

# Debian Contributors

Laura Arjona Reina

`laura.arjona@upm.es`

July 2014

## Abstract

Debian contributors is a new software for gathering, measuring, and crediting the active participants in the Debian Project, for any kind of contribution (whether it is package uploads, system administration, squashing bugs, wiki or website editing, spam fighting, participation in the video team, volunteering in a booth...).

It is a web application with a client reporting tool supporting a wide range of data mining options. It offers comprehensive identity management which allows to credit a person's work across multiple emails, logins or gpg keys. Results for several data sources and proof-of-concepts are visible in <https://contributors.debian.org/> currently retrieving 16 types of contributions (only one of them was previously tracked). Source code is AGPL licensed, and quickly evolving in the public Git repository of the Debian project.

Besides being a technical tool, it is a social experiment: the intent is to change the way the Debian community perceives itself, by making a diverse range of contributions visible, and showing them with the same level of importance.

## What is Debian Contributors?

Debian Contributors is a website with a list of people and teams actually contributing to the Debian Project. Behind this web end, there is a software infrastructure to allow contributions data from different sources, to be automatically gathered in a standard format.

- Website: <https://contributors.debian.org>
- Source code: <http://anonscm.debian.org/gitweb/?p=nm/dc.git>
- Authors: Enrico Zini, Debian New Members Front Desk team.
- Technologies used: Python, Django, JSON.
- More info: <https://wiki.debian.org/DebianContributors/DcDocs>

The infrastructure is designed in a way that each data source is independent and it can be created and maintained for the related team, distributing the work of creating the “big picture”.

Debian Contributors has the goal of making non-packaging contributions more visible in Debian, attracting new members with different skills and showing how diverse the project is. It creates a “new hat” in the Debian project, which is earned or lost depending not on the acquired reputation but on the actual contributions of each person.

Like this, a more accurate personal profile is created of each contributor, even if she is not a member of the Debian project yet. Indeed, it can be used to help asses contributions of applicants to the Debian New Member process, or it can be a tool for the Debian Teams to self-diagnose issues about burden distribution, people entering/leaving the team, etc.

# Development and implementation details

## Data sources

A **Data Source** is a JSON structure with an array of records, containing the identity of the contributor and a list of contributions.

The type of identifier of a contributor, currently, can be debian.org or Alioth<sup>1</sup> login names, emails or OpenPGP. A human-readable version for the identifier, like a person's full name, can be added.

A Contribution is an object detailing one kind of contribution over time:

```
Contribution := { type: string,
                  begin: optional_Date,
                  end: optional_Date,
                  url: optional_string }
Date := "yyyy-mm-dd"
```

The type of contribution is defined in the data source configuration, in the website [contributors.debian.org](https://contributors.debian.org). The url is an URL pointing to a page that will list details of this kind of contributions for this person, if available.

## Examples

A minimal submission (for example, to “manually” credit some contributions that are not automatically tracked):

```
[
  { "id": { "type": "email", "id": "larjona@fsfe.org" },
    "contributions": { "type": "volunteering_in_DebianBooth" } },
  { "id": { "type": "email", "id": "cosasdelio@inventati.org" },
    "contributions": { "type": "volunteering_in_DebianBooth" } }
]
```

A more detailed submission:

```
[
  { "id": [ { "type": "login", "id": "enrico" } ],
    "contributions": [
      { "type": "talking", "begin": "2010-01-01", "end": "2013-11-16",
        "url": "http://talks.debian.org/enrico"
      }
    ]
  },
  { "id": [ { "type": "login", "id": "93sam" } ],
    "contributions": [
      { "type": "talking", "begin": "1990-05-01", "end": "2013-11-16",
        "url": "http://talks.debian.org/93sam"
      }
    ]
  }
]
```

Submissions are sent via a multipart/form-data HTTP POST, manually, or via a commandline tool, `dc-tool`.

