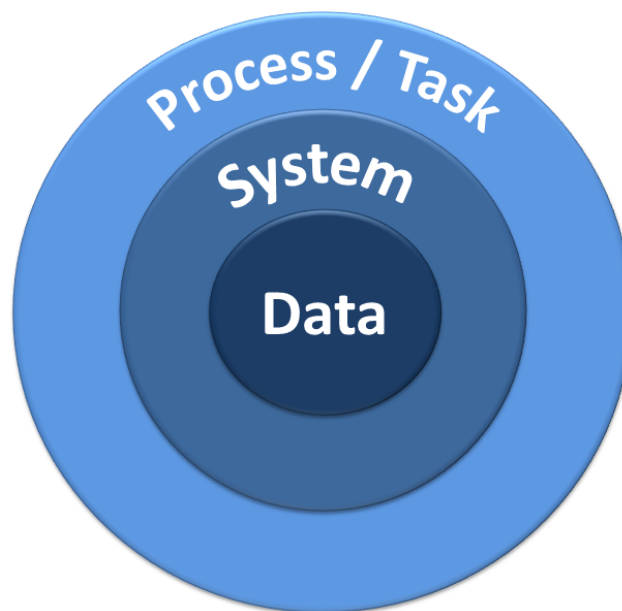
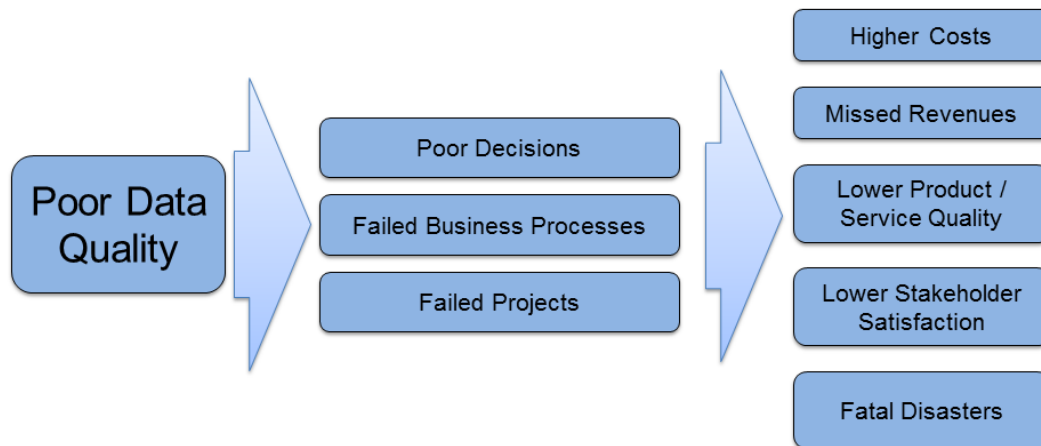


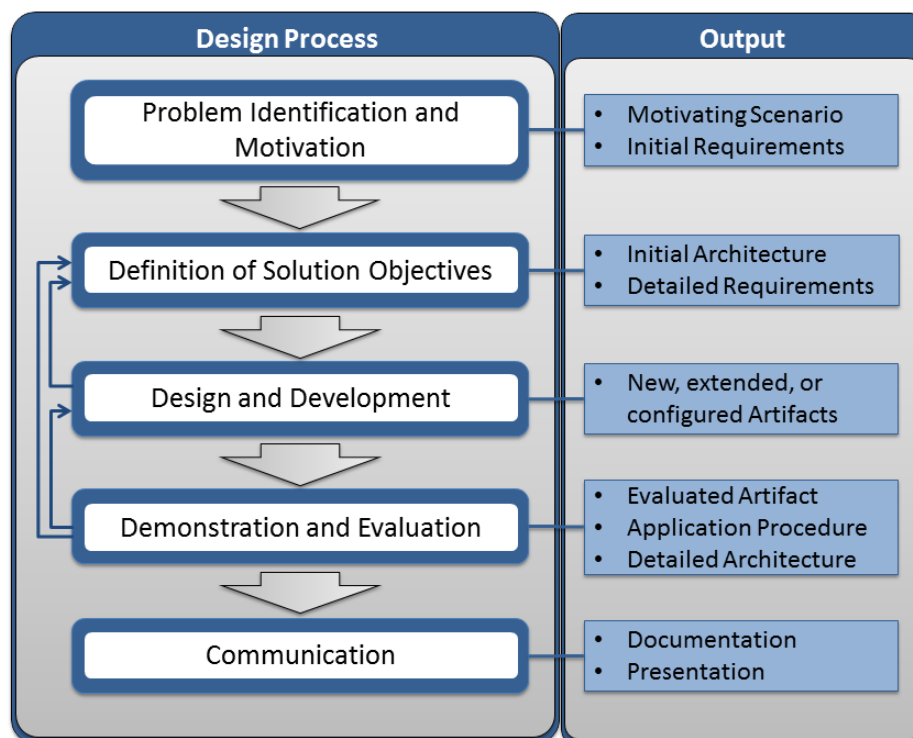
**Figure 1:** Extended DIKW hierarchy (cf. Bodendorf, 2006, p. 1; Rowley, 2007, p. 164)



**Figure 2:** Simplified illustration of the relationship between business processes and data

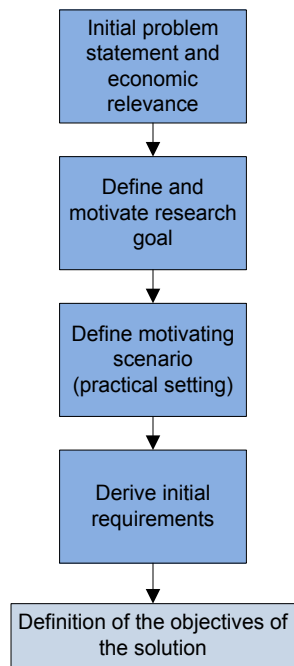


**Figure 3:** Impact of poor data quality on organizational success<sup>1</sup>

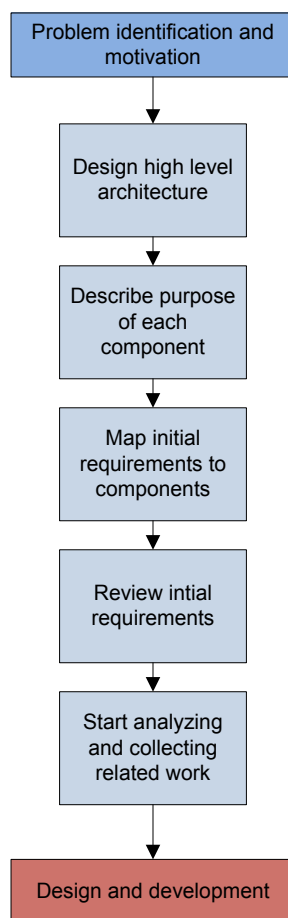


**Figure 4:** Design methodology as applied in this thesis (cf. Peffers et al., 2008)

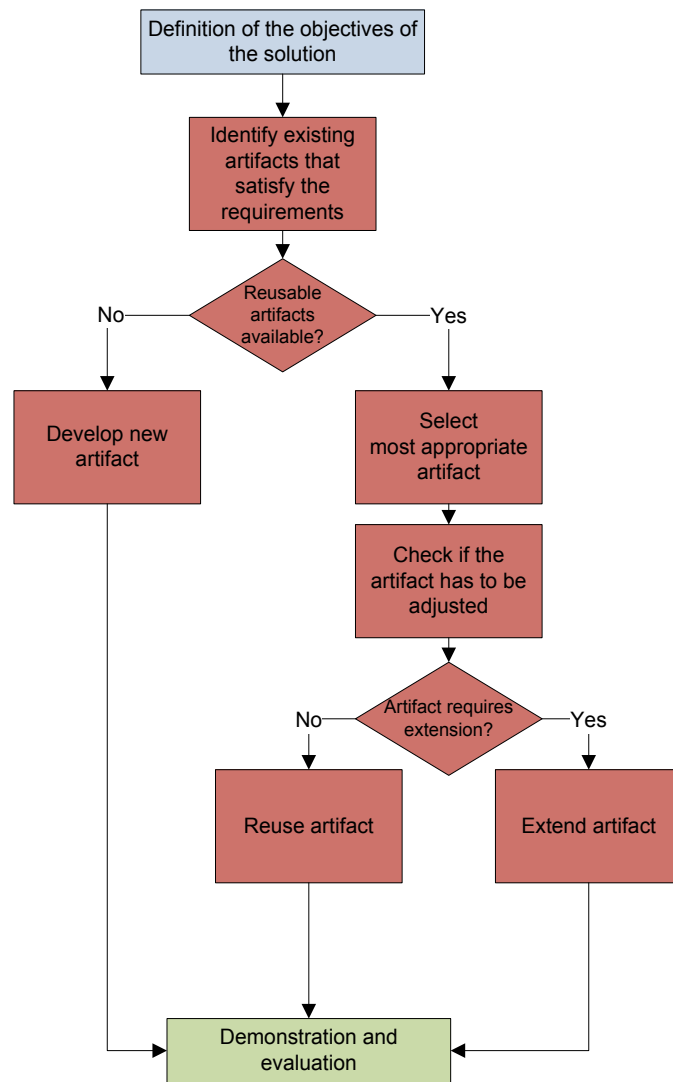
<sup>1</sup> Summary based on (Eckerson, 2002; English, 1999, pp. 3-13,209-212; Redman, 1998)



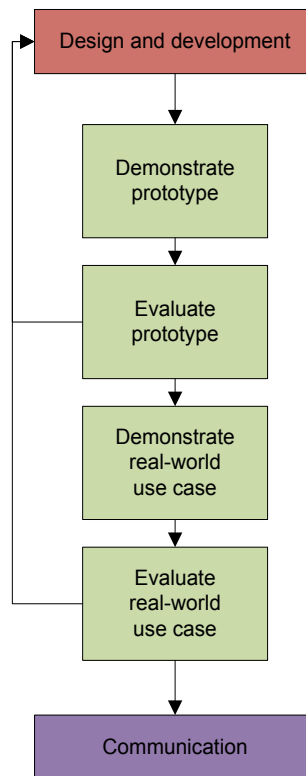
**Figure 5:** Problem identification and motivation process as applied in this thesis



