Evolving the Internet: the Role of Linked Data

Stefan Decker
They had a different idea:

Vannevar Bush
“The memex stores associative trails”

Doug Engelbart
“Augmenting Human Intellect”

Tim Berners-Lee
“Computers […] make sense of what we are doing”
A Network of Knowledge

- Interconnected
- Universal
- All encompassing

- Enable global and local collaboration
- The right information for the right people at the right time
Two Key Ingredients

1. **RDF – Resource Description Framework**
   - Graph based Data – nodes and arcs
     - Identifies objects (URIs)
     - Interlink information (Relationships)

2. **Vocabularies (Ontologies)**
   - provide **shared understanding** of a domain
   - organise knowledge in a **machine-comprehensible** way
   - give an exploitable **meaning** to the data
Why Graphs and Ontologies?

Wikipedia.org

Cities: Dublin

84421km²

Geo: has Largest City

Geo: Island Of Ireland

Geo: located On

EU: Republic Of Ireland

Geo: has Capital

Geo: area

Gov: has Taoiseach

Person: Brian Cowen

Gov: has Department

IE: Department of Finance
Questions:

Q1: Most important innovation? Single most important measurable output. Topics from Objective 4.3?
Q2: Key justification for high level funding?
Q3: Expected impact of proposal.
Q4: Competitive landscape? Others actors? Alternative approaches?
Linked Data Domains
Who is providing the data?
(few (!) selected examples)
Example: data.gov.uk

Unlocking innovation
Working with UK Public Sector information and data

Latest datasets
2 July  Public servants earning over £150,000 now also covers NDPBs
25 June  Costs and other data about central government websites
18 June  Estimated government workforce including consultants
14 June  More COINS public spending data: now covers 2005 to 2010

What we do

data.gov.uk is a key part of the Government’s Transparency programme for the UK public sector as a whole. Working with Sir Tim Berners-Lee, Professor Nigel Shadbolt and Tom Steinberg and other members of the Cabinet Office Ministers’ new Public Sector Transparency Board, this site seeks to give a way into the wealth of government data becoming available. It’s under constant development and we want to work with you to make it better.

Search Data
Enter keyword(s): Search
- education, NHS, crime, transport, environment

Browse for Data
Random dataset
List all datasets
By Public Body
Common tags
Benefits:

- **data reuse**, new business opportunities.
- **direct access** to public services
- **cut costs** and improve processes in the government agencies and between the agencies and the citizens.
- **increase** the citizen engagement and participation.
The 5 star–plan

★ make resources available on the Web

★★ make resources available as structured data
(for example, an Excel sheet instead of image scan of a table)

★★★ use a non-proprietary format
(e.g., a CSV file instead of an Excel sheet)

★★★★ use URIs to identify things and RDF to represent the data

★★★★★ link your data to other data to provide context
## Efforts staring in many other countries and regions

- EU
- Norway
- Ireland
- Trento
- ...

### Table of Efforts

<table>
<thead>
<tr>
<th>Country</th>
<th>Description</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>Opengov.se is a private initiative aiming to collect public datasets. It contains a catalog of 20 government datasets, their formats and usage restrictions. Data is available only in Swedish.</td>
<td>★★★</td>
</tr>
<tr>
<td>New Zealand</td>
<td>Data.govt.nz is a directory of New Zealand government datasets. This directory could also be characterised as a single point of access for governmental datasets since it includes links to other governmental web sites that contain the actual files.</td>
<td>★★★</td>
</tr>
<tr>
<td>Canada</td>
<td>The Open Data Catalogue of the City of Vancouver (data.vancouver.ca) contains about 20 datasets in CSV, XLS, KML and SHP formats, incl. information about schools, libraries, fire halls, water networks, etc.</td>
<td>★★★★</td>
</tr>
<tr>
<td>United States</td>
<td>Data.gov includes over 700 raw datasets in CVS/TXT, XML and SHP formats and more than 100,000 geo-spatial datasets such as administrative and political boundaries. Further, only recently Tetherless World Constellation at Rensselaer Polytechnic Institute undertook the task to transform a number of the raw datasets to Linked Data.</td>
<td>★★★★★</td>
</tr>
<tr>
<td>UK</td>
<td>Perhaps due to the involvement of Tim Berners-Lee, the data.gov.uk effort is one of the most ambitious initiatives on this list, providing also Linked Data connected with the Linked Open Data cloud. At time of writing, more than 3300 datasets and over 60 applications, developed by SMEs and individuals, are available.</td>
<td>★★★★★★</td>
</tr>
</tbody>
</table>
Linked Governmental Data in use …