---

<sup>1</sup>Alioth is the name of the source forge of Debian. It runs FusionForge and it's available at <https://alioth.debian.org/>

## Data mining: the dc-tool

A tool is created to ease the process of gathering contributions. Once the Debian member has created the data source at [contributors.debian.org](https://contributors.debian.org) website, a single command will mine the corresponding source and post the data in the JSON format, to the website:

```
dc-tool --mine myconfigfile --post
```

The method for gathering the data is defined in the file `myconfigfile`. The configuration file follows the usual Debian RFC822/Yaml-like syntax, and it has two sections: the data source configuration section (with the data source name, and the authentication token, as they are defined in the website), and the data mining sections, with one block for each type of contribution that will be posted (type of contribution, method used, options for that method).

Example:

```
# Data source configuration section
source: nm.debian.org
auth_token: @secrets/auth_token.txt

# Data mining section
contribution: committer
method: gitdirs
# Configuration specific for gitdirs method
dirs: /srv/git.debian.org/git/collab-maint/*.git
url: https://alioth.debian.org/users/{user}/
```

## Data mining methods

There are several data mining methods:

- **files**: Recursively scan directories using file attributes to detect contributions. Generates ‘login’ types of identifiers, using the usernames of the system where it is run.
- **gitdirs**: Scan git directories using file attributes to detect contributions. Generates ‘login’ types of identifiers, using the usernames of the system where it is run.
- **gitlogs**: Scan git logs, taking note of committer and author activity. Generates ‘email’ types of identifiers, trusting whatever is in the git log.
- **mailfrom**: Scan email address from From: headers in mailboxes
- **postgres**: Perform data mining using a SQL query on a Postgres database.
- **svndirs**: Scan subversion directories using file attributes to detect contributions. Generates ‘login’ types of identifiers, using the usernames of the system where it is run.

Some options that can be applied to most of the methods are:

- **dirs**: directories to scan (one or more, or shell-style globbing).
- **url**: template used to build URLs to link to people’s contributions.
- **blacklist**, **whitelist**: for email-based methods

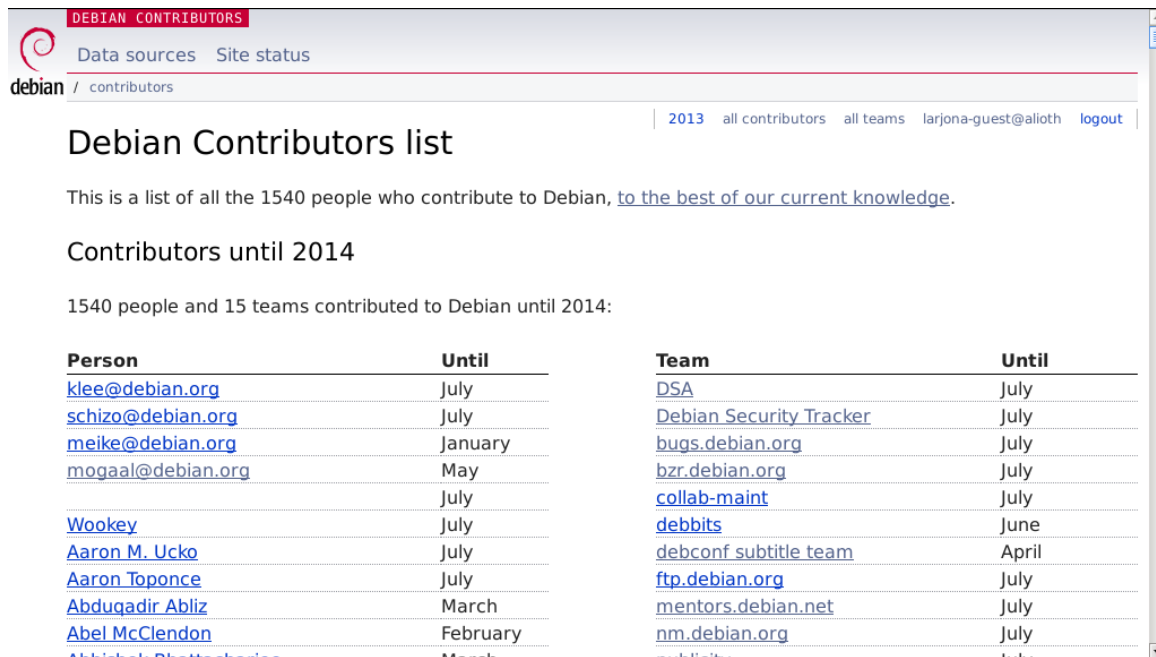
More information about `dc-tool` can be found in the source code repository<sup>2</sup>.

