

# gesis

Leibniz Institute  
for the Social Sciences



## Lessons Learned with Additional Mappings into DDI- FlatDB

*Claus-Peter Klas, Oliver Hopt, Alexander Mühlbauer*

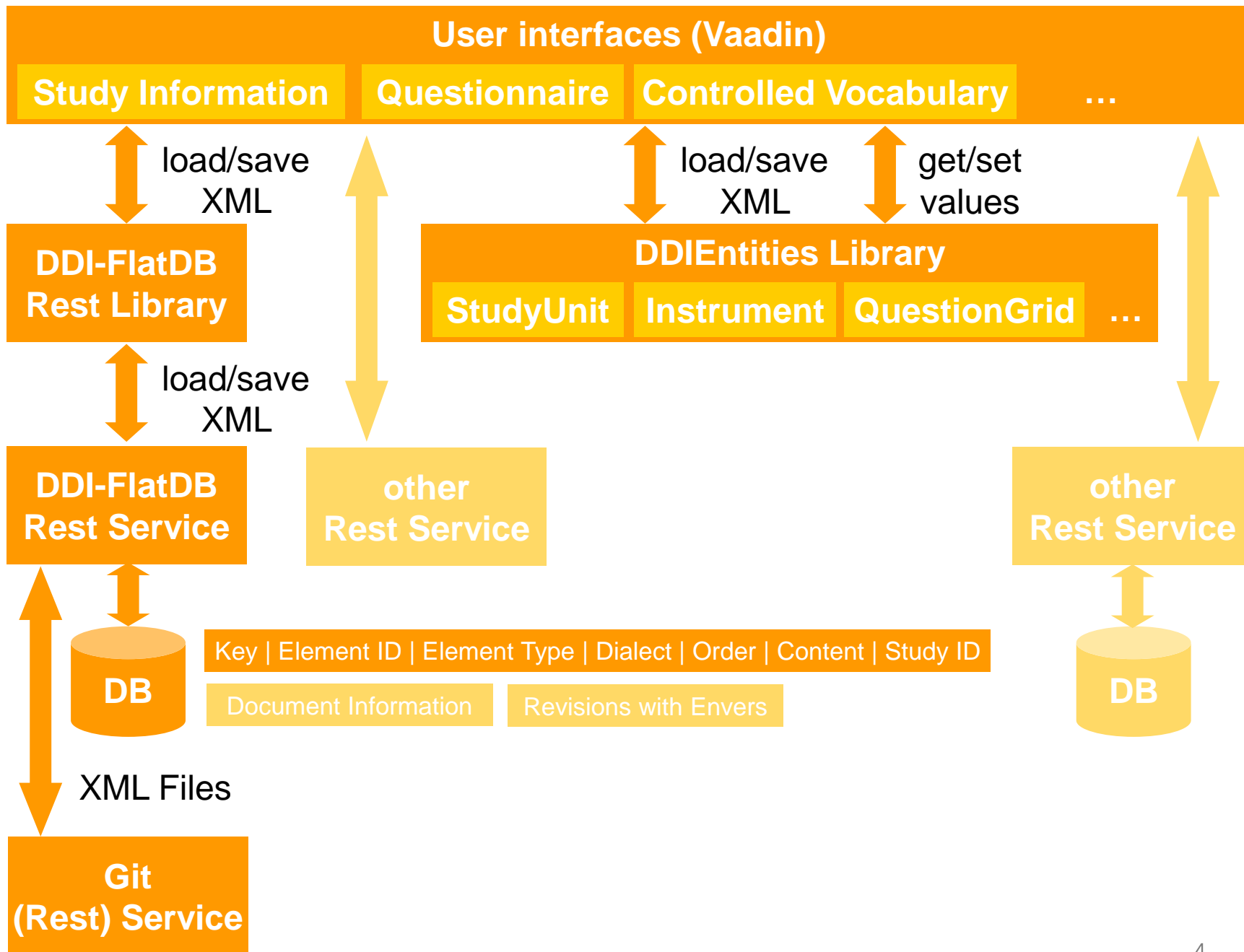
*EDDI 2017 – License [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)*

