
| <input type="checkbox"/> ark     | <input type="checkbox"/> handle | <input type="checkbox"/> pmid |
| <input type="checkbox"/> arXiv   | <input type="checkbox"/> isbn   | <input type="checkbox"/> purl |
| <input type="checkbox"/> bibcode | <input type="checkbox"/> issn   | <input type="checkbox"/> upc  |
| <input type="checkbox"/> doi     | <input type="checkbox"/> istc   | <input type="checkbox"/> url  |
| <input type="checkbox"/> ean13   | <input type="checkbox"/> lissn  | <input type="checkbox"/> urn  |
| <input type="checkbox"/> eissn   | <input type="checkbox"/> lsid   |                               |

For example:

☞ ID Type: doi

Citation: Authors, Year, "Title article", Journal, volume (issue), p.

ID Number: 10.1186/s13690-018-0313-1

URL: <https://doi.org/10.1186/s13690-018-0313-1>

☞ ID Type: isbn

Citation: Authors, Year, "Book title", Publisher.

ID Number: 9789491297687

URL:

## Notes

*Additional important information about the dataset.*

## Language

*Language of the dataset.*

- It is possible to include multiple languages.

## Producer

*Person or organization with the financial or administrative responsibility over this dataset.*

- Generally, this information can be found in the project's contract and/or your employment contract.

For example:

☞ Name: Your University

Abbreviation: YU

Back to [Table of content](#)

Affiliation:

URL: [www.youruniversity.be](http://www.youruniversity.be)

Logo URL: [www.youruniversity.be/logo](http://www.youruniversity.be/logo)

☞ Name: International Consortium

Abbreviation: IC

Affiliation:

URL: [www.InternationalConsortium.com](http://www.InternationalConsortium.com)

Logo URL: [www.InternationalConsortium.com/logo](http://www.InternationalConsortium.com/logo)

☞ Name: Federal Scientific Institution

Abbreviation: FSI

Affiliation:

URL: [www.fsi.gov.be](http://www.fsi.gov.be)

Logo URL: [www.fsi.gov.be/logo](http://www.fsi.gov.be/logo)

## Production Date

*Date when the data collection or other materials were produced (not distributed, published or archived)*

- Year-Month-Day

## Production Place

*The location where the data collection or other materials were produced.*

## Grant Information

*Specify the source of the research funding.*

For example:

☞ Grant Agency: Belgian Science Policy Office (BELSPO)

Grant Number: ABC/###/##

## 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.*

For example:

☞ [A dataset on travel behavior occurring in 2019, collected using a retrospective survey which was administered in January 2020]

Start: 2019-01-01

Back to [Table of content](#)

End: 2019-12-31

## Date of Collection

*Contains the dates when the data were collected.*

For example:

- ↳ [A dataset on travel behavior occurring in 2019, collected using a retrospective survey which was administered in January 2020]  
Start: 2020-01-01  
End: 2020-01-31

## Kind of Data

*Type of data included in the file.*

For example:

- |                               |                           |
|-------------------------------|---------------------------|
| ↳ survey data                 | ↳ experimental data       |
| ↳ census/enumeration data     | ↳ psychological test      |
| ↳ aggregate data              | ↳ textual data            |
| ↳ clinical data               | ↳ coded textual           |
| ↳ event/transaction data      | ↳ coded documents         |
| ↳ program source code         | ↳ time budget diaries     |
| ↳ machine-readable text       | ↳ observation data/rating |
| ↳ administrative records data | ↳ process-produced data   |

## Series

*Information about the data series.*

For example:

- ↳ Name: PAnel Study on Happiness in Adolescence  
Information: This dataset contains the results from the second wave of data collection of the Panel Study on Happiness in Adolescence. The first wave is also available in SODHA (doi:...). For more information: [www.PASHA.be](http://www.PASHA.be)

## Software Used to Create the Dataset

*Information about the software used to create the dataset.*

For example:

- ↳ Name: Qualtrics  
Version: Month(s) and year(s) of use
- ↳ Name: Excel  
Version: 16.16.20
- ↳ Name: Stata

Back to [Table of content](#)

Version: Stata/MP 13.1 for Mac

## Related Material

*Any material related to this dataset.*

For example:

- John, Paul & George, Ringo (2020). “btls: All you need is Data”. R package version 0.7.6. <https://CRAN.R-project.org/package=btls>
- Houston, Whitney (2020). Step by step meta-analysis. Pure Python module. <https://pypi.org/project/meta/sbsma>

## Related Datasets

*Any datasets that are related to this dataset, such as previous research on this topic.*

For example:

- Knowles, Beyoncé (2018). Queen Bee Behavior in Urban settings. First wave. Available in SODHA. Doi: ...

## Other References

*Any references that would serve as background or supporting material for this dataset.*

For example:

- The project website: Belgian Research on E-Leisure [www.BREL.be](http://www.BREL.be)

## Data Sources

*List of books, articles, serials, or machine-readable data files that served as the sources of data collection.*

For example:

- Information about employment per economic sector is retrieved from the official production statistics: National Institute for Statistics: ‘Bulletin mensuel du commerce de l’union économique Belgo-Luxembourgeoise avec Les Pays Etrangers (1948-1994; 1996-1998)’. Bruxelles: Ministère des affaires économiques, available in the Quetelet Fund library.

Back to [Table of content](#)

## Origin of Sources

*For historical materials, information about the origin of the sources and the rules followed by establishing the sources should be specified.*

For example:

- The production statistics were consulted in the Bibliothèque Fonds Quetelet, SPF Economie (contact: quetelet@mineco.fgov.be )

## Characteristic of Sources Noted

*Assessment of characteristics and source material.*

For example:

- The source material was in good condition and photocopied on site. The relevant information was entered in the spreadsheet by the lead investigator. The accuracy and completeness of data entry was checked by two different researchers.
- The source material included inconsistent data records. For example, one female respondent had a recorded diagnosis of prostate cancer. These inconsistent data records are not included in the available dataset.

## Documentation and Access to Sources

*Level of documentation of the original sources.*

For example:

- The original documentation is included in the dataset. However, information on the construction of the original indicator for well-being has not been included in the original documentation.

## Geospatial Metadata

### Geographic Coverage

*Information on the geographic coverage of the data. Includes the total geographic scope of the data.*

For example:

- [Dataset on mobility among disabled persons in Brussels, Ghent and Namur]
  - Country / Nation: Belgium
  - State / Province: Brussels Capital Region
  - City: Brussels

Back to [Table of content](#)

Other:

+ Country / Nation: Belgium  
State / Province: East Flanders  
City: Ghent  
Other:

+ Country / Nation: Belgium  
State / Province: Namur  
City: Namur  
Other:

- ↳ [Dataset with information on the proximity of hospitals in Belgium]  
Country / Nation: Belgium  
State / Province:  
City:  
Other:

## Geographic Unit

*Lowest level of geographic aggregation covered in the dataset (e.g. village, county, region)*

For example:

- ↳ [Dataset on mobility among disabled persons in Brussels, Ghent and Namur]  
City
- ↳ [Dataset with information on the proximity of hospitals in Belgium] exact GPS latitude and longitude coordinates for all Belgian hospitals

## Social Science and Humanities Metadata

### 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 ([https://ddialliance.org/Specification/DDI-CV/AnalysisUnit\\_1.0.html](https://ddialliance.org/Specification/DDI-CV/AnalysisUnit_1.0.html) ).*

For example:

- |                          |                  |
|--------------------------|------------------|
| ↳ Individual             | ↳ EventOrProcess |
| ↳ Organization           | ↳ GeographicUnit |
| ↳ Family                 | ↳ TimeUnit       |
| ↳ Family.HouseholdFamily | ↳ TextUnit       |
| ↳ Household              | ↳ Group          |
| ↳ HousingUnit            | ↳ Object         |

Back to [Table of content](#)

## 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, income, veteran status, criminal convictions, and more.*

*The universe may consist of elements other than persons, such as housing units, 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.*

For example:

- The target population includes all men and women between the ages of 18 and 35 who had their official residence in the Brussels Capital Region on 1 January 2020.
- The population universe entails all prisons in Flanders (Belgium).

