

GESIS' Data Management and Sharing Portal SowiDataNet and Preservation Workflows

Call for Pilot(s)

*Jonas Recker,
GESIS Data Archive*

*CESSDA Training Day
27-28 November 2019, Cologne*

 cessda.eu  @CESSDA_Data



Topics

- Introduction: SowiDataNet|datorium
 - ▶ What is it, who is it for?
- SowiDataNet|datorium and digital preservation
 - ▶ From bitstream preservation to policy-driven digital preservation
- Call for Pilots

I. Introduction: SowiDataNet|datorium

SowiDataNet|datorium

- Data sharing platform for quantitative social science research
 - ▶ Data
 - ▶ Code
 - ▶ Supporting documentation and materials (e.g. stimuli, questionnaires, etc.)
- DSpace 5.7, Vaadin Framework, Elasticsearch

SowiDataNet datorium

Search

[Home](#)

[My submissions](#) ▼

[Terms of Use](#)

[FAQ](#)

[Show all data](#)

[Log in](#)

Welcome to SowiDataNet | datorium

SowiDataNet|datorium is a research data repository for social sciences and economics that enables researchers to easily and securely document, publish and share (quantitative) primary and secondary data. The repository is geared to the needs of the scientific community: a comprehensive [metadata schema](#) provides an opportunity to describe the research data in detail, which is the prerequisite for effective reuse. With the publication in SowiDataNet|datorium, research data become citable because they are given a persistent identifier, namely a "Digital Object Identifier (DOI)". A two-stage review process also guarantees a quality and technical quality check of the data.

SowiDataNet addresses three main target groups:

1. **Researchers in the social and economic sciences** are invited to upload quantitative data, code, metadata and other relevant documents into the repository as data depositors and link them with related publications. By means of selecting an access category and a Creative Commons license depositors can determine under which conditions the data can be accessed.
2. **Research institutes** without their own infrastructure for archiving and publishing their research data, SowiDataNet|datorium suggests a workflow for institutional research data management and offers a reliable platform for data publication.
3. For **data users** interested in re-using data from the social and economic sciences, SowiDataNet|datorium offers convenient search functions and serves as an important point of access.

Contact

SowiDataNet|datorium Team

GESIS Data Archive for the
Social Sciences

Tel.: +49(221)-47694-417
[E-Mail](#)

Who is it for today?

- Individual researchers who want to share data (basic service, free of charge)
- Research institutes or projects who would like to share their data, but do not want to build their own repository (value-added service, fee-based)

