Importing Data into Protégé-OWL

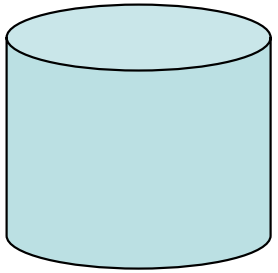
Martin O'Connor
Stanford Center for Biomedical Informatics Research,
Stanford University

Importing Data into Protégé-OWL

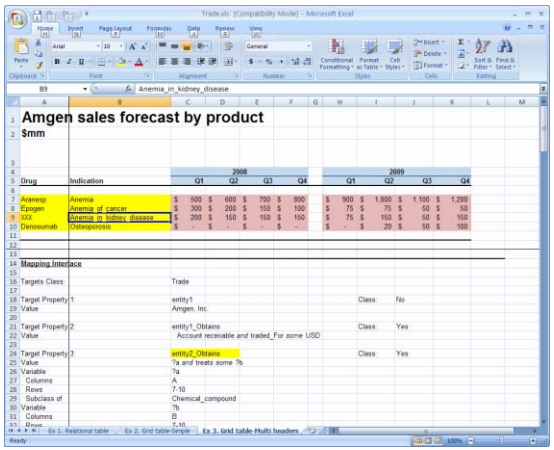
- Most data are not stored in the form of ontology
- Goal: import and reason with data from external sources
- Several Protégé-OWL Plug-ins to support batch importation from relational database, spreadsheets, XML documents
- Ongoing research: dynamic importing

