

Slovak Archive of Social Data (SASD)

Preservation Policy

(1.01 28.1.2021)

Preservation Policy

The Preservation Policy sets out the individual activities and procedures performed within the Slovak Archive of Social Data (hereinafter as the Archive).

The individual phases of preservation are modelled on the principles laid out in the ISO 14721 standard - Open Archival Information System(OAIS)¹

¹[ISO 14721 \(OAIS\)](#)

Contents

1. Pre-Ingest Stage

- 1.1 Planning and search for data to be preserved
- 1.2 Data Provision Agreement
- 1.3 Submission Information Package - SIP
- 1.4 Data formats of ingested data

2. Data Ingestion

- 2.1 Data validation
- 2.2 Archival Information Package - AIP
- 2.3 SIP and AIP audit
- 2.4 AIP descriptive information

3. Archive System

- 3.1 Electronic Materials Archive
 - 3.1.1 SIP Archive
 - 3.1.2 AIP Archive
 - 3.1.3 DIP Archive
- 3.2 Archive of Versions & versions administration
- 3.3 Non-Digital Documents Archive
- 3.4 Storage facilities
- 3.5 Backup storage media

4. User Access Administration

- 4.1 Dissemination Information Package – DIP
- 4.2 Search for data

5. Overview of Job Positions in SASD with respect to performed tasks

- 5.1 The Pre-Ingest Stage & Data Acquisition Administrator
- 5.2 Archive System Administrator
- 5.3 User Access Administrator
- 5.4 Shared activities

1. Pre-Ingest Stage

1.1 Planning and search for data to be preserved

Within the pre-ingest stage, the Archive focuses on identifying potential depositors who possess data suitable for preservation. The Archive Director and the Pre-Ingest Stage & Data Acquisition Administrator draft a plan for gathering new data. This plan contains information on potential institutions in possession of such data i.e. the data originators as well as a time schedule detailing procedures for its fulfilment.

When searching for datasets suitable for preservation, the Pre-Ingest Stage & Data Acquisition Administrator ensures that the basic required information on individual materials is secured. The required information shall include:

- basic information about the project (project title and topic),
- producer name,
- identification of depositor's contact person,
- basic information about the dataset,
- basic information about the documentation.

In case a negotiation with a potential producer does not conclude with an agreement, the Pre-Ingest Stage & Data Acquisition Administrator shall inform the Archive Director of such matter. If an agreement has been reached, the Administrator forwards the gathered information to the Archive Director for assessment.

1.2 Data Provision Agreement

The Archive Director assesses the individual materials acquired in the negotiation process with the producer or their authorised contact person. If the provided information is sufficient, the Director entrusts the Pre-Ingest Stage & Data Acquisition Administrator with drafting a Data Provision Agreement. Conversely, per Archive Director's request, the authorised administrator asks the producer to provide necessary additional data or he/she rejects the potential preservation.

The authorised Archive Administrator forwards the template of draft agreement² to the producer and subsequently negotiates the agreement terms with the producer. Upon producer's proposal, specific requests from the producer's side may be incorporated into the agreement.

An essential part of the Data Provision Agreement is formed by the classification of the provided materials into appropriate data categories with respect to their access rights³.

The authorised Archive Administrator drafts the Data Provision Agreement, forwards it to the Archive Director for approval and subsequently sends the draft agreement to the producer's authorised representative for comments.

²[Draft of Data Provision Agreement](#)

³[Categories of preserved data](#)

Once the comments have been incorporated and approved by both contracting parties, the authorised Archive Administrator prepares two copies of the agreement and ensures the agreement is signed by the authorised representative from both the producer's and Archive's side.

1.3 Submission Information Package - SIP

Following the Data Provision Agreement which specifies the terms of submission as well as the extent of data to be deposited into the Archive, the producer provides the Archive with the Submission Information Package (hereinafter as SIP). The package contains information about the package content and also includes data which facilitates the preservation in the Archive and will allow future users to retrieve, interpret and use the data from the Archive.

The Archive receives the data from the producer in the original form of SIP which contains:

- data files in spss format,
- questionnaires in electronic and printed form,
- frequency tables in electronic form,
- methodological information about the research,
- information about the research scope and contents as well as the list of principal research investigators.

The Archive issues the producer with a confirmation of the receipt of data which includes a list of ingested materials and a request for re-submitting the SIP in the event of error detection.

1.4 Data formats of ingested data

In order to secure a long-term preservation, data interpretability and access, it is essential to provide data in the preferred data format.

