CentERdata
Institute for data collection and research

Preservation and Dissemination Policy of the LISS Data Archive

Date	December 2020
Authors	Marika de Bruijne, Joris Mulder, Maarten Streefkerk
Version	2.0
Classification	standard

Table of Contents

1 Introduction

The LISS Data Archive provides access to the data that are collected in the LISS panel. The panel is an innovative data collection facility intended to boost and integrate research in various disciplines, such as economics, social sciences, life sciences, and behavioral sciences. The Longitudinal Internet studies for the Social Sciences (LISS) panel is a representative panel of about 5,000 households based on a probability sample drawn from the population register. Respondents participate in surveys over the Internet each month.

This document outlines the data preservation and dissemination policy for the LISS Data Archive[1], hereafter 'the archive'. It presents the purpose of the archive and describes how the management tasks as well as the operational archival and dissemination functions are organized. Further, the document describes the measures that have been taken to ensure the security and preservation of the LISS data in the long-term.

[1] https://www.dataarchive.lissdata.nl/

2 Purpose

2.1 Mission

The LISS Data Archive preserves and disseminates the data that are collected in the LISS panel[2]. The LISS panel was established to facilitate research in the social sciences in the Netherlands and abroad. The facility is open to academics and policy makers anywhere in the world for (non-profit) scientific purposes.

2.2 Scope and Objectives

The data which are collected in the LISS panel are made available online for all scientific researchers via the LISS Data Archive. The aim of the archive is to provide reliable and easily accessible information, including data and metadata, on the entire life-cycle of the LISS research projects.

In addition to its own archiving and (meta)data dissemination system, the LISS panel archives its data in EASY, the online archiving system of the Dutch Data Archiving and Networked Services (DANS), to ensure their long-term availability.

[2] https://www.lissdata.nl

3 Legal and Regulatory Framework

The LISS Data Archive is managed and operated by CentERdata, a scientific research institute housed on the campus of Tilburg University in the Netherlands. CentERdata will at all times comply with applicable laws and regulations in the Netherlands, including the General Data Protection Regulation (GDPR; in Dutch: Algemene Verordening Gegevensbescherming). Furthermore, CentERdata uses working methods that are in accordance with the guidelines developed by the Association of Universities in the Netherlands (VSNU) as described in the Code of Conduct for use of personal data in scientific research (VSNU, 2005).

CentERdata is registered at the Dutch Data Protection Agency (*Autoriteit Persoonsgegevens*) under No: FG008875. CentERdata is registered at the Tilburg Chamber of Commerce under the number KvK Tilburg: 41098659.

4 Organization

In 2006, the Dutch Research Council (NWO) granted a proposal entitled *An Advanced Multi-Disciplinary Facility for Measurement and Experimentation in the Social Sciences (MESS)*. The funding was used to set up the LISS panel (see NWO, 2006). The Immigrant panel was set up as an additional panel in a later stage of the project, but was discontinued in 2014. In this document we will refer to both panels by the name LISS panel. The data collected in these panels are preserved and disseminated via the LISS Data Archive. The archive is managed by CentERdata, a partner of Tilburg University, the Netherlands.

Within the organization, specific roles are responsible for the LISS panel and the data archive. Below we describe the roles and responsibilities according to three main functions within the data life-cycle: data production, data archiving & management and data consumption (see also the illustration in Chapter 6, Figure 1). CentERdata both collects and archives the data of the LISS Data Archive, which is why some of the roles can apply to both the data production as well as the data archiving & management tasks.

4.1 Data Production

Director CentERdata

The director of CentERdata makes the strategic decisions concerning the panel and has the final responsibility for data safeguarding.

Head of Survey Research

The Head of Survey Research at CentERdata is responsible for the organizational management of the panel. He/she oversees the planning of data collection and is responsible for the contracts with Client Researchers. He/she reports to the director of CentERdata.

LISS Coordinator

The LISS Coordinator oversees the operational activities related to the LISS panel. He/she is responsible for informing and obtaining the consent of LISS respondents and for maintaining the representativeness, quality and response of the panel. He/she reports to the Head of Survey Research of CentERdata.

Panel Manager

A special department is dedicated to operational management of the panel, including support to and contact with the panel members. The panel manager coordinates these tasks and the employees within this department.

