

MISSY 2

[Microdata Information System – part 2]

Implementing DDI 3: The German Microcensus Case Study

andias.wira-alam@gesis.org

oliver.hopt@gesis.org

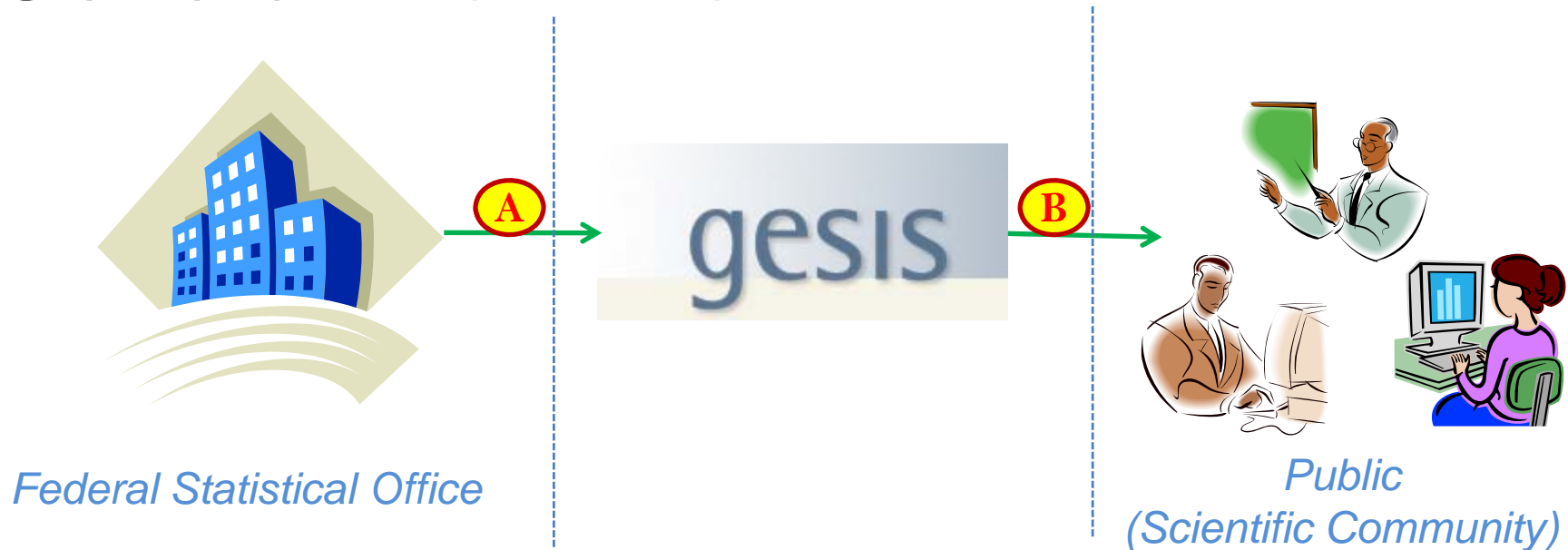
Information Processes in the Social Sciences



Issues:

1. Documenting the Metadata at Variable Level.
2. DDI 3 Editor?
3. Metadata Presentation?

General Work Flow



A

The Federal Statistical Office (Destatis)...

- provides access to microdata for scientific purposes (scientific use files).
- provides survey documents for MISSY.
- in collaboration with, produces scientific use files of the Microcensus (e.g. anonymisation, construction of variable labels).

B

GESIS provides...

- provides metadata for the Microcensus scientific use files on the Web (MISSY): codebook and detailed variable information in **DDI 3 format**.
- study description and study related material (e.g. questionnaires, definitions, interviewer guides).
- examples and recommendations on analysing the metadata.

