

-From Relational to RDF-

Núria Queralt Rosinach

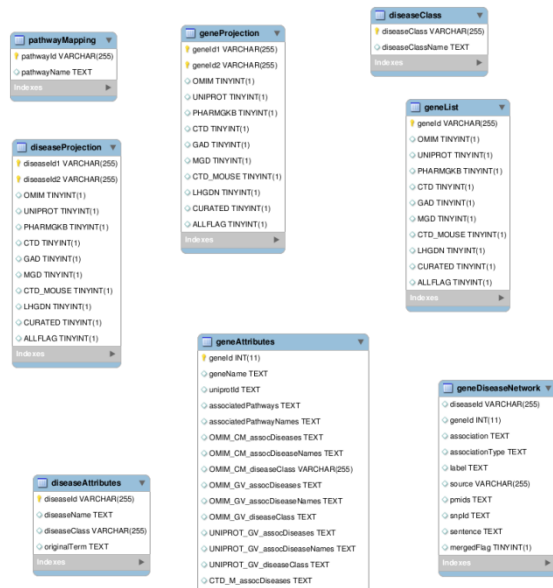
2015-12-04, Utrecht

BioSB /ELIXIR-NL course

Managing & Integration Life Science Information

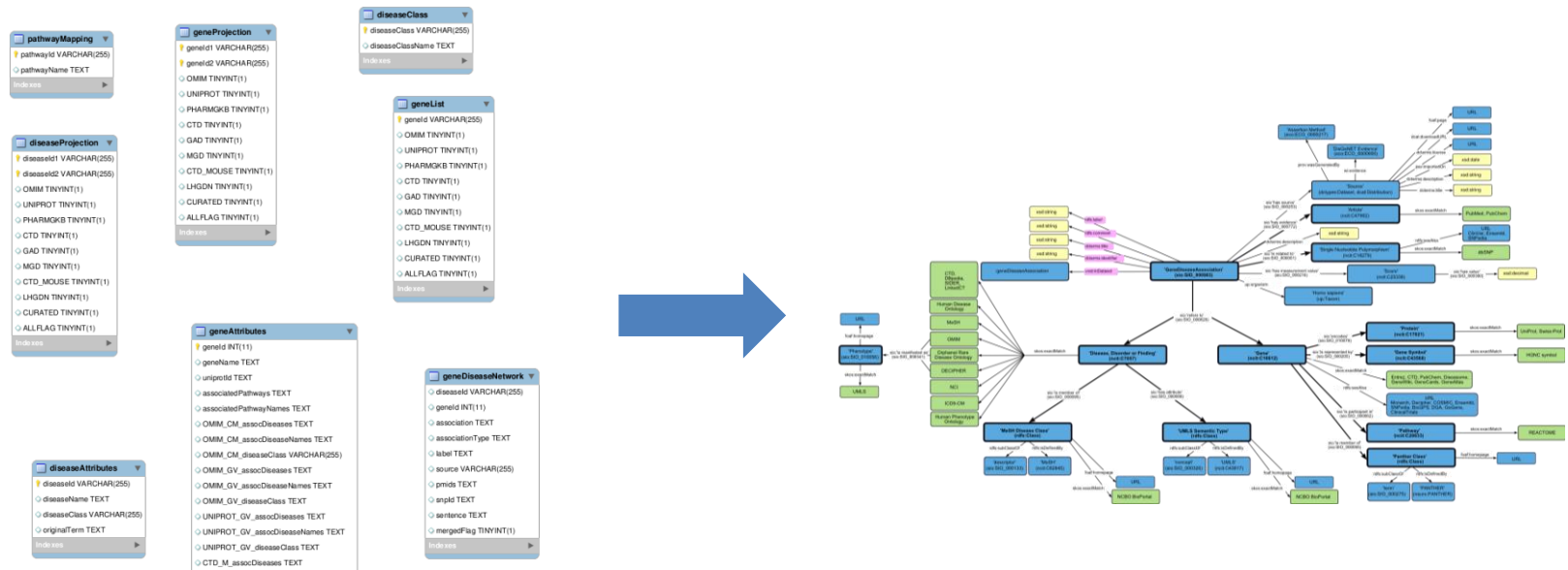
DisGeNET

- Relational Database (MySQL)



