

gesis

Leibniz Institute
for the Social Sciences



Information gathering, holding, and re-use in the
Social Sciences - Lessons for IIR?

Daniel Hienert

Keynote at BIIRRR Workshop at CHIIR, March 14, 2022

Introducing myself



- Working as a postdoctoral researcher at the GESIS department Knowledge Technologies for the Social Sciences
- Graduation in Computer Science
- Service: Technical lead of GESIS search
- Research: Active in IIR

Goal of this talk

- This talk aims to present the gathering, holding, and re-use of information in the social sciences.
- The directly following question is if we can use and adopt some of these patterns, lessons, or insights for data management and re-use in the field of interactive information retrieval.

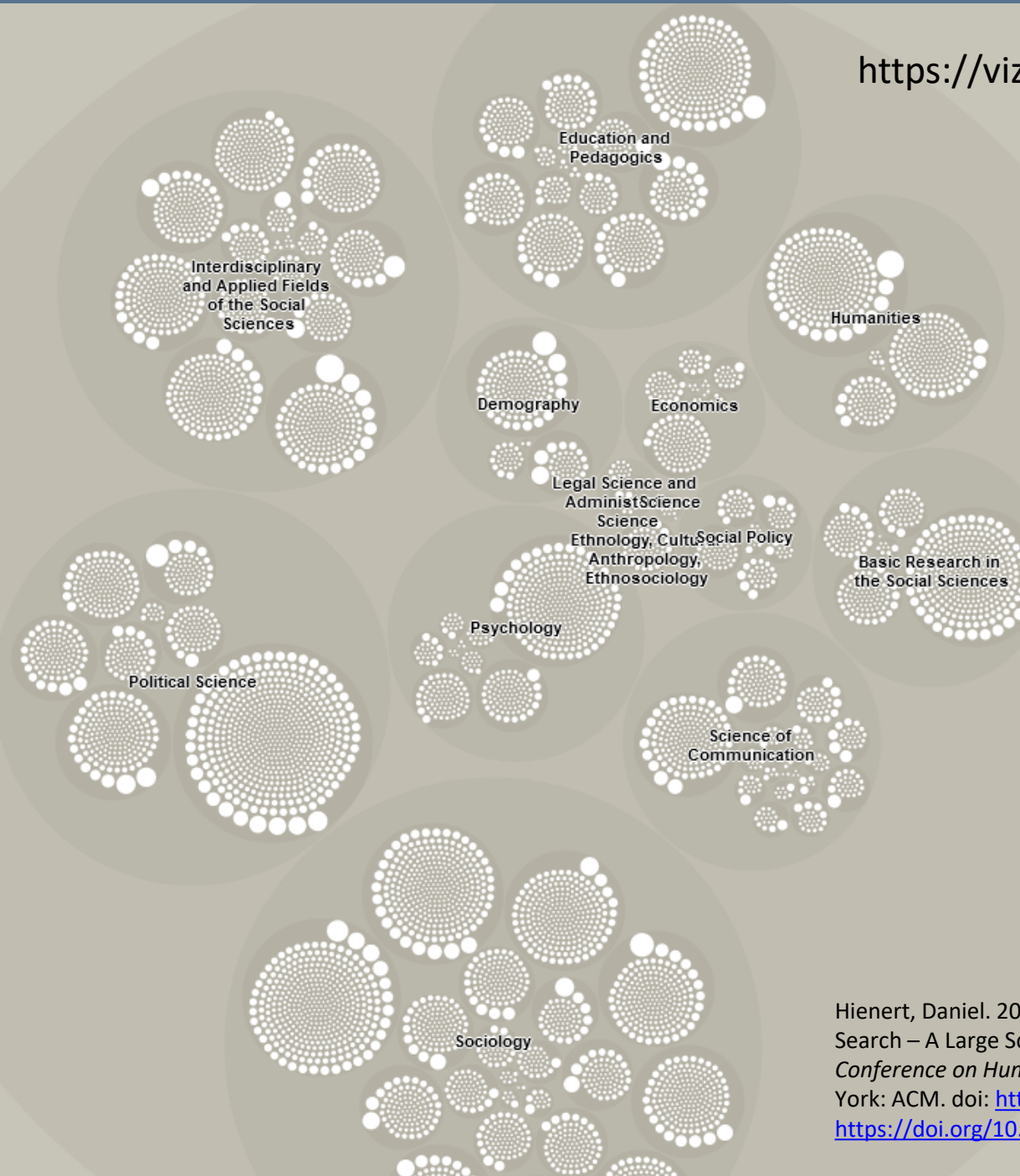
Talk structure

- *Part I:* Addressing the classic approach of information gathering, holding and re-use in the empirical social sciences with survey data.
- *Part II:* Addressing the rather new approach of information management for digital behavioral data.

