SBVR as a Semantic Hub for Integration of Heterogeneous Systems

A Case Study and Experience Report

11th July, 2013

Ling Shi, Dumitru Roman, and Arne J. Berre

E-mail: lish@statsbygg.no
RuleML 2013 Challenge Track

Outline

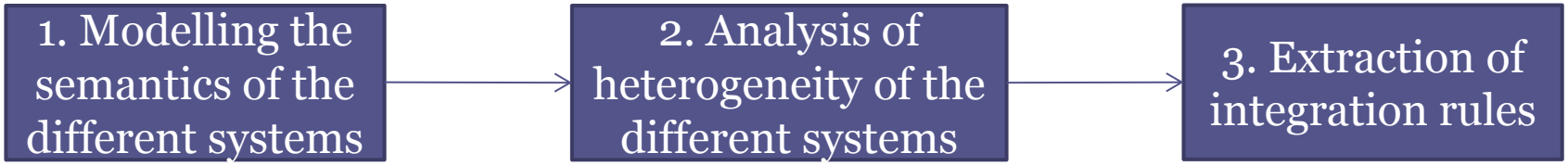
- Introduction
- The Case Study
- Realization of the SBVR Hub
- Conclusion and further work
- Q&A

Introduction

- Challenges in integration of heterogeneous systems
  - Difficult to extract integration rules
- Integration rules
  - Handle the heterogeniety
  - Decide what to integrate, under what conditions and how to integrate

Introduction – Integration Rules

- Extraction of integration rules – 3 steps



- Why SBVR is appealing as a modeling language for the process above?
  - Machine readability provides possibility of utilizing automated reasoning
  - Human readability provides possibility of involving domain experts as well as others without special semantic modeling education in the whole process above
  - SBVR provides both human and machine readability and thereby a good candidate