# Agenda

- Short technical overview of DDI-FlatDB
- Use cases covered
  - ▶ Content
  - ▶ Development performance
- Discussion

## General idea of DDI-FlatDB

- XML document as source of information
- Cut into pieces for „client“ performance
- Pieces cached in a database
- XML processing is configurable
  - ▶ How to cut/recombine
  - ▶ How to read/write data
  - ▶ Supporting several versions/dialects



## Use cases

- GLES questionnaire Editor (QMD)
- Study level editor (DBK case)
- Variable [documentation and] discovery (Explore Data project)
- CESSDA Controlled Vocabulary manager
- (DDILimDAS/CESSDA CMM)

# QMD

Save Abbrechen

Question number: pre001

Question label: Politisches Interesse

Filtercode:

Filterdescription:

Interviewer instruction:

Question text: Einmal ganz allgemein gesprochen: Wie stark interessieren Sie sich für Politik: sehr stark, stark, mittelmäßig, weniger stark oder überhaupt nicht?

Answers: Antworten 1

Open answer label:

Concept:

Runtime:

Answers Instructions Constructs Variables

search for/within answers

Name

Gesamt Anzeige

Antworten 1

Antworten 2

Antworten 3

Antworten 4

Antworten 5

Antworten 6

Antworten 7

Antworten 8

Antworten 9

Antworten 10

Antworten 12

Antworten 14

Antworten 16

Antworten 17

Antworten 18

Antworten 1

- [1] sehr stark

## Model classes involved in QMD

- StudyUnit
- Instrument
- Sequence
- QuestionConstruct
- QuestionItem
- QuestionGrid
- CodeList
- CategoryScheme
- TextDomain
- Universe
- Instruction
- Concept
- Variable

## Effort for QMD case

- High – about a year
  - ▶ XML-consuming
  - ▶ Bean mapping
  - ▶ UI mapping
  - ▶ „Automatic“ UI creation



# Model classes involved in Study Information

- StudyUnit
- Concept
- OtherMaterial
  - ▶ Creator
  - ▶ Contributor
  - ▶ Note

## Effort for Study Information case

- High for concept – one month
  - ▶ Modeling the use case
  - ▶ Functional specialties in UI
- Low for the data management (CRUD) – one month
  - ▶ Uncovered cases within XML consuming

# Variable discovery

**European Values Study 2008: Great Britain (EVS 2008)**

Study Description | [View Data: online analysis](#) | Download data and documents

ALL VARIABLES OF: EUROPEAN VALUES STUDY 2008: GREAT BRITAIN (EVS 2008)

Search ...

Var v50 "dont like as neighbours" ^

**Variable:** v50 "dont like as neighbours: right wing extremists (Q6E)"

**Question wording:** On this list are various groups of people. Could you please tell me any that you would not, generally speaking, like to have as neighbours?

**Response category:**

value	value label	total	freq(%)	cum.
-5	other missing			
-4	question not asked			
-3	nap			
-2	na			
-1	dk			
1	mentioned			
2	not mentioned			

# Model classes involved in Variable Information

- StudyUnit
- QuestionConstruct
- QuestionItem
- QuestionGrid
- CodeList
- CategoryScheme
- TextDomain
- Universe
- Instruction
- Concept
- Variable
- (Statistic)

## Effort for Variable Information case

- Medium for concept – some months
  - ▶ Simple modeling the use case
  - ▶ Functional specialties in UI
  - ▶ How to structure the search index
- Low for the data management (CRUD) – some days
  - ▶ Mostly covered by Questionnaire case and Study Information

