Applications of Research Data Management at GESIS Data Archive for the Social Sciences

Recker, Jonas; Zenk-Möltgen, Wolfgang; Mauer, Reiner

Jonas Recker, Wolfgang Zenk-Möltgen and Reiner Mauer

7 Applications of Research Data Management at GESIS Data Archive for the Social Sciences

Abstract: The chapter "Applications of Research Data Management at GESIS Data Archive for the Social Sciences" explores ways in which an archive – i.e. an organization whose work has a strong focus on preservation and dissemination of digital data – can become involved in research data management (RDM). The Data Archive looks back on a long history of working with researchers to make their data re-usable and accessible since 1960. Today it provides support for Research Data Management across the entire data lifecycle by offering a wide range of tools and services tailored to the needs of different types of stakeholders.

The chapter gives an overview of selected tools and services offered in the areas of metadata and data documentation, data preparation, data publication, and long-term preservation. To illustrate how support for research data management plays out in different settings, three case studies for typical scenarios are presented: 1) The European Values Survey (EVS), a large international longitudinal survey studying basic human values across Europe. 2) The German Longitudinal Election Study (GLES), a national survey program with a comprehensive approach to gain insights into the German federal elections. 3) A data center in the health sector which decided to make data originally collected to support policy-making available to research.

1 Introduction

Research data management (RDM) can only be successful if its strategies and procedures reflect the research process and the characteristics of the data to be managed. Thus, while there are generic aspects of data management that do not vary greatly across the disciplines (e.g. back-up strategies, versioning rules, data security), there are also discipline-specific elements.

In this chapter we will consider RDM strategies for social science data used in empirical social research, a discipline which studies social reality and its phenomena. Typical methods of data collection in social research include surveys, interviews, observations, and experiments. In addition, social science researchers often analyze administrative or transactional data that was not

DOI 10.1515/9783110365634-008

generated primarily for research, but originally collected for another purpose (e.g. official statistics, data from public administration, social media data).

Commonly we distinguish between two general “styles of research” (King, Keohane and Verba 1996, 3): quantitative and qualitative. Due to the specialization of the GESIS Data Archive for the Social Sciences (DAS) the focus of this chapter will be on quantitative social science data. Simply put, the latter can be construed as information captured in alphanumeric codes and analyzable using statistical methods, as opposed to textual or audiovisual information in the case of qualitative data.

Among the characteristics of this data relevant to the planning and implementation of RDM measures, the following are of particular importance from our perspective:

- Often the data contains personal information of participants in the research. Collecting, processing, and disseminating this information is subject to ethical and legal constraints, particularly data protection legislation.
- The data (here meaning the sheer numbers expressing the measured values) typically does not “speak for itself”. It cannot be understood without “documentation”, i.e. further information about how and why it was collected and prepared.
- The accessibility and re-use potential of the data can be considerably increased through intensive, high-quality preparation (cleaning, standardization and harmonization of the data; in-depth documentation down to the level of variables; translation into different languages, etc.).

RDM in empirical social research specifically needs to address these aspects to ensure the understandability and usability of the data in and beyond the project. Measures to accomplish this include informed consent, anonymization, strategies for secure data access, study- and variable-level documentation.

Designing and implementing suitable data management strategies can be challenging for researchers. Fortunately, a support infrastructure exists in the social sciences which provides tools and services for data management. In this support infrastructure data archives play an important role, a circumstance that contributes to archives taking over new and different tasks compared with their “traditional” role in the research process.

Thus, historically archives came into play when the “active life” of an artefact or record ended and it was deemed valuable enough to be preserved. This is true for research data as well. Traditionally, the “research sphere” and the “archival sphere” tended not to overlap, with archives only encountering artefacts after the research ended. However, the digital revolution has consequences for research – among them the so-called “data deluge” (see for example

Lord et al. 2004; Marcum and Georg 2010), the trend towards open science, calls for reproducibility and research data management. All these are changing the role that (data) archives play in this context.

In this chapter, we will explore this changing role from the perspective of a German social science data archive. We will do so by focusing on our specific "involvement" with RDM through services and tools developed for the purpose of supporting different types of stakeholders, ranging from individual researchers to projects and survey programs, to other data-holding organizations. This will be followed by three case studies illustrating the ways in which we support these stakeholders in their RDM activities. In these examples, we will focus on data management measures carried out in the European Values Survey (EVS), the German Longitudinal Election Study (GLES), and for a data center in the health sector.

2 The GESIS Data Archive for the Social Sciences

The GESIS Data Archive for the Social Sciences (DAS) was founded in 1960 as Central Archive for Empirical Social Research (*Zentralarchiv für empirische Sozialforschung*), one of the first social science data archives. Today it is a department of GESIS – Leibniz Institute for the Social Sciences, an infrastructure organization offering research support services to the social science community across the entire research data lifecycle (see Figure 7.1).

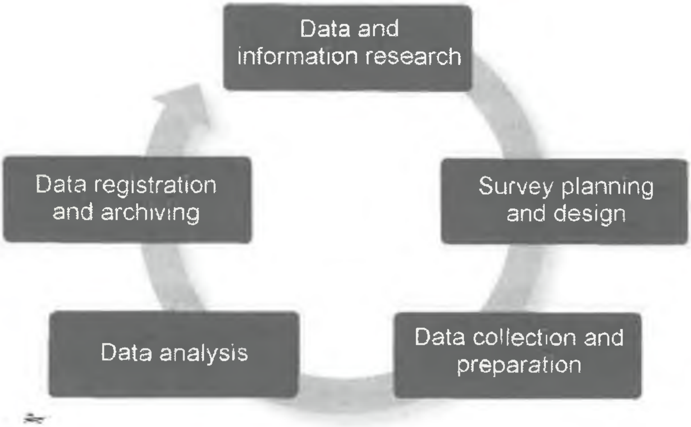
Data and information research
Survey planning and design
Data collection and preparation
Data analysis
Data registration and archiving

Figure 7.1: Research data lifecycle.

Among the services that GESIS offers to the social science community are consultation and support in study planning, data collection and analysis. The Data

Archive in particular is tasked with the curation, registration, dissemination and long-term preservation of data. In accordance with GESIS's statutes[1], the Data Archive collects and disseminates data for social research, with an emphasis on quantitative data that lends itself to the investigation of social change across space and time (see Table 7.1). This includes data sets that were not collected for scientific purposes originally, but are nonetheless of interest for science (e.g. official statistics or data produced in commercial contexts).

While in many ways a "traditional" archive in the sense described above, the GESIS Data Archive is strongly dedicated to partaking in and contributing to the current development towards a more open, reproducible science. Thus our role is no longer merely that of a "safe deposit box" but we are increasingly engaged in research projects long before they end and the data is archived. This involvement entails providing RDM support and data management planning in the form of consultation and training, the development and provision of tools for collaboration, and the preparation, documentation, persistent identification, and publication of data. Moreover, the Data Archive carries out RDM measures as a service for research projects.

Getting involved in the actual research process in these ways has meant reshaping the archive's services (roughly over the course of the past 10 years) and developing new tools and services so that we can offer the support necessary in this changing situation. These are tailored to the needs of different target audiences, including large national or international survey programs or institutional data producers as well as smaller projects using a diversity of methods to generate their data and covering a broad range of topics from the social sciences.