- UK Public Sector Information Data Mashup
  http://map.psi.enakting.org/
- Research Funding Explorer
  http://bis.clients.talis.com/
Linked Governmental Data in use ...

- Comparing US and UK Global Foreign Aid

- Crime statistics
  http://www.jenitennison.com/visualisation/crime.html
Bio2RDF integrates some 40 biomedical information resources (such as GO, Uniprot, etc.) or extracts recoded in RDF

RDF Data Access:
http://atlas.bio2rdf.org/sparql
<table>
<thead>
<tr>
<th>Resources included in Bio2RDF</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Resource</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO</td>
<td>KEGG</td>
</tr>
<tr>
<td>OMIM</td>
<td>HGNC</td>
</tr>
<tr>
<td>PUbMed</td>
<td>INOH</td>
</tr>
<tr>
<td>GeneID</td>
<td>IProClass</td>
</tr>
<tr>
<td>UniProt</td>
<td>MGI</td>
</tr>
<tr>
<td>UniRef</td>
<td>CellMap</td>
</tr>
<tr>
<td>UniParc</td>
<td>BioPAX</td>
</tr>
<tr>
<td>Kegg Pathway</td>
<td>InterPro</td>
</tr>
<tr>
<td>CPATH</td>
<td>Pfam</td>
</tr>
<tr>
<td>Reactome</td>
<td>PROSITE</td>
</tr>
<tr>
<td>Biocyc</td>
<td>Protein</td>
</tr>
<tr>
<td>MeSH</td>
<td>SID</td>
</tr>
<tr>
<td>PDB</td>
<td>CID</td>
</tr>
<tr>
<td>CPD: Kegg Ligand for chemical compound</td>
<td>PubChem</td>
</tr>
<tr>
<td>GL: Kegg Ligand for carbohydrate structure</td>
<td>UniSTS</td>
</tr>
<tr>
<td>EC</td>
<td>Homologene</td>
</tr>
<tr>
<td>RN Kegg Ligand for chemical reaction</td>
<td>DBpedia</td>
</tr>
<tr>
<td>DR: Kegg Ligand for drugs</td>
<td>OBO CheBi</td>
</tr>
<tr>
<td>Taxonomy: NEWT</td>
<td>Affymetrix</td>
</tr>
<tr>
<td>PID</td>
<td>Biocarta</td>
</tr>
</tbody>
</table>
Example: BBC

Nirvana

50s US grunge band

Latest Tracks Played On The BBC

About A Girl
BBC 6 Music | Shaun Keaveny 13/07/2010

Smells Like Teen Spirit
BBC 6 Music | Lauren Laverne 01/07/2010

On A Plain
BBC 6 Music | Shaun Keaveny Are you a chimp or a bonobo?

Love Buzz
BBC 6 Music | Chris Hawkins 23/06/2010

Come As You Are
BBC Radio 2 | Janice Long 22/06/2010

Biography

Nirvana was an American rock band that was formed by singer-guitarist Kurt Cobain and bassist Krist Novoselic in Aberdeen, Washington in 1987. Nirvana went through a succession of drummers, the longest-lasting being Dave Grohl, who joined the band in 1990.