Available Materials

(Example: Code Book and Questionnaire)

Code Book



Mikrozensus Scientific Use File 2005
Häufigkeitsauszählung

Codebuch und Randverteilungen (ungewichtete Fallzahlen)

EF1 - Bundesland

Value	Value Label	Frequency	Total %	Valid %
1	Baden-Württemberg	63131	13.2%	13.2%
2	Bayern	78087	16.4%	16.4%
3	Berlin	19461	4.1%	4.1%
4	Brandenburg	15465	3.2%	3.2%
5	Bremen	3960	0.8%	0.8%
6	Hamburg	10155	2.1%	2.1%
7	Hessen	36750	7.7%	7.7%
8	Mecklenburg-Vorpommern	9535	2%	2%
9	Niedersachsen	38790	8.1%	8.1%
10	Nordrhein-Westfalen	98402	20.6%	20.6%
11	Rheinland-Pfalz	22965	4.8%	4.8%
12	Saarland	6121	1.3%	1.3%
13	Sachsen	26481	5.5%	5.5%
14	Sachsen-Anhalt	15510	3.2%	3.2%
15	Schleswig-Holstein	18032	3.8%	3.8%
16	Thüringen	14394	3%	3%
	Valid Total	477239	100%	100%
	Total	477239	100%	

EF3 - Nr. des Auswahlbezirks (laufend)

Value	Value Label	Frequency	Total %	Valid %
Kategorien werden auf Grund der großen Zahl nicht ausgezählt.				
	Valid Total		100%	100%
	Total	NaN	100%	

Available Materials

(Example: Code Book and Questionnaire)

Questionnaire

Bitte den Erhebungsbogen für alle Haushaltsmitglieder ausfüllen!

	1. Person	2. Person	3. Person	4. Person	5. Person
4a	<p><i>Bitte behalten Sie die Reihenfolge der Personen aus der Namens-Lasche immer bei. Aus Erfahrung bietet es sich bei Mehrpersonen-Haushalten an, erst alle Fragen für die erste Person, dann alle für die zweite Person usw. zu beantworten.</i></p>				
Fragen zu den Personen im Haushalt					
5	<p>Sind Sie in den letzten 12 Monaten hier eingezogen?</p> <p>Ja <input type="checkbox"/> 1 <input type="checkbox"/> 1 <input type="checkbox"/> 1 <input type="checkbox"/> 1 <input type="checkbox"/> 1</p> <p>Nein..... <input type="checkbox"/> 8 <input type="checkbox"/> 8 <input type="checkbox"/> 8 <input type="checkbox"/> 8 <input type="checkbox"/> 8</p>				
6	<p>Bitte geben Sie Ihr Geschlecht an!</p> <p>Männlich..... <input type="checkbox"/> 1 <input type="checkbox"/> 1 <input type="checkbox"/> 1 <input type="checkbox"/> 1 <input type="checkbox"/> 1</p> <p>Weiblich..... <input type="checkbox"/> 2 <input type="checkbox"/> 2 <input type="checkbox"/> 2 <input type="checkbox"/> 2 <input type="checkbox"/> 2</p>				
7	<p>In welchem Jahr und Monat sind Sie geboren?</p> <p>Bitte geben Sie an: Jahr</p> <p style="padding-left: 100px;">Monat</p>				
8	<p>Welchen Familienstand haben Sie?</p> <p>Ledig <input type="checkbox"/> 1 <input type="checkbox"/> 1 <input type="checkbox"/> 1 <input type="checkbox"/> 1 <input type="checkbox"/> 1</p> <p>Verheiratet..... <input type="checkbox"/> 2 <input type="checkbox"/> 2 <input type="checkbox"/> 2 <input type="checkbox"/> 2 <input type="checkbox"/> 2</p> <p>Verwitwet..... <input type="checkbox"/> 3 <input type="checkbox"/> 3 <input type="checkbox"/> 3 <input type="checkbox"/> 3 <input type="checkbox"/> 3</p>				

What goes to where?