Project Leader

Project Leaders in the LISS panel's data collection and dissemination projects are provided by the Survey Research department. For each Submission Information Package (SIP) there is a second reader control by another employee of Survey Research, before the SIP is delivered to the Data Archive Coordinator (see section 4.2). The Head of the Survey Research supervises these employees.

System Administrator

The system administrator performs routine maintenance of the IT infrastructure and monitors the proper functioning of the servers.

4.2 Data Archiving & Management

LISS Coordinator

The LISS Coordinator is responsible for the data archiving and dissemination of the LISS data. He/she oversees the implementation of the archiving, data management and dissemination activities. He/she is also responsible for the contracts with Data Users.

LISS Data Archive Coordinator

The LISS Data Archive Coordinator is responsible for the operational data ingest activities and dissemination of the metadata and data. He/she controls the SIPs and transforms them into Archival Information Packages (AIP). He/she coordinates the data-entry tasks of the Data Archive Operator(s). He/she accepts and publishes data updates on the lissdata.nl archive website and coordinates the depositing of data disseminated via the LISS Data Archive in the EASY online archiving system of Data Archiving and Networked Services (DANS). He/she stays updated on the developments of new data formats and statistical tool versions and will take timely action to safeguard the long-term usability of the data and metadata. He/she reports to the Head of Survey Research and works in close consultation with the LISS Coordinator.

LISS Data Archive Operator(s)

The Data Archive Operator is responsible for entering the data and metadata in the LISS Data Archive. He/she also carries out the submissions of the data deposits in the online repository of Data Archiving and Networked Services (DANS). Publications based on LISS panel data are put on the LISS Data Archive website by the Data Archive Operator. The Operator also receives and controls the signed Contracts for the use of data for Data Users and grants the access rights to Data Users. When in doubt about granting access, he/she seeks advice from the LISS Data Archive Coordinator, the LISS Coordinator or the Head of Survey Research.

Database Manager

The Database Manager develops and maintains the online archival system (front-end) and the related online dissemination application (back-end). He/she also stays updated about the developments in the archival standards such as the DDI. Receives input for implementation of new features via the LISS Data Archive Coordinator.

Information Security and Privacy Officer

The Information Security and Privacy Officer is responsible for the information and physical security measures that are taken to ensure the safety and availability of the archival data stored at CentERdata.

Partner: DANS

For additional long-term preservation guarantee, the data disseminated via LISS Data Archive are also deposited in the repository of DANS.

4.3 Data Consumption

Client Researcher

The Client Researcher gives an assignment to CentERdata to collect data in the LISS panel. Prior to data collection, he/she signs a contract with CentERdata on the data collection project and, prior to receiving the data, signs the Statement for the Use of Data[3].

Data User (Consumer)

Data Users (or Consumers) comply with CentERdata's stated rules on using the data in an appropriate manner by signing the Statement for the Use of Data of the LISS panel before being granted access.

[3] See https://statements.centerdata.nl/liss-panel-data-statement

5 Cooperation

CentERdata cooperates with the following parties to facilitate the LISS Data Archive and long-term preservation of the data:

5.1 DANS

The data that are archived in and disseminated via the LISS Data Archive are also deposited in the online repository of DANS. Data Users have access to the metadata via DANS, but are referred to the LISS Data Archive to download the actual data files and more detailed metadata. The metadata fields in the EASY system are modeled as much as possible by the specifications of Qualified Dublin Core[4].