Part I: Empirical Social Sciences

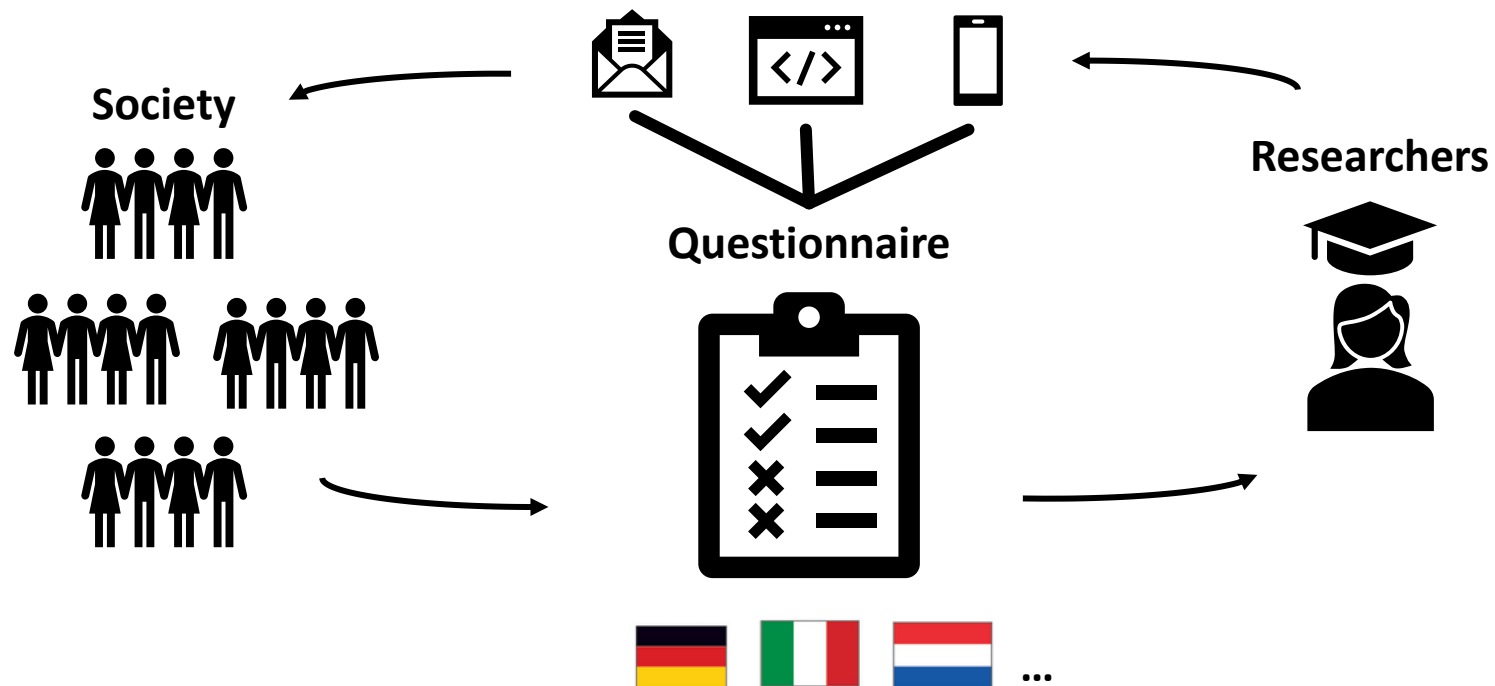
<https://vizgr.org/social-science-topics/>

The Social Sciences



Hienert, Daniel. 2017. "User Interests in German Social Science Literature Search – A Large Scale Log Analysis." In *CHIIR '17 Proceedings of the 2017 Conference on Human Information Interaction and Retrieval*, 7-16. New York: ACM. doi: <http://dx.doi.org/10.1145/3020165.3020168>.
<https://doi.org/10.1145/3020165.3020168>.