The band established itself as part of the Seattle music scene, releasing its first album Bleach for the independent record label Sub Pop in 1989. After signing to major label DGC Records, Nirvana found unexpected success with “Smells Like Teen Spirit”, the band's lead single from its second album Nevermind (1991). Subsequently, Nirvana entered into the mainstream, bringing along with it a subgenre of alternative rock called grunge. As Nirvana's frontman, Kurt Cobain himself referred to the media as the "spokesman of a generation," with Nirvana being considered the "flagship band" of Generation X. Cobain was uncomfortable with the attention and placed his focus on the band's music, believing the band's message and artistic vision to have been misrepresented by the public, challenging the band's audience with its third studio album in Utero (1993).

Nirvana's brief but intense following the death of Cobain in 1994, but the band's influence and

Audio Previews From Latest Album Review

Bleach: Deluxe Edition
1 Blew
2 Floyd the Barber
3 About a Girl

Played By

Siros December 2008

This XML file does not appear to have any style information associated with it. The document tree is shown below.

- <rdf:RDF>
  - <rdf:Description rdf:about="http://musicartists/sb11f4a-a62d-471e-81f6-a69a8278c7da.rdf"/>
    - <foaf:primaryTopic rdf:resource="http://musicartists/sb11f4a-a62d-471e-81f6-a69a8278c7da#artist"/>
    - <foaf:Description>
      - <xmlns:MusicArtist rdf:about="http://musicartists/sb11f4a-a62d-471e-81f6-a69a8278c7da#artist"/>
        - <rdf:type rdf:resource="http://purl.org/eology/mo/MusicGroup"/>
        - <foaf:name>Nirvana</foaf:name>
        - <v:sortLabel>Nirvana</v:sortLabel>
        - <v:comment>90s US grunge band</v:comment>
        - <bioevent>
          - <bio:Birth>
            <bio:chordate rdf:datetype="http://www.w3.org/2001/XMLSchema#date"/>
          </bio:Birth>
        </bioevent>
      </xmlns:MusicArtist>
    </foaf:Description>
  </rdf:Description>
</rdf:RDF>
Example: New York Times

The New York Times

Linked Open Data BETA

data.nytimes.com

For the last 150 years, The New York Times has maintained one of the most authoritative news vocabularies ever developed. In 2009, we began to publish this vocabulary as linked open data.

The Data

As of 13 January 2010, The New York Times has published approximately 10,000 subject headings as linked open data under a CC BY license. We provide both RDF documents and a human-friendly HTML versions. The table below gives a breakdown of the various tag types and mapping strategies on data.nytimes.com.

<table>
<thead>
<tr>
<th>Type</th>
<th>Manually Mapped Tags</th>
<th>Automatically Mapped Tags</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>People</td>
<td>4,978</td>
<td>0</td>
<td>4,978</td>
</tr>
<tr>
<td>Organizations</td>
<td>1,489</td>
<td>1,592</td>
<td>3,081</td>
</tr>
<tr>
<td>Locations</td>
<td>1,910</td>
<td>0</td>
<td>1,910</td>
</tr>
<tr>
<td>Descriptors</td>
<td>498</td>
<td>0</td>
<td>498</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10,467</td>
</tr>
</tbody>
</table>

Browse Individual data records:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

SKOS Files

Download all of the data records as SKOS Files.

- People
- Organizations
- Locations
- Subject Descriptors
Example: Facebook

Build the social and personalized web.

Facebook's powerful APIs enable you to create social experiences to drive growth and engagement on your website.

