# Authoritative linked data descriptions of debian source packages using ADMS.SW

Olivier Berger, Christian Bac

# Authoritative Linked Data descriptions of Debian source packages using ADMS.SW

Olivier Berger[1,2] and Christian Bac[1]

[1] Telecom SudParis, Évry, France
{olivier.berger,christian.bac}@telecom-sudparis.eu
[2] Debian project
obergix@debian.org

**Abstract.** The Debian Package Tracking System is a Web dashboard for Debian contributors and advanced users. This central tool publishes the status of subsequent releases of source packages in the Debian distribution.
It has been improved to generate RDF meta-data documenting the source packages, their releases and links to other packaging artifacts, using the ADMS.SW 1.0 model. This constitutes an authoritative source of machine-readable Debian "facts" and proposes a reference URI naming scheme for Linked Data resources about Debian packages.
This should enable the interlinking of these Debian package descriptions with other ADMS.SW or DOAP descriptions of FLOSS projects available on the Semantic Web also using Linked Data principles. This will be particularly interesting for traceability with upstream projects whose releases are packaged in Debian, derivative distributions reusing Debian source packages, or with other FLOSS distributions.

**Key words:** ADMS.SW, Debian, Linked Data, package, Semantic Web, standard, interoperability, Open Source, Free Software, RDF, DOAP, PTS, FLOSS

## 1 Introduction

*Asset Description Metadata Schema for Software* (ADMS.SW) is a novel ontology developped for describing software packages, releases and projects, which can be applied to describe packages in a Free, Libre and Open Source Software (FLOSS) distribution, using Semantic Web techniques. We consider it is a foundational component that will allow to conduct future Quality Assurance or other large scale FLOSS studies across the Linked Open Data cloud [5].

FLOSS software ecosystems are composed of many different actors collaborating around single programs, from original upstream authors to downstream

This is a revised version of a previous paper [2] which was initially accepted at the 8th International Workshop on Semantic Web Enabled Software Engineering (SWESE 2012), but that the authors weren't able to present physically at the workshop.

packagers in distributions like Debian. Descriptions of FLOSS development artifacts made with standardized and semantic formats like ADMS.SW can help trace some of the process which generally happen in various venues across the ecosystem.

### 1.1 The Need for Linked Data Descriptions of FLOSS

Constructing models of interactions happening along the FLOSS production lines can be interesting, both for researchers and practitioners. Research in empirical software engineering can for instance involve studies conducted by modeling properties and relations between FLOSS production artifacts and actors.

The Semantic Web techniques bring key benefits in terms of semantic interoperability : using a W3C standard like RDF [16] which is natively extensible helps integrate potentially incoherent data, which fits quite well large scale problems. The size of the communities and diversity of actors and tools present in large FLOSS ecosystems qualify well for such an approach [13].

The Linked Data approach [5], can be very convenient to *interlink* resources representing actors or artifacts belonging to different projects described with RDF. It will allow researchers to integrate in the same "triple store" database, description of FLOSS artifacts or actors with variable structures, still relying on common semantics and a harmonized URI nomenclature that reflects the origin of these resources.

But for FLOSS developers alike, these semantic Web Techniques should offer potential interesting applications, in particular to create new global services that need to interconnect different heterogenous project tools [4]. As an illustration, we can imagine a new "global bug tracking system" that aims at correlating similar bug reports filed in different Linux distributions. It can be helpful to offer better support responses, allowing navigation between reports which may have been related to each-other previously. Such a system will require to interface to lots of different bugtracker APIs. Whereas standards like *Open Services for Lifecycle Collaboration* (OSLC) [3] (which rely on extensible semantic formats based on RDF and REST[1] APIs) can help solve some concrete interoperability issues, they only address parts of the problems (and their deployment is not yet spectacular among FLOSS project). Actually, even once semantically compatible data has been collected, it must be integrated in a coherent data store. And therefore, nomenclature, freshness and accuracy issues still represent interesting challenges. Addressing them is a foundational requirement for large scale applications described above.