In addition to working directly with researchers and research projects as well as funders, GESIS also collaborates with national and international organizations contributing to the data infrastructure in the social sciences, among them the Consortium of European Social Science Data Archives (CESSDA ERIC), the International Federation of Data Organizations for Social Science (IFDO), or the German Data Forum (RatSWD). The main focus of these cooperative activities is to

- facilitate access to data both nationally and internationally;
- promote open science;
- support and facilitate the development and implementation of policies for data management and sharing (e.g. journal policies, institutional policies, funder policies).

[1] https://www.gesis.org/institut/der-verein/satzung/, accessed 09292017

Table 7.1: GESIS Data Archive overview

Holdings	• more than 5,700 published studies • ~700,000 individual files amounting to 0.5 TB • most common data file formats: SPSS, Stata • data types: quantitative survey data, time series data
Usage (2016)	• ~64,000 datasets downloaded by more than 12,500 distinct users • ~60% international users from 109 countries worldwide

3 GESIS services for research data management

Just like research itself, data management is not a one-off activity but a process that spans the entire lifetime of a data set. Typically, this lifetime does not end with the project in which the data was collected, and hence it is commonly conceptualized as a "data lifecycle" rather than a finite path (see Figure 7.1). In order not to break this cycle, data has to be curated – i.e. managed – continuously, both within and beyond the active project phase, for as long as it is to remain accessible. In this cycle of data collection, use, and re-use for research purposes it is one important task of data archives to take stewardship of this data in between active use phases in research projects (where data is managed by researchers) to keep it accessible, understandable, and hence usable. But, as described above, the Data Archive also offers tools and services supporting the management of data while the active research is still ongoing. All of these efforts are made with the same fundamental goals in mind: 1) Supporting the transparency of research processes and their outputs and 2) enabling re-use of data for new and innovative research.

In the following, we introduce some of our tools and services for data management and publication, targeting different types of users (ranging from individual researchers to large-scale survey programs), and addressing needs that arise in different phases of the research process.

3.1 Data repository service for the social and economic sciences: datorium

Researchers in the social sciences are increasingly faced with funder requirements, institutional, and journal policies that require them to manage and share their research data. Accordingly, there is a growing demand for tools that support the easy description and publication of data – especially in the so-called "long tail" of research. This phrase was coined in 2004 by Chris Anderson to describe the rising importance of niche products – as opposed to economic "hits" – in the Internet economy (Anderson 2004).

In science, this long tail consists of smaller and medium-sized research projects, often carried out by individual scholars or small groups of researchers producing smaller amounts of data on smaller budgets. It is also data underlying publications, which is often derived from bigger data sets and/or data generated by others.

Often, the data generated in the long tail of science is "dark data" – data "not carefully indexed and stored so it becomes nearly invisible to scientists and other potential users and therefore is more likely to remain underutilized and eventually lost" (Heidorn 2008, 280). Among its characteristics Heidorn lists heterogeneity, uniqueness of procedures, lack of accessibility and visibility, infrequent re-use, and low retention (see Heidorn 2008, 288).

However, there is a general consensus that this data is valuable and that adequately managing and sharing it is beneficial to the overall openness, transparency, and replicability of research. There is a long tradition of sharing data in the social sciences, with the first data archives established in the 1960s. But like other disciplines it lacks infrastructure for long-tail data preservation and sharing. Thus, while institutional repositories exist that also accept datasets, and discipline-independent solutions have been implemented in recent years by both commercial and non-commercial providers (e.g. Zenodo, figshare[2]), these solutions tend not to be particularly well-suited to ingest and effectively disseminate social science research data. Issues exist not only with discoverability, but also with data protection (see Archive and Data Management Training Center 2013).

In light of this situation, the GESIS Data Archive developed datorium (https://datorium.gesis.org/), an online tool for the description and sharing of social science research data specifically geared towards the needs of individual researchers and smaller projects in the "long tail" of science. datorium helps to close the gap between data lost on personal hard drives and large datasets from bigger projects that are extensively curated and widely shared (see below). To accomplish this, datorium was built to be rather generic with regard to the diversity of social science data and to be scalable with respect to the volume of data or usage.

Besides guidance concerning data management and tools to manage, prepare and document their data, data producers primarily seek opportunities to publish their data, i.e. to make it available, citable and referenceable – be it to

[2] Zenodo: www.zenodo.org; figshare: www.figshare.com, accessed 09292017

give their research a higher visibility or to meet requirements of funding agencies and journals, or simply because they believe in the ideas and principles of data sharing and open science. Datorium supports these objectives by offering

- standardized but flexible documentation of data with metadata;
- increased visibility and persistent, unambiguous identification of data;
- a secure publication process supported by data curators;
- usage statistics;
- flexible licensing options for controlled access.

The datorium metadata schema has only five mandatory fields to keep the effort involved in publishing data as low as possible for interested researchers, but comprises of over 20 fields all in all. Thus it offers orientation and guidance to those who seek to document their data more comprehensively and in a standardized form.

For persistent identification, each uploaded data set is registered with da|ra, the German DOI® registration agency for data in the social and economic sciences (www.da-ra.de/en/home/), and receives a digital object identifier which can, for example, be resolved through https://doi.org. This also contributes to increasing the visibility of the data shared through datorium as it can be found through the da|ra and the DataCite search (www.datacite.org).

As mentioned above, legal and ethical considerations play an important role in the publication of social science research data. To make sure that data published through datorium does not infringe data protection regulations and is sufficiently anonymized for sharing, all submitted data is reviewed for potential problems by a data curator. In our experience, anonymization is often an issue that needs to be addressed before publication and frequently causes data to be returned to depositors for revision.

Often researchers have misgivings about sharing data due to fearing a loss of control over the data and resulting consequences for future publications or their careers (see for example van Panhuis et al. 2014; Van den Eynden and Bishop 2014). In response to this issue, datorium allows depositors to flexibly determine under which license they would like to share the data and under which conditions it can be accessed by others. The following access options are offered, paired with a recommendation to make data as openly accessible as possible:

- Free Access (without Registration): Unrestricted download of research data for all users.
- Free Access (with Registration): Unrestricted download of research data for all registered users.

- Restricted Access: Access to data only after authorization by the data depositor.
- Embargo: Publication of submitted data will be delayed for a maximum of twelve months.

To facilitate the sharing of data underlying publications, two sociology journals are currently cooperating with GESIS to offer a service for the publication of replication datasets via datorium[3]. An expansion of the datorium service is being developed within the project SowiDataNet[4] (Linne and Zenk-Möltgen 2017). This service contains additional functions for institutions using the repository for their organizational research data management and the publication of research data.

3.2 Ensuring long-term access to research data: digital preservation

The datorium service is designed as a flexible, fairly generic tool which does not require researchers to invest significant amounts of time and resources if they want to share their data. Offering a low-threshold means for publishing data is important if we want to foster the change towards more openness in “data culture”. Yet it should also be clear to us that there is a tradeoff between lowering the threshold – e.g. by making only minimal demands on documentation and metadata, and by not limiting accepted file formats – and the long-term availability and usability of the data. As a consequence, the Data Archive currently guarantees that the bitstream of the data deposited to datorium will remain available for at least ten years. This reflects the fact that neither do the resources exist to preserve everything forever, nor that it is desirable to do so (see for example Whyte and Wilson 2010). However, while this enables researchers to comply with funder and journal requirements for transparency and reproducibility, it does not in itself constitute “long-term preservation”.

