AUSSDA
AUSTRIAN SOCIAL SCIENCE DATA ARCHIVE

FILE NAMING AND FILE ORGANISATION

Organisation of archival processes for a long-term data preservation under consideration of the OAIS reference model

2.10.2026

Christian Bischof, Iris Butzlaff

Date	2.10.2026
Version	1.2
License	 This work is licensed under a Creative Commons Attribution 4.0 International License.
Access	Public
Suggested citation	Bischof, Christian, Iris Butzlaff (2026) File naming and File organisation. Vienna: The Austrian Social Science Data Archive.

Filenames: all lowercase, no diacritics, no blanks, underline between elements

Numbering of files: 5 digits, first number: 10001, ascending (+1)

Language of file: use of ISO 639-1

For files (e.g. photos) without applicable language, leave empty (e.g. 10001_da _v1_0.jpeg).

	ISO language name	native name
en	english	english
de	german	deutsch

Versioning: two levels: 1_0 (Major-Minor)

Major: deleting or inserting of a variable or case

Minor: significant change (labels, recoding, etc.), not significant change (cosmetic) - not significant changes are only modified in files when significant change is necessary

Code of files: for processed files in AIP and DIP packages

If more than one file with same code use additional 2-digit numbering after code (e.g. 10001_im01, 10001_im02).

For a data file with same content but other data format (e.g. Stata <-> SPSS), use same filename (distinction by file-extension).

Code	Description English	Description German	Comment
ap	anonymisation plan	Anonymisierungsplan	
da	data	Daten	
cf	consent form	Einverständniserklärung	
co	codebook	Codierliste	
fr	field report	Feldbericht	
im	interviewer manual	Interviewerhandbuch	
mr	method report	Methodenbericht	
om	other material	anderes Material	Use “Template_DataLifeCycle_Log.xlsx” for description
qu	questionnaire	Fragebogen	
rr	research report	Forschungsbericht	
sx	code/ syntax	Code/ Syntax	
ta	tabulation report	Tabellenband	

vi	variable identifiers and descriptions	Variablenidentifikation und -beschreibung	

Example: 10001_da_de_v1_0.sav

10001: dataset number

da: data file

de: german file

v1_0: version major 1

Folder Organisation:

	Package Folder	Package Files		Package Content
/dataset	/SIP	from depositor received files		submission information package
	/AIP	processing files, license documents		archival information package, long-term archival data
		/data	processed data	
		/documentation	processed documentation	
	/DIP	/data	data for dissemination	dissemination information package
		/documentation	documentation for dissemination	

Dataset=number of dataset (5 digits)

Folder names lowercase, excepting Package Folder (SIP, AIP, DIP) uppercase.

For old versions in SIP use folder changelog:

/SIP/changelog

In AIP all versions: if new versions, move previously published files from AIP/data or AIP/documentation to the new created folder AIP/data/vX (X = version number) or AIP/documentation/vX:

AIP/data/v1

AIP/documentation/v1

For correspondence with depositor (E-Mail, teamwork report, etc.) use folder communication in AIP:

/AIP /communication

Use folder where all check files are collected:

/AIP /checks

In DIP only current version for dissemination.

Additional files for process documentation

DataLifeCycle_Log:

Use "DataLifeCycle_Log.xlsx" (folder: teamspace\02_acquisition\documents\ name: e.g. 10001_DataLifeCycle_Log.xlxs), store in AIP:

- Used from acquisition as log-file
- To describe renaming of documentation files from SIP to AIP
- For the description of processing of files. Also important to document not so significant changes for the upcoming release of a new version
- To document the current files of the folders AIP and DIP, and the current files uploaded to Dataverse
- To document the disposition of files stored in SIP but not processed and not published

Readme:

Use "AIPTemplate_readme.docx" (store as: name e.g. 10001_readme.pdf, file format PDF/A). One file describes all changes, see template for examples. For description of internal processes use DataLifeCycle_Log file. To store in the documentation folder AIP and DIP, and upload to Dataverse:

- Recommended Dataset Citation: citation from Dataverse
- File Overview: files in Dataverse with file description
- Data curation: to record processes for potential users, e.g. anonymization, changes in new versions (newest record on top)

Designated file formats in AIP and DIP package folders for data and documentation

Filename Extension	File	Comment
.csv	raw data	UTF-8 encoded, long-time preservation
.dta	Stata data	version 118 (14)
.sav	SPSS data	UTF-8 encoded
.do	Stata do	processing code
.sps	SPSS syntax	processing code
.xlsx	data	for not statistical trained communities (for Dataverse double zipped to prevent ingest)
.csv	text	UTF-8 encoded, documentation, data
.pdf	PDF/A version	documentation long-time preservation
.txt, .R, etc.	ASCII text	documentation of programming & syntax of assigned software

Additional file codes for GfK Project

Numbering of files starts with 20001. To use in AIP and DIP package folders, without subfolders data and documentation (all files in main folder AIP and DIP). Only German files as PDF/A-2b in documentation folder.

Code	Description English	Description German	Comment
gr	Graphs	Grafikband	
te	Text report	Textband, Textbericht, Textkommentar	