5.2 Combell

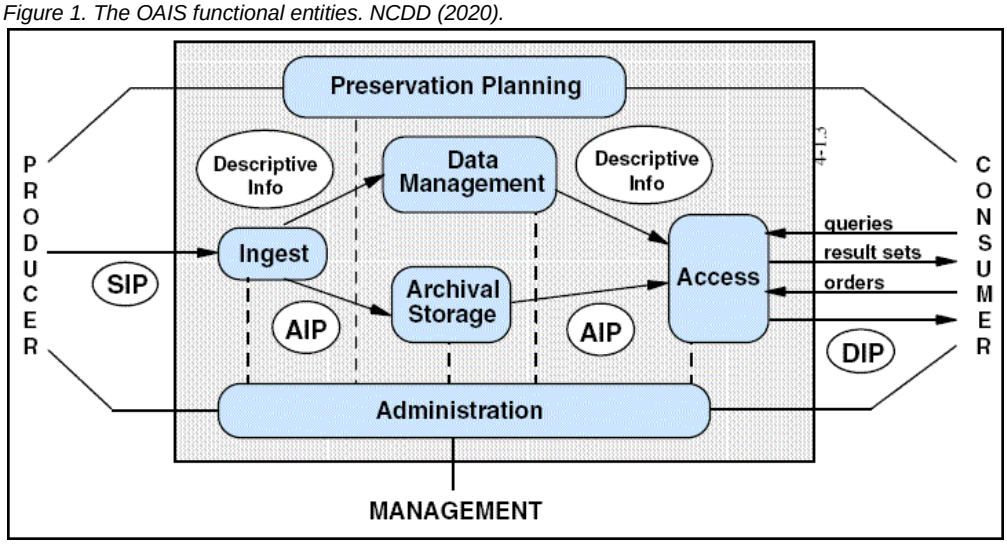
To store the data, CentERdata uses the services of the hosting provider Combell. The Combell data centers are fully ISO 27001 certified, redundant and physically located within the European Economic Area. CentERdata has its own protected partition, to which third parties have no access.

[4] http://dublincore.org/documents/dcmi-terms/

6 Data Process

This chapter describes the different tasks around the LISS Data Archive, applying the Open Archival Information System (OAIS) reference model. According to this model, the data processing can be divided into six functional entities and related interfaces (CCSDS, 2012): ingest, data management, archival storage, administration, access, and preservation planning (see Figure 1). In addition, we describe the pre-ingest processes involved in data collection.

Figure 1. The OAIS functional entities. NCDD (2020).



6.1 Pre-ingest

The data of the LISS Data Archive are collected in the LISS panel. Academic researchers world-wide can request to collect data in the LISS panel. This procedure is described on the panel website[5]. Once a research contract has been signed, a CentERdata employee will act as a Project Leader for the data collection project and will confer with the Client Researcher to coordinate the timing of the fieldwork in the panel, as well as the questionnaire content and details. CentERdata has the right to revise or decline questions that it deems unsuitable for the panel members.

The CentERdata Project Leader is responsible for the correct data collection and processing of the data, and for preparing the data for the LISS Data Archive. After completing the fieldwork, the CentERdata Project Leader delivers the data to the Client Researcher. After this, the study is prepared for archiving.

The different steps of the LISS data archiving are given in Table 1. First, the Project Leader prepares the data into a Submission Information Package (SIP) (1). This SIP contains the dataset and related documentation, such as the codebook, that are prepared in conformity with LISS Data Archive's standard so that the files are ready to be ingested in the archive. The procedure for preparing the SIP is documented in a manual that contains the data and metadata requirements and quality checks. All data processing steps are run by using an SPSS syntax file to ensure a full audit-trail to the original raw data file and a reconstruction

[5] https://www.lissdata.nl

of the data processing. Each SIP is also checked by a second reader (2) and if necessary improved based on the feedback of the check (3), before it is sent to the Data Archive Operator for ingest (4).

Data Quality

Several studies on data quality have been carried out in the LISS panel. Comparing data from the CATI/CAPI recruitment interview for the LISS panel and data collected in the LISS online panel questionnaires, Scherpenzeel (2009) concluded that data collected via the Internet are at least as valid and reliable as those collected in the CATI and CAPI interviews. Validity coefficients estimated in a Multitrait-Multimethod (MTMM) model were higher for the Internet questions than for CAPI questions and similar to the CATI questions. Reliability coefficients obtained with the same model were clearly higher for the Internet questions than for both the CATI and CAPI questions. Using an MTMM model as well, Révilla and Saris (2010) compared the quality of European Social Survey (ESS) questions in the (regular) face-to-face survey with the quality of the same questions in the LISS panel. Their conclusion was that the validity and reliability coefficients for ESS questions were similar in the Internet and the face-to-face mode of data collection. In another comparative research, Rekker, Van der Meer and Van der Brug (2020) showed that the quality of data collected by the LISS panel is comparable to data collected through CAPI and a fresh random CAWI sample drawn by Statistics Netherlands (CBS).