(DDI 3 structure)

Code Book & Questionnaire

missy Mikrozensus Scientific Use File 2005 Häufigkeitsauszählung Seite 4 von 298

Variablenverzeichnis

EF-Nummer	Variablenlabel
EF1	Bundesland
EF3	Nr. des Auswahlbezirks (laufend)
EF4	Nr. des Haushalts im Auswahlbezirk (laufend)
EF5	Nr. der Person im Haushalt (laufend)
EF6	Fragenprogramm
EF12	Bereichsquartal
EF17	F149a Erhebung: Art der Beteiligung
EF25	Nr. der Familie im Haushalt (systemfrei)
EF27	Nr. der Lebensform im Haushalt (systemfrei)
EF29	Erwerbstyp
EF30	Bevölkerung: Haupt- oder Nebenwohnsitz
EF31	Bevölkerung: Privathaushalte
EF32	Bevölkerung: Eurostat-Definition
EF33	Ehe-/Lebenspartner im Haushalt
EF34	Stellung innerh. der Familie (tradit. Familienkonzept) 1)
EF35	Stellung innerh. der Lebensform (Konzept der Lebensformen) 1)
EF36	Stellung der Mütter/des Vaters im Haushalt
EF37	Stellung der Person im Haushalt
EF38	Erwerbstätigkeit: Subgruppen von EF29
EF44	Alter
EF46	F6 Geschlecht
EF47	F7 Geburtsjahr

missy Mikrozensus Scientific Use File 2005 Häufigkeitsauszählung Seite 13 von 298

Codebuch und Randverteilungen (ungewichtete Fallzahlen)

EF1 - Bundesland	Value	Value Label	Frequency	Total %	Valid %
1	Baden-Württemberg	63131	13.29%	13.29%	
2	Bayern	78087	16.48%	16.48%	
3	Berlin	19461	4.18%	4.18%	
4	Brandenburg	15465	3.29%	3.29%	
5	Bremen	3960	0.89%	0.89%	
6	Hamburg	10155	2.18%	2.18%	
7	Hessen	36750	7.79%	7.79%	
8	Mecklenburg-Vorpommern	9535	2%	2%	
9	Niedersachsen	38790	8.19%	8.19%	
10	Nordrhein-Westfalen	98402	20.65%	20.65%	
11	Rheinland-Pfalz	22965	4.88%	4.88%	
12	Saarland	6121	1.3%	1.3%	
13	Sachsen	26481	5.59%	5.59%	
14	Sachsen-Anhalt	19510	4.19%	4.19%	
15	Schleswig-Holstein	18032	3.89%	3.89%	
16	Thüringen	14394	3%	3%	
	Valid Total	477239	100%	100%	
	Total	477239	100%	100%	

EF3 - Nr. des Auswahlbezirks (laufend)	Value	Value Label	Frequency	Total %	Valid %
	Valid Total			100%	100%
	Total		NaN	100%	

Questionnaire

Bitte den Erhebungsbogen für alle Haushaltsmitglieder ausfüllen!

