

Generating Linked Data descriptions of Debian packages in the Debian PTS

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

Sunday 25/11/2012
Debian mini-debconf - 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.



Issues

- A universal Free Software Project description format ;
- A universal distribution Package description format ;
- How to inter-link documents about the same program :
 - Bug reports (upstream and in all distros)
 - Security advisories
- Are we developing Free Software in close silos ?
- What's usually wrong with [XML | JSON | YAML | RFC822] format ?



Linked Open Debian Development Meta-Data

Let's try and make as much *Debian facts* as possible available to *humans + machines*?

Adopting the 5 ★ **Open Data** principles using RDF for Debian 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) |

Aligns quite well with 3. *We will not hide problems* (**Debian social contract**)



What's the Debian PTS

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

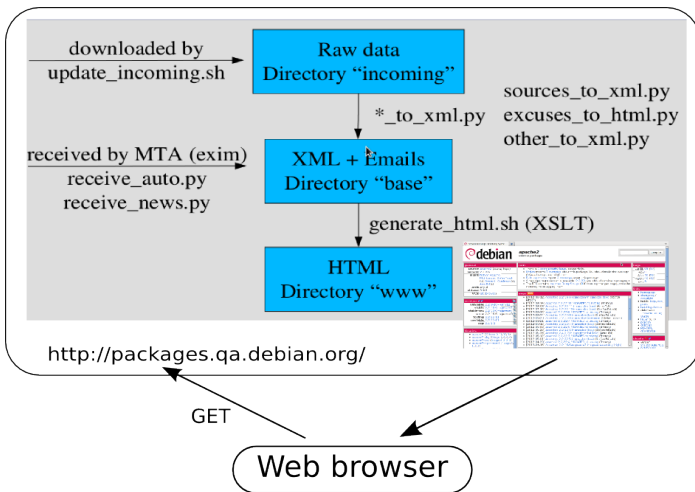
The screenshot shows the Debian Package Tracking System (PTS) web interface for the **apache2** source package. The interface is organized into several panels:

- Header:** Displays the Debian logo, the package name **apache2**, and the label "source package". There is a search box and a "jump to" button.
- general:**
 - source: [apache2](#) (source, httpd)
 - version: 2.2.22-11
 - 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+squeeze7](#)
 - stable-sec: [2.2.16-6+squeeze7](#)
 - s-p-u: [2.2.16-6+squeeze8](#)
 - testing: [2.2.22-11](#)
 - unstable: [2.2.22-11](#)
 - exp: [2.4.2-2](#)
- binaries:**
 - [apache2](#) (33 bugs: 0, 19, 13, 1)
 - [apache2-dbg](#) (0 bugs: 0, 0, 0, 0)
 - [apache2-doc](#) (0 bugs: 0, 0, 0, 0)
 - [apache2-mpm-event](#) (1 bug: 0, 1, 0, 0)
- todo:**
 - There is [1 open security issue](#), please fix it.
 - [Lintian](#) reports [7 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: [2.2.23](#), you should consider packaging it.
 - The BTS contains [patches fixing 6 bugs](#) (8 if counting merged bugs), consider including or untagging them.
- news:**
 - [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)
 - [2012-05-12] [apache2 2.2.22-5 MIGRATED to testing](#) (Britney)
 - [2012-05-01] [Accepted 2.2.22-5 in unstable \(low\)](#) (Stefan Fritsch)
 - [2012-04-23] [apache2 2.2.22-4 MIGRATED to testing](#) (Britney)
 - [2012-04-16] [Accepted 2.2.16-6+squeeze7 in squeeze-security \(high\)](#)
- bugs:**
 - all: [93](#) ([97](#))
 - RC: [0](#)
 - ISN: [50](#) ([52](#))
 - M&W: [40](#) ([42](#))
 - F&P: [3](#)
- links:**
 - [homepage](#)
 - [changelog / copyright](#)
 - [build: logs, exp, ports](#)
 - [build log checks](#)
 - [debcheck: unstable testing stable](#)
 - [lintian](#) (0, 7)
 - [popcon](#)
 - [debtags](#)
 - [RDF/XML description](#)
- ubuntu:**
 - version: [2.2.22-6ubuntu2](#)
 - [patches for](#)



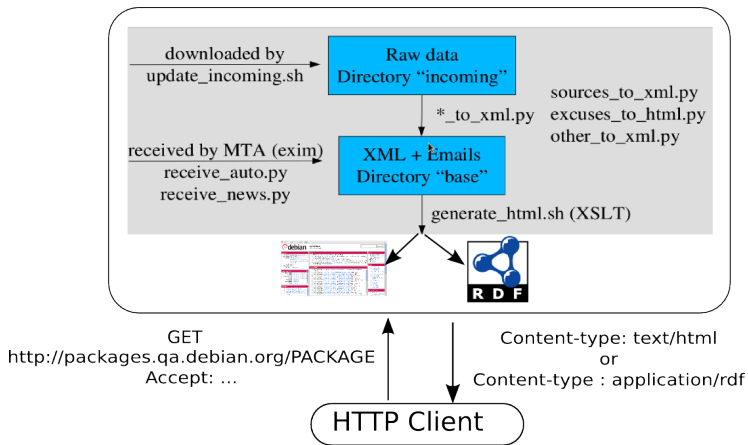
Architecture of the PTS

The PTS used to be for humans



Producing HTML for humans and RDF for machines

NEW : the PTS now generates **RDF** meta-data to be consumed by machines (+/- 1.5 M triples)



