Data Model and Data Standard - a Happy Marriage?

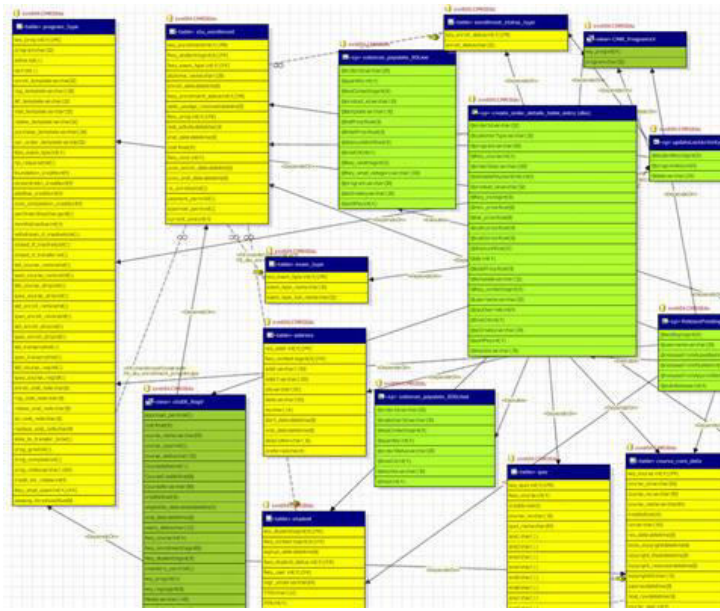
Dr. Brigitte Mathiak
Oliver Hopt
GESIS - Leibniz Institute for the Social Sciences

This talk is about

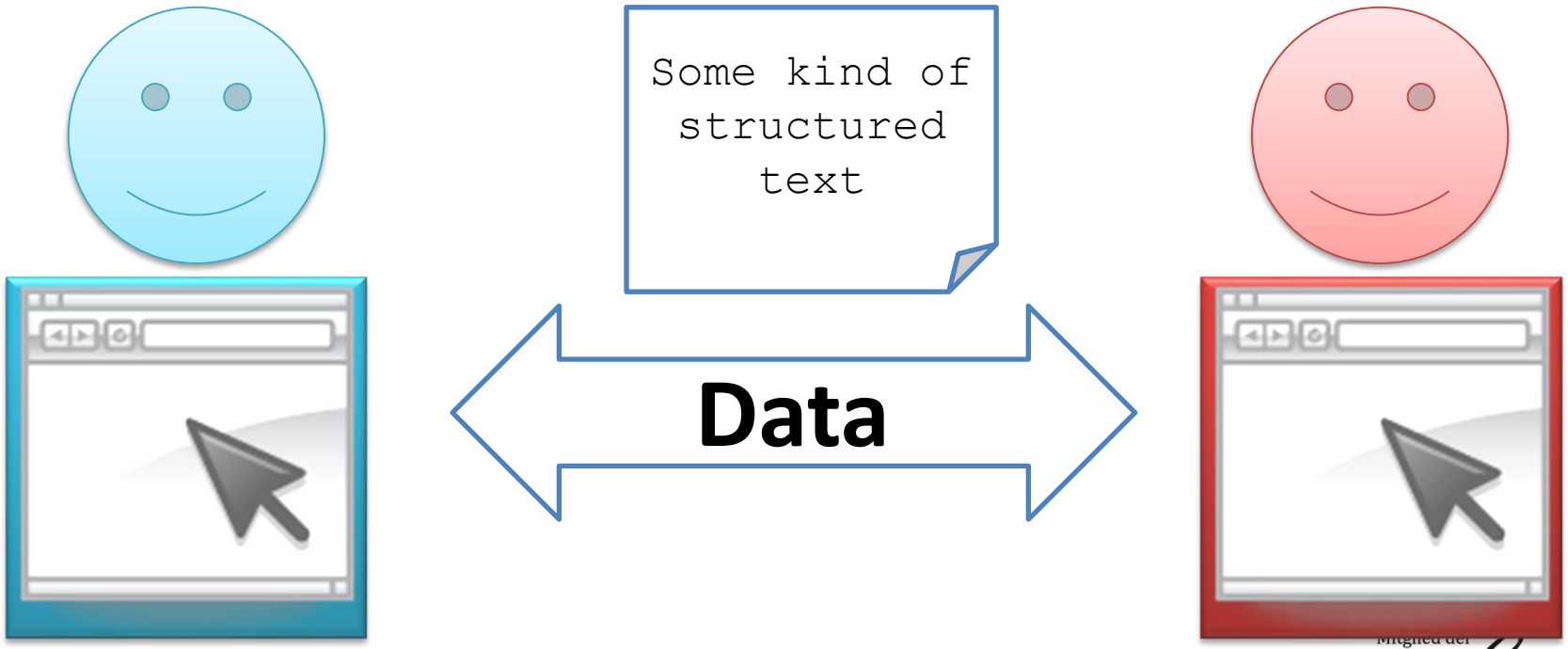
- Problem statement
- DDI decisions
- Technical exploration
- Open issue

What is a Data Model for?

