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

- GESIS Data Archive has predominatly 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) Research data (6,592)

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

GESIS-Panel

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







PIAAC

Challenges for metadata applications

- Higher amounts of data
- Wider variety of data

Leibniz Institute for the Social Science:

- 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

Leibniz Institute for the Social Science:

- 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

		1050											
populerii	rescritere	Sana and	Free and	14					1 contains	-	Manual and State	Managhar	Developmental
dt.	en.	dt.	en.	dt.	en.	(3949)	(3950)	(3951)	Logane	ANL	DBKSearch	(intern)	Crockansien
3950. Citize	Internati nship - It	onal Social SSP 2004	i Survey Pr	ogramme:									
ZA-Stu:	Sennumm	er		3950									
Titel				Internation	al Social Su - ISSP 200	ivey Progra 4	imme:						Bearbeiten
Weitere	Titel												Liste
Produkt	zuordnun	gen											Liste
Gruppe	nzuordnur	ngen		GN0001GN0136	International ISSP - Mod	II Social Su ule Topic: (rvey Programm Ditizenship	e (ISSP)					Liste
FDZ-Zu	ordnunge	n		 FD2000 	3: Internation	ale Umfrag	jeprogramme (F	DZ Int. Umfrag	peprogramm	bei GE	SIS)		Liste
letzte V	ersion			 1.3.0, 26 Update: brien 10.4232 (Version (Version) 	1.05.2012, Ug correction of 1.11372 wurde bereit ten zu dieser	state: com errors (see s registrier Version si	ection of errors (e changes in this t) nd aktuell.)	see changes in a version)	n this version	0			Liste
Intebrugschraum standardiserffer								Liste					



nternational Social Survey Programme: Citizenship - ISSP 2004

Wilson, Shaun; Gibson, Rachel: Hadler, Markus 📖

GESS Data Archive, Colegne. 240980 Onta file Version 1.3.0, https://doi.org/10.4320/1.1322 Abstract: "Critter and state. Topics: Qualities of a good crittere: artitude toward the right of public meetings for religous sertemists, people who want to overthrow the government by force, or people prejudiced against any rical or of enhic group; social and publical participation; membership; importance of different people 's rightma' memory (social participation; membership; prejudiced against and ready in the memory (social participation; membership; possibilities political efficacy); Balenses of counter-action against an unjust law and expected chance: o forsion attention to people 's demand; interest in policity; prosond trust in policitican and people."

Principal Investigator' Authoring Entry, Institution: Wilson, Shaon - Macquarie University, Sydney, Australia J Gibson, Rachel - ACSPRI Centre for Social Research (ACSR) Research School of Social Sciences, The Australian Inational University, Carberra, Australia | Haller, Markas - Institute of Sociology, University of Graz, Austria | Homger, Franz- Nature of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Testute of Sociology, University of Graz, Austria | Haller, Franz- Franz- Franz- Franz-

