

---

# Measuring Software development with GrimoireLab

OpenSouthCode, 24th-25th March

Miguel Ángel Fernández, Valerio Cosentino  
{mafesan, valcos}@bitergia.com  
@Mghfdez, @\_valcos\_

[bitergia.com](http://bitergia.com)

---

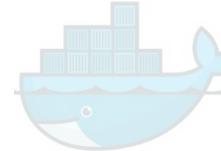


/context



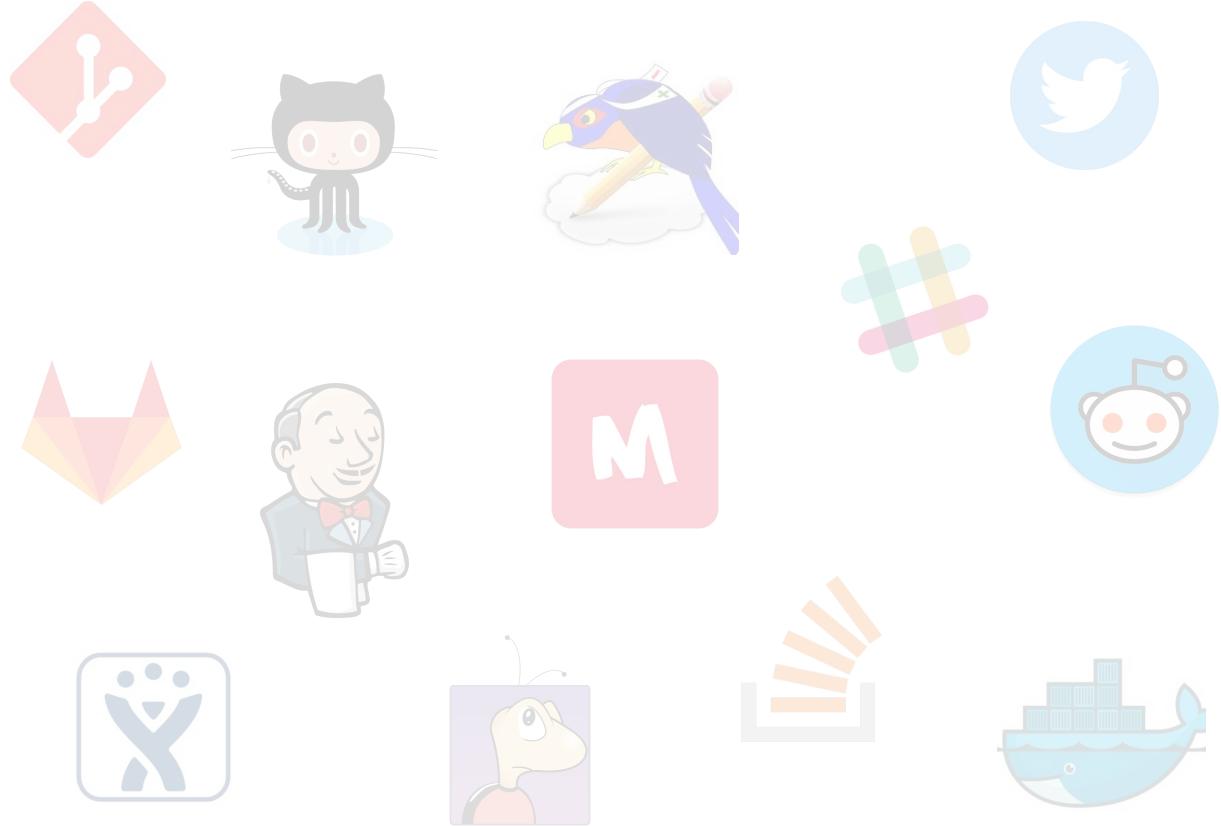


**How can I measure my software project?**



/context

---

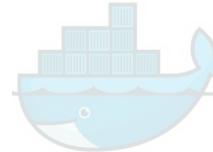


/context



## Activity

**How many contributors do we have ?**



/context

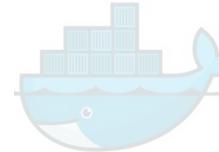


## Activity

**How many contributors do we have ?**

## Community

**How many companies are contributing  
to my project?**



/context



## Activity

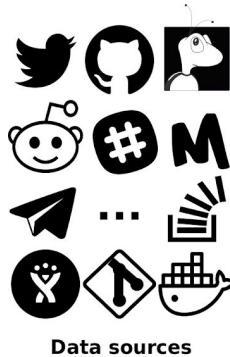
**How many contributors do we have ?**

## Community

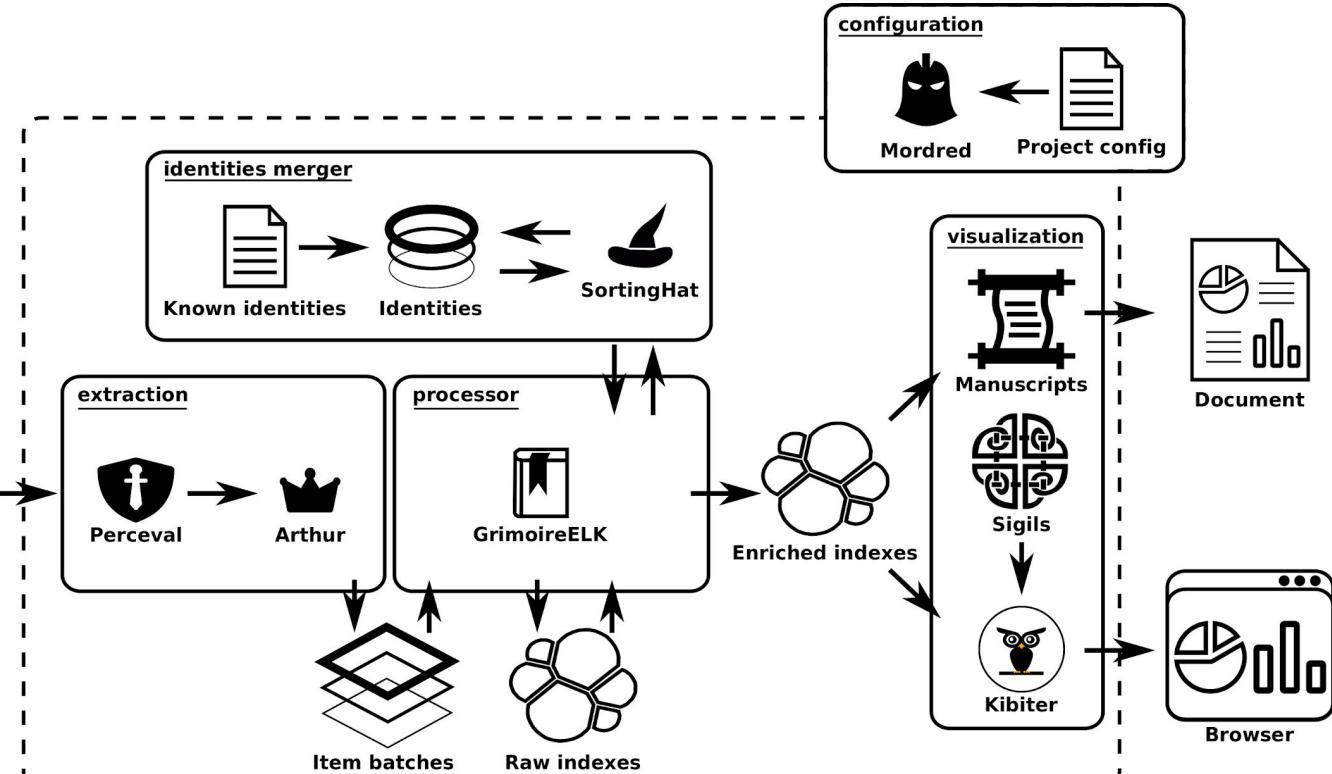
**How many companies are contributing  
to my project?**

## Processes

**How good we are in handling issues?**



Data sources



## Counting Potatoes: the Size of Debian 2.2

*Jesús M. González-Barahona, Miguel A. Ortuño Pérez, Pedro de las Heras Quirós, José Centeno González and Vicente Matellán Olivera*

*Debian is the largest Free Software distribution, with well over 2,800 source packages in the latest stable release (Debian 2.2) and more than 4,000 source packages in the release currently in preparation. But, how large is “the largest”? In this paper, we use David Wheeler’s sloccount system to determine the number of physical source lines of code (SLOC) of Debian 2.2 (aka Potato). We show that Debian 2.2 includes over 56,000,000 physical SLOC (almost twice than Red Hat 7.1, released about 8 months later), showing that the Debian development model (based on the work of a large group of voluntary developers spread around the world) is at least as capable as other development methods (like the more centralized one, based on the work of employees, used by Red Hat or Microsoft) to manage distributions of this size.*

## MetricsGrimoire

Tools for software & development community analytics

 @ GitHub

Update: Check the new GrimoireLab platform

## MetricsGrimoire

[Update]: Currently, main development related with *MetricsGrimoire* has stopped. Main contributors are working on a new platform: [GrimoireLab](#)

**MetricsGrimoire** (pronounced /metrɪksgrɪmwar/) is a toolset to obtain data from repositories related to software development: source code management (aka *version control*), issue tracking (aka *bug reporting*) systems, mailing lists, etc. Data and metadata about the software development processes is retrieved from those repositories (information about commits, ticket management, communication in mailing lists, etc.), and then organized and stored into SQL databases that can later be mined for specific patterns or summaries of activity.

**MetricsGrimoire** tools support many kinds of repositories, including those provided by GitHub (git & GitHub issue tracking).



**MetricsGrimoire** has been already used to analyze many different projects and together with visualization tools like [VizGrimoire](#) it is possible to get dashboards or reports like [the ones provided by Bitergia](#).

# /history/2011-2014

**MetricsGrimoire**

Tools for software & development community analytics

[@ GitHub](#)

Update: Check the new GrimoireLab platform

Hosted on [GitHub Pages](#) using the Dinky theme

**vizGrimoire**

Analyzing and visualizing software development

[GitHub Profile](#)

Hosted on [GitHub Pages](#) using the Dinky theme

## The vizGrimoire project

vizGrimoire (usually pronounced /vɪzgrɪmwar/) is a toolset and framework to analyze and visualize data about software development. Currently, it is focused on data produced by the [MetricsGrimoire tools](#) (CVSAnalY, Bicho and MailingListStats).

vizGrimoire is promoted by [Bitergia](#), the company providing software development analytics services, but it is a project open to the community: you are welcome to participate.

The following is a list of the current set of tools with a brief description. For more information, visit each of tool repositories, and/or read their install and readme files. You can also have a glimpse of the intended [vizGrimoire architecture](#).

### vizGrimoireR

Analyze data in databases produced by MetricsGrimoire tools from software development repositories (source code management systems, issue tracking systems, mailing lists, etc.).