Data



What is a Data Standard for?



Data Model vs. Data Standard

- Individual
- Object orient
- Platform depen
- Modularity for abstraction

- General
- Not object oriented
- Platform independent
- Modularity for content grouping

- (Proprietary)...
- (Open)…

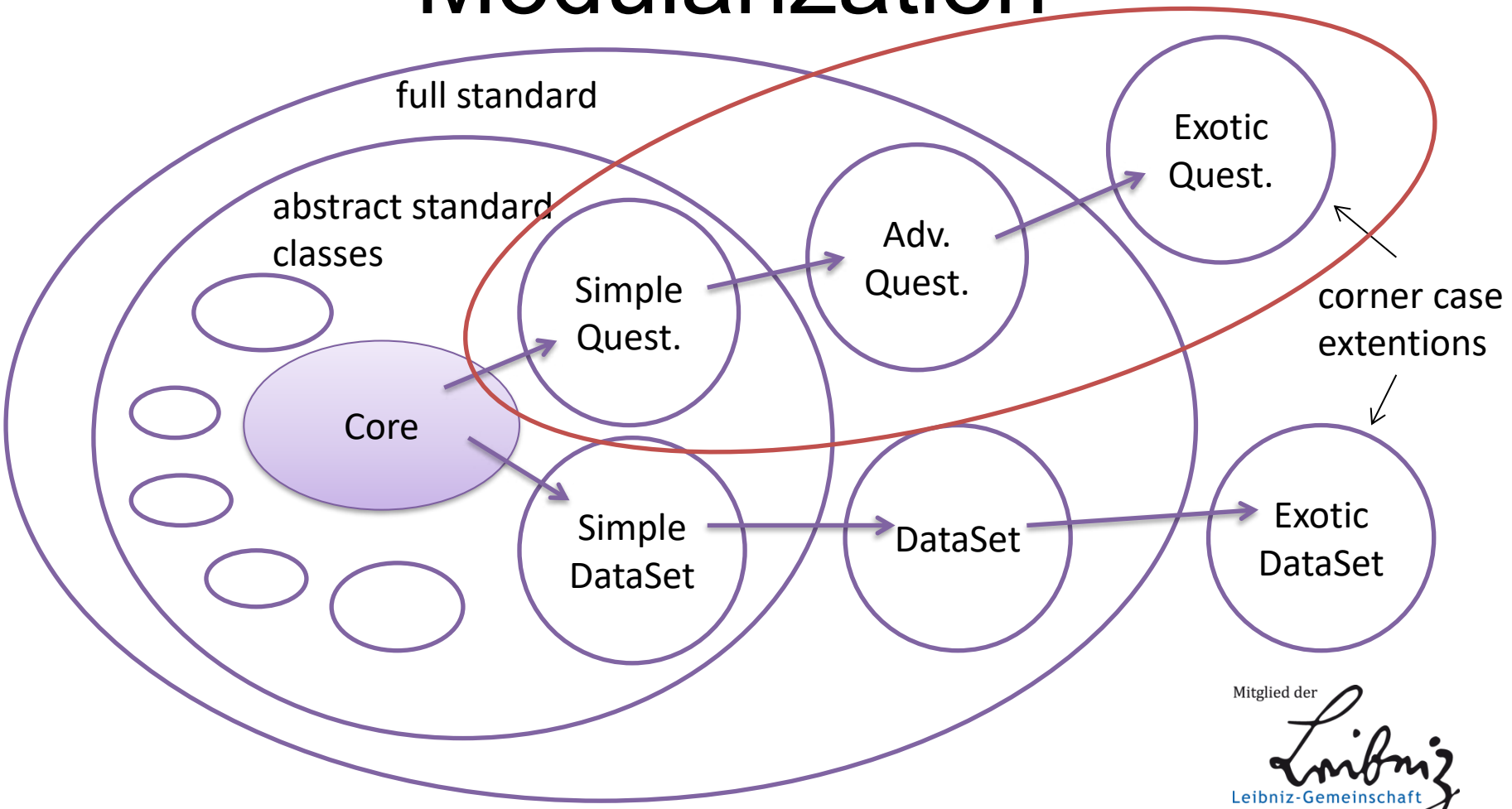
DDI development from the technical point of view

- DDI 2 appears as plain data standard.
- DDI 3 seems to be a data model expressed in data standard language.