[1] REpresentational State Transfer

### 1.2 Authoritative Linked Data Descriptors Produced by FLOSS Projects

We postulate that there are higher chances that meta-data is more accurate and up-to-date when it is produced closest to the very heart of the FLOSS projects, than obtained after a series of collection and conversion activities conducted by third parties. Thus, with the Linked Data principles in mind[2], significant artifacts produced by FLOSS projects ought to be complemented with meta-data available at the very same Web domains, as a minimal set of authoritative RDF resources. URIs naming these resources can then be rooted at the project's domain name, and serve to identify its artifacts unambiguously.

As an illustration, Semantic Web resources describing projects from the Apache foundation would be downloaded from RDF documents available on `http://projects.apache.org/` which would identify them for instance with URIs like `<http://PROJNAME.apache.org/>` (or a variant, like `<http://PROJNAME.apache.org/doap#project>`)[3]. Thus, in the description of the Debian packaging the Apache *geronimo* program, we could reference its upstream project (from Debian's point of view) as the RDF resource `<http://geronimo.apache.org/>`.

Our initiative, coupled with other previous and current efforts, will hopefully help achieve a state when almost every FLOSS project are able to publish on their Web sites or development forges, even minimal, but authoritative Linked Data descriptions of the project or its software artifacts, either as *Descriptions Of A Project* (DOAP) [7] or ADMS.SW.

### 1.3 Goal and Structure of this Paper

This paper will introduce a Linked Data interface based on ADMS.SW, which was deployed on the Debian Package Tracking System (PTS), that produces authoritative meta-data descriptions for the core artifacts produced by the Debian project: *source packages*.

Due to Debian's respected position in the FLOSS ecosystem, such a deployment already covers a great percentage of all FLOSS programs, and can thus be inspirational for many FLOSS projects.

In section 2, we introduce the ADMS.SW specification. Then a brief introduction to the structure of Debian source packages is provided in section 3. Section 4 documents the choices adopted for generating Linked Data representations of Debian source packages and related FLOSS artifacts in the Debian PTS. Section 5 presents a quick review of similar and complementary initiatives, while section 6 illustrates how a trivial project matching can be made with collected Linked Data descriptions.

[2] `http://www.w3.org/DesignIssues/LinkedData.html`

[3] There's actually a DOAP description for Geronimo, linked from `http://projects.apache.org/projects/geronimo.html` — see 5.1

## 2 The ADMS.SW Specification

The *Asset Description Metadata Schema for Software* (ADMS.SW) specification[4] is described as : "[...] *a metadata vocabulary to describe software making it possible to more easily explore, find, and link software on the Web.*"

It is an outcome of the ISA programme (*Interoperability Solutions for European Public Administrations*) of the European Commission, elaborated by a working group of software catalogues and forges experts[5]. Although it is not specifically covering FLOSS software only, ADMS.SW has nevertheless been geared at addressing meta-data of FLOSS projects hosted in public development forges to facilitate their identification and reuse by Public Administrations.

ADMS.SW specifications are published with a complementary OWL ontology, referenced as `http://purl.org/adms/sw/`, to allow the publishing of such meta-data as RDF.

As illustrated in Figure 1, it provides three main entities : *Software Project*, *Software Release*, and *Software Package* to model meta-data about software programs, their versions, and the distribution archives of these.

But it also contains various elements related to *Software Repositories* descriptions in order to facilitate the maintenance of data managed by software catalogues (provenance, timestamping, etc.), based on the RADion common model of ADMS, which describes generic semantic assets.

ADMS.SW 1.0 reuses existing specifications and standards, such as Dublin Core [19], DOAP [7], SPDX™ [9], ISO 19770-2 [8], ADMS [10], and the "Sourceforge Trove software map"[6]. As DOAP is already widely used in practice, ADMS.SW reuses much of its properties. ADMS.SW is also interoperable with the SPDX specification, whose main object, to date, is the description of copyright and license conditions applying to particular software packages or source files.

