

Debsources as a Platform

All your Debian source are belong to us!

Stefano Zacchiroli, Matthieu Caneill

August 18, 2015

DebConf15 (Heidelberg, Germany)

Acknowledgements

Code

Initially developed at IRILL, by Stefano Zacchiroli and Matthieu Caneill. Many people have contributed code and ideas since then.

Infrastructure

Debsources' servers are sponsored by IRILL.

Table of contents

- 1 Introduction
- 2 Features
 - Debsources' features
 - What's new?
 - Roadmap
- 3 Technologies
- 4 Research platform
- 5 Hacking

Table of contents

1 Introduction

2 Features

- Debsources' features
- What's new?
- Roadmap

3 Technologies

4 Research platform

5 Hacking

What is Debsources?

- A web application to browse **the source code** of Debian packages
- The infrastructure behind: **database, plugins, ...**

What is Debsources?

- A web application to browse **the source code** of Debian packages
- The infrastructure behind: **database, plugins, ...**

Play with it!

Navigate to <http://sources.debian.net>

Debian Sources

All Debian source are belong to us — Anonymous []

Browse through the source code of the [Debian](#) operating system. [Read more...](#)

Browse by prefix

[0](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [a](#) [b](#) [c](#) [d](#) [e](#) [f](#) [g](#) [h](#) [i](#) [j](#) [k](#) |
[lib-](#) [lib3](#) [liba](#) [libb](#) [libc](#) [libd](#) [libe](#) [libf](#)
[libg](#) [libh](#) [libi](#) [libj](#) [libk](#) [libl](#) [libm](#) [libn](#) [libo](#)
[libp](#) [libq](#) [libr](#) [libs](#) [libt](#) [libu](#) [libv](#) [libw](#) [libx](#)
[liby](#) [libz](#) [m](#) [n](#) [o](#) [p](#) [q](#) [r](#) [s](#) [t](#) [u](#) [v](#) [w](#) [x](#) [y](#) [z](#)

Search

by package name:

the source code (via [codesearch](#)):

Browse by prefix: [0](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [a](#) [b](#) [c](#) [d](#) [e](#) [f](#) [g](#) [h](#) [i](#) [j](#) [k](#) | [lib-](#) [lib3](#) [liba](#) [libb](#) [libc](#) [libd](#) [libe](#) [libf](#) [libg](#) [libh](#) [libi](#) [libj](#) [libk](#) [libl](#) [libm](#) [libn](#) [libo](#) [libp](#) [libq](#) [libr](#) [libs](#) [libt](#) [libu](#) [libv](#) [libw](#) [libx](#) [liby](#) [libz](#) [m](#) [n](#) [o](#) [p](#) [q](#) [r](#) [s](#) [t](#) [u](#) [v](#) [w](#) [x](#) [y](#) [z](#) | [Browse by page](#)

Debsources — Copyright (C) 2011–2015, [The Debsources developers](#). License: [GNU AGPLv3+](#).
Hosted source files are available under their own [copyright and licenses](#).
Source code: [Git](#) Contact: info@sources.debian.net. Last update: Wed, 08 Apr 2015 22:33:20 -0000.

hosted by



Source code display

File: cowsay

file content (199 lines) | stat: -rwxr-xr-x 4,421 bytes

[parent folder](#) | [download](#) | [duplicates \(3\)](#)

```
1  #!/usr/bin/perl
2
3  ##
4  ## Cowsay 3.03
5  ##
6  ## This file is part of cowsay. (c) 1999-2000 Tony Monroe.
7  ##
8
9  use Text::Tabs qw(expand);
10 use Text::Wrap qw(wrap fill $columns);
11 use File::Basename;
12 use Getopt::Std;
13 use Cwd;
14
15 if (${^UTF8LOCALE}) {
16     binmode STDIN, ':utf8';
17     binmode STDOUT, ':utf8';
18     require Encode;
19     eval { $_ = Encode::decode_utf8($_,1) } for @ARGV;
20 }
21
22 $version = '3.03';

```

Cowsay
See? Cowsay variables are declared here.

```
23 $progname = basename($0);
24 $eyes = 'oo';
25 $tongue = " ";
26 $cowpath = $ENV{'COWPATH'} || '/usr/share/cowsay/cows';
27 @message = ();
28 $thoughts = '';
```

package info (click to toggle)
cowsay 3.03+dfsg1-4

- links: [PTS](#), [VCS](#)
- area: main
- in suites: wheezy
- size: 616 kB
- ctags: 39
- SLOC: perl: 164; sh: 68;
- python: 16; makefile: 7

So what?

Is this really useful?

“I want to check the source code of *cowsay*. What do?”

The old way

```
cd /tmp
apt-get source cowsay
cd cowsay-3.03+dfsg1
...
cd ..
rm -r cowsay-3.03+dfsg1
```

Note that it only works on Debian(-based) systems.

So what?

Is this really useful?

“I want to check the source code of *cowsay*. What do?”

The old way

```
cd /tmp
apt-get source cowsay
cd cowsay-3.03+dfsg1
...
cd ..
rm -r cowsay-3.03+dfsg1
```

Note that it only works on Debian(-based) systems.

The new way

```
lynx http://sources.debian.net/src/cowsay/
```

Almost runs on your typewriter.

Table of contents

1 Introduction

2 Features

- Debsources' features
- What's new?
- Roadmap

3 Technologies

4 Research platform

5 Hacking

Source code browsing

Syntax highlighting

For all languages supported by `highlight.js`: C, C++, Java, Python, Ruby, Makefile... and 112 others.

Source code browsing

Syntax highlighting

For all languages supported by [highlight.js](#): C, C++, Java, Python, Ruby, Makefile... and 112 others.

Included versions

Packages in

hamm, slink, potato, woody, sarge, etch, lenny, squeeze, oldoldstable,
oldstable, stable, testing, unstable, experimental,
oldoldstable-proposed-updates, oldstable-proposed-updates,
proposed-updates, testing-proposed-updates, oldoldstable-updates,
oldstable-updates, stable-updates, wheezy-backports, jessie-backports,
squeeze-lts

are in Debsources.