Thus, as stated in the Data Archive’s preservation policy, preserving data for long-term access and reuse “involves more than physical preservation of the bitstream by means of back-ups. In contrast to traditional (i.e. analog) preservation, digital preservation must address the effects of rapid technological change [...]. Another important issue in digital preservation is ensuring data can be understood now and in the future” (GESIS – Data Archive for the Social Sciences

[3] www.gesis.org/replikationsserver/home/ (in German), accessed 09292017

[4] https://sowidatanet.de/ (in German), accessed 09292017

2015, 4). To address these challenges, the DAS has established standardized and transparent procedures and workflows subject to regular internal and external review[5]. This includes a set of criteria to determine which data should be curated more extensively, procedures for intensive data checks and basic data preparation, the documentation of data with structured and unstructured metadata by professional data curators (see 3.3), and long-term preservation measures. In the following we describe these measures in more detail.

After submission, data and documentation received undergo extensive (intellectual) checks relating to both technical, structural, and content aspects. This entails control of formats, readability and checks for malware as well as an assessment of the completeness and consistency of the data and documentation. An important aspect of this procedure concerns legal and ethical aspects, potential data protection issues and Intellectual Property Rights in particular (see Jensen 2012). In this phase we work closely with the data producer to clarify any open questions, to correct potential errors, or to address potential issues with the anonymization of the dataset.

To minimize the risks associated with software/file format obsolescence, all submitted data sets are "normalized", that is, they are converted to a standard archival format. This enables us to react efficiently to the threat of obsolete file formats by devising a migration plan and transferring digital objects to a new file format in accordance with this plan as necessary.

The authenticity and integrity of data are further protected by means of strict access controls paired with a comprehensive back-up strategy. Moreover, to detect any changes made to data or documentation files, each digital object is associated with a hash sum – a unique string of characters generated with the help of an algorithm based on the individual combination of bit values (i.e. the ones and zeros) that constitute a given digital object. With the help of a checking program which compares the stored hash sum against the current one, even minimal changes can be detected: thus a single altered bit in a digital object will result in a different hash sum. While this procedure does not prevent (authorized or unauthorized/accidental) changes from occurring, it allows us to detect them quickly and decide whether any action needs to be taken (e.g. recover a file from back-up).

[5] See https://assessment.datasealofapproval.org/assessment_116/seal/html/, accessed 09292017 for the evaluation report for the Data Seal of Approval 2014-2017. The GESIS Data Archive is also currently preparing for the nestor Seal certification http://www.langzeitarchivierung.de/Subsites/nestor/EN/Siegel/siegel_node.htm, accessed 09292017

However, these more technology-centered measures have to be complemented by measures aimed at preserving the “meaning” of the data, which will be discussed in the following section.

3.3 Metadata documentation, data curation, and added-value services

As briefly discussed above, a challenge that we face in curating and preserving social science research data for re-use is to ensure the data can be understood. Without this interpretability, it becomes impossible to grasp what the data actually represents and to evaluate the research findings based on them. Without sufficient information about the research process, interpreting the (potential) meaning of data and drawing conclusions from it is near impossible. As stated elsewhere,

> [t]his includes knowledge about how, when, and why the data was created, and how it was processed and analyzed. For example, to test hypotheses in the quantitative social sciences, researchers must, among others, have information on the composition of the target population, the selection of respondents, the field work and the instrument used, and data cleaning procedures and coding schemes. Preserving the data for re-use requires preserving this contextual information along with the actual “data”, for example, the numeric representation of the measurements. (Recker and Müller 2015, 231)

It is an important aspect of Research Data Management to collect all the information necessary for understanding the data and the results throughout the research process, and to capture this information ideally in a standardized form. As Sundgren explains,

> [i]n order to increase chances that receivers of data interpret the data in the way that was intended by the sender of the data, we may extend the data messages with metadata, data that describe and explain the meaning of the communicated data. Since the metadata are themselves data, they also need to be interpreted by the receivers, and these interpretations are of course also subject to errors and uncertainties. However, the metadata introduce some redundancy into the communicated messages, and hence hopefully decrease the variation and errors in the interpretations. (Sundgren 2011, 2)

This metadata can take different forms. Specifically, it can be either “structured” or “semi-/unstructured”. The latter is often referred to as “documentation” in the social sciences. Structured metadata tends to be standardized, machine-readable, and often makes use of keywords and controlled vocabularies. Documentation – contextual material generated in the course of the research project – tends to consist of running text, which is human readable but less standardized.

Typically, data in the social sciences is described with both structured and unstructured metadata.

Much of the contextual information required to understand data has to be captured during the research process. At this point it is of secondary importance whether the data is described with structured or unstructured metadata. What matters is that the information is there and can be used later to create structured metadata from it, which fulfills three main purposes:

- Sharing and re-use: standardized, structured metadata can easily be exchanged between different tools and re-used in more than one project.
- Access: discovery and initial assessment of a data set in terms of its re-use value for a certain research interest (study-level metadata).
- Enhancement: by adding in-depth, searchable description down to the level of individual variables the re-use of the data can be enriched.

In the social sciences, the most sophisticated standard for structured metadata is DDI (Data Documentation Initiative). It began to emerge in the 1990s, three decades after the establishment of social science data archives such as the Roper Center and ICPSR (USA), the Central Archive for Empirical Social Research (Germany), the Norwegian Social Science Data Service, or the UK Data Archive.[6] Its development since then reflects a) the need to create transparency about the process of data collection and preparation, b) the desire to support the discovery and citation of existing data sets for secondary use, and c) the effort to enable machine-actionable documentation of data throughout the research data lifecycle.

Currently, two versions of DDI are commonly used: DDI Codebook (most recent version: 2.5) and the considerably more extensive DDI Lifecycle (most recent version: 3.2). The two versions follow a different logic. Thus DDI Codebook focuses on the description of the completed data set, whereas DDI Lifecycle makes the structured collection of information throughout (and about) all phases of the research data lifecycle possible.

The DAS services concerning documentation of data are twofold and occur in different phases of the research process:

1) We (retrospectively) describe data that is submitted to the archive for dissemination after completion of the project: this includes structured and semi-/unstructured information such as the descriptive metadata listed in Table 2, and in-depth information about the instrument, data collection and preparation in the form of a codebook or methods report.[7]

[6] www.ddialliance.org/what/history.html, accessed 09292017. See also Rasmussen 2013.

[7] See, for example, the catalog entry and documentation for the German Family Panel (pairfam), at http://dx.doi.org/10.4232/pairfam.5678.7.0.0.

For this purpose, the most important metadata standards applied at GESIS are DDI and DataCite. The latter is a generic standard that can be employed to describe data across all disciplines (https://schema.datacite.org/) and which makes it possible to provide a basic description of our research data supporting the identification and citation of these resources by means of a Digital Object Identifier (DOI). On the basis of these standards, additional metadata schemas were developed for the GESIS Data Catalogue DBK (see 3.4 below), datorium, or da|ra and implemented into the systems we use for the management and curation of data.

To provide researchers in the social sciences with structured information that enables them to assess the re-use potential of a given data set in their own research, the generic DataCite standard was expanded for the da|ra system to allow a much more sophisticated documentation of data from the social sciences (Helbig et al. 2014).

