

gesis

Leibniz Institute
for the Social Sciences



Managing Social Science Metadata at the Data Archive

EDDI 2020 – Virtual Conference, 1-2 December 2020

Wolfgang Zenk-Möltgen

GESIS – Leibniz Institute for the Social Sciences, Germany

Outline

- Introduction
- Challenges for metadata applications
- Selection of standards
- Conversions to exchange metadata
- Example applications
- Current state of workflows
- Conclusions

Introduction



Research data
(6,592)

- GESIS Data Archive has predominately survey data: 6592 research data sets
- Mostly from Germany and other European Countries
- With typical topics from Sociology and Political Science

Mass political behaviour, attitudes/opinion (2,017)
Government, political systems and organisation (1,496)
Elections (929)
Economic systems and development (704)
International politics and organisation (695)
Employment (663)
Mass media (542)

GESIS-DBK (6,266)
GESIS-SowiDataNet | datorium (234)
GESIS-MISSY (92)

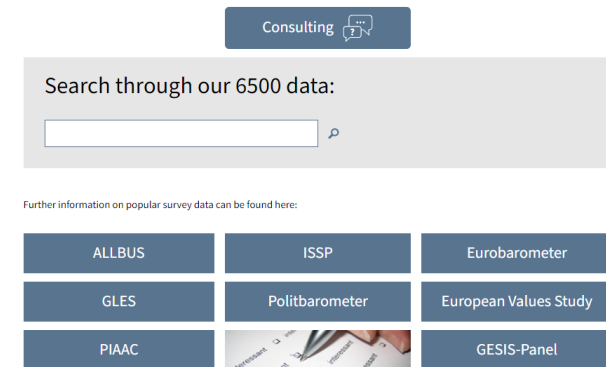
Germany (3,122)
German Democratic Republic (1949-1990) (950)
France (787)
Italy (745)
Netherlands (745)
Spain (700)
Ireland (695)
Belgium (691)
Denmark (682)
Portugal (662)
[more...](#)

Introduction

- Data are coming more and more from other sources than surveys
- In very diverse data formats
- Examples are
 - ▶ Social media (e.g. Twitter, Facebook)
 - ▶ Geographic or spatial data
 - ▶ Free-text (from open ended questions, but also from party programs or other texts)
 - ▶ Code for data preparation or analysis
 - ▶ Link collections

Introduction

- Tasks of the Data Archive include
 - ▶ Creating study descriptions
 - ▶ Offer data deposit and access
 - ▶ Help researchers to create documentation
 - ▶ Create complex documentation
 - ▶ Give advise for research data management
 - ▶ Long-term archiving
- User Services
 - ▶ Formerly done often by special purpose portals
 - ▶ Will be integrated as much as possible into general purpose website



Challenges for metadata applications

- Higher amounts of data
- Wider variety of data
- Context of data collection more often interdisciplinary
- More exchange of data and metadata
- Multi-linguality of metadata
- Expectations of fast provision and access
- Need to comply with various standards

Selection of standards

- Why? Selection builds a ground for development of applications
- Well defined interfaces are needed
- Until now: DDI-Lifecycle as the basis
- Extended and combined with additional standards
- Technically managed with the DDIFlatDB approach¹



¹ Hopt, Oliver, Claus-Peter Klas, and Alexander Mühlbauer. "Flexible DDI storage." *IASSIST Quarterly* 42.2 (2018): 1-12.

Selection of standards

Additional standards used

- ISO standards for date, time, and country
- CESSDA CMM metadata (DDI compliant)
- da|ra and DataCite metadata schema
- Controlled Vocabularies from all these

Need to do version management (metadata, CVs)

- E.g. ISO Countries do change over time
- Also metadata schemata do change, especially versions of CVs may differ

Selection of standards

Additional relevant standards

- Semantic web, RDF, SKOS
- Schema.org
- OpenAire, DCAT
- Sometimes applied several times through chained workflows

Topic Classifications

- ELSST, TheSoz, ZA Categories, CESSDA Topic Classification, Continuity Guides, etc ...

