Data Integration with SMW+
Michael Erdmann, ontoprise Karlsruhe

SMWCon Spring 2011, Arlington VA
Agenda

- Motivation
- General workflow and architecture
- OntoStudio (with demo)
  - From relational database to wiki vocabulary
- SMW+ (with demo)
  - Using external data
  - Enhancing/extending external data
- Summary and Outlook
Motivation

- Enterprises typically manage their data in relational databases.
- Often a landscape of isolated data silos grows.
- In contrast, a wiki provides one platform for sharing information.
- Nevertheless,
  - the databases exist
  - contain valuable information
  - will not be deserted
- Enterprise use cases we see often require access to external data from within the wiki.
  - Usually the external data sources are relational databases.
  - Wiki is used as a data access and visualisation tool.
  - Wiki users should be able to formulate queries.
Workflow for Data Integration with SMW+

### OntoStudio
- Lift database schema
- Model wiki ontology
- Map database ontology to wiki ontology
- Test and refine mappings
- Export ontologies
- Transfer to Wiki server
- Import ontology
- Test and refine ontology

### SMW+ / TSC
- Import ontologies
- Formulate queries (QI)
- Create articles containing inline queries
- Browse database via Non-Existing-Pages
- Potentially, enhance model via annotations
- Adapt and extend wiki ontology
- Export ontology as OBL
Architecture

SMW+

Queries ➔ Results ➔ SMW+ ➔ Annotations

Triple Store Connector

OntoBroker

Ontology / Rules ➔ Semantic Facts

Relational Databases
Database Lifting

- Database schemas are table-based and usually geared towards performance not conceptual clarity
- Schemas are **lifted** to an ontological level automatically
  - Tables become classes
  - Foreign-keys become object properties (in the OWL sense)
  - Other columns become data properties
- The result is an **ObjectLogic ontology** representing the database
- This ontology contains **data-access rules** that populate classes and properties
  - These rules are OBL rules **reaching out into the database**
  - Rules are triggered **at query time**
Live Demo of OntoStudio
OntoStudio: Database Lifting

Edit database connection data

Database: mssqlserver
Database name: OS1.6 Demo DB
Host: data
Port: 1433
User name: sa

Import
Select table owners

Table owner:
- dbo
- demo
- guest
- os_tester
- test

Import
Select tables to import

- card
- name
- company
- type
- countries
- name
- continent
- dt_properties
- prices
- sales
- region
- car
- amount

Refresh

< Back Next >
OntoStudio: Modeling
OntoStudio: Graphical Rule Modeling
OntoStudio: Mapping
OntoStudio: Testing
Query Answering

TSC with OntoBroker

mapping results

manually mappings

query

SMW+

integrated results

generated ontologies

access rules

automatic schema mapping

facts from source

databases

mapping rules

query

Wiki ontology

mapping results

generated ontologies

access rules
Workflow

**OntoStudio**
- Lift database schema
- Model wiki ontology
- Map database ontology to wiki ontology
- Test and refine mappings
- Export ontologies
- Transfer to Wiki server

**SMW+ / TSC**
- Import ontologies
- Formulate queries (QI)
- Create articles containing inline queries
- Browse database via Non-Existing-Pages
- Potentially, enhance model via annotations
- Adapt and extend wiki ontology
- Export ontology as OBL
Import Ontology into Wiki

- The wiki ontology coming from OntoStudio needs to be imported.
- This is a task for the **Deployment Framework (DF)**
- Ontologies can be imported from a file or from a repository
- Technically, the ontology file is converted into the MW XML-format and then imported by wiki tools.
  - The external tool for this task is “onto2mwxml”. It can also handle OWL, RDF/XML, N3, NTRIPLE, OBL
  - `smwadmin -i myOntology.obl`
ObjectLogic

- OBL contains stuff that has no equivalent in the wiki
  - Predicates
  - Mapping rules
  - Queries
- They are stored as a separate file in the wiki and are copied into the TSC as an individual module to be executable.

- TSC:
  - Wikimodule (contains all wiki annotations, incl. the ones imported from OBL) is automatically synched with Wiki
  - TSC-only part from imported OBL (this module is imported by the wikimodule)
  - Other modules, e.g. module created by lifting a database
Wiki Part of Data Integration
Imported Ontology in OntologyBrowser

The ontology browser lets you navigate through the ontology to easily find and identify items in the wiki. Use the Filter Mechanism at the upper left to search for specific entities in the ontology and the filters below each column to narrow down the given results. Initially the flow of browsing is left to right. You can flip the flow by clicking the big arrows between the columns.

