Using techniques for Ontology-based Data Access to autogenerate Daily Drilling reports

Martin Giese, Univ. of Oslo
Jens I. Ornæs, National Oilwell Varco
Lars Overå, PCA
Inge Svensson, Baker Hughes
Arild Waaler, Univ. of Oslo

Department of Informatics

University of Oslo

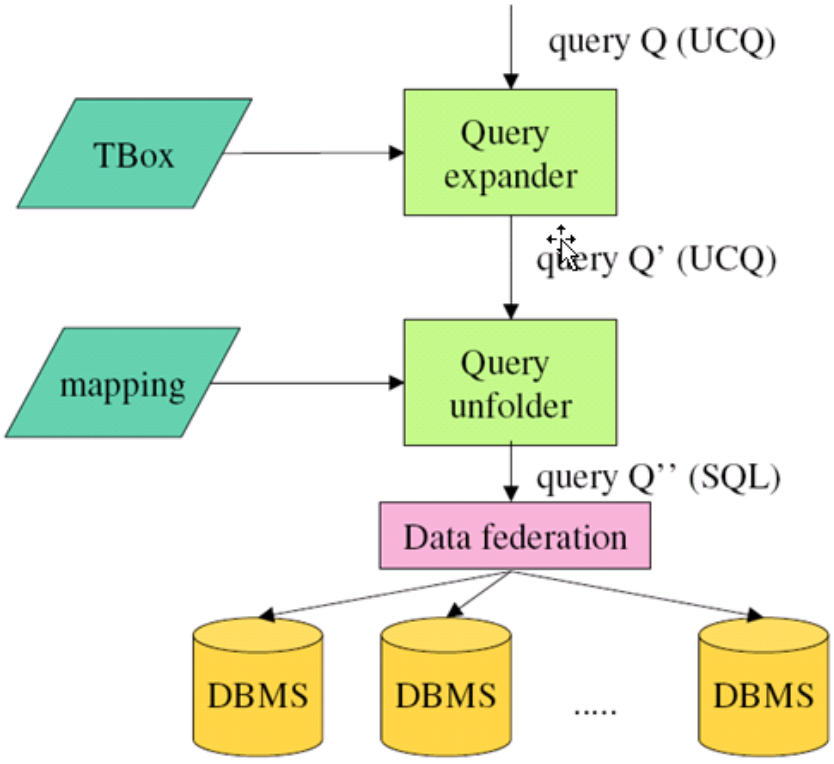
About this work

- It's a case study
- Originated in the Drilling activity of Integrated Operations in the High North in 2008
- Primarily implemented by two MSc students
- Additional funding from PCA / EPIM
- Baker Hughes provided drilling data
- Accepted for presentation at the Society of Petroleum Engineers' Intelligent Energy conference (SPE IE) 2012

Point of departure: Reporting in the O&G sector

- Reports have a clearly specified content
- The descriptions employ standards or "nearby standards"
- Report generation has an element of data integration
- The O&G industry spends significant resources on this
  - Manual processes
  - Integration not supported by the systems they use
- The amount of reporting is increasing

Virtual Ontology Based Data Access (OBDA)