Searching

You can search for:

- Packages
- Files
- File content
 - ▶ ctags
 - ▶ regular expressions through `codesearch.debian.net`

Content indexing

All search data is indexed in Postgres, resulting in pretty decent performances (in spite of sub-optimal disk I/O).

Advanced search

Package search

search for a package and allow to browse through its source code

package

filter by suite

File search

search for a specific source code file and display it

[by SHA256 hash](#)

sha256

within package (optional)

Code search

search all available source code for occurrences of specific features (regex matches, identifiers, etc.)

[via Debian code search](#)

regular expression

[by ctags](#)

tag

within package (optional)

Annotations and highlighting

Use cases

- I'm a **developer**: *I want to share a pointer to a precise location in the source code of package X.*
- I'm a **user**: *An errors point me to line 42 in file Y, and I look for support on IRC.*
- I'm a **static source code checker**: *I found an issue in file Z, line 51.*

Annotations and highlighting

`http://sources.debian.net/src/
package/version/path/to/file.c?hl=a:b&msg=a:b:c#LXX`

package: cowsay
version: 3.03-3
path: cowsay
highlight: 32:36
message: 30:Debian:rocks

`sources.debian.net/src/cowsay/3.03-3/cowsay?hl=32:36&msg=30:Debian:rocks`

```
30  ## One of these days, we'll get it ported to Windows. Yeah, right.  
  
Debian  
rocks  
  
31  
32  if (($^O eq "MSWin32") or ($^O eq "Windows_NT")) { ## Many perls, eek!  
33      $pathsep = ':'  
34  } else {  
35      $pathsep = '.'  
36  }
```

Annotations and highlighting

Developer: I want to share a precise location in the source code of package X.

```
15 $version = "3.03";  
16 $progname = basename($0);  
17 $eyes = "oo";
```

About the eyes...

Hey, should we change the default?

```
18 $tongue = " ";  
19 $cowpath = $ENV{'COWPATH'} || '/usr/share/cowsay/cows';  
20 @message = ();  
21 $thoughts = "";
```

Annotations and highlighting

User: I can't compile software X, it fails at line 42 in the file Y.

```
47 | int   main( int argc, char *argv[] );
48 | void  showDetails( MemoryStoragePtr );
49 | void  showHeader( CSAHeaderPtr )
```

Compilation error

My compiler fails here. What can I do?

```
50 | void  showGroups( CSAGrpHeaderPtr );
51 | void  showSemaphores( CSASemHeaderPtr );
52 |
```

Annotations and highlighting

Static analyzer: I found a semantic error in file Y.

```
247     len *= 2;
248 }
249
250 if ((res == (size_t) -1) || (res == (size_t) - 1))
```

Coccinelle

The same argument is used twice in a condition

```
251 {
252     /* The string cannot be converted. */
253     if (use_malloc)
254     {
255         free (wmessage);
```

Duplicated files

All the files are in the database, along with their checksum.
The duplicates can be computed, for every file.

File: COPYING

file content (674 lines) | stat: -rw-r--r-- 35,147 bytes

[parent folder](#) | [download](#) | [duplicates \(4309\)](#)

```
1 |          GNU GENERAL PUBLIC LICENSE
2 |          Version 3, 29 June 2007
3 |
```

[package info \(click to toggle\)](#)

Duplicated files

Checksum: 8ceb4b9ee5adedde47b31e975c1d90c73ad27b6b165a1dc

4309 results:

- [3depict/0.0.10-1/COPYING](#)
- [3depict/0.0.16-2.1/COPYING](#)
- [3dldf/2.0.3+dfsg-2/COPYING](#)
- [3dldf-doc/2.0.3+ndfsg-2/COPYING](#)
- [4store/1.1.4-2/COPYING](#)
- [4store/1.1.5-1/COPYING](#)
- [4store/1.1.6-1/COPYING](#)
- [aac-tactics/0.2.pl2-7/COPYING](#)
- [aac-tactics/0.4-3/COPYING](#)
- [aaphoto/0.38-2/COPYING](#)
- [aaphoto/0.41-1.1/COPYING](#)
- [aaphoto/0.43.1-1/COPYING](#)
- [aaphoto/0.43.1-3/COPYING](#)
- [abby/0.4.7-1/COPYING](#)

Integration in the Debian infrastructure

Codesearch

`http://codesearch.debian.net` is used for regular expression searches, and redirects back to Debsources to consult results.

Credits: Michael Stapelberg

Integration in the Debian infrastructure

Codesearch

<http://codesearch.debian.net> is used for regular expression searches, and redirects back to Debsources to consult results.

Credits: Michael Stapelberg

Package tracking systems

The old PTS (<http://packages.qa.debian.org>) and the new tracker (<http://tracker.debian.org>) provide links to Debsources (“browse source code”).

Credits: Paul Wise, Raphael Hertzog

Integration in the Debian infrastructure

Codesearch

<http://codesearch.debian.net> is used for regular expression searches, and redirects back to Debsources to consult results.

Credits: Michael Stapelberg

Package tracking systems

The old PTS (<http://packages.qa.debian.org>) and the new tracker (<http://tracker.debian.org>) provide links to Debsources (“browse source code”).

Credits: Paul Wise, Raphael Hertzog

Need to embed code somewhere?

`<iframe>`s embedding of files content is supported (see documentation).