6.2 Ingest

When the Data Archive Operator receives the SIP, he/she checks the data and metadata. The Operator follows a data entry checklist, based on which he/she either accepts or rejects the SIP (5). If the SIP is accepted, the Operator converts it into an Archival Information Package (AIP) by entering the data and documentation in the archive and adding the relevant metadata (6). The data-entry forms that are used for entering (meta)data in the LISS Data Archive contain several systematic checks to prevent entering incorrect or duplicate (meta)data. After this, a colleague Operator checks the AIP and either accepts or rejects it for publication (7).

Table 1. LISS data archiving protocol

	Step	Description	Who	Documentation
1	Prepare SIP	LISS dataset and codebook are processed into a SIP	Researcher A*	LISS dissemination manual
2	SIP check	The SIP is checked	Researcher B*	LISS dissemination checklist
3	Improve SIP	The SIP is improved based on feedback from the second reader	Researcher A	
4	Request to archive	LISS Data Operator is asked to archive the SIP	Researcher A	
5	Accept to archive	LISS Data Operator checks that the SIP conforms with the archival standards and accepts or rejects the SIP	Operator A	LISS Data Archive data entry manual
6	Ingest	A study is created in the LISS Data Archive, the dataset and codebook are entered and metadata are added	Operator A	LISS Data Archive data entry manual
7	AIP accept & publish	A quality check on the AIP: accept or reject for publication	Operator B	LISS Data Archive data entry check-and-publish manual
8	Deposit at DANS	The published study is sent to DANS via SWORD	Operator B	LISS Data Archive data entry check-and-publish manual
9	Inform about data publication	Inform Researcher A and LISS Data Archive Coordinator about the published study	Operator B	LISS Data Archive data entry check-and-publish manual
10	DIP check	The published online version of the study is checked	Researcher A; LISS Data Archive Coordinator	LISS Data Archive data entry check-and-publish manual

**) Researcher A may also delegate the preparation of simple SIPs to an Operator. In this case he/she conducts the check (2) him-/herself.*

The data that are archived in the LISS Data Archive are also deposited in the data repository of DANS for long-term preservation. After the Operator has published the AIP, he/she sends it to the DANS repository via a SWORD interface (Simple Web-service Offering Repository Deposit) (8). Finally, the Operator informs the Project Leader and LISS Data Archive Coordinator that the study has been published (9) and asks them to check (10) the Dissemination Information Package (DIP) online.

6.3 Archival Storage

CentERdata has developed its own application for data archiving and dissemination, formerly called Questasy (De Bruijne & Amin, 2010) and nowadays: RepositoryCTRL. This system forms the technical basis of the LISS Data Archive. All studies that are conducted in the LISS panel are disseminated via this system. RepositoryCTRL is a web application built using a PHP framework that uses a relational database to store data. The application supports the longitudinal nature of the LISS panel data and is based on version 3 of the Data Documentation Initiative (DDI). Version 3 of the DDI introduces a life-cycle approach to documenting survey projects and distinguishes between the metadata of questions (data collection) and variables (dataset). While earlier versions of DDI are widely used, no applications could be found in 2007 which applied version 3 to data as complex as the LISS data, prompting CentERdata to build a new archival application[6].

6.4 Data Management and Administration

Within the context of the OAIS model, data management and administration include, among other things, information on the database requests and events, statistical information needed by the archive administration and management, customer profile information and preservation process history information that tracks the migrations of AIPs, including media replacements and AIP transformations.

The LISS Data Archive system can only be accessed after logging in using a personal account. External data users who are logged in gain limited rights to operate within the system, mainly to download the published datasets and to view and edit parts of their personal account information. Data downloads are logged to monitor user statistics.

Internally, CentERdata employees have to register in order to access the system and, depending on the tasks, a specific role is allocated to the employee. The access rights within the system are dependent on this role. Uploads and changes to the studies are logged and can be traced back to the individual employee. Further, two aspects of data management receive special attention: ensuring data authenticity and version control. Data authenticity concerns the means by which the unchanged meaning and value of the data can be ensured and verified. This is related to the management of data versions and media monitoring.