Importation Functionality

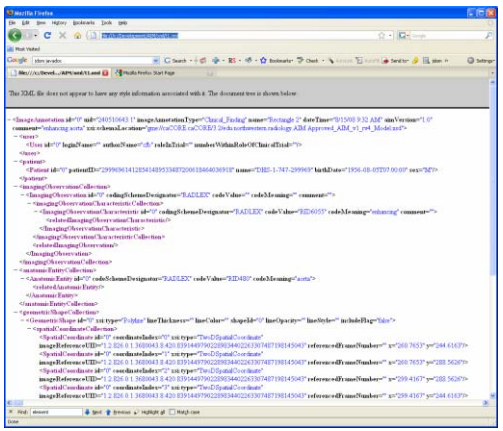
Databases



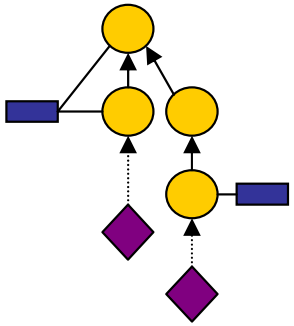
Spreadsheets



XML



OWL Ontology



Classes
Properties
Instances
Data Values

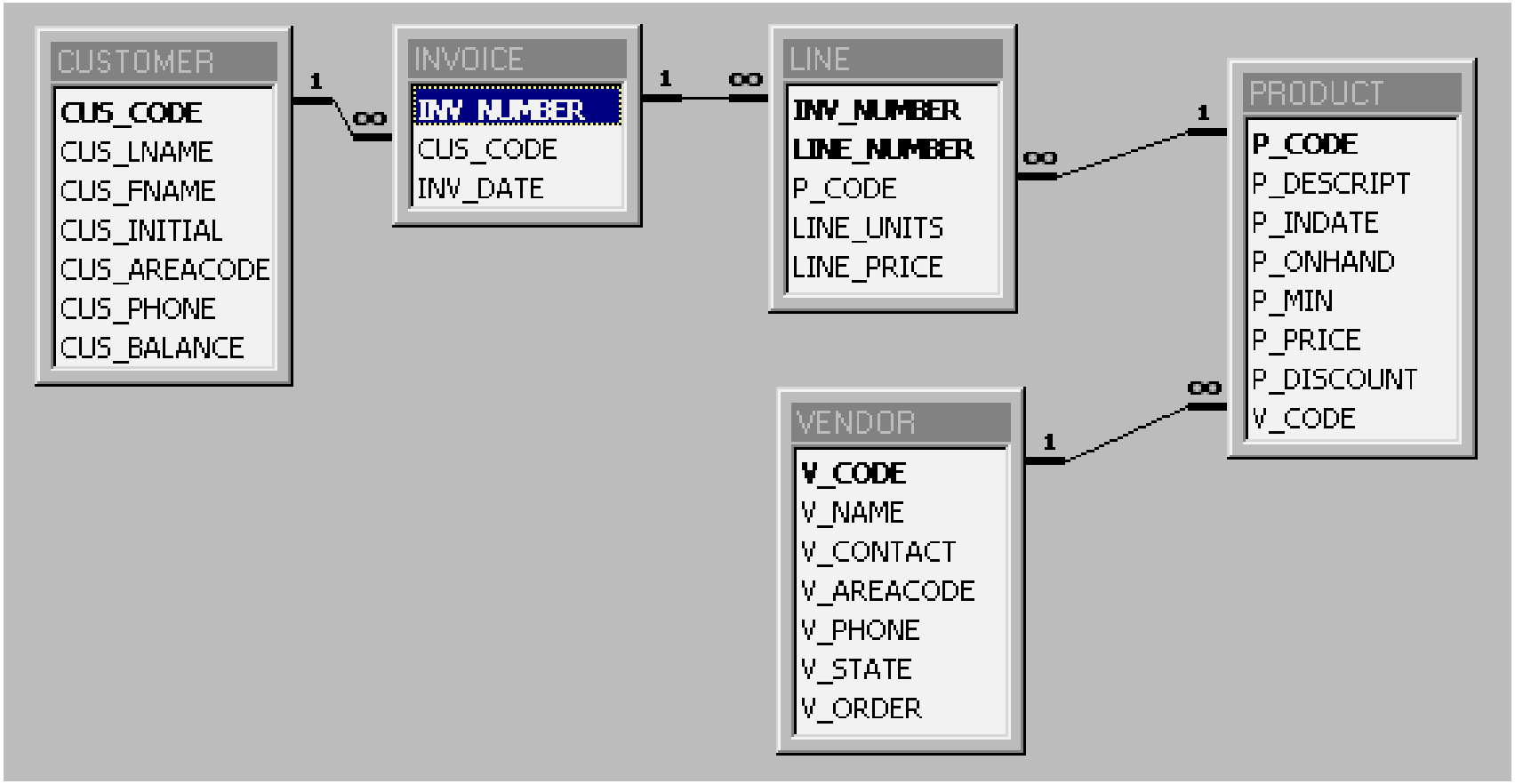
Protégé-OWL
Batch Importing Plug-ins

- DataMaster: relational importing
- SpreadSheetMaster: spreadsheet importing
- XMLMaster: XML importing

DataMaster Plug-in

- Imports relational schema and/or data
- Two import options:
  - Schema import only:
    - Schema as OWL classes
    - Schema as instances of Relational.OWL classes
  - Schema and content importation
- Supports any relational database with JDBC driver
- Java API for embedding in applications