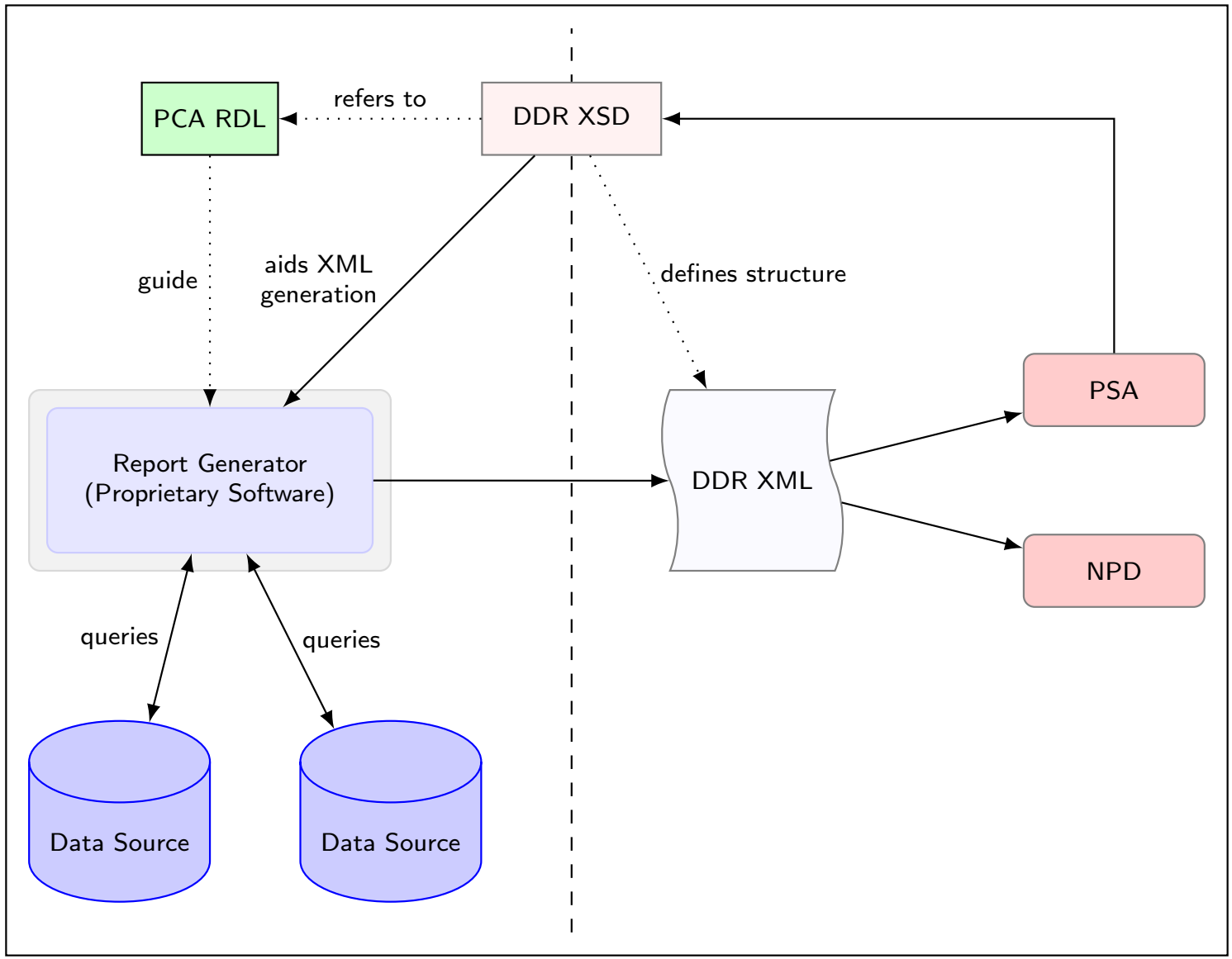


- TBox: End user vocabulary as OWL-QL ontology
- mapping: Ontology concepts map to DBMS by SQL
- Query Q in end user vocabulary
- Query rewriting generates SQL query Q"
- Prototype tools
  - QuOnto
  - Quest