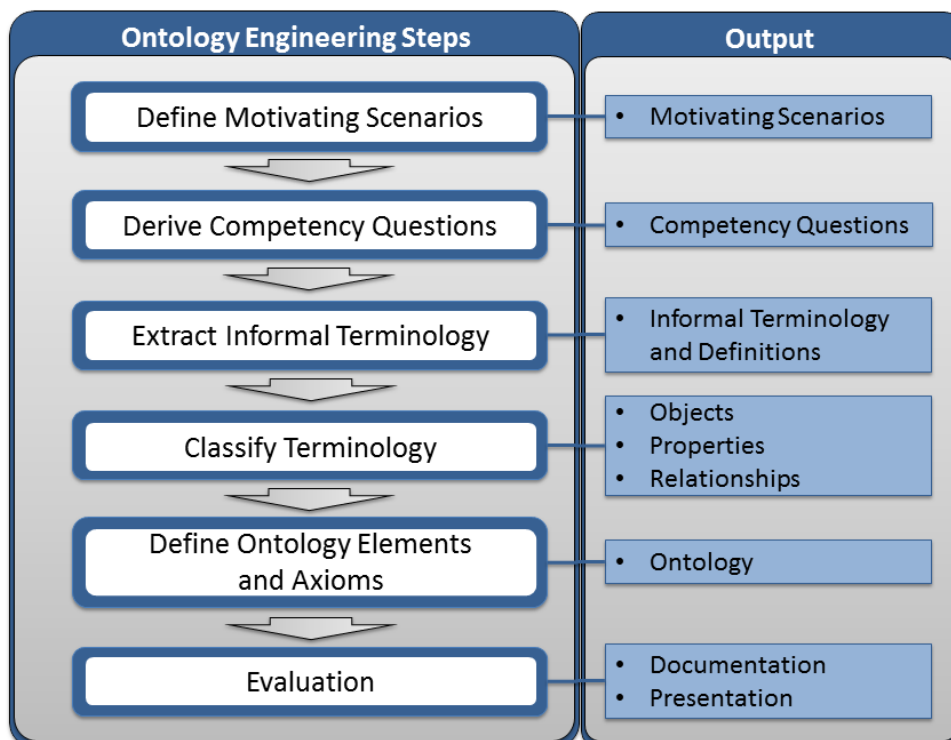
**Figure 6:** Process for the definition of solution objectives as applied in this thesis



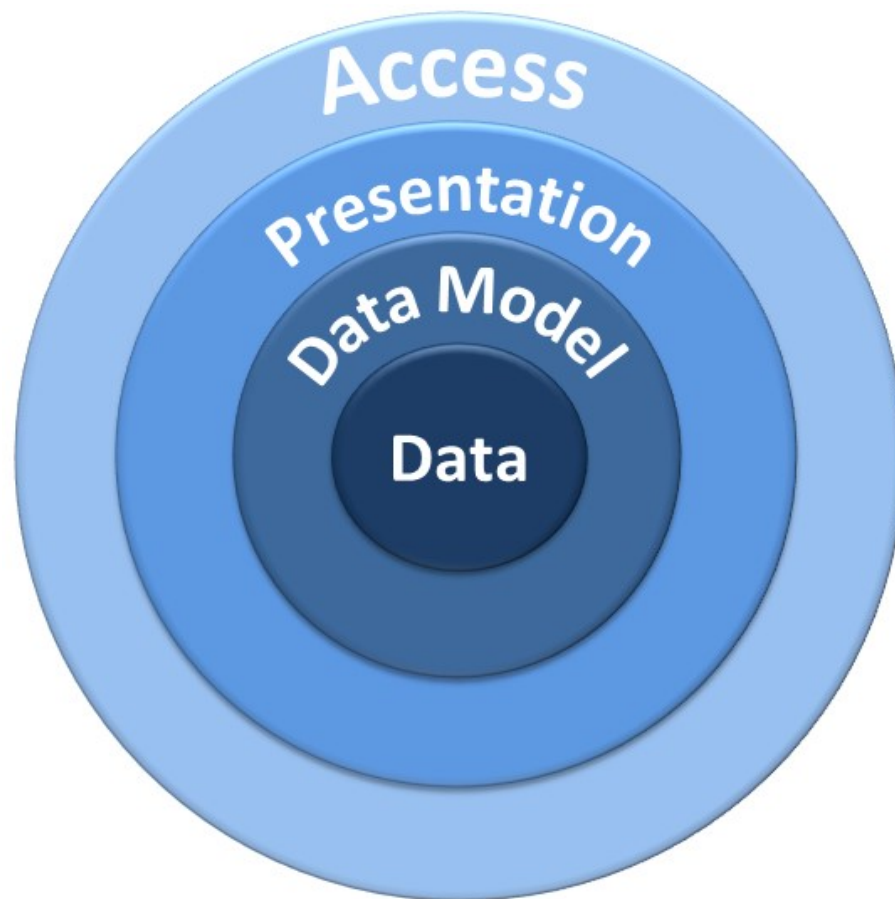
**Figure 7:** Design and development process as applied in this thesis



**Figure 8:** Demonstration and evaluation process as applied in this thesis



**Figure 9:** Ontology engineering methodology as applied in this thesis



**Figure 10:** Layers in the perception of data consumers (inspired by Redman, 2001, p. 72)

Schema	Attribute				
Tuple / Instance	SSN	Last Name	First Name	Street	House No.
	1234567890	Doe	John	Acme St.	2040
	2345678901	Smith	Peter	Wall St.	1010
	...	...	...	...	...

**Figure 11:** Terminology applied to tabular data

<ul style="list-style-type: none"> <li>- Invalid Characters</li> <li>- Character alignment violation</li> <li>- Missing values</li> <li>- False values</li> <li>- Meaningless values</li> <li>- Outdated values</li> </ul>	<ul style="list-style-type: none"> <li>- Embedded values</li> <li>- Out-of-range values</li> <li>- Imprecise values</li> <li>- Unique value violation</li> <li>- Cardinality constraint violation</li> </ul>	<p><b>Not Applicable</b></p>
Single-Source	Integration-specific	

**Figure 12:** Attribute value problems

<ul style="list-style-type: none"> <li>- Functional depdency violations</li> <li>- Referential integrity violations</li> <li>- Incorrect / outdated reference</li> <li>- Conditional missing values</li> <li>- Misfielded values</li> </ul>	<ul style="list-style-type: none"> <li>- Heterogeneity of syntaxes</li> <li>- Heterogeneity of units of measuerment</li> <li>- Data precision conflicts</li> <li>- Default value conflicts</li> </ul>
Single-Source	Integration-specific

**Figure 13:** Multi-attribute quality problems

<ul style="list-style-type: none"> <li>- Inconsistent duplicates</li> <li>- Approximate duplicates</li> <li>- Contradictory relationships</li> </ul>	<ul style="list-style-type: none"> <li>- Heterogeneity in cardinality</li> <li>- Heterogeneity in time reference</li> <li>- Source specific identifiers</li> </ul>
Single-Source	Integration-Specific

**Figure 14:** Instance-related quality problems

<ul style="list-style-type: none"> <li>- Outdated conceptual elements</li> <li>- Missing conceptual elements</li> <li>- Misuse of conceptual elements</li> <li>- Overlapping concepts / Role conflicts</li> </ul>	<ul style="list-style-type: none"> <li>- Heterogeneity of integrity constraints</li> <li>- Schema isomorphism conflict</li> <li>- Schematic descrepancy</li> </ul>
Single-Source	Integration-Specific

**Figure 15:** Quality problems of data models

Employee_No	Revenue	Revenue_Date
1	10,000	2011/03/23
2	15,000	2011/05/06
...	...	...

Employee_No	2011/03/23	2011/05/06
1	10,000	
2		15,000
...	...	...

**Figure 16:** Example of a data value attribute conflict

Emp1_Revenue	Emp1_Revenue_Date	Emp2_Revenue	Emp2_Revenue_Date
10,000	2011/03/23	15,000	2011/05/06
...	...	...	...

Employee No. 1

Revenue	Revenue Date
10,000	2011/03/23
...	...

Employee No. 2

Revenue	Revenue Date
15,000	2011/05/06
...	...

**Figure 17:** Example of an attribute entity conflict

Employee_No	Revenue	Revenue_Date
1	10,000	2011/03/23
2	15,000	2011/05/06
...	...	...

Employee No. 1

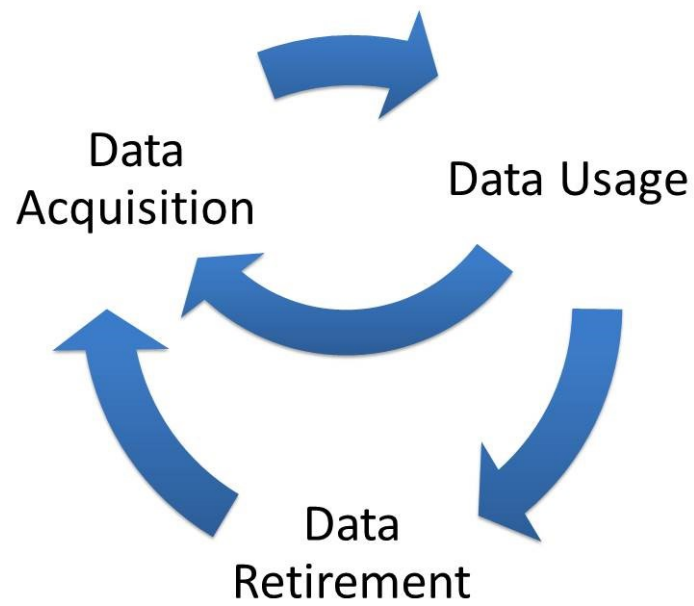
Revenue	Revenue Date
10,000	2011/03/23
...	...

Employee No. 2

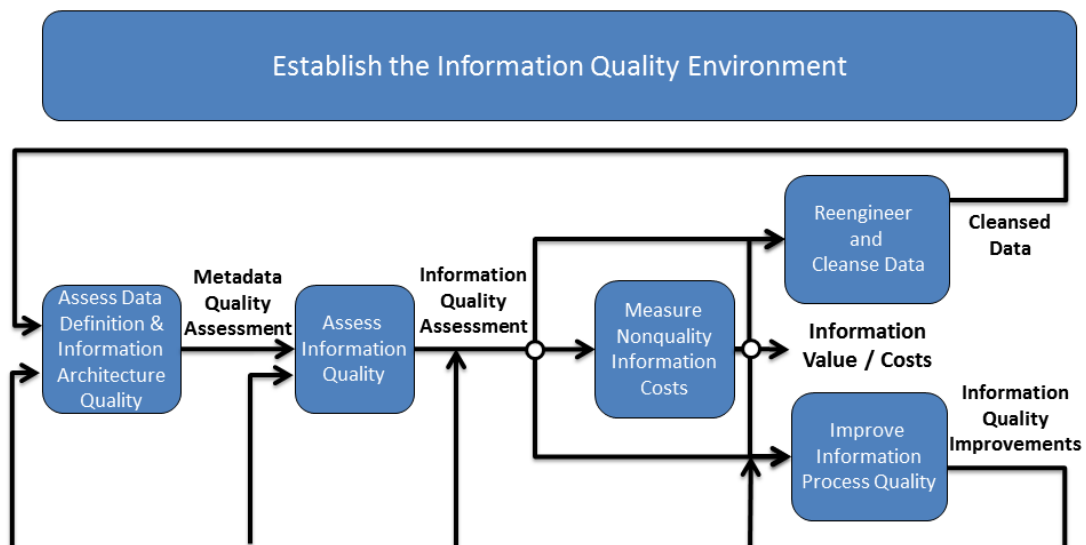
Revenue	Revenue Date
15,000	2011/05/06
...	...

