Capsenta

The Smart Data Company™

Integrating Data using Graphs and Semantics

Juan F. Sequeda

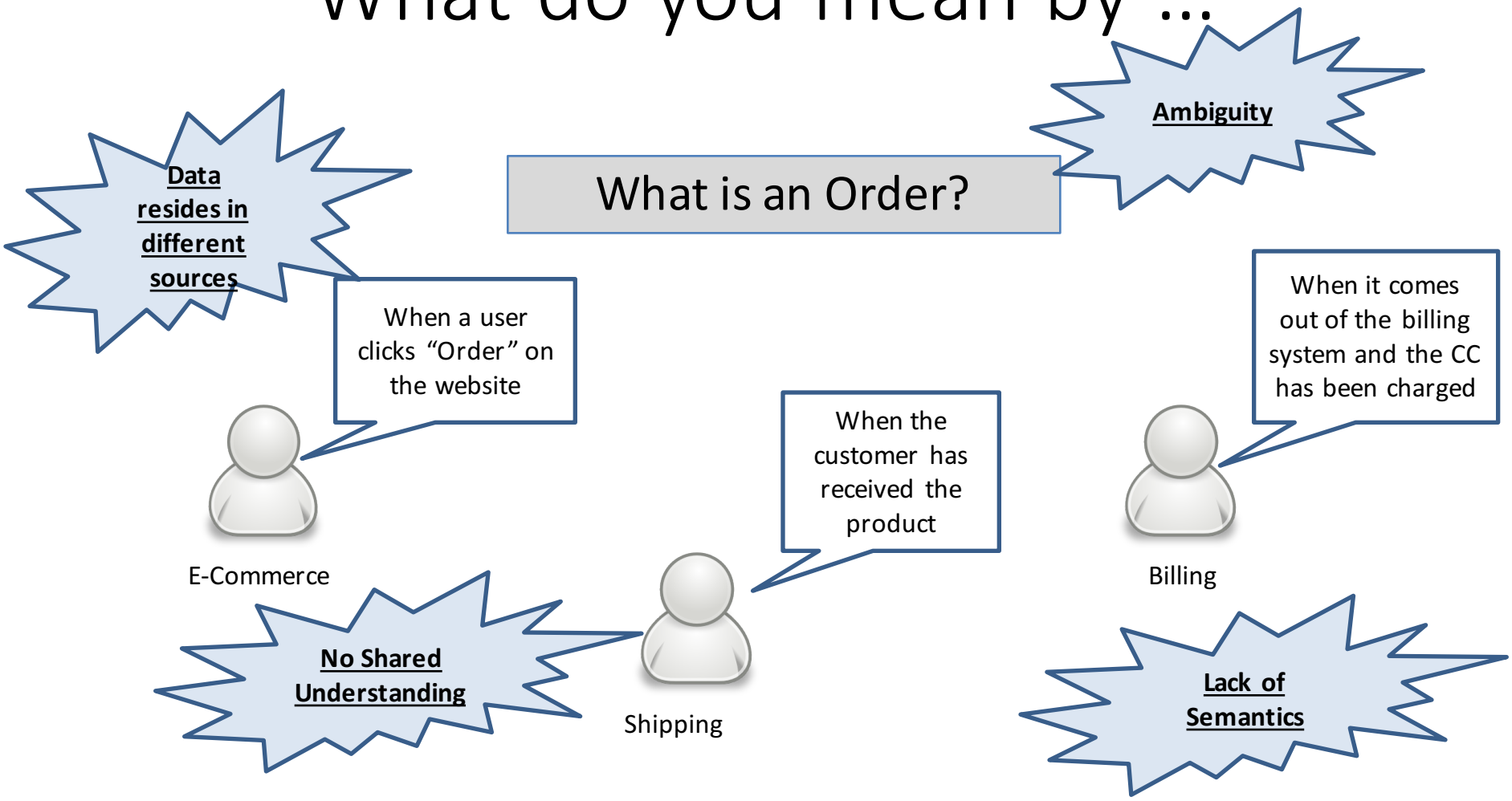
juan@capsenta.com

Gartner Cool Vendor 2016™

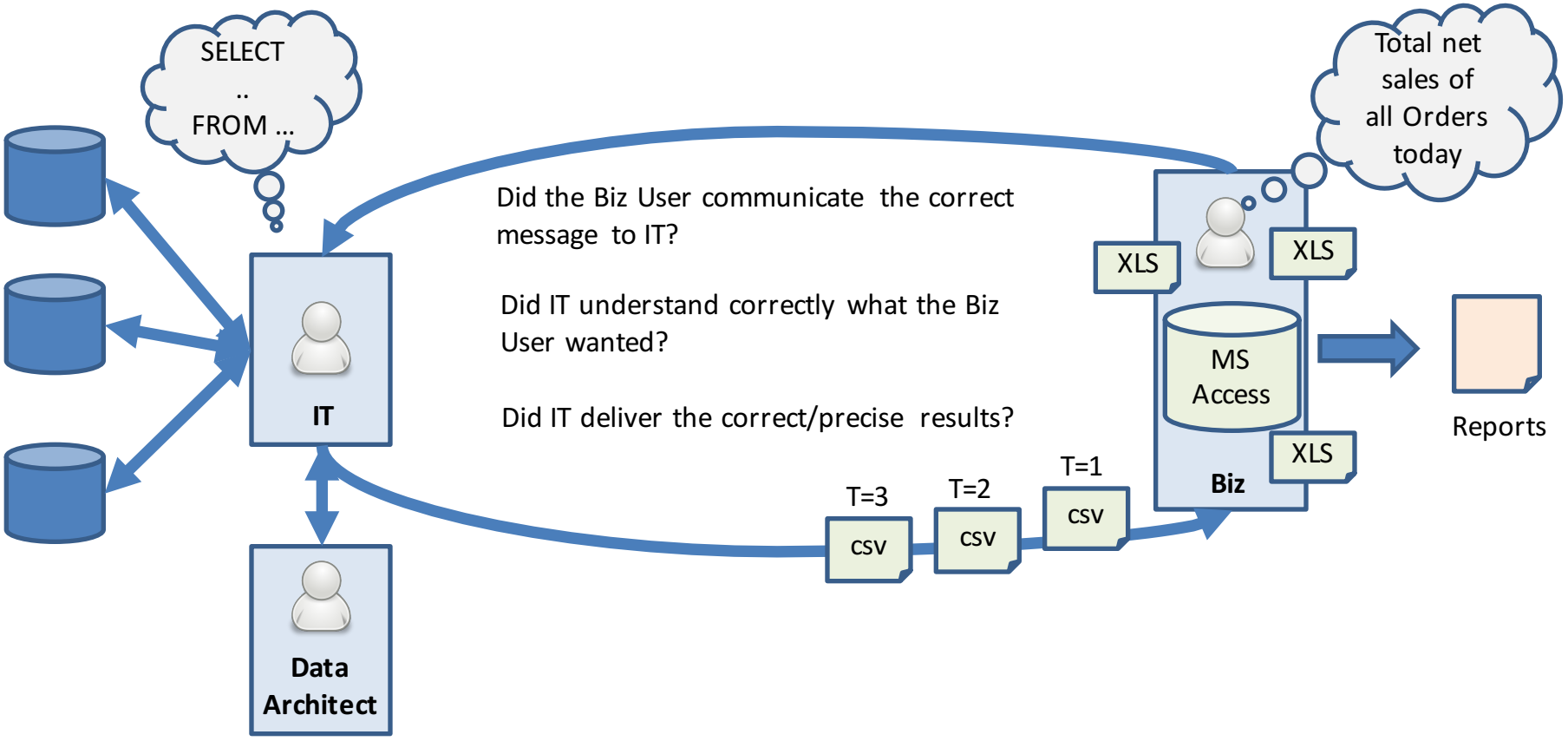

Total net sales of all Orders today
IT
Biz
Reports

What do you mean by …

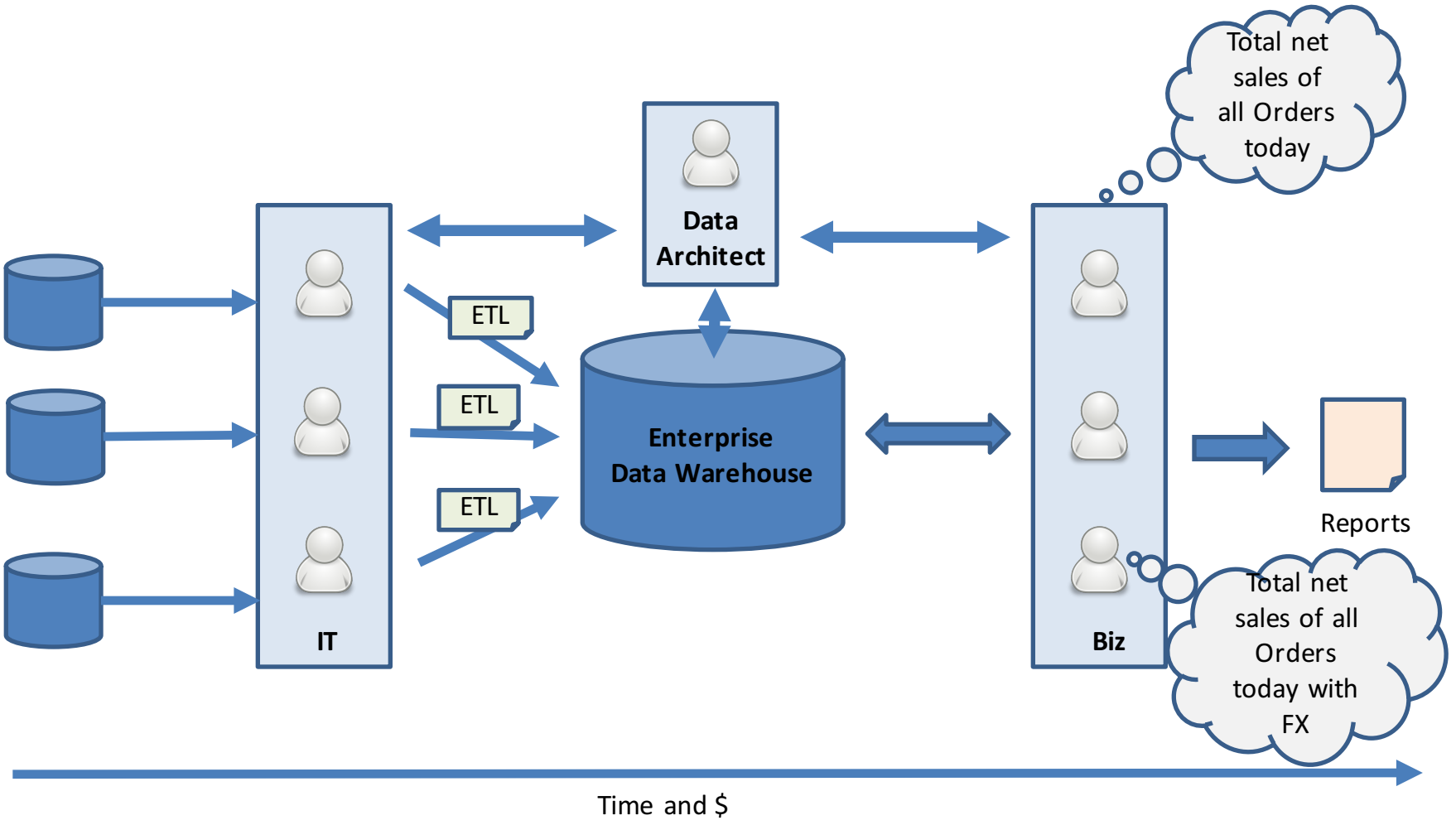
What do you mean by …



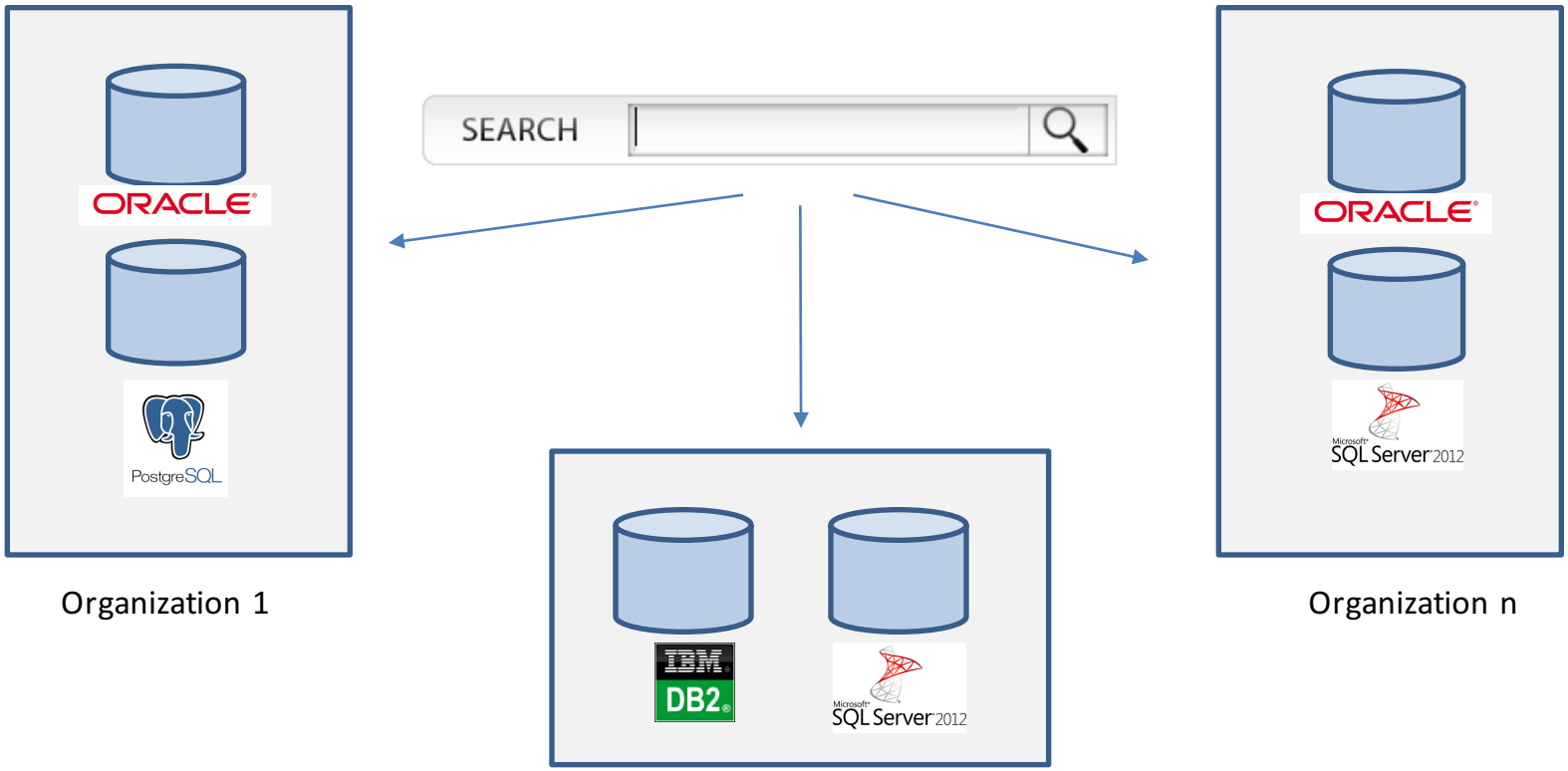
Status Quo 1


SELECT
..
FROM ...
IT
Data Architect
Did the Biz User communicate the correct message to IT?
Did IT understand correctly what the Biz User wanted?
Did IT deliver the correct/precise results?
T=3
csv
T=2
csv
T=1
csv
Total net sales of all Orders today
XLS
XLS
MS Access
XLS
Biz
Reports

Status Quo 2


Total net sales of all Orders today
Data Architect
ETL
ETL
ETL
Enterprise Data Warehouse
IT
Biz
Reports
Total net sales of all Orders today with FX
Time and $