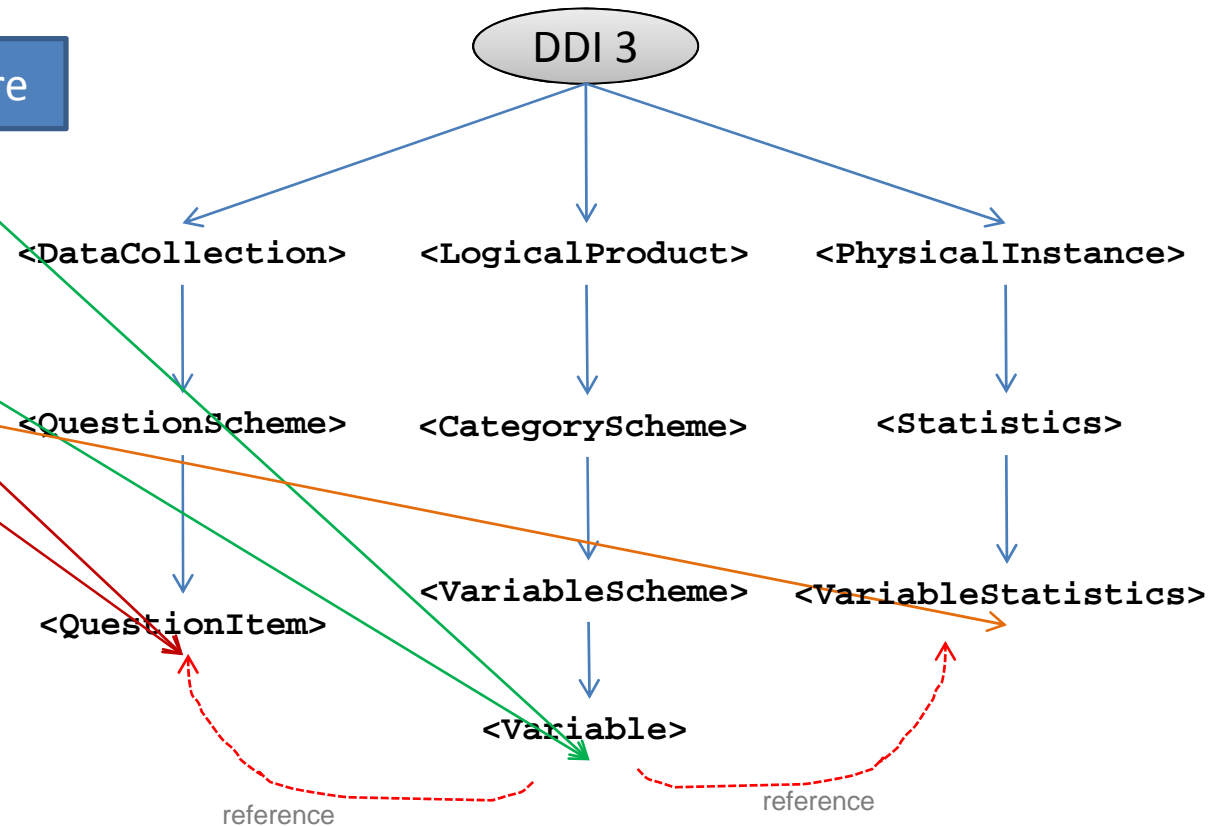
	1. Person	2. Person	3. Person	4. Person	5. Person
4a Bitte behalten Sie die Reihenfolge der Personen aus der Namens-Lasche immer bei. Aus Erfahrung bietet es sich bei Mehrpersonenhaushalten an, erst alle Fragen für die erste Person, dann alle für die zweite Person usw. zu beantworten.					
Fragen zu den Personen im Haushalt					
5 Sind Sie in den letzten 12 Monaten hier eingezogen? Ja <input type="checkbox"/> j1 <input type="checkbox"/> j2 <input type="checkbox"/> j3 <input type="checkbox"/> j4 <input type="checkbox"/> j5 Nein <input type="checkbox"/> n1 <input type="checkbox"/> n2 <input type="checkbox"/> n3 <input type="checkbox"/> n4 <input type="checkbox"/> n5					
6 Bitte geben Sie Ihr Geschlecht an! Männlich <input type="checkbox"/> m1 <input type="checkbox"/> m2 <input type="checkbox"/> m3 <input type="checkbox"/> m4 <input type="checkbox"/> m5 Weiblich <input type="checkbox"/> w1 <input type="checkbox"/> w2 <input type="checkbox"/> w3 <input type="checkbox"/> w4 <input type="checkbox"/> w5					
7 In welchem Jahr und Monat sind Sie geboren? Bitte geben Sie an: Jahr Monat					
8 Welchen Familienstand haben Sie? Ledig <input type="checkbox"/> l1 <input type="checkbox"/> l2 <input type="checkbox"/> l3 <input type="checkbox"/> l4 <input type="checkbox"/> l5 Verheiratet <input type="checkbox"/> v1 <input type="checkbox"/> v2 <input type="checkbox"/> v3 <input type="checkbox"/> v4 <input type="checkbox"/> v5 Verwitwet <input type="checkbox"/> w1 <input type="checkbox"/> w2 <input type="checkbox"/> w3 <input type="checkbox"/> w4 <input type="checkbox"/> w5					

What goes to where?