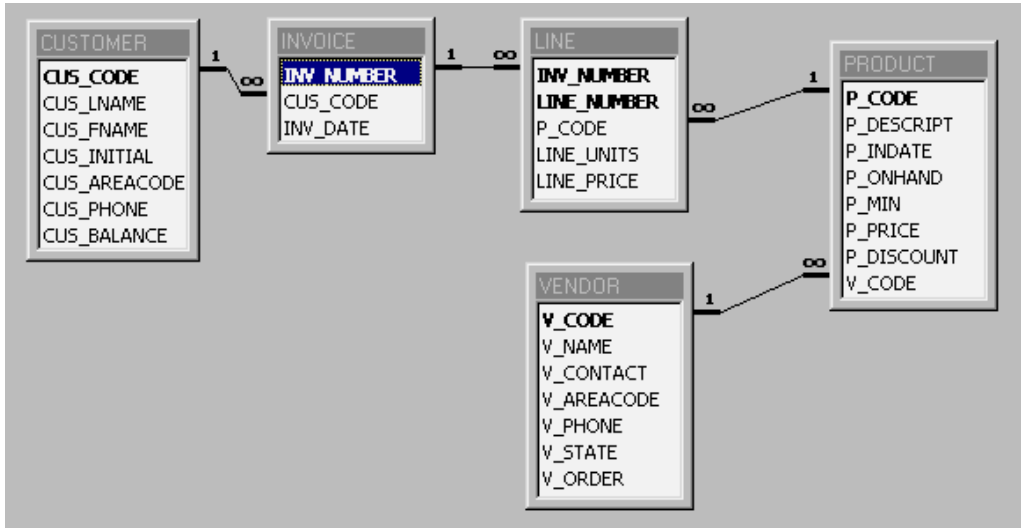
Example Schema



Representing Schema as OWL Classes

Namespace Prefixes

Prefix	Namespace
xsd	http://www.w3.org/2001/XMLSchema#
rdfs	http://www.w3.org/2000/01/rdf-schema#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
owl	http://www.w3.org/2002/07/owl#
dbs	http://www.dbs.cs.uni-duesseldorf.de/RDF/relational.owl#
db	http://biostorm.stanford.edu/db_table_classes?DSN=jdbc:mysql://localhost:3306/trade#



Asserted Hierarchy

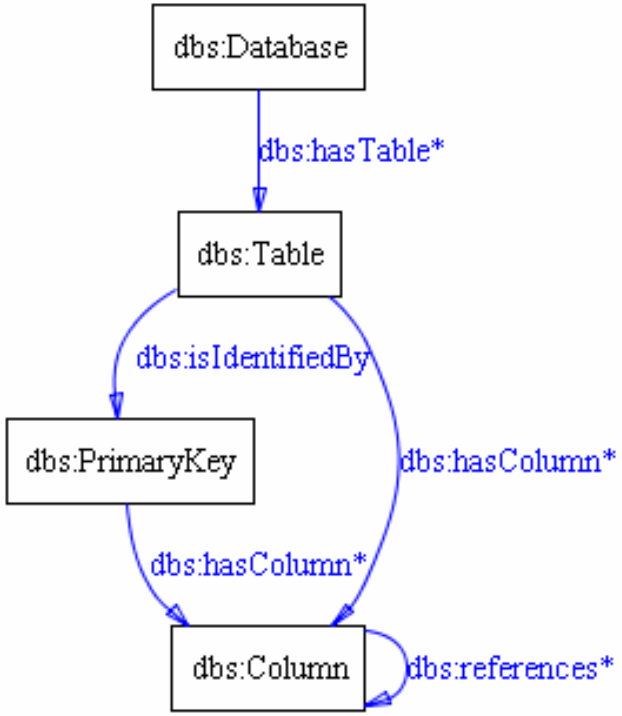
- owl:Thing
  - db:ForeignKey
  - db:CUSTOMER
  - db:INVOICE
  - db:LINE
  - db:PRODUCT
  - db:VENDOR

For Class: db:PRODUCT

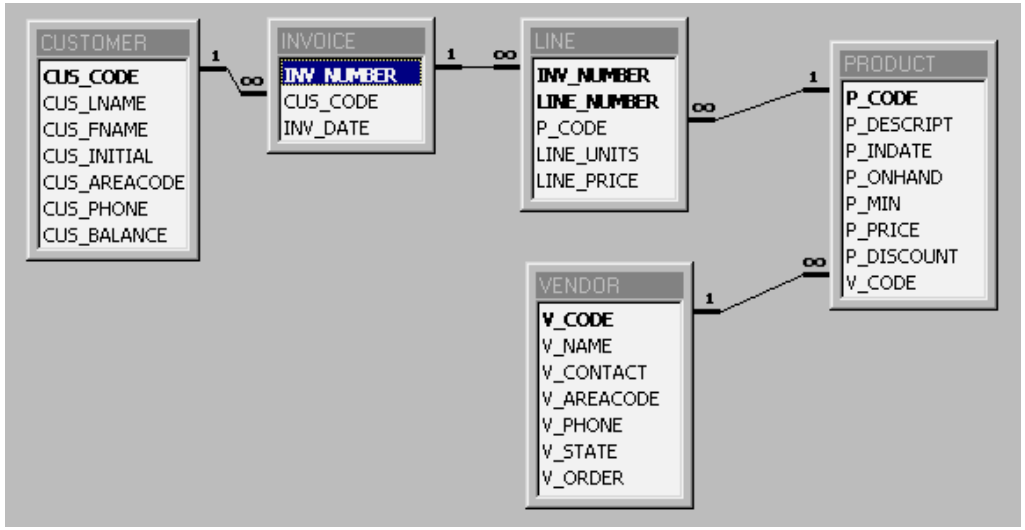
Property	
rdfs:comment	
db:hasForeignKeys	FK: PRODUC
db:hasPrimaryKeyFields	P_CODE
db:isBridgeTable	false

