Data wranglers deliver new value from existing data.
The Problem

Data wranglers deliver new value from existing data.
The Problem

- Spreadsheets
- Biz Apps
- Legacy DB
- Log files
- Cloud APIs
- Open Data

Publish to Website
- Mobile App
- Reporting

Visualisation/Infographics
- Open Data

Stats package
The Problem

- Spreadsheets
- Biz Apps
- Legacy DB
- Log files
- Cloud APIs
- Open Data
- Stats package
- Open Data
- Visualisation/Infographics
- Publish to Website
- Mobile App
- Reporting
- Publishing
The V's of Big Data

1. Volume
2. Velocity
3. Variety
Welcome to the Data Hub!

Find data
the Data Hub contains 3324 datasets that you can browse, learn about and download.

Share data
Add your own datasets to share them with others and to find other people interested in your data.

Collaborate
Find out more about working with open data by exploring these resources:
- GetTheData.org
- DataPatterns.org
- Open Data Manual

Who else is here?

Canada
Datasets for http://www.datadotgc.ca/. DataDotGC, which launched in February 2010, is a Canadian, citizen-led effort to promote open data and help share data that has already been... Canada has 523 datasets.

Bibliographic Data
This group comprises open bibliographic datasets according to the Principles on Open Bibliographic Data and a few not yet really open bibliographic datasets. It

Linking Open Data Cloud
This group catalogs data sets that are available on the Web as Linked Data and contain data links pointing at other Linked Data sets. The descriptions of the data sets in this group are... Linking Open Data Cloud has 326 datasets.

Climate Data
Weather, temperature, carbon, water, soil and all other kinds of climate related open data. Including (but not limited to) data about: the atmosphere the weather...

Linking Open Data
A group for Linking Open Data datasets. The initial import of data for this group was done in October 2009 from the list of RDF datasets dumps provided by the W3C Linking Open Data... Linking Open Data has 81 datasets.

OpenSpending
Datasets to be imported to the OpenSpending.org site. Packages listed here will automatically be available for selection in the OpenSpending web importer.
Namespace registry for RDF/SPARQL users

- Side effect: Create database of vocabularies

prefix.cc
namespace lookup for RDF developers

adms

examples: foaf foaf:knows dc,foaf rdfs,dc,foaf,geo.sparql http://xmlns.com/foaf/0.1/name

popular latest about | prefix.cc

NUI Galway OÉ Gaillimh
DERI Galway
Metadata standards

DCAT — Data Catalog Vocabulary

VoID — Vocabulary of Interlinked Datasets

DDI Discovery Vocabulary

Data Cube Vocabulary
D2RQ

Database-to-RDF mapping
### Type of Contract

<table>
<thead>
<tr>
<th>Contract ID</th>
<th>Contractor Name</th>
<th>Type of Contract</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1939</td>
<td>ASAP SOFTWARE EXPRESS INC DELL MARKETING L.P.</td>
<td>Microsoft Enterprise Agreement</td>
<td>04/01/2000</td>
</tr>
<tr>
<td>1940</td>
<td>BMC SOFTWARE DISTRIBUTION INCORPORATED</td>
<td>Remedy Service Desk Maintenance</td>
<td>04/01/2000</td>
</tr>
<tr>
<td>1941</td>
<td>GOVCONNECTION INCORPORATED</td>
<td>Cisco SmartNet</td>
<td>05/01/2000</td>
</tr>
<tr>
<td>1942</td>
<td>ITS CORPORATION</td>
<td>Time &amp; Materials</td>
<td>12/31/2000</td>
</tr>
<tr>
<td>7490</td>
<td>SENET INTERNATIONAL CORPORA</td>
<td>firm fixed price</td>
<td>05/04/2000</td>
</tr>
</tbody>
</table>

**5200 rows**

**Show as:** rows records

**Show:** 5 10 25 50 rows
### RDF Schema Alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the mapping which column to substitute into which node.

**Base URI**: http://lab.linkeddata.deri.ie/test/edit

#### RDF Skeleton

<table>
<thead>
<tr>
<th>Available Prefixes: rdfs dcat foaf dct myns rdf skos +add prefix +manage prefixes</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Organisation URI</th>
<th>Organisation cell</th>
<th>Name cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>foaf:Organization</td>
<td>foaf.name→</td>
<td>Name cell</td>
</tr>
<tr>
<td>add rdf:type</td>
<td>foaf:member→</td>
<td>Job Title cell</td>
</tr>
<tr>
<td>foaf:Person</td>
<td>foaf.title→</td>
<td>Annual pay rate - including taxable benefits</td>
</tr>
<tr>
<td>add rdf:type</td>
<td>:annualPayRate→</td>
<td></td>
</tr>
</tbody>
</table>

#### Define multiple trees

- Add another root node
Neologism

RDF Schema
Vocabulary
Editor and Publishing System
Great stuff, but...
Great stuff, but…

This is still a full-time job!

- Many separate tools that require setup and have their own learning curve
- Command-line interfaces, complex languages
- Web server configurations for publishing data and vocabularies according to best practices
- Difficult to transfer this knowledge to our partners
ODPP: Open Data Publishing Pipeline

Automated discovery

Catalogue

Reformat

Standardise

Publish
• Web-based data publishing platform
• Map input data (relational, CSV, Excel) to semantic standards
• Visual mapping editor and vocabulary search, based on UML
• Already knows all (?) semantic standards that have been published as RDF Schemas
• Special support for generating SKOS, Data Cube and Schema.org
• Under the hood: proven open-source technologies
Under development since early 2013

Trials with LGMA, Galway and Fingal CCs

Public beta in Q3

How would you use ODPP?