

# An Evaluation of Metadata and Data Quality on Person-Level, Aggregated, Thesauri, Statistical Classifications, and Rectangular Data Sets

Thomas Bosch<sup>1</sup>, Benjamin Zapilko<sup>1</sup>, Joachim Wackerow<sup>1</sup>, and Kai Eckert<sup>2</sup>

<sup>1</sup> GESIS – Leibniz Institute for the Social Sciences, Germany  
{firstname.lastname}@gesis.org,

<sup>2</sup> University of Mannheim, Germany  
kai@informatik.uni-mannheim.de

**Abstract.** From 2012 to 2015 together with other Linked Data community members and experts from the social, behavioural, and economic sciences (*SBE*), we developed diverse vocabularies to represent *SBE* meta-data and data in RDF. The *DDI-RDF Discovery Vocabulary (Disco)* is designed for the purposes of supporting the dissemination, management, and reuse of person-level data (data about individuals) collected in form of responses to studies and archived for research purposes. The *RDF Data Cube Vocabulary (Data Cube, QB)* is a W3C recommendation for representing *data cubes*, i.e. multi-dimensional aggregate data, in RDF. *Physical Data Description (PHDD)* is a vocabulary to represent aggregated and person-level data in a rectangular format. The data could be either represented in records with character-separated values (*CSV*) or in records with fixed length. The *Simple Knowledge Organization System (SKOS)* is a vocabulary to represent knowledge organization systems such as thesauri, classification schemes, subject heading lists, and taxonomies in RDF. *XKOS* is a SKOS extension to describe formal statistical classifications like the *International Standard Classification of Occupations (ISCO)*.

To ensure high quality and trust, RDF data as well as its metadata must satisfy certain criteria - specified in terms of RDF constraints. In this paper, we evaluated the metadata and data quality of large real world aggregated (*QB*), person-level (*Disco*), thesauri (*SKOS*), rectangular (*PHDD*), and statistical classification (*XKOS*) data sets by means of RDF constraints assigned to multiple constraint types. We validated more than 4.2 billion triples and 15 thousand data sets using the *RDF Validator* which is available at <http://purl.org/net/rdfval-demo>.

**Keywords:** RDF Validation, RDF Constraints, DDI-RDF Discovery Vocabulary, Disco, RDF Data Cube Vocabulary, QB, Thesauri, SKOS, Rectangular Data, PHDD, Statistical Classifications, Extended SKOS, XKOS, Linked Data, Semantic Web

## 1 RDF Validation of Metadata and Data

Bosch et al. identified in total 74 requirements to formulate RDF constraints; each of them corresponding to a constraint type. We published a technical report<sup>3</sup> in which we explain each requirement (constraint type) in detail and give examples for each (represented by different constraint languages). The knowledge representation formalism *Description logics (DL)*, with its well-studied theoretical properties, provides the foundational basis for each constraint type. Therefore, this technical report contains mappings to DL to logically underpin each requirement and to determine which DL constructs are needed to express each constraint type [2]. We recently published a technical report in which we describe constraints to validate metadata on person-level, aggregated data, and thesauri. We assign each constraint to constraint types corresponding to RDF validation requirements or to data model specific constraint types<sup>4</sup> [3].

We distinguish two validation types: (1) *Content-Driven Validation*  $\mathcal{C}_C$  contains the set of constraints ensuring that the data is consistent with the intended syntax, semantics, and integrity of given data models. (2) *Technology-Driven Validation*  $\mathcal{C}_T$  includes the set of constraints which can be generated automatically out of data models, such as cardinality restrictions, universal and existential quantifications, domains, and ranges. We determined the default *severity level* (corresponds to requirement *R-158*) for each constraint to indicate how serious the violation of the constraint is. We propose an extensible metric to measure the continuum of severity levels ranging from  $\mathcal{SL}_0$  (informational) via  $\mathcal{SL}_1$  (warning) to  $\mathcal{SL}_2$  (error). Although we provide default severity levels for each constraint, users should be able to specify severity levels of constraints they need to validate for their individual use cases, i.e., users should be able to define use case specific severity levels for constraints.

## 2 Evaluation

We exhaustively evaluated the metadata quality of large real world aggregated (*QB*), person-level (*Disco*), and thesauri (*SKOS*) data sets by means of both  $\mathcal{C}_C$  and  $\mathcal{C}_T$  constraints of the majority of the constraint types. We validated 9,990 / 3,775,983,610 (*QB*), 4,178 / 477,737,281 (*SKOS*), and 1,526 / 9,673,055 (*Disco*) data sets / triples using the *RDF Validator*<sup>5</sup> (available at <http://purl.org/net/rdfval-demo>) in batch mode. That are more than 4.2 billion triples and 15 thousand data sets. We validated, i.a., (1) *QB* data sets published by the *Australian Bureau of Statistics (ABS)*, the *European Central Bank (ECB)*, and the *Organisation for Economic Co-operation and Development (OECD)*, (2) *SKOS* thesauri like the *AGROVOC Multilingual agricultural thesaurus*, the *STW Thesaurus for Economics*, and the *Thesaurus for the Social Sciences (TheSoz)*, and

<sup>3</sup> Available at: <http://arxiv.org/abs/1501.03933>

<sup>4</sup> Requirements/Constraint types and constraints are uniquely identified by alphanumeric technical identifiers like *R-1*

<sup>5</sup> For details about the *RDF Validator* see [1]

(3) *Disco* data sets provided by the *Microdata Information System (Missy)*, the *DwB Discovery Portal*, the *Danish Data Archive (DDA)*, and the *Swedish National Data Service (SND)*. As we evaluated nearly 10 thousand *QB* data sets, we published the evaluation results for each data set in form of one document per SPARQL endpoint<sup>6</sup>. The correctness of all constraints, i.e., the gold standard, has been proved by SBE domain experts. Table 1 shows the evaluation results.

Criteria <sup>7</sup>	<i>Disco</i>	<i>QB</i>	<i>SKOS</i>	Total
<i>Triples</i>	9,673,055	3,775,983,610	477,737,281	4,263,393,946
<i>Data Sets</i>	1,526	9,990	4,178	15,694
<i>CV</i>	3,545,703	45,635,846	5,540,988	54,722,537
<i>CV (SL<sub>0</sub>)</i>	2,437,922 (68.8%)	0 (0%)	2,281,740 (41.2%)	4,719,662 (8.6%)
<i>CV (SL<sub>1</sub>)</i>	473,574 (13.4%)	45,520,613 (99.75%)	3,259,248 (58.8%)	49,253,435 (90%)
<i>CV (SL<sub>2</sub>)</i>	634,207 (17.9%)	115,233 (0.25%)	0 (0%)	749,440 (1.4%)
<i>CT</i>	52 (15 37) <sup>8</sup>	20 (7 13) <sup>8</sup>	14 (4 10) <sup>8</sup>	53
<i>CT (C<sub>C</sub>)</i>	30 (57.7%)	5 (25%)	5 (35.7%)	30 (56.6%)
<i>CT (C<sub>T</sub>)</i>	22 (42.3%)	15 (75%)	9 (64.3%)	23 (43.4%)
<i>C</i>	142 (77 65)	35 (20 15)	35 (17 18)	212
<i>C (C<sub>C</sub>)</i>	72 (50.7%)	16 (45.7% 12 4)	21 (60% 13 8)	109 (51.4%)
<i>C (C<sub>T</sub>)</i>	70 (49.3%)	19 (54.3% 8 11)	14 (40% 4 10)	103 (48.6%)
<i>C (SL<sub>0</sub>)</i>	75 (52.8% 44 31)	4 (11.4% 0 4)	21 (60% 12 9)	100 (47.2%)
<i>C (SL<sub>1</sub>)</i>	9 (6.3% 8 1)	3 (8.6% 3 0)	5 (14.3% 5 0)	17 (8%)
<i>C (SL<sub>2</sub>)</i>	58 (40.8% 25 33)	28 (80% 17 11)	9 (25.7% 0 9)	95 (44.8%)

Table 1: Evaluation

We identified 142 *Disco* constraints ( $\mathcal{C}_C$  and  $\mathcal{C}_T$  constraints to the same extend) assigned to 52 distinct constraint types and implemented 77 of them to actually validate person-level data sets. For *QB*, we specified more  $\mathcal{C}_T$  (54%) than  $\mathcal{C}_C$  constraints; for *SKOS*, however, more  $\mathcal{C}_C$  constraints (60%). We instantiated more  $\mathcal{C}_C$  (58%) than  $\mathcal{C}_T$  constraint types to define *Disco* constraints; for *QB* (75%) and *SKOS* (64%), on the other side, more  $\mathcal{C}_T$  constraint types. In total, we used 53 of overall 82 distinct constraint types (57% of them are  $\mathcal{C}_C$  constraint types) to define 212 constraints (equally  $\mathcal{C}_C$  and  $\mathcal{C}_T$  constraints).

