SODHA DATASET PUBLISHING POLICY

This policy sets out under which conditions datasets that were submitted for review will effectively be published by SODHA.

Its main purpose is to explain the criteria that guide SODHA's editorial policy.

---

1. Overview of the publication criteria

Datasets that were submitted for review must meet the following conditions to be published:

- ☑ If access to one or several data files was restricted, the depositor specified access conditions in the 'Terms of Access' field. (Deposit Agreement, art. 10; Access and Reuse Policy, b.)

- ☑ The study that yielded the dataset is relevant to the social sciences and/or the digital humanities.

- ☑ The data do not contain sensitive information (e.g. personally identifiable information, 'PII').

- ☑ The metadata do not contain ambiguous information.

- ☑ The files of the dataset can be successfully downloaded and opened.

- ☑ When a new version of an already published dataset is created, changes must be documented in an extra instance of the 'Description' field.

- ☑ The dataset does not comprise any executable / binary files.

---

2. Criteria for dataset publication

SODHA follows criteria on the **fundamental** (content) level and on the **formal** level for determining whether submitted datasets are suitable for publication. The question of **sensitive information** is also taken into consideration, as well as that of **version management**.

Fundamentally, data collections must emanate from research domains and scientific disciplines that are relevant to SODHA's scope and missions.

Formally, the metadata that describe datasets have to follow a few basic rules.

In terms of security, datasets cannot contain sensitive information such as personal data.

2.1 Fundamental (content) level

SODHA is a **data archive** for **social sciences** and the **digital humanities**. It is destined for datasets relevant to one or more of the disciplines that constitute these two scientific domains.

To characterize the concepts of social sciences and digital humanities, SODHA refers to authoritative classifications such as BELSPO's project database FEDRA, ECOOM-Hasselt's Flemish Research Discipline Standard, and the report on digital humanities published by the Massachusetts Institute of Technology (MIT).

Notwithstanding this thematic criterion, SODHA has a 'neutral', 'agnostic' perspective on the contents of datasets: the role of our data archive is not to appraise either the quality of the research performed by depositors, the size of datasets, or the research topics chosen by investigators.

Should a dataset not be considered valid for deposit in SODHA in the light of the classifications previously mentioned, SODHA can provide guidance regarding other publication media or repositories (as far as our expertise will allow).

2.2   Sensitive information

At the moment, SODHA unfortunately cannot afford to accept **data** that contain sensitive information, typically (though not exclusively) personally identifiable data, although personal information that cannot, by itself or via cross-referencing, identify individuals is acceptable. The infrastructure is undergoing development so that this becomes a possibility in the future.

However, depositors are welcomed to record **metadata** about sensitive data on the SODHA Dataverse platform. In other words, if they cannot publish research data because they are too sensitive, they are nevertheless encouraged to mention the existence of such data, as this is very valuable information for other researchers in the field. See our FAQ for more information.

2.3   Formal level (metadata)

When depositors set out to deposit a dataset in SODHA, they are asked to fill in a number of mandatory metadata fields.
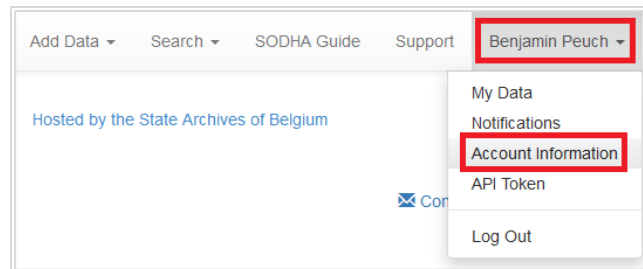
Maintaining the **formal consistency** of the metadata published in SODHA increases the chances that researchers find the data they are looking for when they browse or query the SODHA Web application. That is why the following rules apply:

a. **Names** of individuals mentioned in the fields Author, Contact, and Producer must be fully spelled out (especially the first name, notwithstanding middle initials), separated by a comma, and in this order: **[family name(s), first name, middle initial(s)]**

