# Data Science for Community Managers

J. Manrique López de la Fuente <u>@ismanrique</u> jsmanrique at bitergia dot com https://speakerdeck.com/bitergia





# **Outline**

Introduction

**Open Development Communitites** 

**Open Development Analytics** 

Playing with GrimoireLab

**Extras** 



# Introduction

A bit about me

Why am I here?

Disclaimer

# /me



Manrique Lopez

@jsmanrique

Mobile & Open Source technologies, Friends, Surf and Live passionate. Bilbaíno born in Gijón living in Fuenla and working in @bitergia Hello, my name is Manrique and I am a community junkie

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

Business & marketing *developer* in Bitergia, the software development analytics company



# /why



## **Manrique Lopez**

@jsmanrique

Mobile & Open Source technologies, Friends, Surf and Live passionate. Bilbaíno born in Gijón living in Fuenla and working in @bitergia







# /disclaimer



Manrique Lopez

@jsmanrique

Mobile & Open Source technologies, Friends, Surf and Live passionate. Bilbaíno born in Gijón living in Fuenla and working in @bitergia I am not a computer scientist, how many of you are?

I am not a data scientist, how many of you are?

Presentation focus in *open source* & *inner source* related communities



# Open Development Communities

What is a community?

Community Management

"A community is commonly considered a social unit (a group of people) who **share something in common**, such as norms, values, identity, and often a sense of place that is situated in a given geographical area (e.g. a village, town, or neighborhood). Durable relations that extend beyond immediate genealogical ties also define a sense of community. People tend to define those social ties as important to their identity, practice, and roles in social institutions like family, home, work, government, society, or humanity, at large. Although communities are usually small relative to personal social ties (micro-level), "community" may also refer to large group affiliations (or macro-level), such as national communities, international communities, and virtual communities.

The word "community" derives from the Old French comuneté which comes from the Latin communitas (from Latin communis, things held in common)"



By <a href="https://en.wikipedia.org/wiki/Community">https://en.wikipedia.org/wiki/Community</a>

# It's about people





# It's about people who share























It's about people who share something in common







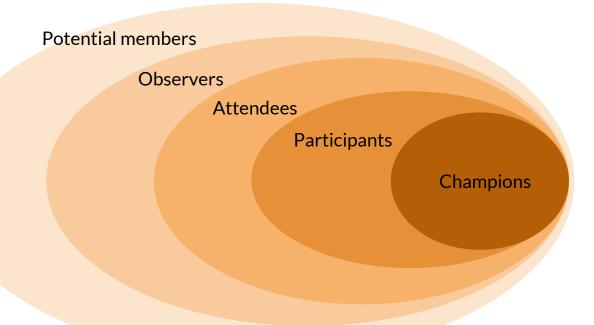


Self-awareness - Governance - Transparency



# /communities se

# **Self-awareness**



# Anatomy of a community

Alex Hillman



# Governance

"Establishment of policies, and continuous monitoring of their proper implementation, by the members of the governing body of an organization. It includes the mechanisms required to **balance the powers** of the members (with the associated **accountability**), and their primary duty of enhancing the prosperity and viability of the organization".



businessdictionary.com

Transparency to the community

Transparency to third parties

Fairness

Trust



# /management

# The Community Manager

Do communities need to be managed?

Do communities need leadership?

Why is becoming so important?

How many of you are community managers, developers relationship (AKA *DevRels*), etc.?



/management

**Community Manager Responsibilities** 

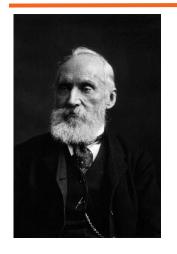
Community health

Community **productivity** 

Community visibility



# /management



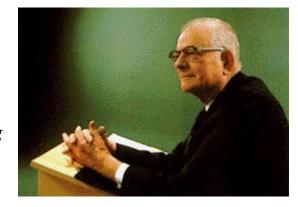
"To measure is to know"

"If you can not measure it, you can not improve it"

Lord Kelvin

"Without data, you are just another person with an opinion"