(or talk to use for more proper integration)

- Source code metrics for every package
- Plugins: disk size, ctags, sloccount

- Source code metrics for every package
- Plugins: disk size, ctags, sloccount

package info (click to toggle)

chromium-browser 41.0.2272.118-1

- links: [PTS](#), [VCS](#)
- area: main
- in suites: jessie, sid
- size: 2,241,900 kB
- ctags: 1,909,592
- SLOC: cpp: 9,691,826; ansic: 3,341,113; python: 712,689; asm: 518,779; xml: 208,926; java: 169,820; sh: 119,353; perl: 68,907; makefile: 28,311; yacc: 13,305; objc: 11,385; tcl: 3,186; cs: 2,225; sql: 2,217; lex: 2,215; lisp: 1,349; pascal: 1,256; awk: 407; ruby: 155; sed: 53; php: 14; exp: 11

Aggregated statistics are available at
<http://sources.debian.net/stats/>.

Metrics

- Disk usage
- SLOC (source lines of code)
- Number of source packages
- Number of files
- Number of ctags (symbols)

Statistics

Currently in *sid*:

Source files	11,787,950
Source packages	23,846
Disk usage	228,599,736 kB
Ctags	127,884,129
Source lines of code	1,082,453,728

Statistics

Currently in *sid*:

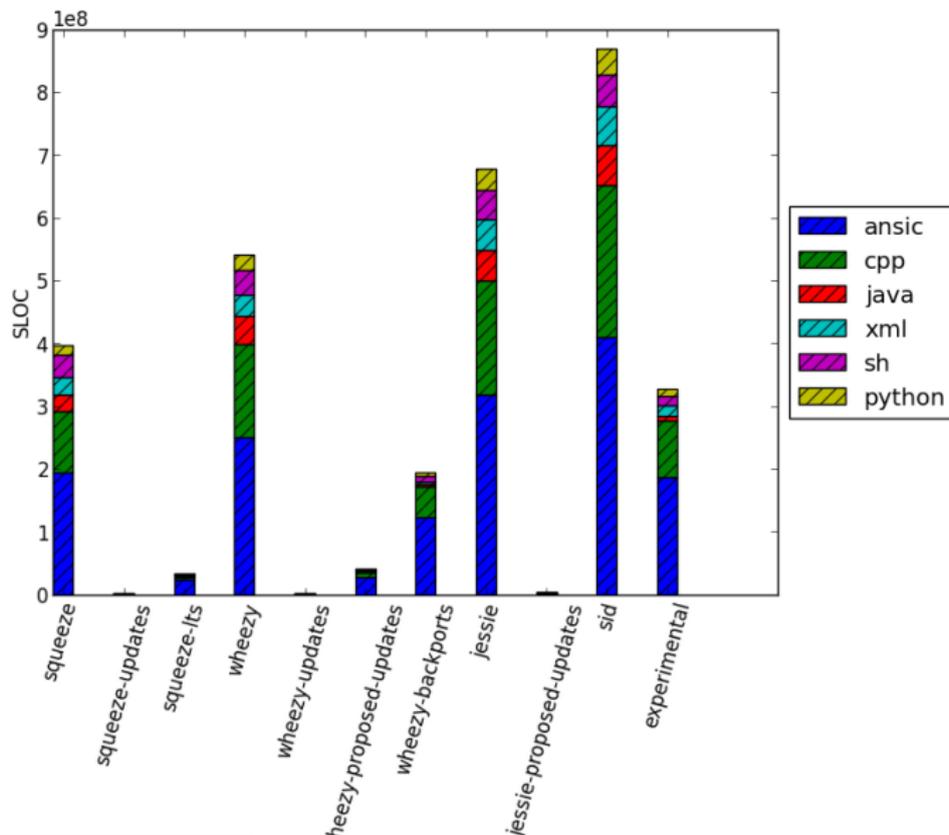
Source files	11,787,950
Source packages	23,846
Disk usage	228,599,736 kB
Ctags	127,884,129
Source lines of code	1,082,453,728

- ① C: 439,197,216
- ② C++: 275,342,652
- ③ Python: 46,067,009

See <http://sources.debian.net/stats/sid/> for more

Statistics

Fancy graphs



API

All functionalities available via the Web UI, are also available via a JSON-based HTTP API:

Examples

```
curl http://sources.debian.net/api/ping/
{ "status": "ok",
  "http_status_code": 200,
  "last_update": "Mon, 17 Aug 2015 10:43:27 -0000" }
```

```
curl http://s.d.n/api/info/package/cowsay/3.03-3/
{
  "pkg_infos": {
    "suites": [
      "woody"
    ]
  }
  ...
}
```

Documentation at <http://sources.debian.net/doc/api/>.

What's new?

What's new?

- Many new features since DebConf14

What's new?

- Many new features since DebConf14
- Outreachy student: Jingjie Jiang (sophiejjj)

What's new?

- Many new features since DebConf14
- Outreachy student: Jingjie Jiang (sophiejjj)
- GSoC students: Clément Schreiner (clemux), Orestis Ioannou (orestis)

What's new?

- Many new features since DebConf14
- Outreachy student: Jingjie Jiang (sophiejjj)
- GSoC students: Clément Schreiner (clemux), Orestis Ioannou (orestis)
- Many new contributors:
 - ▶ Stefano Zacchiroli: 721
 - ▶ Matthieu Caneill: 483
 - ▶ Orestis Ioannou: 118
 - ▶ sophiejjj: 58
 - ▶ Clément Schreiner: 19
 - ▶ Jason Pleau: 12
 - ▶ Akshita Jha: 7
 - ▶ Jingjie Jiang: 5
 - ▶ Paul Wise: 1
 - ▶ tessa joseph: 1
 - ▶ sodamatt: 1
 - ▶ Luciano Bello: 1
 - ▶ Christophe Siraut: 1
 - ▶ James McCoy: 1
 - ▶ Tapasweni Pathak: 1

What's new?

Multiple pop-up messages

Credits: Orestis Ioannou and Jason Pleau.

Application: generated code annotations.

```
22 $version = "3.03";
```

Cowsay

See? Cowsay variables are declared here.

```
23 $programe = basename($0);
```

```
24 $eyes = "oo";
```

Oh hello

I am just another dummy pop-up example

```
25 $tongue = " ";
```

```
26 $cowpath = $ENV{'COWPATH'} || '/usr/share/cowsay/cows';
```

```
27 @message = ();
```

```
28 $thoughts = "";
```

What's new?

