

How open development data benefits your project and your community

Jesus M. Gonzalez-Barahona

jgb@bitergia.com @jgbarah
Bitergia / LibreSoft (URJC)

Linux Tag 2014
Berlin (Germany), May 8th 2014





©2012-2014 Bitergia

Some rights reserved. This presentation is distributed under the
“Attribution-ShareAlike 3.0” license, by Creative Commons, available at
<http://creativecommons.org/licenses/by-sa/3.0/>



Structure of the presentation

- 1 Measuring free / open source software development
- 2 Why open development analytics?
- 3 Areas of interest
- 4 Tools





Bitergia

Measuring free / open source software development



A successful development model?

Free (open source) software
has shown to be a great success

...but there are many details to be understood
...and (a lot of) interest in understanding

...but there is room for improvement
...and (a lot of) interest in improving



A successful development model? (2)

There are *lots* of development models

Common characteristics for many of them:

- Community-based development
- Intensive use of tools, processes for coordination

Open development models
(as opposed to in-house, hidden models)



The importance of the community



[Crowd at FOSDEM 2008, by Jesús Corrius, CC Attribution 2.0]

<http://www.flickr.com/photos/jcorrius/2302302707/>

The importance of the community (2)

Persons (and organizations) with

- different interests
- common goals

Need for coordination, common decision making

Availability of data as a tool:

- Transparency to the community (fairness)
- Transparency to third parties (trust)



Diversity of tools, processes



git



Bazaar



launchpad



Diversity of tools, processes (2)

Despite diversity, a large fraction of projects:

- Use tools & services from a small set

git / svn / hg

Bugzilla / Jira / GitHub tickets

Gerrit

Mailman / Gmane

...

- use similar processes:

bug fixing coordination using tickets

pre-merge code review

general discussion in mailing lists

...

Collection and analysis of data is possible

Publication of data makes sense





Why open development analytics?

Information about code, community, development
for open development projects
can be retrieved, organized, analyzed

Let's publish analytics results & data

Open Development Analytics: A new standard for transparency



Who is interested?

Large & small free software communities

...and thousands of large & small companies,
public administrations, foundations

participating in them,
depending on their software

[Who can afford not to be interested?
It is a key strategic need for many actors]



Why?

Free software produced with open development models is more and more important for IT users, producers, integrators

It is different & complex,
yet transparent,
many details are public,
and it can be improved





Bitergia

Areas of interest



Some areas of interest

- Performance (understanding activity)
- Company participation (beyond copyright notices)
- Transparency (available information)
- Auditing (certify participation, experience, etc.)
- Profiling (key people, companies)
- Neutrality (fair treatment)



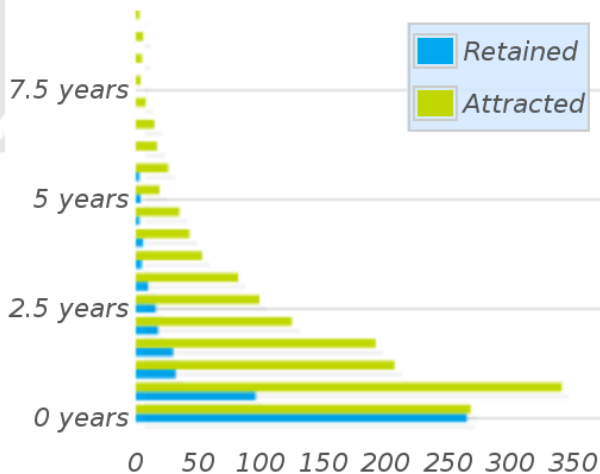


Bitergia

Issues	Parameters
Activity Reliability Sustainability	Raw volume, participants, ... Reaction times, pending issues, ... Growth rate, structure, ...



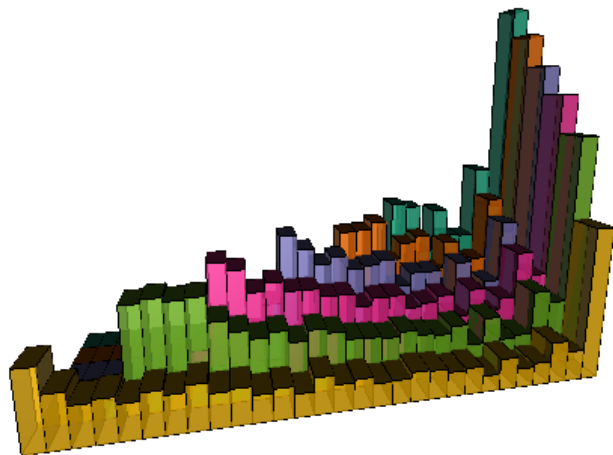
Areas of interest: community management





[Puppet committers community: Attraction / retention]



Areas of interest: community management



[Linux kernel: age of developers per cohort] 
[http://blog.bitergia.com/2013/02/01/](http://blog.bitergia.com/2013/02/01/demographics-of-linux-kernel-developers-how-old-are-they/)

demographics-of-linux-kernel-developers-how-old-are-they/ 

Areas of interest: engineering

99% of tickets open during the corresponding month were closed in less than x days