For *Disco* and *SKOS*, more than the half of the constraints are associated with the weakest severity level  $\mathcal{SL}_0$ . Within the context of *QB*, 80% of the constraints are classified as the most serious ones ( $\mathcal{SL}_2$ ). All in all, there are a little bit more  $\mathcal{SL}_0$  than  $\mathcal{SL}_2$  constraints, whereas  $\mathcal{SL}_1$  constraints are negligible. *Existential quantifications* (32.4%, *Disco*), *data model consistency* (31.4%, *QB*), and *structure* (28.6%, *SKOS*) are the constraint types the most constraints are instantiated from. By validating *QB* data sets, we got the most

<sup>6</sup> Available at: <https://github.com/boschthomas/rdf-validation/tree/master/evaluation/data-sets/data-cube>

<sup>7</sup>  $\mathcal{C}$  (constraints),  $\mathcal{C}_T$  (constraint types),  $\mathcal{C}_V$  (constraint violations)

<sup>8</sup> (implemented | not yet implemented)

constraint violations (more than 45 millions), followed by *SKOS* and *Disco* (with more than 5.5 and 3.5 millions) - consequently, almost 55 million constraint violations were raised during the evaluation which could be used to enhance the metadata quality of these data sets. Close to 70% of all *Disco* constraint violations are caused by violating  $\mathcal{SL}_0$  constraints. For *QB* (nearly 100%) and *SKOS* (almost 60%), the majority of the raised constraint violations are classified to be more serious ( $\mathcal{SL}_1$ ). 80% of all *QB* constraints are  $\mathcal{SL}_2$  constraints leading to less than 1% of all *QB* constraint violations. Altogether, exactly 90% of the constraint violations are assigned to the severity level  $\mathcal{SL}_1$ . These findings are surprising as only 8% of all defined constraints are  $\mathcal{SL}_1$  constraints. The constraints responsible for the largest numbers of constraint violations are *DISCO-C-LABELING-AND-DOCUMENTATION-06* and *DISCO-C-COMPARISON-VARIABLES-02* (both 547,916) (*Disco*), *DATA-CUBE-C-DATA-MODEL-CONSISTENCY-05* (45,514,102) (*QB*), and *SKOS-C-LANGUAGE-TAG-CARDINALITY-01* (2,508,903) (*SKOS*).

## 2.1 Legend

In this section, we describe how the tables in this paper should be read. Table 2 gives an overview over the symbols used in subsequent tables.

Symbol	Description
✓	Validation Successful (without any constraint violation)
$X$	Constraint Violations
$>X$	Poor Performance/Scaling
$\times$	Very Poor Performance/Scaling
(!)	Not Yet Implemented Constraint
( $X$ )	The validation of $X$ data sets could not be finished, due to SPARQL endpoints' technical restrictions (e.g. defined timeouts).
*	default severity level $\mathcal{SL}_0$ (informational)
**	default severity level $\mathcal{SL}_1$ (warning)
***	default severity level $\mathcal{SL}_2$ (error)

Table 2: Legend

- **Constraint Violations.** When constraints are violated,  $X$  indicates the number of raised constraint violation triples.
- **Poor Performance/Scaling.** The performance of the implementation of the underlying SPARQL CONSTRUCT query is too poor to get all resulting constraint violation triples. Therefore, a limit of  $X$  result constraint violation triples is set. It is likely that there are more than  $X$  constraint violations. Although, the result set contains not the whole set of raised constraint violation triples, the constraint can be used as an indicator if there is data

not conforming to the constraint and to resolve constraint violations step by step. As part of future work, the performance will be improved.

- **Very Poor Performance/Scaling.** The performance of the implementation of the underlying SPARQL CONSTRUCT query is too poor to get any results, even though a limit of result constraint violation triples is set. As part of future work, the performance will be improved.

### 3 Evaluation of Person-Level Metadata (*Disco*)

#### 3.1 Evaluation Results

Table 3 shows the results of the evaluation of *Disco* data sets.

Evaluation Criteria	Counts
<i>Validated Triples</i>	9,673,055
<i>Validated Data Sets</i>	1,526
<i>Constraint Violations</i>	3,545,703
<i>Constraint Violations (<math>SL_0</math>)</i>	2,437,922 (68.8%)
<i>Constraint Violations (<math>SL_1</math>)</i>	473,574 (13.4%)
<i>Constraint Violations (<math>SL_2</math>)</i>	634,207 (17.9%)
<i>Constraint (Most Constraint Violations)</i>	<i>DISCO-C-LABELING-AND-DOCUMENTATION-06</i> (547,916)
	<i>DISCO-C-COMPARISON-VARIABLES-02</i> (547,916)
<i>Constraint (Most Constraint Violations (<math>SL_0</math>))</i>	<i>DISCO-C-LABELING-AND-DOCUMENTATION-06</i> (547,916)
<i>Constraint (Most Constraint Violations (<math>SL_1</math>))</i>	<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-46</i> (468,807)
<i>Constraint (Most Constraint Violations (<math>SL_2</math>))</i>	<i>DISCO-C-COMPARISON-VARIABLES-02</i> (547,916)
<i>Constraint Types</i>	52 (15 37) <sup>9</sup>
<i>Constraint Types (<math>C_C</math>)</i>	30 (57.7%)
<i>Constraint Types (<math>C_T</math>)</i>	22 (42.3%)
<i>Constraint Types (Most Constraints)</i>	1. Existential Quantifications: 46 (32.4% 46 0) <sup>10</sup> 2. Data Model Consistency: 7 (1 6) 3. Aggregation: 7 (0 7)
<i>Constraint Type (Most Constraints (<math>SL_2</math>))</i>	Existential Quantifications: 9 (9 0)
<i>Constraints</i>	142 (77 65)
<i>Constraints (<math>C_C</math>)</i>	72 (50.7%)
<i>Constraints (<math>C_T</math>)</i>	70 (49.3%)
<i>Constraints (<math>SL_0</math>)</i>	75 (52.8% 44 31)
<i>Constraints (<math>SL_1</math>)</i>	9 (6.3% 8 1)
<i>Constraints (<math>SL_2</math>)</i>	58 (40.8% 25 33)

Table 3: Evaluation of Disco Data Sets - Evaluation Results

<sup>9</sup> legend: absolute number (absolute number implemented |absolute number not yet implemented)

<sup>10</sup> legend: absolute number (percentage value |absolute number implemented |absolute number not yet implemented)

### 3.2 Data Sets Overview

Tables 4 and 6 give an overview over the evaluated *Disco* data sets, their abbreviations, and publicly available SPARQL endpoints. Table 5 comprehends the number of triples, data sets, and instances of multiple vocabulary-specific classes.

Abbr.	Disco Data Sets
<i>Missy</i>	<i>Microdata Information System</i> <sup>11</sup>
<i>DwB</i>	<i>DwB Discovery Portal</i> <sup>12</sup>
<i>DDA-SND</i>	<i>DDI-RDF</i> <sup>13</sup>
	provided by the <i>Danish Data Archive (DDA)</i> <sup>14</sup> and Swedish National Data Service (SND) <sup>15</sup>

Table 4: Disco Data Sets Abbreviations

Data Sets	Counts									
	triples	disco:StudyGroup	disco:Study	disco:LogicalDataSet	disco:Universe	disco:Variable	disco:Question	disco:SummaryStatistics	disco:CategoryStatistics	skos:Concept
<i>Missy</i>	5,068,838	6	45	159	1,125	21,040	0	0	0	147,193
<i>DwB</i>	2,332,802	0	1,387	1,367	2,796	446,806	0	0	0	0
<i>DDA-SND</i>	2,271,415	0	1,490	0	10,188	80,070	139,237	0	0	290,963
<b>Total</b>	9,673,055			1,526						

Table 5: Disco Data Sets Overview

<sup>11</sup> <http://www.geis.org/missy/eu/missy-home>

<sup>12</sup> <http://dwb-dev.nsd.uib.no/portal>

<sup>13</sup> <http://ddi-rdf.borsna.se/>

<sup>14</sup> <http://samfund.dda.dk/dda/default-en.asp>

<sup>15</sup> <http://snd.gu.se/en>

Data Sets	SPARQL Endpoint
<i>Missy</i>	<a href="http://svko-missy:8181/openrdf-workbench/repositories/native-java-store/summary">http://svko-missy:8181/openrdf-workbench/repositories/native-java-store/summary</a>
<i>DwB</i>	<a href="http://dwb-dev.nsd.uib.no/sparql">http://dwb-dev.nsd.uib.no/sparql</a>
<i>DDA-SND</i>	<a href="http://ddi-rdf.borsna.se/endpoint/">http://ddi-rdf.borsna.se/endpoint/</a>