Blueprints support

Credits: Jingjie Jiang.

Application: new apps plugged.

Blueprints

- Flask apps embedded and plugged together (decentralization)
- Implied a big refactoring
- Enabled the development of new features (GSoC)

What's new?

Detailed directory listing

Credits: Jingjie Jiang

Folder: 3.03+dfsg1-10 [show hidden \(1\)](#)

			.. (parent)
	drwxr-xr-x	4,096	cows
	drwxr-xr-x	4,096	debian
	-rw-r--r--	931	ChangeLog
	-rw-r--r--	385	INSTALL
	-rw-r--r--	1,116	LICENSE
	-rw-r--r--	445	MANIFEST
	-rw-r--r--	1,610	README
	-rw-r--r--	879	Wrap.pm.diff
	-rwxr-xr-x	4,421	cowsay
	-rw-r--r--	4,693	cowsay.6
	-rwxr-xr-x	2,275	install.sh
	-rw-r--r--	631	pgp_public_key.txt

What's new?

File edition in-browser

Credits: Raphael Geissert

File edition

A plugin for Iceweasel and Chromium enables the edition of files in your browser.

A patch ready-to-be-sent™ is generated!

It makes the entire Debian archive editable from a browser.

What's new?

Blueprint app: License information

Credits: Orestis Ioannou (GSoC student)

debian/copyright

Part of the archive uses **machine-readable debian/copyright**

What's new?

Blueprint app: License information

Credits: Orestis Ioannou (GSoC student)

debian/copyright

Part of the archive uses **machine-readable debian/copyright**

- Parse and display this information in the web interface

What's new?

Blueprint app: License information

Credits: Orestis Ioannou (GSoC student)

debian/copyright

Part of the archive uses **machine-readable debian/copyright**

- Parse and display this information in the web interface
- Compute statistics about license usage

What's new?

Blueprint app: License information

Credits: Orestis Ioannou (GSoC student)

debian/copyright

Part of the archive uses **machine-readable debian/copyright**

- Parse and display this information in the web interface
- Compute statistics about license usage
- SPDX (generic license exchange format) export of the copyright file

What's new?

Blueprint app: License information

Credits: Orestis Ioannou (GSoC student)

debian/copyright

Part of the archive uses **machine-readable debian/copyright**

- Parse and display this information in the web interface
- Compute statistics about license usage
- SPDX (generic license exchange format) export of the copyright file
- API

What's new?

Blueprint app: License information

DEBCOPYRIGHT

 Home Search Documentation Stats About

debian / copyright / aircrack-ng / 1:1.2-0~beta3-4

Package: aircrack-ng / 1:1.2-0~beta3-4

Header

Format: <http://www.debian.org/doc/packaging-manuals/copyright-format/1.0/>

Upstream-Name: Aircrack-ng

Upstream-Contact: Thomas d'Otreppe <tdotreppe@aircrack-ng.org>

Source: <http://download.aircrack-ng.org/aircrack-ng-1.2-beta3.tar.gz>

Files

*

Copyright: 2006-2014, Thomas d'Otreppe

License: [GPL-2+](#)

[lib/csharp/MonoExample/NDesk-dbus/*](#)

Copyright: 2006-2007, Alp Tokar <alp@atoker.com>

License: [MIT](#)

What's new?

Blueprint app: Patch tracker

Credits: Orestis Ioannou (GSoC student)

Tracking patches of a package

- Intended as a replacement of the old patch-tracker

What's new?

Blueprint app: Patch tracker

Credits: Orestis Ioannou (GSoC student)

Tracking patches of a package

- Intended as a replacement of the old patch-tracker
- Currently supports 3.0 (quilt) format

What's new?

Blueprint app: Patch tracker

Credits: Orestis Ioannou (GSoC student)

Tracking patches of a package

- Intended as a replacement of the old patch-tracker
- Currently supports 3.0 (quilt) format
- Syntax-highlighting

What's new?

Blueprint app: Patch tracker

Credits: Orestis Ioannou (GSoC student)

Tracking patches of a package

- Intended as a replacement of the old patch-tracker
- Currently supports 3.0 (quilt) format
- Syntax-highlighting
- API

What's new?

Blueprint app: Patch tracker

Package: beignet / 1.0.0-1

Metadata

Package	Version	Patches format
---------	---------	----------------

beignet	1.0.0-1	3.0 (quilt)
---------	---------	-------------

Patch series

Patch	File delta	Description	View	Download
Debian compliant compiler flags handling	CMakeLists.txt 7 +++---- 1 file changed, 3 insertions(+), 4 deletions(-)	respect cflags,etc and be verbose	View	download
Enhance debug output	src/cl_utils.h 1 + 1 file changed, 1 insertion(+)	enhance debug output	View	download

What's new?

Asynchronous updater

Credits: Clément Schreiner (GSoC student)

Asynchronous updater

- Rewrite of the stages of the updater (add-package, compute-stats, gc, etc)
- Using *celery* to spawn independent and asynchronous tasks.

For more information about these features

- Watch the video of the GSoC session (happened earlier this afternoon!)
- Features soon available at `sources.debian.net` (currently being merged and deployed).

What's new?

And many many other features...

- Refactoring

What's new?

And many many other features...

- Refactoring
 - ▶ Debsources as a top-level Python module

What's new?

And many many other features...

- Refactoring
 - ▶ Debsources as a top-level Python module
 - ▶ Configuration loader

What's new?

And many many other features...

- Refactoring
 - ▶ Debsources as a top-level Python module
 - ▶ Configuration loader
 - ▶ Flake8 compliance (Zack, Jingjie Jiang, and others)

What's new?

And many many other features...

