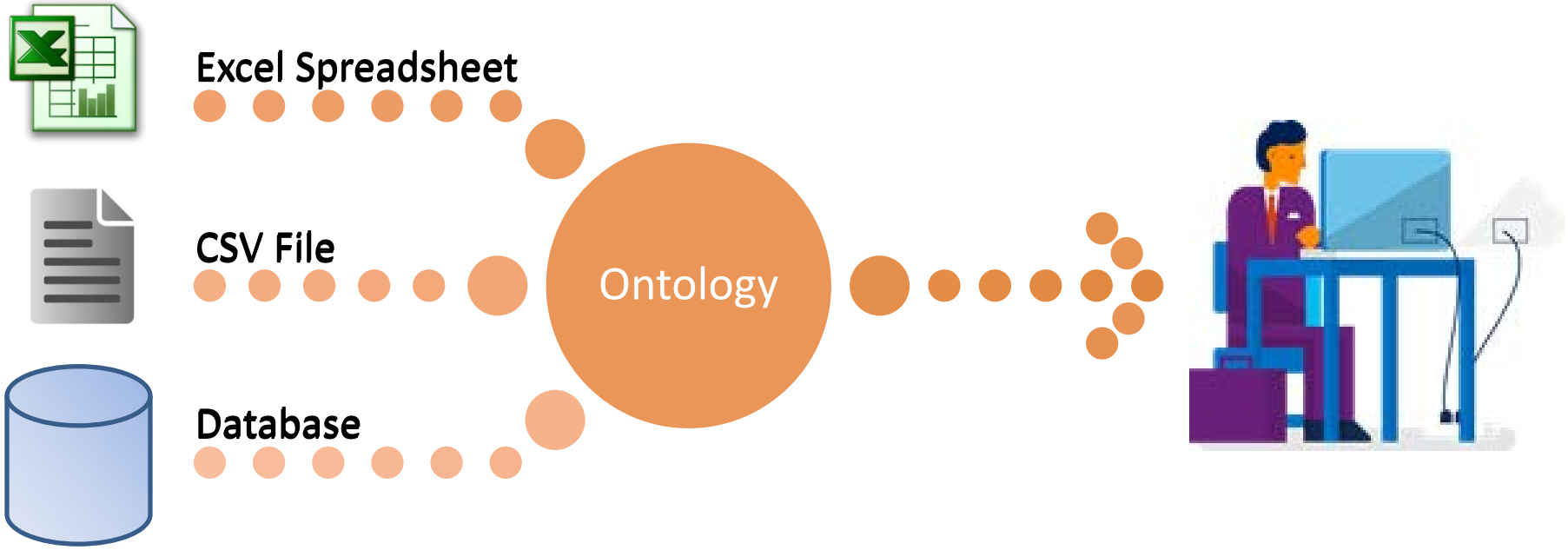
Data access and import from databases and other sources

Csongor Nyulas

Protégé Short Course
October 11, 2017

Scenario

- You have some data that you would like to use as part of your semantic application



Possible Solutions

- By type of the approach:
  - **Extract-Transform-Load (ETL)** – a.k.a. Data Import
  - **Ontology-Based Data Access (OBDA)**

- By type of the implementation:
  - Protégé plug-ins
  - Standalone frameworks or tools
  - Write custom code

Extract-Transform-Load (ETL)
- a.k.a. Data Import -

- Method: Convert the content of your data sources into OWL or RDF
  - Data residing in conventional sources will be instantiated as OWL statements or RDF triples
  - Changes in the original sources will not be (automatically) reflected in your semantic queries

The ETL Approach (1)

APPLICATION LAYER

ETL processing

Ontology

SEMANTIC LAYER

DATA SOURCE

The ETL Approach (2)