**Figure 18:** Example of a data value entity conflict

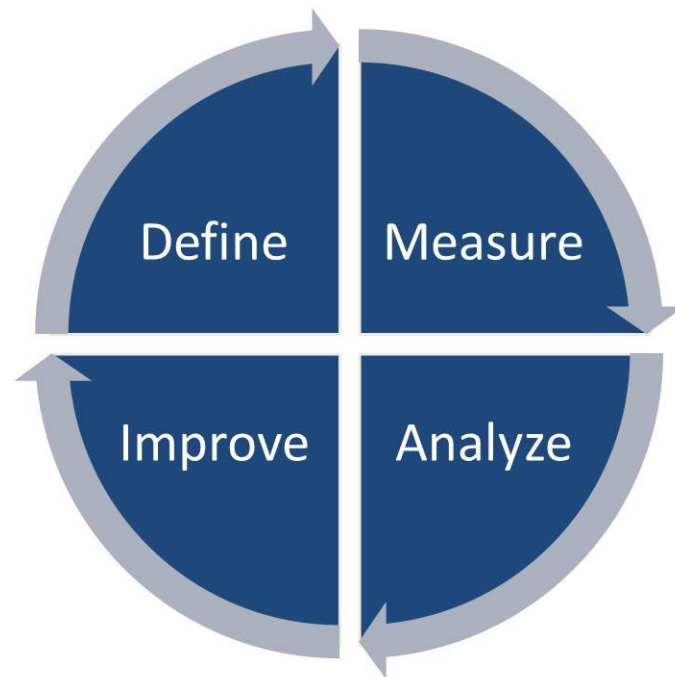




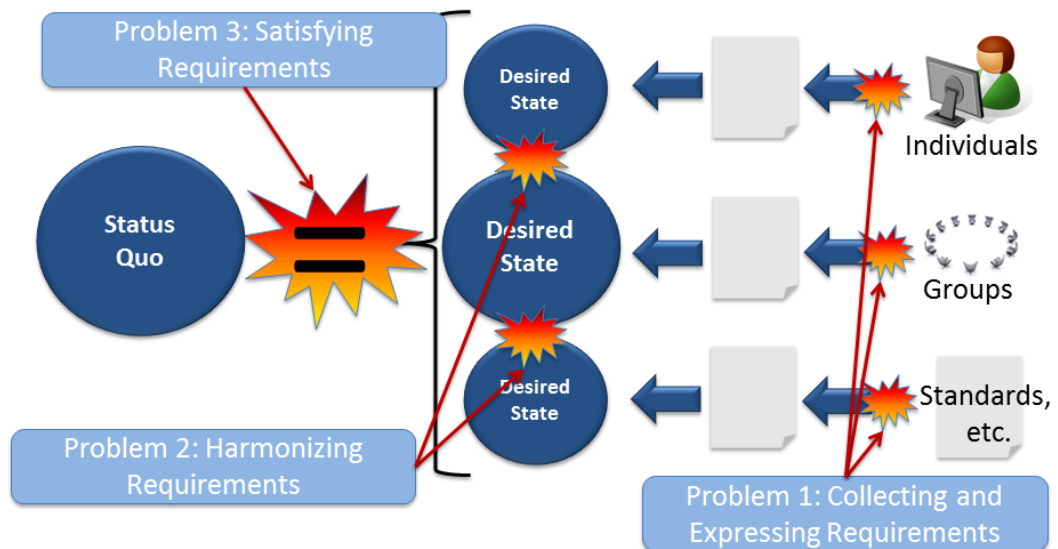
**Figure 19:** Data lifecycle (cf. Redman, 1996, p. 217)



**Figure 20:** Total Information Quality Management (cf. English, 1999, p. 70)

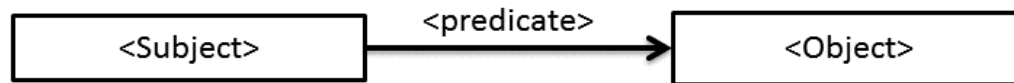


**Figure 21:** Fundamental stages of the TDQM methodology by (Wang, 1998)



**Figure 22:** Challenges of requirement satisfaction

Syntax:



Example:

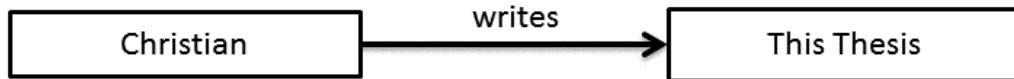


Figure 23: Syntax of RDF triples (cf. Klyne & Carroll, 2004)

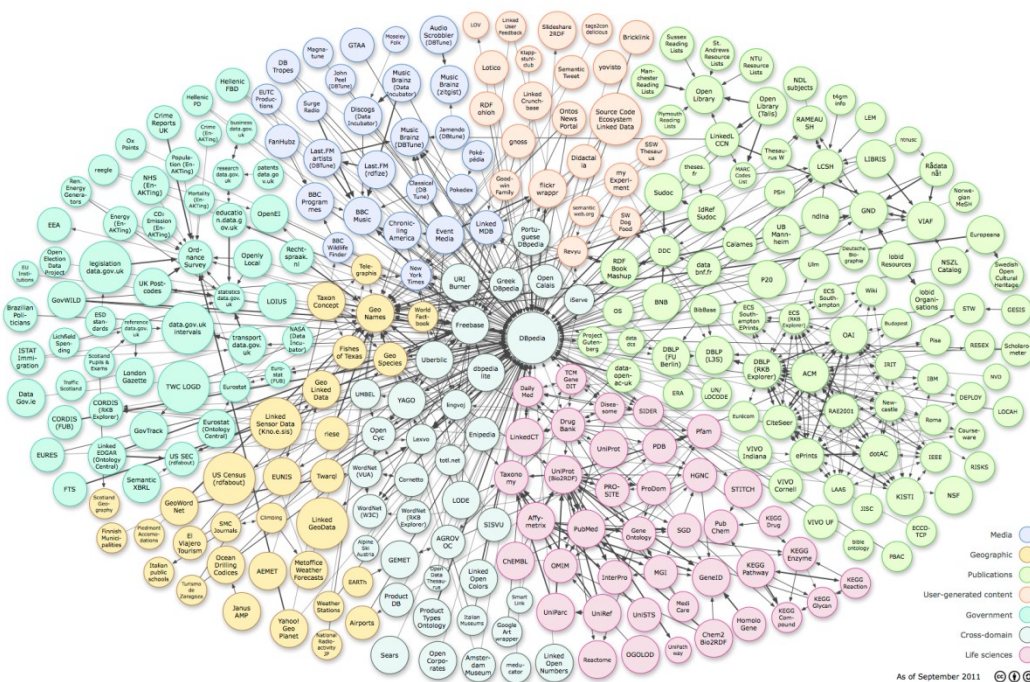
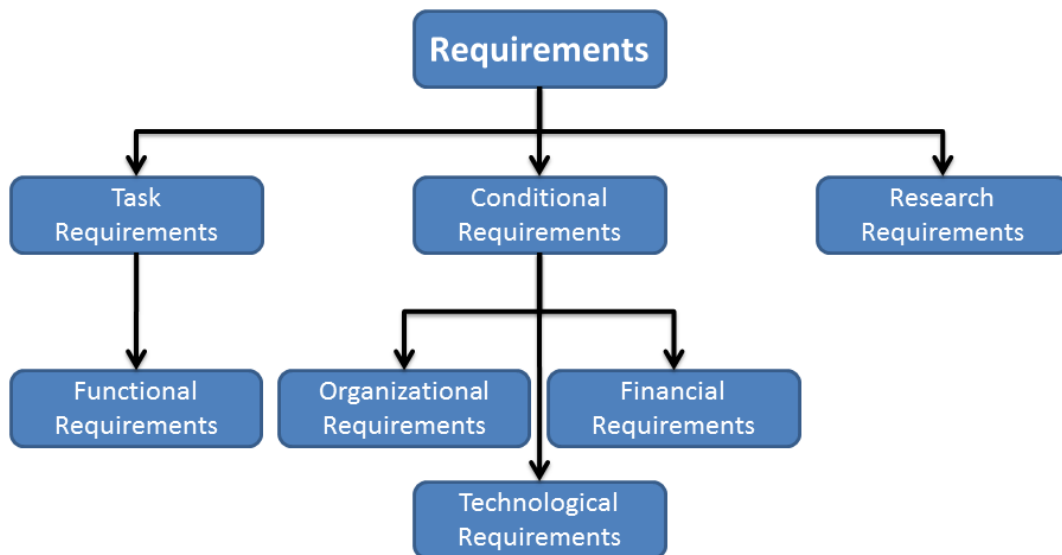
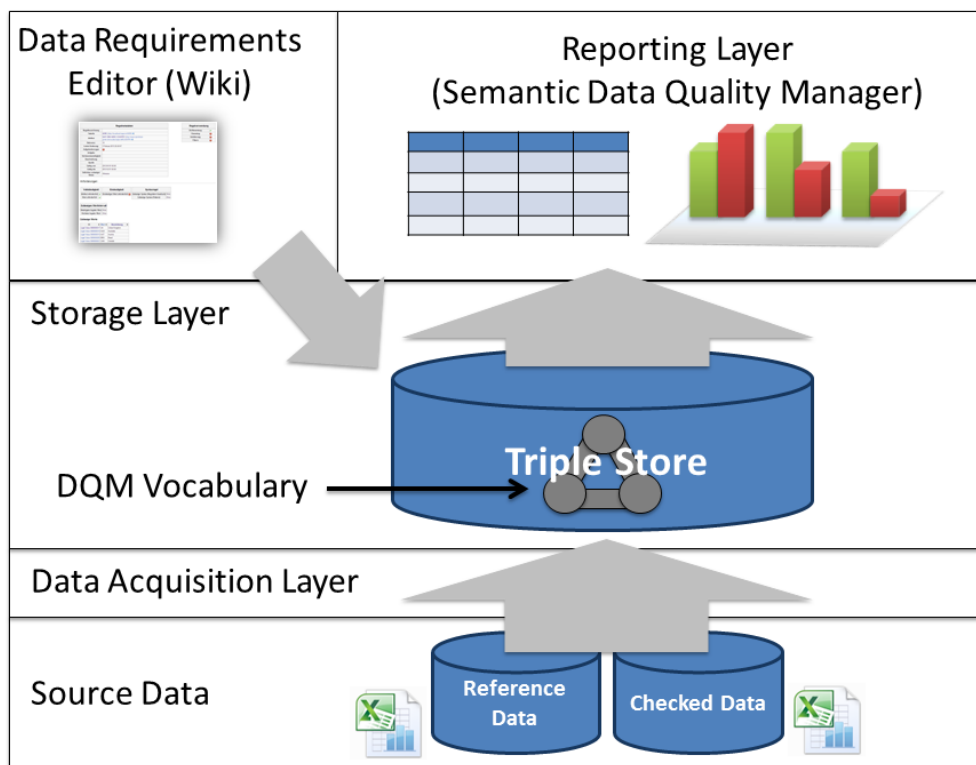


Figure 24: Linking Open Data (LOD) cloud diagram<sup>2</sup> (Cyganiak & Jentzsch, 2011a)

<sup>2</sup> Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/> (Last accessed on April 2<sup>nd</sup> 2012)