Table 6: Disco SPARQL Endpoints

### 3.3 Detailed Evaluation

	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<b>Existential Quantifications (1)</b>			
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-01</i> ***	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-02</i> ***	7	17	1,490
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-03</i> *	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-04</i> *	11,021	445,381	62,260
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-05</i> *	✓	✓	139,237
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-06</i> *	12	1,367	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-07</i> *	6	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-08</i> *	45	1,387	1,490
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-09</i> *	6	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-10</i> *	45	1,387	1,490

Table 7: Evaluation of Disco Data Sets - Existential Quantifications (1)

	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<b>Existential Quantifications (2)</b>			
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-11</i> <sup>*</sup>	6	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-12</i> <sup>*</sup>	6	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-13</i> <sup>*</sup>	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-14</i> <sup>*</sup>	45	1,387	1,490
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-15</i> <sup>*</sup>	45	1,387	1,490
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-16</i> <sup>*</sup>	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-17</i> <sup>*</sup>	159	1,367	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-18</i> <sup>*</sup>	159	1,367	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-19</i> <sup>*</sup>	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-20</i> <sup>*</sup>	✓	1,367	✓

Table 8: Evaluation of Disco Data Sets - Existential Quantifications (2)



	Data Sets		
	<i>Missy</i>	<i>DuB</i>	<i>DDA-SND</i>
<b>Existential Quantifications (3)</b>			
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-21</i> <sup>*</sup>	✓	1,367	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-22</i> <sup>*</sup>	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-23</i> <sup>*</sup>	6	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-24</i> <sup>*</sup>	45	1,387	1,490
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-25</i> <sup>*</sup>	45	1,387	1,490
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-26</i> <sup>*</sup>	45	1,387	1,490
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-27</i> <sup>***</sup>	✓	130	1,490
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-28</i> <sup>**</sup>	159	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-29</i> <sup>**</sup>	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-30</i> <sup>**</sup>	✓	✓	✓

Table 9: Evaluation of Disco Data Sets - Existential Quantifications (3)

	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<b>Existential Quantifications (4)</b>			
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-31</i> **	159	1,367	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-32</i> ***	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-33</i> ***	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-34</i> ***	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-35</i> ***	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-36</i> ***	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-37</i> *	18,625	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-38</i> *	✓	✓	750
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-39</i> ***	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-40</i> *	✓	✓	139,237

Table 10: Evaluation of Disco Data Sets - Existential Quantifications (4)

	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<b>Existential Quantifications (5)</b>			
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-41</i> *	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-42</i> *	✓	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-43</i> *	15,733	446,806	80,070
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-44</i> *	159	✓	✓
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-45</i> *	6,784	446,806	19,221
<i>DISCO-C-EXISTENTIAL-QUANTIFICATIONS-46</i> **	11,550	446,806	10,451

Table 11: Evaluation of Disco Data Sets - Existential Quantifications (5)

Conditional Properties	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-CONDITIONAL-PROPERTIES-01</i> <sup>***</sup>	✓	✓	80,070
<i>DISCO-C-CONDITIONAL-PROPERTIES-02</i> <sup>**</sup>	12	✓	✓
<i>DISCO-C-CONDITIONAL-PROPERTIES-03</i> <sup>**</sup>	90	✓	2,980
<i>DISCO-C-CONDITIONAL-PROPERTIES-04</i> <sup>***</sup>	6	✓	✓
<i>DISCO-C-CONDITIONAL-PROPERTIES-05</i> <sup>***</sup>	45	1,387	1,490
<i>DISCO-C-CONDITIONAL-PROPERTIES-06</i> <sup>***</sup>	✓	✓	✓

Table 12: Evaluation of Disco Data Sets - Conditional Properties

Provenance	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-PROVENANCE-01</i> <sup>*</sup>	6	✓	✓
<i>DISCO-C-PROVENANCE-02</i> <sup>*</sup>	45	1,387	1,490
<i>DISCO-C-PROVENANCE-03</i> <sup>*</sup>	159	1,367	✓
<i>DISCO-C-PROVENANCE-04</i> <sup>*</sup>	✓	1,367	✓

Table 13: Evaluation of Disco Data Sets - Provenance

Labeling and Documentation	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-LABELING-AND-DOCUMENTATION-01</i> <sup>*</sup>	6	✓	✓
<i>DISCO-C-LABELING-AND-DOCUMENTATION-02</i> <sup>*</sup>	45	1,387	1,490
<i>DISCO-C-LABELING-AND-DOCUMENTATION-03</i> <sup>*</sup>	159	1,367	✓
<i>DISCO-C-LABELING-AND-DOCUMENTATION-04</i> <sup>*</sup>	✓	1,367	✓
<i>DISCO-C-LABELING-AND-DOCUMENTATION-05</i> <sup>*</sup>	✓	✓	✓
<i>DISCO-C-LABELING-AND-DOCUMENTATION-06</i> <sup>*</sup>	21,040	446,806	80,070

Table 14: Evaluation of Disco Data Sets - Labeling and Documentation

Data Model Consistency	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-DATA-MODEL-CONSISTENCY-01 (!)</i> <sup>***</sup>			
<i>DISCO-C-DATA-MODEL-CONSISTENCY-02 (!)</i> <sup>***</sup>			
<i>DISCO-C-DATA-MODEL-CONSISTENCY-03 (!)</i> <sup>***</sup>			
<i>DISCO-C-DATA-MODEL-CONSISTENCY-04 (!)</i> <sup>***</sup>			
<i>DISCO-C-DATA-MODEL-CONSISTENCY-05</i> <sup>***</sup>	✓	✓	✓
<i>DISCO-C-DATA-MODEL-CONSISTENCY-06 (!)</i> <sup>***</sup>			
<i>DISCO-C-DATA-MODEL-CONSISTENCY-07 (!)</i> <sup>***</sup>			

Table 15: Evaluation of Disco Data Sets - Data Model Consistency

Comparison	Data Sets		
	<i>Missy</i>	<i>DuB</i>	<i>DDA-SND</i>
<i>DISCO-C-COMPARISON-VARIABLES-01 (!)**</i>			
<i>DISCO-C-COMPARISON-VARIABLES-02***</i>	21,040	446,806	80,070
<i>DISCO-C-COMPARISON-VARIABLES-03 (!)***</i>			
<i>DISCO-C-COMPARISON-VARIABLES-04*</i>	18,625	✓	✓
<i>DISCO-C-COMPARISON-VARIABLES-05***</i>	159	✓	✓

Table 16: Evaluation of Disco Data Sets - Comparison

Mathematical Operations	Data Sets		
	<i>Missy</i>	<i>DuB</i>	<i>DDA-SND</i>
<i>DISCO-C-MATHEMATICAL-OPERATIONS-01 (!)***</i>			
<i>DISCO-C-MATHEMATICAL-OPERATIONS-02 (!)***</i>			
<i>DISCO-C-MATHEMATICAL-OPERATIONS-03 (!)***</i>			
<i>DISCO-C-MATHEMATICAL-OPERATIONS-04 (!)***</i>			
<i>DISCO-C-MATHEMATICAL-OPERATIONS-05 (!)***</i>			

Table 17: Evaluation of Disco Data Sets - Mathematical Operations

Language Tags	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-LANGUAGE-TAG-MATCHING-01 (!)*</i>			
<i>DISCO-C-LANGUAGE-TAG-CARDINALITY-01 (!)*</i>			
<i>DISCO-C-LANGUAGE-TAG-CARDINALITY-02 (!)*</i>			
<i>DISCO-C-LANGUAGE-TAG-CARDINALITY-03 (!)*</i>			

Table 18: Evaluation of Disco Data Sets - Language Tags

Aggregation	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-AGGREGATION-01 (!)*</i>			
<i>DISCO-C-AGGREGATION-02 (!)*</i>			
<i>DISCO-C-AGGREGATION-03 (!)*</i>			
<i>DISCO-C-AGGREGATION-04 (!)*</i>			
<i>DISCO-C-AGGREGATION-05 (!)*</i>			
<i>DISCO-C-AGGREGATION-06 (!)*</i>			
<i>DISCO-C-AGGREGATION-07 (!)*</i>			