- DDI future is planned as a data model with mapping towards a data standard.

→ Can this work?

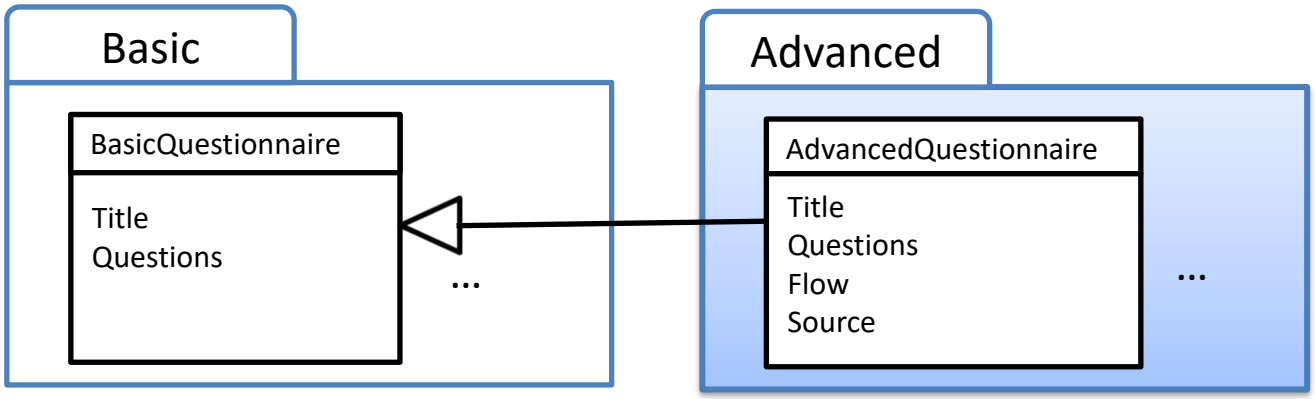
Modularization



The happy part of the marriage

- Technology stack: POJOs, JAXB, JPA
- Generating POJOs from UML
- Direct mapping into relational database (JPA)
- Identification handling (db versus ddi)
- Out of the box (un-)marshalling works!!!