Materials Datasets

Questionnaires

Other documents

↓ Actions

Codebook

Publisher: GESIS Data Archive Study number: 7A3950

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

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

Analysis Unit (ii): Individual

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



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

🧐 Dataset D	ocumentation	Manager 2.6.14											
Datei Ans	icht Datensi	atzkatalog Fra	igenkatalog Ext	tras <u>F</u> er	ister <u>2</u>								
-	\$ 🗟 🗞 😨	6 ar											
Variablen						á X	Variabl	enwerte		φ×	🗭 ZA7505-	E069_18A	4.1
							Wert 8	Label (E	nglische Spra	ache)	Englische Sprach	ie .	
📌 🛈 🎜 🛙) 🖬 Studie 🦯	Variable	Label (Englisch	ne Sprac.	. Frage	Sub-Frage *	-5 🛛	Missing	: Other		Please look at t	his card and tell r	me, for each item listed
	ZA7505	E069_04	Confidence: 1	The Pr	Q49	SubQ5	-4 8	Not ask	ed in surve	y	how much conf	dence you have	in them, is it a great
	ZA7505	E069_05	Confidence: L	abour	. 🔁 Q49	SubQ5	-3 🛛	Not app	olicable		deay quice a loc	not very moon	or none of on
	ZA7505	E069_06	Confidence: 1	The Po	. 🔁 Q49	SubQ5	-2 8	No ansi	wer		Master Questio	1 in EVS5 (Q38J)	r; in WVS7 (Q82):
	ZA7505	E069_07	Parlia	Q49	SubQ5	-1 🗷 Don't know				specific)			
	ZA7505	E069_08	Confidence: 1	The Ci	Q49	€ SubQ6	1	A great	deal				
•	B ZA7505	E069_18A	Confidence: N	Major r	. Q49	O SubQ1	2	Quite a	lot		-5 Missing: Ot -4 Not asked		
۲	ZA7505	E069_18	Confidence: 1	The Eu	. QQ126	O SubQ1	3	Not ver	y much		-3 Not applical	ale	
	ZA7505	E069_11	Confidence: 1	The Go	• Q49	SubQ6	- 4	None at	t all		-2 No answer		
	ZA7505	E069_12	Confidence: 1	The Po	• Q49	SubQ6					1 A great dea		
	ZA7505	E069_13	Confidence: N	Major	0Q49	€ SubQ6 -					2 Quite a lot		
4						•					3 Not very mi 4 None at all	JON	
Variablen-I	nformationer	1				\$ ×							
Typ Text (Englische Sprache)					*	*				Note: Constructed variable combining the following country:			
Text Constructed variable combining the following country-specific major regiona organizations used in the surveys as follows:				onal			٠	881		,	specific major mas follows: The European I Andorra, Cypru	gional organizati Jnion (EU): in all s, Germany, Gre	EVSS countries and rece, Serbia, Romania,
Variablen-G	ruppen			9 × Z	ugewiesen	e Variablen-G	iruppen	(Original	sprache)	₹×	Turkey, Ukraine		(1050)
Name (Englische Sprache) Amer (Englische Sprache					sche Sprache)					The Asa-Pachc Economic Cooperation (APEC): in Japan Malaysia, South Korea The Arab League: in Egypt, Iraq, Jordan, Lebanon,			
Politics and Society Politics and Society					Society								
National Ide	intity			10							Tunsia The Association	of South East A	isian Nations (ASEAN):
Security *				*							Bangladesh, Inc	Jonesia, Myanmi	ar, Philippines, Thailand
-liteorel-moalt	and Barry marked	COM DOTING	DUC MAIL TRAK	1.0.0 30	201112 deden							0.	stanhank-Marginn: 47

Consortium of European Social Science Data Archives 😽 Home 🔍 Search 🔮 Account 👻 Help About											
CESSDA EQB European Question Bank	Q, democracy All languages [all] ~ Clear search										
4131 results found in 8255ms Page size	Add all to basket I≣ Add all to comparison										
10 50 100 500	🖀 📰 English 🗸										
Sort by											
Relevance 👻	Question Text [075: If there was a general election tomorrow, can you tell me if you would yote?]										
Selected filters	[Q75: If there was a general election tomorrow, can you tell the if you would vote?] European Values Study 2008: Integrated Dataset (EVS 2008)										
Interview language 🚱	En openn vunes stady 2000 mitegraten sutaset (Evo 2000)										
al-xk 163	Question details Study details Documentation Response categories										
az-az 163											
View all	台 📰 English 🗸										
Filters											
Question texts [all]	Question Text										
furner in the fault	[Q75: If there was a general election tomorrow, can you tell me if you would vote?]										
C Search in questions texts	European Values Study 2008: Integrated Dataset (EVS 2008)										
Study titles [all]	Question details Study details Documentation Response categories										
Q. Search in study titles											



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



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

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

Thank you! Questions?

Gesis Leibniz Institute for the Social Sciences



Contact: <u>Wolfgang.Zenk-Moeltgen@gesis.org</u>

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.