(DDI 3 structure)

Code Book & Questionnaire

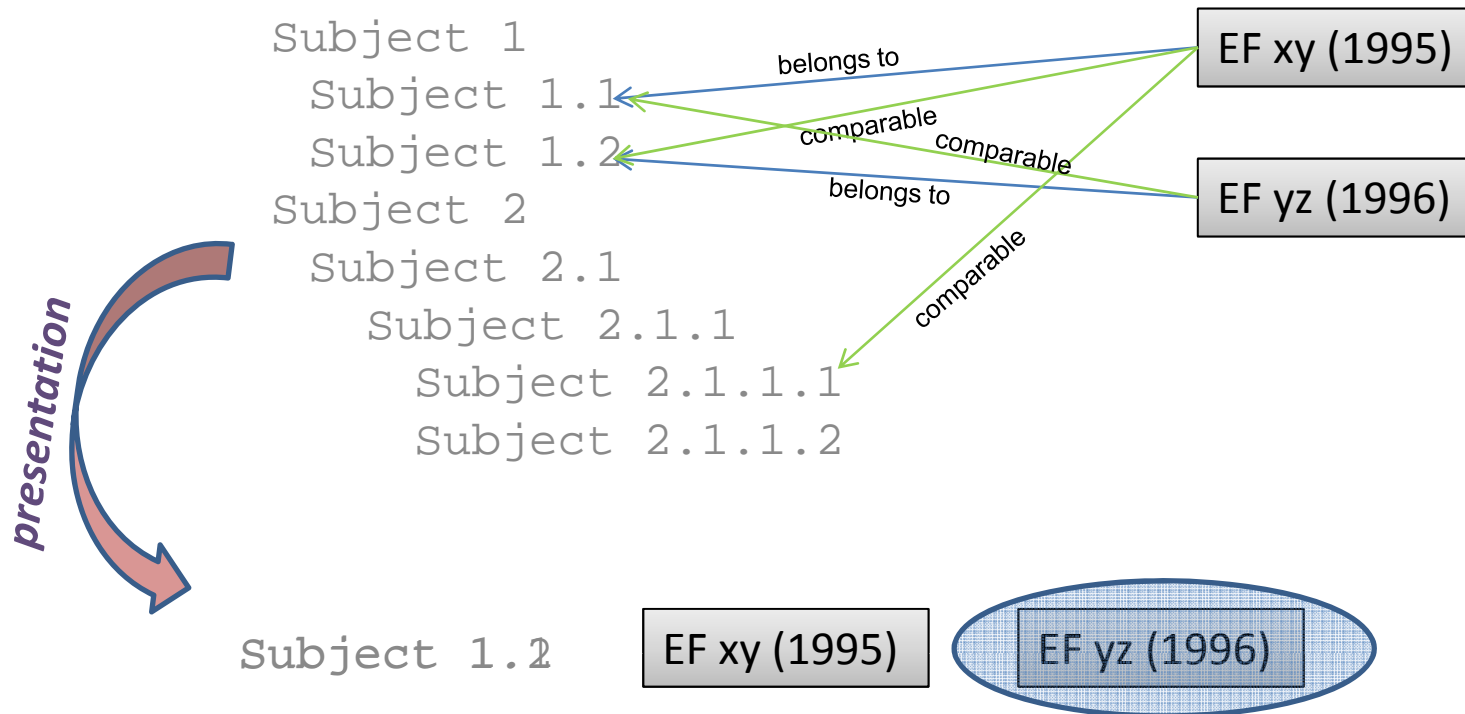
- 1. Variable Name (EF Number)
- 2. Variable Label
- 3. Question Number
- 4. Question Text
- 5. Frequencies



Comparable Variables

(DDI 3 structure ++)

Subjects (Categorization)



missy Editor - Mikrozensus - 1973 - 0

Studie | Variable |

Pool +

Zugang Passwort
ho *****

Anmelden

Survey/Panel +

Mikrozensus

Erhebungszeitraum + d
1973

Version + d i
0

Hilfe Speichern

jLabel1 jT
jLabel2 jT
jLabel3 jT
jLabel4 jT

JTree
cc
sp
fo

Nr. des Haushalts im Auswahlbezirk (systemfrei) aktualisieren

Variablenname EF4

Variablenlabel Nr. des Haushalts im Auswahlbezirk (systemfrei)

