



**Engaging Content**  
Engaging People

# R2RML-F: Towards Sharing and Executing Domain Logic in R2RML Mappings

Christophe Debruyne and Declan O'Sullivan

ADAPT Centre (<http://www.adaptcentre.ie/>)

Trinity College Dublin (<http://www.tcd.ie/>)

2016-04-12 @ Linked Data on the Web (LDOW2016)



The ADAPT Centre is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

The RDB-2-RDF initiatives ultimately led to two W3C Recommendations: a **direct mapping** and **R2RML**.

R2RML allows one to describe **mappings** and assumes the database to conform to the Core SQL 2008 specification.

**But what if the underlying technology does not support certain data manipulations?**

- Underlying technology not expressive enough.
- Procedural domain knowledge part of application.



- The Ordnance Survey Ireland (OSi) – Ireland's National Mapping Agency – engaged with ADAPT to serve Ireland's geospatial information as Linked Data.
- Map points and polygons in Irish Transverse Mercator stored in a table to World Geodetic System 84 in RDF. The relational database did not support conversion between coordinate systems.

id	name	point
10000	CARLOW	POINT(671989.126 676051.233)
...	...	...

```
<http://data.example.com/county/10000>  
<http://www.example.com/ns#point>  
"Point(52.83060100053135 -6.93169839449793)" ;  
  <http://www.example.com/ns#name> "CARLOW" .
```



- When such functionality is not available, one has to resort to **more complex data processing “pipelines”**, which impact transparency and traceability.
  - Preprocessing, RDF transformations, ...
- Bernstein and Melnik argued mapping languages should be more expressive to support various case scenarios and that one could include constructs for declaring user-defined functions\*.

\* Provided tractability is not a problem.



- We argue such procedural knowledge **can and should be included** in the mapping.
- **Advantages** are traceability (e.g., using PROV-O), transparency, exchange of procedural domain knowledge, ...
- The requirements formulated for our approach – dubbed **R2RML-F** – are:
  - A minimal extension of R2RML, and
  - Adopting a standardized programming language.



- Namespace rrf: <http://kdeg.scss.tcd.ie/ns/rrf#>
- *Functions* have a function *name* and *body*.

```
<#Multiply>
  rrf:functionName "multiply" ;
  rrf:functionBody """
    function multiply(var1, var2) {
      return var1 * var2 ;
    }
  """ ;
.
```

- Functions are written in ECMAScript.



- A “function valued” term map *calls a function* and the *parameters* are themselves term maps.

```
<#TriplesMap1>
rr:logicalTable [ rr:tableName "Employee"; ];
rr:subjectMap [ rr:template "http://org.com/employee/{ID}"; ] ;
rr:predicateObjectMap [
  rr:predicate ex:salary ;
  rr:objectMap [
    rr:datatype xsd:double ;
    rrf:functionCall [
      rrf:function <#Multiply> ;
      rrf:parameterBindings (
        [ rr:constant "12"^^xsd:integer ]
        [ rr:column "monthly_salary" ]
      ) ;
    ] ;
  ] ;
]
```

Parameter bindings as  
an RDF Collection.

Parameter bindings can  
be empty.

Term Maps as parameters.



- Current prototype is based on db2triples.  
(<https://github.com/CNGL-repo/db2triples>)
- Functions are evaluated and called using Java's Nashorn JavaScript engine.
- The R2RML-F processor first loads the functions before processing the triple maps.
- Parameters – which are term maps – are evaluated before passed on as arguments to a function.





## Transforming points in ITM to WGS84.

- Loaded the table into a PostgreSQL database with PostGIS enabled, we created a table for the 26 counties of Ireland consisting of an *id*, *name*, *geom* (the geometry object) and *geoms* (the geometry object as string in WKT)

Three mappings were created:

- M1: transforming geoms with function;
- M2: cast geom to string and apply to function;
- M3: transform geom and cast to string;



# Demonstration and Experiment (i)

Each mapping contained function which is loaded to create a “constant” across the three mappings. Each mapping was run 110 times in one thread and only the last 100 execution times were taken into account.

	M1	M2	M3
Average	77.66	78.62	87.86
Standard Deviation	37.69	34.80	40.19
Min	16.00	20.00	15.00
Max	205.00	214.00	205.00

Welch Two Sample T-Test for each pair of mappings to investigate whether differences are significant. The p-values were...

1. M1-M2: 0.85180 > 0.05
2. M1-M3: 0.06564 > 0.05
3. M2-M3: 0.08380 > 0.05

Suggesting that differences are not significant.



## **Multiplying numbers.**

1,000 records with three fields (id, x, and y), where x and y contained random values between 0 and 100.

1. Multiply x with x with function and in query.
2. Multiply x with y with function and in query.

Experiments ran in similar fashion.

- For 1, strong evidence that function is better.
- For 2, strong evidence that query is better.

More experiments are needed to draw more conclusive results.



- Inclusion of (user defined) functions has been studied in fields such as model-to-model transformations for mappings between UML and ER, and even UML and OWL. **We incorporated this notion in a W3C recommendation for uplift, also allowing one to share and annotate these functions.**
- XSPARQL and RML, amongst others, rely on the underlying technology. **Inclusion of functions not depending on those technologies solves potential limitations.**
- KR2RML – “OpenRefine Like” editor on top of an extension of R2RML supporting functions. **Drawbacks are their own serialization format and functions that cannot be reused.**



- Certain data manipulation tasks cannot be captured in R2RML and relies on more complex data processing “pipelines”. If one is willing to trade tractability for richer mappings, user-defined functions can be included in R2RML mappings.
- We presented **R2RML-F**, which extends R2RML with function call term maps adopting ECMAScript as the programming language.
- Our initial results show that our **approach is viable** and that there seems to be little impact on the performance.



- Additional experiment to validate our findings and gather more conclusive results.
- Development and gathering additional use case scenarios motivating this approach.
  - E.g., calling web services
  - E.g., in digital humanities or application of geospatial data



# Acknowledgements

www.adaptcentre.ie

We thank the Ordnance Survey Ireland (OSi) for permitting us to use their boundaries dataset for the purposes of this research project. Within OSi, we are especially grateful for the input and domain expertise provided by Lorraine McNerney and Éamonn Clinton.



OSi Linked Data: <http://data.geohive.ie/>