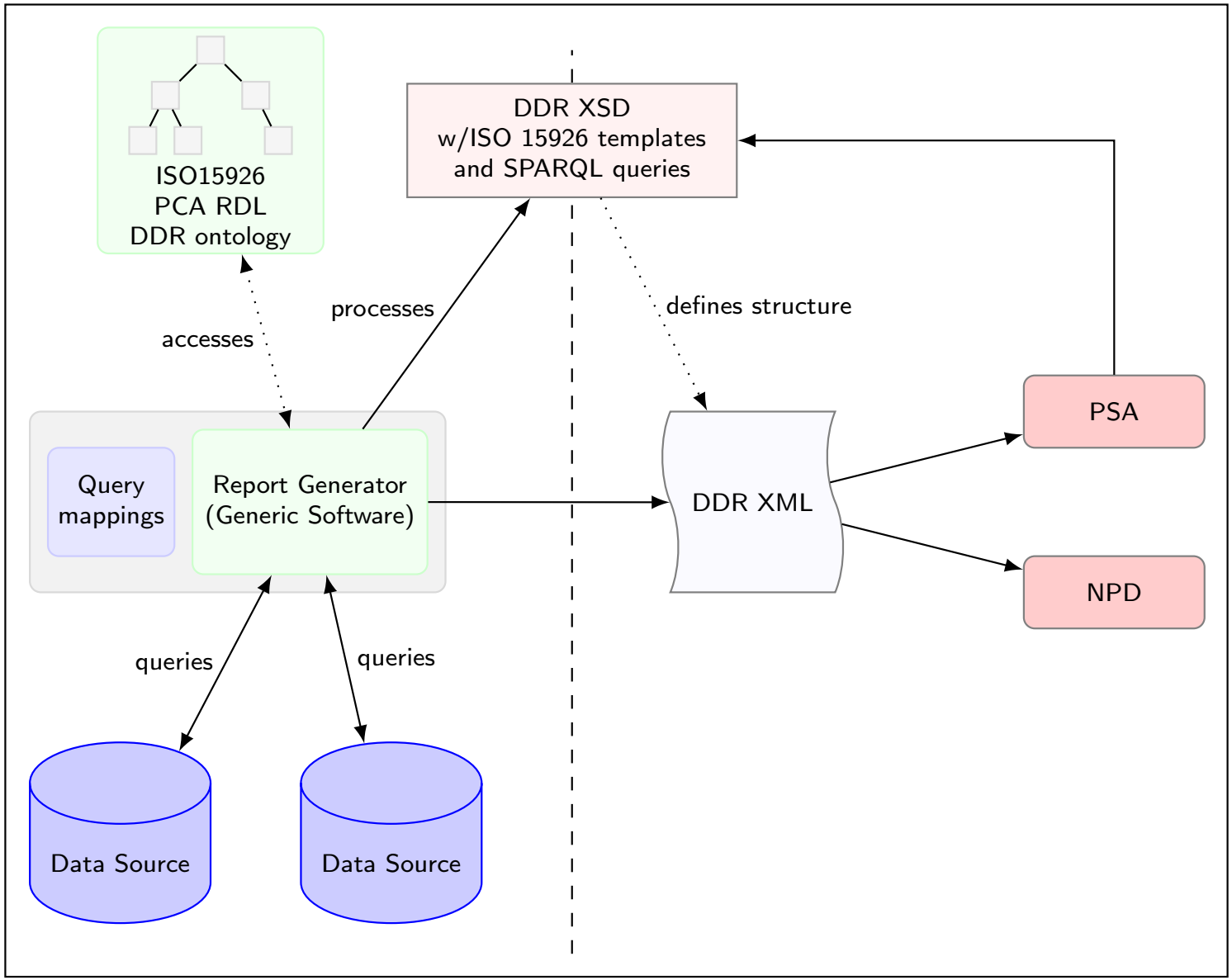
Selling points for semantic technology/declarative methods

- Information quality: What are the sources of error?
- Maintenance. How smoothly do the solutions support changes?
- Genericity: Does the report generation practice easily transfer from one type of report to another?
- Reuse: How easy is it to use the reported data in other contexts?