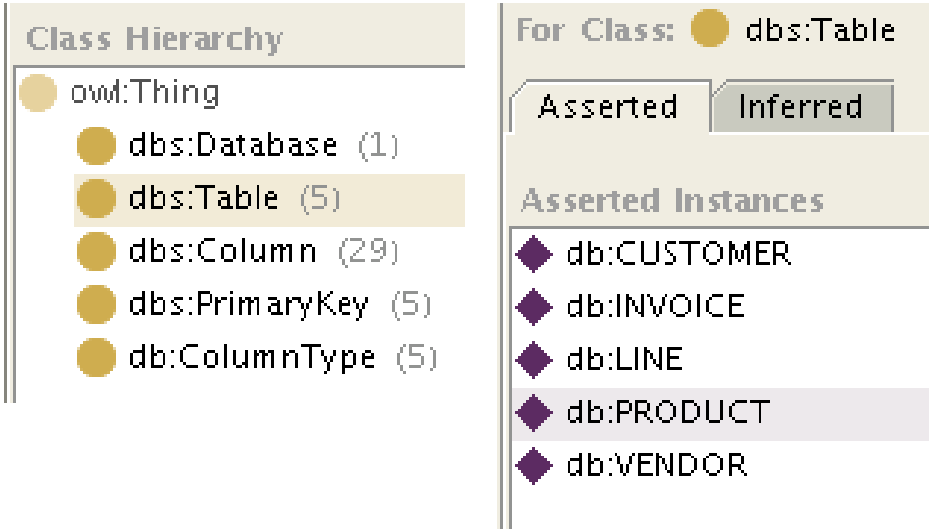
- db:PRODUCT.P_CODE (single int)
- db:PRODUCT.P_DESCRIPT (single string)
- db:PRODUCT.P_DISCOUNT (single float)
- db:PRODUCT.P_INDATE (single date)
- db:PRODUCT.P_MIN (single int)
- db:PRODUCT.P_ONHAND (single boolean)
- db:PRODUCT.P_PRICE (single float)
- db:PRODUCT.V_CODE (single int)

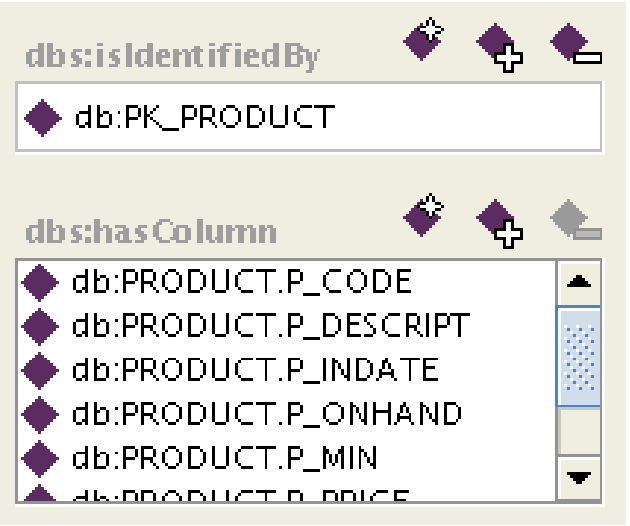
Representing Schema in Relational.OWL Ontology



