

# gesis

Leibniz Institute  
for the Social Sciences



## DDI-FlatDB: Next steps

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

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

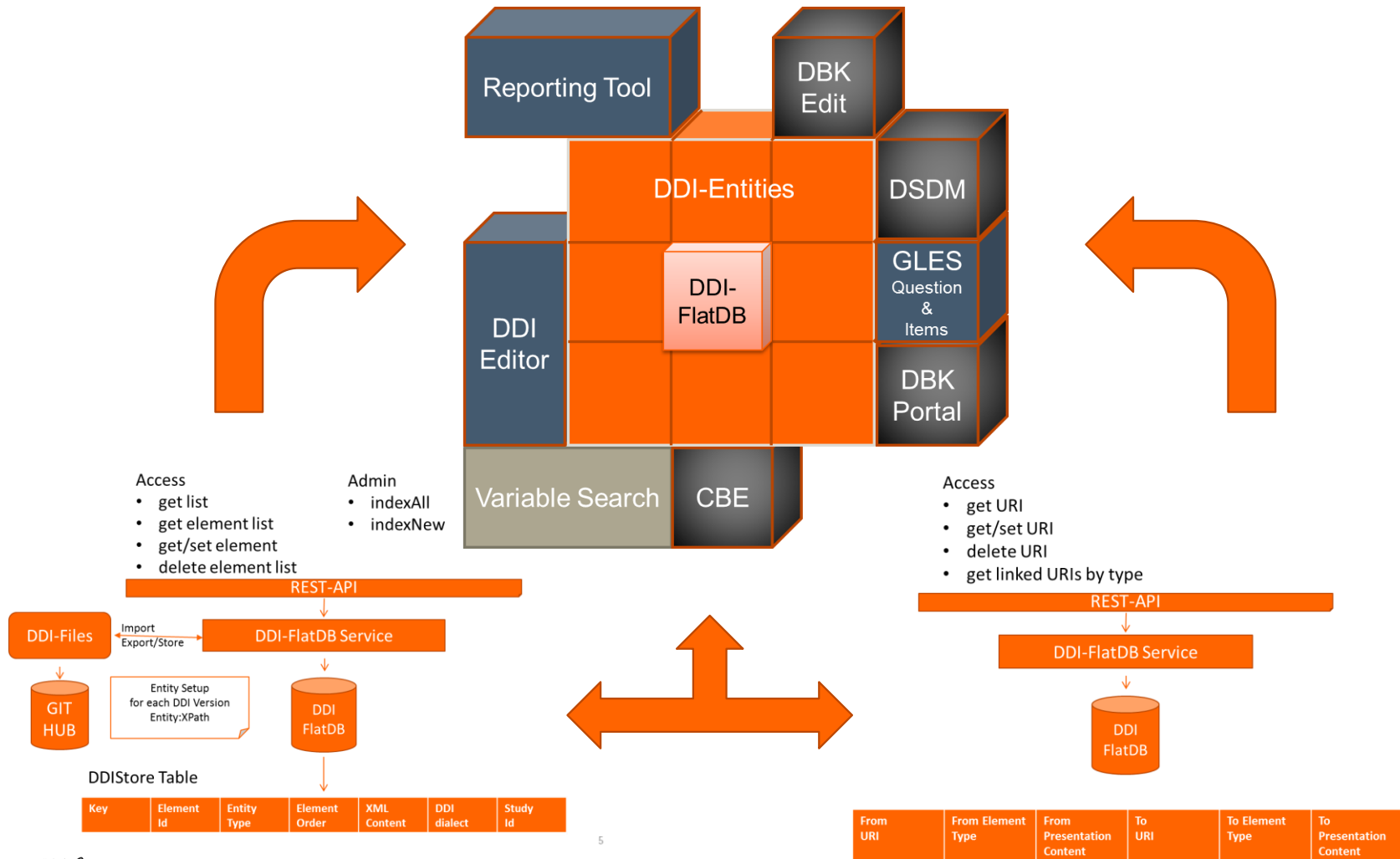
# Overview

- Efficiency and effectiveness using DDI-Flat DB?
  - ▶ Links
  - ▶ Tool to create and edit property file for accessing DDI-FlatDB entites in Java
- Integration into GESIS documentation and long term preservation workflows