## 3 Debian Source Packages

The Debian project[7] creates a Free Software distribution, which contains thousands of FLOSS *binary* packages ready to be installed on various computer architectures. Several versions of the Debian distribution are maintained in parallel, in three main *suites* : 'stable', 'testing' and 'unstable'.

Debian has been studied by many authors, as it represents a good proxy for the entire FLOSS ecosystem, due to the high number of packages it contains, and since its development and Quality Assurance (QA) infrastructure is generally open or easily accessible to researchers in empirical software engineering (see for instance [11] or [6]).

[4] `https://joinup.ec.europa.eu/asset/adms_foss/release/release100`
[5] one of the authors was an active member of the working group.
[6] `http://sourceforge.net/apps/trac/sourceforge/wiki/Software Map and Trove`
[7] `http://debian.org/`

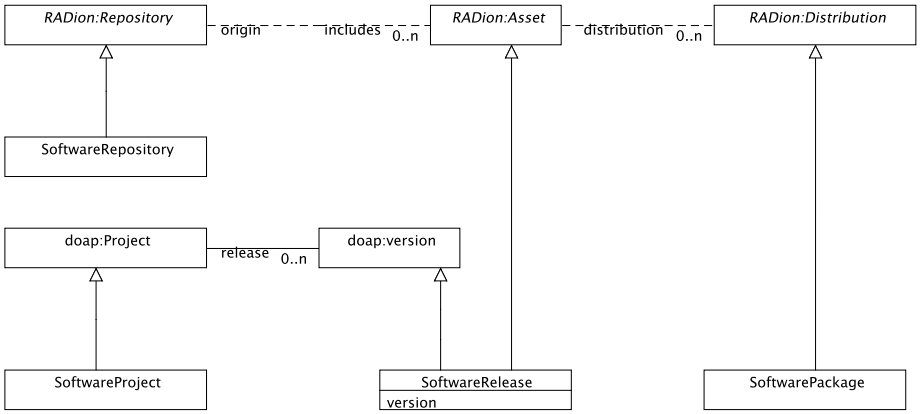


**Fig. 1.** Simplified UML diagram of the main ADMS.SW entities

### 3.1 Structure of Debian Source Packages

Each binary package is actually built from a particular Debian *source* package. Source packages contain "Makefiles" for package generation, `control` files containing different meta-data like versions or package dependency descriptions, and other scripts necessary for installation, configuration, upgrade or removal of the binary packages [15]. In addition, it is quite common to include patches applying to the source code of the packaged program, to adjust it to Debian specificities or to include security fixes backported from later upstream releases.

Each revision of a Debian source package is then generally composed of two file archives : one for the source code of the upstream version of the packaged program (ending in `.orig.tar.gz`), complemented by another one for these Debian specific files (ending in `.debian.tar.gz`)[8]. Only the latter Debian specific files archive, and associate meta-data descriptors change between subsequent revisions of Debian source packages of the same version of an upstream program.

### 3.2 The Debian Package Tracking System

For every Debian source package, the *Debian Package Tracking System* (PTS) provides a Web dashboard (see a screenshot[9] in Figure 2) which displays almost all there is to know about the status of that package [17].

However, its HTML pages are not really exploitable by machines in a direct form, should anyone need to interface the Debian QA system with other services. One such need seems quite obvious for derivative distributions constructed from

[8] as an exception to this general rule, some packages, which are called "native packages", don't have a corresponding upstream project outside Debian and only have Debian specific files.

