D2RCrime: A Tool for Helping to Publish Crime Reports on the Web from Relational Data

Júlio Alcântara Tavares, Vasco Furtado, Henrique Oliveira, Eurico Vasconcelos
University of Fortaleza
Brazil
STIDS, Fairfax, October, 2012
Agenda

- Context and Motivation
- Assisting the mapping of crime relational data to RDF
- The D2RCrime tool
- Future Work
- Conclusion
Motivation

Difficulties to open crime data

- Transparency

- Underreporting
  - more than 50% for thefts and robberies in big metropolis

- Prevention through Access to information
  - “Is here dangerous?”
  - “Beware of pickpocketing”

- Collaborative Crime Mapping
WikiCrimes

www.wikicrimes.org
WikiCrimes Services

- Prevention through Access to information
  - “Is here dangerous?” iPhone and Android
  - “Beware of pickpocketing” QR-Code
  - Receive email about a particular area
  - Stats about crime in the World
  - Safe route
  - Filter information by time, day, type, etc.
Research in WikiCrimes

- Identifying Malicious Activity in Collaborative Maps

- Trust and Credibility of Information
  - *Information Science*, vol. 180, issue 1, 2010

- Pinpointing crimes automatically via Natural Language Processing
  - *IEEE Intelligence and Security Informatics*, Vancouver, 2010 (best paper)

- Open Government and Citizen Participation via Collaborative Mapping
  - *IEEE Intelligent Systems*, vol. 27, issue 4, 2012

- Semantic for Collaborative Maps *(www.wikimapps.com)*
  - *Proceedings of the Brazilian Symposium on AI*, October, 2012
Semantic WikiCrimes

Crime Report
Provenance
Trust and Reputation
New trend in e-gov: Raw data on crime occurrence

Each police department to define its own way to open its data by creating intermediary representations (typically spreadsheets). Alternatively, they develop their own APIs.

See http://www.atlantapd.org/crimedatadownloads.aspx in Atlanta
See http://sanfrancisco.crimespotting.org/api for San Francisco

Several proposals in the industry and academia for mapping RDB to RDF

Lack of a standard for crime and crime report representations
D2R Server

- Data on-demand (*on-the-fly*)
  - No need for replications
  - W3C working group - R2RML
Mapping File

To express the mapping between an ontology and a RDB.

Uses the D2RQ language, *(in N3)*.

Main structures:
- ClassMap
- PropertyBridge
- TranslationTable
- JavaClass
The Mapping File Example

# TranslationTable generated to implement a mapping 1xN (1 Table x N Classes)

map:TypeofCrimeBridge a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:CrimeOccurrence;
  d2rq:uriColumn "tb_cri_crime.TCR_IDTYPE_CRIME";
  d2rq:translateWith map:TypeofCrimeTranslator;
.

map:TypeofCrimeTranslator a d2rq:TranslationTable;
  d2rq:translation [ d2rq:databaseValue "1"; d2rq:rdfValue wm:AttemptRobbery; ];
  d2rq:translation [ d2rq:databaseValue "2"; d2rq:rdfValue wm:AttemptTheft; ];
  d2rq:translation [ d2rq:databaseValue "3"; d2rq:rdfValue wm:Theft; ];
  d2rq:translation [ d2rq:databaseValue "4"; d2rq:rdfValue wm:Robbery; ];
  d2rq:translation [ d2rq:databaseValue "5"; d2rq:rdfValue wm:Murder; ];
.

✓ Steep learning curve!
The D2RCrime Tool

Motivation

- Easy to use
  - SQL metaphor
  - To generate automatically the Mapping File
  - TARGET: The designer/DBA of Police Departments
- Based on a (eventually) standard representation
  - Crime and Crime Report ontology
D2RCrime
The Approach

CRIME ONTOLOGY
An interactive and iterative process that guides the automatic construction of the mapping file in D2RQ based on the metaphor of SQL.

- SQL is a declarative language widely known by DBAs and System Analysts.

- The basic premise is that the mapping can be automatically done from questions to the DBA about how to define the tuples of a DB that describe a certain class (or property) in the Crime Ontology.
D2R Crime Mapping Crime Occurrences

Step 2:Occurrences Mapping

Enter the SQL query used to return all the Thefts:

```sql
SELECT cri.cri_dsc_description AS OCCURRENCE_DESCRIPTION,
       cri.cri_id_crime_type AS OCCURRENCE_TYPE,
       cri.cri_date_time AS OCCURRENCE_DATE_TIME,
       cri.cri_latitude AS OCCURRENCE_LATITUDE,
       cri.cri_longitude AS OCCURRENCE_LONGITUDE
FROM tb_crime cri
WHERE cri.flag_theft = 'Y'
```
D2RCrime Correspondence RDB2RDF

Column OCCURRENCE_DESCRIPTION
- Class Information
- pmlp:hasPrettyString

Column OCCURRENCE_DATE_TIME
- Class DateTimeDescription
  - time:hour, time:day, time:month, time:year

Column OCCURRENCE_LATITUDE/OCCURRENCE_LONGITUDE
- Class Point
  - wgs84_pos:lat, wgs84_pos:long

Column OCCURRENCE_TYPE
- Class specialization
  - d2rq:translationTable
Finalization

Instances Publication Finished

You are now able to open and expose your Data in the Semantic Web!

Your data is published at http://www.wickicrimes.org

Your SPARQL EndPoint is available at:
http://semantic.wikimaps.com:2025/snorql

Your HTML EndPoint is available at:
http://semantic.wikimaps.com:2025/
Research Challenges

- Integration official data to citizen-direct data
  - Trust, reputation, credibility

- Entity disambiguation / Identifiers
  - Same Crime?

- The meaning is the same (e.g. burglary, "roubo a residencia")
  - Is there a right level of granularity to use in the representation?
Conclusion

- Open Crime Data is already there but not linked and not standardised.

- To facilitate the work of “liberators” (of data) is necessary
D2RCrime

Steps to publish RDF crime data

1. Configuration of DB
2. Mapping of crime occurrences based on the Crime Ontology
3. Preview at Google Maps

 ✓ Step 2 is executed for every class of the ontology