Table 19: Evaluation of Disco Data Sets - Aggregation

Disco Constraints	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-ALLOWED-VALUES-01</i> ***	✓	✓	✓
<i>DISCO-C-LITERAL-RANGES-01</i> ***	✓	✓	✓
<i>DISCO-C-INVERSE-FUNCTIONAL-PROPERTIES-01</i> ***	✓	✓	✓
<i>DISCO-C-INVERSE-FUNCTIONAL-PROPERTIES-02</i> ***	✓	✓	✓
<i>DISCO-C-CLASS-SPECIFIC-PROPERTY-RANGE-01</i> ***	✓	✓	✓
<i>DISCO-C-MEMBERSHIP-IN-CONTROLLED-VOCABULARIES-01</i> ***	✓	✓	✗
<i>DISCO-C-LITERAL-VALUE-COMPARISON-01</i> ***	✓	1,299	✓
<i>DISCO-C-CONTEXT-SPECIFIC-VALID-PROPERTIES-01</i> *	21,038	✓	✓
<i>DISCO-C-DATA-PROPERTY-FACETS-01</i> **	✓	✓	✓
<i>DISCO-C-DATA-PROPERTY-FACETS-02</i> **	✓	✓	✓

Table 20: Evaluation of Disco Data Sets - Disco Constraints (1)

Disco Constraints	Data Sets		
	<i>Missy</i>	<i>DuB</i>	<i>DDA-SND</i>
<i>DISCO-C-VALUE-IS-VALID-FOR-DATATYPE-01</i> ***	30	6,932	✓
<i>DISCO-C-VALUE-IS-VALID-FOR-DATATYPE-02</i> ***	✓	✓	✓
<i>DISCO-C-SUBSUMPTION-01 (!)</i> ***			
<i>DISCO-C-CLASS-EQUIVALENCE-01 (!)</i> *			
<i>DISCO-C-SUB-PROPERTIES-01 (!)</i> ***			
<i>DISCO-C-PROPERTY-DOMAIN-01 (!)</i> ***			
<i>DISCO-C-PROPERTY-RANGES-01 (!)</i> ***			
<i>DISCO-C-INVERSE-OBJECT-PROPERTIES-01 (!)</i> ***			
<i>DISCO-C-INVERSE-OBJECT-PROPERTIES-02 (!)</i> ***			
<i>DISCO-C-INVERSE-OBJECT-PROPERTIES-03 (!)</i> ***			
<i>DISCO-C-DISJOINT-PROPERTIES-01 (!)</i> ***			

Table 21: Evaluation of Disco Data Sets - Disco Constraints (2)



Disco Constraints	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-ASYMMETRIC-OBJECT-PROPERTIES-01 (!)</i> ***			
<i>DISCO-C-IRREFLEXIVE-OBJECT-PROPERTIES-01 (!)</i> ***			
<i>DISCO-C-CLASS-SPECIFIC-IRREFLEXIVE-OBJECT-PROPERTIES-01 (!)</i> ***			
<i>DISCO-C-CLASS-SPECIFIC-IRREFLEXIVE-OBJECT-PROPERTIES-02 (!)</i> ***			
<i>DISCO-C-DISJOINT-CLASSES-01 (!)</i> ***			
<i>DISCO-C-EQUIVALENT-PROPERTIES-01 (!)</i> *			
<i>DISCO-C-LITERAL-PATTERN-MATCHING-01 (!)</i> *			
<i>DISCO-C-DISJUNCTION-01 (!)</i> ***			
<i>DISCO-C-UNIVERSAL-QUANTIFICATIONS-01 (!)</i> ***			
<i>DISCO-C-MINIMUM-QUALIFIED-CARDINALITY-RESTRICTIONS-01 (!)</i> ***			

Table 22: Evaluation of Disco Data Sets - Disco Constraints (3)

Disco Constraints	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-MAXIMUM-QUALIFIED-CARDINALITY-RESTRICTIONS-01 (!)</i> ***			
<i>DISCO-C-EXACT-QUALIFIED-CARDINALITY-RESTRICTIONS-01 (!)</i> ***			
<i>DISCO-C-CONTEXT-SPECIFIC-EXCLUSIVE-OR-OF-PROPERTY-GROUPS-01 (!)</i> *			
<i>DISCO-C-IRI-PATTERN-MATCHING-01 (!)</i> *			
<i>DISCO-C-ORDERING-01 (!)</i> *			
<i>DISCO-C-ORDERING-02 (!)</i> *			
<i>DISCO-C-ORDERING-03 (!)</i> *			
<i>DISCO-C-STRING-OPERATIONS-01 (!)</i> *			
<i>DISCO-C-CONTEXT-SPECIFIC-VALID-CLASSES-01 (!)</i> *			
<i>DISCO-C-CONTEXT-SPECIFIC-VALID-PROPERTIES-01 (!)</i> *			

Table 23: Evaluation of Disco Data Sets - Disco Constraints (4)

Disco Constraints	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-DEFAULT-VALUES-01 (!)*</i>			
<i>DISCO-C-WHITESPACE-HANDLING-01 (!)*</i>			
<i>DISCO-C-HTML-HANDLING-01 (!)*</i>			
<i>DISCO-C-HTML-HANDLING-02 (!)*</i>			
<i>DISCO-C-RECOMMENDED-PROPERTIES-01 (!)*</i>			
<i>DISCO-C-HANDLE-RDF-COLLECTIONS-01 (!)*</i>			
<i>DISCO-C-HANDLE-RDF-COLLECTIONS-02 (!)*</i>			
<i>DISCO-C-USE-SUB-SUPER-RELATIONS-IN-VALIDATION-01 (!)*</i>			
<i>DISCO-C-USE-SUB-SUPER-RELATIONS-IN-VALIDATION-02 (!)*</i>			
<i>DISCO-C-STRUCTURE-01 (!)***</i>			

Table 24: Evaluation of Disco Data Sets - Disco Constraints (5)

Disco Constraints	Data Sets		
	<i>Missy</i>	<i>DwB</i>	<i>DDA-SND</i>
<i>DISCO-C-VOCABULARY-01 (!)***</i>			
<i>DISCO-C-HTTP-URI-SCHEME-VIOLATION (!)***</i>			

Table 25: Evaluation of Disco Data Sets - Disco Constraints (6)

## 4 Evaluation of Aggregated Metadata (*Data Cube*)

### 4.1 Evaluation Results

Table 26 shows the results of the evaluation of *Data Cube* data sets.

Evaluation Criteria	Counts
<i>Validated Triples</i>	3,775,983,610
<i>Validated Data Sets</i>	9,990
<i>Constraint Violations</i>	45,635,846
<i>Constraint Violations (<math>SL_0</math>)</i>	0 (0%)
<i>Constraint Violations (<math>SL_1</math>)</i>	45,520,613 (99.75%)
<i>Constraint Violations (<math>SL_2</math>)</i>	115,233 (0.25%)
<i>Constraint (Most Constraint Violations)</i>	<i>DATA-MODEL-CONSISTENCY-05</i> (45,514,102)
<i>Constraint (Most Constraint Violations (<math>SL_0</math>))</i>	-
<i>Constraint (Most Constraint Violations (<math>SL_1</math>))</i>	<i>DATA-MODEL-CONSISTENCY-05</i> (45,514,102)
<i>Constraint (Most Constraint Violations (<math>SL_2</math>))</i>	<i>MINIMUM-QUALIFIED-CARDINALITY-RESTRICTIONS-02</i> (1,556)
<i>Constraint Types</i>	20 (7 13)
<i>Constraint Types (<math>C_C</math>)</i>	5 (25%)
<i>Constraint Types (<math>C_T</math>)</i>	15 (75%)
<i>Constraint Types (Most Constraints)</i>	1. Data Model Consistency: 11 (31.4% 10 1) 2. Existential Quantifications: 4 (11.4% 4 0) Data Model Consistency: 8 (22.9% 7 1)
<i>Constraint Type (Most Constraints (<math>SL_2</math>))</i>	
<i>Constraints</i>	35 (20 15)
<i>Constraints (<math>C_C</math>)</i>	16 (45.7% 12 4)
<i>Constraints (<math>C_T</math>)</i>	19 (54.3% 8 11)
<i>Constraints (<math>SL_0</math>)</i>	4 (11.4% 0 4)
<i>Constraints (<math>SL_1</math>)</i>	3 (8.6% 3 0)
<i>Constraints (<math>SL_2</math>)</i>	28 (80% 17 11)

Table 26: Evaluation of Data Cube Data Sets - Evaluation Results

## 4.2 Data Sets Overview

There are websites giving an overview over available *Data Cube* data sets<sup>16</sup>. Tables 27 and 29 give an overview over the evaluated *Data Cube* data sets, their abbreviations, and publicly available SPARQL endpoints. Table 28 comprehends the number of triples, data sets, and instances of multiple vocabulary-specific classes.