**Figure 25:** Typology of requirements for artifact design



**Figure 26:** High-level architecture of the SDQM framework



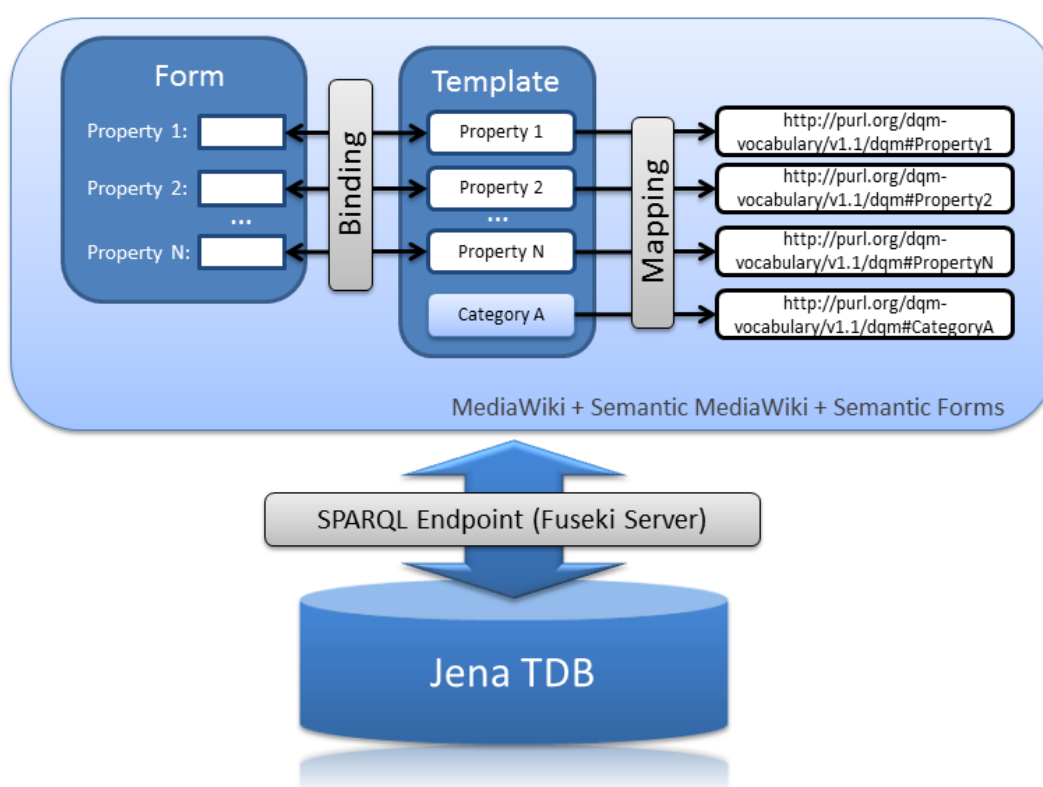
```

{{#ask: [[Category:Location]]
| mainlabel=Wikipage
| ?city
| ?population
| format=table
}}

```

	City	Population
Berlin	Berlin	3,500,000
Bonn	Bonn	325,000
Cologne	Cologne	1,000,000
New York City	New York City	8,100,000

**Figure 28:** Example for an inline query and its result (cf. Dauw et al., 2014)



**Figure 29:** Architecture of SDQM's data requirements wiki

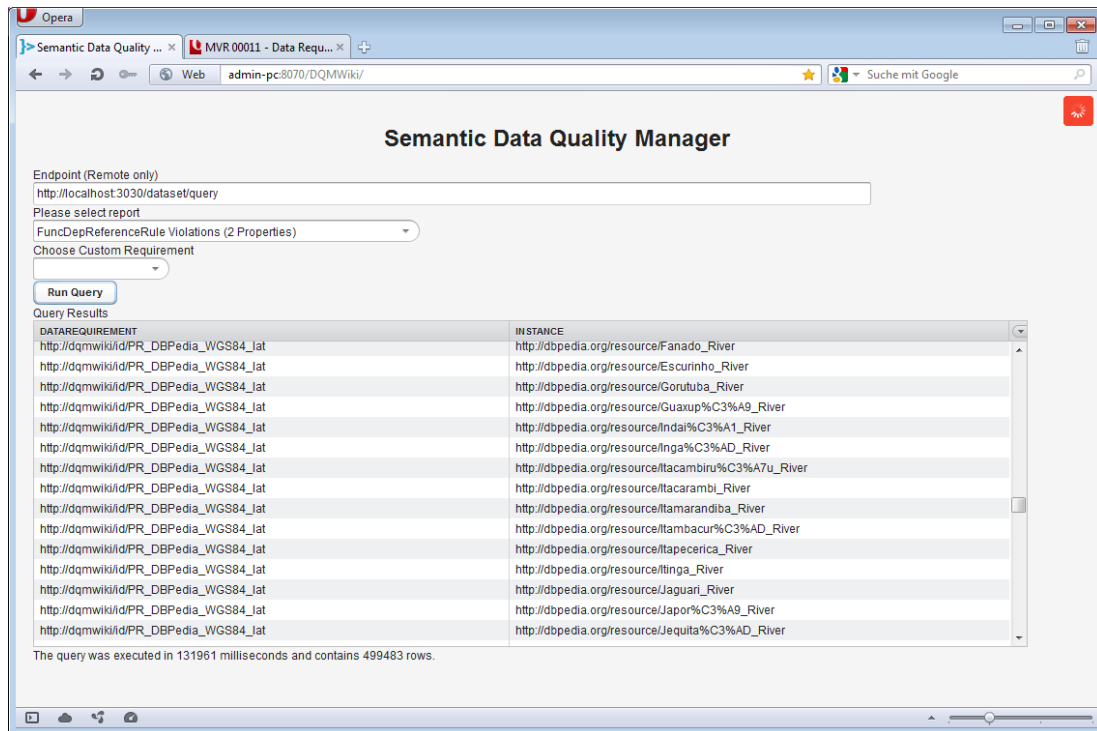


Figure 30: Web-based user interface of the Semantic Data Quality Manager

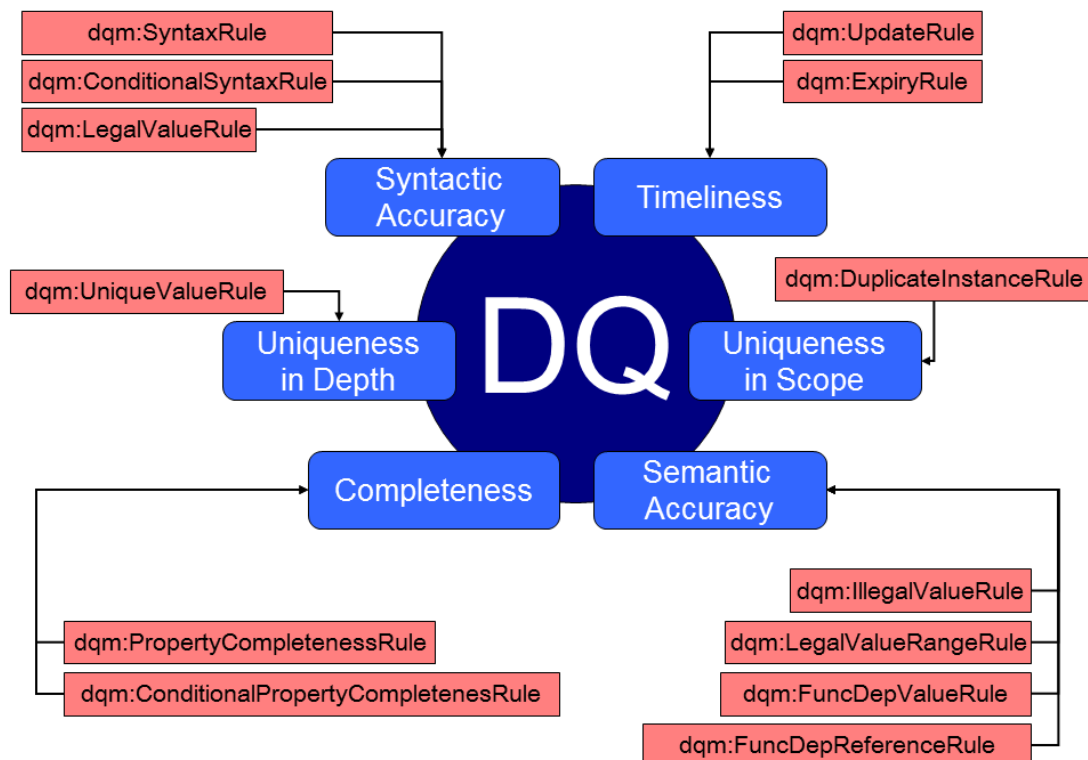
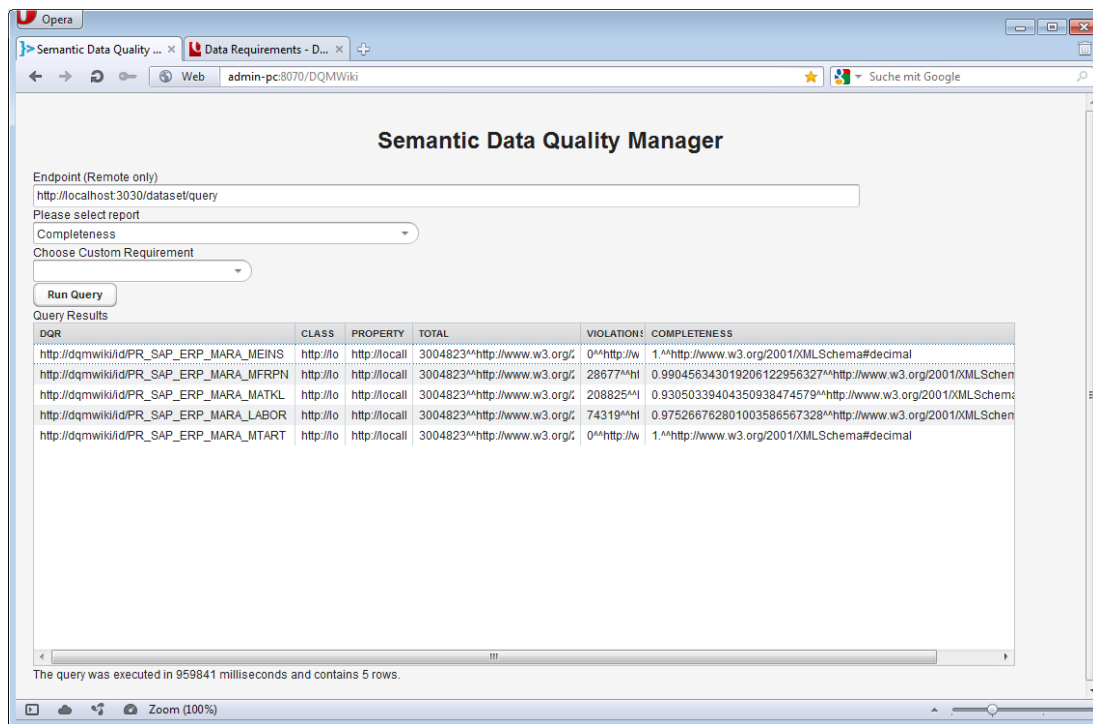


Figure 31: Configuration of data quality assessment reports in SDQM



**Figure 32:** Data quality assessment report of SDQMGr





**Figure 33:** DQM process as supported by SDQM (based on Wang, 1998)

### Create TestedClass

**Label:**

**HasURI:**

**Summary:**

☐ This is a minor edit   ☐ Watch this page

        [Cancel](#)

Categories: [TestedClass](#) | [Prototype](#)

**Figure 34:** SDQM's form to register new tested classes

### Create PropertyRequirement

Name:	Property requirement Supplier ID		Assessment:	<input checked="" type="checkbox"/>
Tested class:	VCARD Organization		Cleansing:	<input type="checkbox"/>
Tested property:	FOO Supplier ID		Validation:	<input type="checkbox"/>
Valid from:	30/05/2012	00:00	Filtering:	<input type="checkbox"/>
Valid until:	31/12/2012	00:00		
Importance:	10		Unit of importance:	
Confidence:	100		Unit of confidence:	
Task dependent:	<input type="checkbox"/>		Applies for task:	

Requirement description: Every record in VCARD Organization must posses a unique supplier ID starting from 1.