Amtliche Klassifikation -

Substichprobe -

Auswahlsatz 1%

Erhebungsbogen Erhebungsbogen 1+E

Anmerkung

Thematische Gliederung * Nr. des Haushalts im Auswahlbezirk
bearbeiten

Vergleichbarkeit (them. Gliederung)
bearbeiten

generierte Variable

missy Editor - Mikrozensus - 1989 - 0

Studie Variable |

Mikrozensus:1989-0

Name	Frage	USP	FB
O003			
EF1			
EF5			
EF6			
EF7			
EF8			
EF15			
EF23			
EF24			
EF25			
EF26			
EF27			
EF28			
EF29			
EF30			
EF31			
EF32			
EF33			
EF34			
EF35	F7		
EF37	F8		
EF38	F10		
EF39	FA3	x	
EF40	F11		
EF41	F13		
EF47	F14		
EF48	F15		
EF49	F16		
EF50	F17		
EF51	F18		
crse	F19		
EF53	F20		
EF56	F23		
EF61	F29		
EF62	F30		
EF63	F31		
EF64	F32		
EF65	F33		
EF66	F34-42ba		

Sortierung
Name <

Speichern

Variableninformationen | Values/Häufigkeiten

FA3 Eheschließungsjahr (FB) aktualisieren

Variablenname EF39

Variablenlabel Eheschließungsjahr

Amtliche Klassifikation -

Substichprobe -

Auswahlsatz 1%

Erhebungsbogen Erhebungsbogen 1

Anmerkung

Thematische Gliederung * Eheschließungsjahr
bearbeiten

Vergleichbarkeit (them. Gliederung)
bearbeiten

generierte Variable

Frage Nummer FA3

Frage Text _Eheschließungsjahr_ der jetzigen bzw. letzten Ehe

Erläuterungen zur Frage im Anhang

Filteranweisung für Verheiratete, Verwitwete, Geschiedene

Filterangabe (wörtlich)

