Data Science for Community Managers

Open Source Summit North America, Los Angeles (USA), Sep. 12th 2017

J. Manrique López de la Fuente
@jsmanrique
jsmanrique at bitergia dot com
https://speakerdeck.com/bitergia
Introduction

A bit about me
Hello, my name is Manrique and I am a community junkie

Involved in: HPCC, AsturLiNUX, HispaLiNUX, GPE, Maemo, Meego, Gnome, GDG, Mozilla, ...

Business, marketing & alliances developer in Bitergia, the software development analytics company
Software Development Analytics
for your peace of mind
Open Development

Everybody 💓 to be Open

Open Development

Community Management

Software Development Management
Come In
WE'RE
OPEN
Who is contributing* to open source software?

How much/what/where are they contributing*?

Who is driving successful projects?

Who is behind an open source project?

* Contributions can be made by code, issues, requests, etc.
Everybody loves community.
“Developers want to be part of something bigger; they want to **contribute** and **influence** where that technology is going”

*The Power of Community in Open Source, The Newstack*
Which projects are more community attractive?
Are projects retaining contributors?
Are projects community-driven or enterprise-driven?
How diverse is your community?
Who are projects top contributors?
Where does my community come from?
Where else my community contributes to?
How openness can be applied to any IT development?

**Inner Source**

*Managing Inner Source Projects* (bit.ly/managing-innersource)

**DevOps**

*DevOps framework*: CALMS (Culture/Collaboration, Automation, Lean, Measurement, and Sharing)

**Agile**

*Agile Manifesto*
How fast are we dealing with feedback (issues, requests, comments)?

How long does it take to add new features?

Are people engaged with the project/company?

Is the project/company allowing innovation?
Open Development Analytics
Let's get answers!

Why?
Collecting Data
Sharing Data
Be aware
Methodology
Why do you need metrics / analytics?

Transparency
Awareness
Governance support
Motivational
“Collecting data is only the first step toward wisdom, …”

<table>
<thead>
<tr>
<th>Source Code</th>
<th>Tickets / issues</th>
<th>Code Review</th>
<th>Continuous Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Git, Baazaar*, Mercurial* (*exporting to gh)</td>
<td>Bugzilla, GitHub Issues, JIRA, Manalm, Redmine</td>
<td>Gerrit, GitHub Pull Requests</td>
<td>Jenkins</td>
</tr>
<tr>
<td>Mailing lists / Forums</td>
<td>Chat</td>
<td>Downloads</td>
<td>Wiki</td>
</tr>
<tr>
<td>Askbot, Discourse, Gmane, Hyperkitty, mbox files, NNTP, Pipermail, Stack Exchange</td>
<td>IRC (Supybot), Slack, Telegram</td>
<td>Apache http logs</td>
<td>Confluence, MediaWiki</td>
</tr>
</tbody>
</table>

+30 different data sources
Easy to extend
"Collecting data is only the first step toward wisdom, but sharing data is the first step toward community."

Henry Louis Gates Jr., IBM/Linux Commercials

grimoirelab.github.io
“Collecting data is only the first step toward wisdom, but sharing data is the first step toward community”

Henry Louis Gates Jr., IBM/Linux Commercials

grimoirelab.github.io

chaoss.community
<table>
<thead>
<tr>
<th>Author</th>
<th>Commits</th>
<th>LoC added</th>
<th>LoC removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvaro del Castillo</td>
<td>1,582</td>
<td>698,280</td>
<td>58,978</td>
</tr>
<tr>
<td>Santiago Dueñas</td>
<td>772</td>
<td>130,883</td>
<td>23,612</td>
</tr>
<tr>
<td>Luis Cañas Díaz</td>
<td>110</td>
<td>16,179</td>
<td>1,091</td>
</tr>
<tr>
<td>Alberto Martín</td>
<td>65</td>
<td>21,921</td>
<td>1,136</td>
</tr>
<tr>
<td>Luis Cañas-Díaz</td>
<td>51</td>
<td>3,352</td>
<td>907</td>
</tr>
<tr>
<td>J. Manrique Lopez de la Fuente</td>
<td>42</td>
<td>23,214</td>
<td>1,070</td>
</tr>
<tr>
<td>Andy Grunwald</td>
<td>36</td>
<td>499</td>
<td>281</td>
</tr>
<tr>
<td>Jesús M. González-Barahona</td>
<td>34</td>
<td>1,505</td>
<td>605</td>
</tr>
</tbody>
</table>