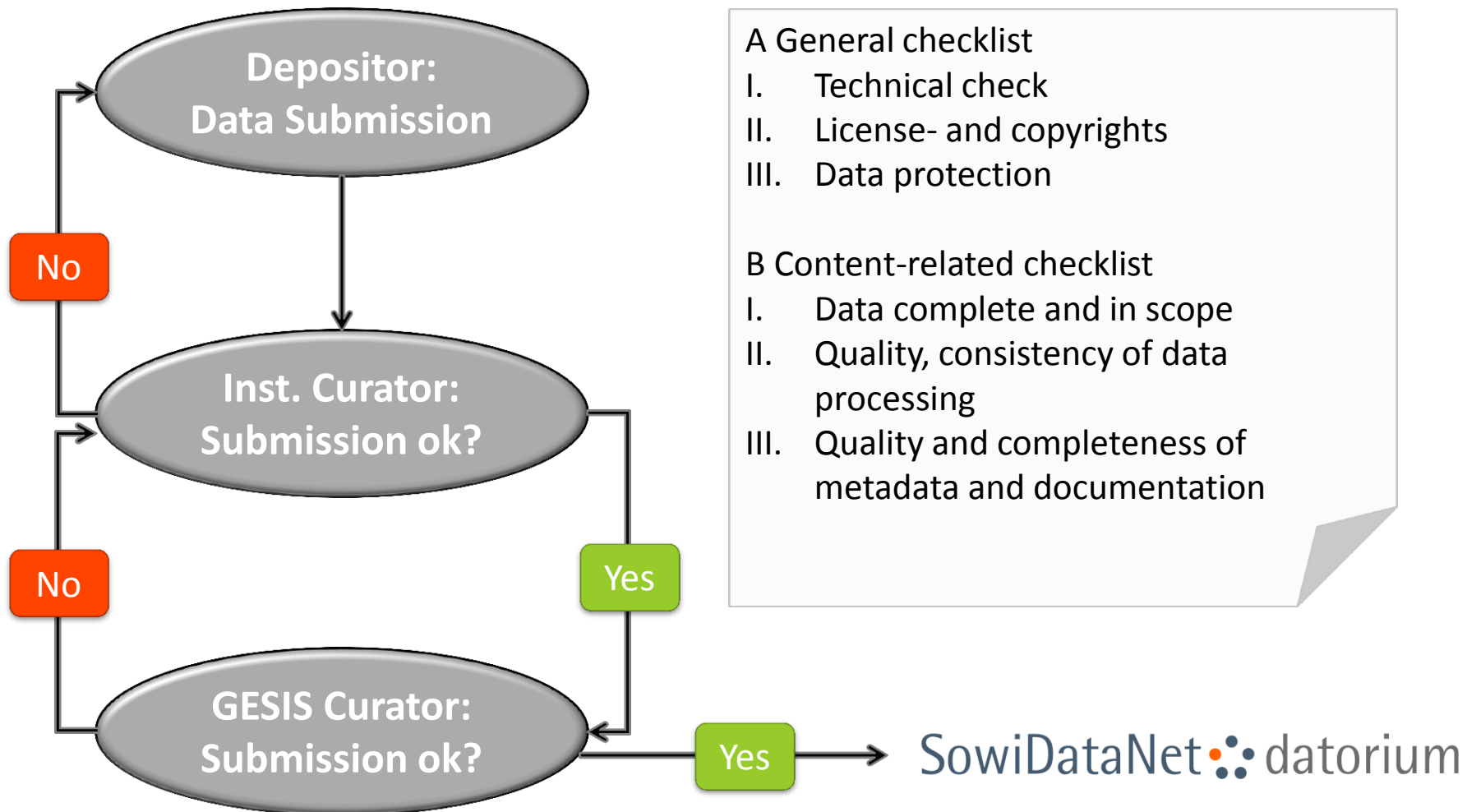
Who could it be for in the future?

- Social Science data archives looking for a trusted Service Provider to host their data and support their digital preservation processes

Features of current institutional service

- Quality assurance for data and documentation
 - ▶ Checklist for curators and two-step curation workflow
 - ▶ Direct support from GESIS curator
- Data discovery and citation
 - ▶ Metadata schema, controlled vocabularies
 - ▶ DDI-registration via da|ra
- Persistence
 - ▶ Guaranteed bitstream 'preservation' for 25 years
- Reporting
 - ▶ Download statistics
- Branding
 - ▶ Integration into customer webpage

Curation workflow









Curator Task Pool







- Home
- Admin
- My submissions ▼
- Terms of Use
- FAQ
- Show all data

Task pool

Owned tasks

Test GESIS Kollektion aus Firefox	Last edited: 2019-11-22	Created by: Submitter Tester	  
This is a test submission	Last edited: 2019-11-22	Created by: Submitter Tester	  

Tasks on the pool

test gesis kollektion	Submitted: 2019-11-08	Created by: Submitter Tester	 
GESIS Test aus CHrome	Submitted: 2019-11-13	Created by: Submitter Tester	 
test (rechte)	Submitted: 2019-11-13	Created by: Jonas Recker	 

Curation backend (I)

This is a test submission

Primary Researcher:	Recker, Jonas
Publisher:	GESIS - Leibniz Institut für Sozialwissenschaften
Publication Year:	2019
Availability:	Free access (without registration)
Other Contributors:	GESIS Datenarchiv für Sozialwissenschaften (Data Curator) GESIS Datenarchiv für Sozialwissenschaften (Distributor)
Project funder:	CESSDA ERIC
Replication Server:	No

Content

Subject Area:	Training, Teaching and Studying, Professional Organizations of Sociology
Thesaurus for the Social Sciences:	basic research
Abstract:	Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.
Internal Comment on Content:	This is an internal comment because I had a question about metadata.