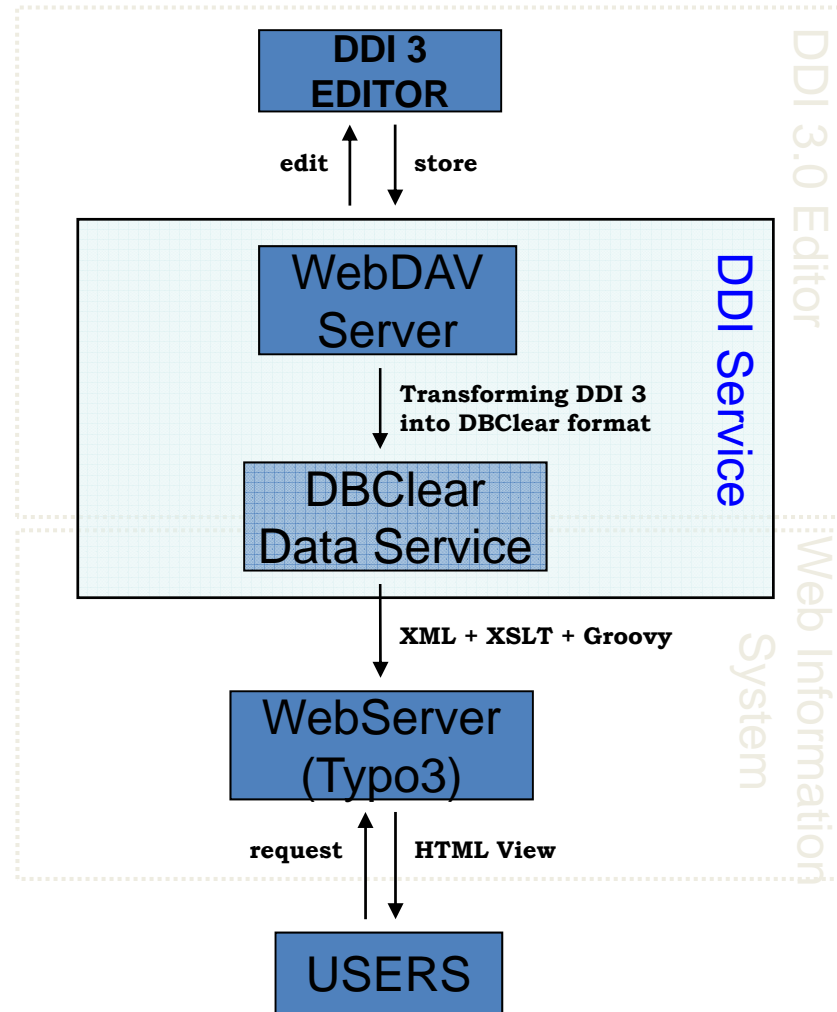
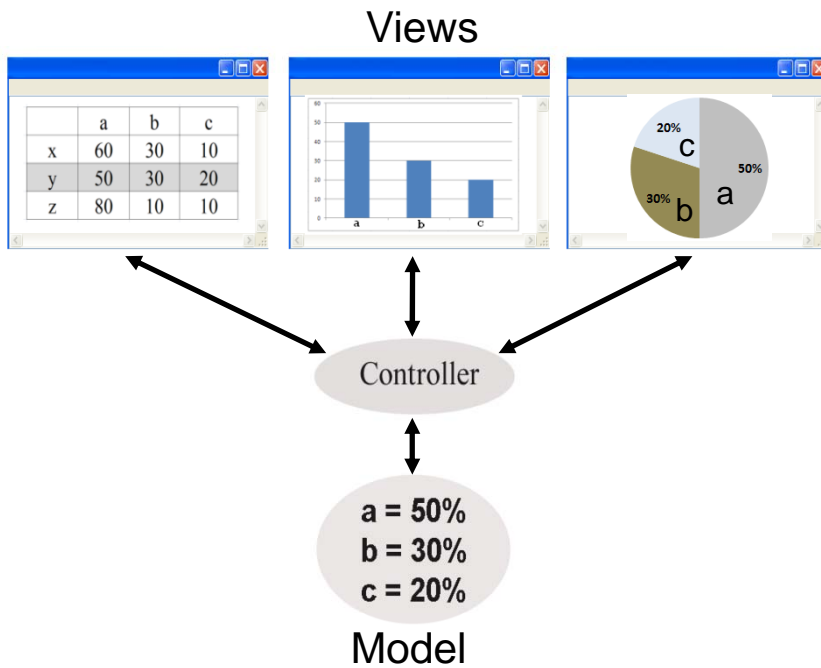
Filterangabe (formal) IF EF38=2 OR EF38=3 OR EF38=4.