[9] taken from `http://packages.qa.debian.org/a/apache2.html`

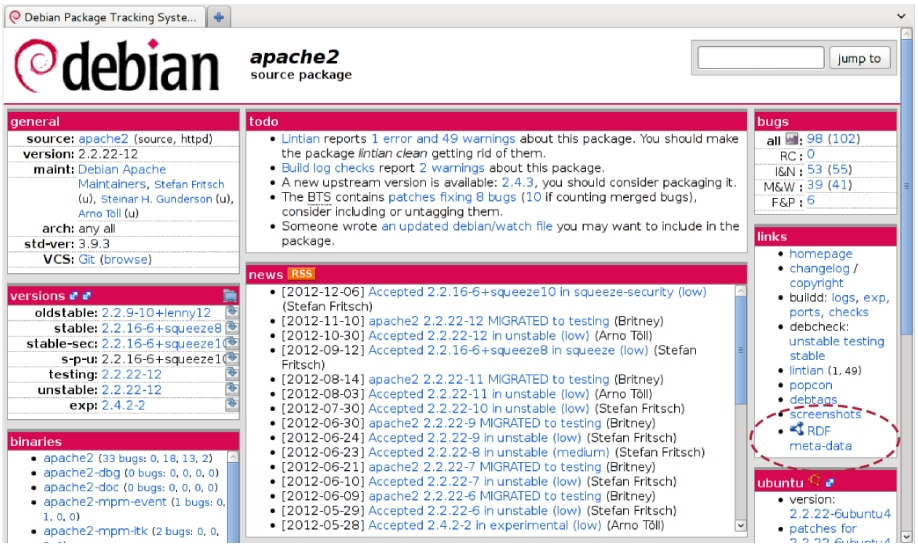


**Fig. 2.** Apache 2 source package status in the Debian PTS

Debian, like Ubuntu. Therefore, the PTS provides a custom SOAP interface[10], but the lack of standard representation of data retrieved from this API may require another ad-hoc converter to be added to every application wishing to interface with it.

As an alternative, we have started implementing a Linked Data [5] interface for the Debian PTS, using the ADMS.SW ontology to represent Debian source package facts with the standard, thus interoperable, RDF model.

## 4 Linked Data Representation of Debian Source Packages

We have improved the Debian PTS to add the generation of RDF descriptions for all Debian source packages.

Every Debian source package, which used to have an HTML page accessible at a URL like `http://packages.qa.debian.org/apache2`, now has a corresponding RDF document available at the same URL, either as Turtle [1] or RDF/XML format. Applying a common Linked Data pattern, HTTP clients will be redirected automatically to the proper HTML or RDF document depending on the *content-type* that is requested by the HTTP client, so that the same URL can be used to represent both the human-readable HTML pages or the machine-processable RDF document.

Thus, each package in Debian can be identified on the Semantic Web with a unique URI like `http://packages.qa.debian.org/SRC-PKG-NAME`, which is dereferenceable as an RDF document.

[10] `http://wiki.debian.org/qa.debian.org/pts/SoapInterface`

```
<http://p.q.d.o/apache2#apache2_2.2.22-12>
    a admssw:SoftwareRelease ;
    rdfs:label "apache2 2.2.22-12" ;
    dcterms:description "Debian apache2 source package version
        2.2.22-12" ;
    doap:revision "2.2.22-12" ;
    dcterms:publisher <http://debian.org/> ;
    admssw:status <http://p.q.d.o/#released> ;
    admssw:project <http://p.q.d.o/apache2#project> ;
    admssw:includedAsset <http://p.q.d.o/apache2#upstreamsrc_2.2.22> ;
    admssw:includedAsset <http://p.q.d.o/apache2#debiansrc_2.2.22-12>;
    admssw:package <http://p.q.d.o/apache2#apache2_2.2.22-12.dsc> ;
    dcterms:relation <https://launchpad.net/ubuntu/+source/apache2
        /2.2.22-6ubuntu4> .
```

**Listing 1.** RDF description available at http://packages.qa.debian.org/a/apache2.ttl of revision 12 of the source package for apache2 version 2.2.22

The example in Listing 1 is an excerpt of such an RDF description, in Turtle, of a particular revision of the Debian source package for `apache2`. Note that URIs based on `http://packages.qa.debian.org/` are converted to `http://p.q.d.o/` for brievity in the rest of this section.

### 4.1 Modelling Debian Source Packages with ADMS.SW

This section presents the modelling choices adopted so that every Debian source package can be modeled as interlinked RDF resources. The version numbers reflected in the resource URIs or file names below respect the Debian package versions numbering convention[11].