Press Ctrl+Alt+Space to use auto-completion (Ctrl+Space in IE).
Rules of an ontology are stored on a separate page for each ontology.
Example of an Imported Page

- Imported page **Property:HasName**
  - Property has two domains: **Product, Region**
  - Type is **String**
  - The **original URI** in the ontology is
    - [http://www.NewOnto1.org/ontology#hasName](http://www.NewOnto1.org/ontology#hasName)
    - This is separately stored in a mapping table to provide fast access whenever wiki names have to be replaced by OB URIs or vice-versa.
  - The (automatically created) bundle is **Ontology-v9**
Query for Integrated Data

{{#ask: 
[[Category:Price]]
[[InRegion.HasName::+]]
[[OfProduct.HasName::+]]
[[OfProduct.HasManufacturer::+]]
| ?Price
| ?InRegion.HasName=Region name
| ?OfProduct.HasName=Product name
| ?OfProduct.HasManufacturer=Manufacturer
| source=tsc
}}
Non-Existing Pages

- An instance created from a database mapping is represented by a red-link.
- In case that there is no real database key, OntoBroker creates one automatically.
- If the user clicks on a red link, the non-existing page handler takes care of it
  - It displays all statements about this instance as well as all statements which have this instance as object.
Non-Existing Pages

This is the NEP page with the default template. It can be user-defined.

Create the article Dbo cars 565041163 Audi A2 with the content displayed below. This resource is stored in the underlying knowledge-base with the following URI: <obt.term%3Chttp://www.NewOnto1.org/dbOntology%23c%3E(%3Chttp://www.NewOnto1.org/dbOntology/dbo%23cars_565041163%3E,%22Audi%20A2%22)>.

Dbo cars 565041163 Audi A2 belongs to the following categories: Cat Car, Cat Product, Cat TopSeller

<table>
<thead>
<tr>
<th>Facts about Dbo cars 565041163 Audi A2</th>
<th>References to Dbo cars 565041163 Audi A2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
</tr>
<tr>
<td>HasManufacturer</td>
<td>dbo_prices_565041163_Audi_A2_France</td>
</tr>
<tr>
<td></td>
<td>dbo_prices_565041163_Audi_A2_Germany</td>
</tr>
<tr>
<td></td>
<td>dbo_prices_565041163_Audi_A2_Italy</td>
</tr>
<tr>
<td></td>
<td>dbo_prices_565041163_Audi_A2_Portugal</td>
</tr>
<tr>
<td></td>
<td>dbo_prices_565041163_Audi_A2_Spain</td>
</tr>
<tr>
<td></td>
<td>dbo_sales_565041163_France_Audi_A2</td>
</tr>
<tr>
<td></td>
<td>dbo_sales_565041163_Germany_Audi_A2</td>
</tr>
<tr>
<td></td>
<td>dbo_sales_565041163_Italy_Audi_A2</td>
</tr>
<tr>
<td></td>
<td>dbo_sales_565041163_Portugal_Audi_A2</td>
</tr>
<tr>
<td></td>
<td>dbo_sales_565041163_Spain_Audi_A2</td>
</tr>
<tr>
<td></td>
<td>dbo_sales_565041163_Sweden_Audi_A2</td>
</tr>
<tr>
<td></td>
<td>dbo_sales_565041163_Switzerland_Audi_A2</td>
</tr>
</tbody>
</table>

O                                      |                                          |
|                                        |                                          |
|                                        |                                          |
Extend the Ontology in the Wiki

- Create a new category
Extend the Ontology in the Wiki

- Create a rule in the wiki with the editor

Derive Category TopSeller by complex rule

**Head**

All articles $X_1$ belonging to Category TopSeller are defined by

**Body**

All articles $X_2$ belong to category RevenueByProduct
AND
All articles $X_1$ belong to category Product
AND
All articles $X_2$ have the property HasProduct with value $X_1$
AND
All articles $X_2$ have the property Value with value a certain value > 100000
AND
Being member of a certain category or property

This rule implies the following:

Generate rule

```
?X1:TopSeller :-
  ?X2:RevenueByProduct AND
  ?X1:Product AND
  ?X2[HasProduct->?X1] AND
  ?X2[Value->?_VALUE0] AND
  ?_VALUE0 > 100000.
```
To re-import the ontology in OntoStudio, export it via the OBL export bot.
Exported ontologies are stored in the wiki as files.

- Upload a new version of this file
- Edit this file using an external application (See the setup instructions for more information)

File links

The following page links to this file:

- Smw exportobjectlogicbot at 2011 3 21 15 48 58
**Workflow**

### OntoStudio
- Lift database schema
- Model wiki ontology
- Map database ontology to wiki ontology
- Test and refine mappings
- Export ontologies
- Transfer to Wiki server

### SMW+ / TSC
- Import ontologies
- Formulate queries (QI)
- Create articles containing inline queries
- Browse database via Non-Existing-Pages
- Potentially, enhance model via annotations
- Adapt and extend wiki ontology
- Export ontology as OBL

**Import ontology**
- Refine and test ontology
Re-imported Ontology

- OS shows new wiki entities
Summary and Outlook

- SMW+ and TripleStoreConnector realize
  - Access to relational data from within SMW
  - Using the wiki vocabulary
  - Live queries
  - Enhance/Extend the data by wiki users

- Let’s see how this combines with SPARQL endpoints
  - SMW’s new SPARQL implementation
  - Federated queries
  - Against different sources
  - Different vocabularies (with and without mapping)
Thank you!

Michael Erdmann
ontoprise GmbH
Karlsruhe, Germany
erdmann@ontoprise.de