Auskunftspflicht nein

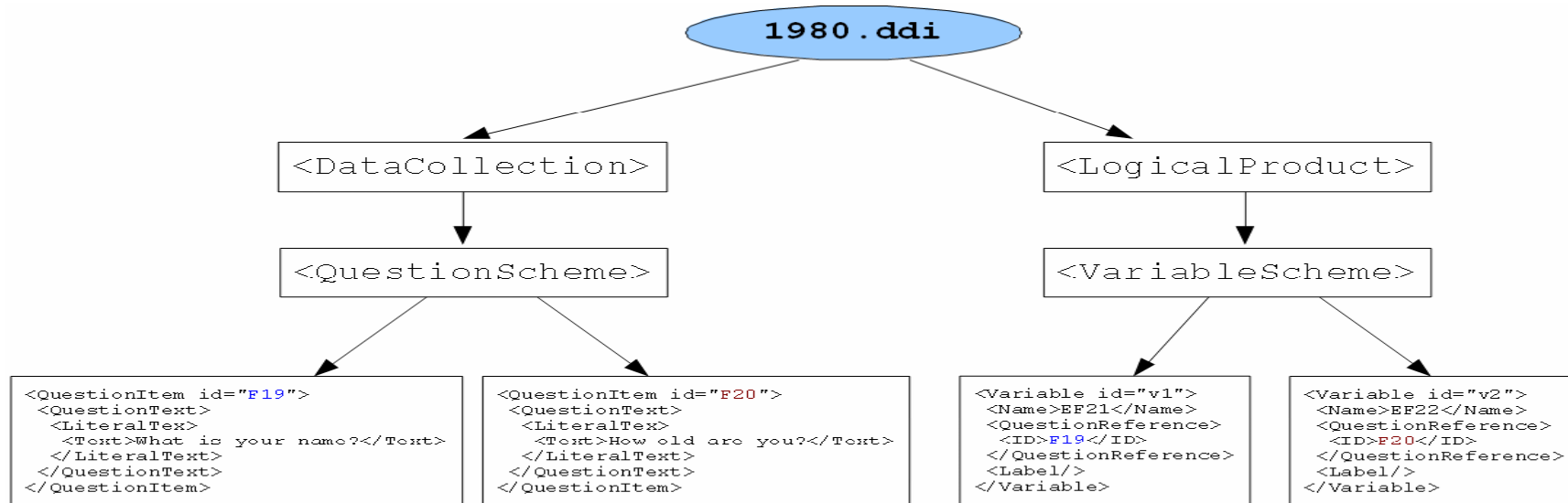
Nr. Interviewer-Handbuch 5/31, 32

Kommentar

Software Architecture



DDI to Metadatabase



<i>variable</i>	<i>census_year</i>	<i>question_number</i>	<i>question_text</i>
EF21	1980	F19	What is your name?
EF22	1980	F20	How old are you?
...

Database Indexing

Lucene Index

Web Information System in MISSY:

(3 detail views of the metadata - dynamic pages)

1. List of Variables (Variablenliste).
2. Subject Outline (Thematische Gliederung).
3. Timeline Matrix of Variables (Variablen-Zeitpunkte-Matrix).

Sie sind hier: [Variablen](#)

Variablen

- Thematische Gliederung
- Variablenliste
- Liste der generierten Variablen
- Variablen-Zeitpunkte-Matrix

© GESIS

Variableninformationen

Zu den Informationen auf Variablenebene erhalten Sie Zugang über

- eine [thematische Gliederung](#), die die Variablen nach Themengebieten sortiert
- eine [Variablenliste](#), die alle Variablen eines Erhebungsjahres auflistet
- eine [Matrix](#), die einen Überblick aller Variablen und Erhebungsjahre enthält

Impressum

21

Gefördert durch:



Sie sind hier: [Variablen](#)

Variablen

- Thematische Gliederung
- Variablenliste
- Liste der generierten Variablen
- Variablen-Zeitpunkte-Matrix

© GESIS

Variableninformationen

Zu den Informationen auf Variablenebene erhalten Sie Zugang über

- eine **thematische Gliederung**, die die Variablen nach Themengebieten sortiert
- eine **Variablenliste**, die alle Variablen eines Erhebungsjahres auflistet
- eine **Matrix**, die einen Überblick aller Variablen und Erhebungsjahre enthält

Impressum

Gefördert durch:



Bundesministerium
für Bildung
und Forschung

Sie sind hier: [Variablen](#)

Variablen

- Thematische Gliederung
- Variablenliste
- Liste der generierten Variablen
- Variablen-Zeitpunkte-Matrix

© GESIS

Variableninformationen

Zu den Informationen auf Variablenebene erhalten Sie Zugang über

- eine thematische Gliederung, die die Variablen nach Themengebieten sortiert
- eine **Variablenliste**, die alle Variablen eines Erhebungsjahres auflistet
- eine **Matrix**, die einen Überblick aller Variablen und Erhebungsjahre enthält

[Impressum](#)

Gefördert durch:



Bundesministerium
für Bildung
und Forschung

Acknowledgement

This project is funded by the BMBF Germany (01UW0707 - "Servicezentrum für Mikrodaten der GESIS / MISSY II").

Future Work

1. Integrating with regional microcensus.
2. Improving our application (e.g. performance, search capabilities, object oriented database, etc) so it can be used also by other studies.
3. Open Research Data.