Additional Notes

Type of License:	CC BY 4.0: Attribution (https://creativecommons.org/licenses/by/4.0/deed.de)
------------------	--

Curation backend (2)

Actions you may perform on this task:

This item is a new submission with the version: 1.0.0

View Curator Checklist

If you have reviewed the item and it is suitable for inclusion in the collection, select "Approve".

If you have reviewed the item and found it is not suitable for inclusion in the collection, select "Reject". You will then be asked to enter a message indicating why the item is unsuitable, and whether the submitter should change something and resubmit.

Select this option to change the items metadata.

Return the task to the pool so that another user may perform the task.

Edit version

Checklist

Approve item

Reject item

Edit metadata

Return task to pool

Checklist (excerpt)

*Are combinations of variables possible that would lead to a de-anonymization/re-identification of data subjects (e.g. occupation and geo-information)?

Yes No Not applicable

Could linking with other datasets lead to a de-anonymization of data subjects?

Yes No Not applicable Not checked

*Do the responses to open-ended questions contain sensitive information falling under data protection regulations?

Yes No

Does the submission contain qualitative data?

Yes No Not checked

Is the chosen access class adequate for the data protection requirements?

Yes No Not checked

Further comments on data protection

+ C Content quality assessment

+ D Summary of the quality assessment

Statistics

Downloaded Datasets

a few seconds ago



Value	%	Count
Top values		
1166	10.73%	60
1460	9.30%	52

Top Search Terms

a few seconds ago



Value	%	Count
Top values		
Behr	54.55%	12
gummer	13.64%	3

II. SowiDataNet|datorium

From bitstream preservation to policy-driven digital preservation

Bitstream 'preservation'

- GESIS securely stores the files and metadata as published via SowiDataNet|datorium
 - ▶ Redundant storage in multiple locations
 - ▶ Mix of storage media
 - ▶ Hash sums with regular fixity checks on files
 - ▶ IT Security policy
 - ▶ Recovery tests

From storage to digital preservation

- Digital preservation is more than secure, long-term storage

“Digital Preservation Refers to the series of managed activities necessary to ensure continued access to digital materials [...] beyond the limits of media failure or technological and organisational change.

Long-term preservation - Continued access to digital materials, or at least to the information contained in them, indefinitely.”

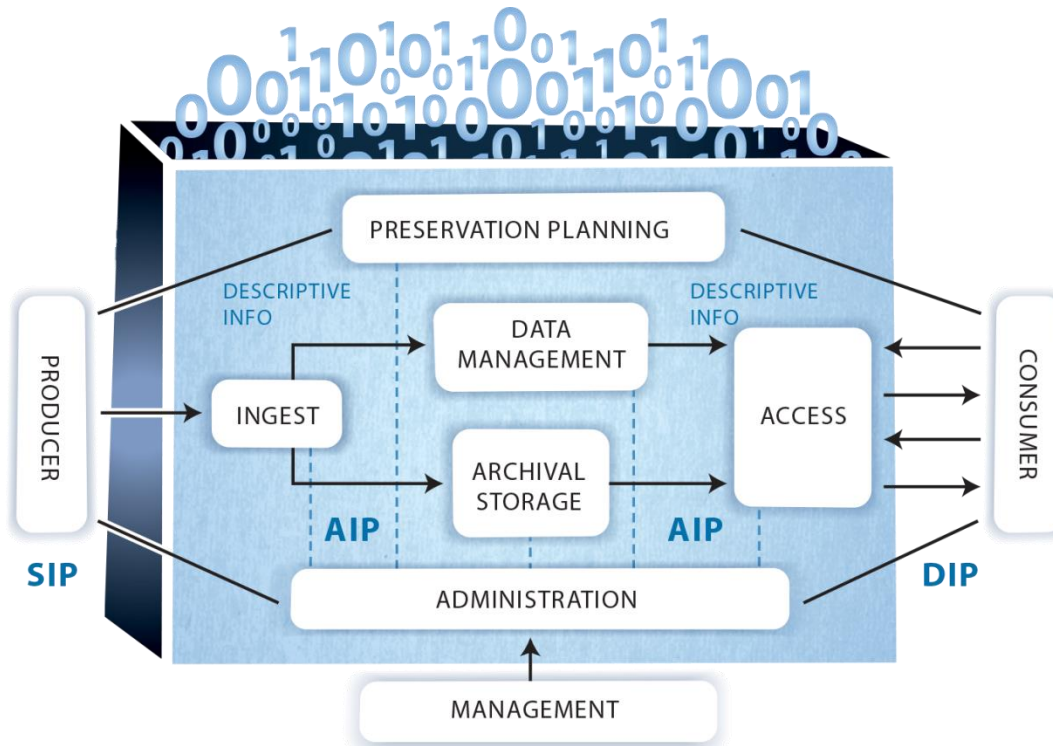
(DPC Handbook, www.dpconline.org/handbook/glossary)