Representing Schema in Relational.OWL Ontology



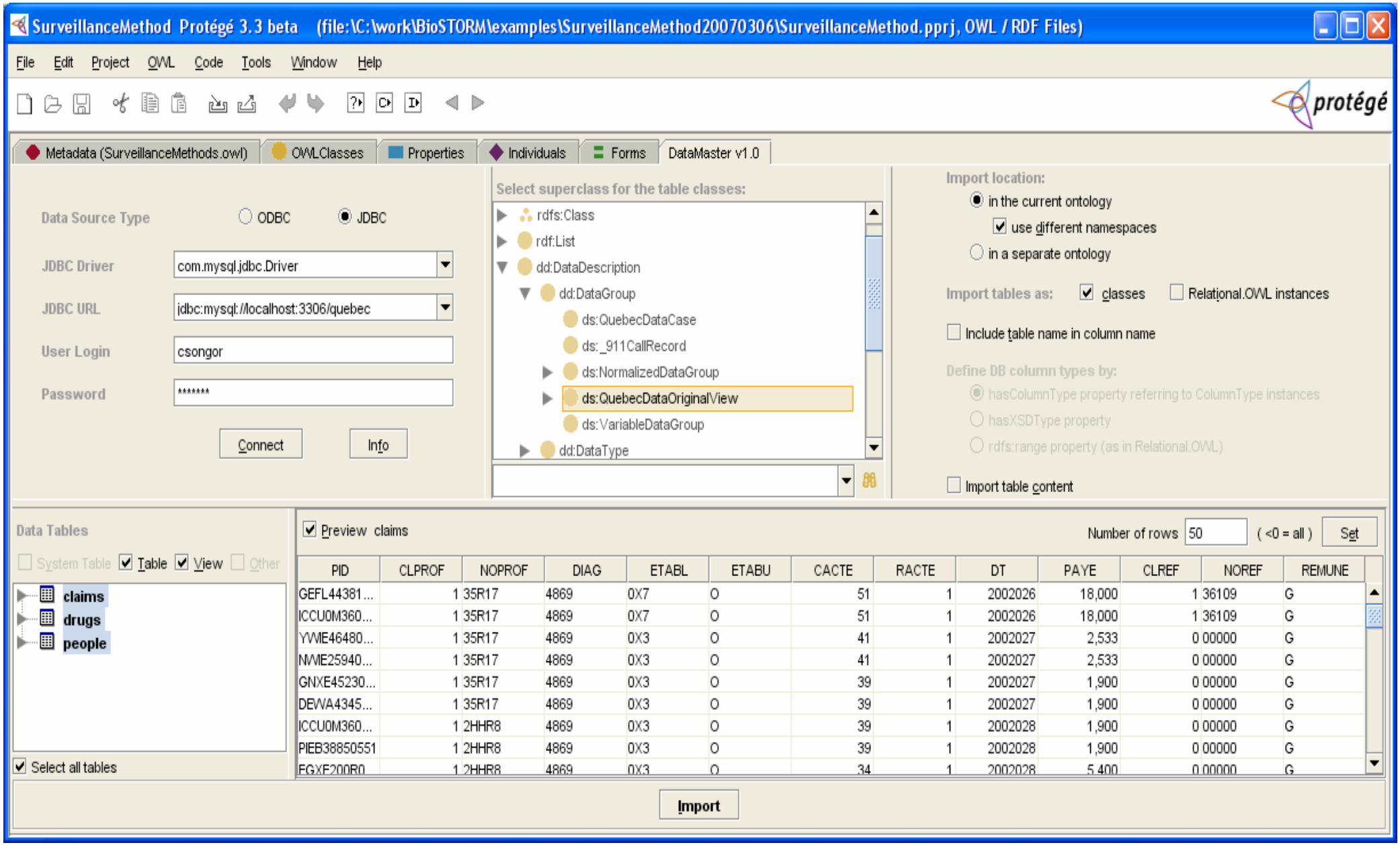




Importing Schema and Content

- Imports relational schema and data
- Two schema representations:
  - Schema as OWL classes
  - Schema as instances of Relational.OWL classes
- Data stored as instances of either OWL classes or Relational.OWL instances