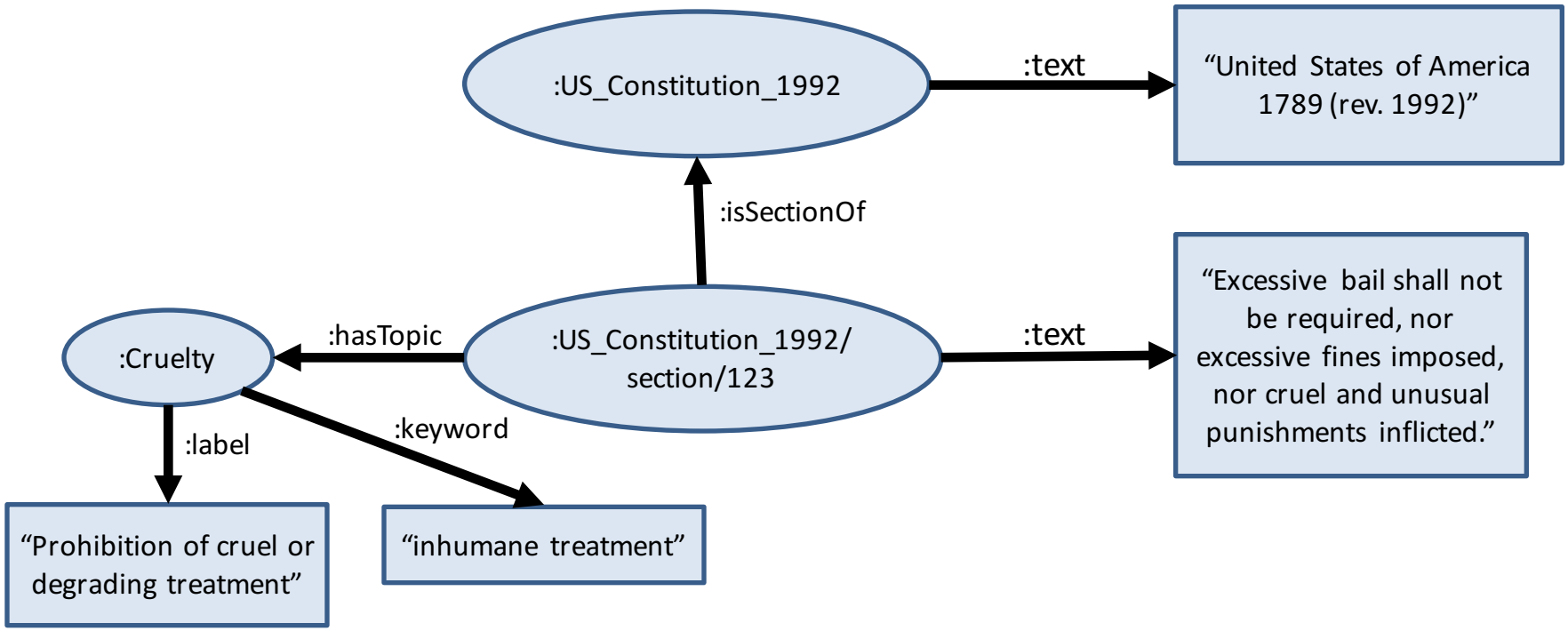
Cross Organizational Data Integration



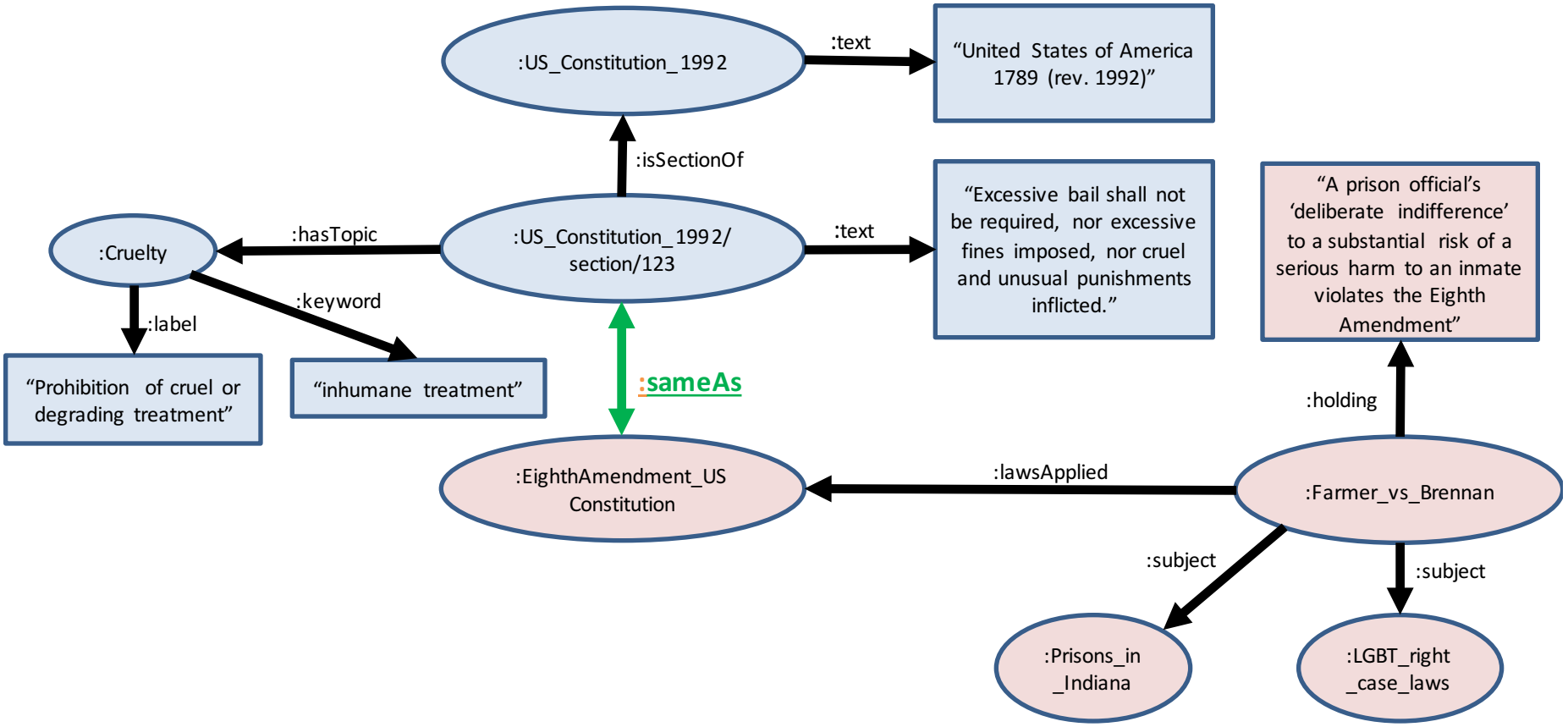

English
ELECTION OFFICIALS CHECK VOTER ROLLS FOR KENYA'S CONSTITUTIONAL REFERENDUM (2010)
CONSTITUTE
THE WORLD'S CONSTITUTIONS TO READ, SEARCH, AND COMPARE
EXPLORE
CONSTITUTIONS
CONSTITUTE ONEBOX ON GOOGLE SEARCH
GOOGLE SEARCHES NOW HARNESS CONSTITUTE DATA TO PROVIDE FULL CONSTITUTIONS RIGHT ON YOUR SEARCH PAGE. READ MORE HERE.
LEARN MORE ABOUT THE PROJECT AND THE PEOPLE BEHIND IT.
SHARE

GRAPHS ARE COOL!

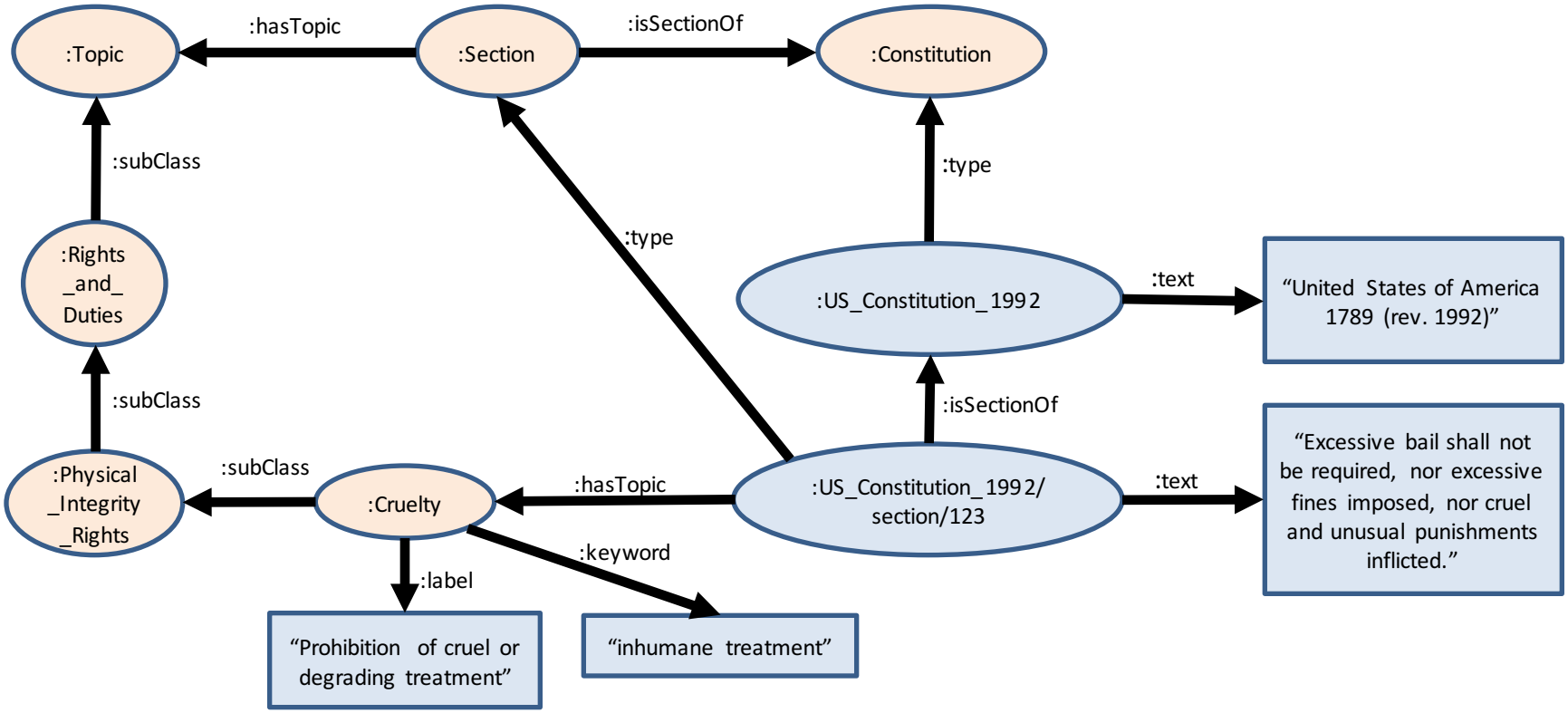
Flexible


:US_Constitution_1992
:text
"United States of America 1789 (rev. 1992)"
:isSectionOf
:Cruelty
:hasTopic
:US_Constitution_1992/ section/123
:text
"Excessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted."
:label
:keyword
"Prohibition of cruel or degrading treatment"
"inhumane treatment"