From storage to digital preservation

- SowiDataNet|datorium:
 - ▶ Only minimum of required metadata and documentation
 - ▶ No file format identification / characterization beyond file extension
 - ▶ No file normalization and migration strategy
- Digital preservation requires additional technical measures and has to be supported by policies

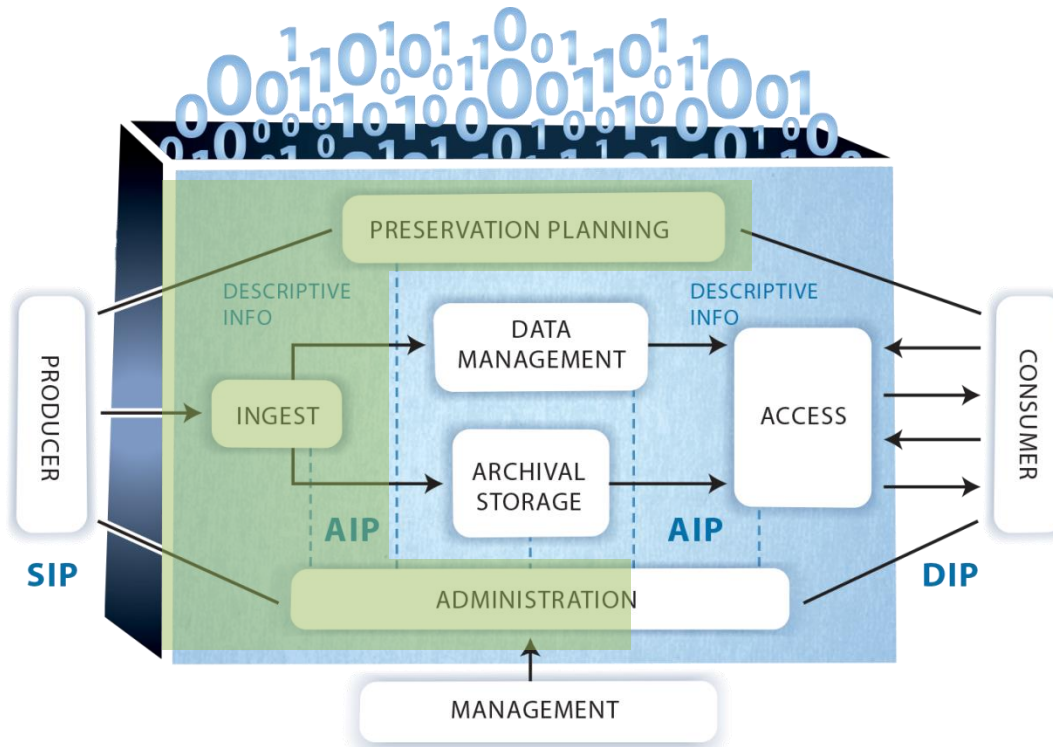
III. Call for Pilots

Preservation in a collaborative setting



Cooperating partners fulfill different digital preservation tasks

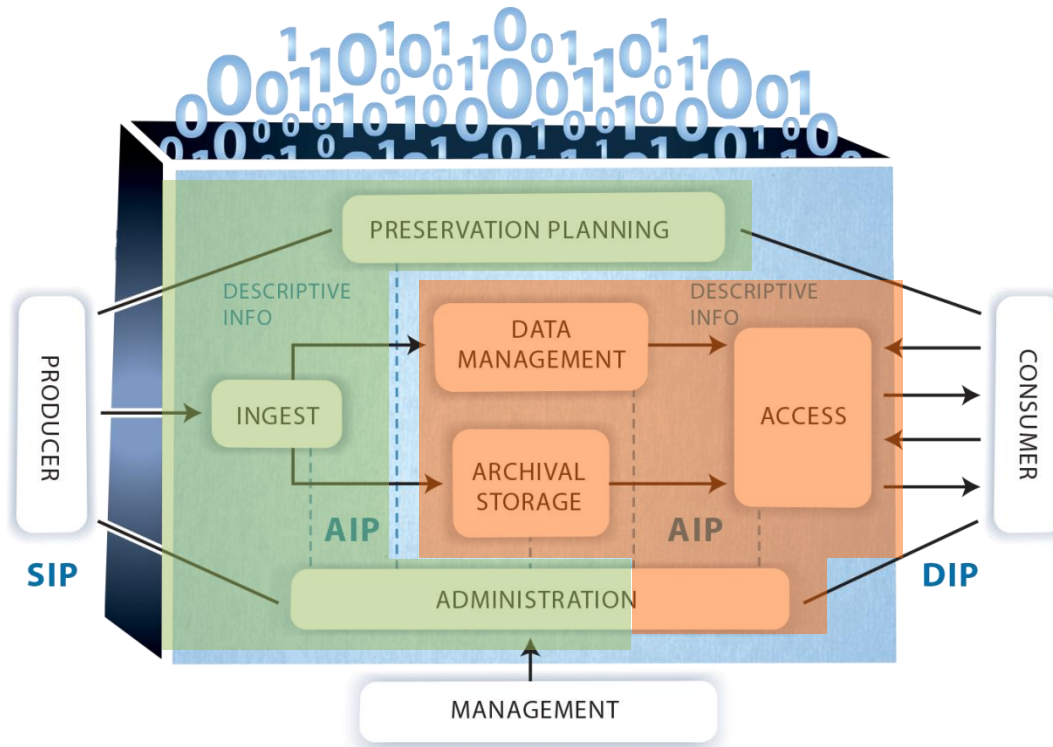
Collaborative setting: Example



Pilot partner (with support from GESIS)

- Ingest, Preservation Planning, Administration
- Communication with consumers and producers on questions relating to the data

Collaborative setting: Example



Pilot partner (with support from GESIS)

- Ingest, Preservation Planning, Administration
- Communication with consumers and producers on questions relating to the data

SowiDataNet|datorium

- Archival Storage, Data Management, Access
- Technical support for producers and consumers

Pilot project

- Define Ingest policies
 - ▶ e.g. file formats, normalization, mandatory metadata and documentation, etc.
- Define and test minimum required metadata
 - ▶ Enough for interpretability?
 - ▶ Provenance metadata
- Preservation policy and plan
- Define workflows, e.g.
 - File format migration
- GESIS waves fees for SowiDataNet|datorium for duration of pilot

Pilot project: Potential goals

- CoreTrustSeal Self-Assessment submitted to CESSDA Trust Group
- Apply for CoreTrustSeal certification in collaborative setting

<https://data.gesis.org/sharing>

Contact: jonas.recker@gesis.org

gesis

Leibniz Institute
for the Social Sciences

Leibniz
Leibniz
Association