Collecting data in the emperical social sciences



Resulting data and information from data collection

- **Study data**
- **Raw data files**
- **Questionnaires**
- **Codebooks** with questions, variables, answers items
- **Other materials** -> Method report, Standards, Classification, Scales, Question texts, Coding
- **Primary and secondary literature**

DDI as a metadata standard for describing these entities

- DDI stands for Data Documentation Initiative
- an international standard for describing the data produced by surveys and other observational methods
- For the social, behavioral, economic, and health sciences
- typically in XML

DDI as a standard for holding these entities

Example of ICPSR Data Catalog Record in DDI 3.1

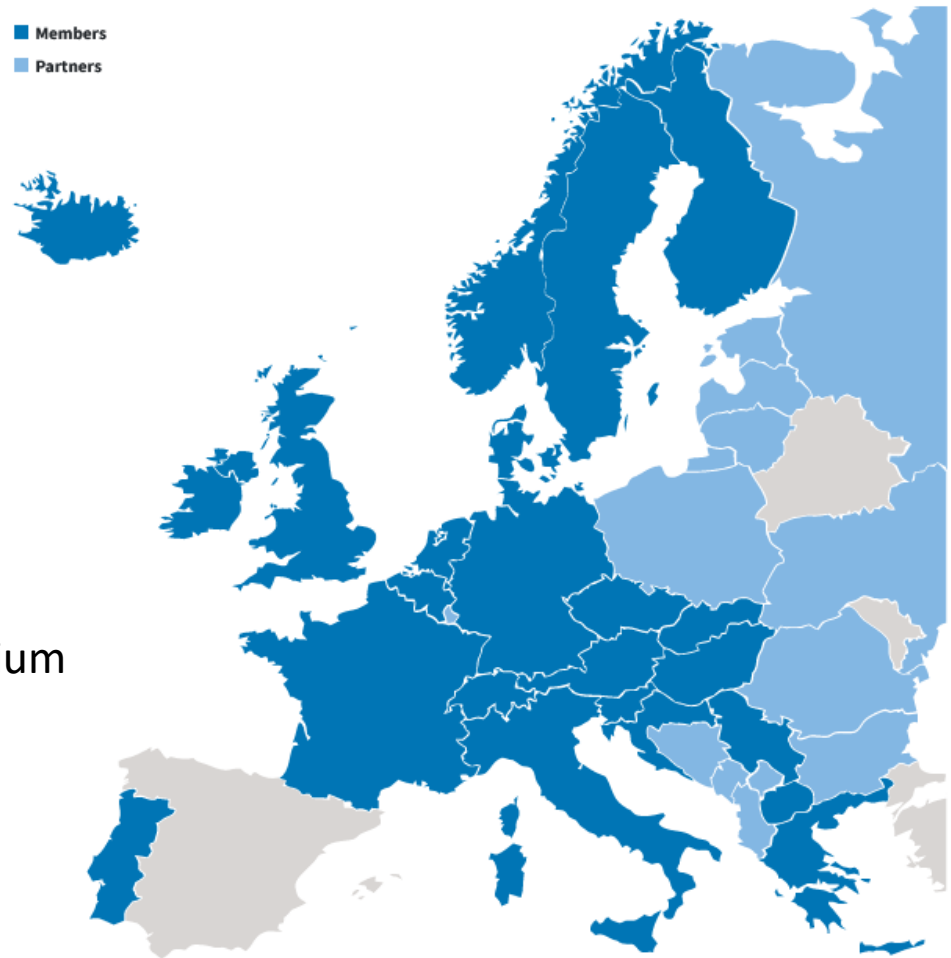
```
<StudyUnit xmlns="ddi:studyunit:3_1" id="StudyUnit36053" versionDate="2015-04-16">
  <Citation xmlns="ddi:reusable:3_1">
    <Title>Cognition and Aging in the USA (CogUSA) 2007-2009</Title>

    <Creator xmlns="ddi:reusable:3_1" affiliation="University of Southern California">McArdle, John</Creator>
    <Creator xmlns="ddi:reusable:3_1" affiliation="University of Michigan">Rodgers, Willard</Creator>
    <Creator xmlns="ddi:reusable:3_1" affiliation="University of Michigan">Willis, Robert</Creator>

    <Publisher>Inter-university Consortium for Political and Social Research</Publisher>
    <Contributor role="distributor">ICPSR</Contributor>
    <PublicationDate>
      <SimpleDate>2015-04-16</SimpleDate>
    </PublicationDate>
    <InternationalIdentifier xmlns="ddi:reusable:3_1" type="ICPSR Number">36053</InternationalIdentifier>
    <InternationalIdentifier xmlns="ddi:reusable:3_1" type="DOI">doi://10.3886/ICPSR36053.v1</InternationalIdentifier>
  </Citation>

  <Abstract isIdentifiable="true" id="Abstract36053">
    <Content xmlns="ddi:reusable:3_1">
      <div xmlns="http://www.w3.org/1999/xhtml" id="Summary36053">Cognition and Aging in the USA (CogUSA) is a national longitudinal
study of cognition, focused on the age-related changes in cognition across cohorts and on the impact of cognition on key health and economic
outcomes. The aim of the CogUSA Study was to evaluate the effectiveness of a variety of tests in assessing cognitive skills on a sample
mirroring the Health and Retirement Survey (HRS) (ICPSR 6854). Data were derived in three waves, with each wave utilizing a variety of
measures. In Waves 1 and 3, these measures included an adaptive number series test. The Woodcock-Johnson (WJ-III) number series test and the
Wechsler Abbreviated Scale of Intelligence (WASI) were used in Wave 2. Waves 1 and 3 were conducted as telephone interviews, while Wave 2 was
conducted as an in-person interview. This collection includes indicators of cognitive abilities and functioning as well as a variety of
demographic, health, and economic variables.
Dataset 1 presents participant demographic information, and Dataset 2 presents scores calculated for the cognitive tests administered to the
respondents in each of the three waves.</div>
    </Content>
  </Abstract>
```

