

# Linking Study Descriptions to the Linked Open Data Cloud

Johann Schaible, Ben Zopilko, Thomas Bosch and  
Wolfgang Zenk-Möltgen

GESIS – Leibniz Institute for the Social Sciences

# Agenda

- Motivation
- Approach
  - Mapping of Entities
  - Generating RDF
  - Link Discovery with Silk (Workflow and Demo)
- Results and Discussion

# Motivation

- Find additional information on Resources inside the study descriptions<sup>1</sup>
- Creating a single point of information access
- Higher visibility for users and other data sources
- Why Linked Open Data?
  - RDF based data is more available (LOD-Cloud)
    - Easy access because of Linked Data Principles
  - Lots of Semantic Web tools for RDF data

1. <http://www.gesis.org/en/services/research/data-catalogue/>

# Approach

1. Map Entities to Vocabularies
2. Use XSLT to transform the XMLs to RDF
3. Identifying external data sources
  - a. Use Link Discovery tools, e.g. Silk, to find equal entities in external data sources
  - b. Store discovered links into our representation

# Mapping of Entities

- Extraction of only most promising entities
- Mapping to commonly used vocabularies

Entity	dc Vocabulary	foaf Vocabulary	swrc Vocabulary
Title	dcterms:title	-	-
Creator	dcterms:creator	-	swrc:name
Funding Agency	-	foaf:fundedBy	-
Topical Coverage	dcterms:subject	-	-
Spatial Coverage	dcterms:coverage	-	-
Temporal Coverage	dcterms:date	-	-

# Generating RDF (Simplified Example)

XML:

```
<titleStmt>
  <title xml-lang="en"> Politbarometer Eastern Germany 2003 </title>
</titleStmt>
```

XSLT:

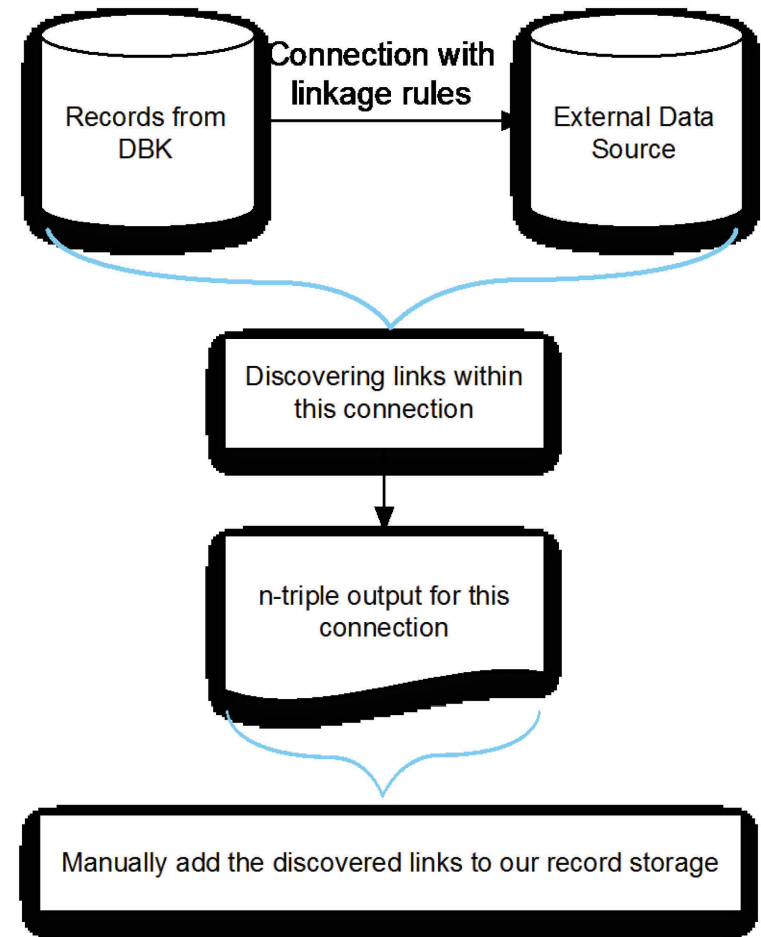
```
<xsl:element name="dcterms:title">
  <xsl:attribute name="xml:lang"> en </xsl:attribute>
  <xsl:value-of select="titleStmt/title"/>
</xsl:element>
```

RDF:

```
<dcterms:title xml:lang="en">
  Politbarometer Eastern Germany 2003
</dcterms:title>
```

## Silk - Workflow

1. Define linkage rules and linking scenrio
2. Silk detects links between the data sources
3. Silk stores the linkage in a n-trilpe file
4. Manual integration of links into our records



## Silk - Demo

The screenshot displays the Silk Workbench interface. At the top, there is a dark header bar with the text "Silk Workbench" on the left, "Workspace: Showcase" in the center, and "About" on the right. Below the header is a toolbar containing several icons and buttons: a folder icon, a document icon labeled "Project", a document icon with a plus sign labeled "Import", a folder icon labeled "Showcase", a wrench icon labeled "Prefixes", a green plus sign labeled "Source", a green plus sign labeled "Task", a green plus sign labeled "Output", a document icon labeled "Link Spec", a document icon labeled "Export", and a red X icon labeled "Remove". On the right side of the interface, there is a sidebar with two sections: "Overview" and "Next Steps".

**Overview**  
Use the workspace to manage different projects. Each project consists of data sources, linking tasks and output tasks.

**Next Steps**  
Add the data sources you want to interlink.

2. <http://www4.wiwiw.fu-berlin.de/bizer/silk/>
3. <http://www.youtube.com/watch?v=CPI5eeyW-Xk>



# Results and Discussion

- Mapping very hard
  - Future Work: For mapping use DDI Ontology
- Silk discovers links in an easy way, but:
  - It occurred a few times, that Silk did not find exact matches
  - Silk had a lot of problems with SPARQL Endpoints
  - Loading RDF Dumps bigger than 200MB takes a long time
  - Future Work: Explore Silk; Develop more complex linkage rules
- Too few data sources with relevant data for Social Science
  - Future Work: Encourage people to publish as Linked Data; Find external data sources automatically

# Q&A

## Thank you