[Repository at GitHub](#)

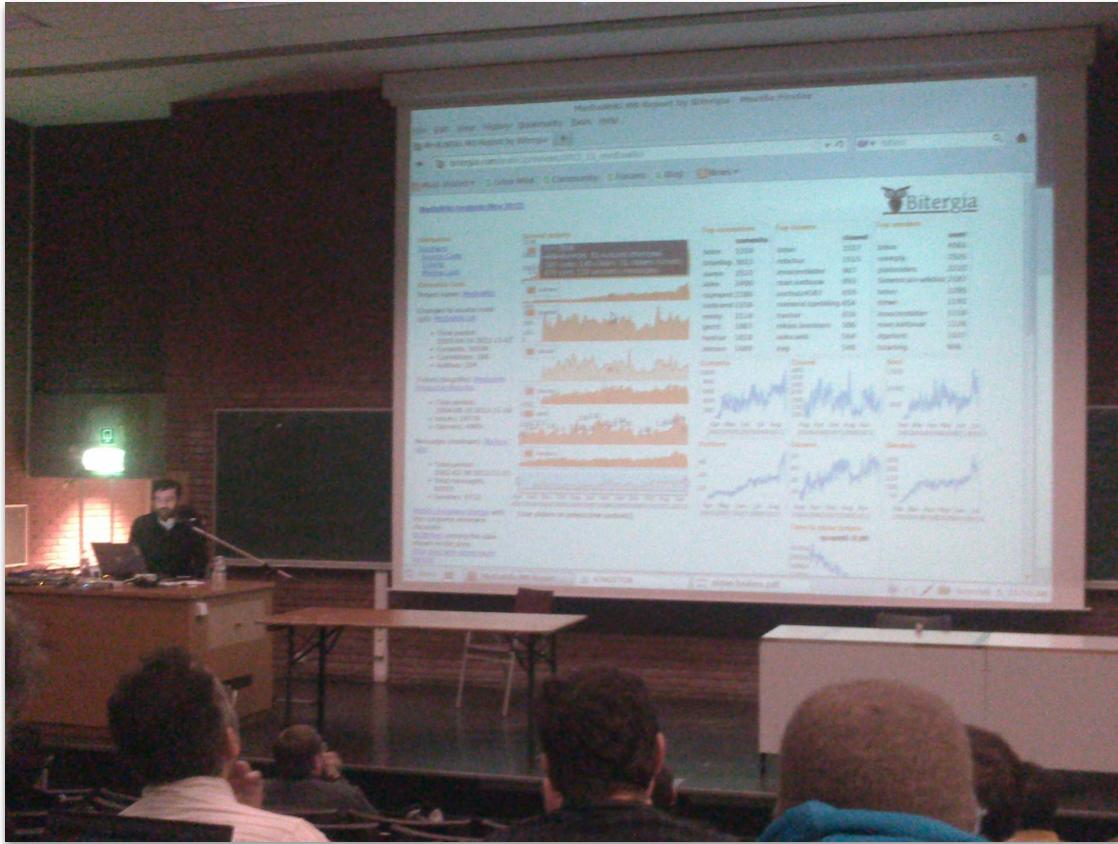
### vizGrimoireJS

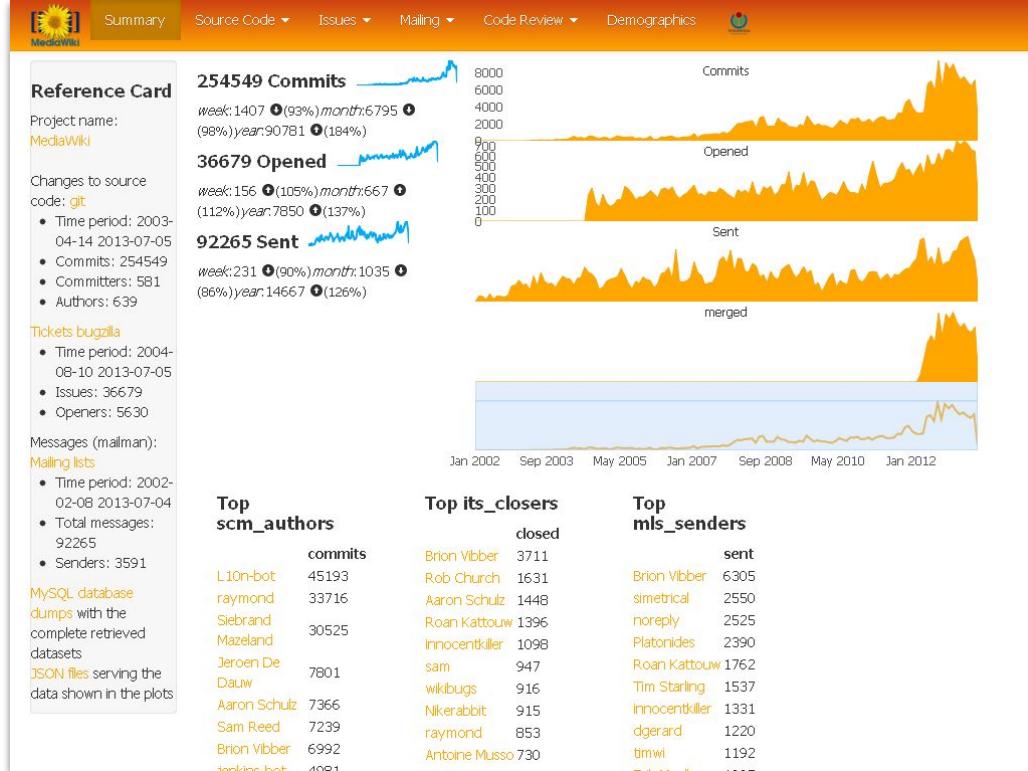
JavaScript framework to create dashboards and interactive reports and visualizations from data produced by vizGrimoireJS

[Repository at GitHub](#)

### Source code, reporting bugs, communication







# /history/recent

Dashboard / Overview

Full screen Share Edit Last 90 days Uses lucene query syntax Actions

NOT Empty Commits NOT Bots Add a filter +

Git Top Authors

Author	Commits	Projects	Added Lines	Removed Lines	Avg. Files
Alvaro del Castillo	185	1	2831	1511	1.784
Valerio Cosentino	175	1	37017	6429	2.08
Jesus M. Gonzalez-Barahona	89	1	16477	3196	6.618
Santiago Dueñas	47	1	1915	2516	2.979
Manrique Lopez	11	1	2	275	1

Organizations

Git Top Projects

Project	Commits	Authors	Repositories
grimoire	535	15	13

Export: Raw Formatted

Git

535 # Commits 15 # Authors 13 # Repositories

Commits

Authors

GitHub Pull Requests

290 # Pull Requests 17 # Submitters 13 # Repositories

Pull Requests

Pull Request Submitters

Github Issues

GitHub Issues

GitHub Issues Submitters



## /history/very-recent

The screenshot shows the CrossMiner homepage. At the top, there's a logo for "CROSSMINER" with the subtitle "Developer-Centric Knowledge Mining from Large Open-Source Software Repositories". Below the logo, a large heading reads "Developer-Centric Knowledge Mining from Large Open-Source Software Repositories". A paragraph below the heading explains that CrossMiner enables monitoring, in-depth analysis, and evidence-based selection of open source components. A "READ MORE" button is at the bottom.

The screenshot consists of two parts. On the left is a slide titled "Extracting Knowledge From CHAOS" which lists metrics like Project lifecycle, Diversity and inclusion, Risk and provenance, Ecosystem impact, and more. It also mentions developing a FLOSS reference implementation and integrating GrimoireLab, Prospector, and Credit. On the right is a photograph of a man speaking on stage at the "OPEN SOURCE SUMMIT NORTH AMERICA" under the "THE LINUX FOUNDATION" banner.



/history/today

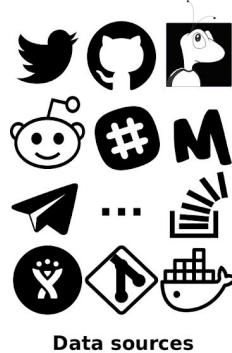
---



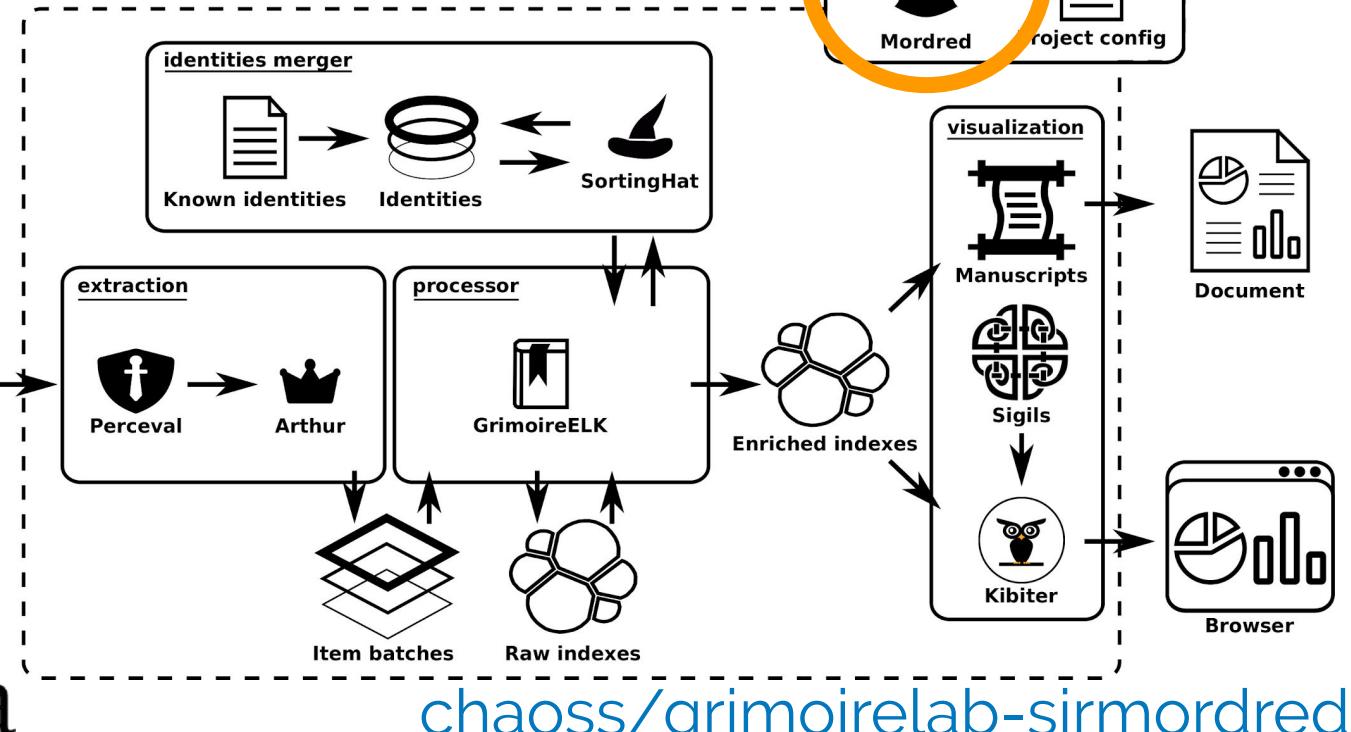
[chaoss.github.io/grimoirelab](https://chaoss.github.io/grimoirelab)