Data archives as institutions documenting and offering this research data



<https://www.cessda.eu/About/Consortium>

Data archives as institutions documenting and offering this research data

▼ Ingest check PLUS

▼ Long-term digital preservation

▲ Documentation on study level

- Preparation of a high quality, qualified study description
- Use of the international standard DDI (Data Documentation Initiative)
- Bilingualism: German, English
- Making the study description available in various formats:
 - online in the desired portal
 - as XML file for import into databases

▼ Data download

How to enable researchers and other end users to search and find all this information

The screenshot shows the gesis website interface. At the top, the gesis logo and navigation links (Login, German, Contact, FAQ, Watchlist) are visible. A search bar contains the text 'za3950'. Below the search bar, a navigation menu includes 'Services', 'Research', and 'Institute'. The main content area displays '462 Hits' and a 'Filter results' section with various dropdown menus (Topic, Person, Year, Geography, Source, Study title, Study group, Collection year, Thematic collection). A 'Sort by:' dropdown is set to 'Relevance'. On the left sidebar, there are buttons for 'Research data (1)', 'Variables (253)', 'Instruments & Tools (0)', 'Publications (115)', 'GESIS Library (86)', and 'GESIS Webpages (7)'. The main results area shows the title 'International Social Survey Programme: Citizenship - ISSP 2004' by Wilson, Shaun; Gibson, Rachel; Hadler, Markus. It includes the GESIS Data Archive, Cologne, ZA3950 Data file Version 1.3.0, and a DOI link. The abstract describes the survey topics. On the right, there is a 'Downloads' section with links for Datasets, Questionnaires, Codebook, and Other documents, and an 'Actions' section with links for Bookmark and Cite. At the bottom, a pagination bar shows '1 - 1 of 1'.

gesis Leibniz Institute for the Social Sciences

Login German Contact FAQ Watchlist (0)

Search za3950

Services Research Institute

462 Hits

Filter results

Topic Person Year Geography Source Study title Study group Collection year Thematic collection show more filters

Sort by: Relevance

Research data (1)

Variables (253)

Instruments & Tools (0)

Publications (115)

GESIS Library (86)

GESIS Webpages (7)

What do you find in this category?

International Social Survey Programme: Citizenship - ISSP 2004

Wilson, Shaun; Gibson, Rachel; Hadler, Markus +

GESIS Data Archive, Cologne. ZA3950 Data file Version 1.3.0, <https://doi.org/10.4232/1.11372>, Date(s) of Data Collection: 03.2003 - 02.2006

Abstract: Citizen and state. Topics: Qualities of a good citizen; attitude toward the right of public meetings for religious extremists, people who want to overthrow the government by force, or people... more

Linked information: Publications (119) Variables (253)

1 - 1 of 1

Downloads

Datasets

Questionnaires

Codebook

Other documents

Actions

Bookmark

Cite

https://search.gesis.org/research_data/ZA3950

Which services are offered to re-use data and tools?

MISSY-Home General Information Find Metadata

You are here: Home Find Metadata EU-LFS Variables over time

National Data

Microcensus (DE)