# Links: Use Cases

- Internal DDI links
  - ▶ Study description to variables
  - ▶ Variables to question texts
  - ▶ ... and vice versa and et al
- External DDI links
  - ▶ Variables in prior waves of studies
  - ▶ Controlled vocabulary, etc.
- Resolving/Presentation layer
- Good thing: <\*Reference>
  - ▶ <r:ExternalURLReference>
  - ▶ <dc:InterviewerInstructionReference>

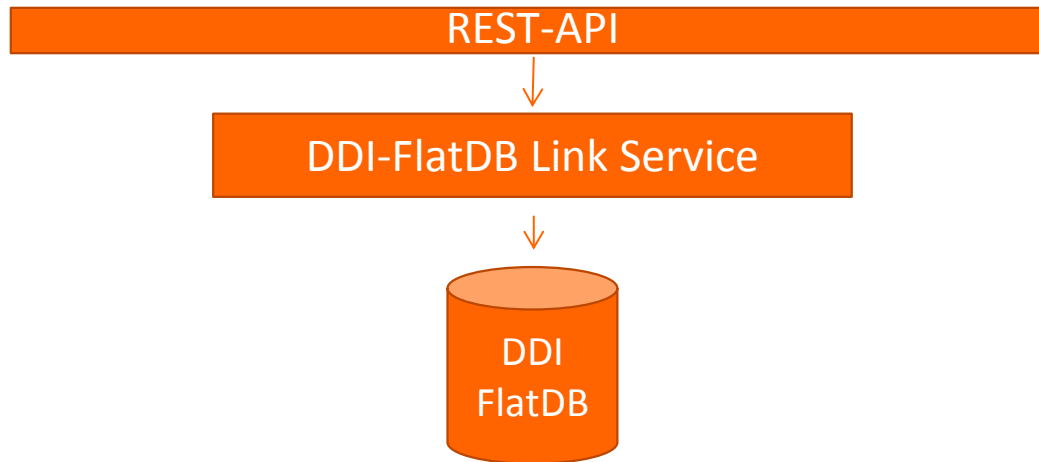
# Current DDI-FlatDB



# DDI-FlatDB Triple Store

## Access

- get URI / resolve URI
- get/set URI
- delete URI
- get linked URIs by type



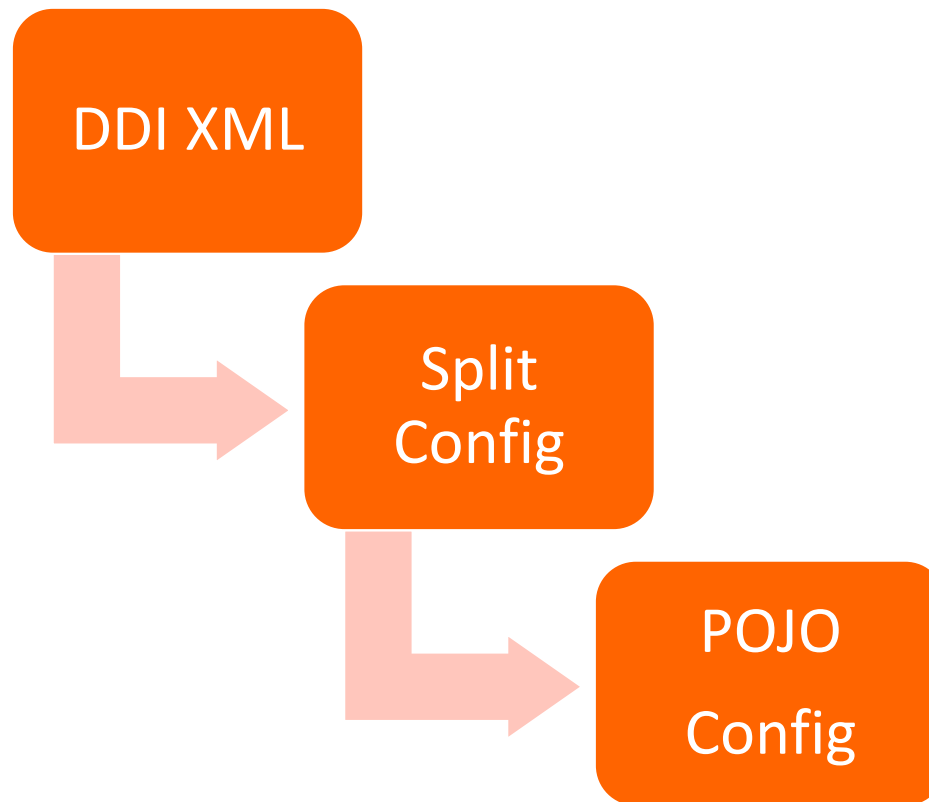
From URI	From Element Type	From Presentation Content	To URI	To Element Type	To Presentation Content
----------	-------------------	---------------------------	--------	-----------------	-------------------------