<sup>16</sup> <http://270a.info/>; [http://datahub.io/de/dataset?tags=format-qb](http://datahub.io/de/dataset?tags=format-qb;); <http://ontologycentral.com/>

Abbr.	Data Cube Data Sets
<i>ECB</i>	<i>European Central Bank</i> <sup>17</sup>
<i>UIS</i>	<i>UNESCO Institute for Statistics</i> <sup>18</sup>
<i>IMF</i>	<i>International Monetary Fund</i> <sup>19</sup>
<i>BFS</i>	<i>Bundesamt für Statistik - Swiss Federal Statistics</i> <sup>20</sup>
<i>FAO</i>	<i>Food and Agriculture Organization of the United Nations</i> <sup>21</sup>
<i>WB</i>	<i>World Bank</i> <sup>22</sup>
<i>FRB</i>	<i>Federal Reserve Board</i> <sup>23</sup>
<i>TI</i>	<i>Transparency International</i> <sup>24</sup>
<i>OECD</i>	<i>Organisation for Economic Co-operation and Development</i> <sup>25</sup>
<i>BIS</i>	<i>Bank for International Settlements</i> <sup>26</sup>
<i>ABS</i>	<i>Australian Bureau of Statistics</i> <sup>27</sup>
<i>IEEE-VIS</i>	<i>IEEE VIS Source Data</i>
<i>ACORN-SAT</i>	<i>Australian Climate Observations Reference Network - Surface Air Temperature Dataset</i>
<i>HDP</i>	<i>HealthData.gov Platform (HDP) on the Semantic Web</i>
<i>Eurostat</i>	<i>The Eurostat Linked Data</i> (SPARQL endpoint unavailable)
<i>Asturias</i>	<i>Nomenclator Asturias</i> (SPARQL endpoint unavailable!)
<i>ISTAT</i>	<i>ISTAT Immigration (LinkedOpenData.it)</i> (SPARQL endpoint unavailable)
<i>ICANE</i>	<i>Statistical Office of Cantabria (Instituto Cántabro de Estadística, ICANE)</i> (SPARQL endpoint unavailable)
<i>EE-2009</i>	<i>European Election Results 2009</i> (SPARQL endpoint unavailable)
<i>EU-B</i>	<i>Standard Eurobarometer</i> (SPARQL endpoint unavailable)
<i>ECB-S</i>	<i>European Central Bank Statistics (PublicData.eu)</i> (SPARQL endpoint unavailable)
<i>CPV-2008</i>	<i>Common Procurement Vocabulary (CPV) 2008</i> (SPARQL endpoint unavailable)
<i>CPV-2003</i>	<i>Common Procurement Vocabulary (CPV) 2003</i> (SPARQL endpoint unavailable)

Table 27: Data Cube Data Sets Abbreviations

<sup>17</sup> <http://www.ecb.europa.eu/home/html/index.en.html>

<sup>18</sup> <http://www.uis.unesco.org/Pages/default.aspx>

<sup>19</sup> <http://www.imf.org/external/index.htm>

<sup>20</sup> <http://www.bfs.admin.ch/>

<sup>21</sup> <http://www.fao.org/home/en/>

<sup>22</sup> <http://www.worldbank.org/>

<sup>23</sup> <http://www.federalreserve.gov/>

<sup>24</sup> <http://www.transparency.org/>

<sup>25</sup> <http://www.oecd.org/>

<sup>26</sup> <http://www.bis.org/>

<sup>27</sup> <http://abs.gov.au/>

Data Sets	Counts				
	triples	qb:DataSet	qb:DataStructureDefinition	qb:Observation	qb:Slice
<i>ECB</i>	468,899,474	55	46	>11,000,000	428,698
<i>UIS</i>	10,400,534	5	5	1,437,651	0
<i>IMF</i>	35,688,446	4	8	3,603,719	0
<i>BFS</i>	1,533,743	0	0	8	0
<i>FAO</i>	53,000,000	10	10	>7,100,000	0
<i>WB</i>	174,006,552	9,466	59	>17,000,000	0
<i>FRB</i>	185,266,900	49	98	>9,500,000	0
<i>TI</i>	52,233	6	6	3,928	0
<i>OECD</i>	304,995,160	136	140	>12,000,000	0
<i>BIS</i>	54,197,482	6	12	3,606,466	47,914
<i>ABS</i>	2,357,400,000	253	257	>11,000,000	0
<i>IEEE-VIS</i>	19,935,340	0	0	1,350	0
<i>ACORN-SAT</i>	98,381,319	0	4	0	0
<i>HDP</i>	12,226,427	0	0	0	0
<b>Total</b>	3,775,983,610	9,990			

Table 28: Data Cube Data Sets Overview

Data Sets	SPARQL Endpoints
<i>ECB</i>	http://ecb.270a.info/sparql
<i>UIS</i>	http://uis.270a.info/sparql
<i>IMF</i>	http://imf.270a.info/sparql
<i>BFS</i>	http://bfs.270a.info/sparql
<i>FAO</i>	http://fao.270a.info/sparql
<i>WB</i>	http://worldbank.270a.info/sparql
<i>FRB</i>	http://frb.270a.info/sparql
<i>TI</i>	http://transparency.270a.info/sparql
<i>OECD</i>	http://oecd.270a.info/sparql
<i>BIS</i>	http://bis.270a.info/sparql
<i>ABS</i>	http://abs.270a.info/sparql
<i>ACORN-SAT</i>	http://lab.environment.data.gov.au/sparql
<i>HDP</i>	http://healthdata.tw.rpi.edu/sparql

Table 29: Data Cube SPARQL Endpoints

### 4.3 Detailed Evaluation

Data Model Consistency	Data Sets						
	<i>ECB</i>	<i>UIS</i>	<i>IMF</i>	<i>BFS</i>	<i>FAO</i>	<i>WB</i>	<i>FRB</i>
<i>DATA-MODEL-CONSISTENCY-01</i> **	✓ (2)	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-02</i> ***	✓ (2)	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-03</i> ***	✓ (2)	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-04</i> ***	✓ (6)	✓	✓	✓	✓	✓	14,372
<i>DATA-MODEL-CONSISTENCY-05</i> **	1,198,352 (50)	✗	✗	✓	✗	✓	16,175,814 (42)
<i>DATA-MODEL-CONSISTENCY-06</i> ***	✓ (2)	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-07</i> ***	✓ (9)	✓	99,091	✓	✓	✓	✓ (1)
<i>DATA-MODEL-CONSISTENCY-08</i> ***	✓ (2)	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-09</i> ***	✓ (2)	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-10</i> *** (!)	-	-	-	-	-	-	-
<i>DATA-MODEL-CONSISTENCY-11</i> **	6,511 (10)	✓	✓	✓	✓	✓	✓

Table 30: Evaluation of Data Cube Data Sets - Data Model Consistency (1)

Data Model Consistency	Data Sets						
	<i>TI</i>	<i>OECD</i>	<i>BIS</i>	<i>ABS</i>	<i>IEEE-VIS</i>	<i>ACORN-SAT</i>	<i>HDP</i>
<i>DATA-MODEL-CONSISTENCY-01</i> **	✓	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-02</i> ***	✓	✓	✓	✓	✓	8	✓
<i>DATA-MODEL-CONSISTENCY-03</i> ***	✓	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-04</i> ***	✓	✓	✓	✓ (6)	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-05</i> **	✓	21,142,838 (116)	✗	6,997,098 (246)	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-06</i> ***	✓	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-07</i> ***	✓	✓	✓	✓ (8)	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-08</i> ***	✓	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-09</i> ***	✓	✓	✓	✓	✓	✓	✓
<i>DATA-MODEL-CONSISTENCY-10</i> *** (!)	-	-	-	-	-	-	-
<i>DATA-MODEL-CONSISTENCY-11</i> **	✓	✓	✓	✓	✓	✓	✓

Table 31: Evaluation of Data Cube Data Sets - Data Model Consistency (2)

Existential Quantifications	Data Sets													
	<i>ECB</i>	<i>UIS</i>	<i>IMF</i>	<i>BFS</i>	<i>FAO</i>	<i>WB</i>	<i>FRB</i>	<i>TI</i>	<i>OECD</i>	<i>BIS</i>	<i>ABS</i>	<i>IEEE-VIS</i>	<i>ACORN-SAT</i>	<i>HDP</i>
<i>EXISTENTIAL-QUANTIFICATIONS-01</i> ***	9	✓	11	7	8	77	8	9	7	8	7	✓	✓	✓
<i>EXISTENTIAL-QUANTIFICATIONS-02</i> ***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>EXISTENTIAL-QUANTIFICATIONS-03</i> ***	✓	✓	✓	✓	✓	59	✓	6	✓	✓	✓	✓	4	✓
<i>EXISTENTIAL-QUANTIFICATIONS-04</i> ***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 32: Evaluation of Data Cube Data Sets - Existential Quantifications