W. Edwards Deming





# /take\_care

"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



# Open Development Analytics

Transparency & objectivity matters

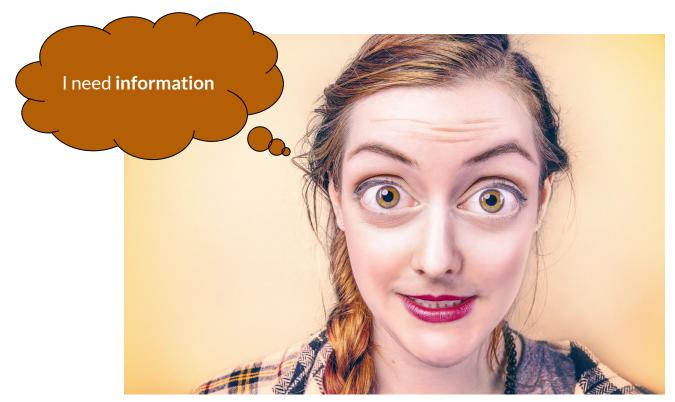
**Data Sources** 

Tools

GrimoireLab

# /data\_sources

The community manager nightmare!





# /data\_sources

# The community manager nightmare:

- Time is limited
- How to take decisions without reliable data?





# /data\_sources

Where is everything happening?























# /data\_sources Data silos





# /tools

# Several interesting approaches

OpenHub

StackOverflow metrics

Stackalytics

GitTorrent, GitHub Archive

GitHub Archive + Google BigQuery

github3.py

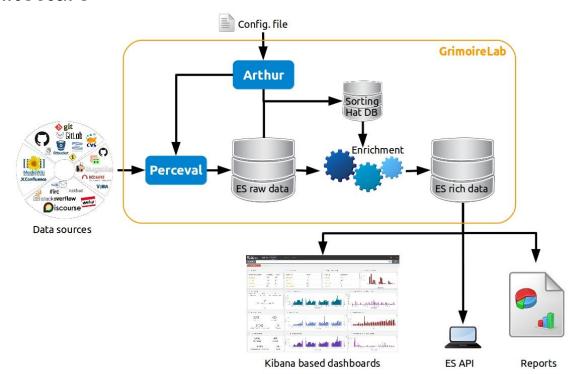






grimoirelab.github.io

## **Architecture**













elasticsearch

### **PERCEVAL**

Sir Perceval goes on the quest to retrieve and gather data from git, GitHub, Bugzilla, JIRA, Gerrit, mbox, pipermail, StackExchange, Discourse, etc.

Check Perceval repository



### **ARTHUR**

King Arthur commands his loyal knight Sir Perceval mananging the tasks to retrieve data for analysis. It manages data incremental update, parallel downloading, etc.

Check Arthur repository



### **SORTING HAT**

Tool to manage people identities information, to merge multiple person identities across different data sources, enrich profile information with affiliation information, etc.

Check Sorting Hat repository



### KIBITER

Custom fork of Kibana to work on new ideas for metrics & data visualization to be used by GrimoireLab Panels.

♠ Check KiBiter repository



### GRIMOIREELK

Playground for testing the whole set of tools as a platform for software development analytics. It's is a prototype of the *Grimoire Open Development* Analytics platform.

○ Check GrimoireELK repository



### **PANELS**

Set of pre-defined widgets and dashboard templates to visualize Elasticsearch indexes generated by GrimoireELK with KiBiter.

Check Panels repository



grimoirelab.github.io































grimoirelab.github.io

# Some features

Drill down

Time frame selection

Sharing / embedding

Data export (CSV...)

Query API (ElasticSearch)

Users can create custom

widgets and panels

Easy validation

Links to real artifacts (commits, tickets, etc.)

Search box



# Playing with GrimoireLab

Showtime!

## Let's start..

```
$ git clone http://.../perceval.git
$ sudo python3 setup.py install
$ git clone http://.../grimoireelk.git
$ elasticsearch/bin/elastic &
$ kibana/bin/kibana &
GrimoireELK/utils$ python3 ./p2o.py --enrich --index git yarn -e http://localhost:9200
--no inc --debug git https://github.com/yarnpkg/yarn.git
2016-10-14 14:09:32,392 Total items enriched 897
2016-10-14 14:09:32,392 Done git
2016-10-14 14:09:32,392 Enrich backend completed
2016-10-14 14:09:32,393 Finished in 0.75 min
GrimoireELK/utils$ python3 ./p2o.py --enrich --index github yarn -e
http://localhost:9200 --no inc --debug github -t *** --owner yarnpkg --repository yarn
2016-10-14 14:20:19,736 Total items enriched 900
2016-10-14 14:20:19,736 Done github
2016-10-14 14:20:19,737 Enrich backend completed
2016-10-14 14:20:19,738 Finished in 7.20 min
```