Figure 3 represents the main resources produced by the PTS for a particular release of the `apache2` Debian source package, as found in `http://packages.qa.debian.org/a/apache2.ttl` (in grey, the "upstream"-related resources).

Every source package has a corresponding *source packaging project* `SoftwareProject` resource, named `<http://p.q.d.o/SRC-PKG-NAME#project>`. The different resource URIs which will be expressed below will be fragments to this base URI. *Revisions* of this source package have corresponding `SoftwareRelease` resources, named as `<#SRC-PKG-NAME_DEB-PKG-VERS>`. Only one of these (the "latest" one known by the PTS) is fully described as containing (`includedAsset`) two sub `SoftwareReleases` :

– one sub `SoftwareRelease` for the *upstream program's version*, named `<#upstreamsrc_UPSTR-VERS>`. It comes with additional resources for all *archive files* of the upstream sources as `SoftwarePackages` (typically named like `<#SRC-PKG-NAME_UPSTR-VERS.orig.tar.gz>`);

[11] as a short rule, the Debian package revision M of version N of an upstream program P is identified in file names as `P_N-M`.

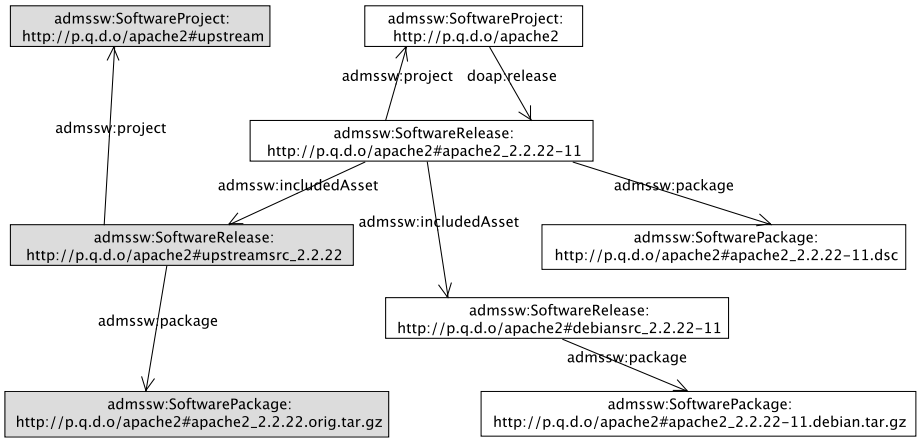


**Fig. 3.** Relations between resources produced for release 2.2.22-11 of the Debian `apache2` source package

– one for the set of *Debian packaging files*, as `<#debiansrc_DEB-PKG-VERS>`, with resources for all files comprising the Debian *package source archive* (typically named like `<#SRC-PKG-NAME_DEB-PKG-VERS.debian.tar.gz>`).
– An additional `SoftwarePackage` resource is generated for its `SRC-PKG-NAME_UPSTR-VERS.dsc` file at a URI like `<#SRC-PKG-NAME_DEB-PKG-VERS.dsc>`.

Also produced is one `SoftwareProject` resource for the *upstream project*, named `<#upstream>` whith its `doap:homepage`, if it's known by the PTS (which means it has been documented by the Debian packager appropriately).

Additional complementary resources are produced, and all resources have RDF properties (as mandated in ADMS.SW, mainly reused from DOAP or Dublin Core), all of which it is useless to describe here in detail.

### 4.2 Deployment on debian.org

The authors have deployed the XSLT stylesheets generating these RDF documents on the PTS service of the Debian project[12]. Alongside the HTML pages of the PTS, the RDF descriptions of Debian source packages are thus refreshed every time new revisions will appear in the Debian archive.

A full RDF dump of all the meta-data is also available to Debian project members[13]. It contains around 2.1 million triples at the time of writing.