Trouble with modularity

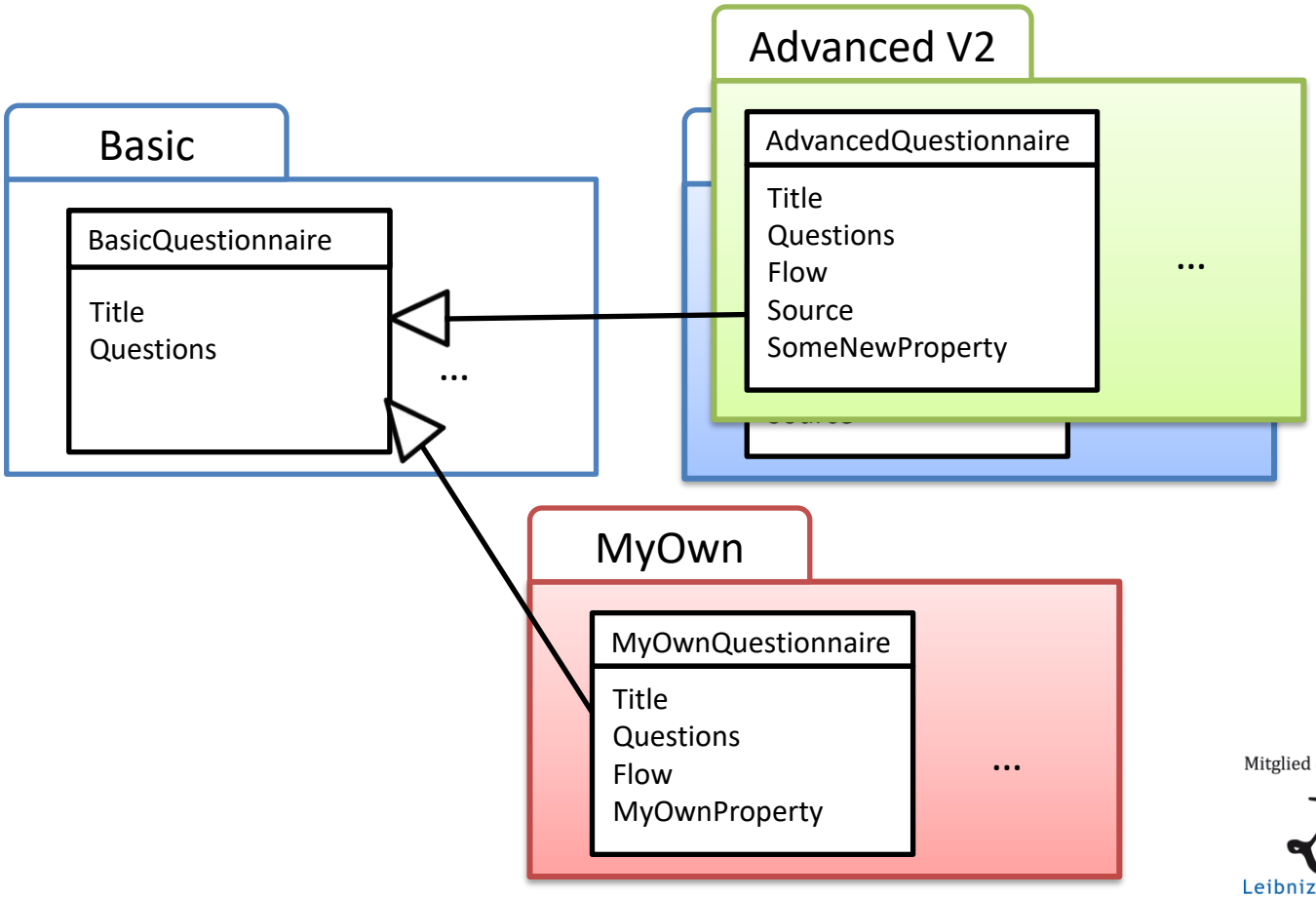


- My system can only deal with BasicQuestionnaires.
- I get delivered a document with AdvancedQuestionnaires.

  → No library to determine the mapping on standard level.

  → Handmade mapping for each and every interaction.

Increasing the trouble



Solution I: Exchange layer

- Provide XSL-snippets for up-casting and version change.
- Compile a style sheet from XSL-snippets based on the used namespaces in source an destination.
- Upstream transformation of each import

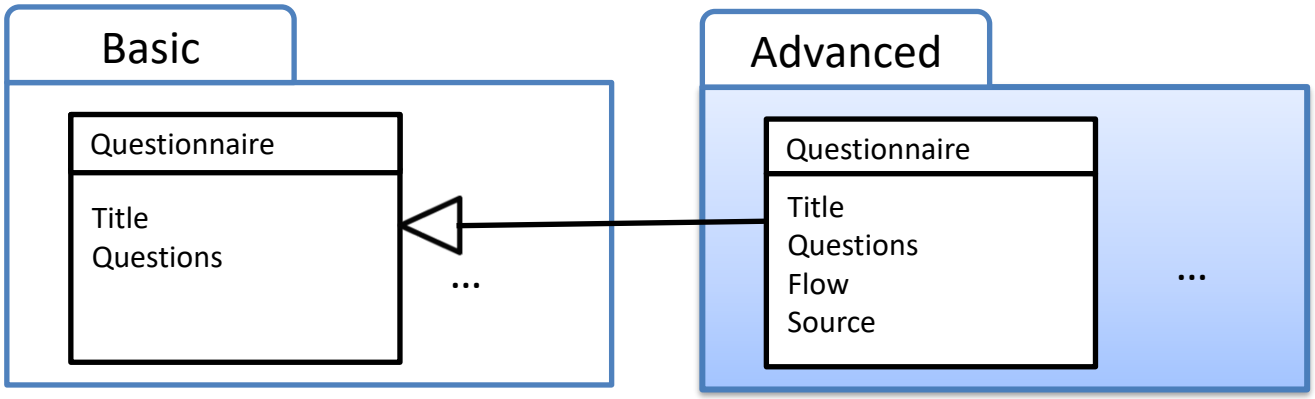
- Lot of work and issues

Solution II: Application layer

- Extend POJOs with methods for up-casting and version change.
- Provide an API with all modules in each version.
- Cast to fitting objects before storing.

- Lot of work and issues

Solution III: Cheating



- Parser are usually implemented to ignore anything they do not understand
- So, an implementation of the Questionnaire from Basic will typically be able to parse an Advanced Questionnaire input, no casting/transformation needed

Issues with solution III

- Somewhat weird to have two objects with different properties, but the same name in the same standard
- Requiring anything not in Basic is impossible
- xs:any needs to be added to everything

- Different modeling philosophy is needed

Time for Discussion

Contact:

brigitte.mathiak@gesis.org

oliver.hopt@gesis.org