/mordred



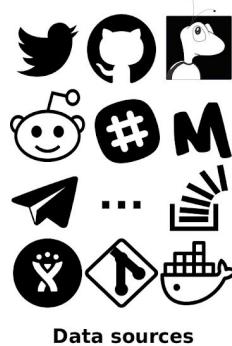
Data sources



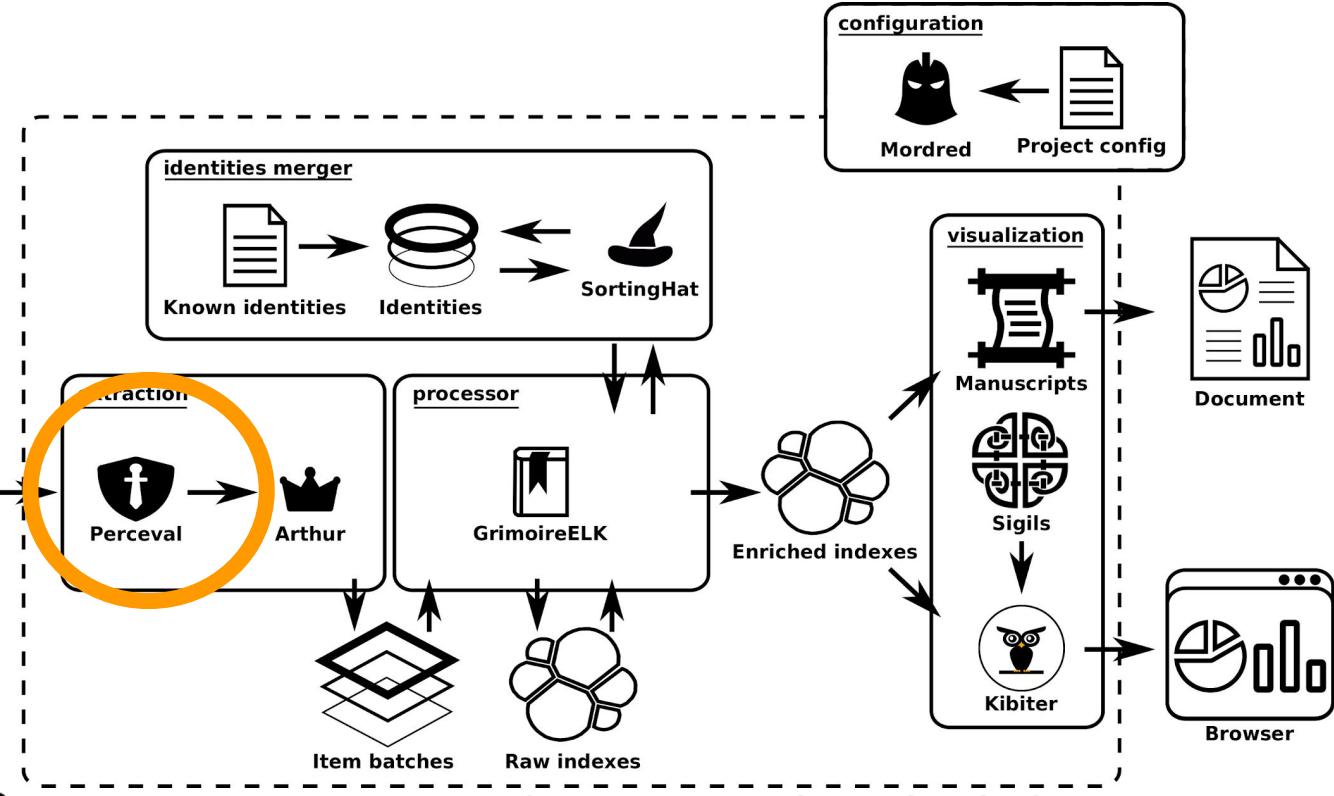
 Bitergia

[chaoss/grimoirelab-sirmordred](https://chaoss/grimoirelab-sirmordred)

/perceval

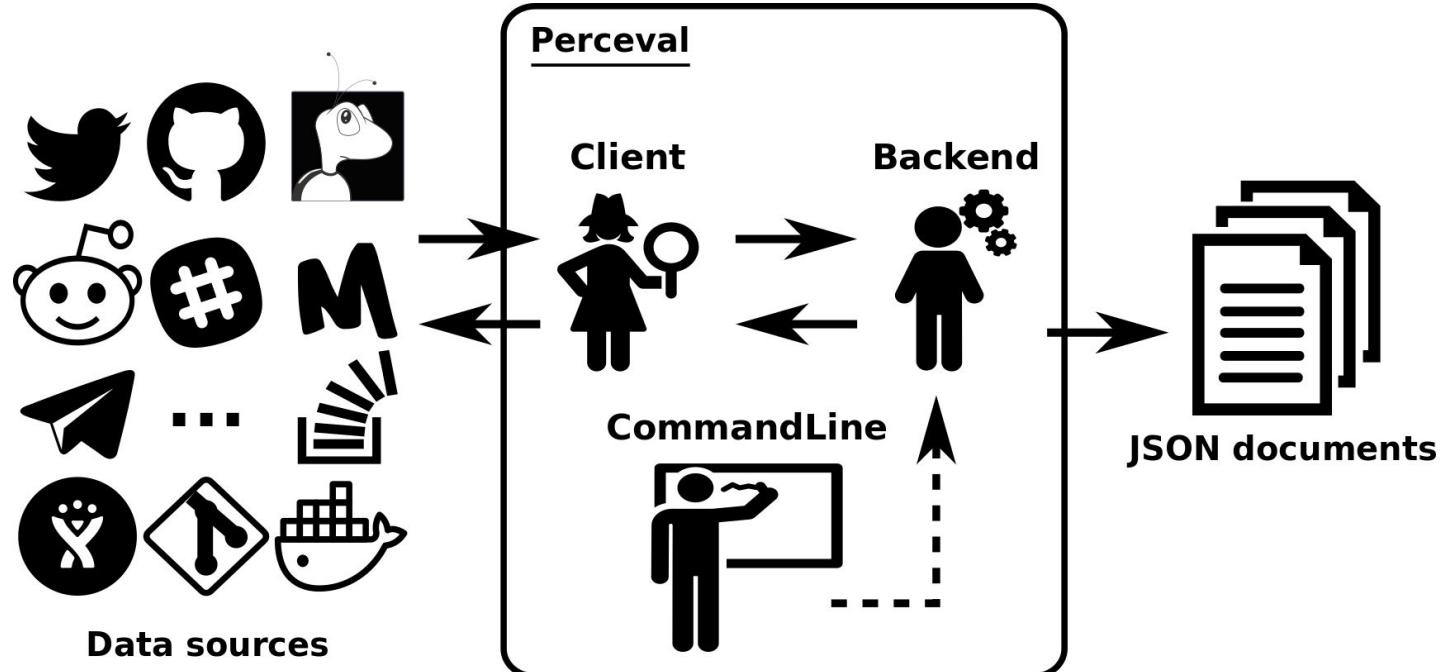


Data sources

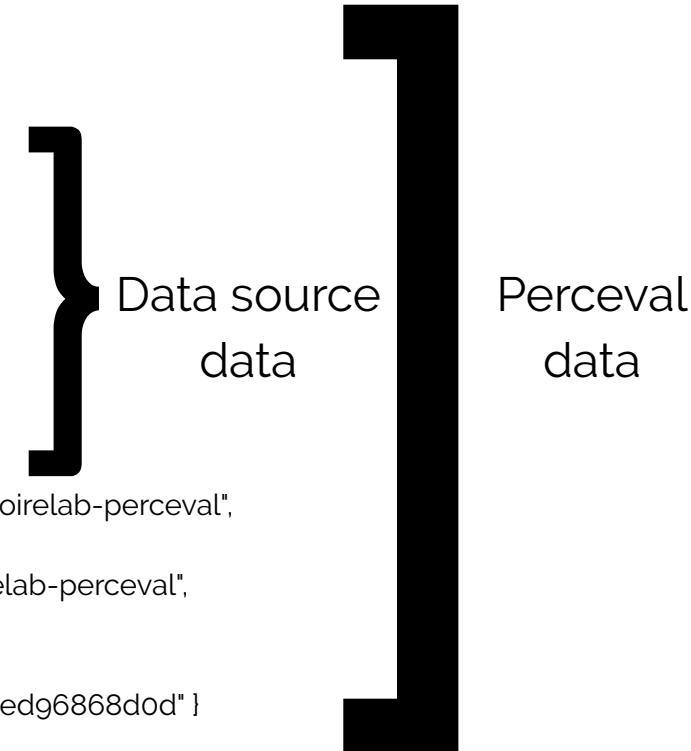


[chaoss/grimoirelab-perceval](https://chaoss/grimoirelab-perceval)

Bitergia



```
[ "backend_name": "Git",
  "backend_version": "0.10.2",
  "category": "commit",
  "data": [
    "Author": "...",
    "AuthorDate": "Mon Sep 3 11:37:40 ...",
    "Commit": "...",
    "CommitDate": "Mon Sep 3 11:37:40 ...",
    "commit": "d3a8e97444dd9a63b0c ...",
    "files": [...], "message": "Merge ...",
    "parents": [...], "refs": [...]
  ],
  "origin": "https://github.com/chaoss/grimoirelab-perceval",
  "perceval_version": "0.11.11",
  "tag": "https://github.com/chaoss/grimoirelab-perceval",
  "timestamp": 1537685906.51857,
  "updated_on": 1535967460.0,
  "uuid": "9b7b4722fef72a8ce69781641a8220ed96868dod" }
```





## As a program

```
$ pip3 install perceval
$ perceval git https://github.com/chaoss/grimoirelab-perceval
--from-date=2018-09-01
[2018-09-23 08:58:25,085] - Sir Perceval is on his quest.
[2018-09-23 08:58:25,088] - Fetching commits ...
... JSON Documents
[2018-09-23 08:58:26,540] - Sir Perceval completed his quest.
```