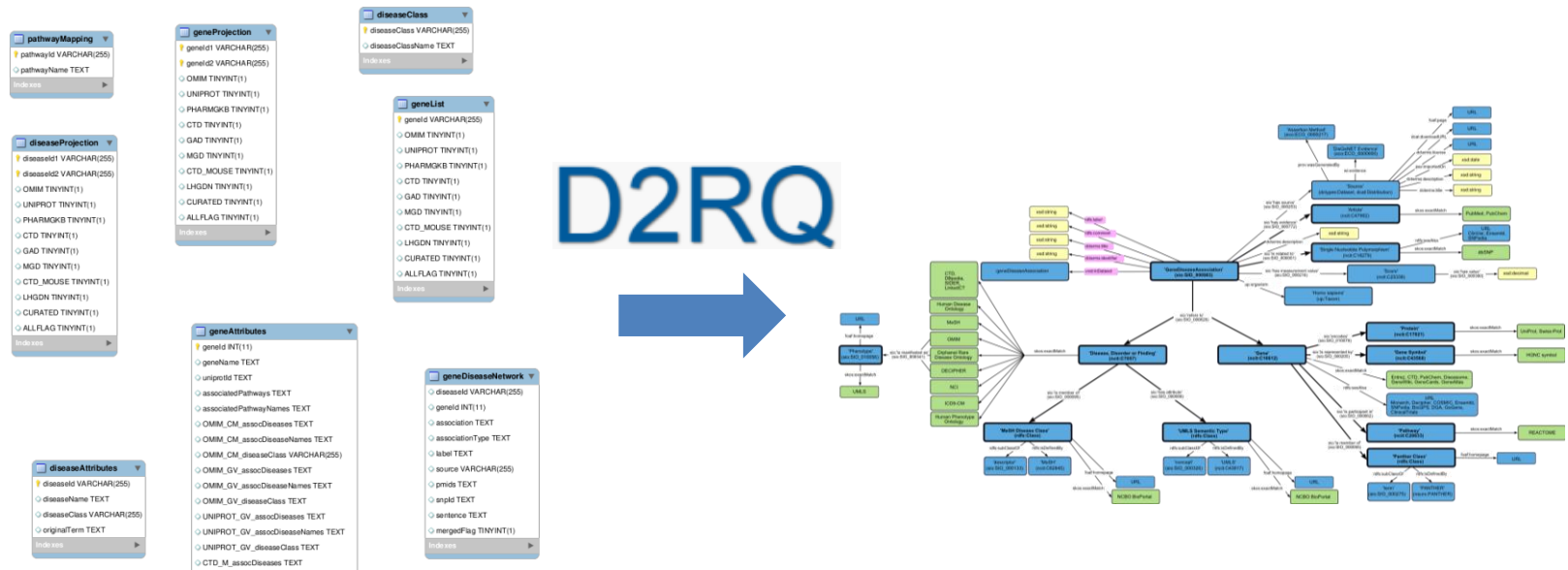
DisGeNET

- Relational Database => RDF Linked Dataset



DisGeNET

- Direct translation of data models



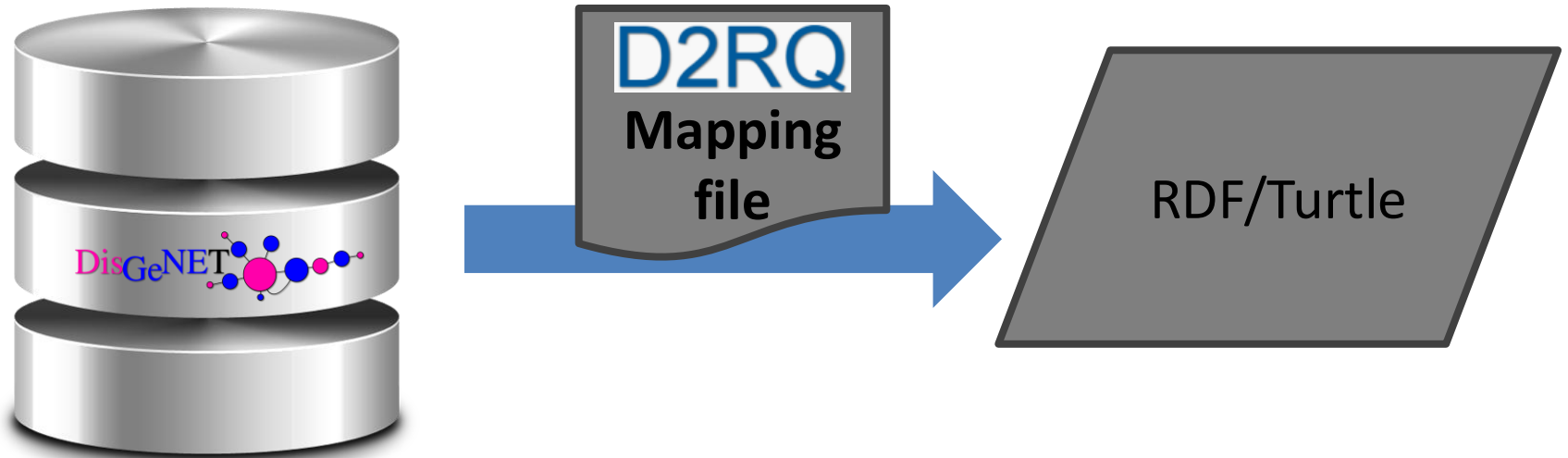
D2RQ

- D2RQ Platform: access relational DBs as virtual RDF graphs
- Open Source Software (Java). Performance issues with large SQL tables.
- Several tools
- Enables serialize relational data into RDF dumps for loading into an RDF store.

