

From the egg to the chicken:

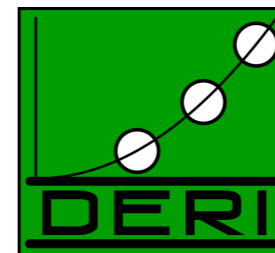
Why building Linked Data applications is hard, and how to make it easier

Richard Cyganiak
Linked Data Research Center, DERI, NUI Galway

Linked Data Camp Vienna, 30 Nov 2009



OÉ Gaillimh
NUI Galway



1. The egg: Linking Open Data
2. The elusive chicken: Linked Data applications
3. Data Consolidators help eggs grow into chickens

2005

The Great Chicken-and-Egg Problem of the Semantic Web

httpRange-14

Tabulator

“Linked Data”

1. Use URIs to identify things
2. Use HTTP URIs so people can look them up
3. When URIs are resolved, provide a description of the thing (in RDF, SPARQL)
4. Include links to related things

Linking Open Data project

“Raw Data Now!”

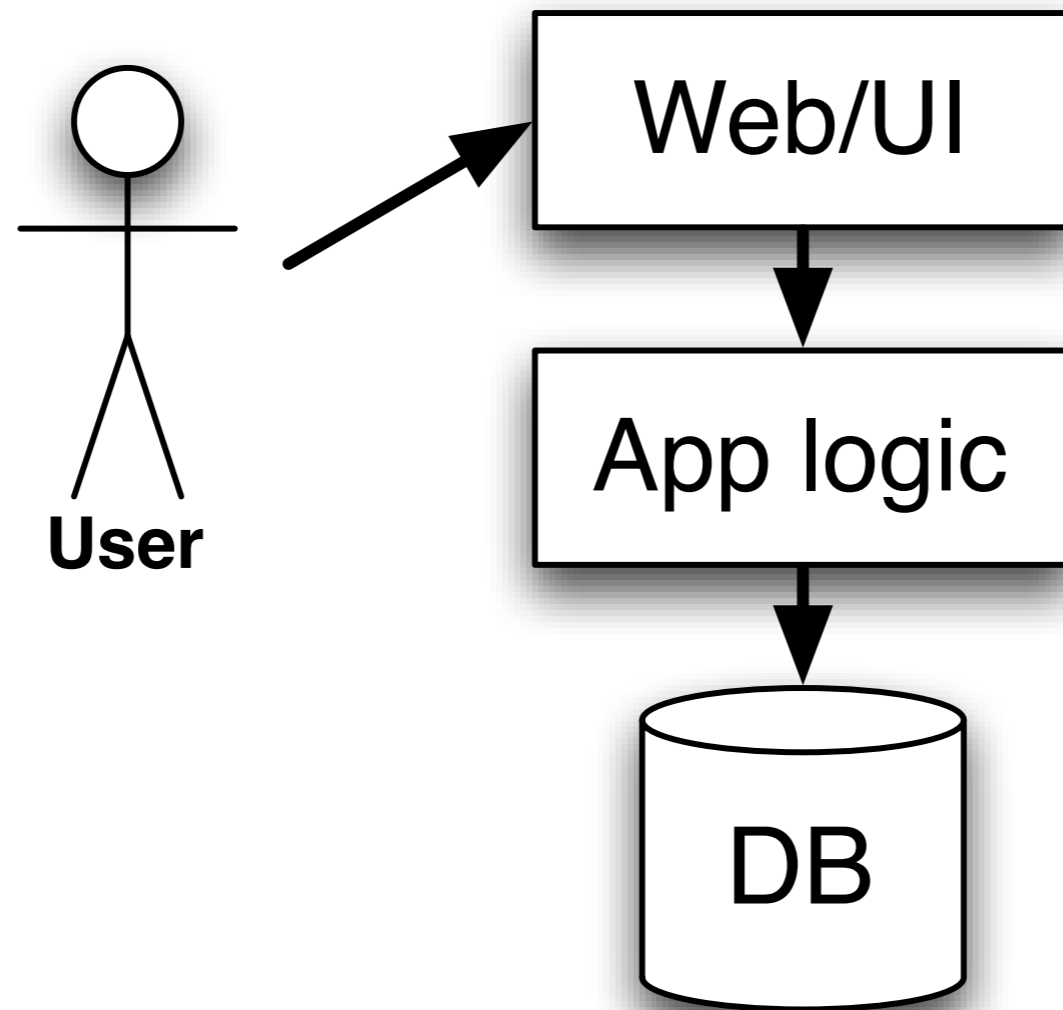
The egg is here

- Millions of RDF files
- Billions of triples
- ~100 organisations and 1000s of individuals
- BBC, New York Times, UK government

So where is
the chicken?

Applications

How we build applications



Building those is easy!

- Everything has its well-known place
- One ID for one thing
- Schema of the data is known
- Quality of the data is known
- Data is trustworthy
- You own the data

Linked Data applications are
faced with heterogeneity.

Sources of heterogeneity in Linked Data

- Location of data
- Identifiers
- Modelling/structure
- Vocabularies
- Quality
- Trustworthiness
- Ownership of data

How to overcome heterogeneity?

- Limit scope of app to 1 vocabulary or 1 data source
- Standardisation (vocabularies)
- Tools and services
- Don't overcome it — expose it to the user
- Let someone else clean up the data

Data Consolidators

What are consolidators?

- Data web sites
- Serving a specific community
- *Owned* by the community – those who are passionate about data
- Aggregate and unify data from different sources
- Re-publish improved data

Consolidators provide value to a community

- Identify and select relevant datasets
- Encourage data owners to make data available
- Filter bad data
- Manage stable, authoritative identifiers
- Single point of access
- Instance and schema reconciliation
- Data analysis

**Consolidators reduce the
cost of building applications**

Examples

- Bio2RDF
- RKB Explorer
- DBpedia
- Linking Open Drug Data
- Google Social Graph API

Tools for building consolidators

- **Remove heterogeneity in data**
- Different location – get it all into one place
- Different identifiers – link discovery, smushing, identifier management
- Different schemas – rules and mapping languages
- Different quality – heuristics for data cleansing

Example tools

- SPARQL CONSTRUCT
- SPARQL I.I aggregates
- Silk – link discovery
- Sindice – unified point of access for search
- Google Social Graph API

Summary

- We have the LOD egg.
- Heterogeneity of web data makes building applications hard.
- Community-operated *data consolidators* make data better and re-publish it
- So apps can innovate without having to deal with all this heterogeneity