## As a Python3 library

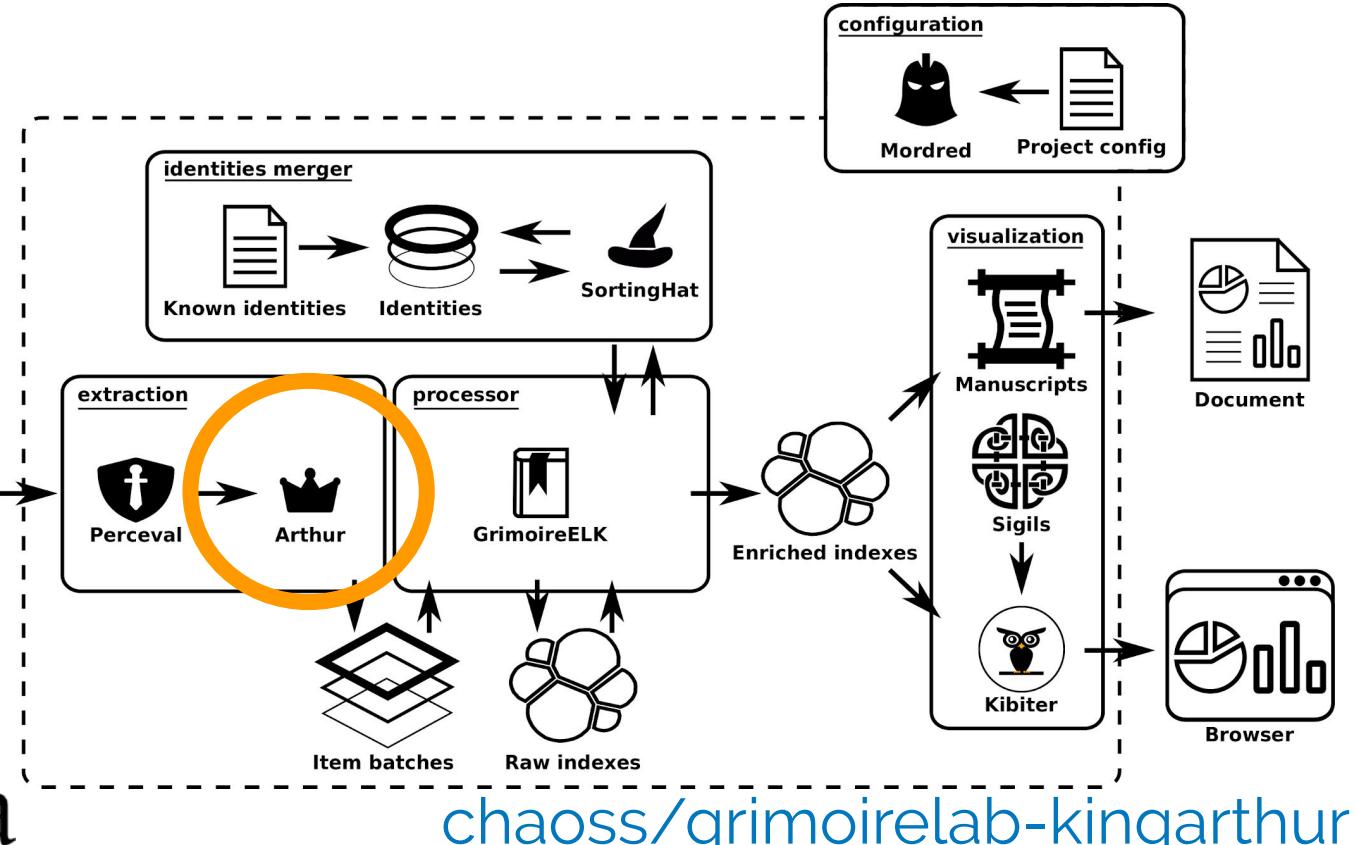
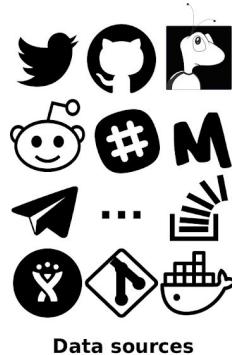
```
from perceval.backends.core.git import Git

from_date = datetime.datetime(2018, 9, 1)

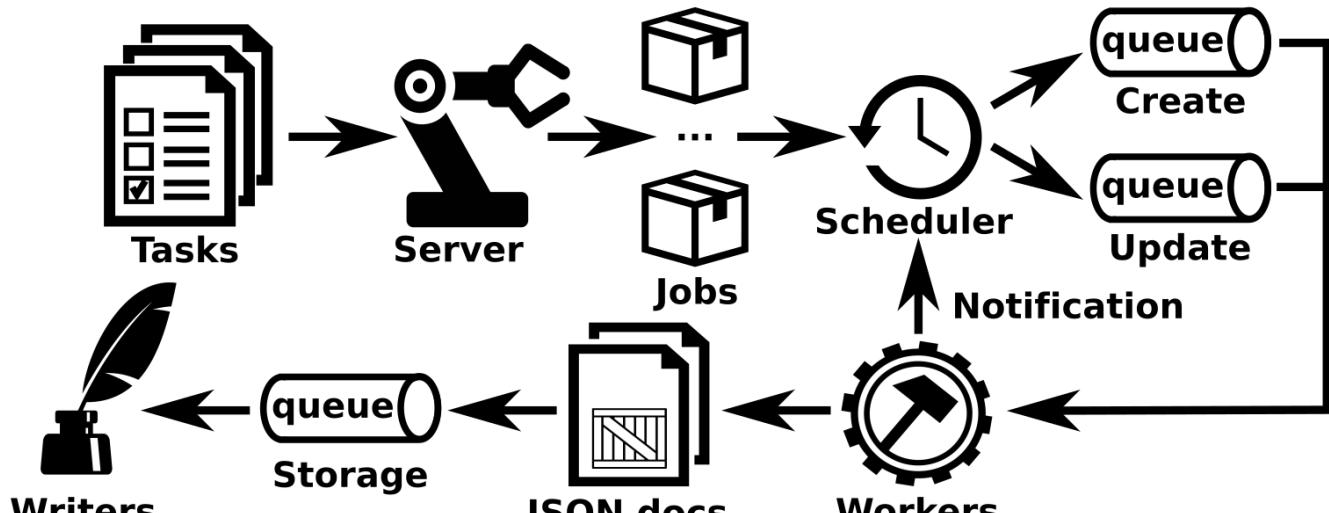
backend = Git(".../grimoirelab-perceval", "/tmp/test")

commits = [commit for commit in backend.fetch(from_date=...)]
```

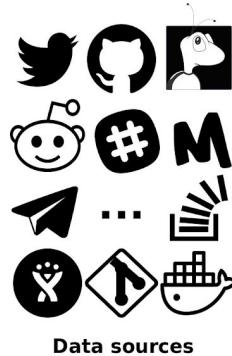
/arthur



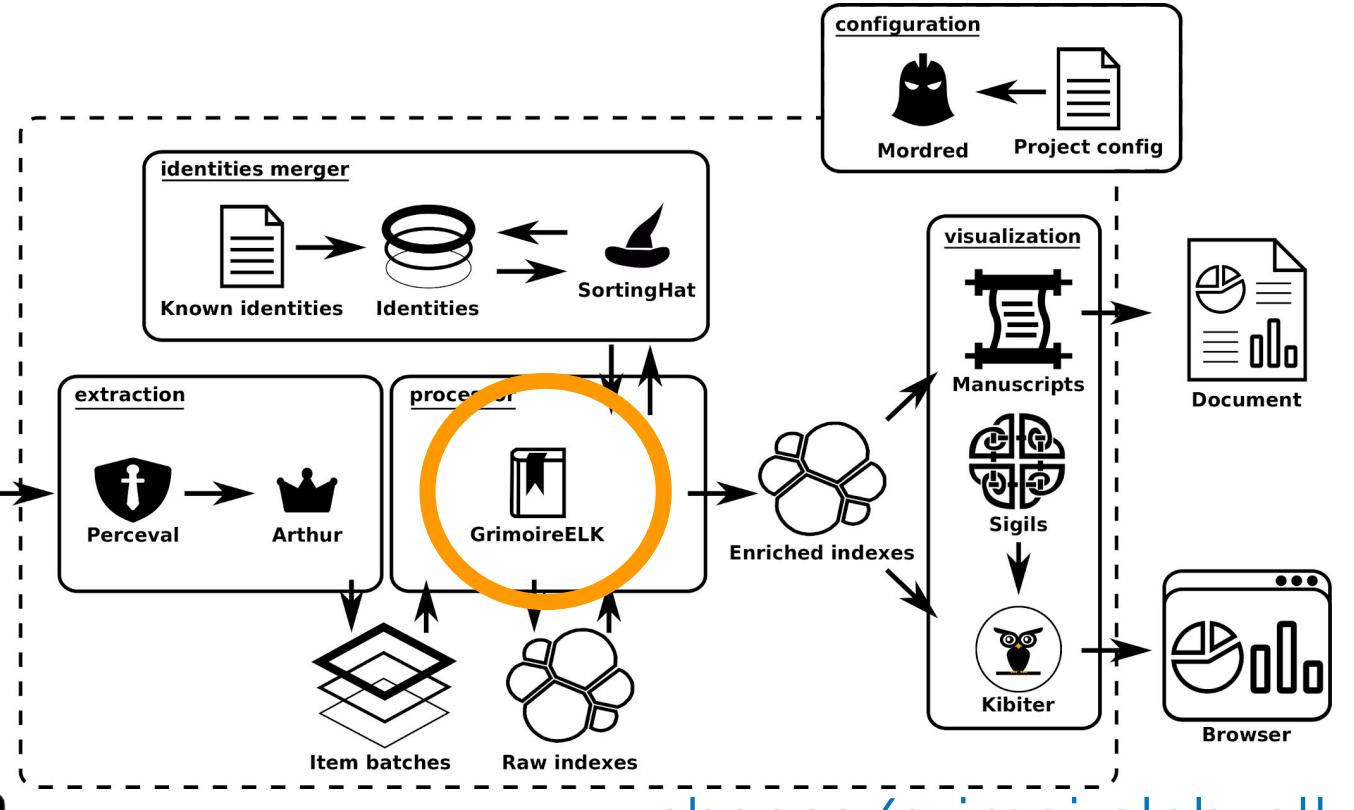
**Arthur** allows to schedule and run Perceval executions at scale through distributed *Redis* queues.



/elk



Data sources

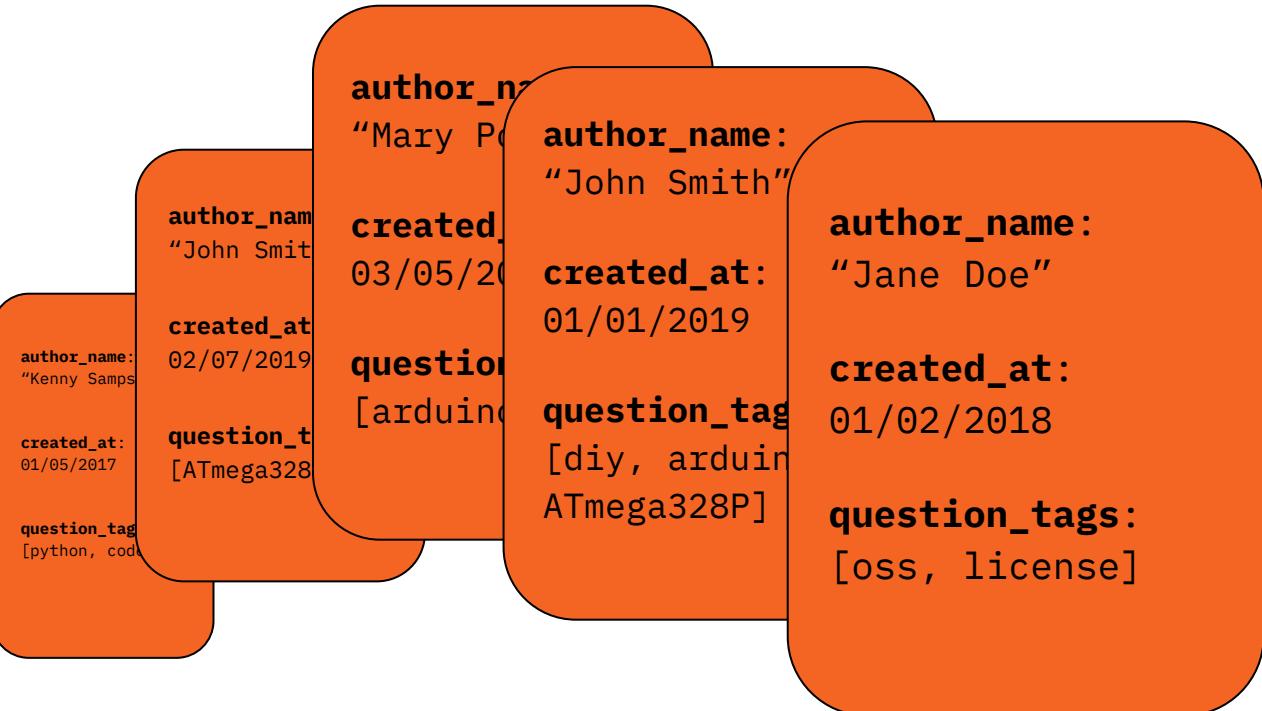


[chaoss/grimoirelab-elk](https://chaoss/grimoirelab-elk)