[12] see : `http://packages.qa.debian.org/common/RDF.html`
[13] on `packages.qa.debian.org:/srv/packages.qa.debian.org/www/web/full-dump.rdf`

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

<http://geronimo.apache.org/>
    a doap:Project .
    doap:name "Apache Geronimo"@en ;
    doap:shortdesc "Java EE Application Server"@en ;
    doap:description "Apache Geronimo is an open source server runtime
        [...]."@en ;
    doap:homepage <http://geronimo.apache.org> ;
    doap:license <http://usefulinc.com/doap/licenses/asl20> ;
```

**Listing 2.** Excerpt from the RDF description of the Apache geronimo project

## 5 Complementary Efforts

In this section, we present a few complementary initiatives which describe software packages with RDF vocabularies, using DOAP or ADMS.SW and which could be interesting for interoperability with the Debian PTS.

### 5.1 DOAP Published by FLOSS Directories

A number of projects maintain public DOAP descriptions of their programs, or other RDF descriptions of meta-data about the releases they produce. They may be interested in complementing descriptions with ADMS.SW, or could offer sources of descriptions that could be interlinked with the ones produced by the Debian PTS.

A quick survey conducted by the authors showed the following sources[14] :

- Gnome project
- Apache project
- PyPI (Python Package Index) directory
- CPAN (Comprehensive Perl Archive Network) directory

Listing 2 shows an excerpt of the DOAP description of the Apache Geronimo project as published by this project[15], and converted to Turtle for readability.

A quick review of samples from these sources showed a lack of consensus on the use of certain meta-data, and that URIs adopted to reference the same projects or programs tend to vary, even for `doap:homepage` URLs (a great portion of these documents are manually crafted, or projects may have various pages that can be considered their *homepage*, in particular when the project is not hosted on its own top level domain).

[14] these are maintained in `https://github.com/edumbill/doap/wiki/Sites`

[15] downloaded from `http://svn.apache.org/repos/asf/geronimo/site/trunk/doap_Geronimo.rdf`

### 5.2 Projects Hosted on FusionForge Forges

An ADMS.SW plugin for the FusionForge 5.2 software development forge has also been created by one of the authors[16], in order to allow the description of projects hosted on FusionForge based development forges. It may be complemented by another FusionForge plugin providing FOAF profiles [12] for project participants, which can enrich the Linked Data representations.

The plugin is still under active development, and targetted at a post 5.2 release of FusionForge, so it will take a certain time until it is deployed on public forges hosting FLOSS projects[17].

### 5.3 Consuming ADMS.SW in the Joinup Portal

The *Joinup* portal[18] of the ISA programme aims at integrating in a single portal FLOSS descriptions available from different Public Administration forges, by harvesting descriptions of projects directly in their development project spaces, as ADMS.SW descriptions[19].

Whereas the current version of Joinup doesn't rely on Semantic Web techniques for collection of the projects descriptions, it is expected to be improved to evolve towards ADMS.SW consuming in the future. FLOSS Project descriptions would then complement other Semantic Assets (standards, documentation) catalogued and made available on the reference portal at Joinup as semantic assets expressed with the ADMS vocabulary.

### 5.4 Interlinked Developer Profiles

Project descriptions aren't the only resources that can be interlinked across the FLOSS ecosystem. Iqbal shows in [14] how developer profiles can also be converted to RDF and interlinked to create a more comprehensive view of the developer communities around a project, for instance. This approach usually involves mining repositories or social sites through custom interfaces (via SOAP for instance), and later converting to RDF. But we believe there would be a great benefit in avoiding such potentially error-prone conversions if development platforms would natively produce DOAP/ADMS.SW (or FOAF) descriptions "out of the box", as explained above.

[16] `http://fusionforge.org/plugins/mediawiki/wiki/fusionforge/index.php/ADMS.SW_Plugin`

[17] like Debian's own Alioth forge operated by FusionForge at `http://alioth.debian.org/`

[18] `http://joinup.ec.europa.eu/`

[19] more details at `https://joinup.ec.europa.eu/software/federated_forge`