## Time Method

*The time method or time dimension of the data collection, such as panel, cross-sectional, trend, time series and other.*

For example:

- Repeated cross-sectional data design [Data collection in 2012, 2015 and 2018]
- Longitudinal cohort study [Yearly data collection of the 1989 birth cohort from 2007 until 2019]

## Data Collector

*Individual, agency or organization responsible for administering the questionnaire or interview of compiling the data.*

For example:

- The organizational name of your project partner/ stakeholder / ...
- Family name, given name of your research partner/ master student/ ...

Back to [Table of content](#)

## Collector Training

*Type of training provided to the data collector.*

For example:

- ↻ An online learning module on the type of data collection and mandatory test
- ↻ A four-hour workshop with role-play exercises
- ↻ A detailed data collection manual

## 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.*

For example:

- ↻ Every three years: 2012, 2015, 2018 [Repeated cross-sectional data design]
- ↻ Yearly: 2007-2019 [Longitudinal cohort study of the 1989 birth cohort]

## Number of Variables

*Total number of variables for all data files that the dataset comprises.*

- Enter an integer

## Sampling Procedure

*Type of sample and sample design used to select the survey respondent to represent the population. May include reference to the target sample size and the sampling fraction.*

For example:

- ↻ Simple random sampling
- ↻ (Dis)proportionate stratified sampling
- ↻ Single stage/ Double stage/Multistage cluster sampling
- ↻ Non-random sampling with quota
- ↻ Expert sampling
- ↻ Systemic sampling



Back to [Table of content](#)

## Target Sample Size / Number of Units

*Specific information regarding the target sample size, actual sample size and the formula used to determine this.*

For example:

- ↳ Actual: Enter an integer representing your study population
- ↳ Formula: Cochran's formula (1963:75) =  $((Z^2 * (p) * (1-p))/c^2)$

## Major Deviations for Sample Design

*Show correspondence as well as discrepancies between the sampled units (obtained) and the available statistics for the population (age, sex, marital status, etc.) as a whole.*

For example:

- ↳ The ratio between men and women in the sample was somewhat lower than observed in the population (0.80 vs. 0.97).
- ↳ Data saturation was reached after 60 of the originally 70 planned interviews.

## Collection Mode

*Methods used to collect the data; instrumentation characteristics (i.e. telephone interviews, mail questionnaires, etc.); ...*

For example:

- ↳ Telephone interview
- ↳ Mail questionnaire
- ↳ Time use diaries
- ↳ Observation
- ↳ Focus group discussion
- ↳ Experiment

## Type of Research Instrument

*Type of data collection instrument used.*

For example:

- ↳ Structured  
indicates an instrument in which all respondents are asked the same questions/tests, possibly with pre-coded 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.

Back to [Table of content](#)

## 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.*

For example:

- Data collection was affected by the lock-down measures in Belgium in response to the COVID-19 pandemic. Before March 13<sup>th</sup> 2020, all data were gathered in face-to-face interviews. Data collection was suspended until April 14<sup>th</sup> 2020, after which data were collected in video interviews.

## 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.*

For example:

- Respondents were offered the opportunity to win a gift voucher of € 500 as an incentive to complete the survey.
- Replacement matching was used to minimize non-response. A detailed overview of the procedure and sensitivity analyses can be found in the technical report.

## Control Operations

*Methods to facilitate data control performed by the primary investigator or by the data archive.*

For example:

- The survey instrument was tested on potential language differences after translations through the Survey Quality Prediction program (SQP).