# Creating and edit property file for accessing DDI-FlatDB entites in Java

# Creating and edit property file for accessing DDI-FlatDB entites in Java

- DDI – File
- DDI – FlatDB and split config
- DDI – FlatDB, config file and POJOs
- Current status
  - ▶ Hand made
  - ▶ Error prone
  - ▶ Long developer turn arounds, because
    - Configured/implemented in application

# Accessing DDI/XML based metadata application depended





# The DDI file

```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!-- -->
<!-- Selection of DDI 3.1 Elements and Attributes used by DBK for DDI3.1 Export -->
<!-- The original DDI 3.1 Schema can be found at http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/XMLSchema/ -->
<!-- A copy of the DDI 3.1 Schema can be found at https://dbk.gesis.org/DDI/3_1 -->
<!-- -->
<!-- created by W. Zenk-Möltgen, 2014-02-03 -->
<!-- from DBK on 2015-02-04 -->
<!-- DDI3.1 export format -->
<!-- -->
<!-- DBK 2.0 to DDI 3.1 Export - 2.4.0 -->
<!-- -->
<ddi:DDIInstance id="gesis_ZA0021" agency="de.gesis" version="1.0.0" versionDate="2015-02-04" xmlns:ddi="ddi:instance:3_1" xmlns:s="ddi:studyunit:3_1"
xmlns:pd="ddi:physicaldataprodu:3_1" xmlns:pi="ddi:physicalinstance:3_1" xmlns:c="ddi:conceptualcomponent:3_1" xmlns:l="ddi:logicalproduct:3_1" xmlns:r="ddi:reusable:3_1"
xmlns:dc="ddi:datacollection:3_1" xmlns:a="ddi:archive:3_1" xmlns:xhtml="http://www.w3.org/1999/xhtml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dce="ddi:dcelements:3_1" xmlns:dc2="http://purl.org/dc/elements/1.1/" xmlns:n1="ddi:physicaldataprodu:ncube_tabular:3_1" xsi:schemaLocation="ddi:instance:3_1
http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1/XMLSchema/instance.xsd">
<r:Citation>
<r:Title>DDI3.1 study level documentation for study ZA0021 Education Expectations and Knowledge Interests (1963) </r:Title>
<r:Creator>GESIS - Leibniz Institute for the Social Sciences</r:Creator>
<r:Publisher>GESIS - Leibniz Institute for the Social Sciences</r:Publisher>
<r:PublicationDate><r:SimpleDate>2015-02-04</r:SimpleDate></r:PublicationDate>
<r:Language>en</r:Language>
<r:InternationalIdentifier type="DOI">doi:10.4232/1.0021</r:InternationalIdentifier>
<r:Copyright>All metadata from GESIS DBK are available free of restriction under the Creative Commons CC0 1.0 Universal Public Domain Dedication. However, GESIS requests that you actively
acknowledge and give attribution to all metadata sources, such as the data providers and any data aggregators, including GESIS. For further information see
https://dbk.gesis.org/dbksearch/guidelines.asp </r:Copyright>
</r:Citation>
<s:StudyUnit id="ZA0021_SU" isMaintainable="true" agency="de.gesis" version="1.0.0">
<r:UserID type="DBK Study Number">ZA0021</r:UserID>
<r:Citation>
<r:Title>Education Expectations and Knowledge Interests (1963)
</r:Title>

<r:Creator>DIVO, Frankfurt</r:Creator>

<r:Publisher>GESIS - Leibniz Institute for the Social Sciences</r:Publisher>

<r:Contributor role="Distributor">GESIS - Leibniz Institute for the Social Sciences</r:Contributor>
<r:PublicationDate>
<r:SimpleDate>1963</r:SimpleDate>
</r:PublicationDate>
<r:InternationalIdentifier type="DOI">doi:10.4232/1.0021</r:InternationalIdentifier>
<dce:DCElements>
<dc2:title>Education Expectations and Knowledge Interests (1963)
</dc2:title>