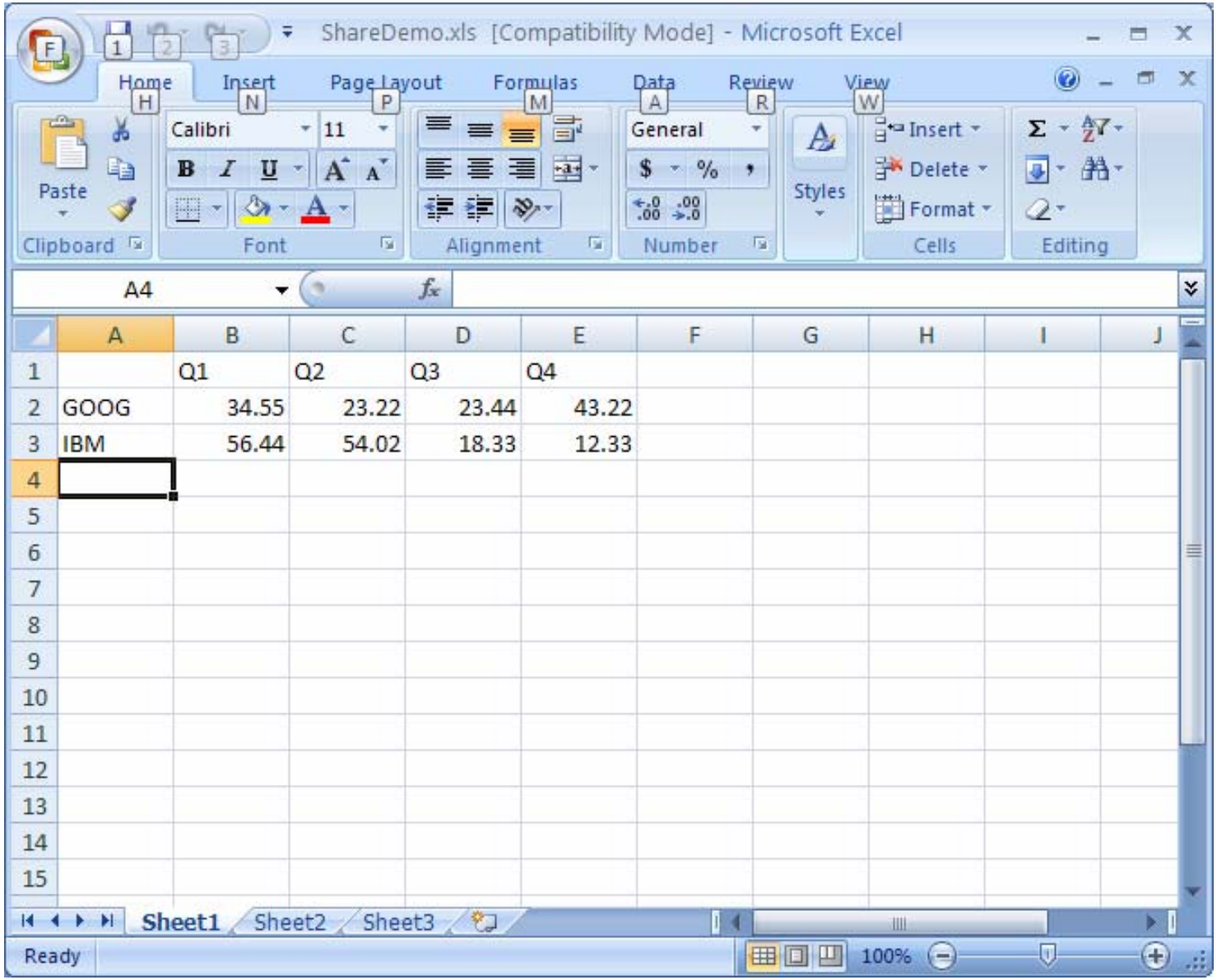
DataMaster Plug-in



SpreadSheetMaster Plug-in

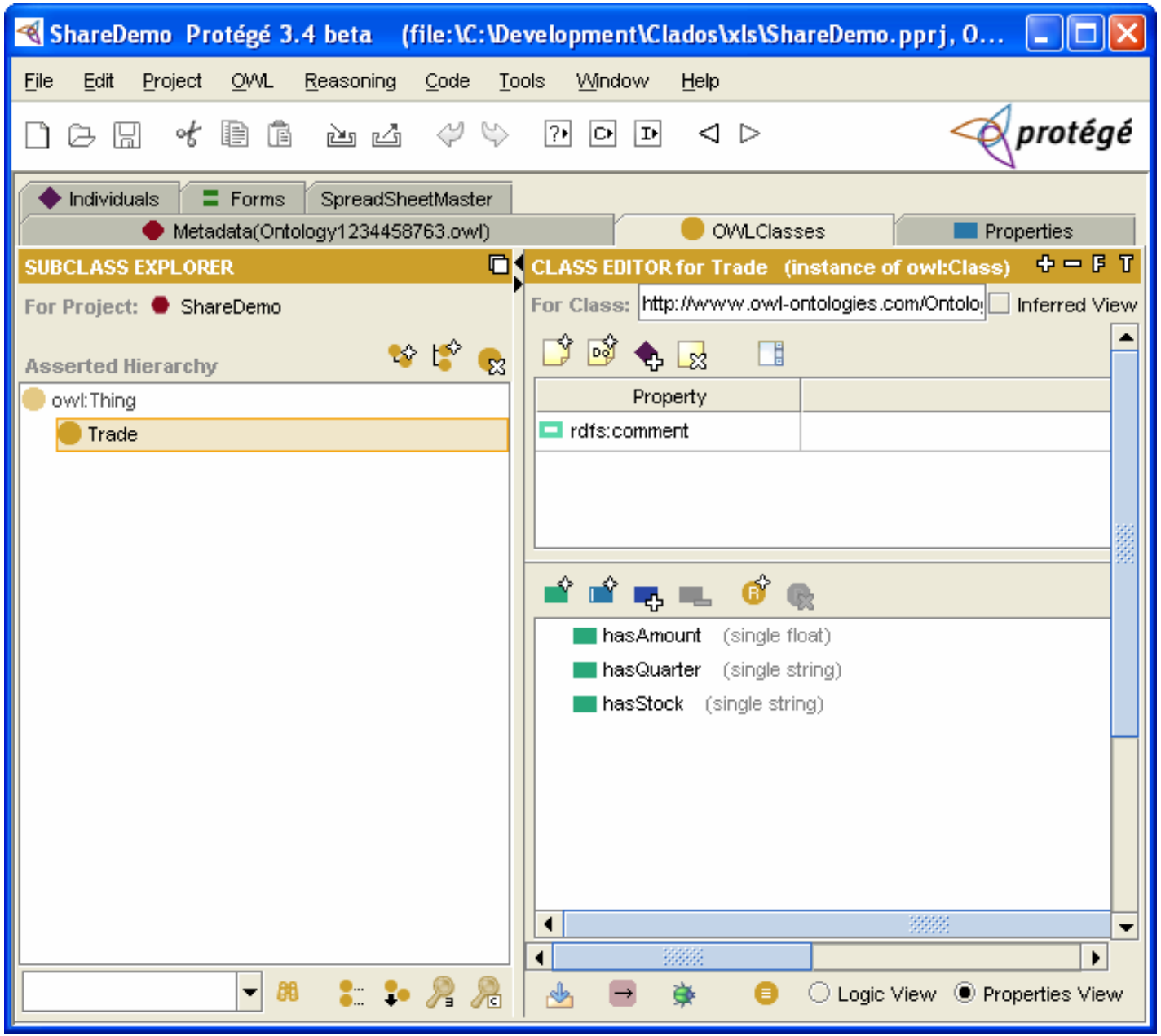
- Protégé-OWL Plug-in for importing spreadsheet content into OWL ontologies
- Can import arbitrary Excel or CSV files
- Also, supports definition of mapping templates to import files of a particular structure
- Mappings saved as an OWL ontology