Cardinality Restrictions	Data Sets									
	<i>ECB</i>	<i>UIS</i>	<i>IMF</i>	<i>BFS</i>	<i>FAO</i>	<i>WB</i>	<i>FRB</i>	<i>TI</i>	<i>OECD</i>	<i>BIS</i>
<i>MINIMUM-QUALIFIED-CARDINALITY-RESTRICTIONS-01 (!)</i> ***	-	-	-	-	-	-	-	-	-	-
<i>MINIMUM-QUALIFIED-CARDINALITY-RESTRICTIONS-02</i> ***	✗	118	8	8	30	✓	30	✓	✗	12
<i>MAXIMUM-QUALIFIED-CARDINALITY-RESTRICTIONS-01</i> ***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>EXACT-UNQUALIFIED-CARDINALITY-RESTRICTIONS-01</i> ***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>EXACT-QUALIFIED-CARDINALITY-RESTRICTIONS-02</i> ***	✓	✓	✓	✓	✓	1	✓	✓	✓	✓

Table 33: Evaluation of Data Cube Data Sets - Cardinality Restrictions (1)

Cardinality Restrictions	Data Sets			
	<i>ABS</i>	<i>IEEE-VIS</i>	<i>ACORN-SAT</i>	<i>HDP</i>
<i>MINIMUM-QUALIFIED-CARDINALITY-RESTRICTIONS-01 (!)</i> ***	-	-	-	-
<i>MINIMUM-QUALIFIED-CARDINALITY-RESTRICTIONS-02</i> ***	✗	1,350	✓	✓
<i>MAXIMUM-QUALIFIED-CARDINALITY-RESTRICTIONS-01</i> ***	✓ (2)	✓	✓	✓
<i>EXACT-UNQUALIFIED-CARDINALITY-RESTRICTIONS-01</i> ***	✓	✓	✓	✓
<i>EXACT-QUALIFIED-CARDINALITY-RESTRICTIONS-02</i> ***	✓	✓	✓	✓

Table 34: Evaluation of Data Cube Data Sets - Cardinality Restrictions (2)

	Data Sets													
Structure	<i>ECB</i>	<i>UIS</i>	<i>IMF</i>	<i>BFS</i>	<i>FAO</i>	<i>WB</i>	<i>FRB</i>	<i>TI</i>	<i>OECD</i>	<i>BIS</i>	<i>ABS</i>	<i>IEEE-VIS</i>	<i>ACORN-SAT</i>	<i>HDP</i>
<i>STRUCTURE-01</i> ***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>STRUCTURE-02</i> ***	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 35: Evaluation of Data Cube Data Sets - Structure

	Data Sets													
Constraints	<i>ECB</i>	<i>UIS</i>	<i>IMF</i>	<i>BFS</i>	<i>FAO</i>	<i>WB</i>	<i>FRB</i>	<i>TI</i>	<i>OECD</i>	<i>BIS</i>	<i>ABS</i>	<i>IEEE-VIS</i>	<i>ACORN-SAT</i>	<i>HDP</i>
<i>PROPERTY-DOMAIN-01 (!)</i> ***														
<i>PROPERTY-RANGES-01 (!)</i> ***														
<i>DISJOINT-PROPERTIES-01 (!)</i> ***														
<i>DISJOINT-CLASSES-01 (!)</i> ***														
<i>EQUIVALENT-PROPERTIES-01 (!)</i> *														
<i>UNIVERSAL-QUANTIFICATIONS-01 (!)</i> ***														
<i>MEMBERSHIP-IN-CONTROLLED-VOCABULARIES-01 (!)</i> ***														
<i>CONTEXT-SPECIFIC-VALID-CLASSES-01 (!)</i> *														
<i>CONTEXT-SPECIFIC-VALID-PROPERTIES-01 (!)</i> *														
<i>RECOMMENDED-PROPERTIES-01 (!)</i> *														
<i>VALUE-IS-VALID-FOR-DATATYPE-01 (!)</i> ***														
<i>VOCABULARY-01 (!)</i> ***														

Table 36: Evaluation of Data Cube Data Sets - Constraints (1)

Constraints	Data Sets
	ECB UIS IMF BFS FAO WB FRB TI OECD BIS ABS IEEE-VIS ACORN-SAT HDP
<i>HTTP-URI-SCHEME-VIOLATION (!)***</i>	

Table 37: Evaluation of Data Cube Data Sets - Constraints (2)

## 5 Evaluation of Thesauri (*SKOS*)

### 5.1 Evaluation Results

Table 38 shows the results of the evaluation of *thesauri*.

Evaluation Criteria	Counts
<i>Validated Triples</i>	477,737,281
<i>Validated Data Sets</i>	4,178
<i>Constraint Violations</i>	5,540,988
<i>Constraint Violations (<math>SL_0</math>)</i>	2,281,740 (41.2%)
<i>Constraint Violations (<math>SL_1</math>)</i>	3,259,248 (58.8%)
<i>Constraint Violations (<math>SL_2</math>)</i>	0 (0%)
<i>Constraint (Most Constraint Violations)</i>	<i>LANGUAGE-TAG-CARDINALITY-01</i> (2,508,903)
<i>Constraint (Most Constraint Violations (<math>SL_0</math>))</i>	<i>LABELING-AND-DOCUMENTATION-06</i> (1,022,362)
<i>Constraint (Most Constraint Violations (<math>SL_1</math>))</i>	<i>LANGUAGE-TAG-CARDINALITY-01</i> (2,508,903)
<i>Constraint (Most Constraint Violations (<math>SL_2</math>))</i>	-
<i>Constraint Types</i>	14 (4 10)
<i>Constraint Types (<math>C_C</math>)</i>	5 (35.7%)
<i>Constraint Types (<math>C_T</math>)</i>	9 (64.3%)
<i>Constraint Types (Most Constraints)</i>	1. Structure: 10 (28.6% 8 2) 2. Labeling and Documentation: 6 (17.1% 5 1) 3. Language Tag Cardinality: 4 (11.4% 4 0)
<i>Constraint Type (Most Constraints (<math>SL_2</math>))</i>	Structure: 1 (0 1)
<i>Constraints</i>	35 (17 18)
<i>Constraints (<math>C_C</math>)</i>	21 (60% 13 8)
<i>Constraints (<math>C_T</math>)</i>	14 (40% 4 10)
<i>Constraints (<math>SL_0</math>)</i>	21 (60% 12 9)
<i>Constraints (<math>SL_1</math>)</i>	5 (14.3% 5 0)
<i>Constraints (<math>SL_2</math>)</i>	9 (25.7% 0 9)

Table 38: Evaluation of Thesauri Data Sets - Evaluation Results

## 5.2 Data Sets Overview

There is a website giving an overview over available *SKOS* data sets<sup>28</sup> and another one giving an overview over available thesauri<sup>29</sup>. Tables 39 and 41 give an overview over the evaluated thesauri, their abbreviations, and publicly available SPARQL endpoints. Table 40 comprehends the number of triples, data sets, and instances of multiple vocabulary-specific classes.

Abbr.	Thesauri
<i>TheSoz</i>	<i>Thesaurus for the Social Sciences</i> <sup>30</sup>
<i>STW</i>	<i>Thesaurus for Economics</i> <sup>31</sup>
<i>AGROVOC</i>	<i>AGROVOC Multilingual agricultural thesaurus</i> <sup>32</sup>
<i>UNESCO</i>	<i>UNESCO Thesaurus</i> <sup>33</sup>
<i>TGN</i>	<i>The Getty Thesaurus of Geographic Names</i> <sup>34</sup>
<i>EARTH</i>	<i>Environmental Applications Reference Thesaurus</i> <sup>35</sup>
<i>ODT</i>	<i>Open Data Thesaurus</i> <sup>36</sup>
<i>SLD</i>	<i>Spanish Linguistic Datasets</i> <sup>37</sup>
<i>SSWT</i>	<i>Social Semantic Web Thesaurus</i> <sup>38</sup>
<i>GBA-GU</i>	<i>Thesaurus of the Geological Survey of Austria (GBA) - Geology Unit</i> <sup>39</sup>
<i>GBA-GTS</i>	<i>Thesaurus of the Geological Survey of Austria (GBA) - Geologic Time Scale</i> <sup>40</sup>
<i>GBA-L</i>	<i>Thesaurus of the Geological Survey of Austria (GBA) - Lithology</i> <sup>41</sup>
<i>GBA-LU</i>	<i>Thesaurus of the Geological Survey of Austria (GBA) - Lithotectonic Unit</i> <sup>42</sup>
<i>GEMET</i>	<i>GEneral Multilingual Environmental Thesaurus</i> <sup>43</sup>
<i>EuroVoc</i>	<i>EuroVoc</i> <sup>44</sup>
<i>CECCT</i>	<i>Clean Energy and Climate Change Thesaurus</i> <sup>45</sup>