```

# DDI-FlatDB: Split a DDI file

- `#/ddi:DDIInstance/g:ResourcePackage[1]/dc:ControlConstructScheme[2]/dc:Sequence[1]`
- `Sequence.path = /DDIInstance/ResourcePackage/ControlConstructScheme/Sequence`
- `Sequence.identifierPath = ./ID`
- `Sequence.parentIdentifierPath = ../ID`
  
- `#/ddi:DDIInstance/g:ResourcePackage[1]/l:CategoryScheme[1]`
- `CategoryScheme.path = /DDIInstance/ResourcePackage/CategoryScheme`
- `CategoryScheme.identifierPath = ./ID`
- `CategoryScheme.parentIdentifierPath = ../ID`
- `CategoryScheme.successorPath = ./CodeListScheme`
  
- `#/ddi:DDIInstance/g:ResourcePackage[1]/l:VariableScheme[1]/l:Variable[1]`
- `Variable.path = /DDIInstance/ResourcePackage/VariableScheme/Variable`
- `Variable.identifierPath = ./ID`
- `Variable.parentIdentifierPath = ../ID`
  
- `#/ddi:DDIInstance/g:ResourcePackage[1]/l:CodeListScheme[1]/l:CodeList[1]`
- `CodeList.path = /DDIInstance/ResourcePackage/CodeListScheme/CodeList`
- `CodeList.identifierPath = ./ID`
- `CodeList.parentIdentifierPath = ../ID`
  
- `#/ddi:DDIInstance/g:ResourcePackage[1]/dc:InstrumentScheme[1]/dc:Instrument[1]`
- `Instrument.path = /DDIInstance/ResourcePackage/InstrumentScheme/Instrument`
- `Instrument.identifierPath = ./ID`
- `Instrument.parentIdentifierPath = ../ID`
  
- `#/ddi:DDIInstance/g:Group[1]/g:SubGroup[1]/s:StudyUnit[1]`
- `StudyUnit.path=/DDIInstance`
- `StudyUnit.identifierPath = ./Group/SubGroup/StudyUnit/ID`