# CESSDA CV Manager

Consortium of European Social Science Data Archives



cessda eric  
CV Manager

en Home Search CVs Login

← Back

**CESSDA Controlled Vocabulary for Analysis Unit**

**Title** Analysis Unit  
**Definition** Describes the entity being analyzed in the study or in the variable.  
**Code** AnalysisUnit  
**Title (en)** erhlkjdf  
**Definition (en)** alsfdj  
**Language** de

**Latest published version** 1.0-de

**Date of publication** 2017-01-07

EN DE

[Details](#)
[Identity, versions and general](#)
[DDI usage](#)
[License and copyright](#)
[Export/download](#)

Descriptive term	Descriptive term (de)	Definition	Definition (de)
Family	Familie	Any individual person, irrespective of demographic characteristics, professional, social or legal status, or affiliation.	Zwei oder mehr Menschen, die miteinander blutsverwandt sind, oder durch Heirat (einschließlich der Verwandtschaft des Ehepartners), Adoption oder Pflege (Pflegeeltern, Pflegekinder etc.) anverwandt sind - unabhängig davon, ob sie zusammenleben oder nicht. Wird zum Beispiel genutzt bei der Erforschung des Ausmaßes, in dem die Menschen ihre Verwandte unterstützen.
Organization	Organisation	Any kind of formal administrative and functional structure - includes associations, institutions, agencies, businesses, political parties, schools, etc.	Jede Art von formaler Verwaltungsstruktur und funktionaler Struktur - schließt Verbände, Institutionen, Agenturen, Unternehmen, politische Parteien, Schulen usw. ein.
Individual	afjd	Any individual person, irrespective of demographic characteristics, professional, social or legal status, or affiliation.	afjd
Family: Household family		A more specific term, refers only to related	

# Model classes involved in Controlled Vocabulary

- CVScheme
- CVConcept

(both based on SKOS)

## Effort for CV case (concept was ready)

- Low for model – few hours
  - ▶ Very simple and preexisting data model
  - ▶ Straight forward concept for functionality/UI
  - ▶ Becoming familiar with SKOS
- Very low for the data management (CRUD) – few hours
  - ▶ Two new beans
  - ▶ One dialect mapping



## Effort for DDILimDAS/CMM case

- Study level, questionnaire, simple variable
- Very low for concept – one hour
  - ▶ Two conceptual changes
    - Externalize Persons and Institutions: three new beans, one bean extended
    - Instrument „inside“ StudyUnit: functional changes in QMD and Variable Discovery
- Very low for the data management (CRUD) – half an hour
  - ▶ Defining paths for another dialect

# Discussion

- DDI-FlatDB is able to provide
  - ▶ Flexible storage across versions/dialects
  - ▶ Hides DDI complexity from end users
  - ▶ (Conversion between versions/dialects)
- Developers can focus on the service
  - ▶ Good UI integration
  - ▶ Efficient (fast) access
  - ▶ Application driven development
  - ▶ Permeable between applications
- Drawbacks
  - ▶ Lets discuss offline ;-)

# DDI-FlatDB and DDI-Entities are available...

- ... as Open Source:
  - ▶ <https://git.gesis.org/stardat/stardat-ddiflatdb>
  - ▶ <https://git.gesis.org/stardat/stardat-ddi-entities>
- ... as docker:
  - ▶ <http://play-with-docker.com/?stack=https://git.gesis.org/stardat/stardat-ddiflatdb/raw/master/doc/play-with-docker/docker-compose-v3.3.1.yml>

## Contact

- Oliver Hopt  
([Oliver.Hopt@gesis.org](mailto:Oliver.Hopt@gesis.org))
- Claus-Peter Klas  
([Claus-Peter.Klas@gesis.org](mailto:Claus-Peter.Klas@gesis.org))
- Alexander Mühlbauer  
([Alexander.Muehlbauer@gesis.org](mailto:Alexander.Muehlbauer@gesis.org))

GESIS – Leibniz-Institute for the Social Sciences

# License

This presentation is offered under license [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

The license does not apply to the following copyrighted material used in this presentation:

- Photograph of the GESIS Main Office
- GESIS logotype