

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

Protégé-OWL

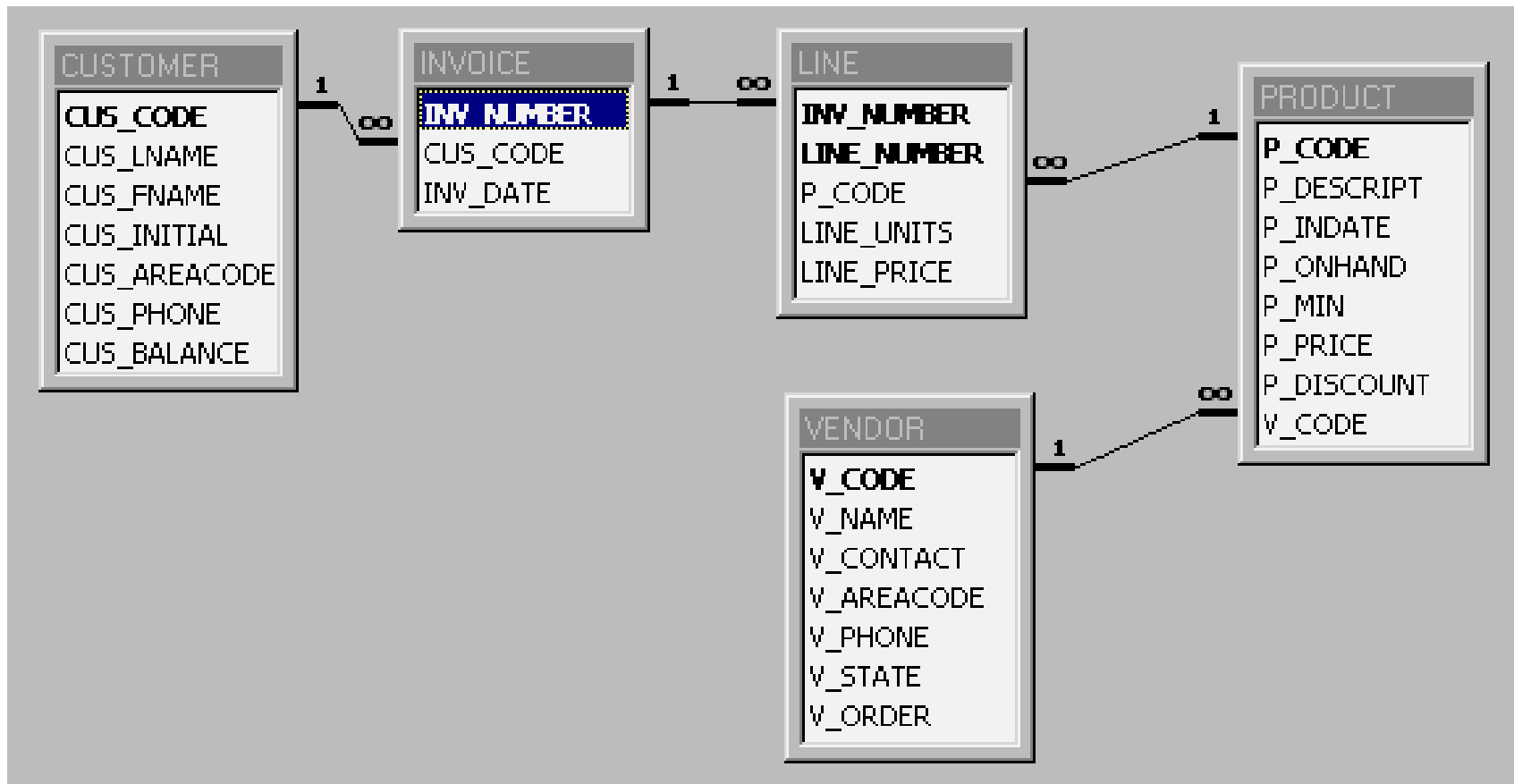
Batch Importing Plug-ins

- 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

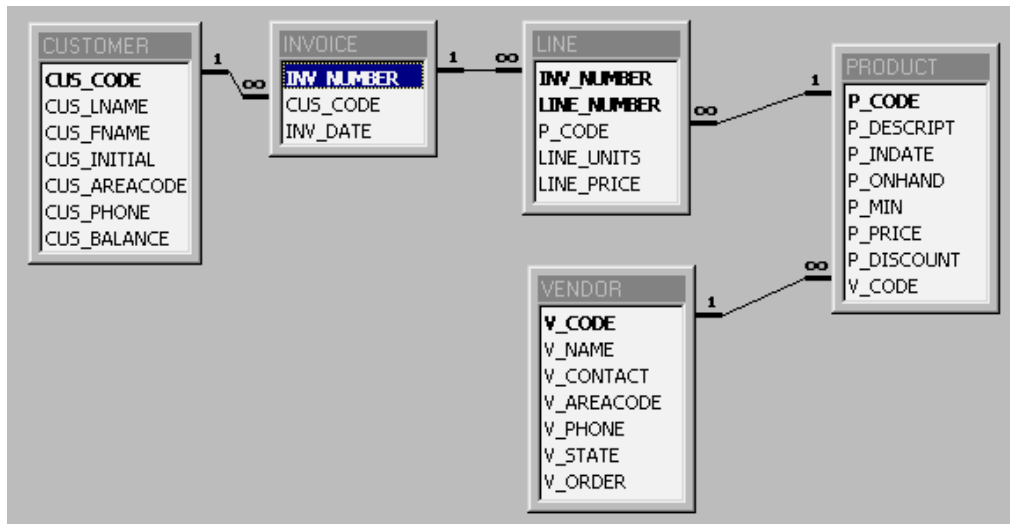
Example Schema



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#
db	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

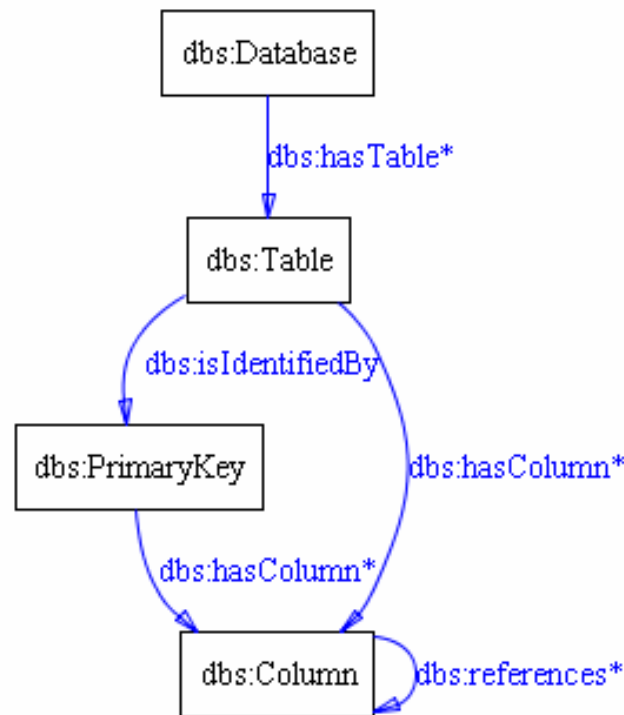
- owl:Thing
 - db:ForeignKey