**1st commit**

- September 30th 2015, 18:15:16.000
- June 30th 2015, 19:52:58.000
- March 10th 2016, 15:59:35.000
- February 9th 2016, 15:56:45.000
- March 10th 2016, 16:59:31.000
- January 20th 2016, 19:04:30.000
- October 11th 2015, 19:08:22.000
- December 31st 2015, 19:16:25.000

**Last commit**

- June 13th 2017, 10:33:15.000
- June 1st 2017, 18:25:17.000
- February 28th 2017, 08:49:14.000
- June 12th 2017, 12:55:50.000
- June 12th 2017, 13:56:42.000
- May 19th 2017, 08:33:27.000
- January 5th 2016, 20:44:26.000
- January 30th 2017, 01:44:03.000
Community dependency KPIs

Onion model
- 7 core
- 40 regular
- 85 casual

ASF Pony factor
- Pony factor: 1

Bitergia Elephant factor
- Elephant factor: 2

Bitergia Zapata factor
- Linux Kernel Zapata factor ~ 200

Bitergia United Fruit Company factor
- Linux Kernel UFCo factor ~ 10

Linux kernel ownership analysis: [linux.biterg.io](http://linux.biterg.io)
More Community Manager Nightmare
Metrics can be very dangerous
“Human beings adjust behavior based on the metrics they’re held against. **Anything you measure will impel a person to optimize his score on that metric.** What you measure is what you’ll get. Period”.

*You Are What You Measure* by Dan Ariely
Governance -> Goals <- Questions <- Metrics

1. Develop a set of corporate, division and project business goals and associated measurement goals for productivity and quality
2. Generate questions (based on models) that define those goals as completely as possible in a quantifiable way
3. Specify the measures needed to be collected to answer those questions and track process and product conformance to the goals
4. Develop mechanisms for data collection
5. Collect, validate and analyze the data in real time to provide feedback to projects for corrective action
6. Analyze the data in a post mortem fashion to assess conformance to the goals and to make recommendations for future improvements
/be_nice!

It's about the people!
Use cases

How people is using open development analytics

- CNCF & Kubernetes
- Puppet Inc.
- OpenStack Foundation
- Samsung Open Source Group
- Liferay Inc.
Increase enterprises involvement in Open Source software development
Reduce single company dependency
Increase participation beyond coding
Increase participation beyond coding
Why am I Using GrimoireLab?

- Samsung relies on open source software for numerous products.
- Investing time and money into software always requires a strategic evaluation.
- It's not enough to evaluate a project solely on its technical merits, the community also has an impact on the strategic value of OSS.
- Public information about OSS communities allows for a much deeper evaluation than proprietary software.

Samsung OSG video from GrimoireCON NA 2017

“Grimoire Lab gives us a well rounded picture of our overall community health to help determine the areas where we are doing good and the areas where we can help improve.”

Jamie Sammons - Developer Advocate at Liferay Inc.
Join the Community

grimoirelab.github.io

How to start playing
Just ask
Cauldron
How to start playing...
$ pip install perceval
...
$ pip install grimoire-elk
...
where_to_ask

GrimoireLab Training Tutorial

GrimoireLab Mailing List
lists.linuxfoundation.org/mailman/listinfo/grimoirelab-discussions

GrimoireLab IRC Channel
#GrimoireLab in Freenode.net
The Cauldron

GitHub organizations analysis

Up to 5 GitHub organizations per user

Latest 30 active repos per organization analysis

FREE

cauldron.io
Your project & community data are already there...
Take care of your project!

grimoirelab.github.io
chaoss.community

jsmanrique@bitergia.com