Integration



Data and Metadata are One



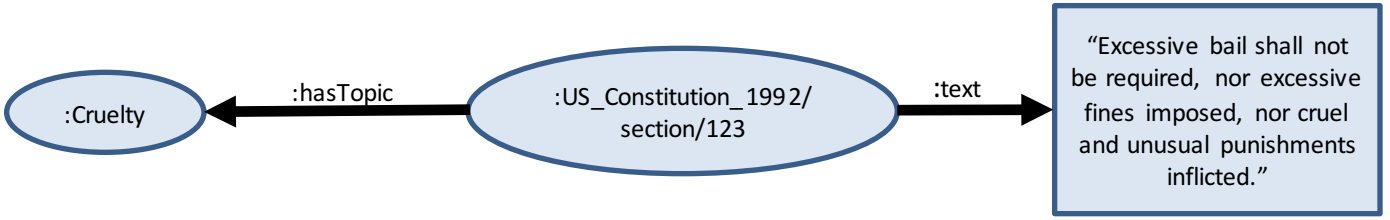
Common denominator

XML

```
<constitution id="US_Constitution_1992">
      <section id="US_Constitution_1992/section/123">
            <text>Excessive bail shall ...</text>
      </section>
      <topic>Cruelty</topic>
</constitution>
```

Text

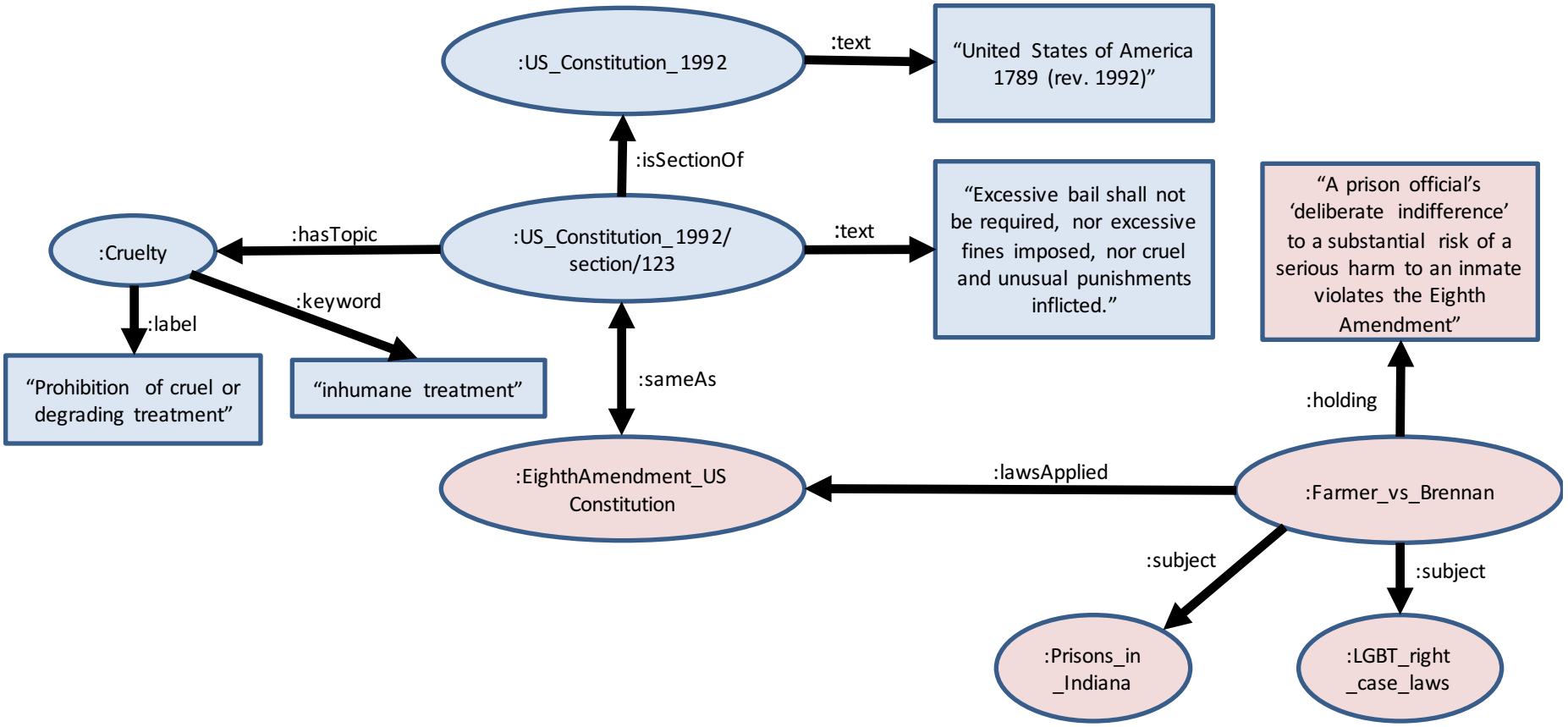
"Excessive bail shall not be required, nor excessive fines imposed, nor **cruel** and unusual punishments inflicted."



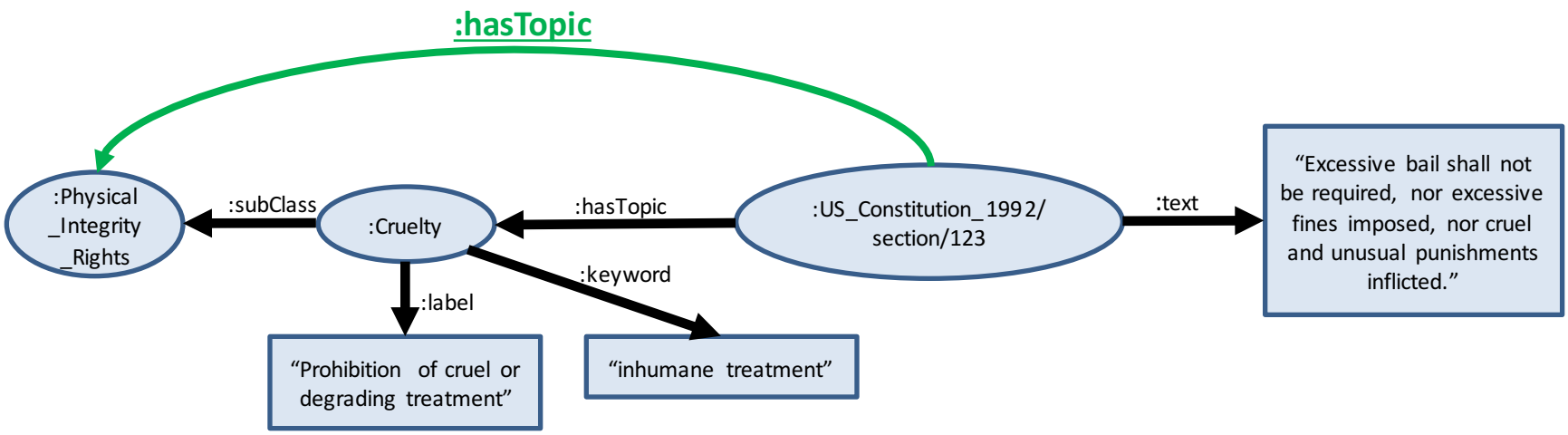
Tabular

id	text	topic
123	Excessive bail shall…	Cruelty

Traversal, Navigation, Reachability



Semantics


:hasTopic
:Physical _Integrity _Rights
:subClass
:Cruelty
:hasTopic
:US_Constitution_1992/ section/123
:text
"Excessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted."
:label
:keyword
"Prohibition of cruel or degrading treatment"
"inhumane treatment"

(Summary) Why are Graphs Cool?