 - db:CUSTOMER
 - db:INVOICE
 - db:LINE
 - db:PRODUCT**
 - db:VENDOR

For Class: db:PRODUCT

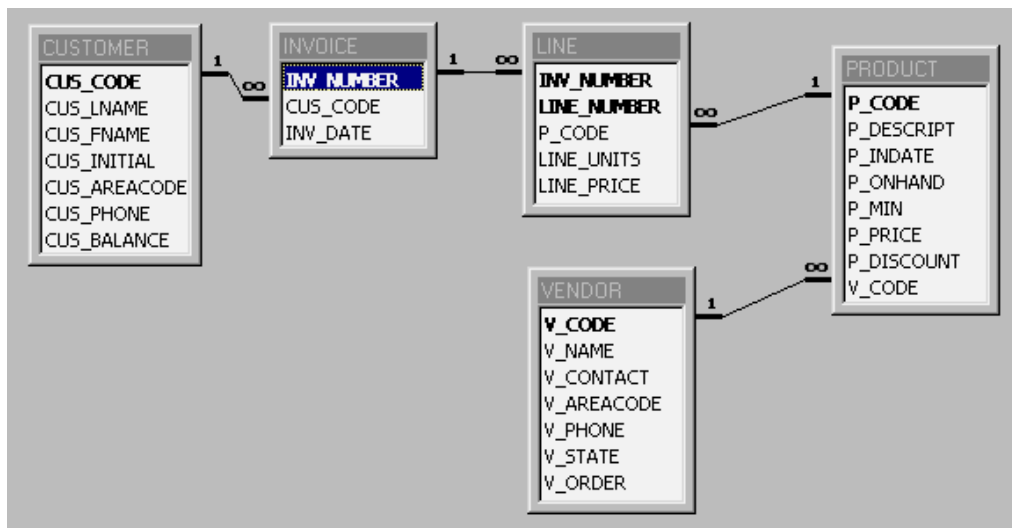
Property	
rdfs:comment	
db:hasForeignKeys	FK: PRODUC
db:hasPrimaryKeyFields	P_CODE
db:isBridgeTable	false

- 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



Class Hierarchy

- owl:Thing
 - db:Database (1)
 - db:Table (5)
 - db:Column (29)
 - db:PrimaryKey (5)
 - db:ColumnType (5)

