



Weaving a Semantic Web across OSS repositories

a spotlight on btsslink, UDD, SWIM

O. Berger and V. Vlasceanu
Institut TELECOM / PFTCR

WOPDASD 2009

Skövde 2009/06/06





In the previous episodes

Cross-repository Integration with RDF and OWL

James Howison
and Kevin Crowston



Hi everyone,
hope the workshop goes well.
Wish I could be there.
Think of me, though, I'm at home
trying to finish this dissertation :)
Thanks to Kevin for presenting.

Cheers,
James



Quoting "Cross-repository data linking with RDF and OWL"

Conclusions :

- *Linking various repositories is hard but achievable*
- *RDF integration offers a way for Repositories to collaborate without enforced standardization*
- *Expose your data to RDF; setup a SPARQL frontend*
- *Take, and extend the flosscomms vocab, together with DOAP and SIOC.*
- <http://floss.syr.edu/ontologies/2008/>

■ **We do agree ! ... and try currently to implement that**



Context : Helios project

- <http://helios-platform.org/>
- Project French R&D cluster System@tic of Paris-Region area (FLOSS special interest group)
- 2 years project (start date : 2008/09/01)
- Academia + Industry
- *Application Lifecycle Management* with Open Source tools



WP3 : bugtracker interoperability

- **Bug traceability**
- **Interoperability between bugtrackers**
- **Final goal : ease of monitoring bug links all over the Open Source ecosystem :**
 - Many packagers for similar software in distributions
 - Increase communication beyond bugtracker silos
- **Application to Internal bugtracker (integrators) monitoring 3rd parties (OSS projects) bugs**



Semantic Web and LinkedData hype

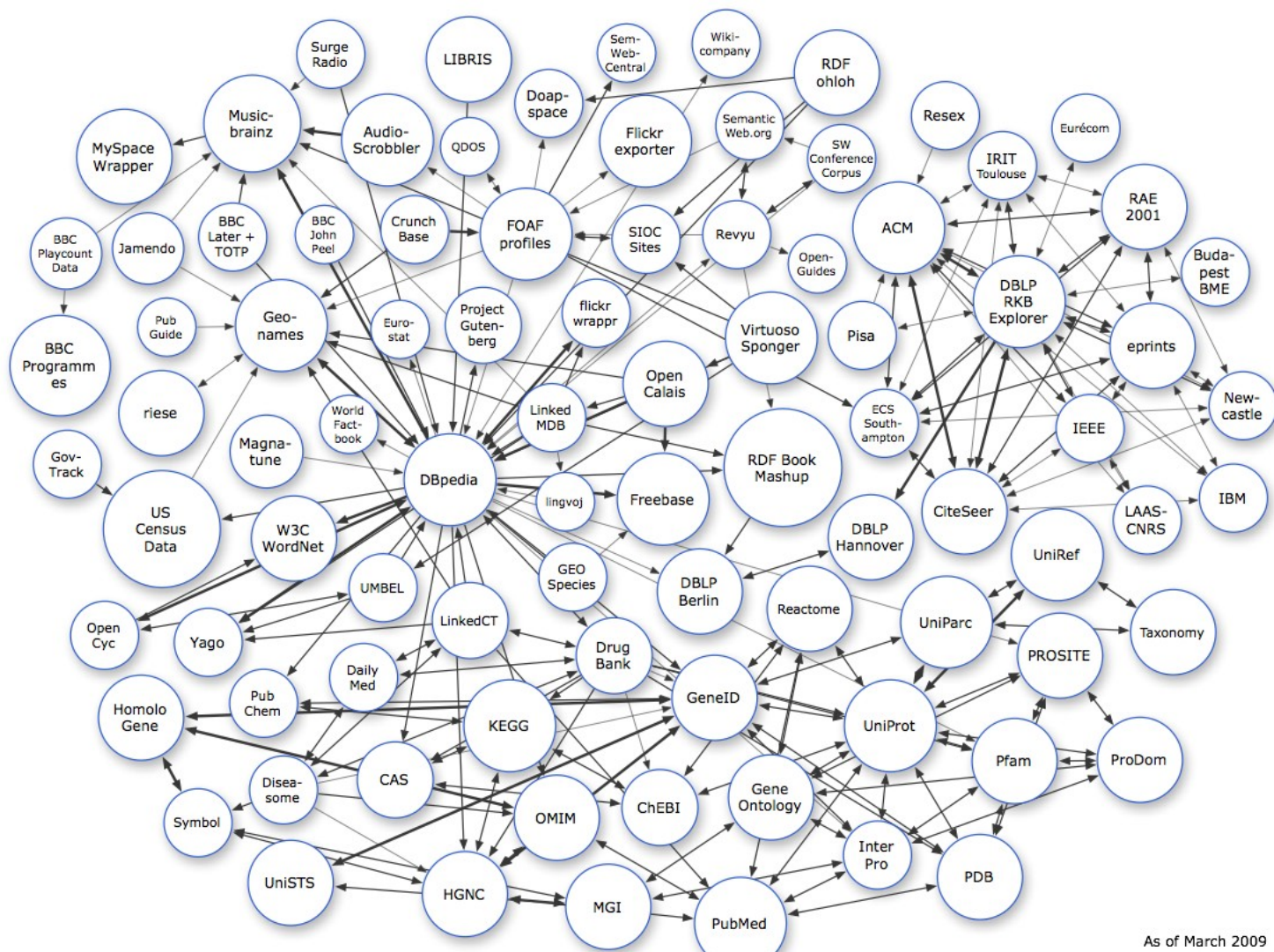
- <http://linkeddata.org/>
- (Open ?) Linked Data
- *“Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods.*

More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF."