The table below lists data formats preferred and accepted by the Slovak Archive of Social Data.

Type of data	Preferred data formats	Acceptable data formats
Primary data matrices	Formats of the SPSS software .sav	MS Excel.xls, .xlsx
Documentation (in text form)	MS Word, PDF, PDF/A .pdf	Libre Office
Graphical data	jpg., gif., png.	

2. Data Ingestion

2.1 Data validation

Each SIP received should meet the Archive's minimum requirements for completeness. Therefore it is essential to subject the submitted data to validation which primarily checks the documentation for its completeness and the formats of electronic files.

In particular, the data is checked in terms of the completeness of the data file and the agreement of the data file with the research (e.g. by comparing the dataset against the questionnaire). Labels are checked for completeness and readability. A preliminary check of data for legal issues is performed, primarily ensuring no breach of provisions of Act on personal data protection and on amendments and supplements to certain acts⁴.

In case the data or documentation are found to contain errors or be incomplete, the Pre-Ingest Stage & Data Acquisition Administrator contacts the producer in order to remove discrepancies from the submitted data and to acquire an updated version of the SIP.

SIP is a tool for maintaining the authenticity and integrity of materials deposited by the producer.

Non-digital SIP files (e.g. a questionnaire, an agreement) are scanned and subsequently stored in the archival system.

2.2 Archival Information Package - AIP

The Pre-Ingest Stage & Data Acquisition Administrator generates Archival Information Package (hereinafter as AIP) which comprises the primary data itself and the metadata describing it. The metadata is created in compliance with CESSDA recommendations and is converted to xml.format via a DDI template. The individual files included in the AIP are converted into standard formats.

The AIP structure shall be designed to be user-friendly and fully readable.

Besides the data itself and the metadata describing it, the AIP also contains information describing the preservation process itself. The outputs of such information are e.g. a Report on the preliminary data validation or a Report on SIP and AIP audit.

2.3 SIP and AIP audit

The Pre-Ingest Stage & Data Acquisition Administrator forwards the generated AIP as well as SIP to the designated archive worker for audit. The worker responsible for such audit has not been involved in the pre-ingest and ingest phase.

Based on the audit results, the AIP may require error corrections or additions. When the audit is completed, a report comprising information on data ingest process is forwarded to the producer. █

⁴Act No. 18/2018 Coll. Act on personal data protection and on amendments and supplements to certain acts

2.4 AIP descriptive information

In order to allow users to identify and search for a particular AIP, the AIP descriptive information has to be generated. This document contains information obtained primarily from AIPs and the essential part of such information are e.g. key words describing the content of the data file.

The European Language Social Science Thesaurus (ELSST) is the primary source for determining key words.

3. Archive System

3.1 Electronic Materials Archive

All materials and documents in digital form are stored in Electronic Materials Archive. It contains all SIPs, AIPs and DIPs together with their individual versions and is administered by the Archive System Administrator.

3.1.1 SIP Archive

SIP Archive contains directories of SIPs.

The Pre-Ingest Stage & Data Acquisition Administrator establishes a new directory in the Electronic Archive when a file suitable for preservation is identified. Within the directory, he/she then creates a folder for an SIP as described in section 1.3 of this document. The directory is given an ID number as its unique identifier. The same identifier is also linked to appropriate files associated with the given dataset which are stored in the Electronic Materials Archive and Non-Digital Documents Archive.

The SIP Archive also holds all records on activities of the Archive performed within the process of acquiring and ingesting data.

3.1.2 AIP Archive

AIP archive contains directories of all AIPs. Each AIP is labelled by a unique ID which is also linked to the appropriate SIP as well as DIP - Dissemination Information Package.

A detailed description of an AIP can be found in section 2.2.

AIP Archive contains all AIP versions.

3.1.2 DIP Archive

DIP Archive contains directories of all DIPs together with their individual versions.

DIPs are formatted in line with a client's request and are generated in the .zip format.

A detailed description of a DIP can be found in section 4.1.

3.2 Archive of Versions & versions administration

The Archive of Versions contains all versions of individual files in SIPs, AIPs and DIPs as well as information about all changes made to these files.

A new version is issued following any changes in the information package. New versions are approved by the Archive Director. Each version is allocated a unique DOI code.

All versions are included in the AIP.

Following users' comments or at the Archive Administrator's or depositor's initiative, the Archive Director checks whether it's necessary to perform an update or error correction. In case the Archive Director decides an update is required, the Pre-Ingest Stage & Data

Acquisition Administrator carries out an AIP and SIP update. He/she then informs the Archive System Administrator of such changes who in turn updates the DIP.