---

<sup>2</sup><http://anonscm.debian.org/gitweb/?p=nm/python-debiancontributors.git>

## Results: the contributors.debian.org website

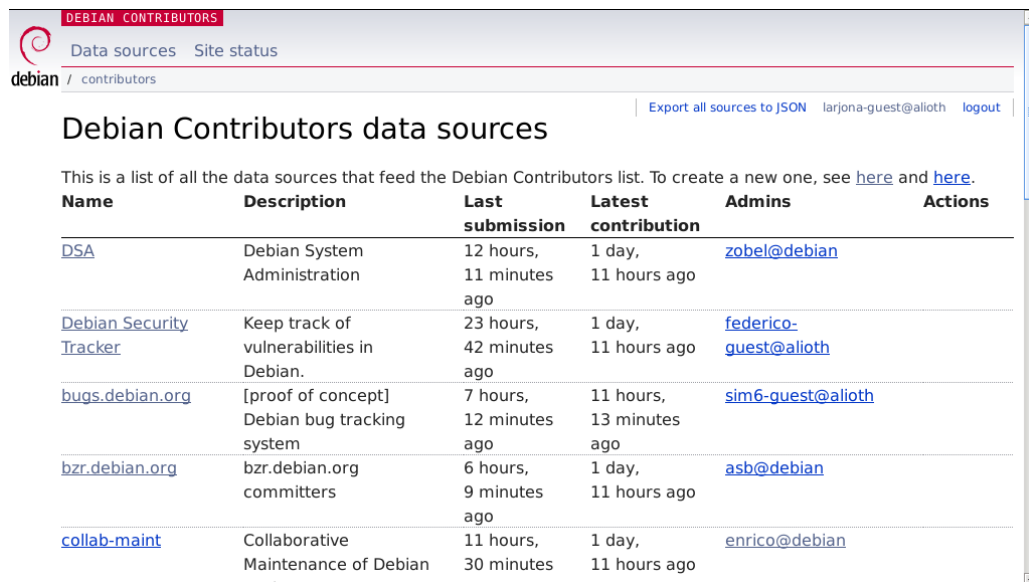
Two list of contributors are shown: for current year, and for the whole tracked history. The list consists on names or nicknames, the initial contribution, and the last contribution (see Figure 1).



Person	Until	Team	Until
<a href="mailto:klee@debian.org">klee@debian.org</a>	July	<a href="#">DSA</a>	July
<a href="mailto:schizo@debian.org">schizo@debian.org</a>	July	<a href="#">Debian Security Tracker</a>	July
<a href="mailto:meike@debian.org">meike@debian.org</a>	January	<a href="http://bugs.debian.org">bugs.debian.org</a>	July
<a href="mailto:mogaal@debian.org">mogaal@debian.org</a>	May	<a href="http://bazaar.debian.org">bazaar.debian.org</a>	July
	July	<a href="#">collab-maint</a>	July
<a href="#">Wookey</a>	July	<a href="#">debbugs</a>	June
<a href="#">Aaron M. Ucko</a>	July	<a href="#">debconf subtitle team</a>	April
<a href="#">Aaron Toponce</a>	July	<a href="http://ftp.debian.org">ftp.debian.org</a>	July
<a href="#">Abdugadir Abliz</a>	March	<a href="http://mentors.debian.net">mentors.debian.net</a>	July
<a href="#">Abel McClendon</a>	February	<a href="http://nm.debian.org">nm.debian.org</a>	July
<a href="#">Abhishek Bhattacharya</a>	March		

Figure 1: The Debian Contributors website

There is a list of teams/data sources, with name, description, last submission, latest contribution, and the person administrating the data source (see Figure 2).