<http://d2rq.org/>

DisGeNET

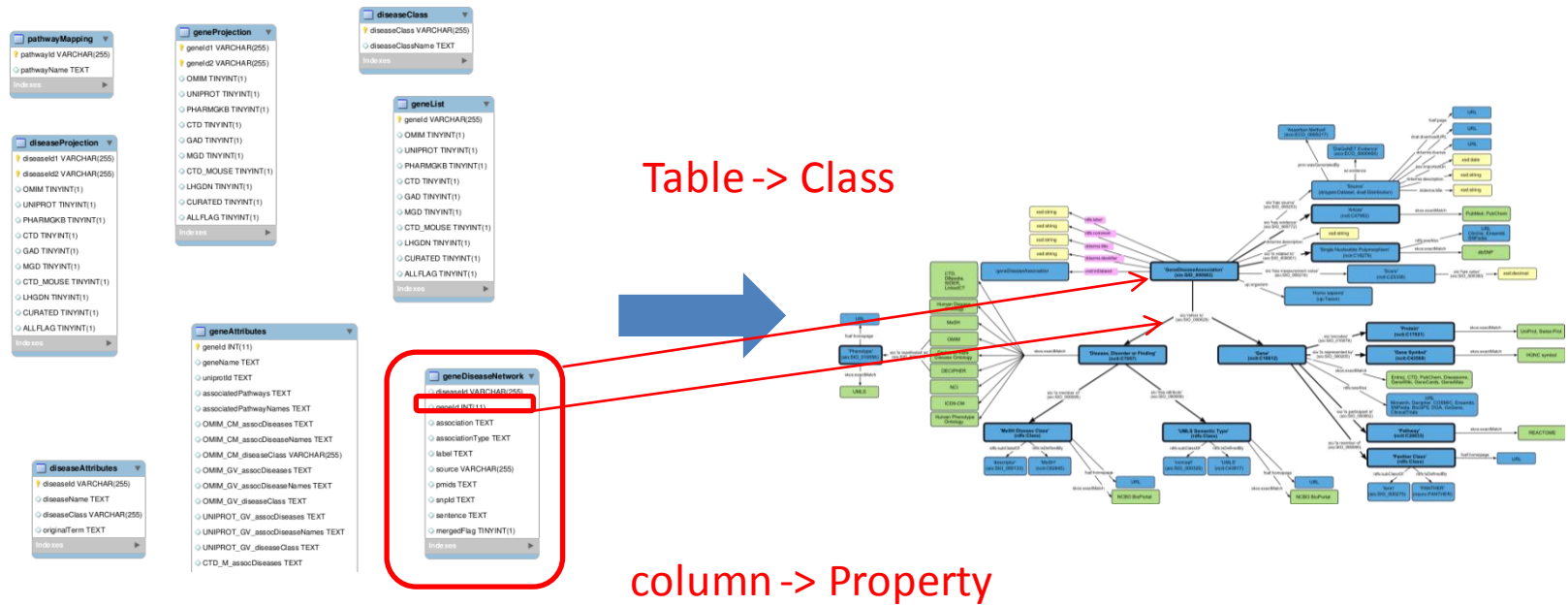
- Direct translation of data models



Mapping-based approach

DisGeNET

- Direct translation of data models



DisGeNET

- Direct translation of data models



\$ cat mapping.ttl

D2RQ Mapping Language

- Describes the mapping
- Re-structures content
- Writes RDF syntax

```
# D2RQ Namespace
@prefix d2rq: <http://www.wiwiw.fu-berlin.de/suhl/bizer/D2RQ/0.1#> .
# Namespace of the ontology
@prefix : <http://annotation.semanticweb.org/iswc/iswc.daml#> .

# Namespace of the mapping file; does not appear in mapped data
@prefix map: <file:///Users/d2r/example.ttl#> .

# Other namespaces
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

map:Database1 a d2rq:Database;
  d2rq:jdbcDSN "jdbc:mysql://localhost/iswc";
  d2rq:jdbcDriver "com.mysql.jdbc.Driver";
  d2rq:username "user";
  d2rq:password "password";
.

# -----
# CREATE TABLE Conferences (ConfID int, Name text, Location text);

map:Conference1 a d2rq:ClassMap;
  d2rq:dataStorage map:Database1;
  d2rq:class :Conference;
  d2rq:uriPattern "http://conferences.org/comp/confno@Conferences.ConfID@@";
.

