

Metadata Requirements to document Data Analyses and Syntax Files in a Virtual Research Environment (VRE)

-

The use case soeb 3

Uwe.Jensen@gesis.org

EDDI14 – 6th Annual European DDI User Conference December 2-3, 2014 in London, UK

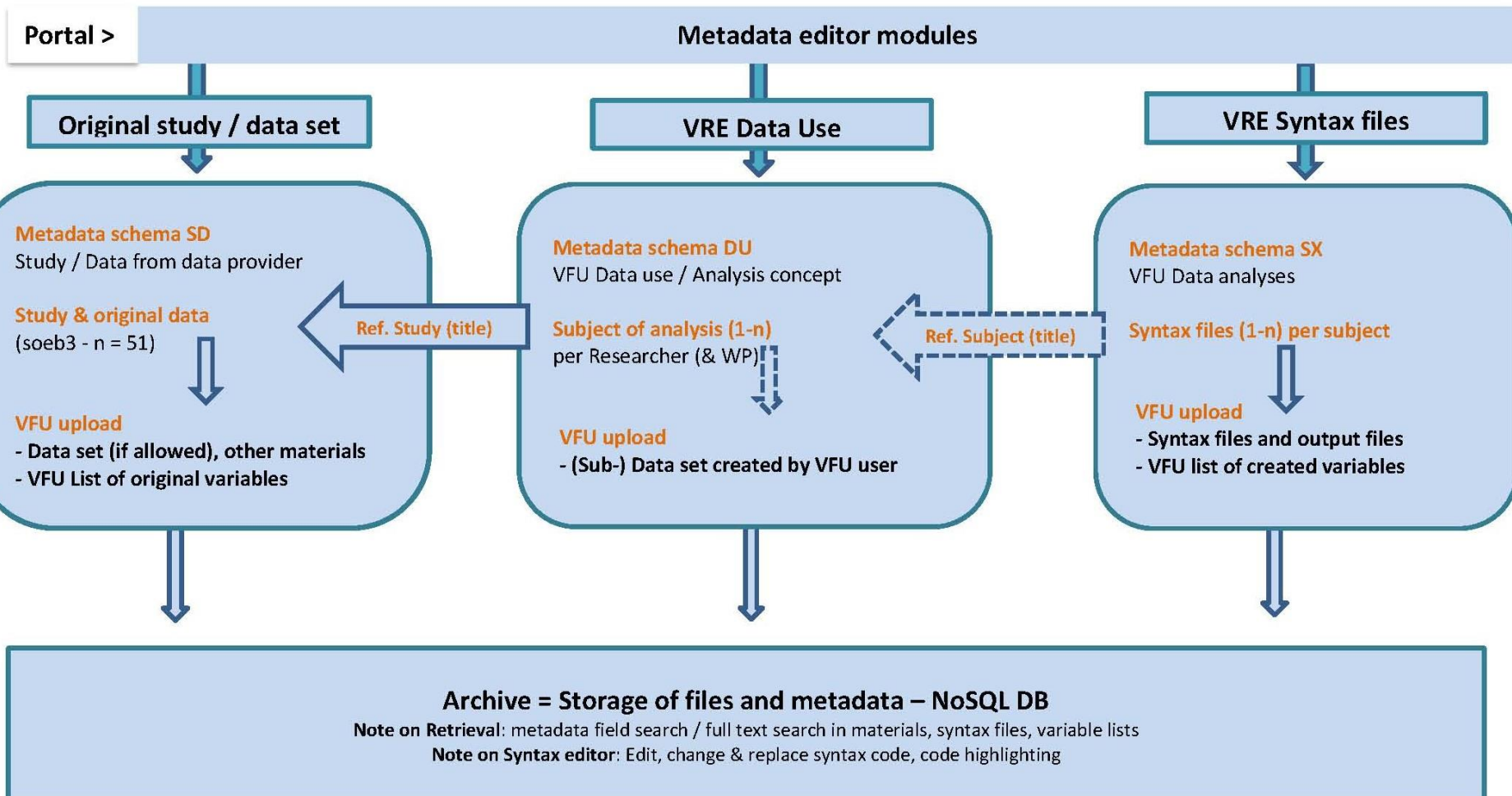
Overview

- Project & technical system
- Metadata workflow VRE soep3
- Commonalities of the 3 MD - Schema
- Metadata Study & data - SD schema
- Metadata Data Use concept - DU schema
- Metadata Syntax - SX schema
- VRE metadata & DDI - An ad hoc sketch & questions

The soeb project & VRE technical system

- **soeb** Reporting on socio-economic development in Germany / 3rd. Report started autumn 2013
 - 30+ researcher, 18 WPs; 50 research data sets (cross-section, panel data, indicator & indices)
 - <http://www.soeb.de> – “soeb3” & “VRE soeb3” are funded by Federal Ministry of Education and Research
- **VRE soeb3** to support collaborative research – Project time 06.2012 – 02.2014
 - **Presentation on Metadata focus** data use & data analyses
- **Technical System & metadata editor**
 - Portal: Liferay / Archive: NoSQL /
 - Three web based metadata editing tools / Metadata represented as JSON objects
 - **SD editor**: document original data
 - **DU editor**: concept description of data usage
 - **SX editor**: syntax files & related output files