- <http://data.gov/> ...



LinkedData



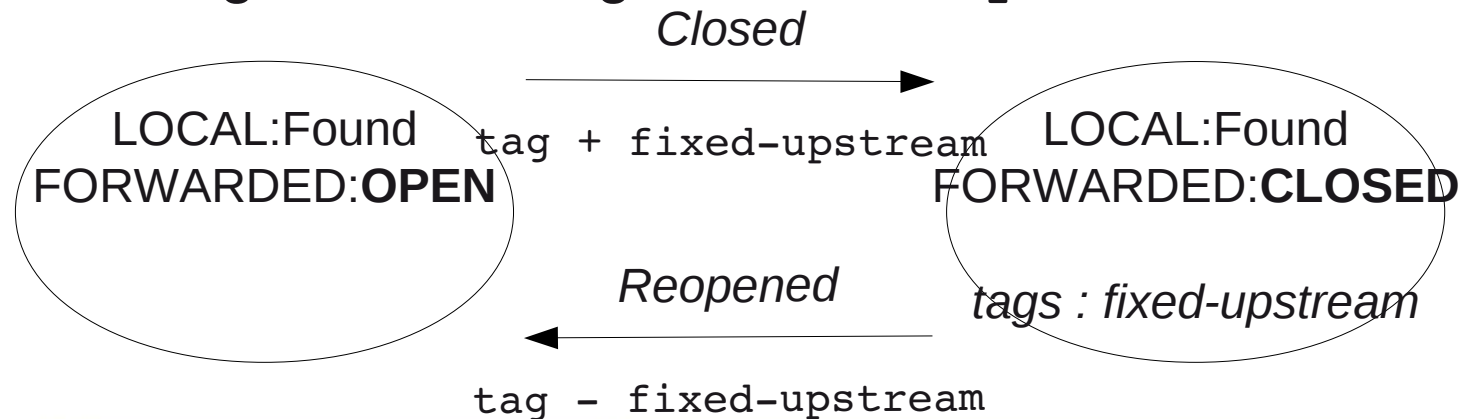
As of March 2009

Source: http://www4.wiwiwiss.fu-berlin.de/bizer/pub/lod-datasets_2009-03-05.html



Bts-link : monitoring bug status change

- Debian tool for package maintainers
- Bug links (**forwarded-to**):
 - Distribution (Debian) package bugs
 - “Upstream” project bugtrackers bugs
- **Monitoring status changes on upstream bugs**
 - Notification for Debian packagers (or people monitoring Debian bugs)
 - Tags Debian bugs : `fixed-upstream`





Improving bts-link

- <http://bts-link.alioth.debian.org/>
- Supports lots of upstream bugtracker types :
bugzilla (and issuezilla), gnats, launchpad, mantis, savane (from savanah), sourceforge trackers, trac, gforge (and fusionforge most probably), google code
- At the moment works only over debugs
- Make it more generic and not Debian specific
- Either :
 - custom bugtracker data gatherer
 - or standard for bugtracker data interchange : none yet



Bug/Issue representation

- Interchange of data representing Bugs/Issues
- Ontology, Schema (Semantik Web standards)
- Standard proposed and community build :
baet1e <http://code.google.com/p/baetle/>
- Reuse of EvoOnt BOM
<http://www.ifi.uzh.ch/ddis/evol>
- Semantic web techniques : extensible
- Currently working on modeling : DebBugs, BugZilla, Mantis, LaunchPad, Gforge/FusionForge, Trac (and more ?)



UDD : Ultimate Debian Database

<http://udd.debian.org/>

- Yet another archive : Debian and Ubuntu facts
- By developers for developers ?
- DB schema
 - persons (carnivore)
 - uploaders
 - upload history
 - source packages & binary packages
 - popcon
 - Bugs
- Postgres relational DB (300 Mb)
- Dumps available (snapshots, no history ?)
- Privacy issues ?