Table 7.2: Descriptive metadata for study-level description

Information type	Example
Bibliographic information	Study number, study title, current version, date of collection, principal investigators, authoring institution, persistent identifier (DOI), topic classification, etc.
Information on study content	Abstract, topics, demographic information, etc.
Information on methodology	Geographic coverage, selection method, mode of data collection, data collector, etc.
Information on data and available documents	Number of units and variables in the dataset, analysis system(s) used, access modalities, available files, etc.
Information on errata and versions	Errata in current version, versions list, etc.
Further information[8]	Comparable or related studies, related publications and study groups, etc.

2) For selected, large-scale surveys we offer value-added documentation. Thus, when it comes to the use of data which other researchers have collected, lots of detailed methodological and content related questions arise. Deep and comprehensive added-value documentation, covering the study design, the data collection, the data cleaning, and the resulting dataset can help answer those questions. At GESIS we provide a detailed Variable Report

[8] Further structural and administrative metadata is added for internal use. Among others, this provides relevant technical and provenance information (see Zenk-Möltgen and Habbel 2012 for detailed information on the metadata schema [in German]).

for value-added studies, containing extensive information on the study, methodology, questionnaire, and dataset.[9] In addition, original documents like project reports, methodological reports, codebooks, and original language questionnaires are made available to secondary users.

3.4 Tools for research data management and collaboration

Research Data Management during the active phase of the research comprises of many different tasks to be completed by researchers. The available tools for this play an important role for the effectiveness and adequateness of the work undertaken. At GESIS we provide tools for different tasks which complement tools from other providers. In practice, there will often be a mix of standard software tools and special purpose tools, as well as a mix of free, open source software and commercial products. It can be a challenge for researchers to identify suitable tools and to evaluate if they are appropriate to fulfill the tasks in the project to be conducted in a way that matches their needs. To help with this issue, in the following we briefly describe some of the tools provided by GESIS to support Research Data Management activities in the areas of documentation, data preparation, and collaboration.

3.4.1 Documentation

Tools for the documentation of research projects in the social sciences provided by GESIS are

- the Data Catalogue DBK for study-level information,
- the Dataset Documentation Manager (DSDM) for dataset information, and
- the CodebookExplorer for more complex information about dataset collections.

The Data Catalogue DBK fulfills several functions pertaining to access to and documentation of research data. It is the main entry point for researchers looking for data to re-use at the GESIS Data Archive.[10] It holds detailed descriptions of all archived studies and provides functions to carry out simple or advanced searches, access data and additional documents, and find information about the content, methodology, data and documents, versions, publications etc. Data

[9] See, for example, GESIS-Variable Reports Nr. 2013/9: ALLBUS/GGSS: German General Social Survey – Cumulation 1980–2012, Study No. 4580, version: 1.0.0, doi:10.4232/1.11952.

[10] https://dbk.gesis.org/dbksearch/, accessed 09292017

citation is facilitated by means of the provided DOI names and the bibliographic metadata. The datasets for each study are usually available via direct download or ordering for registered users. In some cases special usage agreements have to be accepted to access the data.

The metadata schema of the DBK is compliant to the DDI Codebook standard (Zenk-Möltgen and Habbel 2012) and can be accessed in a variety of formats including Dublin Core, DDI Codebook and DDI Lifecycle via an OAI-PMH interface (a standard protocol for harvesting metadata: https://www.openarchives.org/pmh/).[11] GESIS licenses the DBK metadata under a Creative Commons CC0 1.0 Universal Public Domain Dedication[12] but encourages users to give attribution to the metadata sources wherever possible.

A specific version of the DBK software is available as DBKfree[13] for redistribution and modification under the terms of the GNU General Public License. This allows other institutions to re-use the software for building their own catalogs.

One component of the DBK software is DBKEdit, a web-based multi-user interface for editing and publishing study-level metadata (Zenk-Möltgen 2013). To allow researchers to create simple study-level metadata themselves, GESIS provides DBKForm – a HTML/JavaScript application that can run locally in any browser and saves the metadata in DDI Codebook format. The GESIS Data Archive uses DBKForm[14] to obtain study-level metadata for the ingest process of studies directly from researchers in the correct format.

The desktop application Dataset Documentation Manager DSDM has been available for some years (Zenk-Möltgen 2006) and facilitates the documentation of simple and complex datasets on variable level according to the international metadata standard DDI (Mühlbauer, Kratz, and Solanes Ros 2014).[15] With DSDM, metadata from SPSS or DDI can be imported and enhanced with variable and survey question documentation. The results can be exported into different dissemination formats for publication and long-term preservation. DSDM is compatible with the DDI Codebook standard and provides functions to export the metadata in DDI Codebook XML format. To facilitate the production of Enhanced Publications, some of the metadata can be exported into DDI Lifecycle format (Granda et al. 2009). Several studies may be documented within one local database supporting the re-use of documented questions across all studies. Original

[11] https://dbk.gesis.org/dbkoai/?verb=Identify, accessed 09292017

[12] https://creativecommons.org/publicdomain/zero/1.0/, accessed 09292017

[13] https://dbk.gesis.org/DBKfree2.0/, accessed 09292017

[14] DBKForm can be downloaded at https://dbk.gesis.org/dbkform/

[15] The DSDM software is provided as freeware at https://dbk.gesis.org/software/dsdm.asp?db=E, accessed 09292017

language documentation is supported by an approach of allowing two languages to be used for documentation (usually English and an original language) with a broad support of character sets. A direct interface to export the compiled documentation into CodebookExplorer databases is provided.

The software CodebookExplorer was developed to provide comprehensive information about collections of complex datasets.[16] It supports the collection of study- and dataset-level metadata as well as the creation of hierarchical category systems for topics, trends, question scales and other information. Within a given database users may search for keywords in study or variable descriptions, compare question and answer texts between studies or languages, and conduct analyses like frequencies, crosstabs, descriptive statistics, or even comparative analyses. Users may also create their own category system of variables and create reports from within the program.

Examples of CodebookExplorer databases are the Continuity Guide of the German Election Studies (Zenk-Möltgen and Mochmann 2000), the European Values Study 1999/2000 (Luijkx, Brislinger, and Zenk-Möltgen 2003), or the collection Childhood, Adolescence and Becoming an Adult (Reitzle and Brislinger 2005).

3.4.2 Data preparation

In contrast to the tools for documentation of existing datasets, data preparation tools support the manipulation of the raw data in the data files. Tools for data preparation used in the social sciences and specifically at GESIS are mostly commercial software packages like SPSS[17] and Stata[18], but also the open-source software R[19]. In addition, there are recommendations for syntax code documentation to be used in conjunction with these tools, e.g. from the project VFU (see below) or from the datorium service (https://datorium.gesis.org/). Also, a collection of small scale programs is used, e.g. to create and validate file hash sums like MD5.

The question of how to document syntax files was specifically considered within the "VFU soeb 3" project (Jensen 2014; Jensen and Schweers 2016). This project focused on developing and implementing a virtual research environment

[16] The CodebookExplorer is available as freeware and can be downloaded at https://dbk.gesis.org/software/cbe.asp?db=E, accessed 09292017

[17] https://www.ibm.com/analytics/us/en/technology/spss/, accessed 09292017

[18] www.stata.com/, accessed 09292017

[19] www.r-project.org/, accessed 09292017

(VRE) for the joint research project soeb3 ("Berichterstattung zur sozioökonomischen Entwicklung Deutschlands"), reporting on Socio-Economic development in Germany and funded by the German Federal Ministry of Education and Research (BMBF). A very general metadata schema was developed within this project to connect the three areas of study description, data usage, and syntax files. The output is a very detailed and structured metadata schema for syntax files, documenting file properties such as name, authors, software used, description, terms of use, as well as the newly created variable names and labels, and analysis methods used.