Add Facebook to my site

Apps and more
- Mobile apps: Add Facebook to your mobile app
- Apps on Facebook.com: Integrate with our core experience
- Open Source: Build with the same technologies we use

What's new
- Expanding Your Application's User Base
  Yesterday by Huan Yang
- Data Model Migration Under Way
  June 30 by Arun Vijayvergiya
- Reminder: Upgrade to New Data Model by June 30
  June 22 by Bret Taylor

Showcase
- CNN
- The New York Times
- Pandora
- WSJ.com
- Simply Hired
- Levi's

See how companies make their sites and social with Facebook
Example: Best Buy

XML file does not appear to have any style information associated with it. The document tree is shown below:

```
<owl:RDF xmlns:base="http://stores.bestbuy.com/semanticweb">
  <owl:AnnotationProperty rdf:about="http://www.w3.org/2001/vcard-rdf/3.0#ADR="/>
  <owl:AnnotationProperty rdf:about="http://www.w3.org/2001/vcard-rdf/3.0#Street="/>
  <owl:AnnotationProperty rdf:about="http://www.w3.org/2001/vcard-rdf/3.0#Pcode="/>
  <owl:AnnotationProperty rdf:about="http://www.w3.org/2001/vcard-rdf/3.0#City="/>
  <owl:AnnotationProperty rdf:about="http://www.w3.org/2001/vcard-rdf/3.0#Country="/>
  <owl:AnnotationProperty rdf:about="http://www.w3.org/2001/vcard-rdf/3.0#TEL="/>
  <owl:AnnotationProperty rdf:about="http://www.w3.org/2001/vcard-rdf/3.0#EMAIL="/>
  <owl:Ontology rdf:about="/">
    <owl:Imports rdf:resource="http://purl.org/goodrelations/v1="/>
  </owl:Ontology>

  <g:BusinessEntity rdf:ID="#BusinessEntity">Best Buy Co., Inc. <g:legalName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Best Buy Co., Inc.</g:legalName>
    <rdfs:seeAlso rdf:resource="http://bestbuyinc.com"/>
    <v:Card:A:DR rdf:parseType="Resource">
      <v:Card:Street>7601 Penn Avenue South</v:Card:Street>
      <v:Card:Pcode>55423</v:Card:Pcode>
      <v:Card:City>Richfield</v:Card:City>
    </v:Card:A:DR>
    <v:Card:TEL rdf:parseType="Resource">
      <rdf:value>1-612-292-6397</rdf:value>
      <rdf:resource">http://www.w3.org/2001/vcard-rdf/3.0#work"/>
      <rdf:resource">http://www.w3.org/2001/vcard-rdf/3.0#voice"/>
    </v:Card:TEL>
  </g:BusinessEntity>
  <g:Offer rdf:resource="#BestBuyStore_840_offering"/>
```

What to do with this data?
Ireland is the third-largest island in Europe and the twentieth-largest island in the world. It lies to the northwest of continental Europe and is surrounded by hundreds of islands and islets. To the east of Ireland is the Island of Great Britain, separated from it by the Irish Sea. The Republic of Ireland covers five-sixths of the island. Northern Ireland, a part of the United Kingdom, covers the remainder and is located in the northeast of the island.

Ireland is the third-largest island in Europe and the twentieth-largest island in the world. It lies to the northwest of continental Europe and is surrounded by hundreds of islands and islets. To the east of Ireland is the Island of Great Britain, separated from it by the Irish Sea. The Republic of Ireland covers five-sixths of the island. Northern Ireland, a part of the United Kingdom, covers the remainder and is located in the northeast of the island. The population of Ireland is estimated to be 6.2 million.
Exploiting and enhancing Linked Data: Semantic Collaboration
Vision: Networked Collaborative Knowledge

- Free the Silos - Bridge the Islands
- Interconnected Knowledge
- Ensure local and global Collaboration
Human burden to hold everything together

Organising a project showcase at a conference: involved Documents, contacts, time-schedule

E-Emails

Meeting Presentations

Travel Website

Time Schedule

Conference Website
Personal Semantic Web: a semantically enlarged intimate supplement to memory

Social protocols and distributed search
Social semantic peers

friend
acquaintance
colleague

Semantic Desktop

Digital Enterprise Research Institute
www.deri.ie
Personal Information Model

E-Emails

Meeting Presentations

Travel Website

Conference Website

Interconnect disperse Information

http://www.semanticdesktop.org/ontologies/
Semantic Search

http://nepomuk.kde.org/
Link Text to Knowledge

Semantic Notes - smile meeting minutes

smile meeting minutes

date/time/location: wed 2008-08-06 / 3-5 pm / meeting room A

attendance: Siegfried, Sven, Vit, Tudor, Simon, Laura, Ismael, Oszkar, Brian Davis
regrets: Knud (holiday)
scribe: Laura

Review of last meeting minutes
no issues

Trip Reports

meeting, minute, smile

Warning: this is a premature representation of Nepomuk resources for testing purposes!