Example Source Excel Spreadsheet

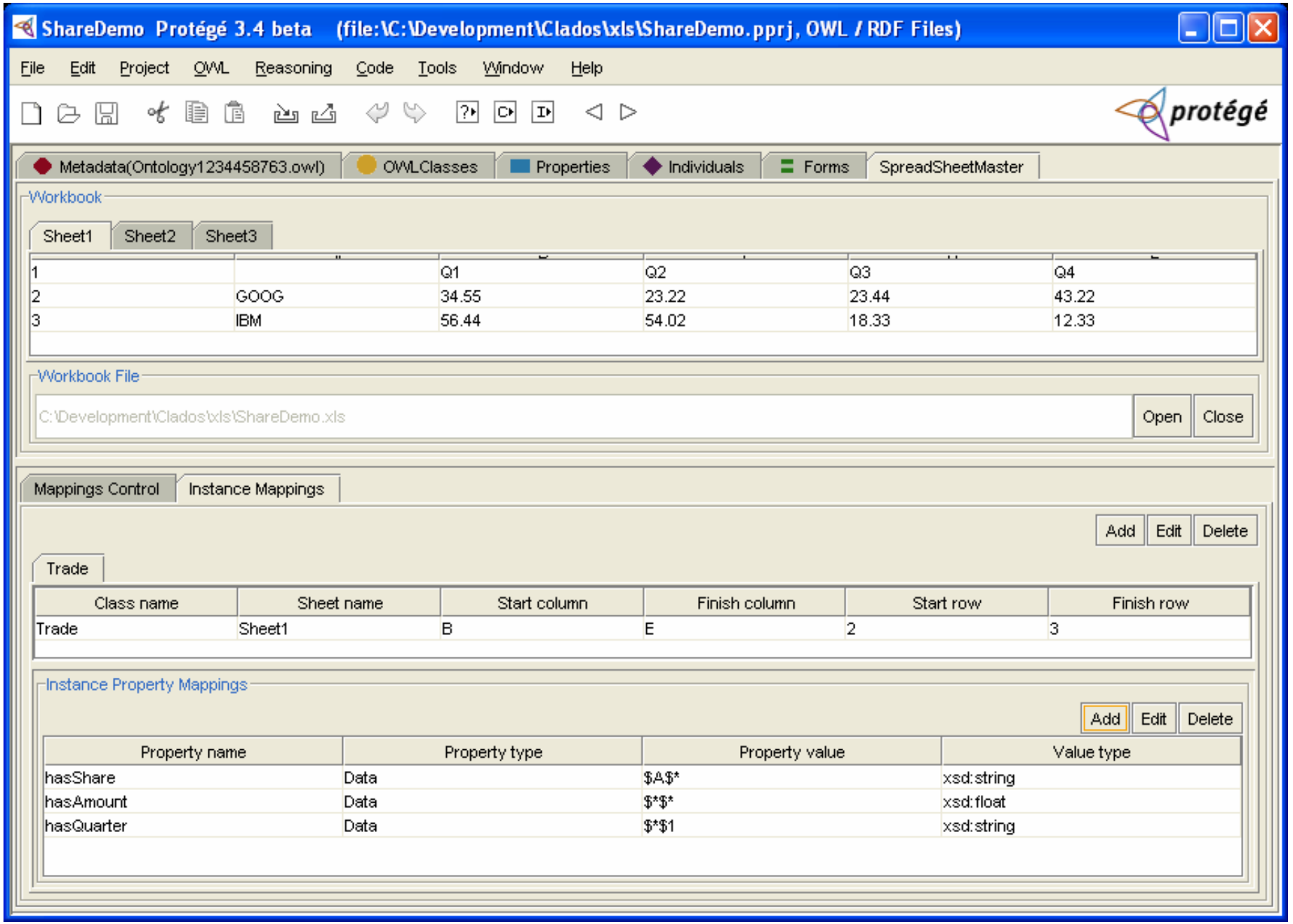


	A	B	C	D	E
1		Q1	Q2	Q3	Q4
2	GOOG	34.55	23.22	23.44	43.22
3	IBM	56.44	54.02	18.33	12.33
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					

Example Target Ontology



SpreadSheetMaster Screenshot



SpreadSheetMaster Features

- Can import content as classes, properties, instances, or data values
- Define complex mappings interactively with custom expression language
- Can save mappings and reuse them on documents with the same structure
- Java API for embedding in applications
- Available in a month

XMLMaster Plug-in

- Protégé-OWL Plug-in for importing XML documents into OWL ontologies
- Can import arbitrary XML documents
- Also supports definition of mappings templates to import XML documents of a particular type
- Mappings saved as an OWL ontology
- Java API for embedding in applications
- Available in a few months!

Dynamic Importing

- DataMaster, SpreadSheetMaster, and XMLMaster perform batch importation
- *Dynamically* importing content is desirable in some cases
- Entities are mapped to OWL axioms on demand
- Significant scalability advantages

Current Work: Dynamic Importing Plugin

- Supports mappings to import relational, spreadsheet and XML data *on demand* when executing SWRL rules
- Mappings saved as an OWL ontology
- Available later this year!

Conclusion

- DataMaster: available Protégé 3.4
- SpreadSheetMaster: available Protégé 3.4 end of July
- XMLMaster: available Protégé 3.4 in a few months
- Dynamic importing tools: available Protégé 4 towards end of year
- Other tools will be ported to Protégé 4