For datorium (see 3.1) there are specific recommendations on how to document all deposited files to accomplish certain levels of reproducibility (Ebel 2016). In addition to the general metadata schema for datorium (Zenk-Möltgen and Linne 2013), which covers the standardized descriptions to be provided when researchers upload data into the repository, published guidelines on when and how to document the data preparation help to ensure that other researchers can use the data to replicate the original results or make secondary use of the data.

These recommendations (Ebel 2016) cover many details relevant to the practical work with datasets, including the organization of the work in a way that helps other users understand what has been done with the data. The main idea behind the suggested practice is that the original dataset from the data collection phase remains unchanged, and all steps of the data preparation and cleaning are carried out with the help of the syntax files. This results in an analysis data file on which all steps of the analysis can be performed by consecutive syntax files. A master syntax file can start the whole process and enables other users to re-conduct all steps until the resulting analysis tables are calculated. In addition, the syntax code should be documented with information about what is done and why, rather than about how it is done. There is a general recommendation to use one of the many code style guides available, e.g. the Google R Style Guide[20].

3.4.3 Collaboration

At the GESIS DAS, tools for collaboration are mostly used within data collection projects to facilitate common work across organizational borders. Examples are the project portal for the European Values Study (see 4.1) and the development of a project portal within the EU-funded project SERISS.

The first project portal for the EVS was built to enable direct communication and file exchange between the project partners in all countries conducting the

[20] https://google.github.io/styleguide/Rguide.xml, accessed 09292017

survey in 2008, the Methodology group, and the GESIS Data Archive. It was built using the general purpose portal software DotNetNuke (DNN)[21].

Each partner was given an individual account associated with specific rights. The portal administrator created all pages and defined the user roles, e.g. for upload or downloads of survey data by specific countries. In this way, the administrators could give information about the data processing to all partners at the same time. Thus the stages of the data preparation by all countries and the resulting data files became a transparent process. This helped all project members to better understand the overall progress of the project (Brislinger and Zenk-Möltgen 2011; Brislinger and Zenk-Möltgen 2012; Brislinger et al. 2011). In addition, notification emails were sent out by the system to inform important actors about certain deliverables of the data preparation process. While the advantage of such a system is that it can provide data security and transparent and comprehensive information about all stages of the work, the disadvantage lies in a substantive amount of work that needs to be invested for preparation and maintenance of such a site.

Based on the experiences and lessons learned with the EVS portal, a new collaboration portal is currently being developed within the SERISS[22] project, funded under the EU Horizon 2020 program to support several survey programs: "Survey Project Management Portal SMAP" (Task 4.4 in work package 4). SMAP will have components for managing workflows, documents, communication, and roles, and for monitoring and quality assurance. It will use sustainable software to build a more generic tool for large scale and distributed survey research projects like the EVS. After being tested successfully for the next EVS wave in 2017, the tool will be made configurable with project portal templates for other survey programs and will be released as an open-source solution.

4 Examples of Research Data Management at GESIS

4.1 An international survey program: the European Values Study

The European Values Study (EVS)[23] is a large-scale, cross-national, and longitudinal survey research program on basic human values across Europe. Its main

[21] http://www.dnnsoftware.com/, accessed 09292017. Note that a similar portal was built using DNN to support the project planning for the EVS wave in 2017.

[22] Synergies for Europe's Research Infrastructures in the Social Sciences, http://seriss.eu/, accessed 09292017

[23] www.europeanvaluesstudy.eu/, accessed 09292017

focus is to gather insights into beliefs, values and opinions of citizens, paying special attention to topics such as life, family, work, religion, politics and society. It started in 1981 with 16 participating countries, and continued with waves in 1990, 1999, and 2008 when altogether 47 countries/regions took part. Currently, the 2017 wave is being prepared. The EVS is managed by the Council of Program Directors of the participating countries. The preparation of the questionnaire is carried out by a dedicated Theory Group and quality issues are addressed by the Methodology Group.

The GESIS Data Archive is the official archive of the European Values Study and provides access to data and documentation files. Moreover, the Data Archive has a close cooperation with the EVS at Tilburg University, the teams in the EVS member countries, and the scientific community for secondary research to ensure a high-quality data management for best results. Since the EVS project plans data collection activities every nine years, experiences and results of previous waves can be incorporated into newer waves very well (see Figure 7.2). However, at the same time it is a challenge to keep future EVS waves in mind when they are nearly a decade away. For the 2017 wave, a close collaboration between Tilburg University, EVS member countries, survey organizations, and the GESIS Data Archive has allowed incorporation of considerations on which kind of data and information should be produced, which review procedures are necessary to improve their quality, and which types of outcomes may facilitate user-friendly access (Brislinger et al. 2011).

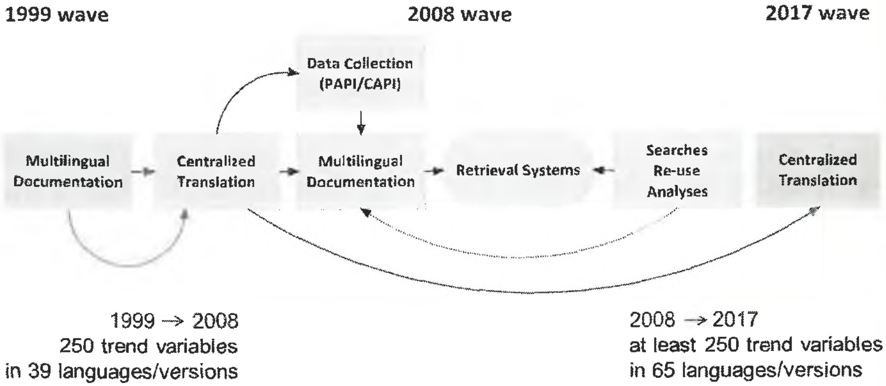

Figure 7.2: Re-use of question items within and between EVS waves.

The preparation of an international, multilingual survey like this needs a sophisticated workflow to ensure the quality of the data. EVS partners designed guidelines documenting data and metadata management principles to be applied during all phases of the survey lifecycle (see Figure 7.3). Special attention was given to documentation of methodology and survey questions. Metadata was captured to describe the master questionnaire in English, covering so-called "trend questions" repeated from previous waves, changes from previous waves, and newly created questions. In addition, translations for the field questionnaires in different languages were documented both for repeated and new questions. Reviews of the questionnaires and back-translations were also noted as modifications between waves and between countries sharing the same language. Metadata for subsequent workflow steps was captured, e.g. for fieldwork monitoring, interviewer training, data verification, data processing, and cleaning.

One of the challenges faced in the project was managing the metadata in a way that enables re-use in later project phases. Questions that arise are whether the software will still work in the future, if the knowledge about processes will be available, or if the documentation will be understandable and comprehensive.

Some of those challenges can be met by using well-defined and documented standards for metadata, since this preserves the knowledge and creates independence from specific software tools. Thus, using the DDI standard enabled the project to start a new wave re-using the question documentation from previous waves. Brislinger and Zenk-Möltgen (2009; 2011; 2012) show how this was done between the waves from 1999 and 2008 as well as the plan for the wave in 2017.