## 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.*

Back to [Table of content](#)

For example:

- A stratified sampling procedure was followed with non-proportional stratification at regional level. This unequal selection probability at regional level affects the representativeness at country level. In order to correct this, users should weigh the data during estimation processes using the variable `Be_weight` (i.e. the inverse of the sampling probability based on the known size of the stratum size).

## Cleaning Operations

*Methods used to clean the data collection, such as consistency checking, wild code checking, or other.*

For example:

- Records containing inconsistencies were removed from the available dataset (e.g. respondent 001: number of children=0; age of the child= 14)
- One expert probably provided the gross monthly salary for teachers in their country (in contrast to the requested gross annual salary). This entry is flagged using the variable `inconsist_tag` (inconsistency=1).

## Study Level Error Notes

*Note elements used for any information annotating or clarifying the methodology and processing of the study.*

For example:

- Due to the low response rate of elderly persons, the dataset may not reflect a representative view of the neighborhood perspective.

## Response Rate

*Percentage of sample members who provided information.*

For example:

- Information on official entry requirements for teacher training programs and in-service support policies is available for over 70% of the 36 participating countries. Salary information by level of experience was obtained for 62% of these 36 countries.

Back to [Table of content](#)

## Estimates of Sampling Error

*Measure of how precisely one can estimate a population from a given sample.*

For example:

- ↪ The relative standard errors (SE/R) for the means and proportions is 3.5% on average and range between 0.3% and 30.0 %.

## Other Forms of Data Appraisal

*Other issues pertaining to the data appraisal. Describe issues as response variance, non-response rate, testing for bias, interviewer and response bias, confidence levels, question bias, or similar.*

For example:

- ↪ The number of respondents for some specific occupational groups are very small (e.g. female truck drivers, male kindergarten teachers, ...), resulting in large confidence intervals when estimating turnover rates.
- ↪ The average survey response rate was 41%. A response bias is possible for lower-educated women as only 5% (20 out of 400) of contacted women in this group answered the questionnaire.

## Notes

*General notes about the dataset.*

- Type
- Subject
- Text

## Historical Metadata

## History of the Data

*Background information about the dataset, especially if it was previously kept by one or several institutions other than the SODHA data archive. Can also include information about what motivated the depositor(s) to entrust SODHA with their data.*

For example:

- ↪ This dataset was previously archived in the university archive. In order to enhance the findability, accessibility, interoperability and reuse of this dataset, a duplicate has been made available in SODHA.

Back to [Table of content](#)

## Biography of the Investigator/Contributor

*Background about (one of) the primary investigator(s) or contributor(s) of the dataset. This element is more relevant for senior than for younger researchers, as the latter's career is more likely to change with time. Still, depositors should feel free to include all information they think will help document the dataset. Use the "+" button to add more fields in case there are several investigators and/or contributors.*

*For example:*

- The principal investigator, Paul Delvaux, is one of the leading researchers in the field of the social perception of surrealist art. Throughout his career, he has spearheaded many data collection efforts including major data sources such as the TRAIN-study.

## History of the Research Center

*Background information about (one of) the research center(s) the primary investigator(s) worked for at the time of the study. Especially relevant if said research center(s) is (are) no longer active or if its (heir) name was changed. Use the "+" button to add more fields in case there are several research centers.*

*For example:*

- Social Mobility and Urbanization Research Faction (SMURF) was previously known as Social Change and Urban Mobility unit (SCUM).

## History of the Grant Agency

*Background information about (one of) the grant agency(ies) which funded the study. Especially relevant if said grant agency(ies) is (are) no longer active or if its (their) name was changed. Use the "+" button to add more fields in case there are several grant agencies.*

*For example:*

- This project was funded by Support for PROjects on Urban Trends (SPROUT). This new organization redistributes the financial funds of 20 private persons and companies to research and innovation projects through a competitive selection procedure. For more information, see [www.sproutgrants.be](http://www.sproutgrants.be)