Name	Description	Last submission	Latest contribution	Admins	Actions
<a href="#">DSA</a>	Debian System Administration	12 hours, 11 minutes ago	1 day, 11 hours ago	<a href="mailto:zobel@debian">zobel@debian</a>	
<a href="#">Debian Security Tracker</a>	Keep track of vulnerabilities in Debian.	23 hours, 42 minutes ago	1 day, 11 hours ago	<a href="mailto:federico-guest@alioth">federico-guest@alioth</a>	
<a href="http://bugs.debian.org">bugs.debian.org</a>	[proof of concept] Debian bug tracking system	7 hours, 12 minutes ago	11 hours, 13 minutes ago	<a href="mailto:sim6-guest@alioth">sim6-guest@alioth</a>	
<a href="http://bazaar.debian.org">bazaar.debian.org</a>	bazaar.debian.org committers	6 hours, 9 minutes ago	1 day, 11 hours ago	<a href="mailto:asb@debian">asb@debian</a>	
<a href="#">collab-maint</a>	Collaborative Maintenance of Debian packages	11 hours, 30 minutes ago	1 day, 11 hours ago	<a href="mailto:enrico@debian">enrico@debian</a>	

Figure 2: Data sources

The list of currently implemented data sources is in Table 1. A list of proposed data sources, pending to implement, can be found in Table 2.

Table 1: List of currently handled data sources

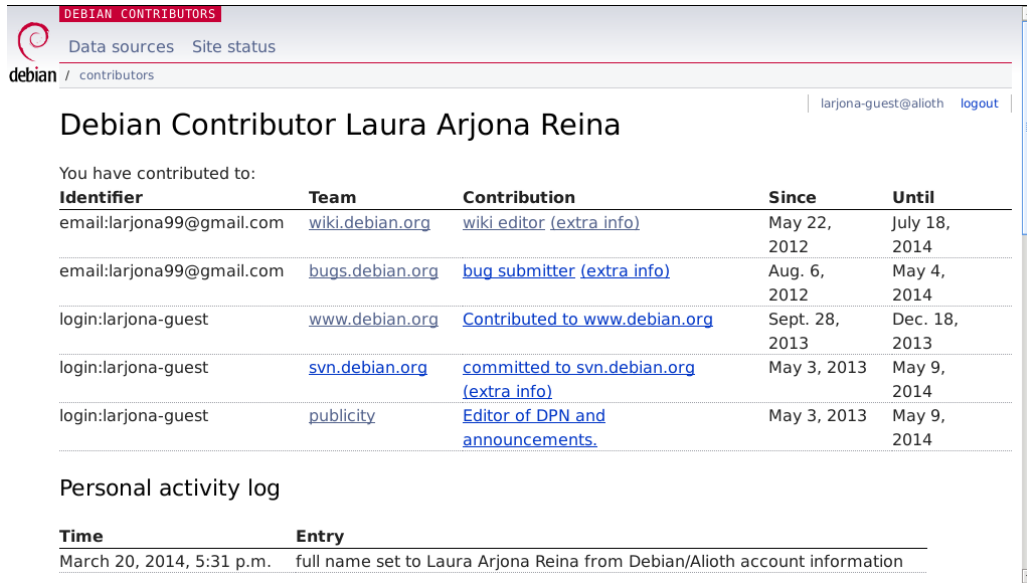
Name of Data Source	Description
DSA	Debian System Administration
Debian Security Tracker	Keep track of vulnerabilities in Debian.
bugs.debian.org	Debian bug tracking system
bzr.debian.org	bzr.debian.org committers
collab-maint	Collaborative Maintenance of Debian packages
debbits	the official blog of the Debian project
debconf subtitle team	Subtitles for DebConf videos.
ftp.debian.org	Debian package uploads
git.debian.org	git.debian.org committers
mentors.debian.net	collaboration platform for new Debian package maintainers
nm.debian.org	Debian New Member Process
publicity	Editing of project news and various announcements.
spamreview	Reviewers of messages reported as spam in the mailing lists
svn.debian.org	svn.debian.org committers
wiki.debian.org	Contributions from the Debian wiki
www.debian.org	Debian website committers