Table 39: Thesauri Abbreviations

<sup>28</sup> <http://datahub.io/de/dataset?tags=format-skos>

<sup>29</sup> <http://datahub.io/de/dataset?tags=thesaurus>

- 
- <sup>30</sup> <http://www.ecb.europa.eu/home/html/index.en.html>
- <sup>31</sup> <http://zbw.eu/stw/versions/latest/about>
- <sup>32</sup> <http://202.45.139.84:10035/catalogs/fao/repositories/agrovoc>
- <sup>33</sup> <http://skos.um.es/sparql/>
- <sup>34</sup> <http://vocab.getty.edu/sparql>
- <sup>35</sup> <http://linkeddata.ge.imati.cnr.it/resource/EARTh/>
- <sup>36</sup> <http://vocabulary.semantic-web.at/PoolParty/wiki/OpenData>
- <sup>37</sup> <http://linguistic.linkeddata.es>
- <sup>38</sup> <http://vocabulary.semantic-web.at/PoolParty/wiki/semweb>
- <sup>39</sup> <http://resource.geolba.ac.at/>
- <sup>40</sup> <http://resource.geolba.ac.at/>
- <sup>41</sup> <http://resource.geolba.ac.at/>
- <sup>42</sup> <http://resource.geolba.ac.at/>
- <sup>43</sup> <http://www.eionet.europa.eu/gemet/>
- <sup>44</sup> <http://open-data.europa.eu/de/data/dataset/eurovoc>
- <sup>45</sup> <http://data.reegle.info/thesaurus/guide>

Thesauri	Counts						
	triples	skos:ConceptScheme	sko:Concept	skos:broader	skos:narrower	skos:hasTopConcept	skos:inScheme
<i>TheSoz</i>	439,153	1	8,426	13,705	13,706	0	48,529
<i>STW</i>	221,668	1	13,468	13,732	13732	7	13,180
<i>AGROVOC</i>	6,080,477	1	32,310	33,507	33,507	25	32,310
<i>UNESCO</i>	288,346	9	26,714	20,028	20,028	607	32,009
<i>TGN</i>	16,112,321	8	2,898,775	0	0	0	1,453,767
<i>EARTh</i>	9,287,364	11	295,375	288,208	93,827	479	295,376
<i>ODT</i>	3,290	6	108	93	93	30	0
<i>SLD</i>	7,629,211	0	31,195	0	0	0	0
<i>SSWT</i>	64,698	9	2,127	2,300	2,301	38	0
<i>GBA-GU</i>	25,718	3	878	1,005	1,005	14	0
<i>GBA-GTS</i>	7,875	3	213	208	208	5	0
<i>GBA-L</i>	9,317	1	249	249	249	4	0
<i>GBA-LU</i>	9,504	3	364	359	359	7	0
<i>GEMET</i>	372,889,229	3,680	414,659	62,193	21,685	30,806	409,290
<i>EuroVoc</i>	64,477,774	439	79,557	6,922	0	532	14,428
<i>CECCT</i>	191,336	3	3,419	3,761	3,762	28	0
<b>Total</b>	477,737,281	4,178					

Table 40: Thesauri Overview

Thesauri	SPARQL Endpoints
<i>TheSoz</i>	<a href="http://lod.gesis.org/thesoz/sparql">http://lod.gesis.org/thesoz/sparql</a>
<i>STW</i>	<a href="http://zbw.eu/beta/sparql/stw/query">http://zbw.eu/beta/sparql/stw/query</a>
<i>AGROVOC</i>	<a href="http://202.45.139.84:10035/catalogs/fao/repositories/agrovoc">http://202.45.139.84:10035/catalogs/fao/repositories/agrovoc</a>
<i>UNESCO</i>	<a href="http://skos.um.es/sparql/">http://skos.um.es/sparql/</a>
<i>TGN</i>	<a href="http://vocab.getty.edu/">http://vocab.getty.edu/</a>
<i>EARTH</i>	<a href="http://linkeddata.ge.imati.cnr.it:8890/sparql">http://linkeddata.ge.imati.cnr.it:8890/sparql</a>
<i>ODT</i>	<a href="http://vocabulary.semantic-web.at/PoolParty/sparql/OpenData">http://vocabulary.semantic-web.at/PoolParty/sparql/OpenData</a>
<i>SLD</i>	<a href="http://linguistic.linkeddata.es/sparql">http://linguistic.linkeddata.es/sparql</a>
<i>SSWT</i>	<a href="http://vocabulary.semantic-web.at/PoolParty/sparql/semweb">http://vocabulary.semantic-web.at/PoolParty/sparql/semweb</a>
<i>GBA-GU</i>	<a href="http://resource.geolba.ac.at/PoolParty/sparql/GeologicUnit">http://resource.geolba.ac.at/PoolParty/sparql/GeologicUnit</a>
<i>GBA-GTS</i>	<a href="http://resource.geolba.ac.at/PoolParty/sparql/GeologicTimeScale">http://resource.geolba.ac.at/PoolParty/sparql/GeologicTimeScale</a>
<i>GBA-L</i>	<a href="http://resource.geolba.ac.at/PoolParty/sparql/lithology">http://resource.geolba.ac.at/PoolParty/sparql/lithology</a>
<i>GBA-LU</i>	<a href="http://resource.geolba.ac.at/PoolParty/sparql/tectonicunit">http://resource.geolba.ac.at/PoolParty/sparql/tectonicunit</a>
<i>GEMET</i>	<a href="http://semantic.eea.europa.eu/sparql">http://semantic.eea.europa.eu/sparql</a>
<i>EuroVoc</i>	<a href="http://open-data.europa.eu/de/linked-data">http://open-data.europa.eu/de/linked-data</a>
<i>CECCT</i>	<a href="http://poolparty.reegle.info/PoolParty/sparql/glossary">http://poolparty.reegle.info/PoolParty/sparql/glossary</a>

Table 41: Thesauri SPARQL Endpoints

### 5.3 Detailed Evaluation

	Data Sets
Data Model Consistency	<i>TheSoz</i> <i>STW</i> <i>AGROVOC</i> <i>TGN</i> <i>UNESCO</i> <i>ODT</i> <i>SSWT</i> <i>GBA-GU</i> <i>GBA-GTS</i> <i>GBA-L</i> <i>GBA-LU</i> <i>CECCT</i>
<i>DATA-MODEL-CONSISTENCY-01 (!)*</i>	
<i>DATA-MODEL-CONSISTENCY-02 (!)*</i>	
<i>DATA-MODEL-CONSISTENCY-03 (!)*</i>	

Table 42: Thesauri Evaluation - Data Model Consistency (1)

Data Model Consistency	Data Sets
	<i>Earth</i> <i>GEMET</i> <i>Euro Voc</i> <i>SLD</i>
<i>DATA-MODEL-CONSISTENCY-01 (!)*</i>	
<i>DATA-MODEL-CONSISTENCY-02 (!)*</i>	
<i>DATA-MODEL-CONSISTENCY-03 (!)*</i>	

Table 43: Thesauri Evaluation - Data Model Consistency (2)

Labeling and Documentation	Data Sets											
	<i>TheSoz</i>	<i>STW</i>	<i>AGROVOC</i>	<i>TGN</i>	<i>UNESCO</i>	<i>ODT</i>	<i>SSWT</i>	<i>GBA-GU</i>	<i>GBA-GTS</i>	<i>GBA-L</i>	<i>GBA-LU</i>	<i>CECCT</i>
<i>LABELING-AND-DOCUMENTATION-01*</i>	8,426	11,508	19,829	1,110	✗	36	1,475	5	2	✓	107	486
<i>LABELING-AND-DOCUMENTATION-02*</i>	>1	✗	>100	287	✗	✓	✓	✓	✓	✓	✓	✓
<i>LABELING-AND-DOCUMENTATION-03*</i>	✓	✓	1	14,114	✗	✓	✓	1	✓	✓	1	✓
<i>LABELING-AND-DOCUMENTATION-04 (!)*</i>												
<i>LABELING-AND-DOCUMENTATION-05*</i>	✓	✓	4	✓	1	2	2	1	✓	✓	✓	7
<i>LABELING-AND-DOCUMENTATION-06*</i>	975,340	✓	✓	2	✓	✓	✓	✓	✓	✓	✓	✓

Table 44: Thesauri Evaluation - Labeling and Documentation (1)



	Data Sets			
	<i>EARTh</i>	<i>GEMET</i>	<i>Euro Voc</i>	<i>SLD</i>
<b>Labeling and Documentation</b>				
<i>LABELING-AND-DOCUMENTATION-01</i> <sup>*</sup>	264,687	✗	54,911	31,195
<i>LABELING-AND-DOCUMENTATION-02</i> <sup>*</sup>	✗	✗	✗	✓
<i>LABELING-AND-DOCUMENTATION-03</i> <sup>*</sup>	2	✗	55,556	31,195
<i>LABELING-AND-DOCUMENTATION-04 (!)</i> <sup>*</sup>				
<i>LABELING-AND-DOCUMENTATION-05</i> <sup>*</sup>	39	✗	✗	978
<i>LABELING-AND-DOCUMENTATION-06</i> <sup>*</sup>	302	46,718	✓	✓