Requirement source: Business policy SOP 30/6

#### Completeness / Uniqueness

Property required: ☒ Values must be unique: ☒

RequiredValue: ☒

#### Syntax Rule

Regular Expression:

Pattern Expression:

#### Legal Value Range

Lowest legal value: 1 Highest legal value:

#### Legal Value Rule

Class with legal values:

Property with legal values:

#### Illegal Value Rule

Class with illegal values:

Property with illegal values:

Summary:

**Figure 35:** SDQM's property requirement form

```

{{#ask: [[Category:<CategoryOfLegalValueList>]]
| mainlabel=Page name
| ?legalValue
}}
{{#formlink:form=LegalValue|link type=button|link text=Add Value|popup}}

```

**Figure 36:** Code for a wiki page to maintain lists in the data requirements wiki

Page [Discussion](#) [Read](#) [Edit](#) [View history](#)  [Go](#) [Search](#)

## Legal Value List Suppliers

[Add Value](#)

**Figure 37:** Example of new wiki page for the maintenance of legal value lists

### Create LegalValue

**Legal value:**

☐ None  
☒ LegalSupplier  
☐ LegalValueCountry

☐ This is a minor edit
 ☐ Watch this page

[Save page](#)
[Show preview](#)
[Show changes](#)
[Cancel](#)

Categories: [LegalValue](#) | [Prototype](#)

**Figure 38:** Example of SDQM's form to add legal values

Page [Discussion](#) [Read](#) [Edit](#) [View history](#)  [Go](#) [Search](#)

## Legal Value List Suppliers

Page name	
<a href="#">LegalSupplier Exotic Liquids</a>	Exotic Liquids

[Add Value](#)

**Figure 39:** Example of legal value list in SDQM's data requirements wiki

## Create Condition: Country Germany

**Conditional Property:**

**Operator:**

**Value:**

☐ This is a minor edit ☐ Watch this page

[Cancel](#)

Figure 40: SDQM's form to define conditions

**Create ConditionalRule**

**Name:**  **Assessment:** ☒

**Valid from:**     **Cleansing:** ☐

**Valid until:**     **Validation:** ☐

**Importance:**  **Filtering:** ☐

**Confidence:**  **Unit of importance:**

**Task dependent:** ☐ **Unit of confidence:**

**Requirement description:**  **Applies for task:**

**Requirement source:**

**Conditions (IF)**

**Condition 1:**

**Condition 2:**

**Condition 3:**

**Condition 4:**

**Condition 5:**

**Consequences (THEN)**

**Tested class:**

**Tested property:**

**Completeness consequence**

**Property required:** ☒

**Value required:** ☒

**Syntactic consequence**

**Regular Expression:**

**Pattern Expression:**

**Semantic consequence**

**Operator:**

**Value:**

**Summary:**

☐ This is a minor edit ☐ Watch this page

[Cancel](#)

**Category:**

Figure 41: SDQM's conditional requirement form

## Functional Dependency Reference Rule: MVR 00001

Name:	Locality Country Combinations		Assessment:	<input checked="" type="checkbox"/>
Valid from:	30/05/2012	00:00	Cleansing:	<input type="checkbox"/>
Valid until:	31/12/9999	00:00	Validation:	<input type="checkbox"/>
			Filtering:	<input type="checkbox"/>
Importance:	10		Unit of importance:	
Confidence:	90		Unit of confidence:	
Task dependent:	<input type="checkbox"/>		Applies for task:	
Requirement description:	Locality country combinations in VCARD Organization must be correct.			
Requirement source:				
Tested class:	VCARD Organization			
Tested property 1:	VCARD Locality			
Tested property 2:	VCARD Country Name			
Tested property 3:				
Tested property 4:				
Tested property 5:				
Reference class:	TrustedClass Locality Country Comb			
Reference property 1:	TrustedProperty Locality			
Reference property 2:	TrustedProperty Country			
Reference property 3:				
Reference property 4:				
Reference property 5:				

**Figure 42:** SDQM's functional dependency reference rule form

## Create OutdatedInstanceRule

---

<b>Name:</b>	<input type="text" value="Expiry rule for VCARD Organization"/>	<b>Assessment:</b>	<input checked="" type="checkbox"/>
<b>Tested class:</b>	<input type="text" value="VCARD Organization"/>	<b>Cleansing:</b>	<input type="checkbox"/>
<b>Tested property:</b>	<input type="text" value="FOO Valid Until"/>	<b>Validation:</b>	<input type="checkbox"/>
<b>Valid from:</b>	<input type="text" value="30/05/2012"/> <input type="button" value="📅"/>	<b>Filtering:</b>	<input type="checkbox"/>
	<input type="text" value="00:00"/> <input type="button" value="🕒"/>		
<b>Valid until:</b>	<input type="text" value="31/12/9999"/> <input type="button" value="📅"/>		
	<input type="text" value="00:00"/> <input type="button" value="🕒"/>		
<b>Importance:</b>	<input type="text" value="9"/>	<b>Unit of importance:</b>	<input type="text"/>
<b>Confidence:</b>	<input type="text" value="100"/>	<b>Unit of confidence:</b>	<input type="text"/>
<b>Task dependent:</b>	<input type="checkbox"/>	<b>Applies for task:</b>	<input type="text"/>
<b>Requirement description:</b>	<div>Instances of <u>VCARD</u> Organization must not be elder than their date of expiration.</div>		
<b>Requirement source:</b>	<input type="text"/>		

---

### Timeliness

---

**Expected Update Interval:**

**Instance Expires:** ☒

☐ This is a minor edit   ☐ Watch this page

Category: [Prototype](#)

**Figure 43:** SDQM's form for timeliness requirements

## Create DuplicateInstanceRule

**Name:**  **Assessment:** ☒

**Tested class:**

**Valid from:**

**Valid until:**

**Cleansing:** ☐

**Validation:** ☐

**Filtering:** ☐

**Importance:**  **Unit of importance:**

**Confidence:**  **Unit of confidence:**

**Task dependent:** ☐ **Applies for task:**

**Requirement description:**

Instances that have identical values for contact name, position title, street address, locality name, and postal code are suspicious to be duplicates.

**Requirement source:**

**Duplicate, if following properties have identical values**

**Tested property 1:**

**Tested property 2:**

**Tested property 3:**

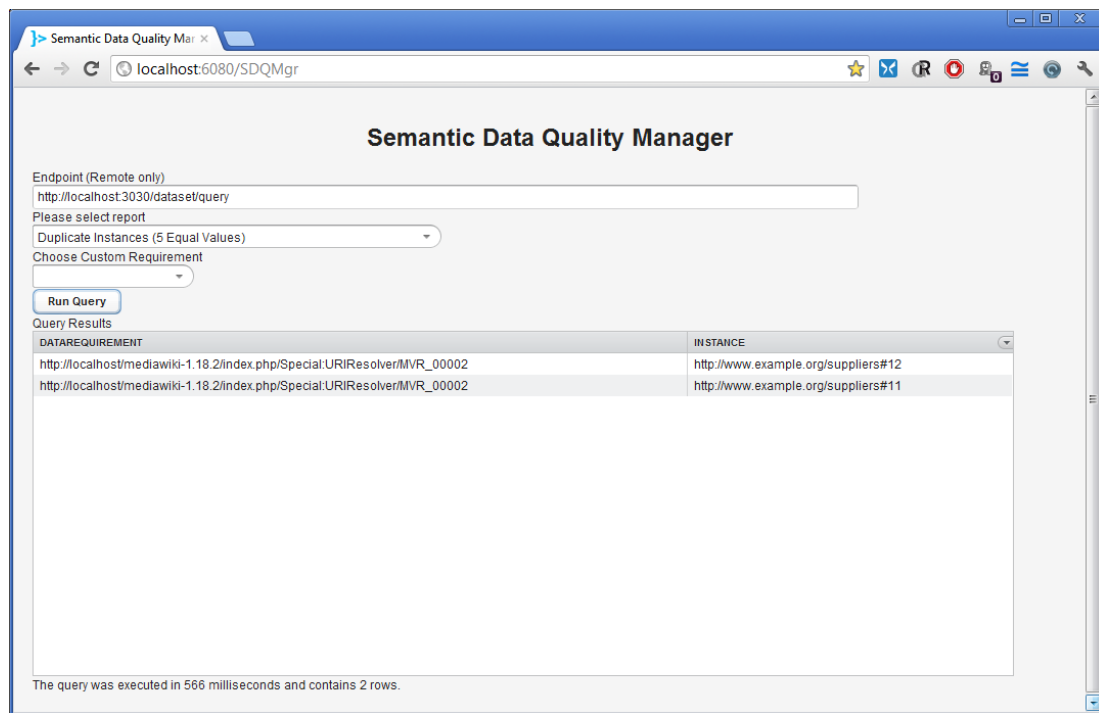
**Tested property 4:**

**Tested property 5:**

☐ This is a minor edit ☐ Watch this page

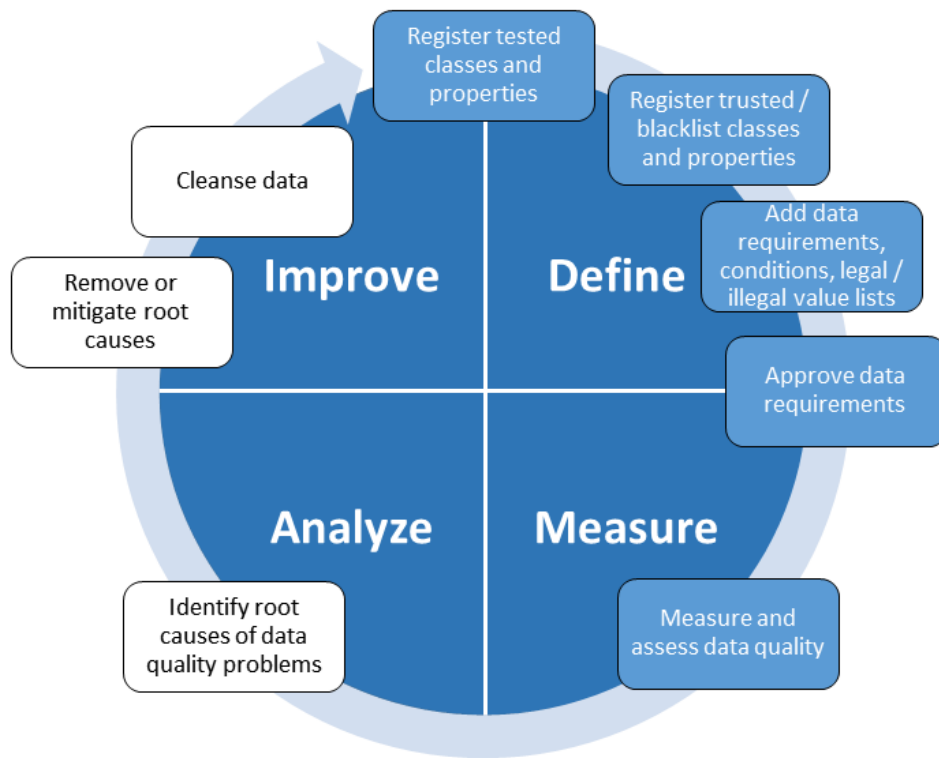
[Cancel](#)