Conversions to exchange metadata

Mappings or Cross-walks between different standards

- Efficiency
 - ▶ document once, export to many portals
 - ▶ Automatic procedure
- Problems
 - ▶ semantic meanings do not always support this
 - ▶ Higher or lower granularity may be needed
 - ▶ Context might change



Example applications

- Study Description
 - ▶ Created in DBKEdit, saved in MySQL
 - ▶ Exported to DDI-C XML for CESSDA CDC
 - ▶ Exported to another DDI-C for ZACAT Nesstar
 - ▶ Exported to DDI-L for DDIFlatDB
 - Used e.g. by ExploreData and CESSDA EQB
 - ▶ Exported to da|ra for DOI metadata
 - ▶ Indexed by Solr Search for GESIS Search

Studie	Studie	Entwurf	Entwurf	Vorgang	Vorgang	vorgige	aktuelle	nächste	Logdatei	XML	Vorschau in	Vorschau	Druckansicht
titel	titel	titel	titel	titel	titel	(2012)	(2012)	(2012)			DDI-Search	(print)	
2002: International Social Survey Programme: Citizenship - ISSP 2004													
ZA-Studiennummer		3950											
Titel		International Social Survey Programme: Citizenship - ISSP 2004											Bearbeiten
Weitere Titel													Liste
Produktionsbedingungen													Liste
Gruppenzuordnungen													Liste
FDZ-Zuordnungen													Liste
Wichtige Version													Liste
Erhebungszeitraum standardisiert													Liste

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>

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 prejudiced against any racial or ethnic group; social and political participation; memberships; importance of different people's rights in democracy (scale); estimation of political influence possibilities (political efficacy); likelihood of counter-action against an unjust law and expected chance of serious attention to people's demand; interest in politics; personal trust in politicians and people..."

Principal Investigator/ Authoring Entity, Institution: Wilson, Shaun - Macquarie University, Sydney, Australia | Gibson, Rachel - ACPSPRI Centre for Social Research (ACSR) Research School of Social Sciences, The Australian National University, Canberra, Australia | Hadler, Markus - Institute of Sociology, University of Graz, Austria | Hoellinger, Franz - Institute of Sociology, University of Graz, Austria | ...

Publisher: GESIS Data Archive

Study number: ZA3950

Current Version: 1.3.0, 2012-06-26, <https://doi.org/10.4232/1.11372>

DOI: 10.4232/1.11372 | 10.4232/1.10078

Publication year: 2012

Date of Collection: 03.2003 - 02.2006

Geographic coverage: Australia | Brazil | Chile | Bulgaria | Denmark | Germany | Finland | Vlaams Gewest | France | Great Britain | Ireland | Israel | Japan | Canada | Latvia | Mexico | Netherlands | New Zealand | Norway | Austria | Philippines | Poland | Portugal | Russian Federation | Sweden | Switzerland | Slovakia | Slovenia | Spain | South Africa | Korea, Republic of | Taiwan | Czech Republic | H... | ...

Universe: Age: 18 years old and older, exclusion: Finland (15), Japan (16), South Africa (16) and the Netherlands (16)

Analysis Unit (3): Individual

Sampling Procedure (3): Probability Sampling procedures differ for the individual countries: Partly simple, partly multi-stage stratified

Materials

[Datasets](#)

[Questionnaires](#)

[Codebook](#)