- Flexible
- Integration
- Data and Metadata are one
- Common Denominator
- Traversal, Navigation, Reachability
- Semantics

Survey of Graph Database Models

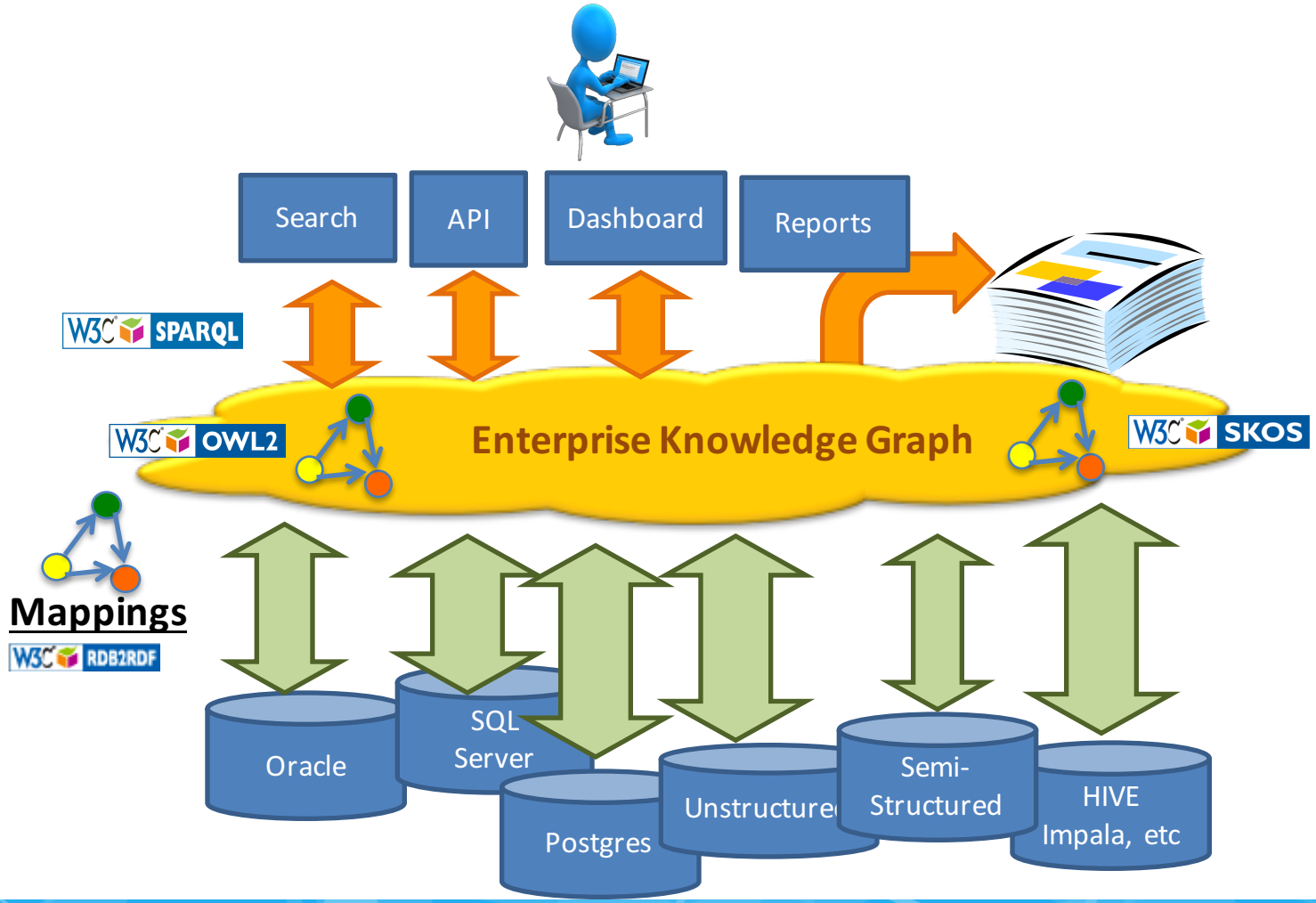
RENZO ANGLES and CLAUDIO GUTIERREZ

Universidad de Chile

Graph database models can be defined as those in which data structures for the schema and instances are modeled as graphs or generalizations of them, and data manipulation is expressed by graph-oriented operations and type constructors. These models took off in the eighties and early nineties alongside object-oriented models. Their influence gradually died out with the emergence of other database models, in particular geographical, spatial, semistructured, and XML. Recently, the need to manage information with graph-like nature has reestablished the relevance of this area. The main objective of this survey is to present the work that has been conducted in the area of graph database modeling, concentrating on data structures, query languages, and integrity constraints.

ACM Computing Surveys 2008

Integrating Data using Graphs and Semantics



MAPPING RELATIONAL DATABASES TO GRAPHS

Relational Database to RDF (RDB2RDF)

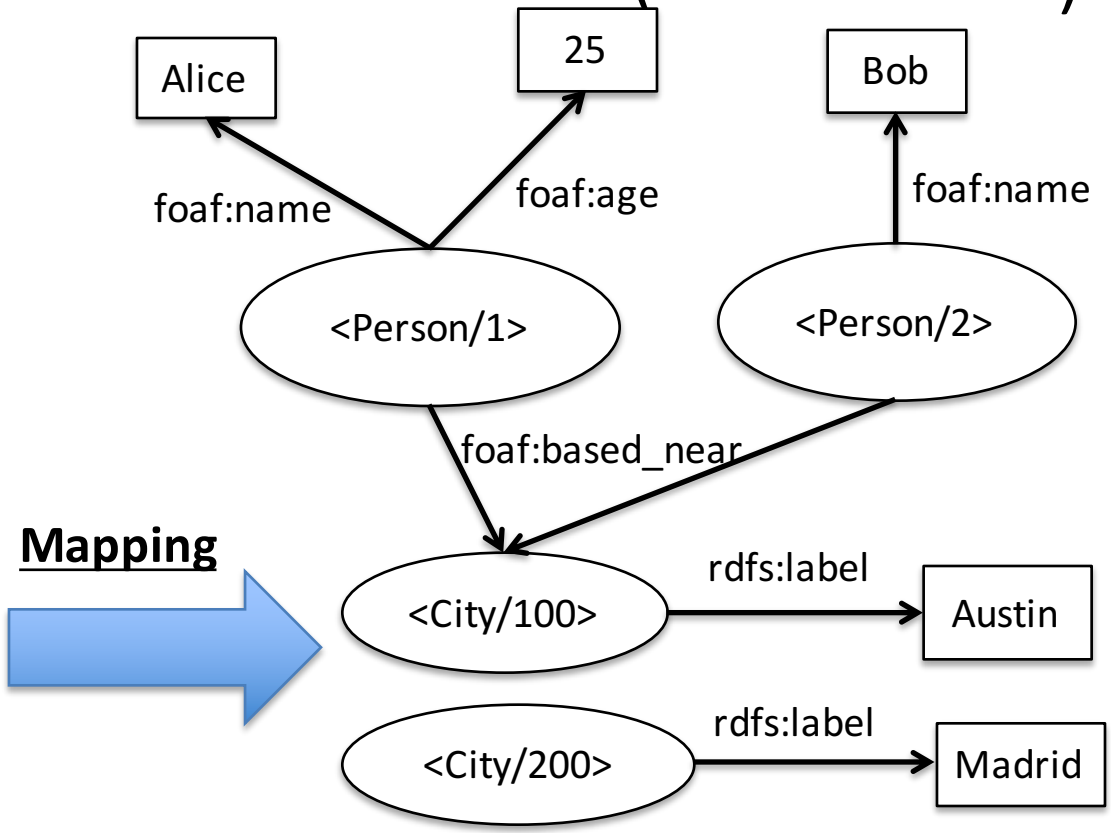
Person

ID	NAME	AGE	CID
1	Alice	25	100
2	Bob	NULL	100

City

CID	NAME
100	Austin
200	Madrid

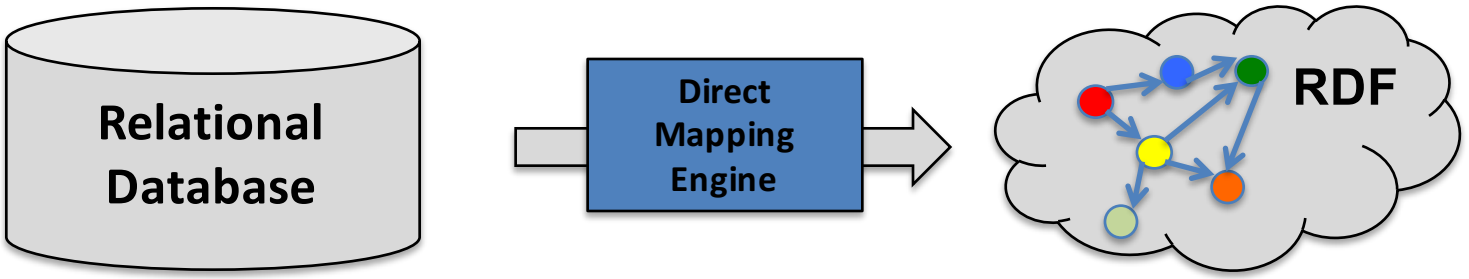
Mapping



W3C RDB2RDF Standards

- Standards to map Relational Data to RDF
- A Direct Mapping of Relational Data to RDF
  - Default automatic mapping of relational data to RDF
- R2RML: RDB to RDF Mapping Language
  - Customizable language to map relational data to RDF