EU-Data

AES

CIS

EU-LFS

Browse

Matrix

Variables over time

Datasets over Time & Countries

Setups

Materials

Tools

EU-SILC

SES

Variables over time: EU-LFS

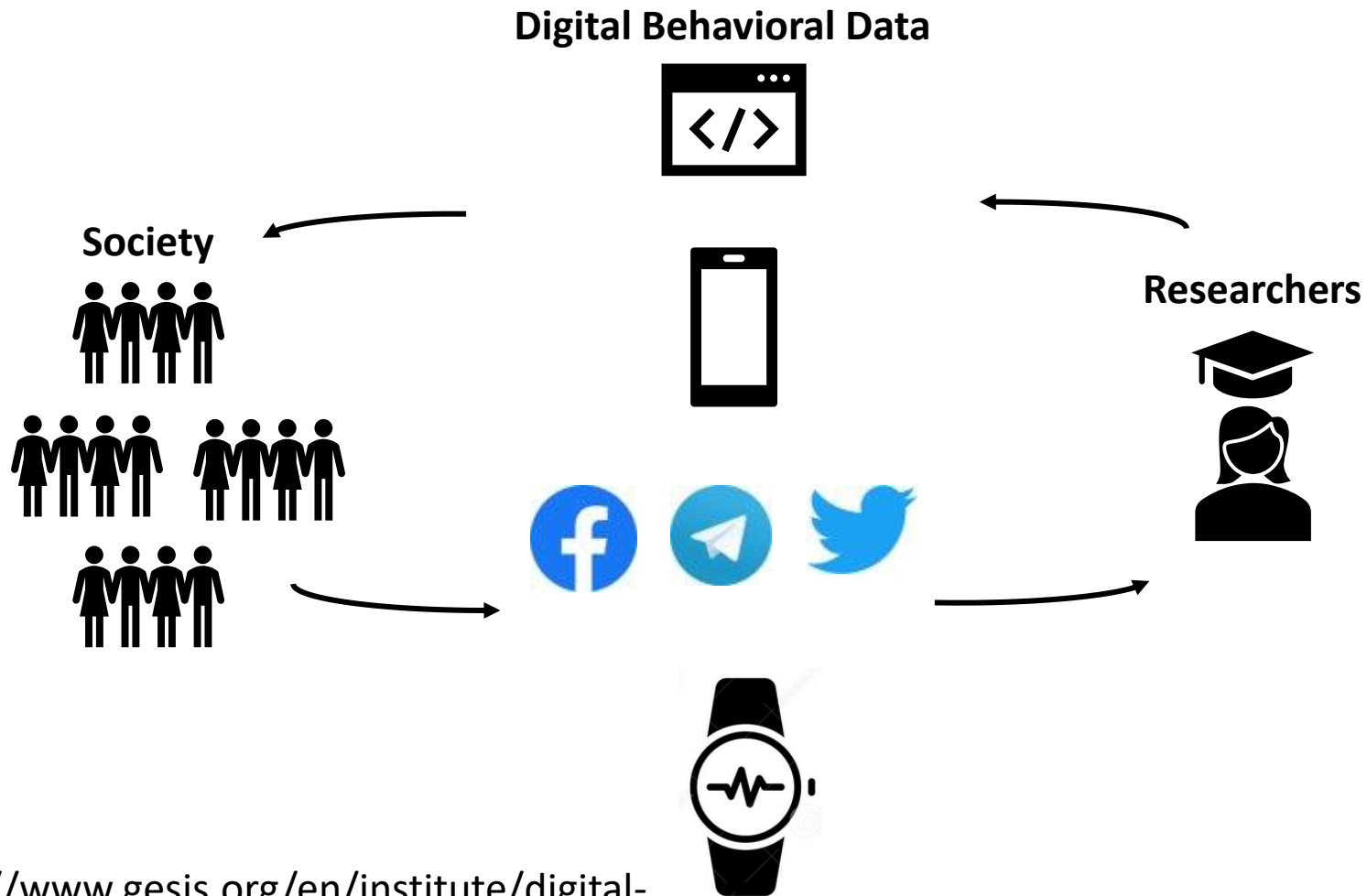
Yearly

Filter Options

	2019	2018	2017	2016	2015
1 Demographic background					
1.1 Basic demographic data					
Sex	SEX	SEX	SEX	SEX	SEX
Age	AGE	AGE	AGE	AGE	AGE
Marital status	MARSTAT	MARSTAT	MARSTAT	MARSTAT	MARSTAT
1.2 Population group					
1.3 Region					
2 Nationality and Migration					
Nationality	NATIONAL	NATIONAL	NATIONAL	NATIONAL	NATIONAL
Years of residence in this country	YEARESID	YEARESID	YEARESID	YEARESID	YEARESID
Age at which person last established their usual residence in the country	AGERESID	AGERESID	AGERESID	AGERESID	AGERESID
Country of residence one year before	COUNTR1Y	COUNTR1Y	COUNTR1Y	COUNTR1Y	COUNTR1Y
Region of residence one year before (NUTS)	REGION1Y	REGION1Y	REGION1Y	REGION1Y	REGION1Y
Country of birth	COUNTRYB	COUNTRYB	COUNTRYB	COUNTRYB	COUNTRYB