```
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
  }
}
```

**Listing 3.** SPARQL query matching Apache and Debian projects by common homepages

## 6 Applications

As with every Linked (Open) Data initiatives, the use of standard representations and their availability on the Semantic Web can lead to lots of different uses.

An obvious case of using such ADMS.SW description of Debian source package is the *matching* of Debian packages with other packages/projects described in their respective projects directories, allowing more interlinking of resources.

### 6.1 Matching Projects / Software Across Repositories

The `doap:homepage` of the "upstream" `SoftwareProject` resources generated by the Debian PTS can be an obvious matching key, provided that one has a database of upstream project descriptions (as DOAP[7]).

As an illustration, we demonstrate this by loading DOAP descriptions of projects of the Apache foundation[20], together with a dump of the Debian source package descriptions in a single triple store (virtuoso). The example SPARQL query in Listing 3 shows how to query for such matches between Debian and Apache.

Such a query currently returns 62 matched source packages and Apache upstream projects (see an excerpt in table 1, where URLs have been compacted for brievity).

But the reliability of this matching method isn't very good in practice. There may be many more Apache foundation projects packaged in Debian, but the maintainers may have forgotten to add a homepage link in the package descriptors. Or the URLs mentioned may not be matching, as project homepage naming conventions can vary (and evolve in time).

[20] collected from `projects.apache.org` (see `http://projects.apache.org/docs/index.html`)

**Table 1.** Matching upstream 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/ |
| ... | ... | ... |

An alternate matching method could be based on project name litterals, but that isn't always feasable either, due to homonimy for instance. One will refer to [18] for an analysis of this problem.

The *distromatch* project[21], started in 2011, intended to try and help solve these project/packages matching issues, although it is unfortunately not maintained anymore at the time of writing.

In any case, this first quick proof of concept allows us to plan further developments based on such meta-data, which will be tested on real life cases, for instance in constructing RDF harvesters and meta-data aggregators, and eventually merging with initiatives like distromatch.

### 6.2 Large Scale Perspective

The RDF-ization of the Debian PTS has just started. Next steps will include modelling of relations between source and binary packages. These will probably require extending ADMS.SW or integrating complementary ontologies.

When deployments of ADMS.SW have been made on software forges (like FusionForge servers), software catalogues (like Joinup) or in other FLOSS distributions, it will become one of the tools allowing automated traceability at large scale of FLOSS releases and associated artefacts, by interlinking their Linked Data resources.

Some interlinking of security advisories, patches, or bug reports (for instance combined with the OSLC-CM standard[22]) should then be easier, and diminish manual intervention needs, for the benefits of all actors along the FLOSS production chains.

[21] `http://www.enricozini.org/2011/debian/distromatch/`
[22] `http://open-services.net/wiki/change-management/`

### 6.3 Future Developments

We believe the current early result can be a driving force for more deployments around ADMS.SW as a standardization core. However there seems to be a reluctance in adopting RDF in FLOSS projects, to some extent, maybe linked to an erronous perception that RDF must be expressed as XML (which is certainly not the case, with representations of the RDF model like Turtle [1], which has been adopted as a default for the Debian PTS).

We can foresee that only when novel inter-project "killer" applications making use of such Linked Data will have been developed, will it be possible to convince FLOSS projects that adoption of Linked Data standards descriptions can really be for their own benefit.