**Figure 44:** SDQM's duplicate instance rule form

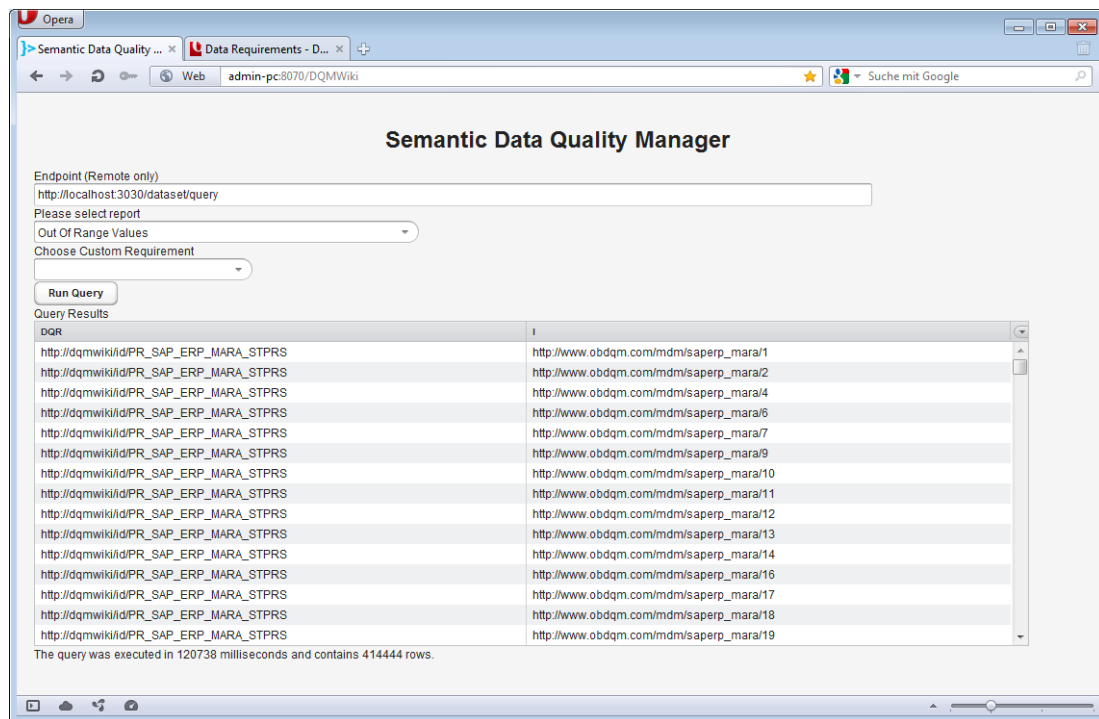


**Figure 45:** Data quality monitoring report of SDQMGr

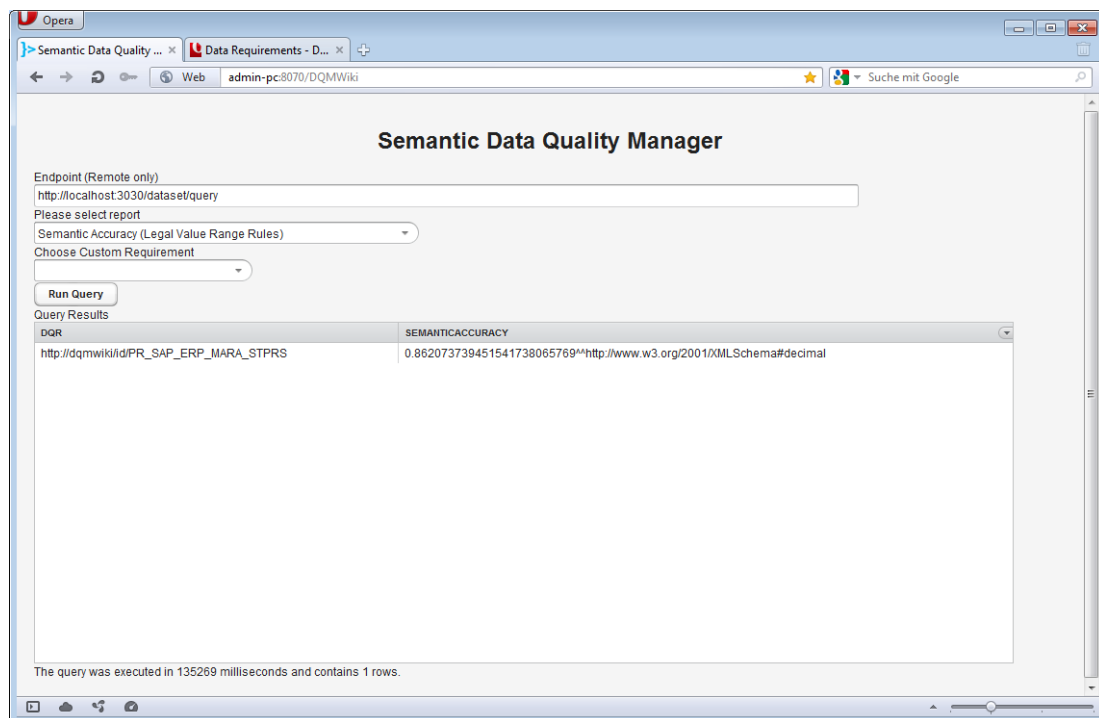




**Figure 46:** SDQM application procedure (based on Wang, 1998)



**Figure 47:** Report with legal value range violations



**Figure 48:** Report with semantic accuracy score based on value range requirement

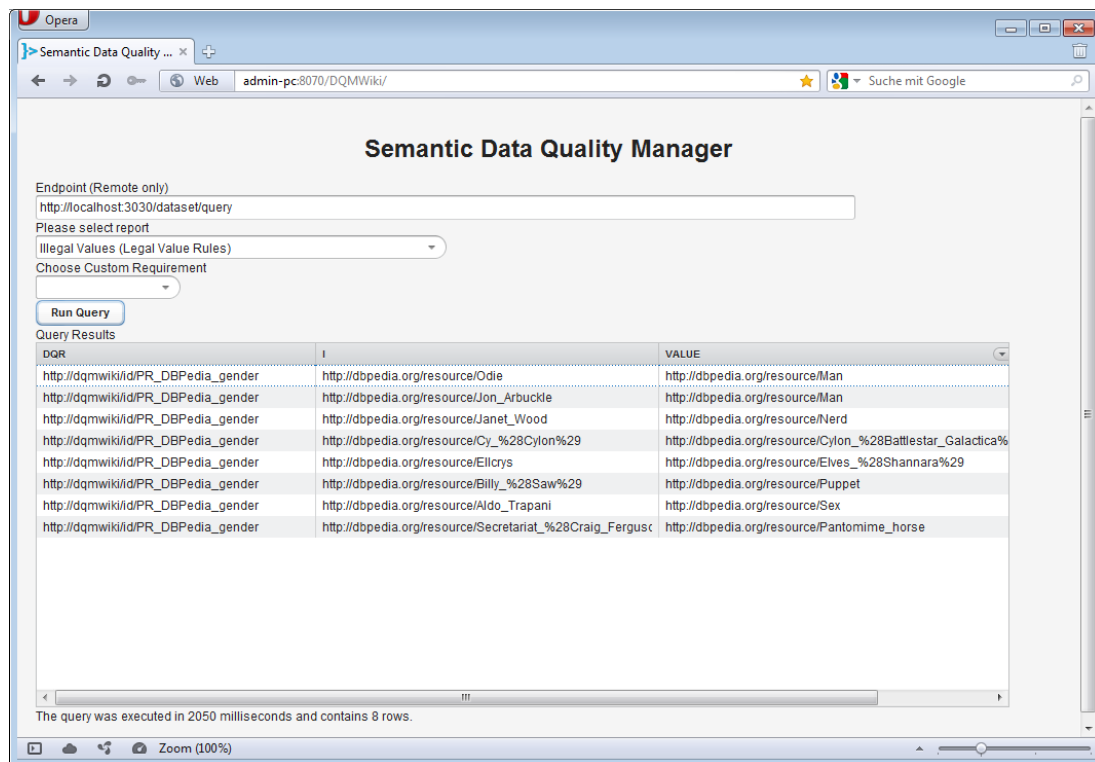


Figure 49: Result of legal value requirement analysis in DBpedia

```

{{Infobox character
| name      = Janet Wood
| image     = [[Image:Janet Wood 1982.png|220px]]
| caption   = Joyce DeWitt as Janet Wood
| first     = "A Man About the House"
| last      = "Friends and Lovers"
| nickname  =
| alias     =
| species   =
| gender    = [[Nerd|Female]]
| occupation = Florist, Aerobics instructor
| title     =
| family    = Roland Wood (father)<br>Ruth Wood (mother)<br>Jenny Wood (sister)<br>unnamed brother
| children  =
| relatives =
| portrayer = [[Joyce DeWitt]]
| creator   =
}}

```

Figure 50: Infobox source code of Wikipedia page "Janet Wood" as of June 27, 2011