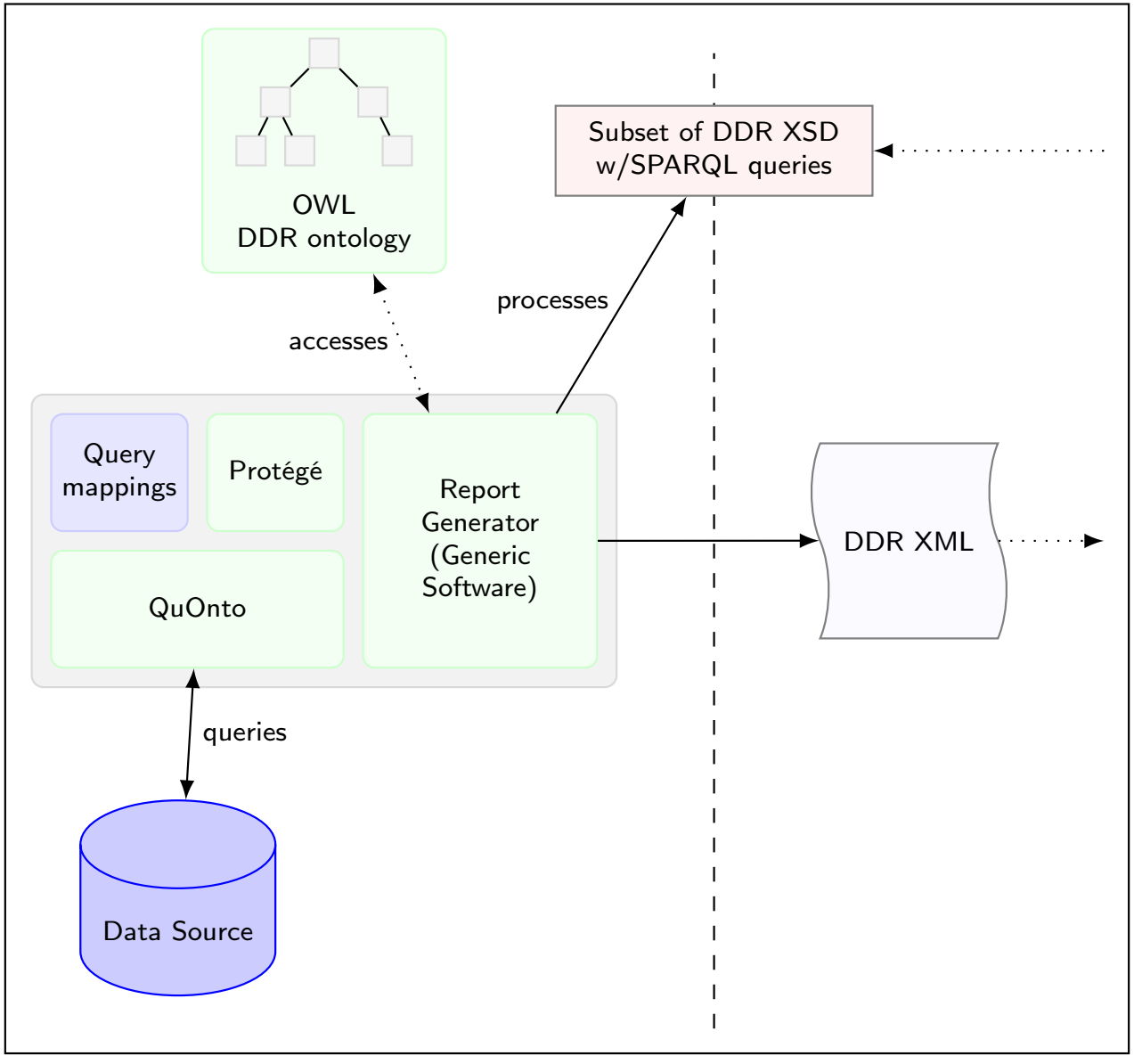
Current practice for DDR generation



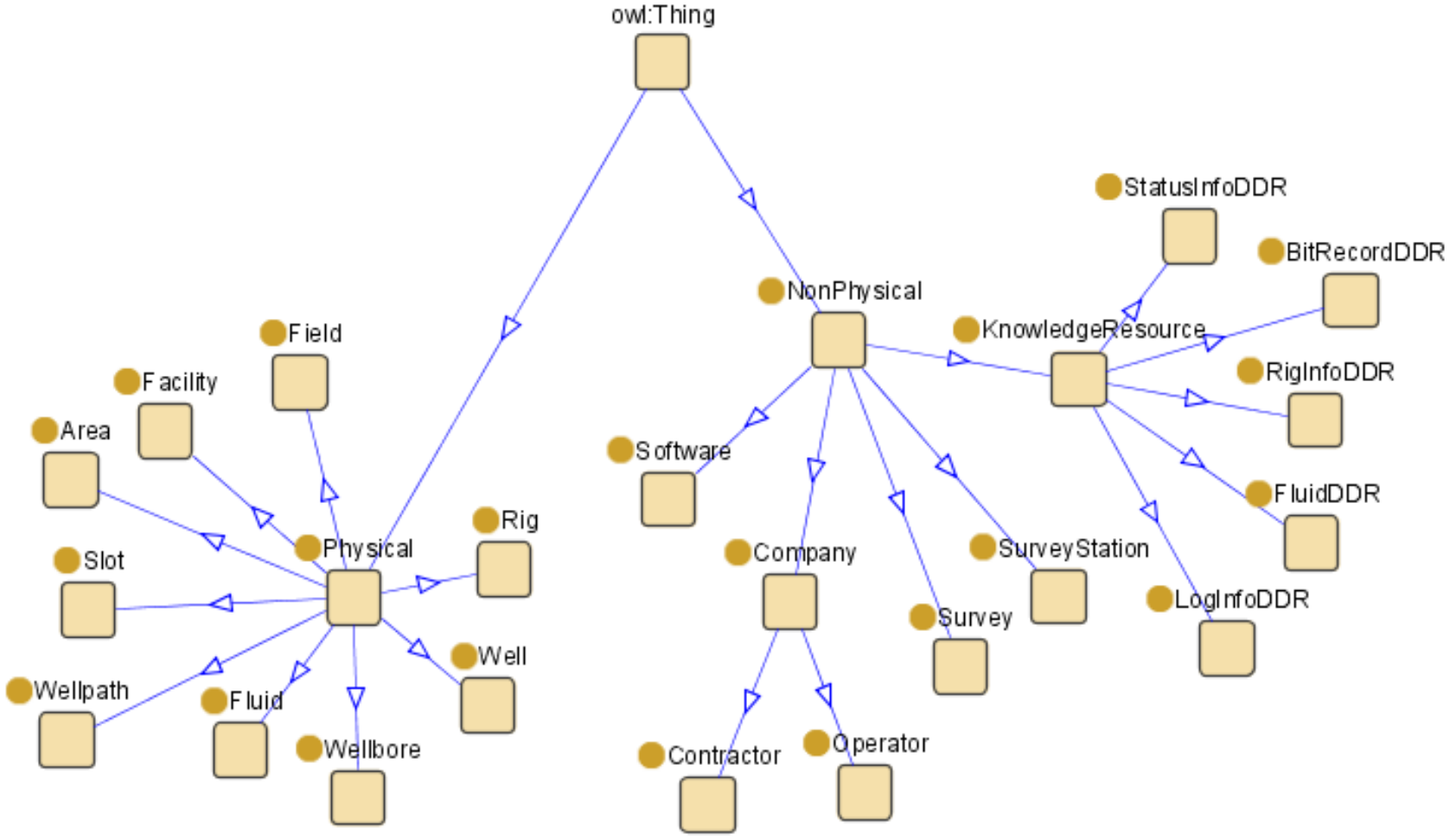
Architecture for a solution that exploits W3C technology



Implemented solution



The OWL DDR ontology



Mappings

```
<mapping id="Wellbore(Class)">
  <CQ string="Wellbore(getWellbore($ID))"/>
  <SQLQuery string="SELECT ID FROM wellbore"/>
</mapping>

<mapping id="tvd(Property)">
  <CQ string="tvd(getSurveyStation($Wellbore,$MD),$tvd)"/>
  <SQLQuery string="SELECT DISTINCT(Wellbore) AS Wellbore,
                    MD, tvd FROM trajectoryStation"/>
</mapping>
```

Annotated XSD

```
<xsd:complexType name="cs_drillReportSurveyStation">
  <xsd:annotation>
    <xsd:documentation>WITSML - Trajectory Station
    Component Schema</xsd:documentation>
    <xsd:appinfo>Modified-in-version=1.4.0,
    By-issue=1.3.1-33, Change=Added</xsd:appinfo>
    <xsd:appinfo>
    <sparql_query:quontoquery>
    SELECT $md $tvd $azimuth WHERE { $x :md $md .
$x :tvd $tvd .  $x :azimuth $azimuth }
    </sparql_query:quontoquery>
    </xsd:appinfo>
  </xsd:annotation>
  <xsd:sequence>
....
```

What did we learn?

- We managed to autogenerate a correct DDR report from realistic data
- DDR was well suited for a simple proof of concept
  - A good case for a MSc project
  - A well-structured terminology
- Extracting a useful ontology from the DDR XSD:
  - Straightforward with domain experts available
  - Time-consuming
- OWL-QL was sufficient
- The QuOnto technology was immature

How can we exploit this further?

- DDR is not an ideal case for OBDA
  - The reporting format is fixed and unlikely to change
  - There degree of integration is too low
  - All companies have implemented procedures for DDR generation
- An ideal case would be environmental data reports
  - The understanding of the domain is likely to change
  - Requires integration from many sources
  - There is a need for standardization of terminologies