It is likely that even when lots of Linked Data descriptions of FLOSS artifacts are made available by major FLOSS projects, achieving effective interoperability will require many implementation efforts, far beyond a single actor's reach. More standardisation will be needed, and services will have to be provided to establish trusted reference catalogues of Semantic project descriptors (in the direction set by *Joinup* of the *distromatch* project for instance). Such actors would provide FLOSS "semantic hubs", or project matching "brokers" which could maintain reference interlinking relations for the concurrent Semantic descriptions which were produced in the many venues of the FLOSS ecosystem.

## 7 Conclusion

We have presented a first significant deployment of an ADMS.SW 1.0 implementation, which illustrates the potential for interlinking large sets of FLOSS project descriptions on the Semantic Web. ADMS.SW allows us to describe relations between projects, programs and their releases so that such entities become part of the Linked Open Data "cloud".

In [4], we envisioned some novel uses of Linked Data representations of FLOSS development artefacts, both for software engineers and researchers observing their efforts. But to achieve the full potential of that approach, the Linked Data representations must be semantically interoperable, authoritative, accurate, and using standard naming schemes for the same resources. We have achieved a first concrete step in this direction, through the current results for the Debian PTS.

The way we did it for the Debian PTS can be inspirational for other FLOSS distributions, either independant, or derived from Debian. By integrating such meta-data generation in the heart of the technical infrastructure of Debian, we hope to establish such an authoritative reference for Debian source packages identification on the Semantic Web.

## References


1. Dave Beckett and Tim Berners-Lee. Turtle - terse RDF triple language, W3C team submission, 2008. See: http://www.w3.org/TeamSubmission/turtle/.
2. Olivier Berger. Linked data descriptions of debian source packages using ADMS.SW. In Elisa F. Kendall, Jeff Z Pan, Ljiljana Stojanovic, and Yuting Zhao, editors, *SWESE 2012: 8th International Workshop on Semantic Web Enabled Software Engineering*, pages 43–55, Nara, Japan, 2012.
3. Olivier Berger, Sabri Labbene, Madhumita Dhar, and Christian Bac. Introducing OSLC, an open standard for interoperability of open source development tools. In *ICSSEA*, pages ISSN–0295–6322, Paris, France, 2011.
4. Olivier Berger, Ion Valentin Vlasceanu, Christian Bac, Quang Vu Dang, and Stéphane Lauriere. Weaving a semantic web across OSS repositories: Unleashing a new potential for academia and practice. *International Journal of Open Source Software and Processes (IJOSSP)*, 2(2):29–40, 2010.
5. Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 3 2009.
6. E Gabriella Coleman. *Coding Freedom: The Ethics and Aesthetics of Hacking*. Princeton University Press, 2012.
7. Edd Dumbill. Decentralizing software project registries with DOAP. In *Intelligent Search on XML Data - XML*, 2004.
8. *unspecified authors*. ISO/IEC 19770-2: Software identification tag standard.
9. *unspecified authors*. Software Package Data eXchange specification, 2011.
10. *unspecified authors*. Asset Description Metadata Schema specification 1.00, 2012.
11. Jesus M. Gonzalez-Barahona, Gregorio Robles, Martin Michlmayr, Juan José Amor, and Daniel M. German. Macro-level software evolution: a case study of a large software compilation. *Empirical Software Engineering*, 14(3):262–285, June 2009.
12. Mike Graves, Adam Constabaris, and Dan Brickley. FOAF: Connecting People on the Semantic Web. *Cataloging & classification quarterly*, 43(3):191–202, April 2007.
13. James Howison. Cross-repository data linking with RDF and OWL: Towards common ontologies for representing FLOSS data. In *WoPDaSD (Workshop on Public Data at International Conference on Open Source Software)*, 2008.
14. Aftab Iqbal and Michael Hausenblas. Integrating developer-related information across open source repositories. In *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on*, pages 69 –76, aug. 2012.
15. Ian Jackson, Christian Schwarz, et al. Debian policy manual. version 3.9.4.0, 2012-09-19 - `http://www.debian.org/doc/debian-policy/`.
16. Ora Lassila, Ralph R. Swick, and World Wide Web Consortium. Resource description framework (RDF) model and syntax specification, 1998. W3C Recommendation.
17. Martin Michlmayr. Managing debian. *AUUGN, The journal of AUUG Inc.*, 25(3), 9 2004.
18. Megan Squire. Integrating projects from multiple open source code forges. *IJOSSP*, 1(1):46–57, 2009.
19. Stuart L. Weibel, John A. Kunze, Carl Lagoze, and Misha Wolf. Dublin core metadata for resource discovery, 1998. RFC 2413.