3.3 Non-Digital Documents Archive

Non-Digital Documents Archive contains all materials and documents in non-digital form which are related to datasets and to the Archive operations. In particular, these are mainly documents associated with the research itself (e.g. questionnaires, encryption keys, methodology questionnaires etc.) and documents related to the Archive operations (e.g. Data Provision Agreements). Such documents are preserved in individual files, labelled by a unique ID number which is in turn allocated to all related files and documents.

Each file includes a list of all documents associated with a particular research.

Documents containing sensitive personal data or confidential information are kept in a locked metal cabinet.

Any borrowings of such documents are documented in the Record of Borrowings which is administered by the Pre-Ingest Stage & Data Acquisition Administrator .

3.4 Storage facilities

Individual directories stored on the server of the Institute for Sociology represent the main storage facility of the Archive. The Slovak Academy of Sciences administers and regularly backs up this server.

An additional storage facility is the NESSTAR Publisher where all AIP versions are stored.

Furthermore, the entire contents of the Data Archive are also stored on an external data storage (an external hard drive) which is kept at the Archive offices and is regularly backed up.

3.5 Backup storage media

Individual archive files are stored on the server of the Institute for Sociology of the Slovak Academy of Sciences (hereinafter also as IS SAS). The server is regularly backed up by the Computing Centre of the SAS Centre of Operations.

A back-up server is being prepared at the premises of IS SAS.

The System Administrator ensures that the individual data files are regularly duplicated and stored on an external storage device (an external hard drive). One copy of the data files is kept outside the premises of IS SAS.

4. User Access Administration

4.1 Dissemination Information Package – DIP

Dissemination Information Package (hereinafter as DIP) is generated by the User Access Administrator who is responsible for both its complete and updated version.

DIP is automatically generated upon a user query submitted via the Archive's Online Archive System and is made available to the user in .zip format.

4.2 Search for data

The Archive's Online Archive System allows its users to search for data via the search or advanced search functions.

The search function allows the users to search for research information and for data files according to selected criteria – e.g. research field and topic, type of research programme, representative research questions, the year of research execution and others.

The Online Archive's data catalogue provides users with access to the list of archived research results and the research documentation categorized by the year of data collection from the latest to the oldest. Apart from the year of data collection and the research title, the research documentation (e.g. used questionnaires, a methodological description of the research, sociopolitical context of data collection) and data file frequency tables can also be directly accessed via the catalogue.

The non-representative and other research function allows users to access the list and description of data as well as the methodology of data collection from researches which are not based on a representative questionnaire type of social data collection.

Anyone who wishes to have access also to data files in .spss format and receive them online must fill in the Data Access Form. In case of requesting data from researches unavailable online, the Data Access Form including the requester's signature and company seal has to be sent by post to the address of the Slovak Archive of Social Data (hereinafter also as SASD). The form also defines the extent of accessibility of stored data files and thus the conditions under which the data can be obtained.

The Archive's Online Archive System and its catalogue are tools based on the DDI standard. The DDI documentation standard together with the DTD standard allow for the conversion of data files' electronic code lists into a multipurpose XML format.

5 Overview of job positions in SASD with respect to performed tasks

5.1 Archive Director

The Archive Director is responsible for managing activities and processes within the Slovak Archive of Social Data. From the legal perspective, the Archive Director is liable to a statutory body of the Institute for Sociology of SAS.

5.1 The Pre-Ingest Stage & Data Acquisition Administrator is responsible for the following activities within the Archive:

- planning the search for data,
- monitoring the potential research fields and projects suitable for preservation,
- contacting potential depositors and negotiating data submission agreements with them,
- concluding Data Submission Agreements,
- receiving and checking the submitted data in the SIP,
- generating the AIP,
- updating SIP and AIP versions as necessary.

5.2 Archive System Administrator

The Archive System Administrator ensures that the individual datasets within the Archive and all information about them are properly administered throughout all stages, from their acquisition up to their delivery to the final user.

The Archive System Administrator is responsible for the following:

- administering SIP, AIP and DIP Archives,
- updating SIP, AIP and DIP versions,
- ensuring back-ups of SIPs, AIPs, DIPs and their versions.

5.3 User Access Administrator

User Access Administrator is responsible for

- generating and updating DIPs,
- verifying that data access conditions are being fulfilled by users,
- verifying the functionality of the Archive's Online Archive System,
- communicating with users.

5.4 Shared activities

- proposing procedures for protection and preservation of archived data,
- processing and updating internal documentation regarding standards.