W3C Direct Mapping

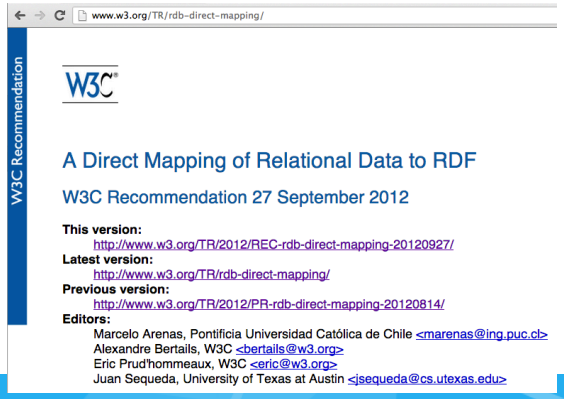


Input:

- Database (Schema and Data)
- Primary Keys
- Foreign Keys

Output

- RDF graph


www.w3.org/TR/rdb-direct-mapping/

W3C Recommendation

A Direct Mapping of Relational Data to RDF

W3C Recommendation 27 September 2012

This version:
http://www.w3.org/TR/2012/REC-rdb-direct-mapping-20120927/
Latest version:
http://www.w3.org/TR/rdb-direct-mapping/
Previous version:
http://www.w3.org/TR/2012/PR-rdb-direct-mapping-20120814/
Editors:
Marcelo Arenas, Pontificia Universidad Católica de Chile <marenas@ing.puc.cl>
Alexandre Bertails, W3C <bertails@w3.org>
Eric Prud'hommeaux, W3C <eric@w3.org>
Juan Sequeda, University of Texas at Austin <jsequeda@cs.utexas.edu>

