

Using RDF metadata for traceability among projects and distributions

Olivier Berger <<mailto:obergix@debian.org>>
Debian + Télécom SudParis

Thursday 04/04/2013
Distro Recipes - Paris



Quick Introduction

Short bio

Olivier BERGER

`<mailto:olivier.berger@telecom-sudparis.eu>`

`<mailto:obergix@debian.org>`

Research Engineer at TELECOM SudParis, expert on software development forges, and interoperability in Libre Software development projects. Contributor to FusionForge, **Debian**, etc.



How much duplication ?

- Think about all the duplicate bug reports, and the number of people involved. . . usually not the ones who can help, btw
- Can you navigate between all of these ?
- Does google search help ?



Some issues

- Universal Free Software *Project description format*; (*no*)
- Universal distribution *Package description format*; (*no*)
- Common Semantics? (*probably*)
- How to *inter-link* documents about the same program :
 - Bug reports (upstream and in all distros)
 - Launchpad
 - Security advisories
 - Debian's Homepage: control field
- Are we maintaining Free Software distributions in close *silos*?
- External directories (FSF, Freshmeat/Freecode, JoinUp, etc.)
- What's usually wrong with [XML | JSON | YAML | RFC822] format?



An approach

Linked Open Development Meta-Data

- Let's try and make as much *distro facts* as possible available to *humans + machines*?
- Adopting the 5 ★ Open Data principles using RDF for distro meta-data :

| | |
|-------|---|
| ★ | make your stuff available on the web |
| ★★ | make it available as structured data |
| ★★★ | non-proprietary format |
| ★★★★ | use URLs to identify things, so that people can point at your stuff (RDF) |
| ★★★★★ | link your data to other people's data to provide context (Linked RDF) |



Project/program/package traceability over the FLOSS ecosystem

- Assembling a graph of descriptions of packages/projects published as Linked Data (DOAP or ADMS.SW) on their forges / project portals.

For instance :

- For Debian, from the Debian PTS (already *Linked Data* proof)
- For Apache, Gnome, Pypi, from **DOAP** files (not yet all *Linked Data*, but close)
- ... *Add your preferred upstream* ...
- Consumed by developer/maintainer/packager tools : following links between packages, (and their bugs, security alerts), all in **semantic interoperable** meta-data formats (RDF)!



Matching project/package descriptions

Example (SPARQL query to match packages by their homepages)

```
PREFIX doap: <http://usefulinc.com/ns/doap>
```

```
SELECT * WHERE
```

```
{  
  GRAPH <http://packages.qa.debian.org/>  
  {  
    ?dp doap:homepage ?h  
  }  
  GRAPH <http://projects.apache.org/>  
  {  
    ?ap doap:homepage ?h  
  }  
}
```

“Semantic query” : Trying to match source packages in Debian whose upstream *project's homepages* match those of the Apache project's DOAP descriptors.



Matching packages

Results : 62 matching Apache projects packaged in Debian (for which maintainers did set the Homepage Control field consistently).

Example (Matching upstream Apache project homepages with Debian source packages')

| dp | h | ap |
|------------------------------|--------------------------|--------------------------|
| ivy | ant.a.o/ivy/ | ant.a.o/ivy/ |
| apr | apr.a.o/ | apr.a.o/ |
| apr-util | apr.a.o/ | apr.a.o/ |
| libcommons-cli-java | commons.a.o/cli/ | commons.a.o/cli/ |
| libcommons-codec-java | commons.a.o/codec/ | commons.a.o/codec/ |
| libcommons-collections3-java | commons.a.o/collections/ | commons.a.o/collections/ |
| libcommons-collections-java | commons.a.o/collections/ | commons.a.o/collections/ |
| commons-daemon | commons.a.o/daemon/ | commons.a.o/daemon/ |
| libcommons-discovery-java | commons.a.o/discovery/ | commons.a.o/discovery/ |
| libcommons-el-java | commons.a.o/el/ | commons.a.o/el/ |
| libcommons-fileupload-java | commons.a.o/fileupload/ | commons.a.o/fileupload/ |
| commons-io | commons.a.o/io/ | commons.a.o/io/ |
| commons-jci | commons.a.o/jci/ | commons.a.o/jci/ |
| libcommons-launcher-java | commons.a.o/launcher/ | commons.a.o/launcher/ |
| ... | ... | ... |

Matching program *names* gives more results but is ambiguous



My current experiment

mining project descriptions

- Running on my laptop ATM
- Currently use Python to harvest meta-data from DOAP files for :
 - Gnome
 - Apache
 - Pypi.python.org
 - Debian
 - (*add you ?*)
- Down to a virtuoso *Triple store* : > 2.5 M triples ATM
- A python app to perform queries
- May be published as a public service some day