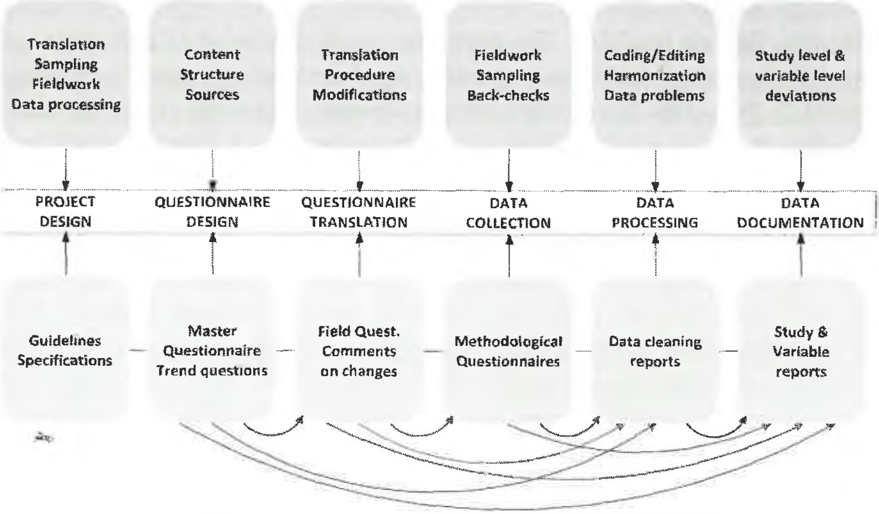

Figure 7.3: Metadata generated across the survey lifecycle.

The resulting collection comprises of national datasets for one wave as well as an integrated dataset for each wave and an integrated longitudinal dataset for all four waves. The project created a detailed plan for releasing data and documentation on different portals, and also for long-term preservation within the GESIS Data Archive. Data and metadata can be accessed for secondary use via the GESIS Data Catalogue DBK[24]. An extended search and browse function also including the original language documentation is provided by the Variable Overview[25] as well as by the Study Overview[26] for the EVS. The English documentation for EVS datasets can also be accessed through ZACAT, the GESIS portal for high quality metadata, where the data can also be analyzed online.[27] A general documentation on available data and metadata is provided on the GESIS website[28], in addition to the comprehensive information on the EVS project website.

4.2 A national level project: the German Longitudinal Elections Study

The German Longitudinal Election Study (GLES)[29] is a project funded by the German Research Foundation (DFG), started for the 2009 federal election and next two elections in Germany. GLES is the largest and most ambitious election study held in Germany to date, following previous election studies which had to apply for funding individually every year. The GLES is directed by six principal investigators from different universities together with the German Society for Electoral Studies (DGfW)[30]. The long time span covered by GLES makes it possible to research attitudes and behavior of respondents over an extended period of time. This is an important improvement compared to the previous situation in which for each national election a separate funding proposal had to be submitted (Schmitt-Beck et al. 2010). Since the GLES has a comprehensive approach and encompasses cross-sectional surveys, short and long-term panels, a candidate survey, TV debate analysis, and media content analyses, research into the electoral process in Germany can be done in a broader way than ever.[31]

[24] https://dbk.gesis.org/dbksearch/GDesc2.asp?no=0009, accessed 09292017

[25] https://dbk.gesis.org/EVS/Variables/, accessed 09292017

[26] https://info1.gesis.org/EVS/Studies/, accessed 09292017

[27] https://zacat.gesis.org/webview/

[28] https://www.gesis.org/en/services/data-analysis/international-survey-programs/european-values-study, accessed 09292017

[29] www.gesis.org/en/elections-home/gles/, accessed 09292017

[30] www.dgfw.info/en/, accessed 09292017

[31] http://gles.eu/wordpress/english/design/, accessed 09292017

The role of the GESIS Data Archive for GLES is to prepare the datasets, provide long-term preservation, and make data and documentation quickly available to the user community. The self-conception of the GLES as a "public data project" (Weßels et al. 2014, 319) leads to the requirement that the release time for data should be short to give all researchers equally fast access to the data. Also, the data collection methods and procedures should be transparent and well-documented in detail, e.g. response rates, field events or non-response issues (Blumenberg, Roßmann, and Gummer 2013). Given the many survey components of the GLES, this makes providing high-quality and comprehensive documentation a special challenge.

Data and documentation of the different GLES components are provided within nearly one hundred datasets up to now, and they can be accessed through the GESIS Data Catalogue DBK.[32] Because of the short timeframe for releasing the data to the public, the project established an internal workflow to provide well-curated SPSS and Stata datasets and to create the documentation as pdf documents, e.g. study descriptions, questionnaires, codebooks, and additional material. These documents are also provided for download via the DBK.

In contrast to the EVS, the metadata for the GLES studies is currently not provided within other portals. However, there are projects underway at GESIS to establish a question database for the GLES studies[33] and to create a question editor for future components (Blumenberg, Klas, and Zenk-Möltgen 2015). Both portals are built on standardized DDI Lifecycle documentation from GLES, derived from project-specific databases and documents available so far. The project can build upon previous work undertaken to align the workflow of the GESIS Data Archive with the DDI Lifecycle standard (Mühlbauer 2014; Mühlbauer, Linne, and Zenk-Möltgen 2010). The idea is to create general purpose software components and release them with an open-source license so that they can also be applied in other survey programs.

4.3 Tailored services for data centers

Over the past decade, growing awareness of the value and importance of research data has led a growing number of data-producing academic institutions, government agencies and public authorities, as well as long-term projects to explore how to increase the accessibility and visibility of their data. However, delivering data and services to end users does not belong to the day-to-day tasks

[32] https://dbk.gesis.org/dbksearch/GDESC2.asp?no=0011, accessed 09292017

[33] https://gles.gesis.org/?al=en, accessed 09292017

of public authorities or public research institutions. Thus they usually do not have the appropriate infrastructure or expertise at hand to offer the required services. What is more, data is organized and prepared in a way that serves the respective organizational purposes and core objectives, e.g. to provide an empirical basis for policy-making. The needs of academic researchers doing secondary analyses are not in the focus of their work.

Projects, on the other hand, are by nature temporary enterprises. They typically concentrate on building infrastructure components to collect and manage data and to exchange it between project partners. Distributing data and delivering support for secondary users is often beyond their capabilities. Accordingly, the development described above resulted in increasing requests from institutional users and long-term projects for the GESIS Data Archive (DAS) to support them with different data management and archiving services. Various institutional settings and requirements lead to different forms of cooperation.

One example for such cooperation is work carried out with a government-funded organization in the health sector which for many years has produced – and continues to produce – significant amounts of data highly relevant to social research. To fully exploit the value of data, collaborative work with DAS was carried out to explore possibilities for publishing the data for re-use by external researchers.

The data in question was primarily collected and prepared for internal use in the organization to fulfill its mandate – e.g. to write reports for the government or the general public, or to carry out public campaigns. In the past this data was accordingly not systematically prepared and documented for use in research by third parties. Among the issues that had to be addressed to make the data sharable are the following:

- data contains undocumented variables;
- data has not been completely anonymized;
- data collected over long periods of time is very heterogeneous, making it difficult to create an integrated dataset (e.g. for comparisons over time).

These are partly a result of the fact that internal data management lacked standardized processes and workflows, as well as standards for data preparation and documentation. This is not surprising, as those responsible for the data are experts in their respective fields but not professional data managers, and because the focus of data preparation was on the needs of the organization rather than on enabling external researchers to answer entirely new research questions with the help of the data.