As discussed in the data archiving protocol, all data processing steps are documented in SPSS syntax files when data files are processed for archiving. These syntax files are stored in the same internal directory as the data files, which is a secured environment and only accessible to CentERdata employees. If the data file or documentation needs to be

[6] More information on the application can be found at the following sources:
http://www.ddialliance.org/sites/default/files/QuestasyDocumentingAndDisseminatingLongitudinalDataUsingDDI3.pdf
http://www.centerdata.nl/sites/default/files/bestanden/factsheet_ddi.pdf

corrected after publication, then the following procedure applies. A copy of the original file is modified by the CentERdata Project Leader, using the same documentation procedure as for the first version, that is, using a syntax file that includes the modifications of the data file. Data file names include an extension which stands for the version number and which is updated for the new version. The changes in the data file are documented in a log file. The file naming and version logging of the related documentation such as codebooks follow the same procedure as for the data file.

After this, the Project Leader delivers a new SIP version to the Data Archive Operator together with the log file. The Data Archive Operator enters the new version of the data in the archive and enters information on the modifications in specified AIP fields which are visible for the Data Users. Older versions of data files remain stored in the database, but only the newest version of any file, such as the data file or codebook, is disseminated.

To be able to check data file integrity, MD5 and SHA1 checksums are calculated whenever any file (data files, codebooks, images etc.) is uploaded to the server. It is possible to check the integrity of the data file by recalculating the checksum of the current files on the server and comparing those values with the checksum determined during file upload. These checksums are calculated by the system but not displayed externally by default. Upon request they can be provided to the Data User to check the integrity of the data file he/she has downloaded.

6.5 Access

Access to the LISS data is open to every academic researcher, both in the Netherlands and abroad. Starting from six months after delivery to the original Client Researcher, the data are made available by CentERdata to scientific researchers through the LISS data website[7].

An extensive set of metadata on the whole life-cycle of the research project are accessible to the public on this website, including information on the study objectives, details on data collection, the entire questionnaire and metadata on the data file and individual variables. Information on publications related to the data is also provided when available. Users can search the database in several ways, for instance by keyword search, browsing lists of studies, or a topic or concept-based search.

While access to all metadata is unrestricted, users must register before being able to download actual data. The Data User needs to sign and comply with the rules of the Statement Concerning the Use of Data of the LISS Panel.[8] The signed statement is checked by the Data Archive Operator, who sends the login information by e-mail following access approval. The Data User can then download all published datasets available in the archive. The archive also includes a so-called shopping basket, in which data users can put selections of variables from different LISS panel datasets and automatically obtain the merged data.

To enable users to harvest the metadata of the LISS Data Archive, the repository supports the OAI-PMH protocol[9]. Dublin Core metadata information about published study units can be harvested here. The LISS Data Archive metadata can also be searched by Google.

[7] https://www.lissdata.nl
[8] https://statements.centerdata.nl/liss-panel-data-statement
[9] Base-url is: https://www.oai-pmh.centerdata.nl/lissdata/oai2.php

To increase the visibility and accessibility of the LISS Data Archive studies, the repository is connected to NARCIS[10]. NARCIS is the National Academic Research and Collaborations Information System, and the main national portal for scientific information.

6.6 Preservation Planning

Combell, CentERdata's hosting partner and an ISO 27001 certified data center, is responsible for the operational management of the server park and performs the tasks of the administration functional entity. An incremental backup of the LISS Data Archive application (including data) is made daily to the back-up storage. The hosting partner also performs the updates of the software packages.

To address the risk of file format obsolescence, CentERdata monitors the development of the used software packages (SPSS and STATA) and also stores the data internally as CSV files. In addition, the source code of the surveys, which includes the metadata used to label the data, is stored internally.

In addition to its own system, CentERdata archives the published data files and codebooks in the repository of DANS, including the study level metadata as contained in the EASY system. While these data files are currently accessible for Data Users only via the LISS Data Archive, CentERdata has signed an agreement with DANS to grant access in case the existence of the LISS Data Archive is ever jeopardized. While the primary goal is to guarantee long-term preservation by good management of the LISS Data Archive, this additional measure also aims to create maximum trust in long-term preservation.