- Refactoring
 - ▶ Debsources as a top-level Python module
 - ▶ Configuration loader
 - ▶ Flake8 compliance (Zack, Jingjie Jiang, and others)
- Test coverage (Jingjie Jiang, Clément Schreiner, and others)

What's new?

And many many other features...

- Refactoring
 - ▶ Debsources as a top-level Python module
 - ▶ Configuration loader
 - ▶ Flake8 compliance (Zack, Jingjie Jiang, and others)
- Test coverage (Jingjie Jiang, Clément Schreiner, and others)
 - ▶ 84% now!

What's new?

And many many other features...

- Refactoring
 - ▶ Debsources as a top-level Python module
 - ▶ Configuration loader
 - ▶ Flake8 compliance (Zack, Jingjie Jiang, and others)
- Test coverage (Jingjie Jiang, Clément Schreiner, and others)
 - ▶ 84% now!
- Case-insensitive package name search (Akshita Jha)

What's new?

And many many other features...

- Refactoring
 - ▶ Debsources as a top-level Python module
 - ▶ Configuration loader
 - ▶ Flake8 compliance (Zack, Jingjie Jiang, and others)
- Test coverage (Jingjie Jiang, Clément Schreiner, and others)
 - ▶ 84% now!
- Case-insensitive package name search (Akshita Jha)
- Better statistics charts (Orestis Ioannou)

What's new?

And many many other features...

- Refactoring
 - ▶ Debsources as a top-level Python module
 - ▶ Configuration loader
 - ▶ Flake8 compliance (Zack, Jingjie Jiang, and others)
- Test coverage (Jingjie Jiang, Clément Schreiner, and others)
 - ▶ 84% now!
- Case-insensitive package name search (Akshita Jha)
- Better statistics charts (Orestis Ioannou)
- Python3 support (Zack)

Static analysis

- Automatic runs of static analysis tools (e.g. clang, coccinelle) on all Debian packages
- Statistics gathering on bugs evolution
- → Debile, Firewoes

Roadmap

And many smaller items

- more live statistics (about licenses, patches, etc)

Roadmap

And many smaller items

- more live statistics (about licenses, patches, etc)
- file name search

Roadmap

And many smaller items

- more live statistics (about licenses, patches, etc)
- file name search
- binary package → source package redirection

Roadmap

And many smaller items

- more live statistics (about licenses, patches, etc)
- file name search
- binary package → source package redirection
- tarball-in-tarball support

Roadmap

And many smaller items

- more live statistics (about licenses, patches, etc)
- file name search
- binary package → source package redirection
- tarball-in-tarball support
- 100% test suite coverage (one day...)

Roadmap

And many smaller items

- more live statistics (about licenses, patches, etc)
 - file name search
 - binary package → source package redirection
 - tarball-in-tarball support
 - 100% test suite coverage (one day...)
 - file-level deduplication
 - ▶ **select count(*) from** checksums; → 35'370'653
 - ▶ **select count(distinct sha256) from** checksums; → 15'822'745
- ⇒ deduplicated core: ≈ 45%

Table of contents

1 Introduction

2 Features

- Debsources' features
- What's new?
- Roadmap

3 Technologies

4 Research platform

5 Hacking

Technologies

What languages and technologies do we use?

- **Code base:** (almost) entirely in Python

Technologies

What languages and technologies do we use?

- **Code base:** (almost) entirely in Python
- **Web application:** Flask framework, Jinja2 templates, HTML/CSS/Javascript

Technologies

What languages and technologies do we use?

- **Code base:** (almost) entirely in Python
- **Web application:** Flask framework, Jinja2 templates, HTML/CSS/Javascript
- **Database:** PostgreSQL

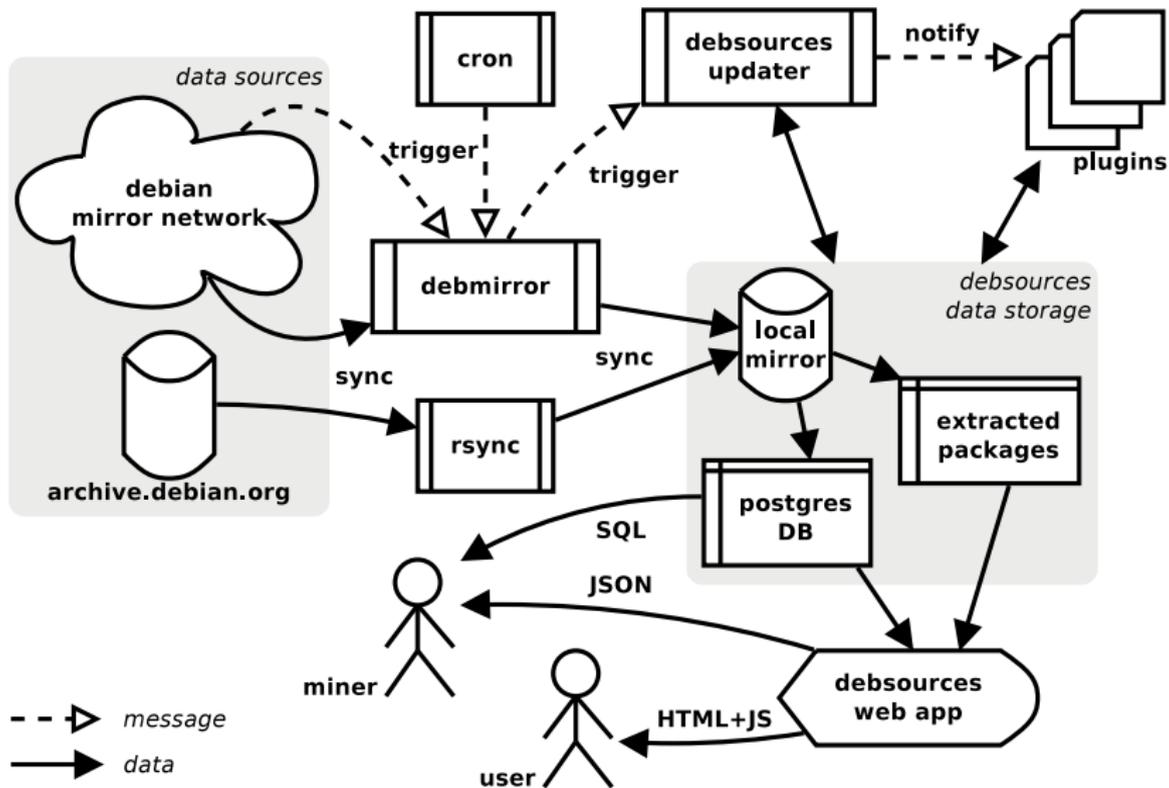
Technologies

What languages and technologies do we use?