# https://jgbarah.gitbooks.io/grimoirelab-training/

# Type to search GrimoireLab Training Introduction Perceval First steps Git GitHub Mail archives Summarizing Producing Kibana dashboards with ... Installation A simple dashboard Managing dashboards with kidash Python scripting Persistent storage in ElasticSearch Querying ElasticSearch ElasticSearch DSL Querying GrimoireELK indexes

Tools and tips



GrimoireLab is free, open source software for software development analytics. It allows you to retrieve data from many kinds of systems with information related to software development, and produce analysis and visualizations with it.

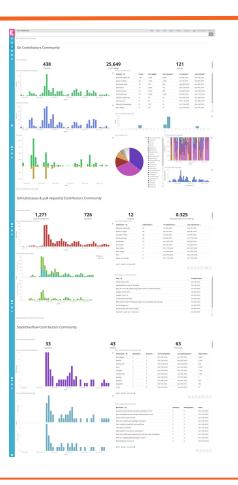


GrimoireLab supports more than 20 different kinds of data sources, from git reposiories or GitHub projects, to issue trackers such as Jira or Bugzilla, including messaging systems such as IRC, Slack or mailing lists, or other types of systems such as StackOverflow or Jenkins.

With the tools in the GrimoireLab toolset you can retrieve data, store it in databases, analyze it, and produce nice dashboards for interactively visualizing the resulting information.







Git data

GitHub data

StackOverflow data



# **Extras**

Some analysis and metrics provided with Grimoire Lab

Network analysis

Dependency

Geographical analysis

Demography

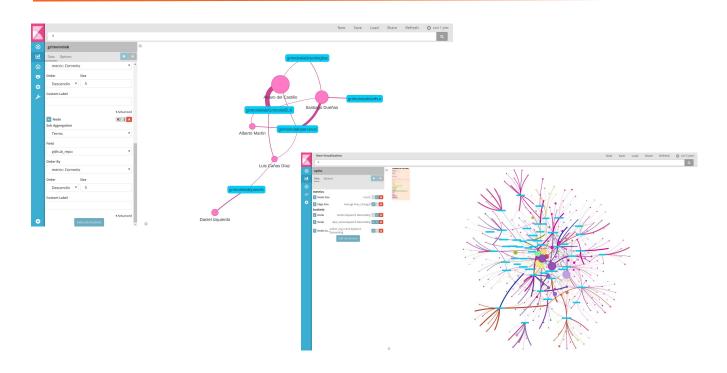
Gender diversity

Contributors review

Dealing with issues

Contributors funnel

# /network

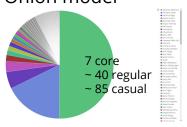




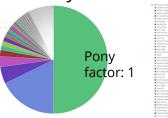
Open Source Kibana network visualization plug-in: https://github.com/dlumbrer/kbn\_network