Figure 51: Wikipedia page "Cy (Cyclon)" as of June 10, 2012

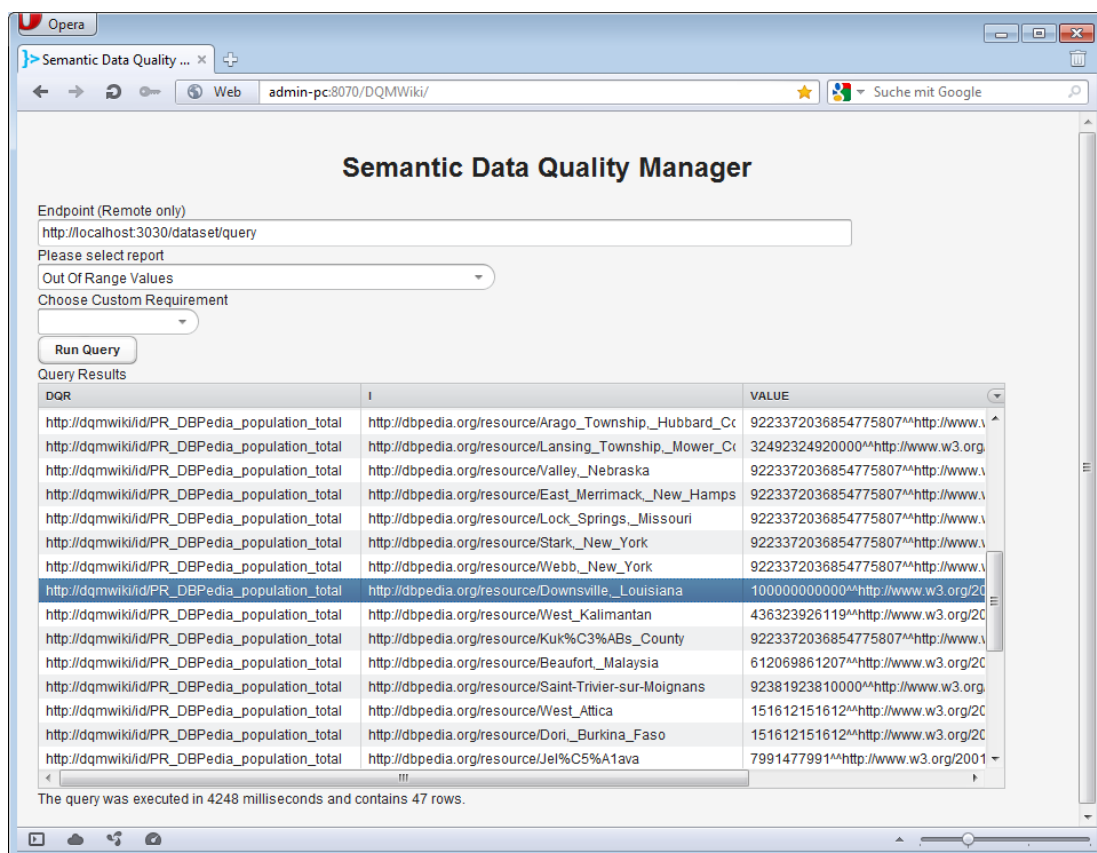


Figure 52: Out of range values for property "population" in DBpedia



Figure 53: Wikipedia page "Downsville, Louisiana" as of June 19th 2011

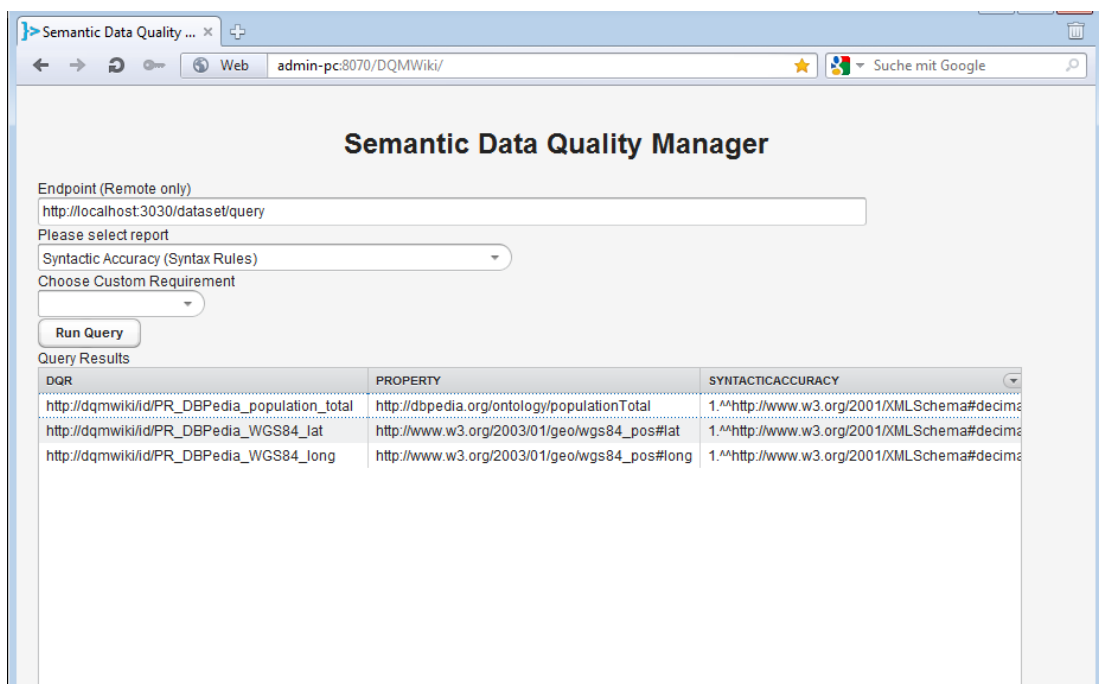


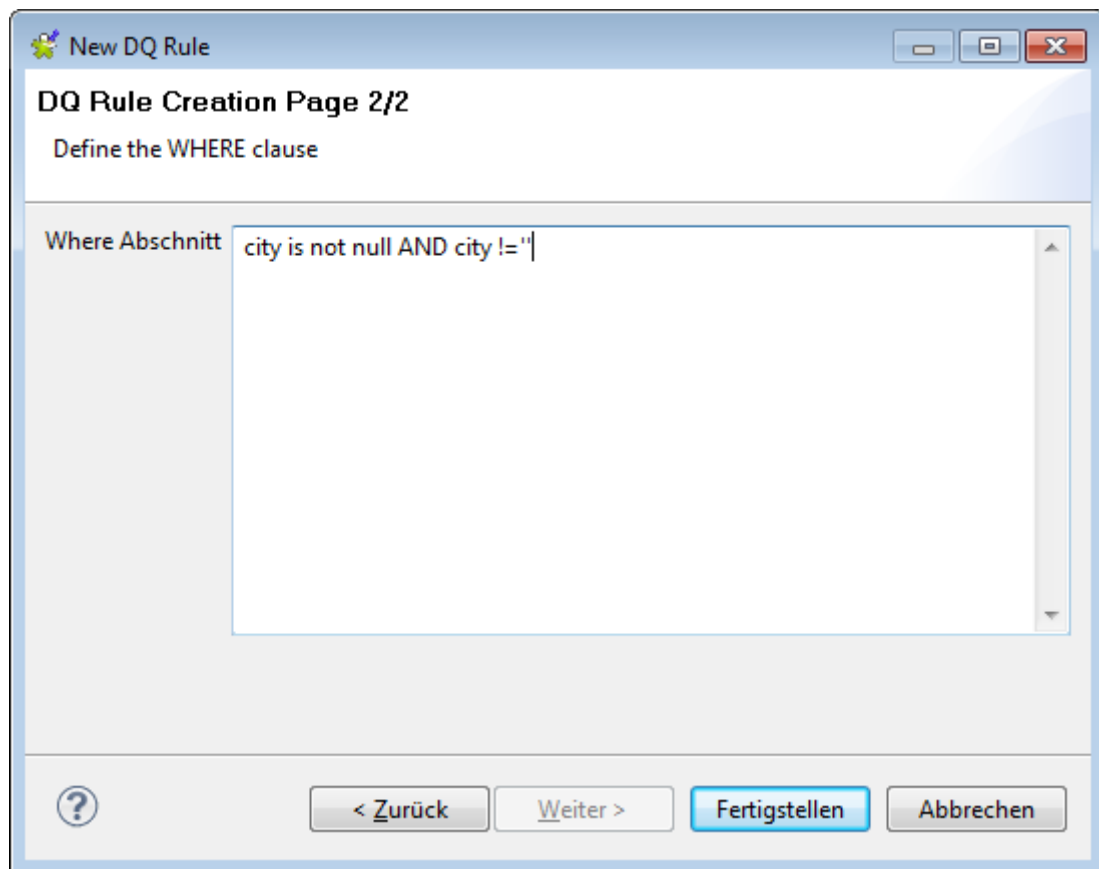
Figure 54: Data quality assessment report displaying syntactic accuracy results

<pre> PREFIX dqm:&lt;http://purl.org/dqm-vocabulary/v1.1/dqm#&gt; SELECT (?tclassURI1 AS ?Class) (?tpropURI1 AS ?Property) (?dr1 AS ?DataRequirement) WHERE {   ?dr1 a dqm:PropertyRequirement .   ?dr1 dqm:testClass ?tclass1 .   ?tclass1 dqm:hasURI ?tclassURI1 .   ?dr1 dqm:testProperty1 ?tprop1 .   ?tprop1 dqm:hasURI ?tpropURI1 . } ORDER BY ?tclassURI1 ?tpropURI1 ?dr1 </pre>		
Class	Property	DataRequirement
"http://www.w3.org/2006/vcard/ns#Organization"	"http://www.example.org/suppliers#supplierID"	<http://localhost/mediawiki-1.18.2/index.php/Special:URIResolver/PR_Organization_EXAMPLE_Supplier_ID>
"http://www.w3.org/2006/vcard/ns#Organization"	"http://www.example.org/suppliers#supplierID"	<http://localhost/mediawiki-1.18.2/index.php/Special:URIResolver/PR_Organization_FOO_Supplier_ID>
"http://www.w3.org/2006/vcard/ns#Organization"	"http://www.example.org/suppliers#supplierID"	<http://localhost/mediawiki-1.18.2/index.php/Special:URIResolver/PR_Organization_Supplier_ID>
"http://www.w3.org/2006/vcard/ns#Organization"	"http://www.w3.org/2006/vcard/ns#country-name"	<http://localhost/mediawiki-1.18.2/index.php/Special:URIResolver/PR_VCARD_Country_Name>
"http://www.w3.org/2006/vcard/ns#Organization"	"http://www.w3.org/2006/vcard/ns#fn"	<http://localhost/mediawiki-1.18.2/index.php/Special:URIResolver/PR_VCARD_Formatted_Name>

Figure 55: SPARQL query and result displaying duplicate property requirements

<pre> PREFIX dqm:&lt;http://purl.org/dqm-vocabulary/v1.1/dqm#&gt; SELECT (?dr1 AS ?UniqueValueRequirement) (?dr2 AS ?InconsistentRequirement) WHERE{   ?dr1 a dqm:UniqueValueRule .   ?dr1 dqm:testClass ?tclass1 .   ?tclass1 dqm:hasURI ?tclassURI1 .   ?dr1 dqm:testProperty1 ?tprop1 .   ?tprop1 dqm:hasURI ?tpropURI1 .   OPTIONAL{     ?dr2 a dqm:PropertyRequirement .     ?dr2 dqm:testClass ?tclass2 .     ?tclass2 dqm:hasURI ?tclassURI2 .     ?dr2 dqm:testProperty1 ?tprop2 .     ?tprop2 dqm:hasURI ?tpropURI2 .     FILTER(str(?tpropURI1) = str(?tpropURI2) &amp;&amp; str(?tclassURI1) = str(?tclassURI2) &amp;&amp; ?dr1 != ?dr2)   }   MINUS{     ?dr2 a dqm:UniqueValueRule   } } FILTER(bound(?tpropURI2)) </pre>		
UniqueValueRequirement	InconsistentRequirement	
<http://localhost/mediawiki-1.18.2/index.php/Special:URIResolver/PR_Organization_FOO_Supplier_ID>	<http://localhost/mediawiki-1.18.2/index.php/Special:URIResolver/PR_Organization_EXAMPLE_Supplier_ID>	
<http://localhost/mediawiki-1.18.2/index.php/Special:URIResolver/PR_Organization_Supplier_ID>	<http://localhost/mediawiki-1.18.2/index.php/Special:URIResolver/PR_Organization_EXAMPLE_Supplier_ID>	