# Model the POJO: Getting the splitted information

- StudyGroup.fieldNames=Label,Description,StudyReference
- StudyGroup.defaultDialect=gesisGroup32
- StudyGroup.Label.ldep=true
- StudyGroup.Label.multi=false
- StudyGroup.Label.complex=false
- StudyGroup.Description.ldep=true
- StudyGroup.Description.multi=false
- StudyGroup.Description.complex=false
- StudyGroup.StudyReference.ldep=true
- StudyGroup.StudyReference.multi=true
- StudyGroup.StudyReference.complex=false
- gesisGroup32.StudyGroup.ID.path=/ddi:DDIInstance/g:Group/r:ID
- gesisGroup32.StudyGroup.Label.path=/ddi:DDIInstance/g:Group/r:Citation/r:Title/r:String
- gesisGroup32.StudyGroup.Description.path=/ddi:DDIInstance/g:Group/r:Abstract/r:Content
- gesisGroup32.StudyGroup.StudyReference.path=/ddi:DDIInstance/g:Group/r:StudyUnitReference/r:ID
- gesisGroup32.StudyGroup.StudyReference.multiplyPath=/ddi:DDIInstance/g:Group
- gesisGroup32.StudyGroup.skeleton=<ddi:DDIInstance isMaintainable="true" versionDate="2012-09-18" scopeOfUniqueness="Maintainable" xmlns:ddi="ddi:instance:3\_2" xmlns:a="ddi:archive:3\_2" xmlns:c="ddi:conceptualcomponent:3\_2" xmlns:cm="ddi:comparative:3\_2" xmlns:d="ddi:datacollection:3\_2" xmlns:g="ddi:group:3\_2" xmlns:l="ddi:logicalproduct:3\_2" xmlns:p="ddi:physicaldatapoint:3\_2" xmlns:pi="ddi:physicalinstance:3\_2" xmlns:pr="ddi:ddiprofile:3\_2" xmlns:r="ddi:reusable:3\_2" xmlns:s="ddi:studyunit:3\_2" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xhtml="http://www.w3.org/1999/xhtml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="ddi:instance:3\_2 http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema-instance.xsd"><r:Agency xmlns:r="ddi:reusable:3\_2">org.gesis</r:Agency><r:ID xmlns:r="ddi:reusable:3\_2">DDI</r:ID><r:Version xmlns:r="ddi:reusable:3\_2">1.0.0</r:Version><g:Group xmlns:g="ddi:group:3\_2"><r:Agency xmlns:r="ddi:reusable:3\_2">org.gesis</r:Agency><r:ID xmlns:r="ddi:reusable:3\_2">tobeset</r:ID><r:Version xmlns:r="ddi:reusable:3\_2">1.0.0</r:Version><r:Citation><r:Title><r:String xml:lang="de"></r:String></r:Title></r:Citation><r:Abstract><r:Content xml:lang="de"></r:Content></r:Abstract><r:StudyUnitReference><r:Agency xmlns:r="ddi:reusable:3\_2">org.gesis</r:Agency><r:ID xmlns:r="ddi:reusable:3\_2">tobeset</r:ID><r:Version xmlns:r="ddi:reusable:3\_2">1.0.0</r:Version><r:TypeOfObject>StudyUnit</r:TypeOfObject></r:StudyUnitReference></g:Group></ddi:DDIInstance>

# Model DDI Data

## Input

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!-- -->
<!-- Selection of DDI 3.1 Elements and Attributes used by
DBK for DDI3.1 Export -->
<!-- The original DDI 3.1 Schema can be found at
http://www.ddialliance.org/Specification/DDI-
Lifecycle/3.1/XMLSchema/ -->
<ddi:DDIInstance id="gesis_ZA0021" agency="de.gesis"
version="1.0.0" versionDate="2015-02-04"
xmlns:ddi="ddi:instance:3_1" xmlns:s="ddi:studyunit:3_1"
xmlns:pd="ddi:physicaldataprodu:3_1"
xmlns:pi="ddi:physicalinstance:3_1"
xmlns:c="ddi:conceptualcomponent:3_1"
xmlns:l="ddi:logicalproduct:3_1"
xmlns:r="ddi:reusable:3_1"
xmlns:dc="ddi:datacollection:3_1"
xmlns:a="ddi:archive:3_1"
xmlns:xhtml="http://www.w3.org/1999/xhtml"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:dce="ddi:dcelements:3_1"
xmlns:dc2="http://purl.org/dc/elements/1.1/"
xmlns:n1="ddi:physicaldataprodu:ncube_tabular:3_1"
xsi:schemaLocation="ddi:instance:3_1
http://www.ddialliance.org/Specification/DDI-
Lifecycle/3.1/XMLSchema/instance.xsd">
<r:Citation>
<r:Title>DDI3.1 study level documentation for study
ZA0021 Education Expectations and Knowledge Interests
(1963) </r:Title>
<r:Creator>GESIS - Leibniz Institute for the Social
Sciences</r:Creator>
<r:Publisher>GESIS - Leibniz Institute for the Social
Sciences</r:Publisher>
```

## Config