VRE soeb3 Metadata Workflow > Original data > Data use concept > Syntax files



Commonalities of schema **SD, DU, SX**

- **Administration of object SD, DU, SX**
 - Responsible person / plus related soep3 WP at DU & SX
 - Object title, Object ID, Version, Ref. with related object SD<DU<SX
 - Publish flag / Duplicate flag
- **File Upload into VRE > relationship to parent object: SD / DU / SX**
 => **SD**: Material + Data / **DU**: self generated Data / **SX**: Syntax & Output
- **Metadata** regards:
 - **Admin**: Person (+ WP); Parent Title + Ref. to related SD or DU or SX
 - **Object**: Title, Type (CV), Content description, File name,
 - **Sharing of Data file (SD & DU) and Syntax file (SX) by**
 - > Conditions of use > e.g. potential restriction, recommend citation, ...
 - > Date of approval, Access right per person (+ WP) (based on VRE Rights & Roles concept)

1. Metadata on Original Data – SD schema

- **Objective VRE:**
 - document provenance of original data (> Preservation / Public service)
 - log & control use of restricted data > support contract management

Metadata sections & elements:

- **Original Data Source & Data access:**
 - Provider name & URL
 - Access conditions (CV: Free, Controlled; Contract) & URL
 - Access type (CV: PUF, SUF, Remote access, Onsite)
 - VFU Data use: per Researcher & WP > Contract (yes/no) > details ...
- **Study description (subset of DDI elements):**
 - Bibliography; Methods; Data description; Materials
 - Mainly used: CV / Free text & links to web source
 - Problem: little scope of data with PIDs

2. MD on Data Use concept – **DU** schema

- **Objective VRE:**
 - Describe the concepts for Data Analyses per Researcher / WP
 - Share info on analysis concepts & strategies among Researchers

Metadata sections & elements:

- **Subject of analysis / analysis purpose:**
 - **Subject Title / Subject Description**
 - **Data file type** (work file(s)) (CV: subset, matched from complex data, original data, self generated data);
 - **Data characteristics & data use re. subject** (dataset structure, wide/long format,)
 - Ref. with Study (> SD title) and original Data Set(s);
- **Analysis items:**
 - time period covered; geographic coverage, smallest regional unit (CV), unit of observation considered analysis methods (> specific syntax)

3. MD about Syntax files – SX schema (1)

- **Objective VRE:**
 - Document general characteristics & functions “about the syntax”
 - Detailed comments are still embedded “in syntax code” (> Parser; / Full-text search)
 - Upload syntax & output files > search, edit & share syntax in VRE

Metadata sections & elements:

- **Sec 1. Formal syntax characteristics**
 - Ref to related subject at Data Use (if applicable)
 - Type of syntax structure - CV: Singular syntax / Syntax set / Syntax Module (e.g. household typology)
 - Statistic software + additional module
 - Syntax title
 - Syntax description (very general > specific later)
 - Syntax file name
 - Storage Date (in VRE)

3. MD on Syntax files – **SX** schema (2)

- **Sec 2. Describe general syntax purpose along four phases**
 - CV on 4 phase to select:
 > Phase 1: Select data (input file) > Phase 2: Process variables > Phase 3: Create data output file > Phase 4: Run data analyse
 - Describe general purpose of the syntax file per phase (FT);
 - Indicate status of the syntax development (CV)
- **Sec 3. Annotate Syntax operations at selected phase**
 - P1: Inform on characteristics of input data file(s) (FT);
 - P2: Describe derivation or change of single variables:
 - » Describe scientific purpose & actions (used var(s), recodes ...) (FT)
 - » Document the new variable (Var name, Values & labels)
 - P3 Inform on characteristics of analysis file (FT)
 - P4 Describe methods & procedures used in the data analyses (FT)

Note: FT: free text / CV: controlled vocab

VRE schema & DDI modules – Ad hoc sketch & Questions

VRE schema (no. of all elements)	DDI 2.5	DDI 3.1 / 3.2
SD, DU, SX Commonalities		
Sec VFU Administration (19/12/12) Sub-Schema VFU Upload (10) & Admin (4/5/5)	> to check Other materials > to check	Administrative metadata: to check Archive Specific: to check Organization Scheme: to check
Schema SD – Original data (59) applied set ~25		
Sec Original Data source & Access conditions (7)	Codebook 2.5 > Elements available Missing: > to check	Archive > Elements available Missing: > to check
Sec Bibliography (18), Methods information (15), Data description (18), Materials (10)	Codebook 2.5 > Elements available Missing: > to check	Study Unit > Elements available Data collection > Methodology elements Logical Data Product ...
DU – Data Use / Data Analysis concept (10)	Generic metadata to document Data Analysis concepts. A new component / module?	
Sec Subject of investigation (5)	No elements	Elements on subject, used data ...?
Sec Analysis items (5)	No elements	Elements on analysed items?
SX – Syntax files (22)	Generic metadata on syntax / syntax files. A new event & file type to document?	
Sec Formal syntax characteristics (8)	Elements not available	Elements on syntax characteristics?
Sec General syntax purpose at phase 1-4 (3)	Elements not available	Elements on syntax life-cycle events?
Sec Syntax operations at selected phase (2)	Elements not available	Elements on syntax operations at events?
- Processing variables (9)	- Variable elements	Data processing / Coding instruction > to check

Resources

- VFU soeb3: <http://www.soeb.de/vfu-soeb-3/>
- Uwe Jensen, Stefan Schweers
[Das erweiterte Metadatenschema der VFU soeb 3](#). GESIS-Technical Reports 2014-13
- Uwe Jensen, Stefan Schweers, Zeljko Carevic
[Die Metadateneditoren der VFU soeb 3](#). GESIS-Technical Reports 2014-14

Thank you!

Uwe.Jensen@gesis.org