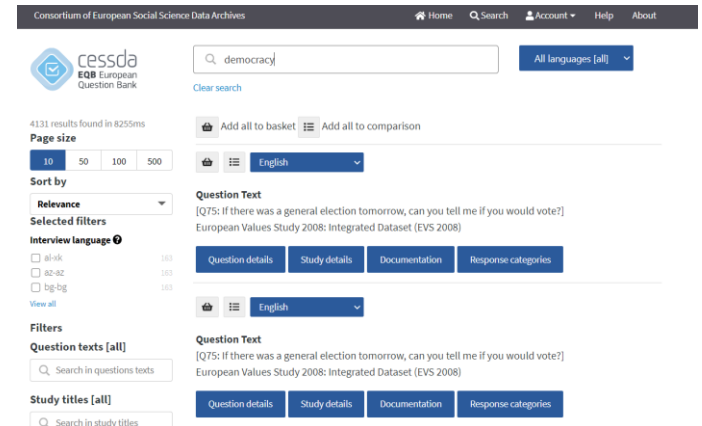
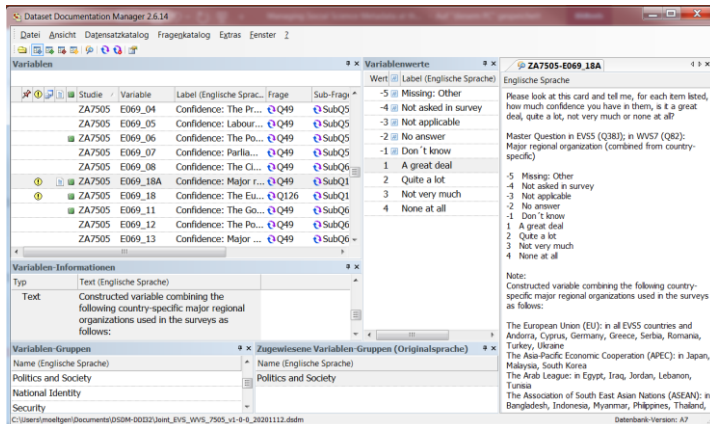
[Other documents](#)

Actions

[Cite](#)

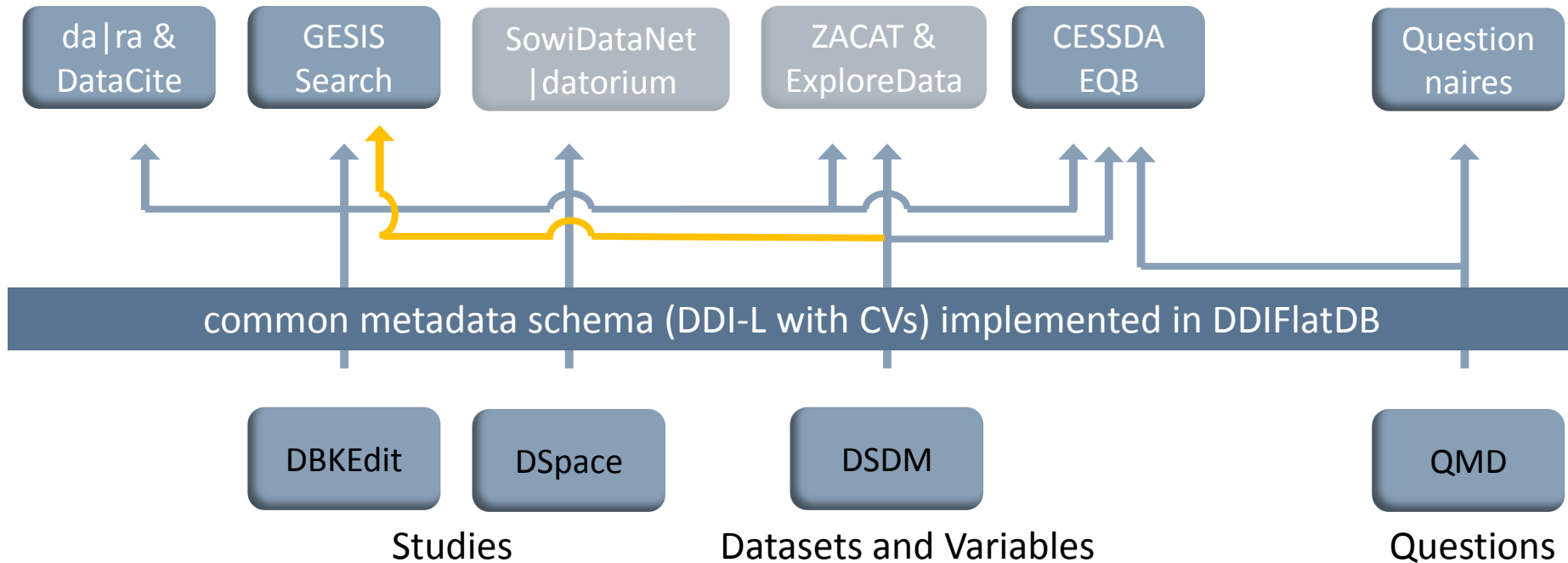
Example applications

- Variable and Question Documentation
 - ▶ Created in DSDM, saved in MS Access
 - ▶ Exported to CodebookExplorer for VariableReports
 - ▶ Exported to DDI-C for ZACAT Nesstar
 - ▶ Exported to DDI-L for DDIFlatDB
 - Used e.g. by ExploreData and CESSDA EQB
 - Indexed by ElasticSearch for GESIS Search