```
#/ddi:DDIInstance/g:ResourcePackage[1]/l:Var-
iableScheme[1]/l:Variable[1]
Variable.path =
/DDIInstance/ResourcePackage/VariableSche-
me/Variable
Variable.identifierPath = ./ID
Variable.parentIdentifierPath = ../ID

#/ddi:DDIInstance/g:ResourcePackage[1]/Co-
deListScheme[1]/l:CodeList[1]
CodeList.path =
/DDIInstance/ResourcePackage/CodeListSche-
me/CodeList
CodeList.identifierPath = ./ID
CodeList.parentIdentifierPath = ../ID

#/ddi:DDIInstance/g:ResourcePackage[1]/dc:In-
strumentScheme[1]/dc:Instrument[1]
Instrument.path =
/DDIInstance/ResourcePackage/InstrumentSch-
eme/Instrument
Instrument.identifierPath = ./ID
```

## Output

### StudyUnit

\* <xml>

### Variable

\* <xml>

### QuestionItem

\* <xml>

# Model DDI Data

## Input

StudyUnit

\* <xml>

Variable

\* <xml>

QuestionItem

\* <xml>

## Config

Type: Variable

Field: VariableLabel

Languages: True

Multi: True

Xpath: /

## Output

Beruf Partner/In (D11.1)

Welchen Beruf übt Ihr Partner/Ihre Partnerin aus? Darf ich Sie bitten, seinen/ihr

Antworten 13

- [1] sehr wahrscheinlich
- [2] ziemlich wahrscheinlich
- [3] eher unwahrscheinlich
- [4] sehr unwahrscheinlich
- [98] weiß nicht
- [99] keine Angabe

Intervieweranweisung[offene Antwort(en) anhand vorgegebener Parteiliste f

EingangsfilerDies ist der neue Filter

if vw17=3

# Goals

- Efficient and error less creation of config files
- Better modelling
- Faster turn-around for developer

# GESIS: DDI Datasources

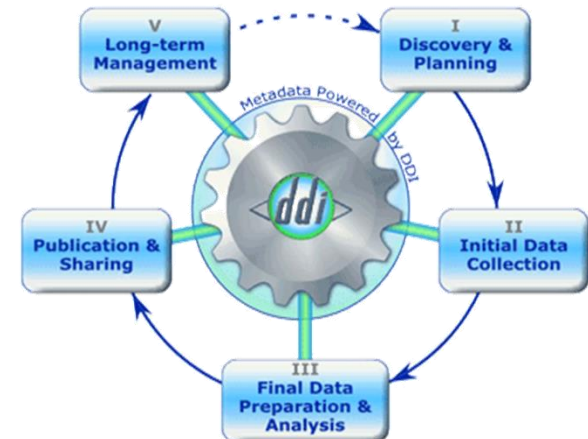
- DBK: StudyUnit
- DSDM: Variables
- CodeBookExplorer: Questions/Variables
- German Election Studies: StudyUnit/Questions
- CVMManager (CESSDA):

***Application dependend DDI-FlatDBs***

***Next step: Connect/Integrate***

# New application several input points

- Early input
  - ▶ Creation of questionnaire
    - Existing wave
    - New study
- Late input
  - ▶ Existing study; new wave
  - ▶ Complete new study; (almost) only statistical file





## Decisions to be discussed

- Central resource packages or resources packages per study?
  - ▶ Duplicate data
- GESIS wide URI or DDI wide URI?
- How to SIP?
- ...

# Summary

- DDI FlatDB (2018)
  - ▶ Links will be handled
  - ▶ Developers will be supported using modelling DDI
  - ▶ Adopted into „real“ internal GESIS workflow

Thank you

# 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