For Class: ● db:Table

Asserted Inferred

Asserted Instances

- db:CUSTOMER
- db:INVOICE
- db:LINE
- db:PRODUCT
- db:VENDOR

db:i sIdentifiedBy

- db:PK_PRODUCT

db:s hasColumn

- db:PRODUCT.P_CODE
- db:PRODUCT.P_DESCRIPT
- db:PRODUCT.P_INDATE
- db:PRODUCT.P_ONHAND
- db:PRODUCT.P_MIN
- db:PRODUCT.P_PRICE

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

The screenshot shows the Protegé 3.3 beta interface with the DataMaster v1.0 plug-in active. The window title is "SurveillanceMethod Protégé 3.3 beta (file:IC:\work\BioSTORM\examples\SurveillanceMethod20070306\SurveillanceMethod.pprj, OWL / RDF Files)". The menu bar includes File, Edit, Project, OWL, Code, Tools, Window, and Help. The toolbar contains various icons for file operations and navigation. The main workspace is divided into several panes:

- Metadata (SurveillanceMethods.owl):** Shows configuration for the data source type (JDBC selected), driver (com.mysql.jdbc.Driver), URL (jdbc:mysql://localhost:3306/quebec), user login (csongor), and password (masked).
- OWLClasses:** A tree view showing the class hierarchy. The selected superclass for the table classes is "ds:QuebecDataOriginalView".
- Import location:** Options for where to import the data (in the current ontology, use different namespaces, or in a separate ontology).
- Import tables as:** Options for how to import the data (as classes or Relational.OWL instances).
- Define DB column types by:** Options for how to define the column types (hasColumnType property, hasXSDType property, or rdfs:range property).
- Import table content:** A checkbox to import the table content.

The bottom pane shows the "Data Tables" section with a "Preview claims" checkbox checked. The "Number of rows" is set to 50. The table below shows the preview of the data:

PID	CLPROF	NOPROF	DIAG	ETABL	ETABU	CACTE	RACTE	DT	PAYE	CLREF	NOREF	REMUNE
GEFL44381...	1 35R17	4869	0X7	0	51	1	2002026	18,000	1 36109	G		
ICCU0M360...	1 35R17	4869	0X7	0	51	1	2002026	18,000	1 36109	G		
YWME46480...	1 35R17	4869	0X3	0	41	1	2002027	2,533	0 00000	G		
NWME25940...	1 35R17	4869	0X3	0	41	1	2002027	2,533	0 00000	G		
GNXE45230...	1 35R17	4869	0X3	0	39	1	2002027	1,900	0 00000	G		
DEWA4345...	1 35R17	4869	0X3	0	39	1	2002027	1,900	0 00000	G		
ICCU0M360...	1 2HHR8	4869	0X3	0	39	1	2002028	1,900	0 00000	G		
PIEB38850551	1 2HHR8	4869	0X3	0	39	1	2002028	1,900	0 00000	G		
EGXF200R0	1 2HHR8	4869	0X3	0	34	1	2002028	5,400	0 00000	G		

An "Import" button is located at the bottom center of the interface.

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

The screenshot shows the Microsoft Excel interface with the following data in the spreadsheet:

	A	B	C	D	E	F	G	H	I	J
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

ShareDemo Protégé 3.4 beta (file:VC:\Development\Clados\xls\ShareDemo.pprj, O...

File Edit Project OWL Reasoning Code Tools Window Help

Individuals Forms SpreadsheetMaster

Metadata(Ontology1234458763.owl) OWLClasses Properties

SUBCLASS EXPLORER

For Project: ShareDemo

Asserted Hierarchy

- owl:Thing
 - Trade

CLASS EDITOR for Trade (instance of owl:Class)

For Class: <http://www.owl-ontologies.com/Ontology1234458763.owl#Trade> Inferred View

Property	
rdfs:comment	

- hasAmount (single float)
- hasQuarter (single string)
- hasStock (single string)

Logic View Properties View

SpreadSheetMaster Screenshot

ShareDemo Protégé 3.4 beta (file:IC:\Development\Clados\xls\ShareDemo.pprj, OWL / RDF Files)

File Edit Project OWL Reasoning Code Tools Window Help

Metadata(Ontology1234458763.owl) OWLClasses Properties Individuals Forms SpreadSheetMaster

Workbook

Sheet1 Sheet2 Sheet3

		Q1	Q2	Q3	Q4
1					
2	GOOG	34.55	23.22	23.44	43.22
3	IBM	56.44	54.02	18.33	12.33

Workbook File

C:\Development\Clados\xls\ShareDemo.xls Open Close

Mappings Control Instance Mappings

Trade

Class name	Sheet name	Start column	Finish column	Start row	Finish row
Trade	Sheet1	B	E	2	3

Instance Property Mappings

Property name	Property type	Property value	Value type
hasShare	Data	\$\$* ^A	xsd:string
hasAmount	Data	\$\$*	xsd:float
hasQuarter	Data	\$\$* ¹	xsd:string

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