Currently, DANS creates persistent identifiers, in this case URNs and DOIs, for the LISS data files when they are ingested by the EASY repository. The DOIs are also made available via the LISS Data Archive. The LISS Data Archive is planning to mint its own DOIs in the near future.

[10] http://www.narcis.info/

7 Data Safeguarding

The LISS Data Archive is stored at the Combell data center. Access to Combell's server rooms is limited by physical and organizational access measures, and the space is fitted with fire protection and power continuity systems.

Functional access to the system is limited to the relevant system administrators. Concerning logical access security, the system is protected by passwords and IP restrictions. Account passwords are always hashed in the database after the first login. At CentERdata, access to the LISS Data Archive application is based on role-specific authorization, and only dedicated IT employees can access the underlying database. Concerning programmable security measures, the configurations and log files of the servers and applications used by CentERdata are periodically checked and updated when necessary. CentERdata uses an internal incident registration system.

The CentERdata application servers are protected by firewalls and measures have been taken to detect any irregularities on the network. The servers are also protected against DDOS attacks.

The security and risk management of the CentERdata research databases, including the LISS Data Archive system, is described in the CentERdata handbook on Information Security and Privacy. This document is based on the ISO 27001 standard.

8 Definitions

AIP
Archival Information Package. Submission Information Package is ingested by the archive and processed into an Archival Information package, which may contain more metadata than the SIP. An AIP conforms to the archive's data formatting and documentation standards (NCDD, 2020; CCSDS, 2012).

DDI
The Data Documentation Initiative (DDI) is an international standard for describing data in the social, behavioral, and economic sciences. The DDI metadata specification supports the entire research data life-cycle. (DDI Alliance, 2020).

DIP
Dissemination Information Package. When a Data User requests information, the archive creates an information package containing the requested data and metadata. (NCDD, 2020; CCSDS, 2012).

OAI-PMH
The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability (Open Archives Initiative, 2020).

OAIS
Open Archival Information System. An archive that has assumed responsibility for preserving information and making it available for its designated community. The term 'Open' means that the system-related recommendations and standards are developed in open forums, not that the access to the archive is unrestricted. (CCSDS, 2012).

SIP
Submission Information Package. The data and the metadata that are sent to the archive by the Data Producer. (NCDD, 2020; CCSDS, 2012).

9 References

CCSDS (2012) Reference Model for an Open Archival Information System (OAIS). Recommended practice, Issue 2. Washington, DC, USA.

De Bruijne M., & Amin A. (2010). Questasy: Online Survey Data Dissemination Using DDI 3. *IASSIST Quarterly*, *33*(1-2), 10. https://doi.org/10.29173/iq645.

DDI Alliance (2020). Website of the DDI Alliance. Information retrieved on October 3, 2020 from https://ddialliance.org/.

NWO (2006). An Advanced Multi-Disciplinary Facility for Measurement and Experimentation in the Social Sciences (MESS). Retrieved on October 3, 2020 from https://www.nwo.nl/projecten/1760102005017-0.

NCDD (2020). Website Netherlands Coalition for Digital Preservation (NCDD). Information retrieved on October 3, 2020 from http://www.ncdd.nl/blog/?page_id=447.

Open Archives Initiative (2020). Website of the Open Archive Initiative. Information retrieved on October 3, 2020 from http://www.openarchives.org/pmh/.

Révilla, M.A. & Saris, W.E. (2010). Comparison of surveys using different modes of data collection: European Social Survey versus LISS Panel. Working paper, New Developments in Survey Methodology – seminar series. Universitat Pompeu Fabra, Spain.

Rekker, R., van der Meer, T., & van der Brug, W. (2020). Dutch Parliamentary Election Study 2017. A comparison of three different survey modes. University of Amsterdam (UvA) and the Dutch Electoral Research Foundation (SKON) from https://www.lissdata.nl/about-panel/composition-and-response.

Scherpenzeel, A.C. (2009). Online interviews and data quality: A multitrait-multimethod study. Working paper, CentERdata, Tilburg University.

VSNU (2005). Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek. Retrieved on October 3, 2020, from https://www.vsnu.nl/en_GB/code-personal-data.