## /dependency

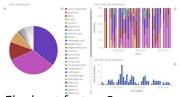
#### Onion model



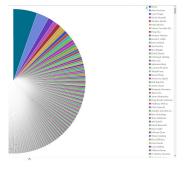
#### ASF Pony factor



#### 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: <a href="mailto:linux.biterg.io">linux.biterg.io</a>



## /geoanalysis





Mozilla Commits by timezone

Node GitHub Pull Requests

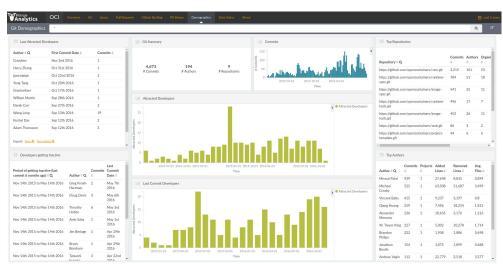


Mozilla Reps Activities



## /demography

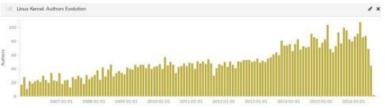






Open Containers Initiative Demography panel

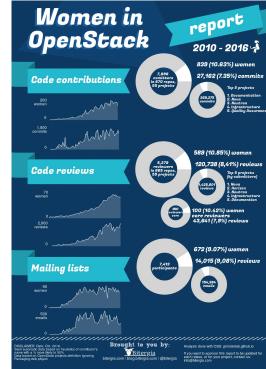
## /gender



Gender-diversity Analysis of the Linux Kernel Technical Contributions

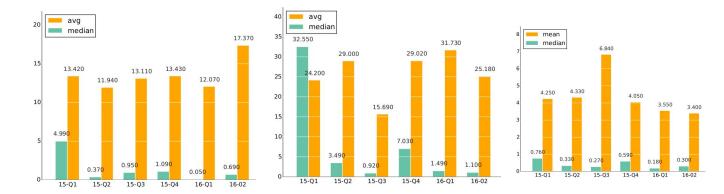
Women in OpenStack report for WOO Meeting

Gender-diversity analysis of technical contributions in the Hadoop Ecosystem





#### /reviews



Open NFV code review stats

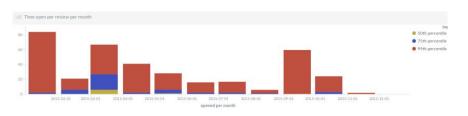
#### **Development cycle analysis**

Idea, request  $\rightarrow$ Iterations  $\rightarrow$ Testing  $\rightarrow$ Deployment



#### /reviews

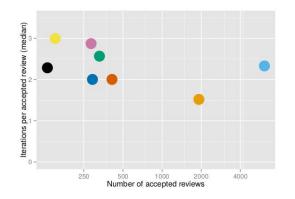
#### Some reviewers are more equal than others



Source: Some developers are more equal than others

(Bitergia's blog, 2015)

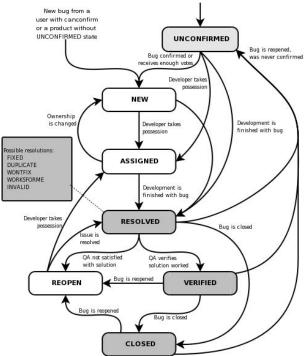
#### **Neutrality?**

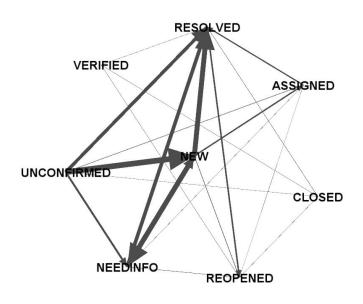


Source: <u>Understanding How Companies</u> <u>Interact with Free Software Communities</u> (IEEE, 2013)



## /issues







### /funnel





## /grimoirelab

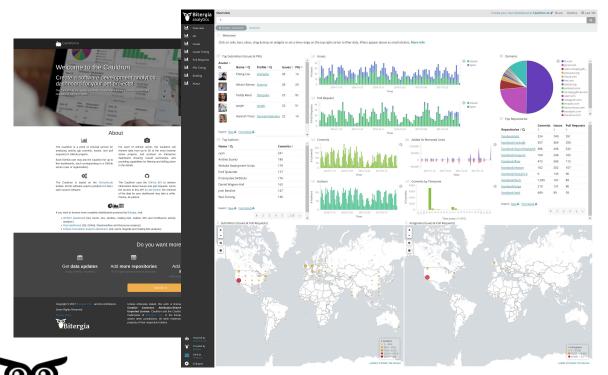
#### Data driven community management





## That's not all...

#### /bonus



itergia

#### **The Cauldron**

GitHub organizations analysis

Up to 5 GitHub organizations per user

Latest 30 active repos per organization analysis

Snapshot (*no data update*)

Kibana 5.x based dashboards

#### **FREE**

http://cauldron.io

### /bonus

Are you a community manager consultant?

Are you working as developer advocate for other companies?

# We are looking for partners!!

Ask us about our partnership program: info@bitergia.com



# Data Science for Community Managers

J. Manrique López de la Fuente <u>@ismanrique</u> jsmanrique at bitergia dot com https://speakerdeck.com/bitergia