In the pilot project with the organization in question, the GESIS DAS used six person months to provide support and services in two main areas:

- data preparation and documentation for archiving and re-use;
- development of standards and tools.

During the project, ten datasets were handed over to DAS for preparation, anonymization and documentation. For each dataset, a Public and a Scientific Use File was created and the data was ingested into the archive for long-term preservation and dissemination.

To support the organization in managing and preparing data that will be collected in the future, DAS first established the status quo of workflows as well as of the existing collection of data. We then responded with the development of standards and guidelines for data preparation and documentation, and provided tools in the form of syntaxes for data preparation and checking.

One of the concerns of the organization was that offering data for re-use would at the same time create a demand for a user support service. As data dissemination is not one of the core tasks of this organization, and accordingly there are no dedicated resources for this task, DAS disseminates the data and offers basic user support through its own help desk. Only if users have substantive questions relating to the content or meaning of the data they are referred to the data producer.

Experiences with this distribution of work are very positive so that this model will be used for the dissemination of future data collected and prepared by this organization in accordance with the guidelines and standards provided by DAS. This not only benefits the research community, but also the data producing organization in that the standardized and improved data preparation and documentation procedures help to increase the usability of the data for the organization's core tasks.

Currently, a follow-up project is being planned, which will include the preparation of additional existing datasets and the creation of guidelines and recommendations for contracts with commercial service providers. These providers often manage and carry out the data collection for big or complex surveys in the social sciences and complete data preparation and documentation tasks according to the specifications made by the respective client. In this manner, the cost of preparing data for sharing and archiving can be reduced further in the long term.

This is only one example of how DAS cooperates with other data producers and data holding organizations. As services can be offered in a tailored and flexible fashion, other modes of cooperation are possible as well. These differ in the extent to which DAS is involved in different lifecycle steps, such as data preparation, dissemination, and preservation. For example, DAS cooperates with projects and data centers for whom it offers only digital preservation services, while preparation and dissemination of the data is done by the data producer; alternatively it offers preservation and dissemination services, including on-site use of sensitive data in the Data Archive's Secure Data Center and the organization of "Meet the data" workshops for researchers interested in re-using the data. In this way, together with the cooperating partners we can ensure that data is adequately managed throughout the lifecycle while at the same time not duplicating infrastructure or "re-inventing the wheel".

5 Conclusion

The social science community, and the scientific community at large, is in the midst of profound change towards increased openness and data sharing. As part of this process, we are witnessing a growing awareness of the importance of good data management as a means to transparent and replicable research. Although research data management and data management plans have not become a fully integral, routine element of the research process in the social sciences yet, we have made considerable progress towards this goal over the past years. In Germany, we are seeing a highly increased demand for training in RDM, also driven by the fact that funders, universities, and other research organizations have begun to address the question of data sharing and responsible data management in policies and funding requirements.

For organizations such as the GESIS Data Archive this is a great opportunity to actively shape this process and support the community in moving towards more openness and sustainability. The development of services and tools to help researchers face data management challenges is at the core of what GESIS and similar infrastructure organizations were established to do.

To meet this demand, the GESIS Data Archive has developed the tools and services to support long-term preservation and responsible re-use of data discussed above. As the demand for support and services increases, one challenge we are currently facing is the considerable heterogeneity that we see in the needs of research projects and their data. It is due to this heterogeneity that there are no "one size fits all" solutions for the tasks and challenges discussed in this chapter. Thus there will always be a need for individual approaches and solutions. While some standardization will certainly occur as the community agrees on a standard set of tools, policies, and requirements, we have not yet reached this stage in Germany. This means that we – and the community as a whole – will have to find a balance between generic, standardized procedures and custom-made, tailored solutions. As part of this process, the Data Archive is currently in the process of restructuring its offers into a modularized portfolio. This makes it possible to offer both standardized bundles and highly individual combinations of services, depending on the needs of the respective data producer.

References

Anderson, Chris. 2004. "The Long Tail." *Wired*. http://www.wired.com/2004/10/tail/. Accessed 09292017.

Archive and Data Management Training Center. 2013. "Self-Archiving Platforms and Data Verification." *Archive and Data Management Training Center Blog*. https://admtic.wordpress.com/2013/11/12/self-archiving-platforms-and-data-verification/, accessed 09292017.

Blumenberg, Manuela, Claus-Peter Klas, and Wolfgang Zenk-Möltgen. 2015. "Implementing DDI-Lifecycle for Data Collection within the German GLES Project." In *EDDI15 – 7th Annual European DDI User Conference, Kopenhagen, December 2-3, 2015*. http://www.eddi-conferences.eu/ocs/index.php/eddi/eddi15/paper/view/217, accessed 09292017.

Blumenberg, Manuela S., Joss Roßmann, and Tobias Gummer. 2013. "Bericht zur Datenqualität der GLES 2009." 2013|14. Technical Reports. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2013/TechnicalReport_2013_14.pdf, accessed 09292017.

Brislinger, Evelyn, Emile DeNijsBik, Karoline Harzenetter, Kristina Hauser, Jara Kampmann, Dafina Kurti, Ruud Luijkx, et al. 2011. "European Values Study. EVS 2008 Project and Data Management." 2011|14. Technical Reports. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2011/TechnicalReport_2011-14.pdf, accessed 09292017.

Brislinger, Evelyn, and Wolfgang Zenk-Möltgen. 2011. "Findings of the Original Language Documentation for European Values Study (EVS) 2008." In *IASSIST 2011 Conference, Vancouver, BC, May 31-June 3, 2011*. http://www.iassistdata.org/conferences/2011/presentation/2856, accessed 09292017.

Brislinger, Evelyn, and Wolfgang Zenk-Möltgen. 2012. "Re-Using the Structured Metadata of the European Values Survey." In *RC33 Eighth International Conference on Social Science Methodology, Sydney, July 9-13, 2012*.

Ebel, Thomas. 2016. "Einreichung von Syntaxen in Datorium (Replikationsserver)." Köln. https://www.gesis.org/fileadmin/upload/Replikationsserver/Einreichung_Syntaxen_2016-02-29.pdf, accessed 09292017.

GESIS – Data Archive for the Social Sciences. 2015. "Digital Preservation Policy Principles of Digital Preservation at the Data Archive for the Social Sciences." http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/DAS_Preservation_Policy_eng_1.4.8.pdf, accessed 09292017.

Granda, Peter, Joachim Wackerow, Meinhard Moschner, Wolfgang Zenk-Möltgen, and Mary Vardigan. 2009. "Managing the Metadata Lifecycle. The Future of DDI at GESIS and ICPSR." In *IASSIST/IFDO Conference, Tampere, Finland, 26-29 May, 2009*. http://www.fsd.uta.fi/iassist2009/presentations/D1_Granda.ppt, accessed 09292017.

Heidorn, P. Bryan. 2008. "Shedding Light on the Dark Data in the Long Tail of Science." *Library Trends* 57 (2): 280–299. doi:10.1353/lib.0.0036.

Helbig, Kerstin, Brigitte Hausstein, Ute Koch, Jana Meichsner, and Andreas Oskar Kempf. 2014. "Da|ra Metadata Schema. Version 3.1." 2014|17. Gesis Technical Reports. doi:10.4232/10.mdsdoc.3.1.

Jensen, Uwe. 2012. "Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten." 2012/07. GESIS Technical Reports. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf, accessed 09292017.