map:eventTitle a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:Conference;
  d2rq:property :eventTitle;
  d2rq:column "Conferences.Name";
  d2rq:datatype xsd:string;
.

map:location a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:Conference;
  d2rq:property :location;
  d2rq:column "Conferences.Location";
.

d2rq:datatype xsd:string;
```


DisGeNET

- The **generation** is command-line:

- 1: Generate mapping

```
$ generate-mapping default-mapping.ttl &
```



Manual edition

```
$ vi default-mapping.ttl -> :w mapping.ttl
```

- 2: Generate the dump

```
$ dump-rdf -f turtle -o disgenet-rdf.ttl mapping.ttl &
```

DisGeNET

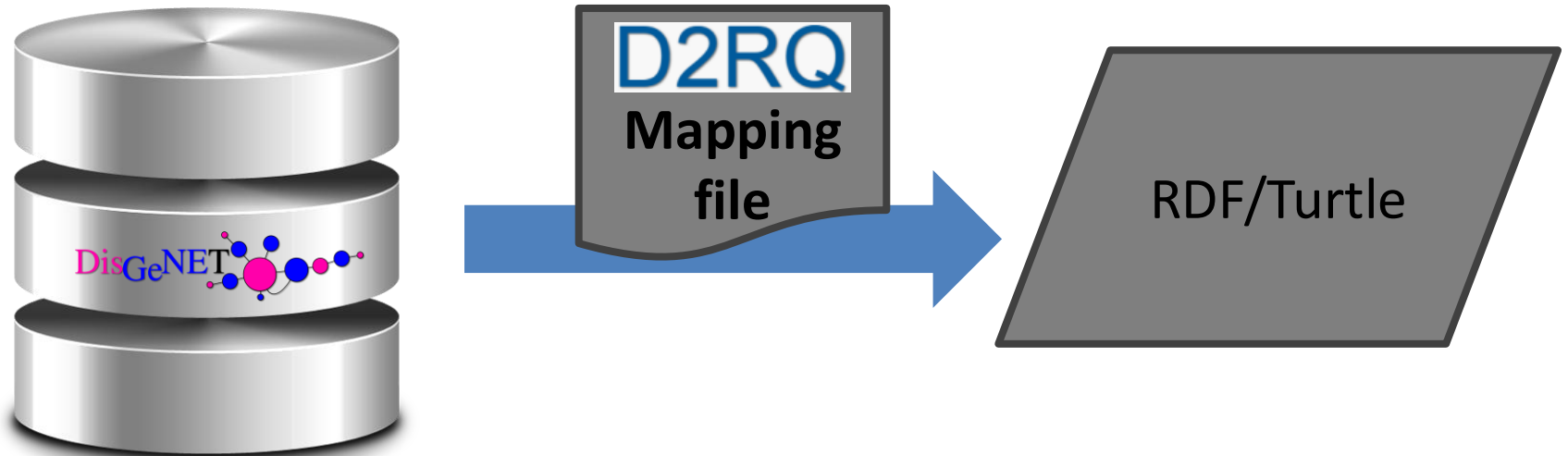
- Dump serialized in RDF/Turtle format

```
$ cat disgenet-rdf.ttl
```

```
<http://rdf.disgenet.org/resource/gda/DGNf5cb3969d758  
71f05a5d5f984f8dfc34>  
    a sio:SIO_001122 ;  
    sio:SIO_000628  
<http://identifiers.org/ncbigene/4728> ,  
<http://linkedlifedata.com/resource/umls/id/C0023264> .
```

DisGeNET

- Direct translation of data models



```
$ dump-rdf -f turtle -o disgenet-rdf.ttl mapping.ttl &
```

Thanks for your attention!
Questions are welcome