95% of tickets open during the corresponding month were closed in less than x days



Liferay: time-to-close tickets (quantiles) <http://blog.bitergia.com/2012/10/25/preview-of-the-analysis-of-liferay/>

Areas of interest: engineering

58286 Opened

week:260 (-6%) month:1223 (6%) year:15195 (58%)

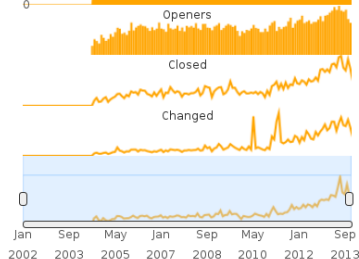
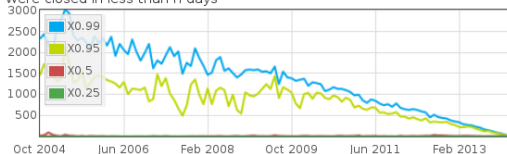
46046 Closed

week:199 (16%) month:1004 (1%) year:11481 (47%)



Time to Fix

X% of tickets open during the corresponding month were closed in less than n days



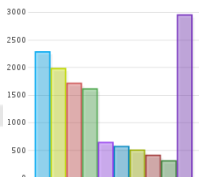
[MediaWiki community: tickets-related parameters]



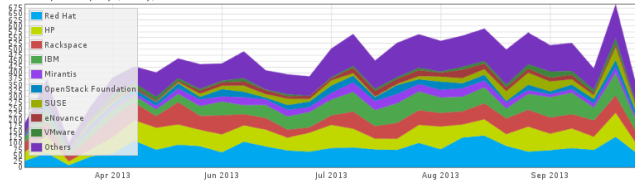
Areas of interest: company participation

Overall View of the Changes to the Source Code (Git)

Commits per company (aggregated)

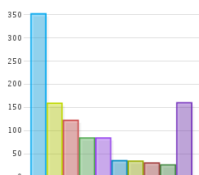


Commits per company (weekly)

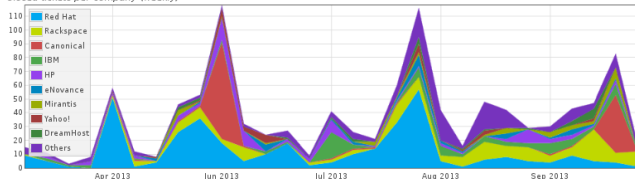


Overall View of the Closing Ticketing Activity (Launchpad)

Closed tickets per company (aggregated)



Closed tickets per company (weekly)



[Main companies in OpenStack Havana (partial view)]
<http://activity.openstack.org/dash/releases/>

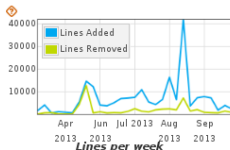
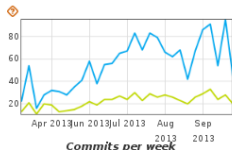


Areas of interest: company participation

GITS ACTIVITY

IBM

Commits: 1614
Start: 2013-04-04
End: 2013-10-16
Committers: 81
Authors: 85
Files: 3165
Branches: 6
Repositories: 37
Files actions: 6205
Lines Added: 205757
Lines Removed: 56723
Commits/week: 59.7778
Files/week: 117.2222
Commits/author:
18.9882
Files/author: 37.2353

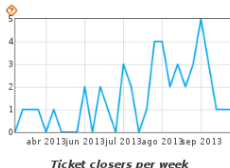


	commits
Dan Smith	182
Chris Yeoh	127
Matthew	102
Sean Dague	88
He Jie Xu	86
Brant Knudson	77
Matt Riedemann	71
Liang Chen	55
Davanum Srinivas	54
Avishay Traeger	48

LAUNCHPAD ACTIVITY

IBM

Opened: 734
Openers: 57
Start: 2013-04-04
End: 2013-10-16
Closed: 80
Closers: 22
Changed: 1659
Changers: 62
Trackers: 32



	closed
Igs	16
Fergal Mc Carthy	14
Viacheslav Biriukov	9
Youcef Laribi	8
Victor Galkin	4
Tanis De Luna	4
Ewan Mellor	4
rpedde	3
Hazoor Hussain	3
Abhijeet Jain	2

[IBM participation in OpenStack Havana (partial view)]

<http://activity.openstack.org/dash/releases/company.html?company=IBM>



- Development communities: companies and developers working together
- Policies, procedures, tools, source code... and development data
- Do they really provide enough data to enable assessment?
- Analysis of all repositories (data sources)...
- ...and associated information (eg: affiliation)



Areas of interest: auditing

Top Authors

	commits
Mark McLoughlin	374
Monty Taylor	344
Robert Collins	339
James E. Blair	330
Andreas Jaeger	260
Clint Byrum	215
Clark Boylan	211
Dirk Mueller	202
Diane Fleming	202
Steve Baker	189