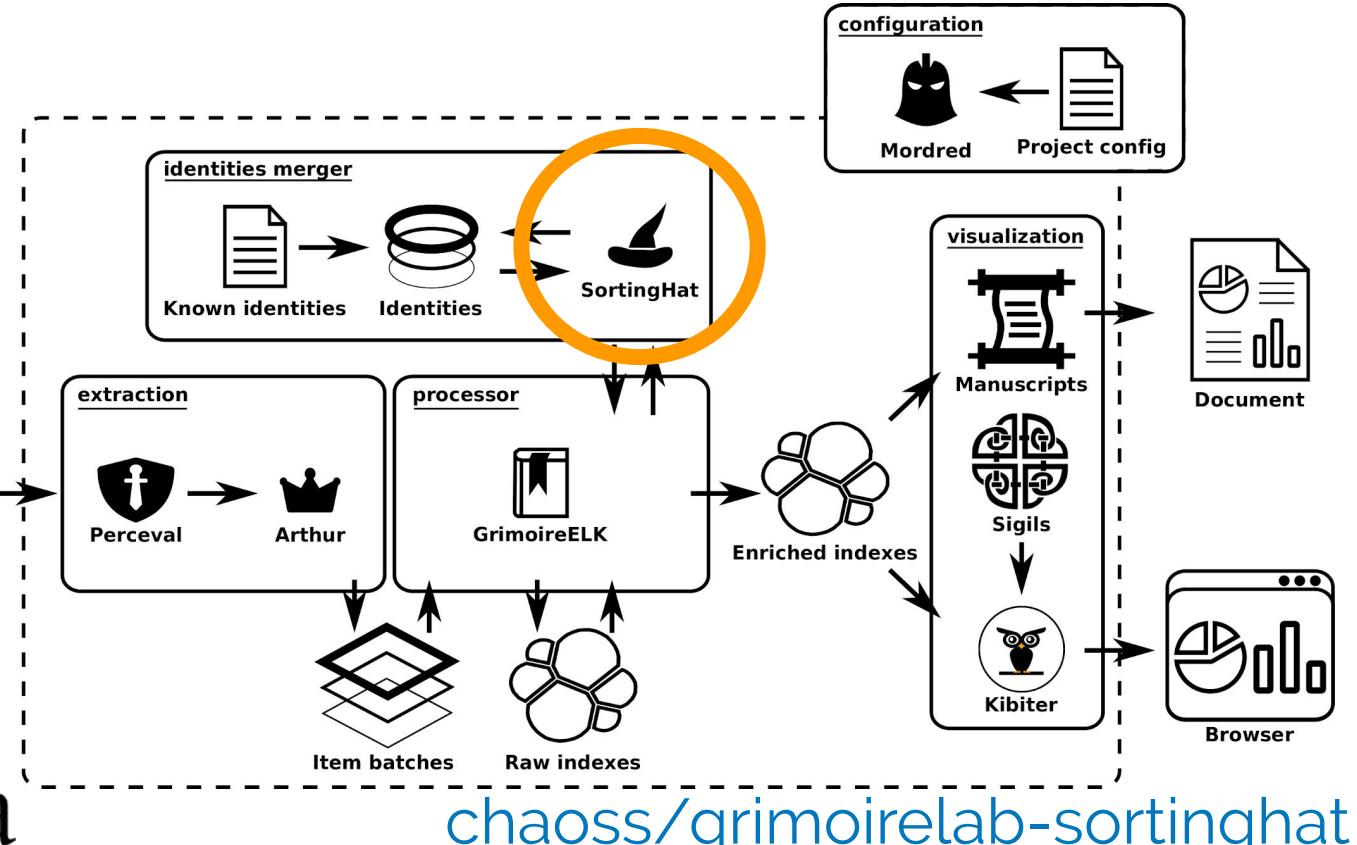
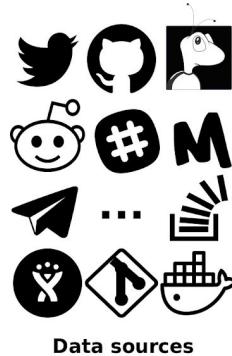
 **Bitergia**

# Fields Documents Indexes

## ElasticSearch



/sortinghat

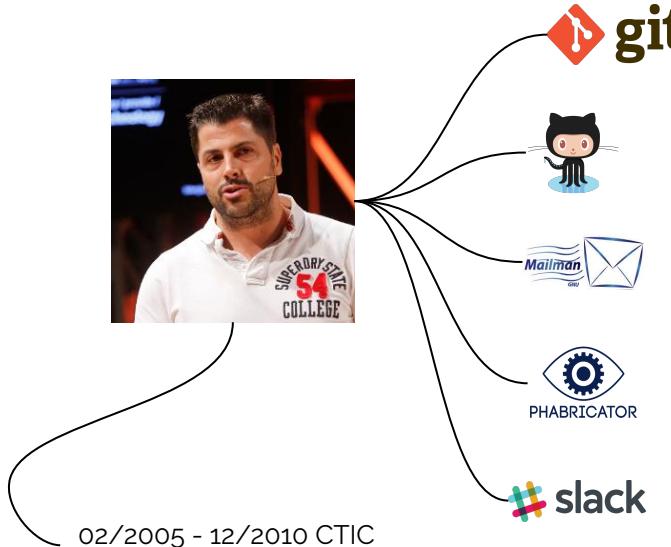


/sortinhat

---



# /sortinghat



02/2005 - 12/2010 CTIC  
01/2010 - 12/2012 Andago  
01/2013 - 06/2013 TapQuo  
07/2013 - 12/2015 freelance (ASOLIF, CENATIC)  
07/2013 - now Bitergia

Manrique López <[jsmanrique@bitergia.com](mailto:jsmanrique@bitergia.com)>  
José Manrique López de la Fuente <[jsmanrique@gmail.com](mailto:jsmanrique@gmail.com)>  
Manrique López <[jsmanrique@gmail.com](mailto:jsmanrique@gmail.com)>

jsmanrique

[jsmanrique@gmail.com](mailto:jsmanrique@gmail.com)  
[jsmanrique@bitergia.com](mailto:jsmanrique@bitergia.com)  
[correo@jsmanrique.es](mailto:correo@jsmanrique.es)

[jsmanrique@bitergia.com](mailto:jsmanrique@bitergia.com)

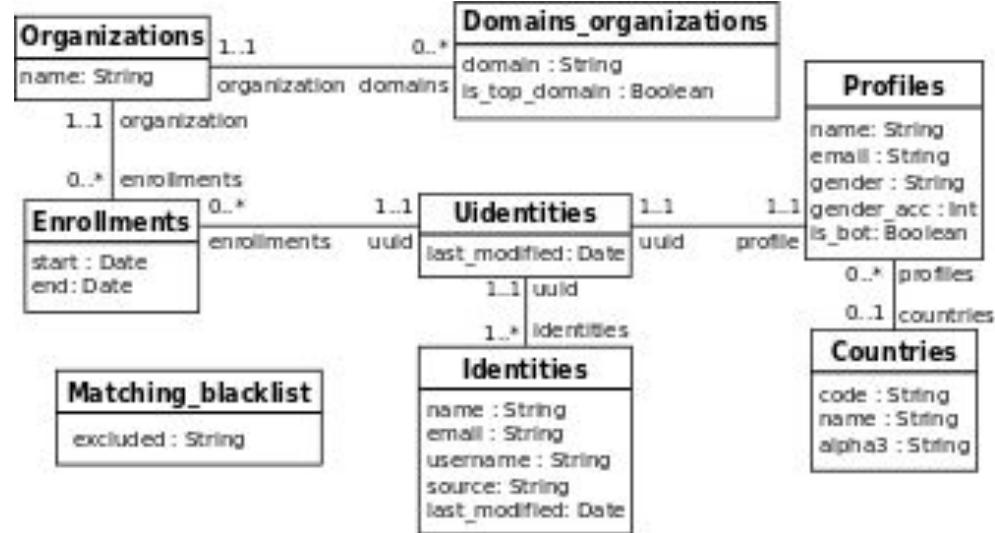
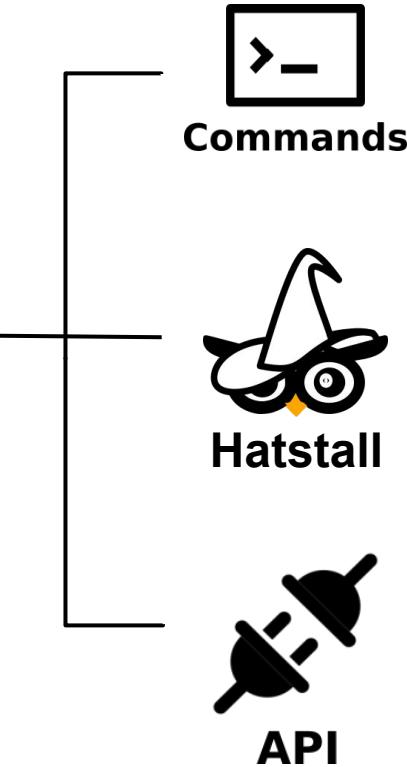
jsmanrique

/sortingshat

---



/sortinghat



# /hatstall



SortingHat:  
Wizardry on Software  
Project Members

aCHAOSproj  
aGimoretLab  
aBitergia  
<https://bitergia.com>

Bitergia

Hatstall Profiles Organizations About

## Profile info / 427f0e8e3d9e9cbcc5ac32b8b84e6f8da

Name Miguel Ángel Fernández

e-mail None

Bot?