Adding RDF to the Debian Package Tracking System

<http://packages.qa.debian.org/>

Debian Package Tracking System

debian **apache2** source package

jump to

general

source: [apache2](#) (source, httpd)

version: 2.2.22-12

maint: [Debian Apache](#)
 Maintainers, [Stefan Fritsch](#) (u), [Steinar H. Gunderson](#) (u), [Arno Töll](#) (u)

arch: any all

std-ver: 3.9.3

VCS: [Git](#) ([browse](#))

versions

oldstable: [2.2.9-10+lenny12](#)

stable: [2.2.16-6+squeeze8](#)

stable-sec: [2.2.16-6+squeeze1](#)

s-p-u: [2.2.16-6+squeeze1](#)

testing: [2.2.22-12](#)

unstable: [2.2.22-12](#)

exp: [2.4.2-2](#)

binaries

- [apache2](#) (33 bugs: 0, 18, 13, 2)
- [apache2-dbg](#) (0 bugs: 0, 0, 0, 0)
- [apache2-doc](#) (0 bugs: 0, 0, 0, 0)
- [apache2-mpm-event](#) (1 bugs: 0, 1, 0, 0)
- [apache2-mpm-itk](#) (2 bugs: 0, 0, 0, 0)

todo

- [Lintian](#) reports **1 error** and **49 warnings** about this package. You should make the package *lintian* clean getting rid of them.
- [Build log checks](#) report **2 warnings** about this package.
- A new upstream version is available: [2.4.3](#), you should consider packaging it.
- The BTS contains [patches fixing 8 bugs](#) (10 if counting merged bugs), consider including or untagging them.
- Someone wrote an [updated debian/watch](#) file you may want to include in the package.

bugs

all **98** (102)

RC: **0**

I&N: **53** (55)

M&W: **39** (41)

F&P: **6**

links

- [homepage](#)
- [changelog / copyright](#)
- [buildd: logs, exp, ports, checks](#)
- [debcheck: unstable testing stable](#)
- [lintian](#) (1.49)
- [popcon](#)
- [debtags](#)
- [screenshots](#)
- [RDF meta-data](#)

news [RSS](#)

- [2012-12-06] [Accepted 2.2.16-6+squeeze10 in squeeze-security](#) (low) (Stefan Fritsch)
- [2012-11-10] [apache2 2.2.22-12 MIGRATED to testing](#) (Britney)
- [2012-10-30] [Accepted 2.2.22-12 in unstable](#) (low) (Arno Töll)
- [2012-09-12] [Accepted 2.2.16-6+squeeze8 in squeeze](#) (low) (Stefan Fritsch)
- [2012-08-14] [apache2 2.2.22-11 MIGRATED to testing](#) (Britney)
- [2012-08-03] [Accepted 2.2.22-11 in unstable](#) (low) (Arno Töll)
- [2012-07-30] [Accepted 2.2.22-10 in unstable](#) (low) (Stefan Fritsch)
- [2012-06-30] [apache2 2.2.22-9 MIGRATED to testing](#) (Britney)
- [2012-06-24] [Accepted 2.2.22-9 in unstable](#) (low) (Stefan Fritsch)
- [2012-06-23] [Accepted 2.2.22-8 in unstable](#) (medium) (Stefan Fritsch)
- [2012-06-21] [apache2 2.2.22-7 MIGRATED to testing](#) (Britney)
- [2012-06-10] [Accepted 2.2.22-7 in unstable](#) (low) (Stefan Fritsch)
- [2012-06-09] [apache2 2.2.22-6 MIGRATED to testing](#) (Britney)
- [2012-05-29] [Accepted 2.2.22-6 in unstable](#) (low) (Stefan Fritsch)
- [2012-05-28] [Accepted 2.4.2-2 in experimental](#) (low) (Arno Töll)

ubuntu

- version: [2.2.22-6ubunt4](#)
- patches for [2.2.22-6ubunt4](#)

Adding ADMS.SW for FusionForge

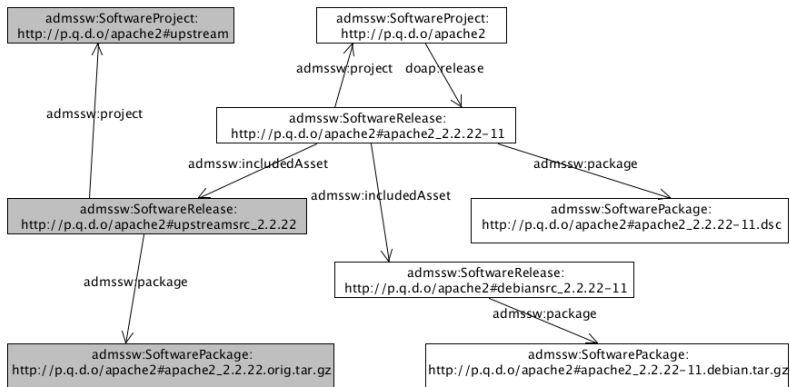
- Adding ADMS.SW support in FusionForge
 - Projects
 - Releases
 - Trove categories
- Expected deployment on :
 - Cenatic
 - Adullact
 - Debian's Alioth



Model : Graph of RDF resources

Reference **Linked Data** resources with *canonical* URI like

<http://packages.qa.debian.org/PACKAGE#RESOURCE_ID>



The greyed resources correspond to upstream components



Forget about RDF/XML

Yes, RDF can be expressed :

- as XML
- as **Turtle (PREFERRED)** : text (close to YAML / RFC 822)
- as JSON

Turtle example :

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .

<http://people.debian.org/~obergix/foaf.ttl#me>
  a foaf:Person ;
  foaf:name "Olivier Berger" ;
  foaf:nick "obergix" ;
  foaf:mbox "mailto:obergix@debian.org" ;
  foaf:homepage <http://people.debian.org/~obergix/> ;
  owl:sameAs <http://www-public.telecom-sudparis.eu/~berger_o/foaf.rdf#me> .
```

See also [RDF Primer — Turtle version](#)



Apache2 Debian packaging as RDF

`http://packages.qa.debian.org/a/apache2.ttl`

`@prefix doap: <http://usefulinc.com/ns/doap#> .`

`@prefix admssw: <http://purl.org/adms/sw/> .`

`<http://p.qa.d.o/apache2#project>`

```

a admssw:SoftwareProject ;
doap:name "apache2" ;
doap:description "Debian␣apache2␣source␣packaging" ;
doap:homepage <http://packages.d.o/src/apache2> ;
doap:homepage <http://p.qa.d.o/apache2> ;
doap:release <http://p.qa.d.o/apache2#apache2_2.2.22-11> ;
schema:contributor [
  a foaf:OnlineAccount ;
  foaf:accountName "Debian␣Apache␣Maintainers" ;
  foaf:accountServiceHomepage <http://qa.d.o/developer.php?login=
    debian-apache@lists.d.o>
] .

```

`<http://p.qa.d.o/apache2#apache2_2.2.22-11>`

```

a admssw:SoftwareRelease ;
rdfs:label "apache2␣2.2.22-11" ;
doap:revision "2.2.22-11" ;
admssw:package <http://p.qa.d.o/apache2#apache2_2.2.22-11.dsc> ;
admssw:includedAsset <http://p.qa.d.o/apache2#upstreamsrc_2.2.22> ;
admssw:includedAsset <http://p.qa.d.o/apache2#debiansrc_2.2.22-11> .

```



The ADMS.SW ontology

Asset Description Metadata Schema for Software (ADMS.SW)

- Pilot : EC / Interoperability Solutions for European Public Administrations (ISA) - cf. Joinup site
- Exchanging project / packages / releases descriptions across development platforms and directories



Open specifications

- Not too much NIH syndrom : reuses :
 - **ADMS / RADion** (generic meta-data for semantic assets indexing)
 - **DOAP** (*Description of a project*)
 - **SPDX™** (*Software Package Data Exchange* ®)
 - **W3C Government Linked Data (GLD) Working Group**

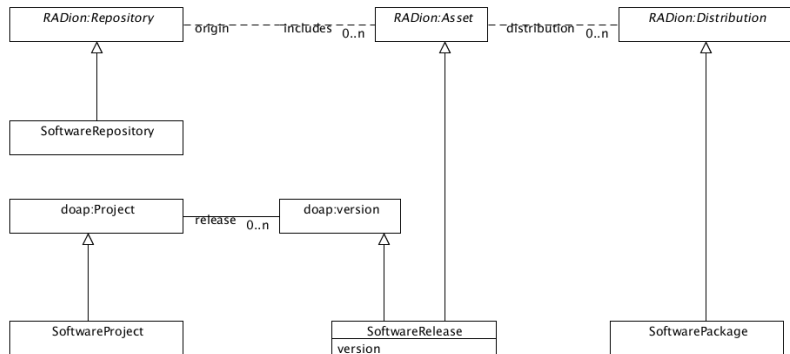
Version 1.0 issued 2012/06/29

- RDF Validator available
- RDF is extensible : ADMS.SW / DOAP core + distro-specific extensions



ADMS.SW main concepts

Project (= Program), Release, Package



Modeling Debian will require other complements



Related initiatives

Related initiatives about package meta-data :

- [AppStream](#) (Software Center, DEP11, etc.)
- [Umegaya](#) (upstream meta-data, links with research publications, etc.)
- [DistroMatch](#) (match package names across distributions)



Recommendations

- Upstream authors : please create **DOAP** descriptions for your projects
<https://github.com/edumbill/doap/wiki>
- Distributions : join the ADMS.SW bandwagon to documents package releases
- Followup-to : <distributions@lists.freedesktop.org> ?



Fin

More details at :

- <http://wiki.debian.org/qa.debian.org/pts/RdfInterface>
- [Linked Data descriptions of Debian source packages using ADMS.SW](#)
- *Authoritative Linked Data descriptions of Debian source packages using ADMS.SW*, to appear at [OSS 2013](#) (pre-print available on demand)

Contact :

Micro-blogging : @oberger <http://identi.ca/oberger/>

Email : <mailto:obergix@debian.org>

Blog : http://www-public.telecom-sudparis.eu/~berger_o/weblog/

Copyright 2013 Institut Mines Telecom + Olivier Berger

License of this presentation : Creative Commons Share Alike (except illustrations which are under copyright of their respective owners)