Introduction – The SBVR Hub

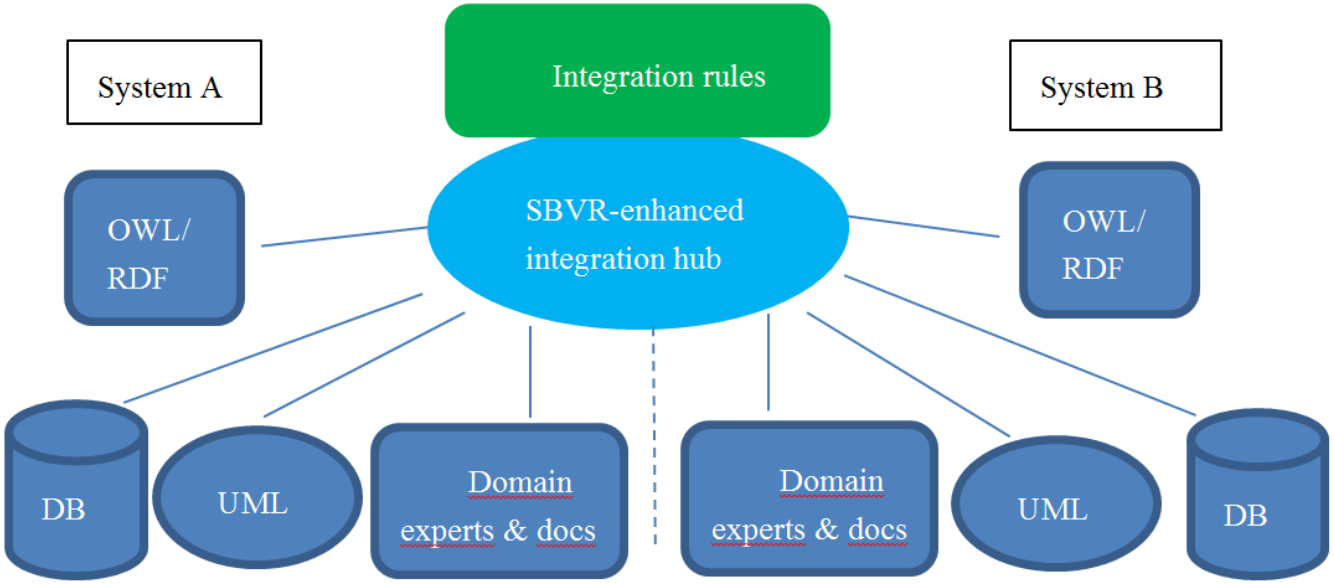
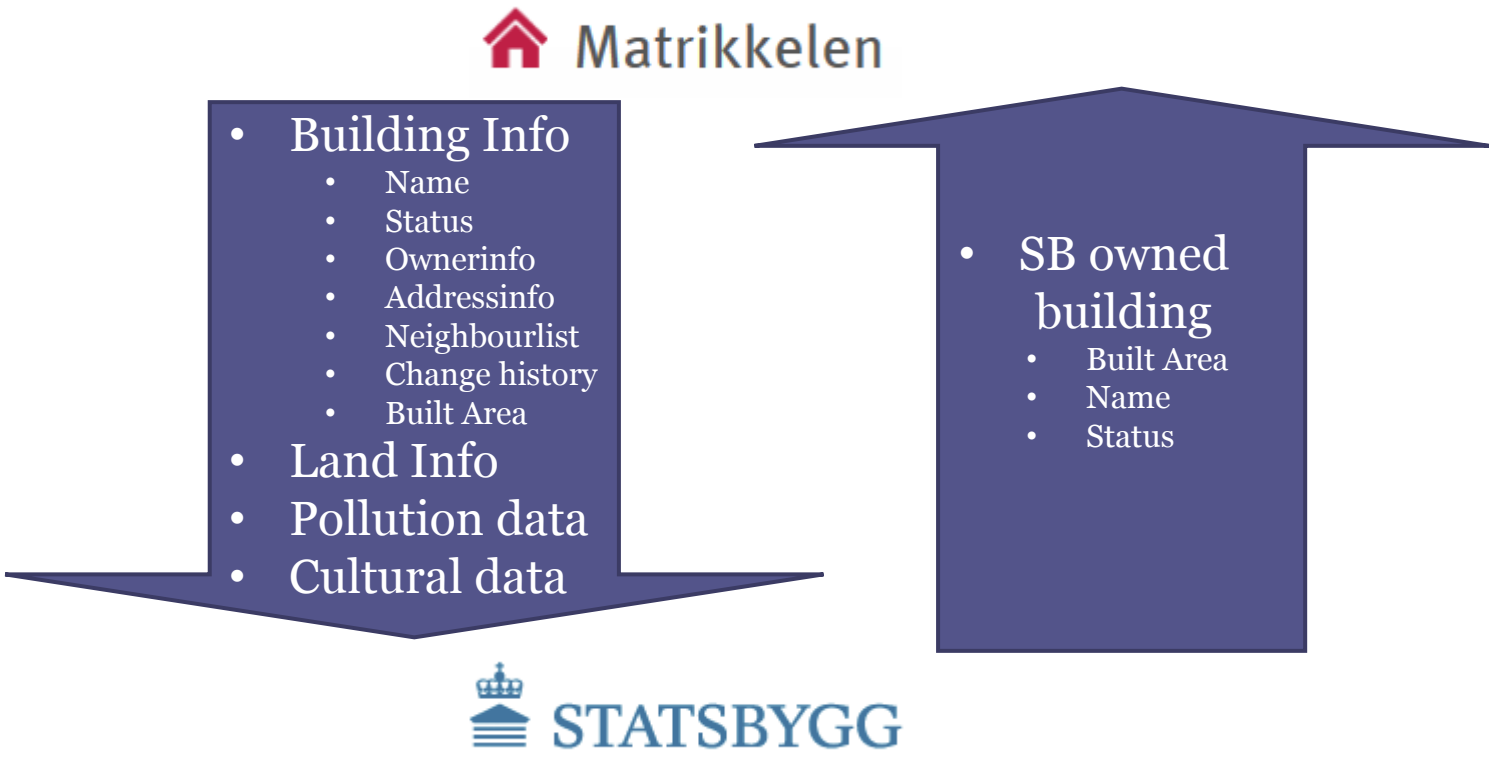


Fig. 1. SBVR-enhanced integration hub

The Case Study



The Case Study

- Matrikkel information is needed by real property planing, developing, acquisitions, maintaining, i.e. by all the major business functions in SB
  - Rental income represents 50% of SBs total income. Rental price is currently decided by the usage area, location and other building related informations.
  - Real property planning needs to retrieve among others cultural heritage information on the building and around the building.
  - Acquisition or sale of buildings or lands need building and land information
  - Building development needs the neighbour list to send them formal information

