Reportnet – a Case Study

Søren Roug
European Environment Agency
What is Reportnet?

- 14 year old system to receive data from countries
  - About 400 reporting obligations
  - Is also an archive with persistent URLs
- Countries upload (mainly) XML files with a web browser to a repository
- Online web questionnaire using XForms for small data amounts
  - Field validation
Automatic QA

- Countries can run automatic QA on XML files before final delivery
  - Uses the XQuery language
  - Rules selected by the schema identifier in the XML file
  - Feedback is HTML

- Very popular
  - Improves quality of deliveries
  - Decreases mundane work for humans
  - Clients very imaginative in devising new rules
Automatic QA popularity problem

- Need for checking against large code lists
- Need for comparison with previous delivery
- Not necessarily same schema
- XQuery and XForms can load code lists in XML format into memory
- But not on this scale
- Idea: Load the data into a database and implement a ReST web service
- Very heterogeneous data
Solution: Conversion to RDF and import

- Very easy for code lists
- When a delivery is made it is automatically converted to RDF and imported into the triple store
  - Can be done with XSL-T
- SPARQL is the ReST web service
- New problem: RDF modelling
- New problem: Small code lists
- New opportunity: linking of vocabularies
New possibilities – new ideas

- What can we do when the deliveries are in the triple store?
  - More QA?
  - Auto-correction?
  - Linking the data?
  - Extract a reference dataset?
  - Inference?
- Post-harvest SPARQL scripts to insert, replace or delete triples
Post delivery pipeline

Reportnet triple store → Create ref. data → Semantic data service

Eurostat data → World Bank

Semantic data service → Website → Charts
Visualisation

Per capita — World Bank CO2 statistics 2008

- America (all income levels)
- India
- Arab World
- China
- European Union
- Japan
- Russian Federation
- North America

Per capita emissions

DaViz
Visualise your data

European Environment Agency