- **Code base:** (almost) entirely in Python
- **Web application:** Flask framework, Jinja2 templates, HTML/CSS/Javascript
- **Database:** PostgreSQL
- Apache web server, SQLAlchemy, ...

Overview

Architecture



Overview

data model (excerpt)

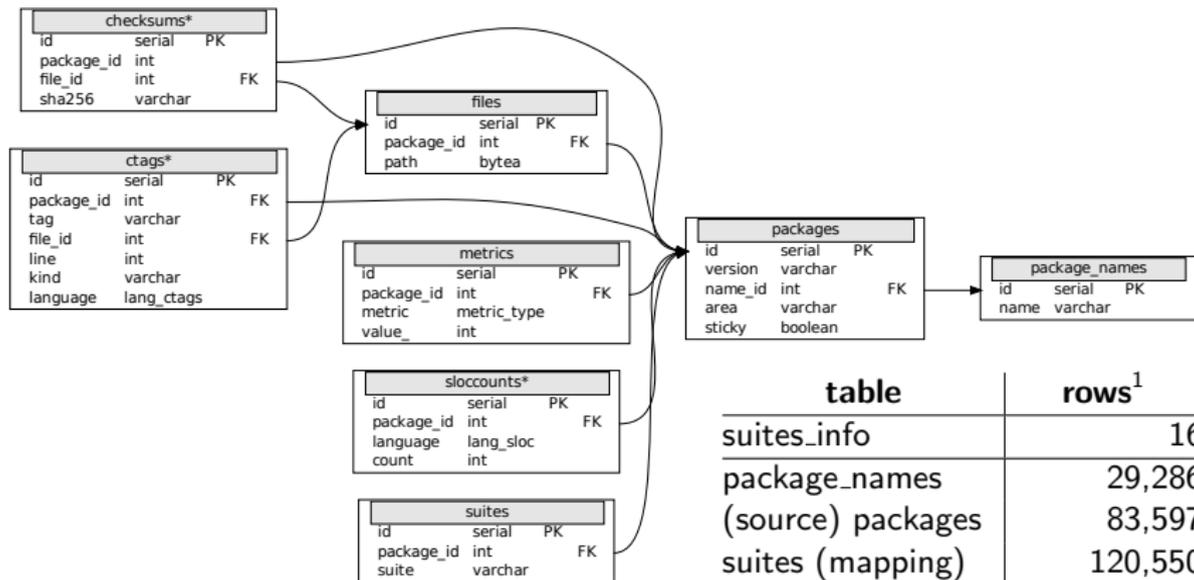


table	rows ¹
suites_info	16
package_names	29,286
(source) packages	83,597
suites (mapping)	120,550
metrics (e.g., du)	83,597
sloccounts	298,360
checksums	35,370,653
ctags	358,773,259

¹snapshot, 31 July 2014

Disk usage

- unpacked sources: 805 GB
- PostgreSQL DB: 145 GB
- Source mirror: 135 GB

Hosting requirements: \approx 1.1 TB

(17 August 2015)

Disk usage

- unpacked sources: 805 GB
- PostgreSQL DB: 145 GB
- Source mirror: 135 GB

Hosting requirements: \approx 1.1 TB

(17 August 2015)

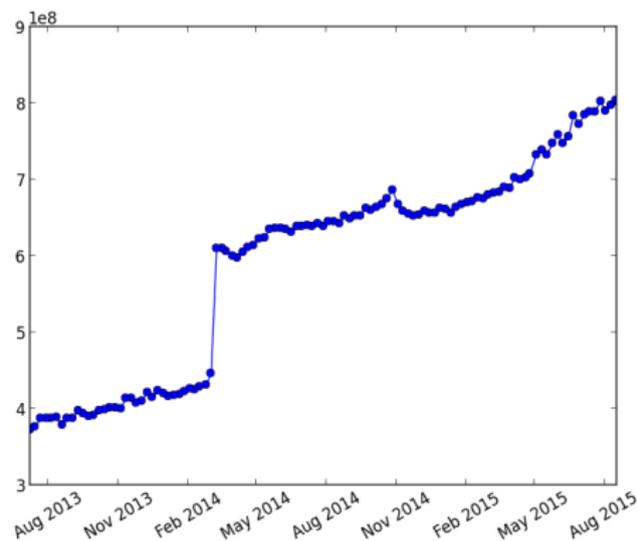


Figure: unpacked sources trend (peek due to archive.d.o injection)

Table of contents

1 Introduction

2 Features

- Debsources' features
- What's new?
- Roadmap

3 Technologies

4 Research platform

5 Hacking

Facts

- Debsources is a **huge** software collection.
- Homogeneous: all software follow Debian's packaging format.
- It is up-to-date.

Research platform

Facts

- Debsources is a **huge** software collection.
- Homogeneous: all software follow Debian's packaging format.
- It is up-to-date.

Software evolution

- 20 years of source code evolution.
- Plugins to compute stats.

Research platform

Facts

- Debsources is a **huge** software collection.
- Homogeneous: all software follow Debian's packaging format.
- It is up-to-date.