W3C Direct Mapping Result

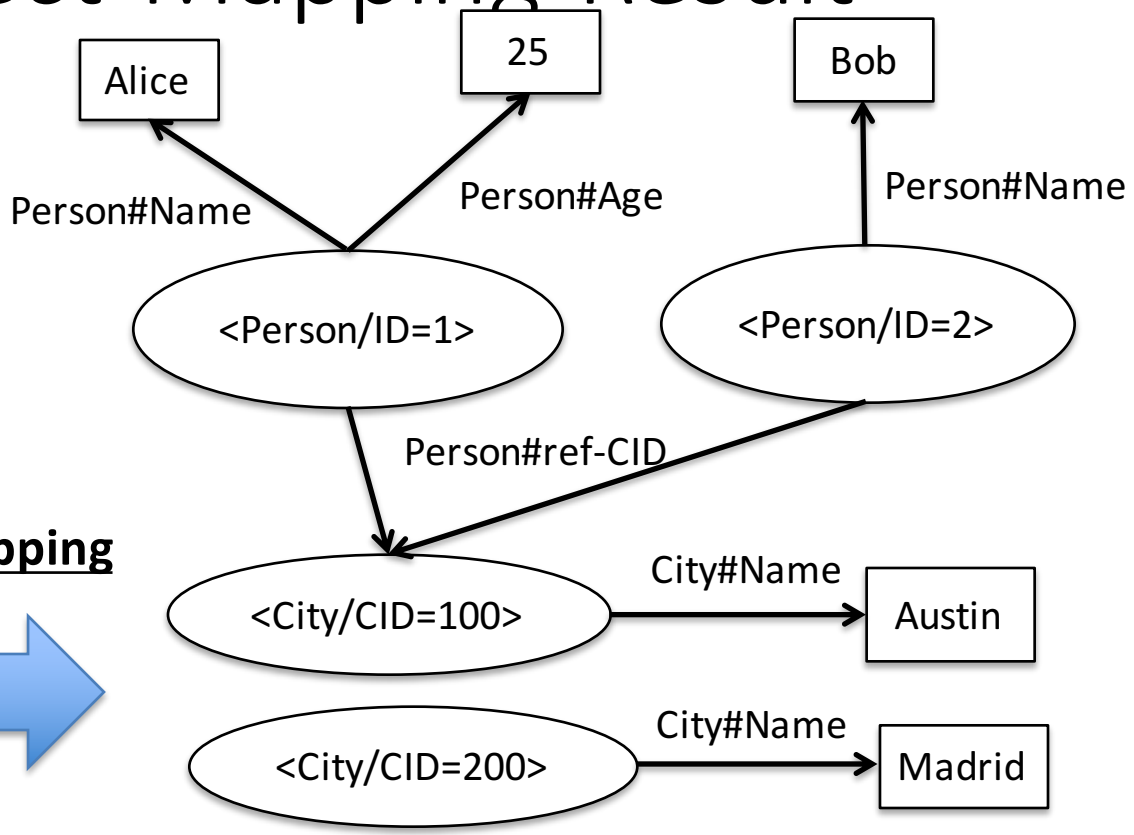
Person

ID	NAME	AGE	CID
1	Alice	25	100
2	Bob	NULL	100

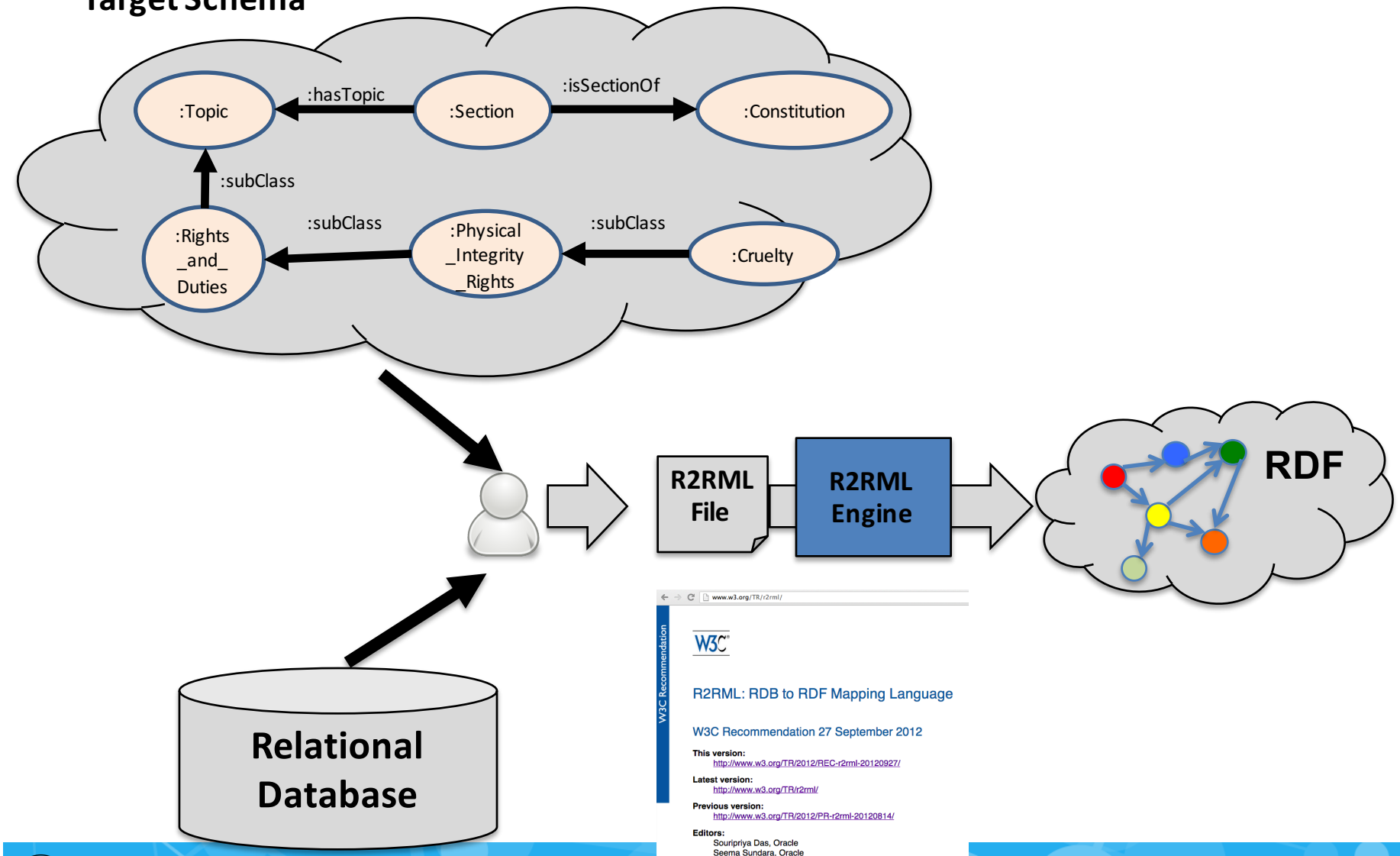
City

CID	NAME
100	Austin
200	Madrid

Direct Mapping

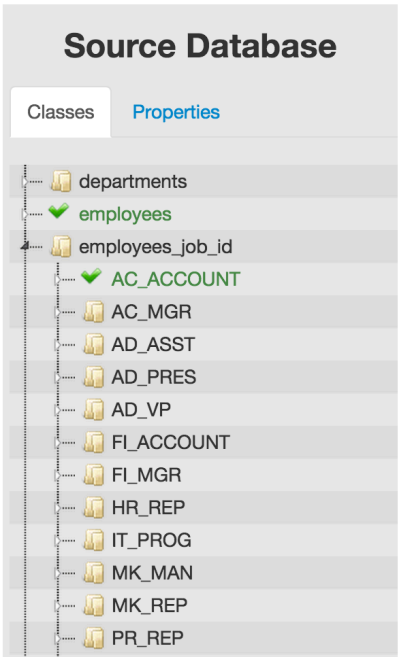


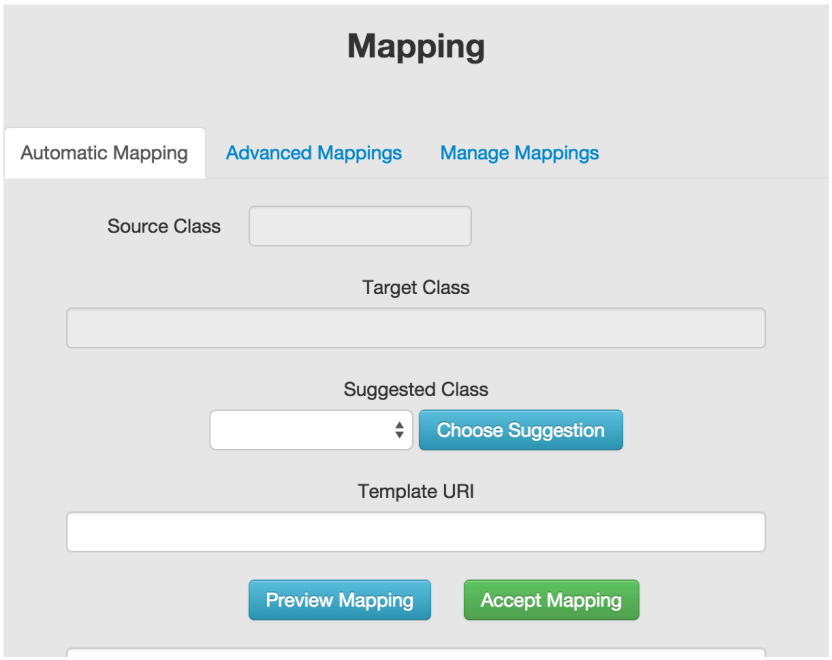
R2RML

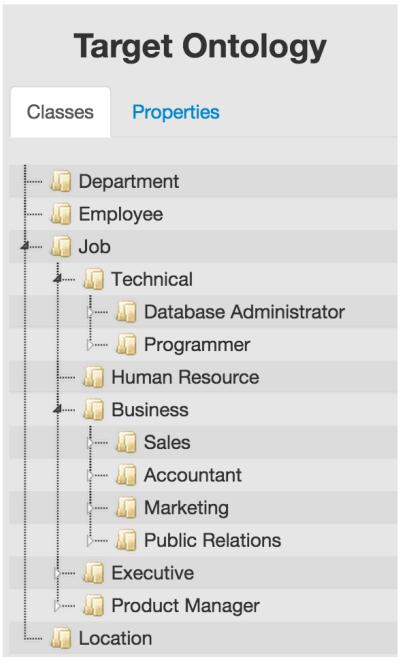

Target Schema
:Topic
:hasTopic
:Section
:isSectionOf
:Constitution
:subClass
:Rights _and_ Duties
:subClass
:Physical _Integrity _Rights
:subClass
:Cruelty
R2RML File
R2RML Engine
RDF
Relational Database
www.w3.org/TR/r2rml/
W3C Recommendation
W3C
R2RML: RDB to RDF Mapping Language
W3C Recommendation 27 September 2012
This version:
http://www.w3.org/TR/2012/REC-r2rml-20120927/
Latest version:
http://www.w3.org/TR/r2rml/
Previous version:
http://www.w3.org/TR/2012/PR-r2rml-20120814/
Editors:
Souripriya Das, Oracle
Seema Sundara, Oracle
Richard Cyganiak, DERI, National University of Ireland, Galway

Example R2RML

```
<TriplesMap1>
    a rr:TriplesMap;

    rr:logicalTable [ rr:tableName"Person" ];

"ht

foa
```

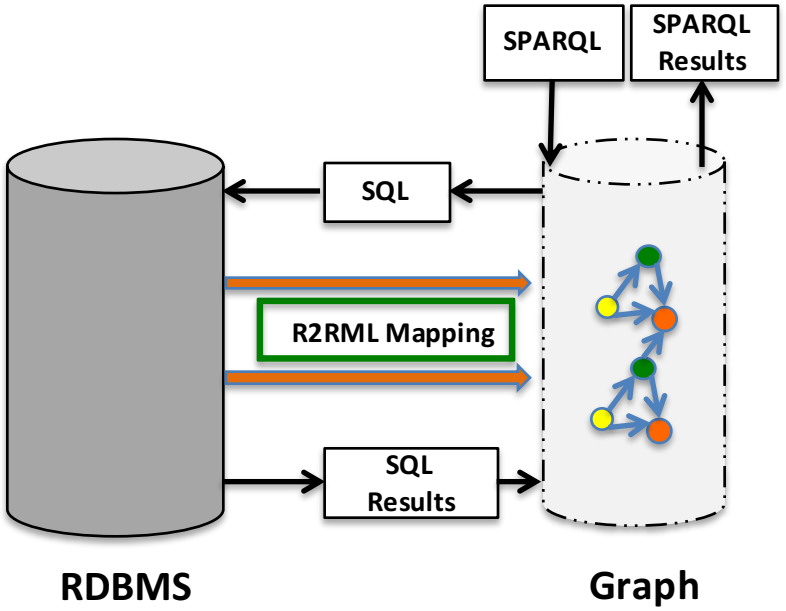






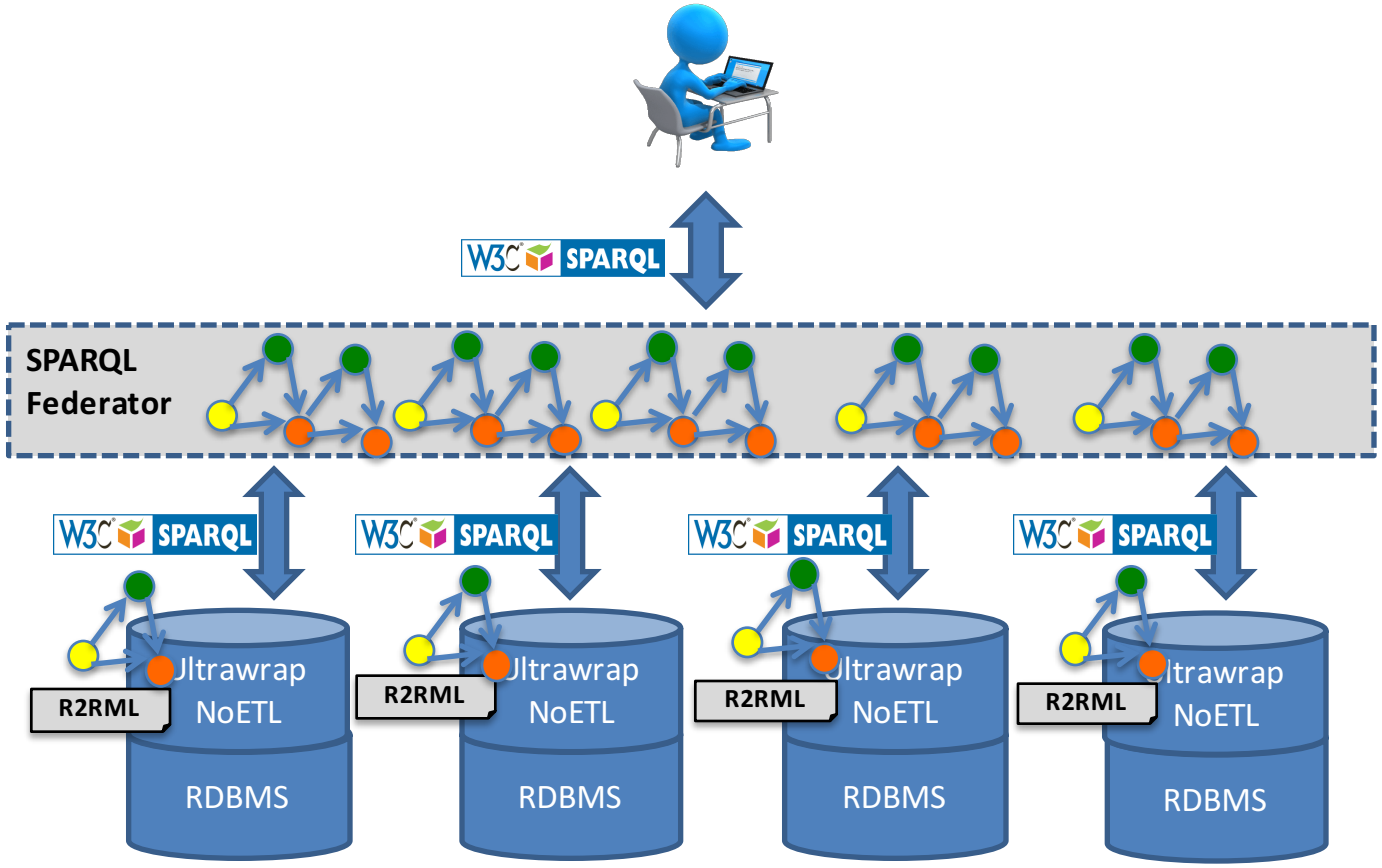

```
    rr:subjectMap [ rr:template "http://ex.com/City/{CID}";
                    rr:class ex:City ];

    rr:predicateObjectMap [
       rr:predicate  foaf:name;
       rr:objectMap  [ rr:column "TITLE" ]
    ]
    .
```