- Data quality in the internal property management system "Propman" is essential for SB. Integration of Propman and the national cadastral system "Matrikkel" is one of the most important integration tasks in SB

The Case Study – earlier integrations
Integration 2007: ignore the mismatch

- Based on the matching Matrikkel's building number in Propman though 25% buildings in Propman did not have the key registered

- The quality of registered Matrikkel's building numbers was not good either
  - Some numbers were used several times on different buildings in Propman
  - Some numbers did not belong to the right building
  - There were several kinds of errors after the updating process according to the feedback from the property managers

The Case Study – earlier integrations

Integration 2012: handle the mismatch manually

- It tried to cover all the buildings including those without a Matrikkel's building number
- Mismatches of data in the two systems were listed up and suggestions of modifications were delivered to the property management administrator who forwarded them to each property manager individually
- Each property manager then made an evaluation based on his/her domain knowledge and the corrections were therefore not 100% consistent.
- Major reasons for this time and resource consuming process included
  - Lack of a suitable vocabulary that could define terms like building, building's built area, building's address, etc.
  - Lack of clear structures and rules that could be used to decide which building should be included in the register and what to do if mismatches between Matrikkel and Propman occurred

The Case Study – A better approach

- To improve the quality of the integration, we have to deal with the mismathes instead of ignoring them
- To handle the mismathes consistently and effectively, we need a better approach that minimize the inconsistency caused by individual evaluations by domain experts
- A centralized semantic hub would provide the needed consistency and effectiveness

Realization of the SBVR Hub Overall Aproach

- Part 1: Establishing the SBVR hub
  - Building the vocabularies, fact models, and business rules for the source systems.

- Part 2: Extracting Integration Rules
  - Identifying the mismatches of term definitions in the vocabulary, business rules and fact models; Building the integration rules after the mismatches.

Realization of the SBVR Hub

- Part 1: Establishing the SBVR Hub
  - Building the vocabularies using a standard template. There are 92 terms in Matrikkel vocabulary and over 500 in the SB vocabulary.
  - Building the fact models based on UML diagrams and database schemas
  - Building the business rules by extracting rules from existing definitions and knowledge from domain experts

Realization of the SBVR Hub

- Part 2: Extracting Integration Rules
  - Identifying the mismatches of term definitions in the vocabulary, business rules and fact models
  - Terminological heterogeneity: two properties with different names deal with the same information
  - Conceptual heterogeneity: e.g. heterogeneity in coverage
  - Building the integration rules after the mismatches

Realization - Integration Rule Contents

- What part of resource/property X in source system A should integrate with what part of resource/property Y in source system B with which integration keys under which conditions?
  - X
  - Y
  - Integration Key(s)
  - Condition(s)

Realization - Integration Rule Ex.

- To handle the terminological heterogeneity, an integration rule should define the original source of a data property in case of duplications
  - This kind of integration rule can be interpreted as an ontology rule using, e.g., rdfs:subPropertyOf or owl:sameAs.

- ***R-Int-bygg-1:***
  - *A Building's property Area_Built in Propman is the same as a Building's property Built_Area in Matrikkel with the integration key "Matrikkel's Building number".*

Realization - Integration Rule Ex.

- One alternative way to handle the "Difference in coverage" type of conceptual heterogeneity is to define the overlapping part as integration part
  - For example, the Building in Matrikkel and the Building in Propman are overlapping when both have the same Matrikkel's building number, then the integration rule below defines that only building with Matrikkel's building number can be integrated

- ***R-Int-bygg-2:***
  - *Each building with Matrikkel's buildingnummer in SB can integrate with a building in Matrikkel on Matrikkel's building number*

Conclusion

- The case study presented in this paper is a preliminary attempt to design and implement an integration analysis framework based on machine and human readable SBVR.

- Benefits of an SBVR Hub
  - Reusability (2N vs. NxN)
  - Machine readability provides possibility of automated reasoning
  - Human readability
  - Improve the consistency and effectiveness

- Possible limitations of an SBVR Hub
  - An extra layer above the original models of UML, DB schemas etc., (information may be lost in the transformation between original models and SBVR)
  - Resource demanding to maintain the hub

Future work

- Automate the transformation process from source models to SBVR
- Review existing support for tools
- Standardize the comparison and analysis of heterogeneity
- Design the extraction of rules as an ontology reasoning process
- Evaluate RuleML for automated generation of integration rules

Thank you!

Q & A