Part II: Digital Behavioral Data

Collecting data with digital behavioral data

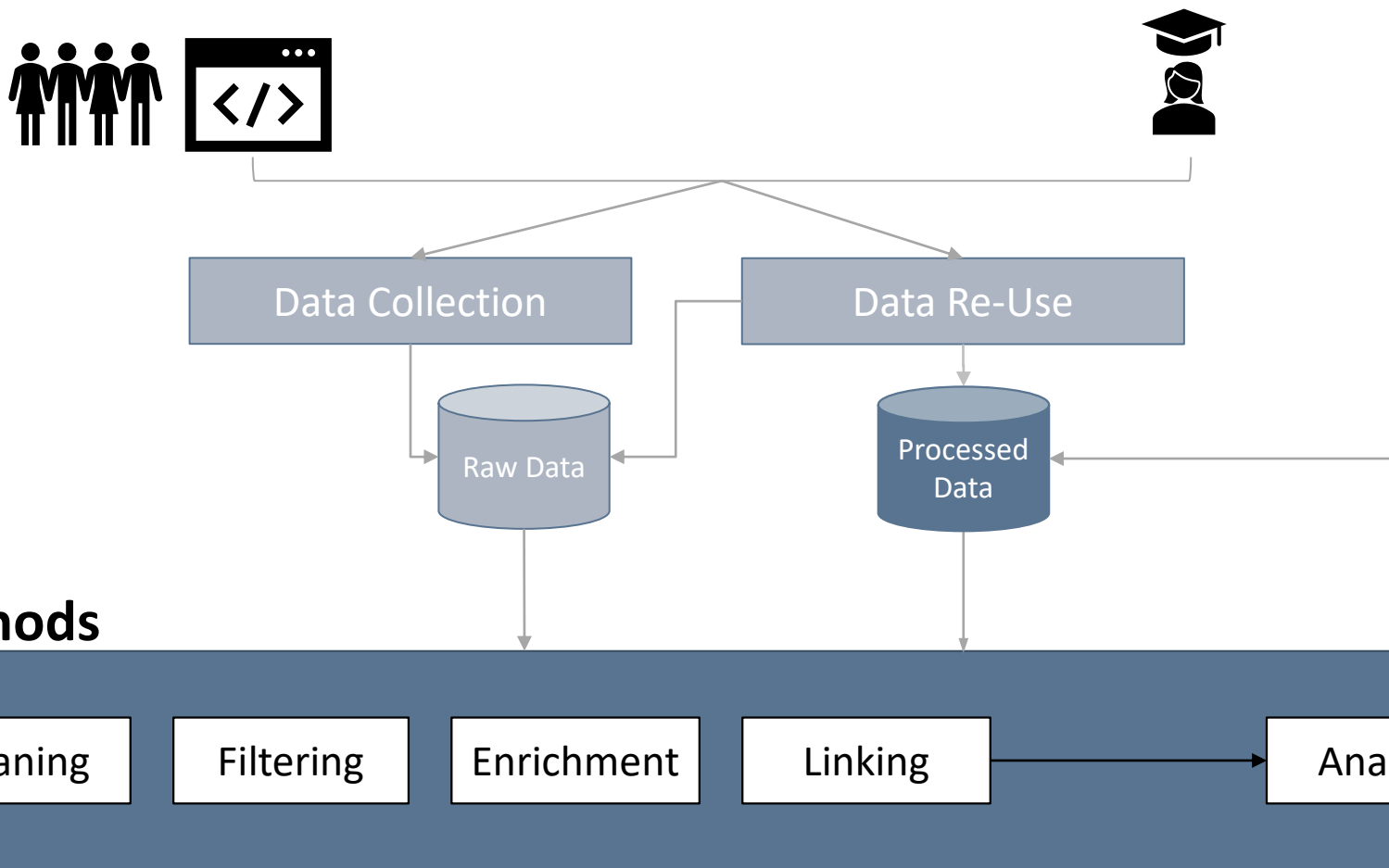


Resulting data and information from data collection

- Study data
- Raw data:
 - Visited Webpages
 - Smartphone data, e.g. little questionnaires
 - Social media data, e.g. liked posts, conversations
 - Sensor data, e.g. location, contacts
- Processed data