Graph Data Virtualization

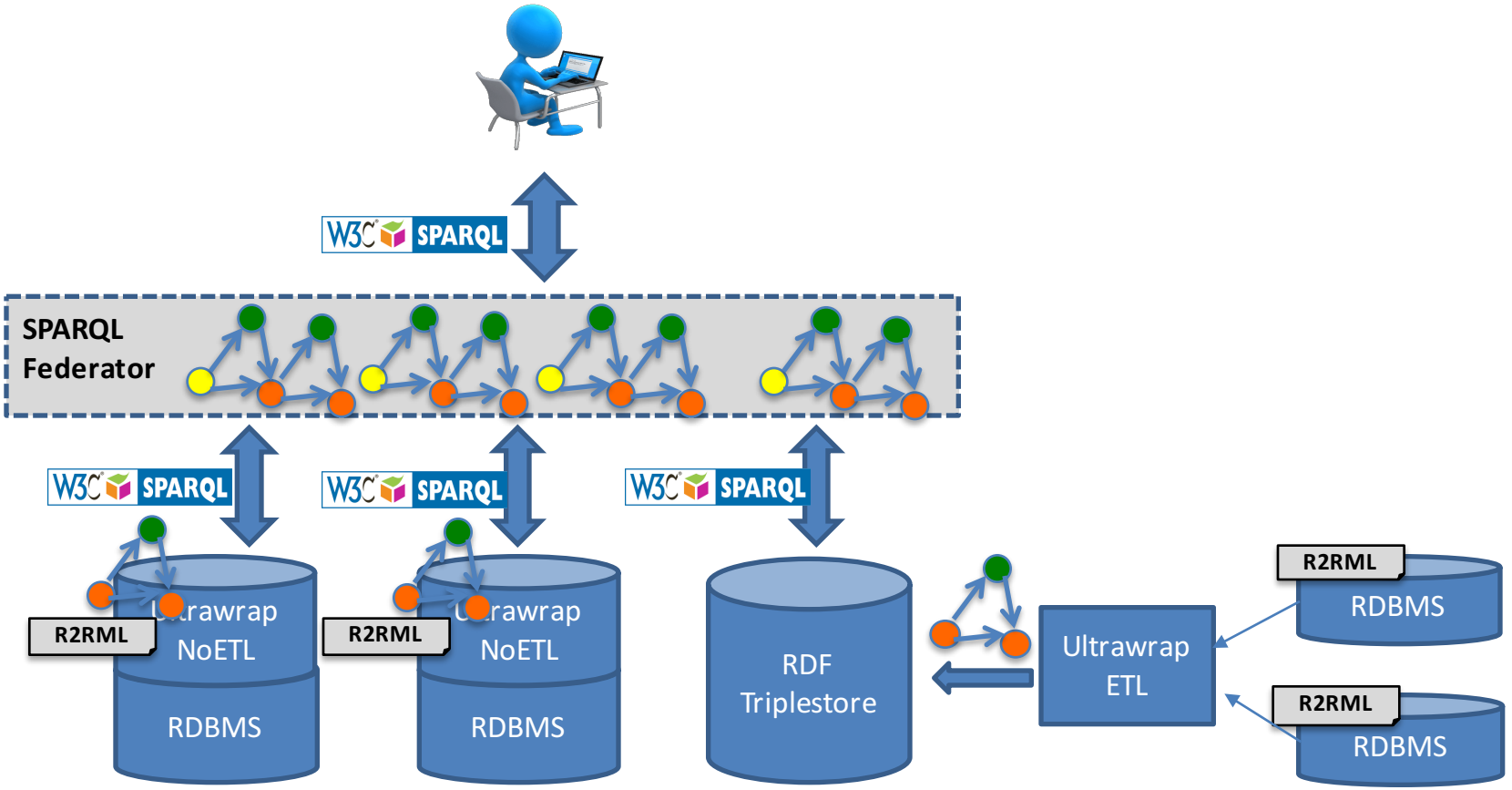


ULTRAWRAP

NoETL Architecture

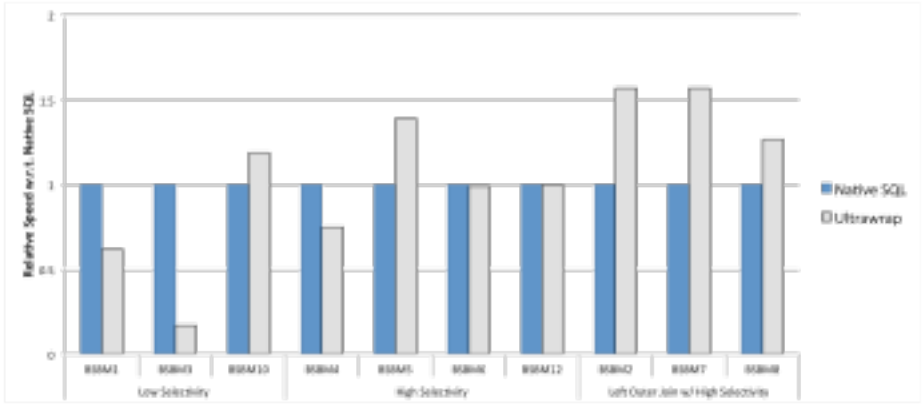


Hybrid NoETL and ETL Architecture



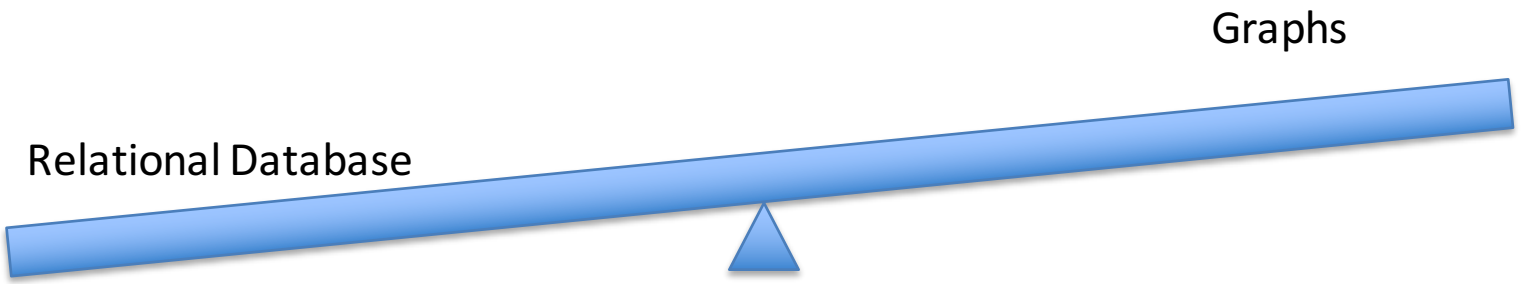
Scalability

- Seconds vs Months
- Reuse existing relational infrastructure
  - 30+ years of optimizations
  - Semantic Query Optimizations
- Result: SPARQL as fast as SQL under mappings



Sequeda & Miranker. Ultrawrap: SPARQL Execution on Relational Data. J. of Web Semantics 2013

The Tipping Point Problem



- Flexible
- Integration
- Data and Metadata are One
- Common Denominator
- Traversal, Navigation, Reachability
- Semantics

An overarching theme is the need to create systematic and real-world benchmarks in order to evaluate different solutions for these features.

Sequeda (2015) Integrating Relational Databases with the Semantic Web