Top Closers

	closed
Akihiro MOTOKI	84
Salvatore Orlando	76
Gary Kotton	73
Steve Baker	62
Aaron Rosen	61
john-griffith	56
Monty Taylor	51
Armando Migliaccio	51
Matt Riedemann	45
Eugene Nikanorov	43

Top Senders

	sent
Thierry Carrez	391
Russell Bryant	329
Monty Taylor	267
annegentle	266
Robert Collins	227
Sean Dague	223
Mark McLoughlin	220
Joshua Harlow	194
Dolph Mathews	189
jaypipes	179

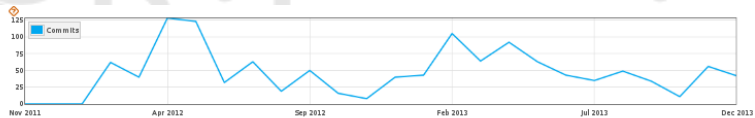
[OpenStack top contributors (December 2013)]



Areas of interest: profiling

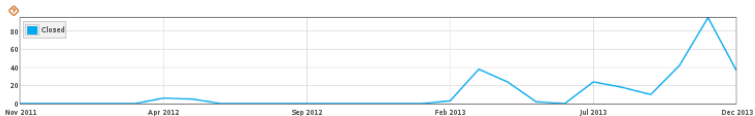
scm

Start 2012-02-07
End 2013-12-12
Commits 1218



its

Start 2012-04-01
End 2013-12-15
Closed 303



mls

Start 2012-01-25
End 2013-12-17
Sent 10950



scr

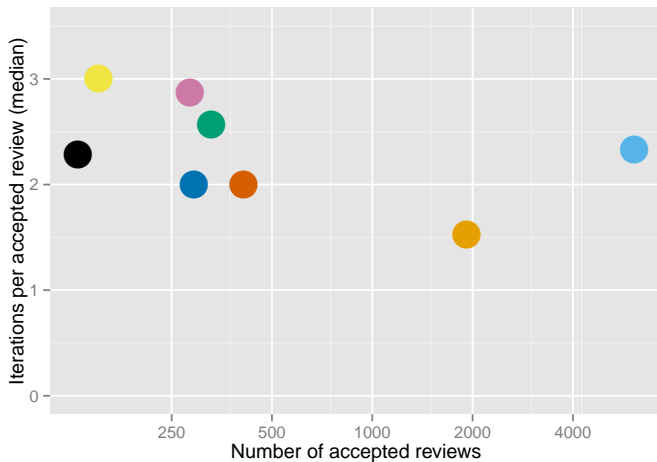
Start 2012-11-25
End 2013-12-17
Closed 3106



[oVirt developer profile]



Areas of interest: neutrality



[WebKit code review data per company (2012)]





Bitergia

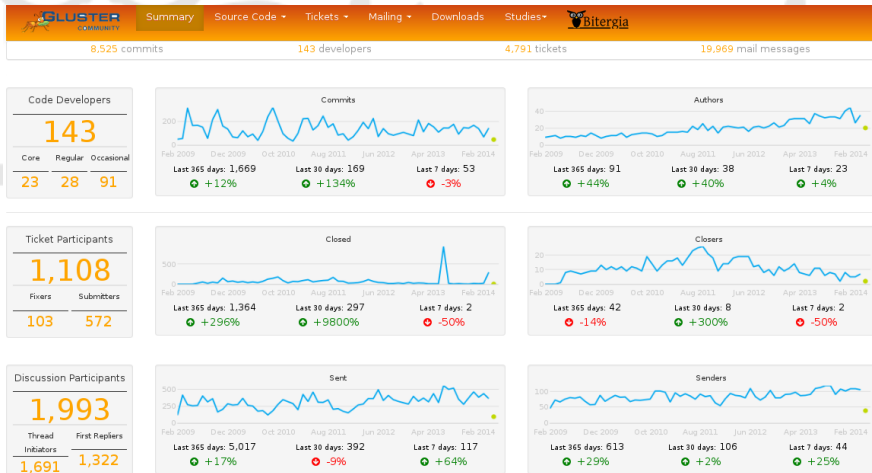
Tools



- MetricsGrimoire:
Free software for retrieving data from repositories
- vizGrimoire:
Free software for analyzing, visualizing data
- Grimoire Dashboard:
Many panels, different views of the project
(charts, summaries, statistic analysis)
- Commercially supported by Bitergia



Tools: Grimoire Dashboard



[Dashboard for the GlusterFS project
<http://projects.bitergia.com/redhat-glusterfs-dashboard/>





Bitergia

Let's go one step further:
Open Development Analytics



Markos:

- License analyzer
- New tools for software development analysis
- Production of linked open data

PROSE:

- Software development analytics to track results of R&D projects
- Open Source Projects Europe forge: development analytics facilities

<http://www.markosproject.eu/>

<http://www.ict-prose.eu/>

<https://opensourceprojects.eu/>



This is the end



Bitergia

[Questions, comments...]