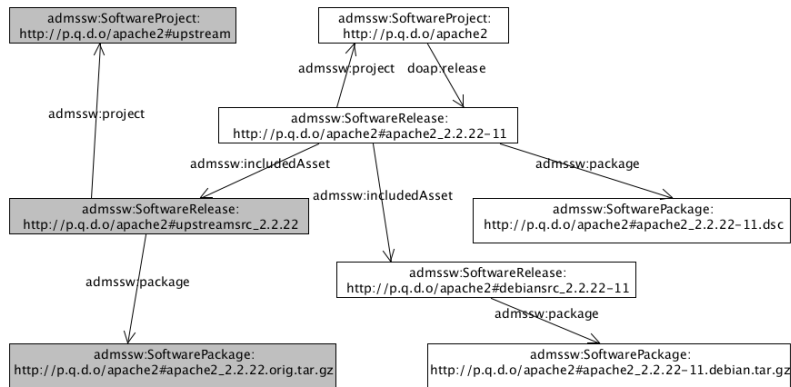
We use content-negotiation to redirect to the proper document.



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



Apache2 packaging as RDF (Turtle notation)

<http://packages.qa.debian.org/a/apache2.ttl> (soon)

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

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

<<http://packages.qa.debian.org/apache2#project>>

a admssw:SoftwareProject ;

doap:name "apache2" ;

doap:description "Debian␣apache2␣source␣packaging" ;

doap:homepage <<http://packages.debian.org/src/apache2>> ;

doap:homepage <<http://packages.qa.debian.org/apache2>> ;

doap:release <http://packages.qa.debian.org/apache2#apache2_2.2.22-11> ;

schema:contributor [

a foaf:OnlineAccount ;

foaf:accountName "Debian␣Apache␣Maintainers" ;

foaf:accountServiceHomepage <<http://qa.debian.org/developer.php?login=debian-apache@lists.debian.org>>

] .

<http://packages.qa.debian.org/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://packages.qa.debian.org/apache2#apache2_2.2.22-11.dsc> ;

admssw:includedAsset <http://packages.qa.debian.org/apache2#upstreamsrc_2.2.22> ;

admssw:includedAsset <http://packages.qa.debian.org/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**

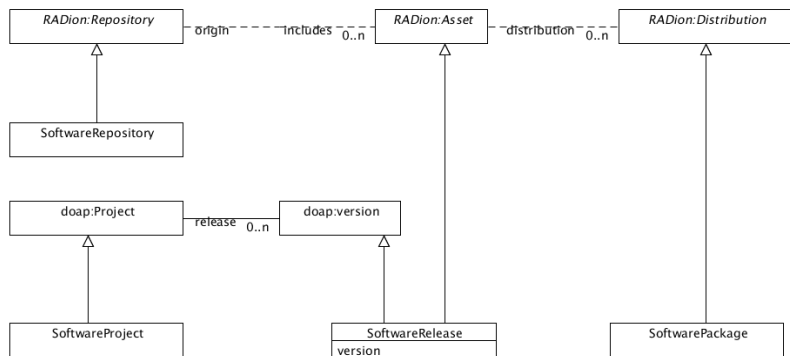
Version 1.0 issued 2012/06/29

RDF Validator issued a few weeks ago



ADMS.SW main concepts

Project (= Program), Release, Package



Modeling Debian will require other complements



Traceability over the FLOSS ecosystem

Collecting descriptions of packages/projects from their source projects as Linked Data (DOAP or ADMS.SW) into an RDF triple store.

For instance :

- For debian, from PTS ADMS.SW full dump from `p.q.d.o:/srv/packages.qa.debian.org/www/web/full-dump.rdf`
- For apache, from DOAP files (see <http://projects.apache.org/docs/index.htm>)
- ... *Add your preferred upstream* ...

Following links between bugs, packages, security alerts, all in **semantic interoperable** meta-data formats!



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/
...



Related initiatives

About ADMS.SW :

- [Adding ADMS.SW support in FusionForge](#) (hence soon in Alioth)

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)

Let's discuss further steps (replacing YAML by Turtle, etc. ;-)



Fin

More details at :

- <http://wiki.debian.org/qa.debian.org/pts/RdfInterface>
- Linked Data descriptions of Debian source packages using ADMS.SW to be published soon in SWESE 2012 proceedings

Contact :

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

Email : <mailto:obergix@debian.org>

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

Copyright :

Copyright 2012 Institut Mines Telecom + Olivier Berger

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