Table 45: Thesauri Evaluation - Labeling and Documentation (2)

Structure	Data Sets											
	<i>TheSoz</i>	<i>STW</i>	<i>AGROVOC</i>	<i>TGN</i>	<i>UNESCO</i>	<i>ODT</i>	<i>SSWT</i>	<i>GBA-GU</i>	<i>GBA-GTS</i>	<i>GBA-L</i>	<i>GBA-LU</i>	<i>CECCT</i>
<i>STRUCTURE-01</i> **	1	1,074	✓	✓	1	5	1	✓	✓	✓	✓	✓
<i>STRUCTURE-02 (!)</i> *												
<i>STRUCTURE-03</i> **	✓	✓	✓	✓	84	✓	✓	✓	✓	✓	✓	✓
<i>STRUCTURE-04</i> *	2,906	8,046	726	✓	3,840	12	124	84	256	68	22	2,422
<i>STRUCTURE-05</i> *	✓	✓	✓	✓	✗	90	5,150	✓	✓	✓	✓	9,864
<i>STRUCTURE-06</i> *	1,457	37	✓	✓	✗	✓	4	1	1	64	✓	136
<i>STRUCTURE-07</i> **	40	5,370	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓
<i>STRUCTURE-08 (!)</i> ***												
<i>STRUCTURE-09</i> *	7,897	19,844	99	✓	552	2	16	26	✓	✓	✓	82
<i>STRUCTURE-10</i> **	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 46: Thesauri Evaluation - Structure (1)

Structure	Data Sets			
	<i>EARTh</i>	<i>GEMET</i>	<i>Euro Voc</i>	<i>SLD</i>
<i>STRUCTURE-01</i> **	18,240	✗	55,757	31,195
<i>STRUCTURE-02 (!)</i> *				
<i>STRUCTURE-03</i> **	39	4,244	✓	✓
<i>STRUCTURE-04</i> *	11,286	74	✓	✓
<i>STRUCTURE-05</i> *	✓	✗	✓	✓
<i>STRUCTURE-06</i> *	239,346	✗	13,876	✓
<i>STRUCTURE-07</i> **	110,015	✗	366,155	155,975
<i>STRUCTURE-08 (!)</i> ***				
<i>STRUCTURE-09</i> *	107,195	32	✓	✓
<i>STRUCTURE-10</i> **	27	2,122	✓	✓

Table 47: Thesauri Evaluation - Structure (2)

Language Tag Cardinality	Data Sets											
	<i>TheSoz</i>	<i>STW</i>	<i>AGROVOC</i>	<i>TGN</i>	<i>UNESCO</i>	<i>ODT</i>	<i>SSWT</i>	<i>GBA-GU</i>	<i>GBA-GTS</i>	<i>GBA-L</i>	<i>GBA-LU</i>	<i>CECCT</i>
<i>LANGUAGE-TAG-CARDINALITY-01</i> **	9,435	13,468	98,894	✓	541	10,147	5,117	2,061	1,742	2,272	15,550	
<i>LANGUAGE-TAG-CARDINALITY-02</i> *	8,222	36,936	✗	✓	265	3,627	2,212	635	631	1,253	9,607	
<i>LANGUAGE-TAG-CARDINALITY-03</i> *	8,222	✓	135	✓	✓	✓	✓	✓	✓	✓	✓	✓
<i>LANGUAGE-TAG-CARDINALITY-04</i> *	✓	476	✗	50	✓	✓	✓	✓	✓	✓	✓	✓

Table 48: Thesauri Evaluation - Language Tag Cardinality (1)

Language Tag Cardinality	Data Sets			
	<i>EARTh</i>	<i>GEMET</i>	<i>Euro Voc</i>	<i>SLD</i>
<i>LANGUAGE-TAG-CARDINALITY-01</i> **	✗	2,318,895	✗	30,781
<i>LANGUAGE-TAG-CARDINALITY-02</i> *	✗	✗	✗	✗
<i>LANGUAGE-TAG-CARDINALITY-03</i> *	224,206	✗	✗	31,195
<i>LANGUAGE-TAG-CARDINALITY-04</i> *	✗	✗	✓	✓

Table 49: Thesauri Evaluation - Language Tag Cardinality (2)

Constraints	Data Sets									
	<i>TheSoz</i>	<i>STW</i>	<i>AGROVOC</i>	<i>TGN</i>	<i>UNESCO</i>	<i>ODT</i>	<i>SSWT</i>	<i>GBA-GU</i>	<i>GBA-GTS</i>	<i>GBA-L</i> <i>GBA-LU</i> <i>CECCT</i>
<i>PROPERTY-DOMAIN-01 (!)</i> ***										
<i>PROPERTY-RANGES-01 (!)</i> ***										
<i>DISJOINT-PROPERTIES-01 (!)</i> ***										
<i>DISJOINT-PROPERTIES-02 (!)</i> ***										
<i>DISJOINT-CLASSES-01 (!)</i> ***										
<i>EQUIVALENT-PROPERTIES-01 (!)</i> *										
<i>UNIVERSAL-QUANTIFICATIONS-01 (!)</i> ***										
<i>CONTEXT-SPECIFIC-VALID-CLASSES-01 (!)</i> *										
<i>CONTEXT-SPECIFIC-VALID-PROPERTIES-01 (!)</i> *										
<i>RECOMMENDED-PROPERTIES-01 (!)</i> *										
<i>VOCABULARY-01 (!)</i> ***										
<i>HTTP-URI-SCHEME-VIOLATION (!)</i> ***										

Table 50: Thesauri Evaluation - Constraints (1)

Constraints	Data Sets			
	<i>Earth</i>	<i>GEMET</i>	<i>Euro Voc</i>	<i>SLD</i>
<i>PROPERTY-DOMAIN-01 (!)</i> ***				
<i>PROPERTY-RANGES-01 (!)</i> ***				
<i>DISJOINT-PROPERTIES-01 (!)</i> ***				
<i>DISJOINT-PROPERTIES-02 (!)</i> ***				
<i>DISJOINT-CLASSES-01 (!)</i> ***				
<i>EQUIVALENT-PROPERTIES-01 (!)</i> *				
<i>UNIVERSAL-QUANTIFICATIONS-01 (!)</i> ***				
<i>CONTEXT-SPECIFIC-VALID-CLASSES-01 (!)</i> *				
<i>CONTEXT-SPECIFIC-VALID-PROPERTIES-01 (!)</i> *				
<i>RECOMMENDED-PROPERTIES-01 (!)</i> *				
<i>VOCABULARY-01 (!)</i> ***				
<i>HTTP-URI-SCHEME-VIOLATION (!)</i> ***				

Table 51: Thesauri Evaluation - Constraints (2)

## 6 Evaluation of Rectangular Data (*PHDD*)

### 6.1 Evaluation Results

### 6.2 Data Sets Overview

### 6.3 Detailed Evaluation

## 7 Evaluation of Statistical Classifications (*XKOS*)

### 7.1 Evaluation Results

### 7.2 Data Sets Overview

Abbr.	Statistical Classifications
<i>NAF</i>	<i>Nomenclature d'activités française</i> <sup>46</sup>
<i>PCS</i>	<i>Nomenclature des Professions et Catégories Socioprofessionnelles</i> <sup>47</sup>
<i>CJ</i>	<i>Nomenclature des catégories juridiques</i> <sup>48</sup>
<i>ISIC</i>	
<i>ISCO</i>	

Table 52: Statistical Classifications Abbreviations

*Nomenclature d'activités française (NAF)* is the French refinement of the *NACE* classification expressed in *XKOS* having explanatory notes. *Nomenclature des Professions et Catégories Socioprofessionnelles (PCS)* and *Nomenclature des catégories juridiques (CJ)* are French classifications expressed in *XKOS*. The statistical classification *ISIC* has explanatory notes too.

### 7.3 Detailed Evaluation

## References

1. Thomas Bosch and Kai Eckert. Towards description set profiles for rdf using sparql as intermediate language. *Proceedings of the DCMI International Conference on Dublin Core and Metadata Applications (DC 2014)*, 2014.
2. Thomas Bosch, Andreas Nolle, Erman Acar, and Kai Eckert. Rdf validation requirements - evaluation and logical underpinning. 2015.
3. Thomas Bosch, Benjamin Zapilko, Joachim Wackerow, and Kai Eckert. Rdf constraints to validate metadata on person-level, aggregated, thesauri, and statistical classifications data sets and rectangular data. 2015.

<sup>46</sup> <http://rdf.insee.fr/codes/index.html>

<sup>47</sup> <http://rdf.insee.fr/codes/index.html>

<sup>48</sup> <http://rdf.insee.fr/codes/index.html>