Jensen, Uwe. 2014. "Metadata Requirements to Document Data Analyses and Syntax Files in a Virtual Research Environment (VRE) -The Use Case Soeb 3." In *EDDI14 – 6th Annual European DDI User Conference, London, UK, December 2–3, 2014*. http://www.eddi-conferences.eu/ocs/index.php/eddi/eddi14/paper/download/140/112, accessed 09292017.

Jensen, Uwe, and Stefan Schweers. 2016. "The Extended Metadata Schema of the VRE soeb3." 2016|06. GESIS Papers. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_papers/2016/GESIS-Papers_2016-06.pdf, accessed 09292017.

King, Gary, Robert O. Keohane, and Sidney Verba. 1996. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.

Linne, Monika, and Wolfgang Zenk-Möltgen. 2017. "Strengthening institutional data management and promoting data sharing in the social and economic sciences." *Liber Quarterly* (27,1): 58–72. doi:10.18352/lq.10195.

Lord, Philip, Alison Macdonald, Liz Lyon, and David Giaretta. 2004. "From Data Deluge to Data Curation." In *Third All Hands Meeting (AHM 2004) Nottingham, August 31st–September 3rd, 2004*. http://www.allhands.org.uk/2004/proceedings/papers/150.pdf, accessed 09292017.

Luijkx, Ruud, Evelyn Brislinger, and Wolfgang Zenk-Möltgen. 2003. "European Values Study 1999/2000 – A Third Wave: Data, Documentation and Database on CD-ROM." *ZA-Information* (52): 171–182. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-52.pdf, accessed 09292017.

Marcum, Deanna B., and Gerald Georg. 2010. *The Data Deluge: Can Libraries Cope with E-Science?* Santa Barbara, California: Libraries Unlimited.

Mühlbauer, Alexander. 2014. "Working with the STARDAT DDI-Lifecycle Library." In *EDDI14 – 6th Annual European DDI User Conference, London, UK, December 2–3, 2014*. http://www.eddi-conferences.eu/ocs/index.php/eddi/eddi14/paper/view/141, accessed 09292017.

Mühlbauer, Alexander, Sophia Kratz, and Ivet Solanes Ros. 2014. "Einführung in Die Variablendokumentation mit DSDM 2.6." Archivreihe (Internal Working Paper).

Mühlbauer, Alexander, Monika Linne, and Wolfgang Zenk-Möltgen. 2010. "The STARDAT Project. Integrating DDI Tools at the GESIS Data Archive." In *2nd Annual European DDI Users Group Meeting, Utrecht, 08.12.2010*. http://www.iza.org/conference_files/eddi10, accessed 09292017.

Rasmussen, K. B. (2013). Social science metadata and the foundations of the DDI. IASSIST Quarterly, 37(1–4), 28–35. http://www.iassistdata.org/sites/default/files/iqvol371_4_rasmussen.pdf, accessed 09292017.

Recker, Astrid, and Stefan Müller. 2015. "Preserving the Essence: Identifying the Significant Properties of Social Science Research Data." *New Review of Information Networking* 20 (1–2): 229–235. doi:10.1080/13614576.2015.1110404.

Reitzle, Matthias and Evelyn Brislinger. 2005. "Childhood, Adolescence and Becoming an Adult. Data – Documents – Databank on CD-ROM Now Available in a Bilingual (German/English) Version." *ZA-Information / Zentralarchiv Für Empirische Sozialforschung* 57: 132–139.

Schmitt-Beck, Rüdiger, Hans Rattinger, Sigrid Roßteutscher, and Bernhard Weßels. 2010. "Die deutsche Wahlforschung und die German Longitudinal Election Study (GLES)." In *Gesellschaftliche Entwicklungen im Spiegel der Empirischen Sozialforschung*, 141–172. Wiesbaden: VS Verlag für Sozialwissenschaften. doi:10.1007/978-3-531-92590-5_7.

Sundgren, Bo. 2011. "Communicating in Time and Space – How to Overcome Incompatible Frames of Reference of Producers and Users of Archival Data." In *EDDI 2011, 5–6 December 2011, Gothenburg, Sweden*. http://dx.doi.org/10.3886/DDIOtherTopics03, accessed 09292017.

Van den Eynden, Veerle, and Libby Bishop. 2014. "Sowing the Seed: Incentives and Motivations for Sharing Research Data, a Researchers' Perspective." http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf, accessed 09292017.

van Panhuis, Willem G, Proma Paul, Claudia Emerson, John Grefenstette, Richard Wilder, Abraham J Herbst, David Heymann, and Donald S Burke. 2014. "A Systematic Review of Barriers to Data Sharing in Public Health." *BMC Public Health* 14 (1): 1144. doi:10.1186/1471-2458-14-1144.

Weßels, Bernhard, Hans Rattinger, Sigrid Roßteutscher, and Rüdiger Schmitt-Beck. 2014. "Appendix: Study Description and Data Sources." In *Voters on the Move or on the Run*, edited by Bernhard Weßels, Hans Rattinger, Sigrid Roßteutscher, and Rüdiger Schmitt-Beck, 319–320. Oxford: Oxford University Press.

Whyte, Angus, and Andrew Wilson. 2010. "How to Appraise & Select Research Data for Curation. A Digital Curation Centre and Australian National Data Service 'Working Level' Guide." http://www.dcc.ac.uk/resources/how-guides/appraise-select-data, accessed 09292017.

Zenk-Möltgen, Wolfgang. 2006. "Dokumentation von Umfragedaten in Länder vergleichender Perspektive mithilfe des ZA Dataset Documentation Managers (DSDM)." *ZA-Information / Zentralarchiv Für Empirische Sozialforschung* (59): 159–170. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-198427, accessed 09292017.

Zenk-Möltgen, Wolfgang. 2013. "Metadata Management for Research Data with DBKfree." In *EDDI13 – 5th Annual European DDI User Conference. Réseau Quetelet – French Data Archives for Social Sciences, December 3–4, 2013*. http://www.eddi-conferences.eu/ocs/index.php/eddi/eddi13/paper/view/80, accessed 09292017.

Zenk-Möltgen, Wolfgang, and Evelyn Brislinger. 2009. "Original Language Documentation for the European Values Study." In *IASSIST/IFDO Conference, Tampere, Finland, 26–29 May, 2009*. http://www.fsd.uta.fi/iassist2009/presentations/D4_Brislinger.ppt, accessed 09292017.

Zenk-Möltgen, Wolfgang, and Norma Habbel. 2012. "Der GESIS Datenbestandskatalog und sein Metadatenschema. Version 1.8." 2012|01. Technical Reports. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-292372, accessed 09292017.

Zenk-Möltgen, Wolfgang, and Monika Linne. 2013. "Datorium – Ein neuer Service für Archivierung und Zugang zu sozialwissenschaftlichen Forschungsdaten." In *Beiträge des Workshops "Digitale Langzeitarchivierung" auf der Informatik 2013, Koblenz, 20.09.2013. Nestor Edition – Sonderheft 1*, 14–22. nestor-Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland. http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2014012419, accessed 09292017.

Zenk-Möltgen, Wolfgang, and Ekkehard Mochmann. 2000. "Der Continuity Guide der deutschen Wahlforschung und der ZA CodebookExplorer." In *50 Jahre Empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten*, edited by Markus Klein, Wolfgang Jagodzinski, Ekkehard Mochmann, and Dieter Ohr, 596–614. Heidelberg: Springer.