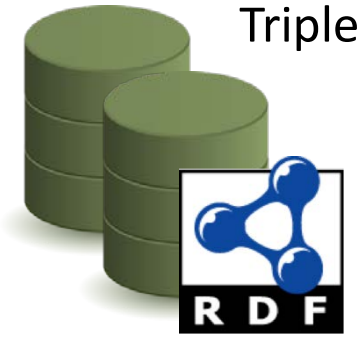
APPLICATION LAYER

Ontology

SEMANTIC LAYER

DATA SOURCE

ETL pre-processing



Triplestore

DATA LAYER

Ontology-based Data Access (OBDA)

- A method that allows the creation of a conceptual view over an existing data asset by mapping the ontology concepts to your data sources and hiding the details of your database schema and the complexity of DB query
- Transparent data access that focuses on the use of common terminology for retrieving data/information.
  - Semantic queries will be translated to SQL queries on the fly, and the results of the SQL queries will be translated back to RDF or OWL statements
- Non-intrusive solution that sits on top of your current data asset.
  - Data will continue to live in the database

The OBDA Approach

APPLICATION LAYER

OBDA Model

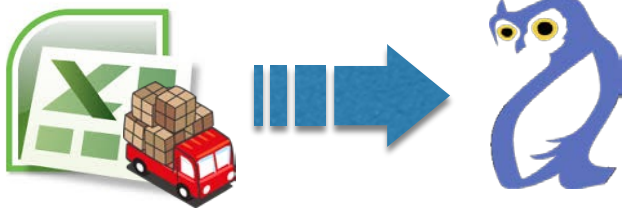
Ontology

SEMANTIC LAYER

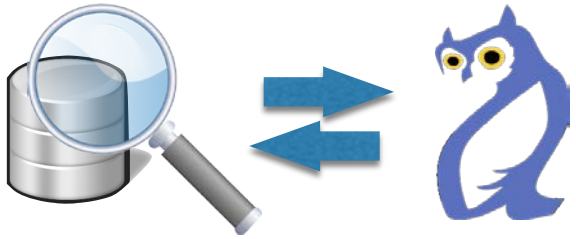
DATA LAYER

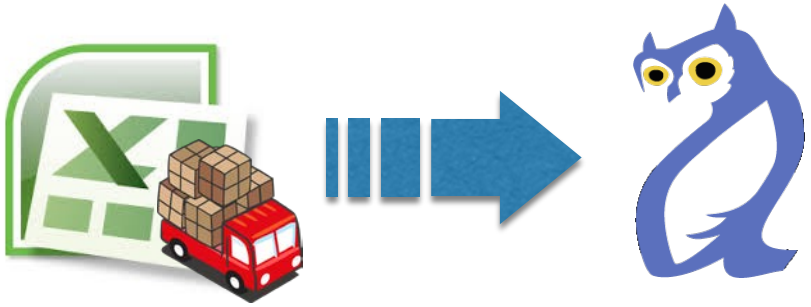
DATA ASSET

ETL
(Cellfie Plugin)



OBDA
(Ontop Plugin)





ETL

Import from spreadsheets with Cellfie

Cellfie

- Protégé plug-in bundled with distribution since version 5.0
- Based on earlier work on the Protégé 3.x plug-in called MappingMaster
- Uses the MappingMaster DSL
- Available at: https://github.com/protegeproject/cellfie-plugin/

DEMO

OBDA

Map and query data using ontologies with -ontop-

- -ontop- is a platform to query relational database over OWL ontology using SPARQL (instead of using SQL).
- Developed by Free University of Bozen-Bolzano in Italy.
- Supports several RDBMS: MySQL, PostgreSQL, H2, SQL Server, Oracle, IBM DB2.
- Distributed under open-source license http://ontop.inf.unibz.it/

DEMO

Other Solutions

- WebProtégé
  - Upload CSV
- Protégé Desktop (v. 4.x or 5.x):
  - **Ontop**, **Cellfie**
- Protégé 3.x:
  - DataMaster, MappingMaster, ProtegeScriptTab
- Frameworks:
  - D2RQ, Ontop
- Standalone tools:
  - OWLPopulous, ROBOT
- Programmatic:
  - Tawny-OWL
- RDF based tools:
  - Tarql, Convert2RDF, RDF123,
  - and many more http://www.w3.org/wiki/ConverterToRdf