Table 2: Some proposed data sources, pending to implement

Name	Details
Debconf Participation	Data at the Debconf website could be processed
Debian Perl team	Could be done adapting PET (Package Entropy Tracker)
Debian blends	An activity tracker for the Blends can be adapted
l10n (packages, website)	A bot already tracks mails with specific tags in the subject
l10n (manuals, documentation)	Subversion repository
l10n (package descriptions)	web form submissions
l10n (debian installer)	debian installer repository, filter by .po files
Voters	List of voters for each votation is recorded in a txt file

## Contributors: handling identifiers

Clicking on a contributor in the website, a list of all their contributions is shown (Figure 3).



DEBIAN CONTRIBUTORS				
Data sources Site status				
debian / contributors				
Debian Contributor Laura Arjona Reina				
You have contributed to:				
Identifier	Team	Contribution	Since	Until
email:larjona99@gmail.com	<a href="http://wiki.debian.org">wiki.debian.org</a>	<a href="#">wiki editor (extra info)</a>	May 22, 2012	July 18, 2014
email:larjona99@gmail.com	<a href="http://bugs.debian.org">bugs.debian.org</a>	<a href="#">bug submitter (extra info)</a>	Aug. 6, 2012	May 4, 2014
login:larjona-guest	<a href="http://www.debian.org">www.debian.org</a>	<a href="#">Contributed to www.debian.org</a>	Sept. 28, 2013	Dec. 18, 2013
login:larjona-guest	<a href="http://svn.debian.org">svn.debian.org</a>	<a href="#">committed to svn.debian.org (extra info)</a>	May 3, 2013	May 9, 2014
login:larjona-guest	<a href="#">publicity</a>	<a href="#">Editor of DPN and announcements.</a>	May 3, 2013	May 9, 2014
Personal activity log				
Time	Entry			
March 20, 2014, 5:31 p.m.	full name set to Laura Arjona Reina from Debian/Alioth account information			

Figure 3: List of all the contributions made by a certain person

For each contributor, an Identifier is defined in each data source: an object identifying the contributor in one of many possible ways (debian.org or alioth login names, emails or OpenPGP key fingerprints). The data source may add a human-readable version for the identifier, like a person's full name, if available.

The debian contributors software matches identifiers to users using UIDs from GPG keyrings, maps login names and emails to User UIDs, and maps unmapped fingerprints to debian accounts.

For the contributors not having Debian or Alioth account, the plan is to develop a dashboard to manually claim identifiers. When this dashboard is deployed, those contributions will be hidden by default and an opt-in system will be deployed to contact recent contributors, explain about the Debian Contributors site, and allow them to unhide their contributions if they want them to be shown in the list.

The contributor can login in the site using their Alioth or Debian credentials, and access a dashboard with some settings about the data displayed already implemented (Figure 4).

The site provides some statistics about the data which is not currently shown (see table 3), as well as some data source health information (last upload for each data source, date of latest contribution reported from each data source, admin contact information).

Table 3: Number of emails, fingerprints or login names for which there are contributions

Type	Count (with Alioth or Debian account)	Hidden (without Alioth or Debian account)
email	5.177	33.688
fpr	1.076	627
login	2.636	106