UDD queries

```
SELECT * FROM
  carnivore_emails AS e,
  ( ( carnivore_names AS n
      LEFT OUTER JOIN
      carnivore_login AS l ON l.id = n.id)
    LEFT OUTER JOIN carnivore_keys k
    ON l.id = k.id)
  WHERE e.id = n.id
  AND n.name = 'Olivier Berger';
```

| id | email | id | name | id | login | id | key | key_type |
|-----|-------------------------------|-----|----------------|-------------|-------------|-------------|-------------|-------------|
| 925 | olivier.berger@it-sudparis.eu | 925 | Olivier Berger | <i>NULL</i> | <i>NULL</i> | <i>NULL</i> | <i>NULL</i> | <i>NULL</i> |



Looking for Jesus

| id | email | name | login | key | key_type |
|-----|-------------------------|----------|-------|--|----------|
| 388 | xxx@hispalinux.es | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | keyring |
| 388 | xxx@hispalinux.es | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | ldap |
| 388 | xxx@hispalinux.es | Jesus C. | mooch | BB6423391CAA7064E4297E1866FC1D7F86946D69 | removed |
| 388 | xxx@cc.hut.fi | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | keyring |
| 388 | xxx@cc.hut.fi | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | ldap |
| 388 | xxx@cc.hut.fi | Jesus C. | mooch | BB6423391CAA7064E4297E1866FC1D7F86946D69 | removed |
| 388 | xxx@gmail.com | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | keyring |
| 388 | xxx@gmail.com | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | ldap |
| 388 | xxx@gmail.com | Jesus C. | mooch | BB6423391CAA7064E4297E1866FC1D7F86946D69 | removed |
| 388 | xxx@hispalinux.es | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | keyring |
| 388 | xxx@hispalinux.es | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | ldap |
| 388 | xxx@hispalinux.es | Jesus C. | mooch | BB6423391CAA7064E4297E1866FC1D7F86946D69 | removed |
| 388 | xxx@debian.org | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | keyring |
| 388 | xxx@debian.org | Jesus C. | mooch | A103263A1923373E122746A06DEA9281DA034410 | ldap |
| 388 | xxx@debian.org | Jesus C. | mooch | BB6423391CAA7064E4297E1866FC1D7F86946D69 | removed |
| 510 | xxx@barrapunto.com | Jesus B. | jgb | 62811BD865C6AFB944BBA275377165C5870AE0C9 | keyring |
| 510 | xxx@barrapunto.com | Jesus B. | jgb | 62811BD865C6AFB944BBA275377165C5870AE0C9 | ldap |
| 510 | xxx@computer.org | Jesus B. | jgb | 62811BD865C6AFB944BBA275377165C5870AE0C9 | keyring |
| 510 | xxx@computer.org | Jesus B. | jgb | 62811BD865C6AFB944BBA275377165C5870AE0C9 | ldap |
| 510 | xxx@debian.org | Jesus B. | jgb | 62811BD865C6AFB944BBA275377165C5870AE0C9 | keyring |
| 510 | xxx@debian.org | Jesus B. | jgb | 62811BD865C6AFB944BBA275377165C5870AE0C9 | ldap |
| 510 | xxx@gsyc.esctet.urjc.es | Jesus B. | jgb | 62811BD865C6AFB944BBA275377165C5870AE0C9 | keyring |





Other queries

- **select b.source, count(*) from bugs as b where b.submitter like '%olivier.berger@it-sudparis.eu%' group by b.source order by count(*) desc**
- **select source, count(*) from (select * from bugs union select * from archived_bugs) as all_bugs where submitter like '%olivier.berger@it-sudparis.eu%' group by source order by count(*) desc**



SWIM : a Semantic Web enabled Issue Manager

- Mandriva, result from Nepomuk EC FP 6 project
- HTML + RDF for Mandriva facts :
 - Applications / packages
 - Experts (developers, ...)
 - Bugs
 - Annotations by users
 - ...
- Reuse ontologies also used on Desktop tools (KDE 4.2)
- <http://club.mandriva.com/xwiki/bin/view/swim/>
- soon : Doc4@Mandriva



UDD going semantic

- Using D2R :
<http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/>
- Relational → RDF
- HTML + RDF (Content negociation)
- Export Debian facts for Semantic Web
- Prototype :
 - FOAF
 - EvoOnt + *HeliosBT*
- USDD : Ultimate *Semantic* Debian Database ?
- Prototype online (ask me)



USDD query

Mapping SPARQL to SQL:

```
SELECT DISTINCT * WHERE {  
  ?s foaf:name "Olivier Berger"  
}
```

```
SELECT DISTINCT  
"d2r_t_bugsubmitters"."submitter" FROM  
"d2r_t_bugsubmitters" WHERE  
"d2r_t_bugsubmitters"."email_phrase" = 'Olivier  
Berger'
```

■ See demo...



Ohloh going semantic

- RDFohloh <http://rdfohloh.wikier.org/>
- Exporting facts about projects and developers registered at Ohloh as RDF
- See demo



Proposals, ideas, future work...

- All archives exporting RDF data
- All Bugtrackers exporting RDF data
- All forges exporting RDF data
- Interlinking all that together with users' data
- Collaboration on Python BTS client tools



Demo



Copyright, etc.

- This work is Copyright 2009 by Institut TELECOM, Olivier Berger and Valentin Vlasceanu, published under a Creative Commons ShareAlike license
- This work was conducted in the HELIOS project with support of the Paris-Region authority in the frame of the **System@tic** cluster
- The trip to come present that work was partially “carbon-compensated” with the **Action Carbone** program.