Figure 56: SPARQL query for identification of inconsistent property requirements



**Figure 57:** SQL business rule in Talend OS for Data Quality

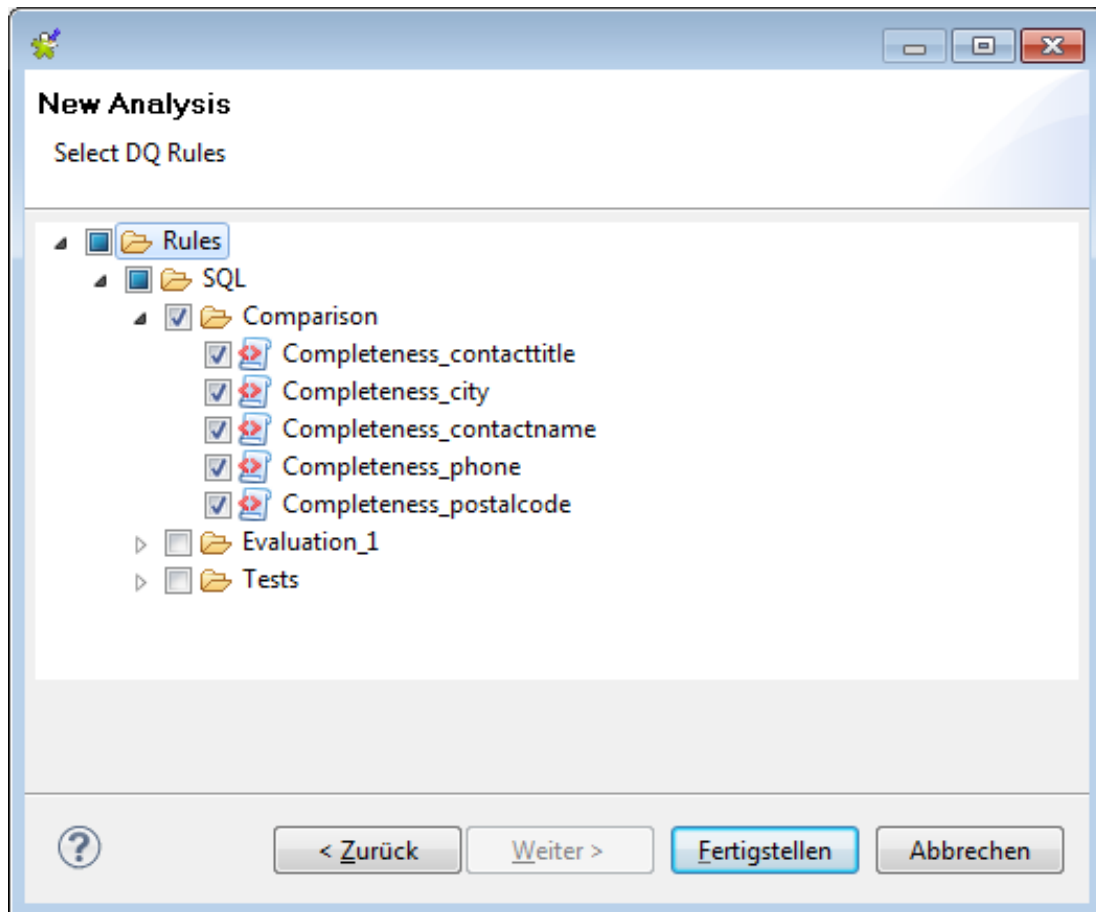


Figure 58: Selecting SQL business rules in Talend OS for Data Quality

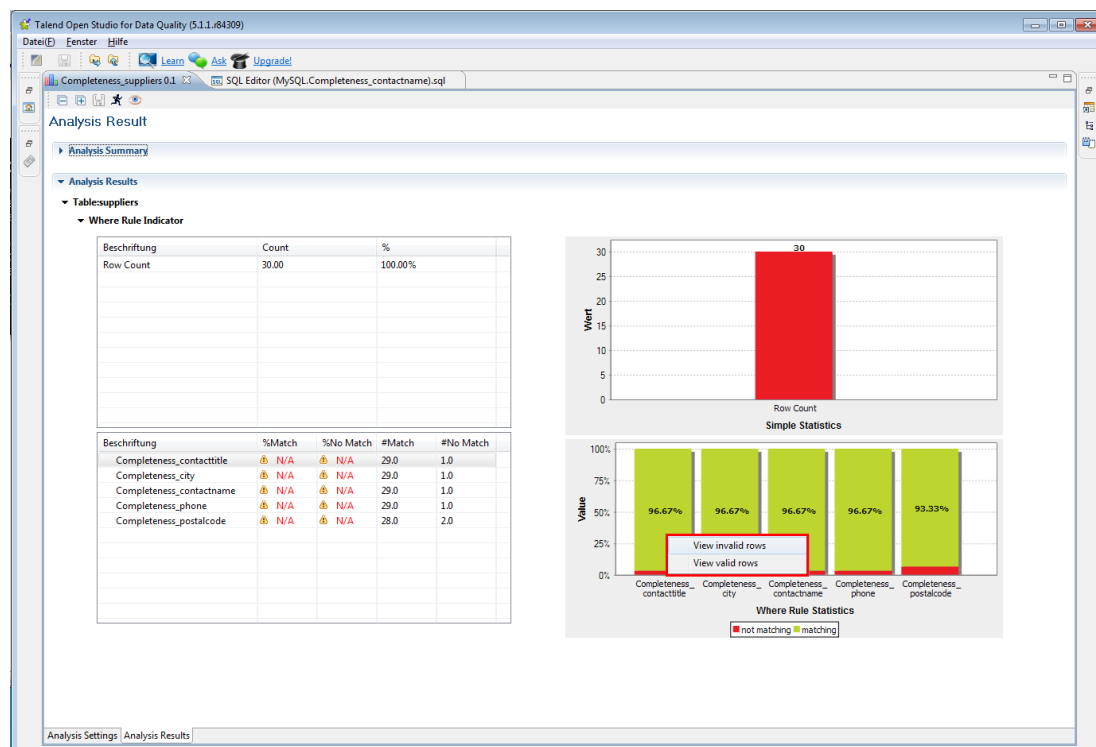
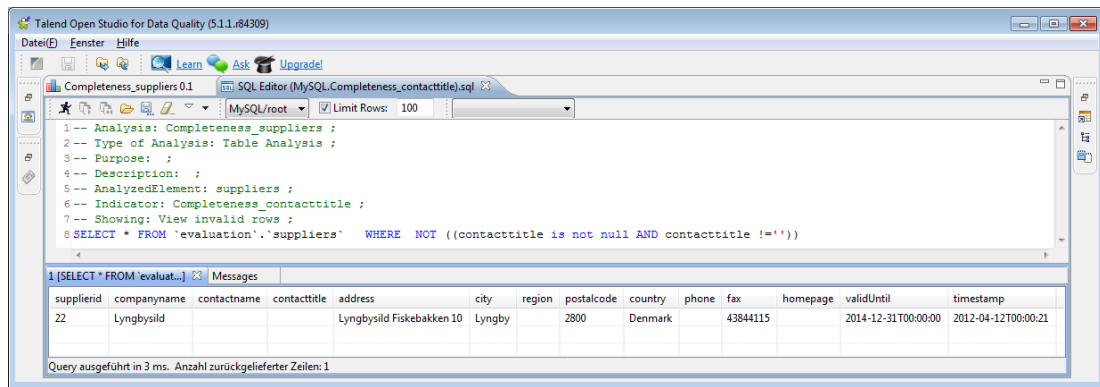
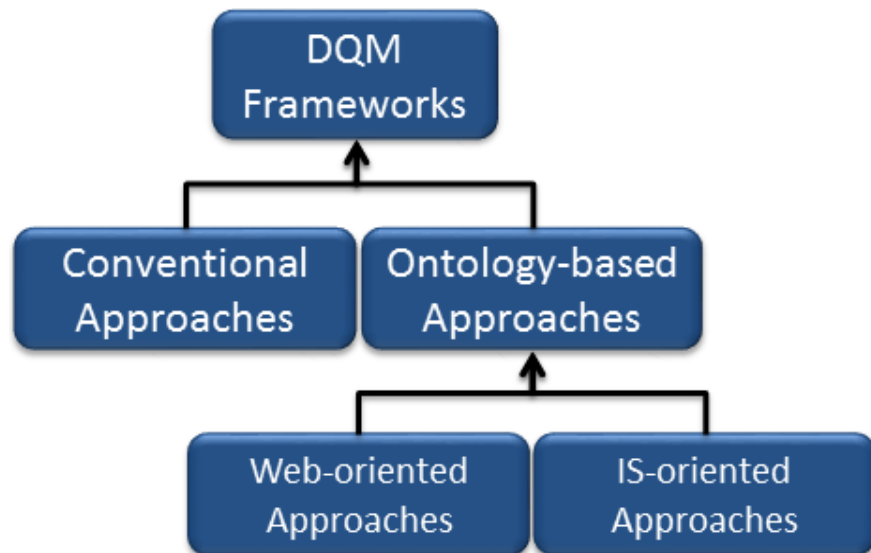


Figure 59: Data quality assessment report in Talend OS for Data Quality





**Figure 60:** Data quality monitoring report of Talend OS for Data Quality



**Figure 61:** High-level classification of DQM frameworks

Category	Subcategory				
Data Lifecycle Step	Data Acquisition		Data Usage		Data Retirement
Data Representation	Relational Data	XML Data	RDF Data	Text / Web Site	Other
Data Quality Task	Data Profiling	Data Requirements Management	Data Quality Monitoring	Data Quality Assessment	Data Cleansing
	Data Validation	Information Filtering	Data Integration	Master Data Management	

**Figure 62:** Categorization schema for related work

Author										Supported Data Lifecycle Step				Supported Data Representation				Supported Data Quality Task								Role of Ontologies																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
Acquisition				Usage		Retirement		Relational		XML		Text / Unstructured Data		RDF		Other		Data Profiling		Data Requirements Management		Data Quality Monitoring		Data Quality Assessment		Data Cleansing		Data Validation		Information Filtering		Data Integration		Master Data Management		Quality-relevant Metadata		Representation of Domain Knowledge		Representation of Provenance Metadata		SWRL / Reasoning rules																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
Conventional Rule-based Approaches																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
		Loshin, 2002																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		</

Figure 63: Own classification of related wor

Data Quality Management with Semantic Technologies

Fürber, C.

2016, XXVII, 205 p. 63 illus., Softcover

ISBN: 978-3-658-12224-9