Software evolution

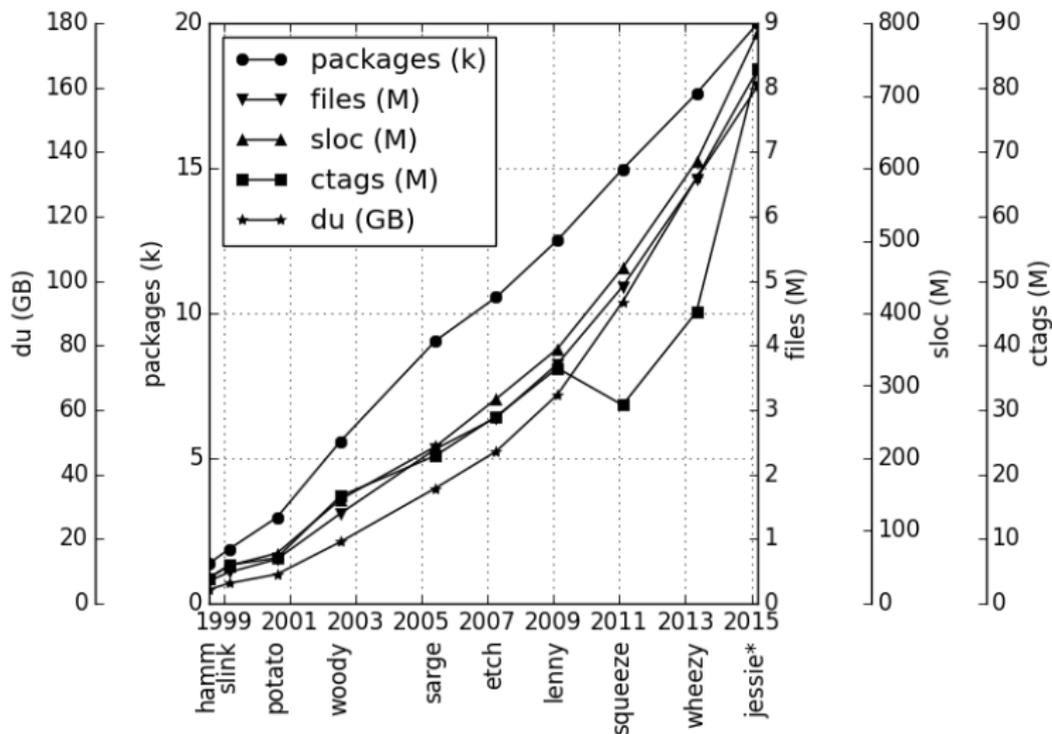
- 20 years of source code evolution.
- Plugins to compute stats.

Nice charts can be computed with this!

Example: What are the trending programming languages?