Country None

Edit Add

### Enrollments

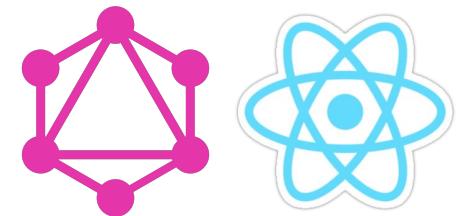
Organization	Dates	Actions
Bitergia	2017-10-02 to 2100-01-01	<button>Update</button> <button>un-enroll</button>
Universidad Rey Juan Carlos	1900-01-01 to 2017-10-01	<button>Update</button> <button>un-enroll</button>

Add

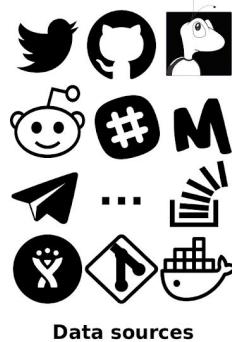
### Profile Identitites

Name - email - Username - Source
Miguel Ángel Fernández - mafesan@bitergia.com - mafesan - github
Miguel Ángel Fernández - None - mghfdez - github
None - mafesan@bitergia.com - None - chaoss:manual

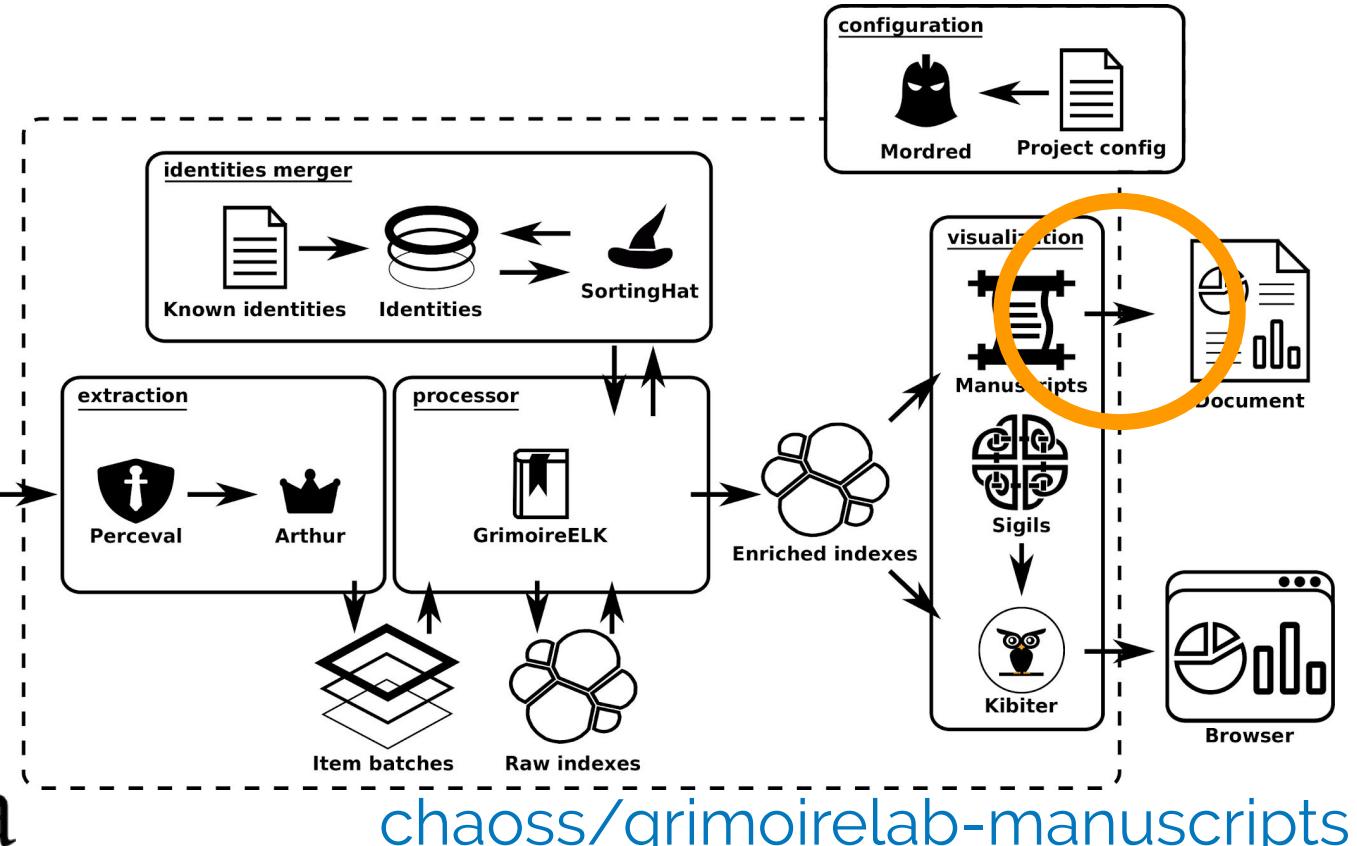
[unmerge](#) [unmerge](#)



/manuscripts



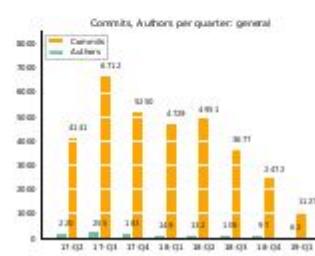
Data sources



## 2 Activity

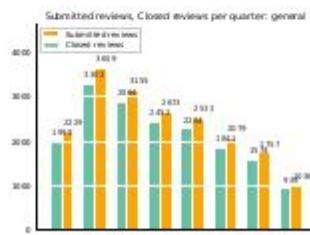
This section covers contributions in the different data sources.

The bar chart below shows the evolution of the number of commits and authors in Git through time, grouped by quarters.



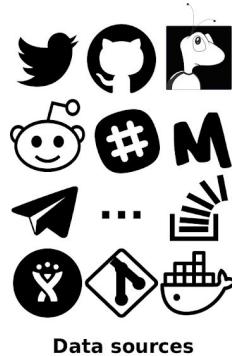
Period	Commits	Authors
17-Q2	4141	220
17-Q3	6712	255
17-Q4	5250	183
18-Q1	4729	149
18-Q2	4951	132
18-Q3	3677	109
18-Q4	2472	97
19-Q1	1127	62

Following chart shows activity in Gerrit data source. It is based on comparing the number of opened and closed reviews grouped by quarters.

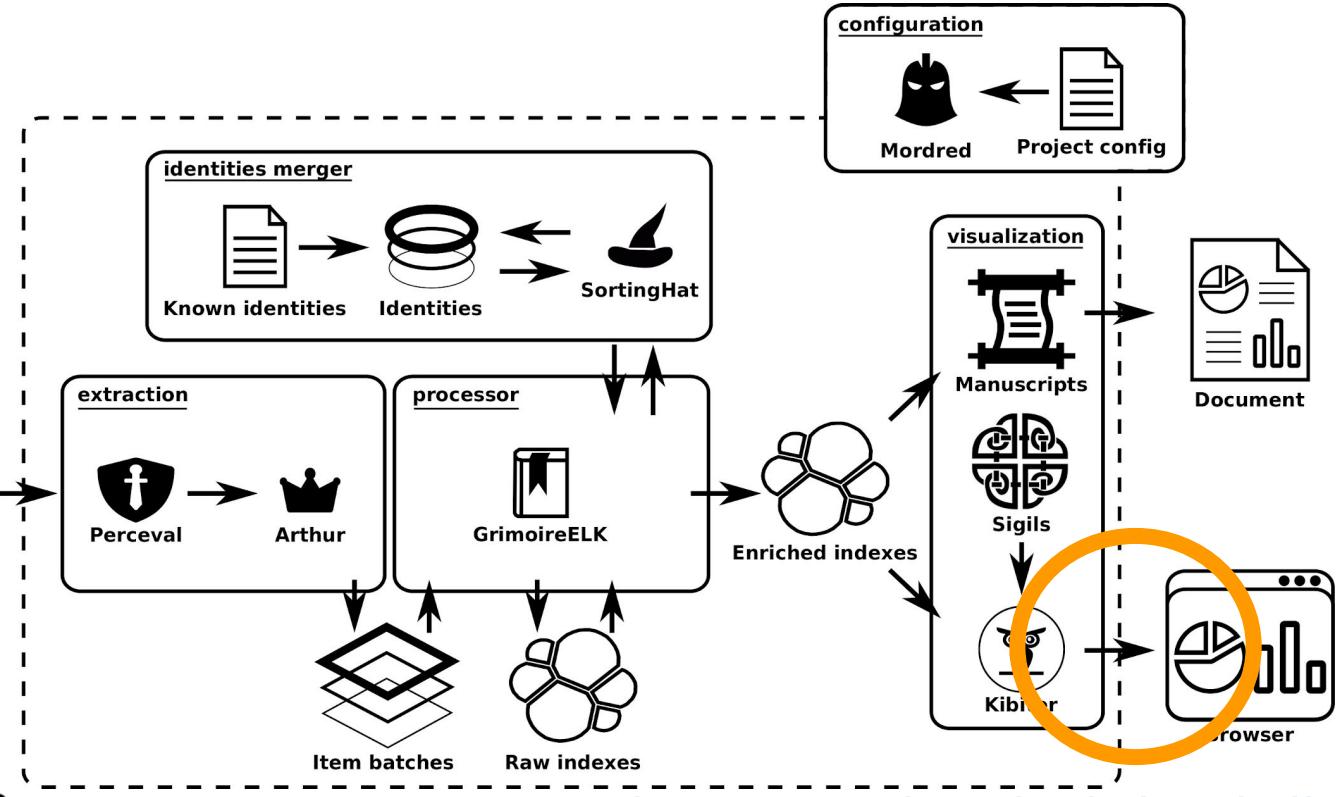


Period	Opened	Closed
17-Q2	2229	1998
17-Q3	3659	3303
17-Q4	3155	2894
18-Q1	2673	2452
18-Q2	2533	2284
18-Q3	2079	1841
18-Q4	1757	1578
19-Q1	1036	939

/sigils



Data sources



[chaoss/grimoirelab-sigils](https://chaoss/grimoirelab-sigils)

Bitergia

# /sigils

chaoss / grimoirelab-sigils

Code Issues 11 Pull requests 5 Projects 0 Wiki Insights Settings

Branch: master grimoirelab-sigils / json / Create new

alpgarcia Add help markdown widget to git demographics panel ...

..

emtech [adoption] Panel for Rust adoption

about.json [about] Update panel to include release date

affiliations-index-pattern.json [affiliations] Add affiliations panel for kibana 6

affiliations.json [affiliations] Set limit author entries due to performance issues

all\_onion-index-pattern.json [onion] Add Kibana 6 Community Structure panels

apache-index-pattern.json [apache] Add Apache panels

apache.json [apache] Add Apache panels

askbot-index-pattern.json [askbot] Update Askbot panel to be similar to Stackoverflow/Discourse

askbot.json Fix badly stacked bar bug

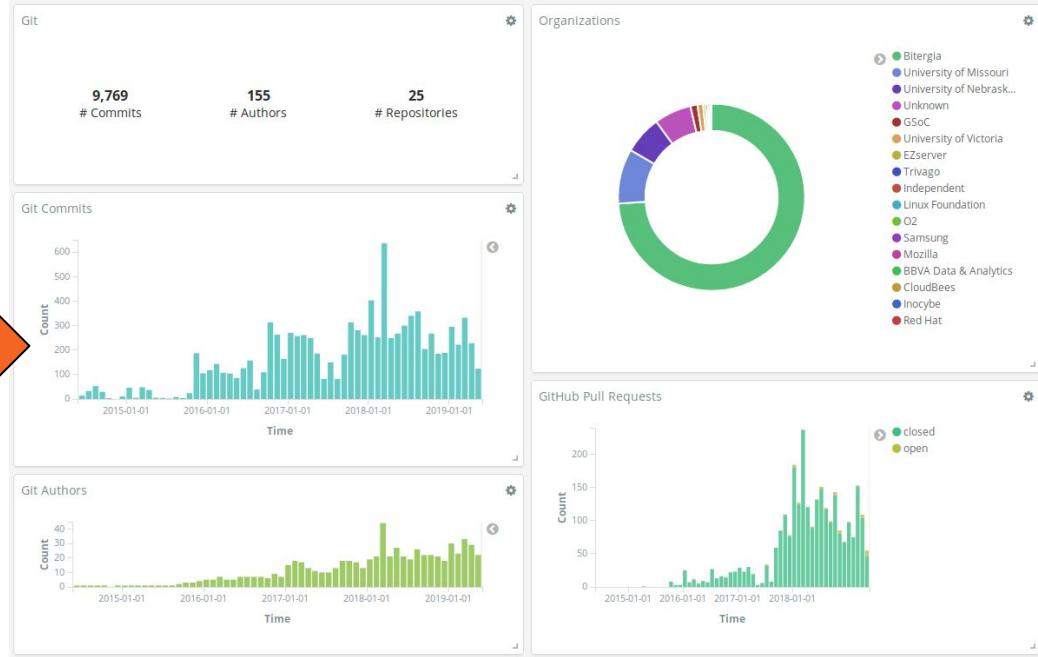
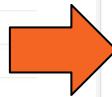
bugzilla-index-pattern.json [bugzilla] Update widget organization