   Examples   Pirlot, Mélanie
                      Van den Broeck, Jan-Albert
                      Márquez Gutíerrez, Emelina Soraya
                      de la Vega, Christian
                      Von Humboldt Zimmerman, Dietrich
                      du Val, Anne-Sophie Angélique
                      Smith, Christian A.
                      Pham, Van Duc
                      Zhang, Ziyi

Exceptions can be made if, for example, in the case of legacy datasets, the first name(s) of authors cannot be identified with certainty. In such cases, please notify SODHA administrators via sodha@arch.be.

Because the name of the depositor is automatically documented in the field of the same name, 'Depositor', as based on the depositor's account information, depositors are asked to spell out their first name in full, either when they register on SODHA or by later modifying their personal information (which they can do in the 'Account Information' menu):

b. To display **linebreaks** in metadata fields, depositors should use the HTML tag **\<br\>** and not just the Enter key. The method is the following:

In order to obtain this:

> This dataset was produced between 2012 and 2014.
>
> Data were collected between 3 July 2012 and 6 January 2013.

\<br\> tags must be added like so:

> This dataset was produced between 2012 and 2014.**\<br\>\<br\>**
>
> Data were collected between 3 July 2012 and 6 January 2013.

c. **Paragraphs** are spaced out with an empty line, as shown above.

d. **Titles** of **works cited**, for example in the field 'Related Publication', can be displayed in italics by the use of the tags **\<i\>\</i\>**. The method is the following.

In order to obtain this:

> D'hont, Ekaterina, *Emotions: Psychology and the Social Sciences*, Brussels, Psykon Press, 2016.

\<i\>\</i\> tags must be added like so:

> D'hont, Ekaterina, **\<i\>**Emotions: Psychology and the Social Sciences**\</i\>**, Brussels, Psykon Press, 2016.

e. When **more than one element** must be encoded in a **repeatable field** (as shown by the '+' but-
ton), the field should be repeated as many times as necessary, rather than encoding everything
in one instance of the field:

Repeating the fields enhances indexing and therefore increases the publicity of your
data.

f. Because SODHA promotes open, **linked data**, if administrators of the SODHA platform come by
additional information such as a DOI of a work cited, the ORCID of an author, etc., they will
suggest adding those elements to the depositors' metadata. This helps enhance the online visi-
bility of your data.

2.4   Version management

If a dataset has already been published once and depositors create a new version of it, they should
document the changes they have brought to the earlier version in a new instance of the
'**Description**' field, alongside the original one(s). They can do so by clicking on the '+' button to re-
peat the field:

2.5   Executable / binary files

For security reasons, SODHA cannot accept datasets that contain executable files.

2.6   Final considerations

When a depositor submits a dataset for review, a SODHA administrator will proofread the metadata
before publishing it.

If the metadata formally deviate from the rules outlined above, SODHA administrators will take the
liberty of making **minor** corrections in this respect. **Apart from these rules, SODHA administrators
will not modify the depositors' input in any way, and depositors will always be notified of any
change made to their metadata by SODHA administrators**.

If, on the other hand, certain elements in the metadata are viewed as ambiguous (e.g. because of typographical mistakes), SODHA administrators will contact depositors and ask them to clarify things prior to publication.

3.   Version history of this document

Date	Version number	Changes made
07-09-2020	1.0	First version
03-02-2021	2.0	New version history section, minor layout edits, correction of typos
03-03-2021	2.1	Brackets removed from the full name of SODHA
27-04-2021	2.2	Missing word on the first page ("*can* be successfully") added
06-05-2021	2.3	Font of the document changed from Arial to Calibri
05-08-2021	3.0	New subsections 2.4 'Version management' and 2.5 'Executable / binary files'
07-03-2022	4.0	Definition of 'sensitive data / information' no longer includes personal information that does not directly identify individuals (as opposed to personally identifiable information)