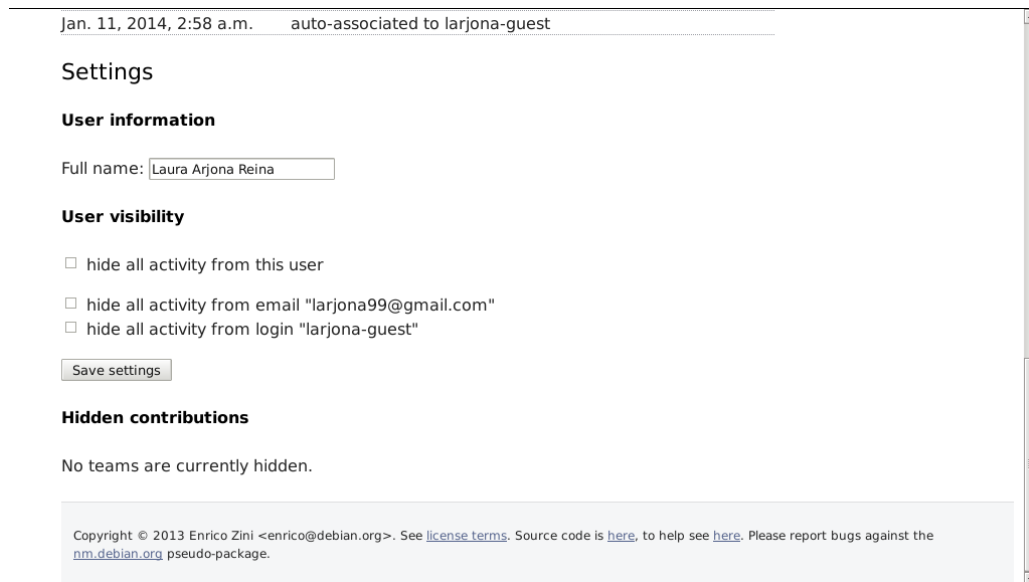


Figure 4: Contributor's dashboard

## Further work

The infrastructure has been developed flexible enough to allow the distribution of the work of maintaining the data sources: currently, 13 Debian Developers from different teams are maintaining 17 data sources, and it is expected that during this year, more teams are getting involved.

For teams, Debian Contributors can be used in coordination with the MIA (Missing In Action) team, to report about people leaving the project, and to better welcome new people and assist them when getting involved in the team tasks, as well as a tool to measure how burden is distributed inside the team.

The main goals for the future work on the project is to increase promotion inside Debian, so more data sources are tracked, and to implement the identity claims and opt-in infrastructure so people not having Debian or Alioth account can be credited while respecting their privacy preferences.

## Reproducibility, use in other projects

Any other free software project facing the challenge of making diversity skills more visible inside the community and for the public, is free to use the Debian Contributors' approach and source code. The code itself is licensed under AGPLv3+ license, and any kind of feedback (wether bugs reports, new ideas, or experiences about applying this system to any kind of community) is highly appreciated. Some links and contact details follow:

- Website: <https://contributors.debian.org>
- Source code: <http://anonscm.debian.org/gitweb/?p=nm/dc.git>
- Authors: Enrico Zini, Debian New Members Front Desk team.
- Bug reports: <http://bugs.debian.org/nm.debian.org>
- Mailing List: <https://lists.alioth.debian.org/mailman/listinfo/newmaint-site>
- Wiki page: <https://wiki.debian.org/DebianContributors/DcDocs>

## About this document

### Purpose

This document was written for the FLOSS Community Metrics Meeting in July 2014, in Portland (OR, USA), which aims to bring together people interested in FLOSS (free, libre, open source software) metrics to discuss the FLOSS tools that are currently available in this area.

More information about the meeting in its website: [flosscommunitymetrics.org](http://flosscommunitymetrics.org).

### License

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

### About the author

Laura Arjona Reina is MSc in Free Software by the Universidad Rey Juan Carlos (Madrid, Spain) and works as Information Technologies assistant at the Universidad Politécnica de Madrid.

She is a Debian contributor, participating in non-coding activities (translations, spam fighting, wiki editions, bug triaging) for several years. She has done some research about translations in free software projects and is interested in how software projects care (or not) about outreach and their non-coding contributors. She currently helps Enrico Zini, main developer of the “Debian contributors” tool, in order to promote the usage and extension of the system inside the Debian project.

Contact details and more information: <https://wiki.debian.org/LauraArjona>.