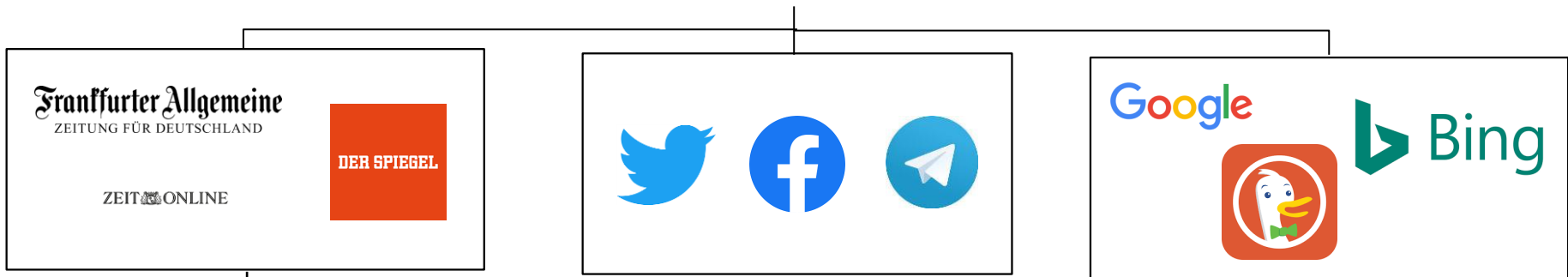
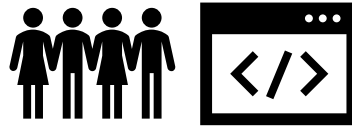
New data – new tools, what is needed for this kind of data?

Method Hub in DBD



New data – new tools, what is needed for this kind of data?