Brian Davis
Type: PersonContact

Relations:

- has tag: eHealth
- preferred label: Brian Davis
- fullname: Brian Davis
- nameGiven: Brian
- nameFamily: Davis
- hasEmailAddress: mailto:brian.davis@deri.org
- addressLocation: Research Assistant
- addressLocation: Ph.D. Student
- addressLocation: Semantic Web
- addressLocation: DERI
- addressLocation: nepomuk:2091572166
- label: Brian Davis

http://smile.deri.ie/projects/semn
Browse Interconnected Knowledge
Linux/KDE Semantic Desktop

- **Open Source**
  - Involvement of KDE developers – Early Adopters
  - Semantic OS

- **Core Technology of the Semantic Desktop:**
  - PIMO, RDF Data store, Wrapper

- **Part of the official KDE 4**
  - Millions of installations
  - Semantic Web Technology on every KDE computer
  - Virulent effect
  - Creation of a market
Semantic Desktop Applications
most popular *Electronic Communication* means

**Virtual Workplace**
- Collaborative Environment
- Knowledge creation, management and sharing

**Lacks clear structure & real semantics**
- Email Tracking
- Email Classification
- Email Retrieval
- *Email Overload*
Semantic Email

- Annotated Email
  i. *Thread metadata* - Email Sequence, Social, Temporal Metadata

  ![Diagram of email sequence with RDF symbols]

  ii. *Content metadata* – Action Items in written dialogue (based on Speech Act Theory)

  ![Diagram of email content with RDF symbols]
Email Action Items

- **Speech Act Model:** [Action, Object, Subject]

- **Example**  “...Please make sure you have the document ready!..”
Email Workflows

- Email Conversations carry out concurrent, implicit, well-formed *Workflows*

- Email Action Item = Start/Continuation of a Workflow

- Example:
  - *Request* Meeting
  - *Amend* Meeting properties
  - *Approve* Meeting
  - *Announce* Meeting

- *Workflow Artefacts* – Events, Tasks, People, Projects…
Integrating Desktop Items

• Linking Emails in Threads

• Linking Emails with generated Tasks/Meetings
What the @#$%^ are we doing?
Establishing Interoperability… for sending data (internetworking)

1. Physical Layer
2. Data Link Layer
3. Network Layer
4. Transport Layer
5. Session Layer
6. Presentation Layer
7. Application Layer
Future Internet Topics

- Smart Health
- Smart Energy
- Security
- Foundations of Trust
- Search
- Smart Cities
- Search
- Future Content Networks
- Real World Internet (Sensor Networks)
Data Interoperability

- Internetworking is solved - sending data

- Interdataworking is the new challenge - understanding data
Current Challenge: Establishing Interoperability... for using data

1. Physical Layer
2. Data Link Layer
3. Network Layer
4. Transport Layer
5. Session Layer
6. Presentation Layer
7. Application Layer
Linked Data to the rescue....

- Physical Layer
- Data Link Layer
- Network Layer
- Transport Layer
- Session Layer
- Presentation Layer
- Application Layer

Linked Data
Challenges

- What does it mean to add a layer?
- Standardise access methods
- Clarify relationships to other layers
- Plan for the future (network extensions)
- Put security, trust, identity, services on top of it
Conclusion
A **global** task ...

- many trends point into the same direction
- supported by many individual efforts
- collaboration across domains / disciplines is essential (Web Science)
- Future Internet requires international collaboration and dedicated effort