bugzilla.json [bugzilla] Update widget organization

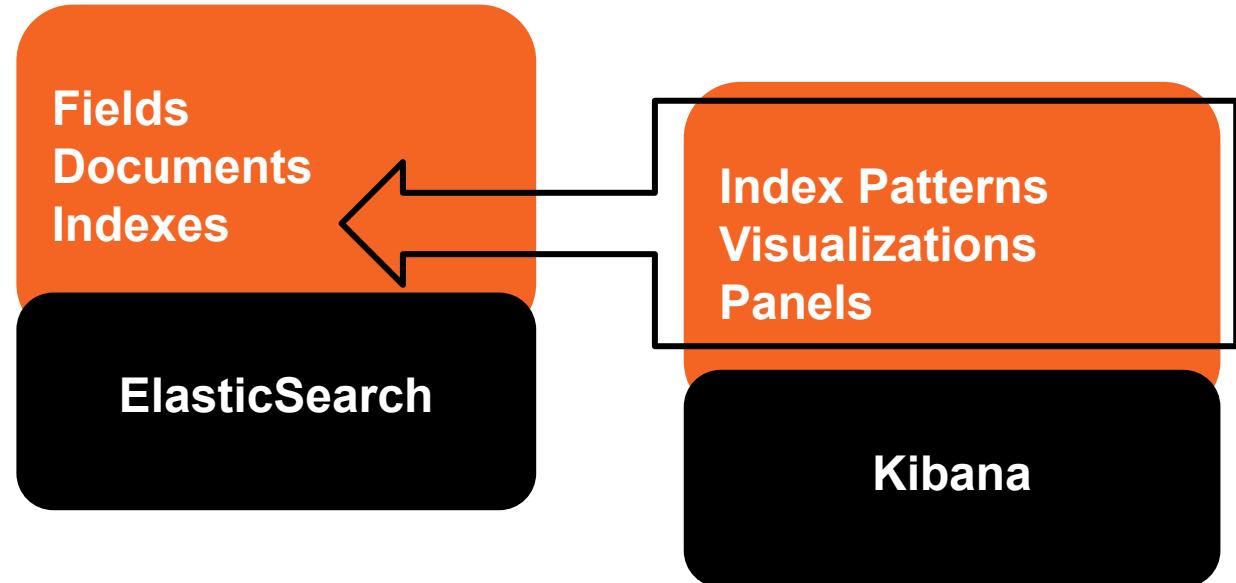
bugzilla\_backlog.json [bugzilla] Update widget organization

bugzilla\_timing.json [bugzilla] Add Kibana 6 Bugzilla-Timing panel

confluence-index-pattern.json [confluence] Add Kibana 6 Confluence panels



Kibiter is a soft-fork of Kibana tailored to GrimoireLab needs



# /index-pattern

35,084 hits

Search... (e.g. status:200 AND extension:PHP)

New Save Open Share < Last 2 years >

Uses lucene query syntax

gitlab

Add a filter +

Selected Fields

? \_source

Available Fields

Popular

t gitlab\_repo

t repository

t \_id

t \_index

# \_score

t \_type

assignee\_bot

t assignee\_domain

t assignee\_gender

# assignee\_gender\_acc

t assignee\_id

? assignee\_location

t assignee\_name

? assignee\_org

t assignee\_org\_name

t assignee\_user\_name

t assignee\_username

t assignee\_uuid

Logout

Collapse

grimoire\_creation\_date per week

November 26th 2016, 11:36:33.960 - November 26th 2018, 11:36:33.960 — Auto

Time \_source

November 23rd 2018 [assignee\_id: 3b6d805aa83112da254938a4bf57ac865428cc4 origin: https://gitlab.com/gitlab-org/gitlab-ce url\_id: gitlab-org/gitlab-ce/issues/54411 updated\_at: November 23rd 2018, 12:07:58.053 author\_bot: false metadata\_timestamp: November 23rd 2018, 21:57:32.472 closed\_at: - author\_org\_name: GitLab time\_to\_first\_attention: - author\_name: Ramya Aauthappa tag: https://gitlab.com/gitlab-org/gitlab-ce url: + Info author\_gender\_acc: 0 author\_id: 3b6d805aa83112da254938a4bf57ac865428bc44 labels: Quality, test plan gitlab\_repo: gitlab-org/gitlab-ce time\_to\_close\_days: - repository: https://gitlab.com/gitlab-org/gitlab-ce uuid: 6d5d89437383cbad695f23bfffdb9fb03f75569 time\_open\_days: 0.42 assignee\_domain: gitlab.com metadata updated\_on: November 23rd 2018, 12:07:58.053 assignee\_org\_name: GitLab id ad695f23bfffdb9fb03f75569]

November 23rd 2018 [assignee\_id: origin: https://gitlab.com/gitlab-org/gitlab-ce updated\_at: November 23rd 2018, 11:53:01.195 author\_bot: false metadata\_timestamp: November 23rd 2018, 21:57:10.563 closed\_at: - assignee\_org\_name: Unknown author\_org\_name: Unknown time\_to\_first\_attention: - author\_name: Dmitry Mekhov tag: https://gitlab.com/gitlab-org/gitlab-ce url: + Info author\_gender\_acc: 0 author\_id: 4b006efb9e123d29896a7085ec8942a29981905 labels: gitlab\_repo: gitlab-org/gitlab-ce time\_to\_close\_days: - repository: https://gitlab.com/gitlab-org/gitlab-ce uuid: 93b4861df490a59dd0eccc362e034cd0ff8f3 time\_open\_days: 0.43 assignee\_domain: metadata updated\_on: November 23rd 2018, 11:53:01.195 assignee\_org: - id: 16,065,326 is gitlab issue: 1 metadata enriched\_on: November 23rd 2018, 22:11:38.484 author\_id 16,065,326]

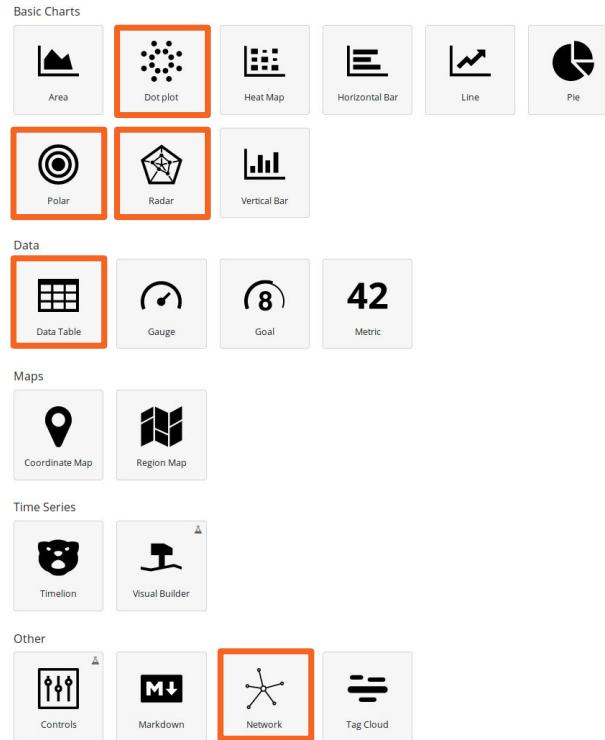
November 23rd 2018 [assignee\_id: origin: https://gitlab.com/gitlab-org/gitlab-ce url\_id: gitlab-org/gitlab-ce/issues/54409 updated\_at: November 23rd 2018, 11:51:28.449 author\_bot: false project: Main closed\_at: - author\_org\_name: Unknown time\_to\_first\_attention: - author\_name: Martin Sefcik tag: https://gitlab.com/gitlab-org/gitlab-ce url: + Info author\_gender\_acc: 0 author\_id: 63de7d446f5d37c2309c31021380f99c7177c4f8 labels: author\_uuid: 63de7d446f5d37c2309c31021380f99c7177c4f8 time\_to\_close\_days: - repository: https://gitlab.com/gitlab-org/gitlab-ce uuid: 5a759f9f7f6fb5674fa3413fe050f0f0264ca7a time\_open\_days: 0.43 assignee\_domain: metadata updated\_on: November 23rd 2018, 11:51:28.449 assignee\_org: - id: 16,065,388 is gitlab issue: 1 metadata enriched\_on: November 23rd 2018, 22:11:37.348 author\_id 16,065,388]

November 23rd 2018 [assignee\_id: origin: https://gitlab.com/gitlab-org/gitlab-ce url\_id: gitlab-org/gitlab-ce/issues/54408 updated\_at: November 23rd 2018, 11:56:14.694 author\_bot: false project: Main closed\_at: - author\_org\_name: Unknown time\_to\_first\_attention: 0.01 author\_name: drunkmogli tag: https://gitlab.com/gitlab-org/gitlab-ce url: + Info author\_gender\_acc: 0 author\_id: 01aa77a91c1880c3422d535db1626465958fb8c5 labels: api, awaiting feedback, support request author\_uuid: 01aa77a91c1880c3422d535db1626465958fb8c5 time\_to\_close\_days: - repository: https://gitlab.com/gitlab-org/gitlab-ce uuid: 8e02fe8b7f70747c66cc7d0966f8cafddc4223b time\_open\_days: 0.43 assignee\_domain: metadata updated\_on: November 23rd 2018, 11:56:14.694 assignee\_org: - id: 16,065,211 is gitlab issue: 1 metadata enriched\_on: November 23rd 2018, 22:11:37.348 author\_id 16,065,211]

November 23rd 2018 [assignee\_id: origin: https://gitlab.com/gitlab-org/gitlab-ce url\_id: gitlab-org/gitlab-ce/issues/54407 updated\_at: November 23rd 2018, 11:56:15.003 author\_bot: false project: Main closed\_at: - author\_org\_name: Unknown time\_to\_first\_attention: 0.01 author\_name: drunkmogli tag: https://gitlab.com/gitlab-org/gitlab-ce url: + Info author\_gender\_acc: 0 author\_id: 01aa77a91c1880c3422d535db1626465958fb8c5 labels: api, awaiting feedback, support request author\_uuid: 01aa77a91c1880c3422d535db1626465958fb8c5 time\_to\_close\_days: - repository: https://gitlab.com/gitlab-org/gitlab-ce uuid: 8e02fe8b7f70747c66cc7d0966f8cafddc4223b time\_open\_days: 0.43 assignee\_domain: metadata updated\_on: November 23rd 2018, 11:56:15.003 assignee\_org: - id: 16,065,210 is gitlab issue: 1 metadata enriched\_on: November 23rd 2018, 22:11:37.348 author\_id 16,065,210]