Method Hub in DBD



Analysis: Browsing Behaviour

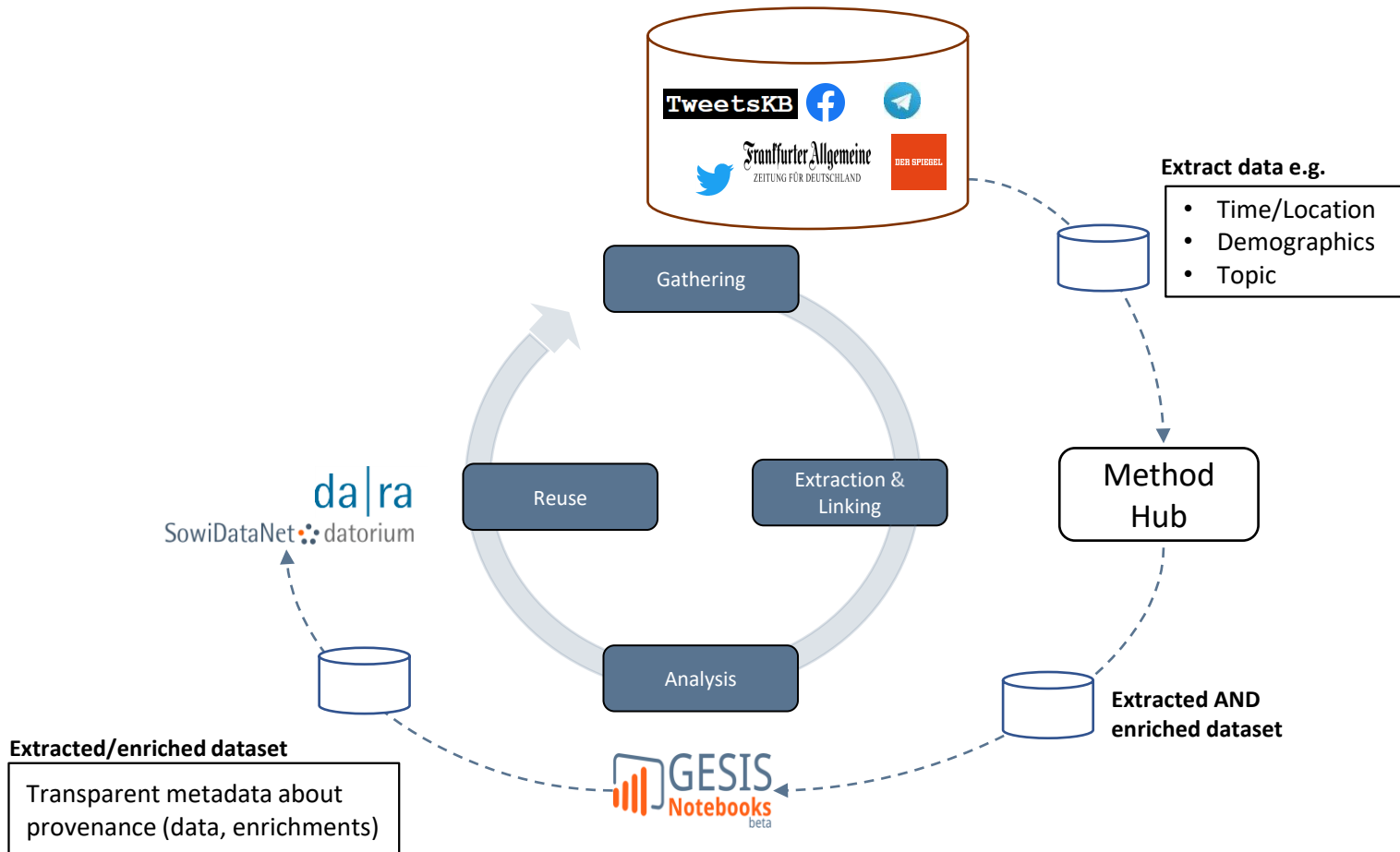
Example Websites

Tasks

Spiegel.de/Politik/top33.html
Spiegel.de/Politik/top14.html
https://www.zeit.de/Wahlrend
https://www.wahl-o-mat.de/
https://www.spd.de/zukunftsprogramm/

- **Topic Classification**
- ..
- Scroll Behavior
- Dwell Time
- Click Analysis
- Mouse Tracking

The Lifecycle with DBD



Conclusion

- The Social science have a well-established model of data gathering, holding, and re-use
- But also have the same new challenges for DBD data

Conclusion

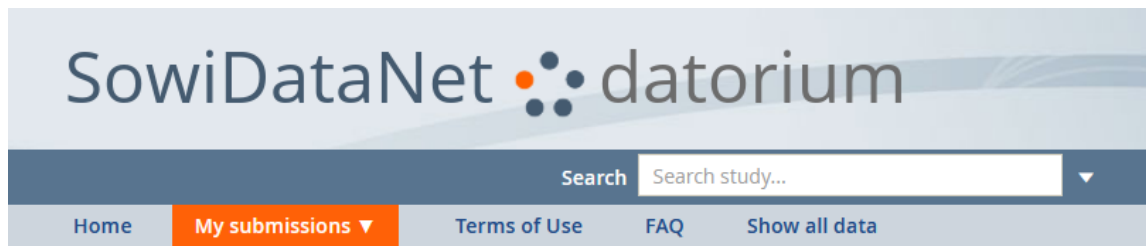
- Essential points are:
 - **Study design** as a basis
 - **Resulting data** from data collection, processing, and research
 - A good **metadata schema** to describe the data
 - **Institutions**, organization, groups, rules, portals to **document** and **hold the data**
 - Portals and search engines to **find the data for re-use**
 - [Organization, portals **helping researchers to re-use and analyzing the data**]



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.

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



Enter your project

Please begin by describing your research project as a whole. Individual files can be documented under "

- ☐ Basic Information
- ☐ Content
- ☐ Methodology
- ☐ Additional Notes

New feature: It is now possible to assign the availability "Free access (without registration)" to supporting materials such as codebooks, questionnaires, or syntax files, even if the availability of the data itself is more restricted.
To activate this option, simply check the box "Set a different availability for this file" after uploading a resource in the area below.

Files (max. 1 GB)

If you wish to upload files bigger than 1 GB, please contact curator2@gesis.org.

Drop files here to upload.

[Click to Upload](#)

[Save](#)

[Review and Submit](#)

References

Hienert, Daniel, Dagmar Kern, Katarina Boland, Benjamin Zapilko, and Peter Mutschke. 2019. "A digital library for research data and related information in the social sciences." In *Proceedings of 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 148-157. Piscataway, NJ: IEEE. doi: <http://dx.doi.org/10.1109/JCDL.2019.00030>.

gesis

Leibniz Institute
for the Social Sciences



GESIS and others infrastructure institutions as providers of other information types related to this research data as publications, scales, questionnaires, syntax files, guidelines, consulting

Services for the Social Sciences

Research projects generally go through a process of multiple phases – the research data cycle. Each individual phase of this cycle requires specific know-how to obtain significant results. GESIS has the know-how and optimizes utilization of this by offering a unique package of services to accompany the entire research data cycle.

To see what GESIS offers please select the current corresponding phase of your research project, or any other phase, and have a look at what we can offer you for your research project:

Phase 1: Planning studies & collecting data

We help you to develop an adequate project design and to ensure the quality of the implementation.

Phase 2: Finding & accessing data

We offer over 6,500 studies for secondary analysis in our data archive.

Phase 3: Processing & analyzing data

Our services support you in the modification, linking and analysis of data.

Phase 4: Archiving & sharing

With us you can archive, register and share your data and publications on a long-term basis.



<https://www.gesis.org/en/services>