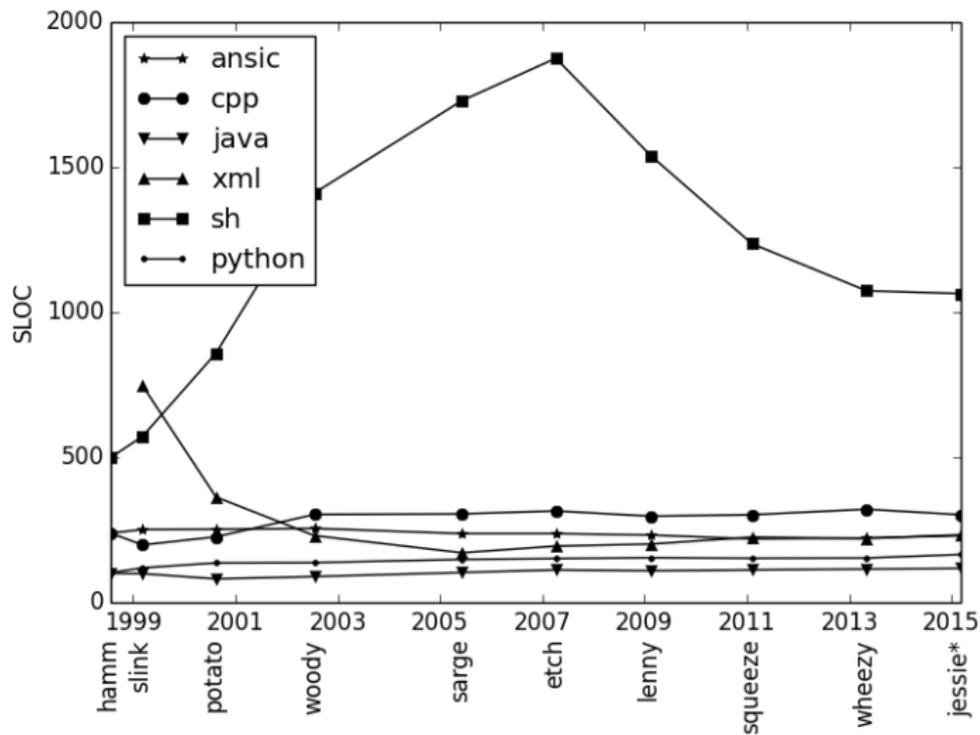
Research platform

Software metrics evolution over Debian releases



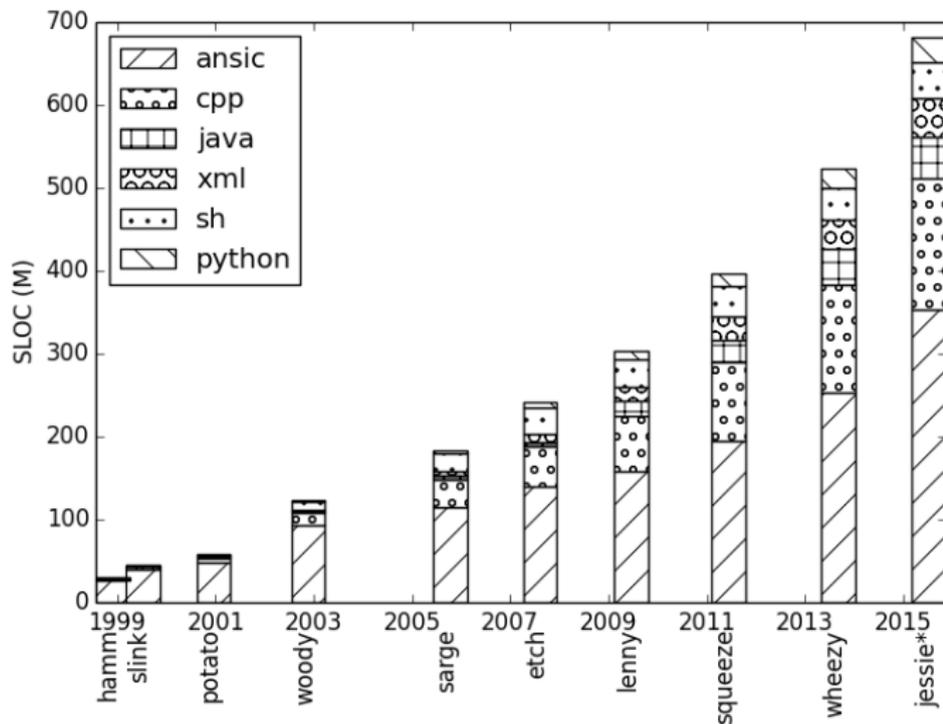
Research platform

File size per language, evolution over Debian releases



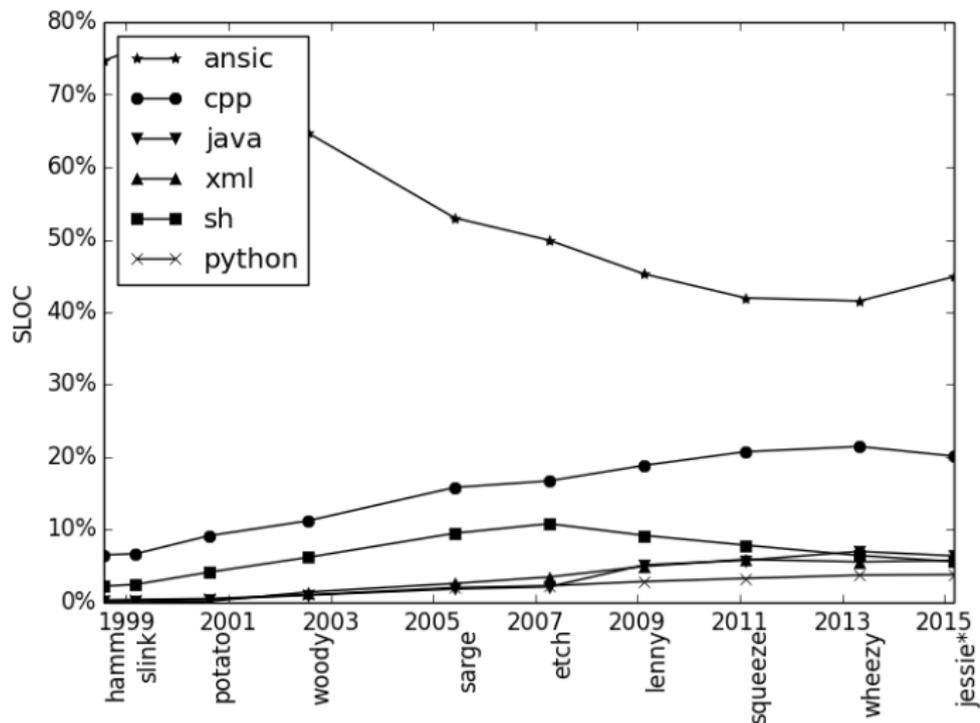
Research platform

Absolute evolution of SLOC per language, over Debian releases



Research platform

Relative evolution of SLOC per language, over Debian releases



- Matthieu Caneill, Stefano Zacchiroli. **Debsources: Live and Historical Views on Macro-Level Software Evolution.** In proceedings of ESEM 2014: 8th International Symposium on Empirical Software Engineering and Measurement.
- Stefano Zacchiroli. **The Debsources Dataset: Two Decades of Debian Source Code Metadata.** To appear in proceedings of MSR 2015: The 12th Working Conference on Mining Software Repositories.

You can find the PDFs of the articles on <http://sources.debian.net/doc/>.

Table of contents

- 1 Introduction
- 2 Features
 - Debsources' features
 - What's new?
 - Roadmap
- 3 Technologies
- 4 Research platform
- 5 Hacking

Hacking

How can I contribute?

Step 1: clone Debsources git repository

```
git clone git://anonscm.debian.org/qa/debsources.git
```

Hacking

How can I contribute?

Step 1: clone Debsources git repository

```
git clone git://anonscm.debian.org/qa/debsources.git
```

Step 2: Set-up a development environment

- Follow the instructions in the **HACKING** file,
- Or use Docker! `bin/debsources-docker-build &&`
`bin/debsources-docker-run` will setup a **Docker container** with all batteries included: dependencies, database, test data, configuration.

Hacking

How can I contribute?

Step 1: clone Debsources git repository

```
git clone git://anonscm.debian.org/qa/debsources.git
```

Step 2: Set-up a development environment

- Follow the instructions in the **HACKING** file,
- Or use Docker! `bin/debsources-docker-build` &&
`bin/debsources-docker-run` will setup a **Docker container** with all batteries included: dependencies, database, test data, configuration.

Step 3: open your editor and hack!

Bugs list: <https://bugs.debian.org/cgi-bin/pkgreport.cgi?pkg=qa.debian.org;tag=debsources>

or: implement your own **plugin** (see examples), add **features**, etc.

Thanks! Questions?

Stefano Zacchiroli, Matthieu Caneill

info@sources.debian.net

Slides: Licensed CC-BY-SA.

<http://upsilon.cc/~zack/talks/2015/20150818-dc15-debsources.pdf>