Current state of workflows

Portals for end-users



Metadata production tools

Current state of workflows

- Using a common metadata store to feed many applications
- Using legacy editor tools to fill this store, but enabling a stepwise migration of these
- Integrated search and access for end-users
- Special functions may still require extra portals
- Support for partner portals where necessary
- Limited support for different metadata versions

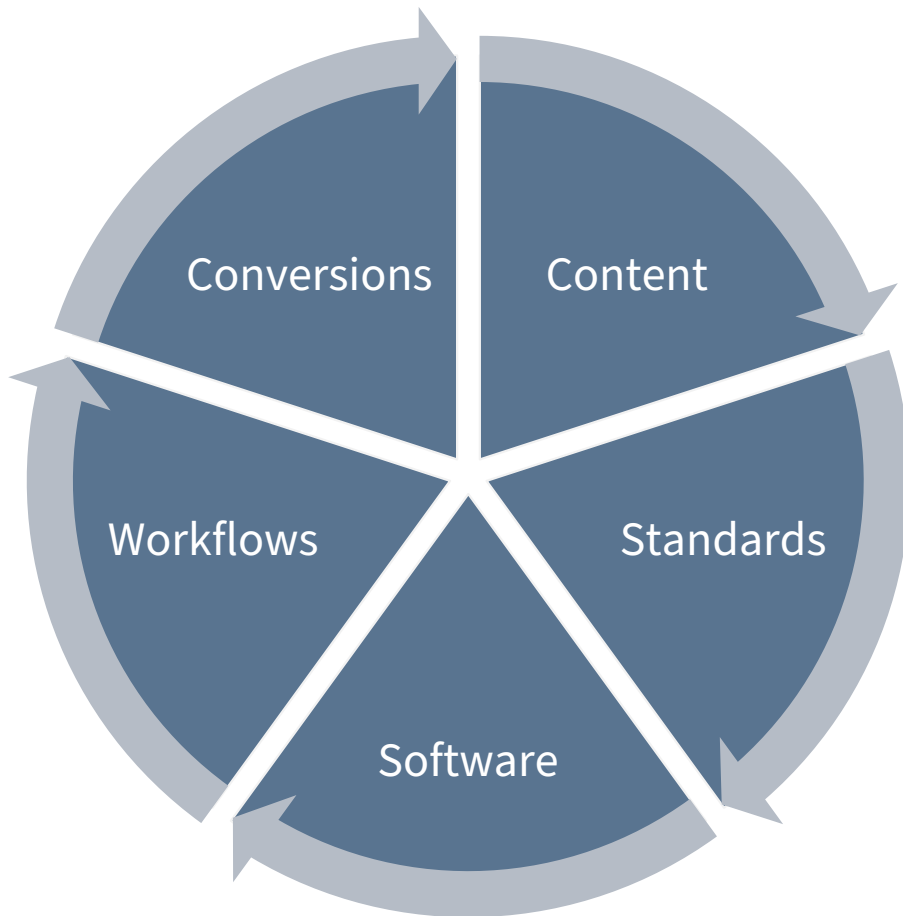
Conclusions

We need to do

- Use the common metadata store to feed all the remaining applications
- Evaluate state-of-the-art editor tools to manage the metadata workflow
- Do even more integration for end-users
- Complete the concept for managing different versions of CVs
- Include the conversion functions into the concept, assuming that there will be ongoing change in the metadata schemata and CVs

Conclusions

You need to manage



... to meet the challenges
of more and
heterogeneous data types

- Extended content
- Additional standards
- More exchange
- More languages
- Fast access

Thank you! Questions?



Contact: Wolfgang.Zenk-Moeltgen@gesis.org

This presentation is offered under the CC-BY 4.0 license. This license does not apply to the following copyrighted material used in this presentation: the logo of GESIS, the logo of the Leibniz Association.