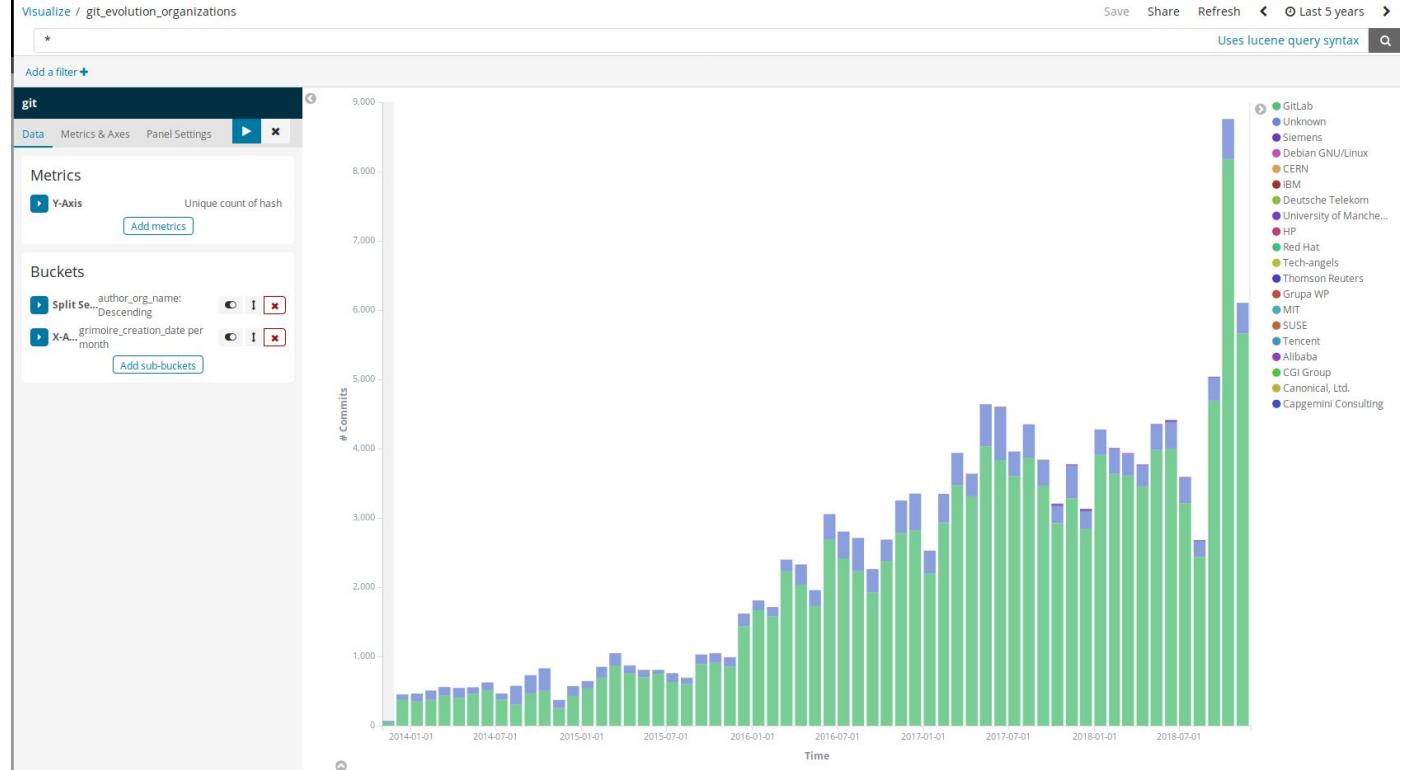
# /visualizations



Q. Search...	Type
☐ Title I	Data Table
☐ affiliations_authors	Pre
☐ affiliations_authors_organizations	Tag Cloud
☐ affiliations_data_sources	Data Table
☐ affiliations_organizations_table	Data Table
☐ affiliations_top_projects	Data Table
☐ all_onion_developers_active_quarters_by_project	Data Table
☐ all_onion_global_community_bar_chart	Vertical Bar
☐ all_onion_global_contributions	Vertical Bar
☐ all_onion_global_developers_active_quarters_by_project	Data Table
☐ all_onion_global_roles	Data Table
☐ all_onion_organizations	Vertical Bar
☐ all_onion_organizations_by_contributions	Data Table
☐ all_onion_orgs_community_bar_chart	Vertical Bar
☐ all_onion_orgs_contributions	Vertical Bar
☐ all_onion_orgs_developers_active_quarters	Data Table
☐ all_onion_projects	Data Table
☐ all_onion_projects_by_contributions	Pre
☐ all_onion_projects_community_bar_chart	Vertical Bar
☐ all_onion_projects_contributions	Vertical Bar

\* Non-standard Kibana visualizations

# /visualizations



## /visualizations

git

Data Metrics & Axes Panel Settings

Metrics

Y-Axis

Aggregation

Unique Count

Field

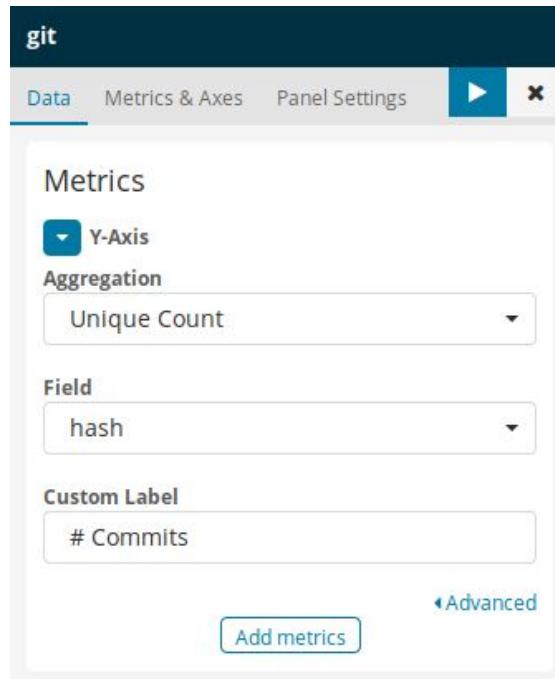
hash

Custom Label

# Commits

Add metrics

Advanced



Buckets

Split Series

Aggregation

Terms

Field

author\_org\_name

Order By

metric: # Commits

Order

Descendi

Size

20

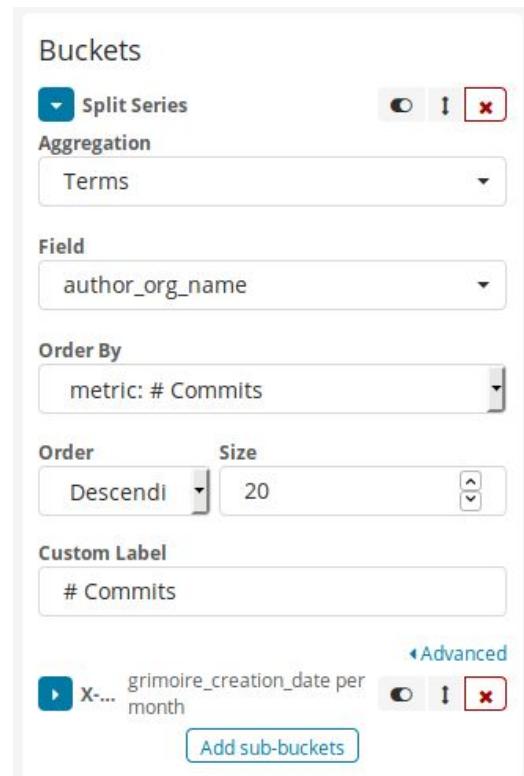
Custom Label

# Commits

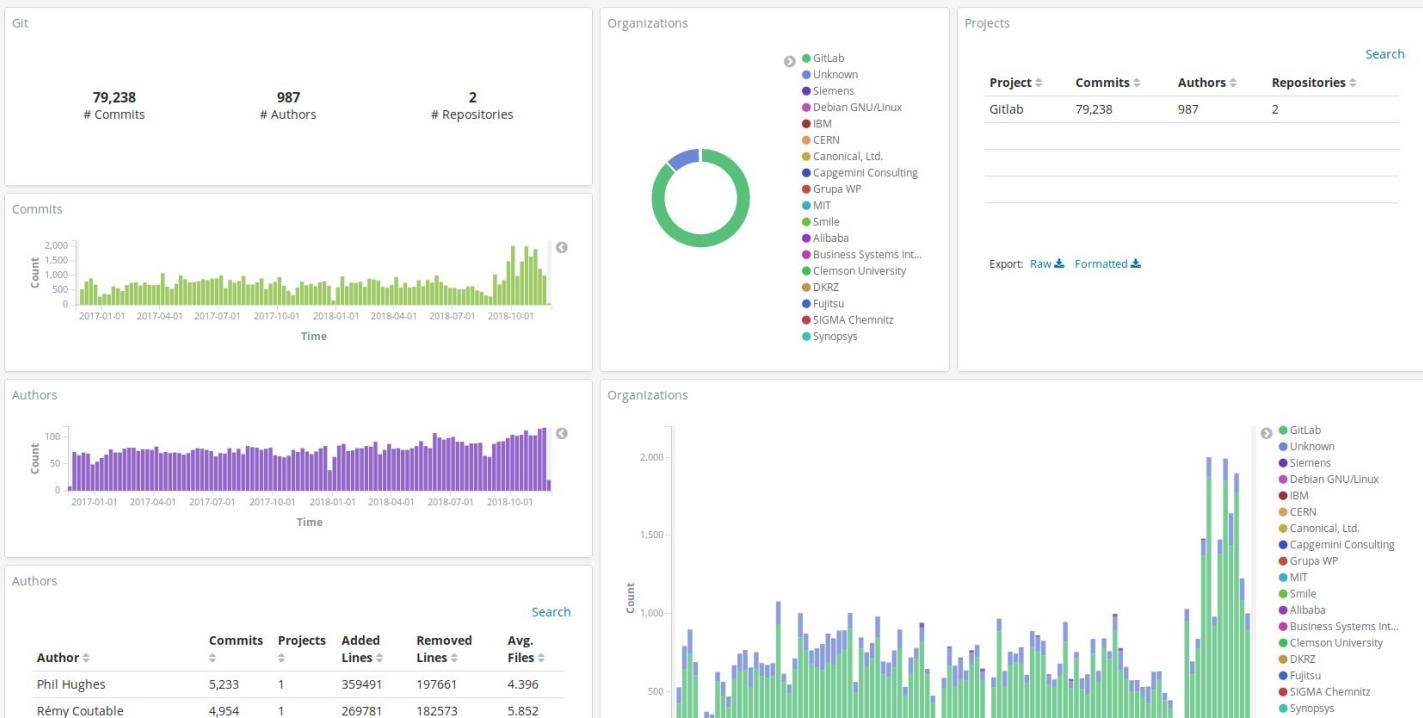
Advanced

X... grimoire\_creation\_date per month

Add sub-buckets



The visualization panel on the left displays a single metric configuration for the 'git' dataset. The Y-axis is set to 'Unique Count' for the field 'hash', and the custom label is '# Commits'. The bucket configuration panel on the right shows a main bucket for 'author\_org\_name' using 'Terms' aggregation, ordered by 'metric: # Commits' in descending order with a size of 20. A sub-bucket for 'grimoire\_creation\_date per month' is also defined.





How can I measure my software project?





## Activity

**How many contributors we have ?**

## Community

**How many companies are contributing  
to my project?**

## Processes

**How good we are in handling issues?**

# Answering questions with visualizations

Data from CHAOSS project  
[chaoss.biterg.io](http://chaoss.biterg.io)



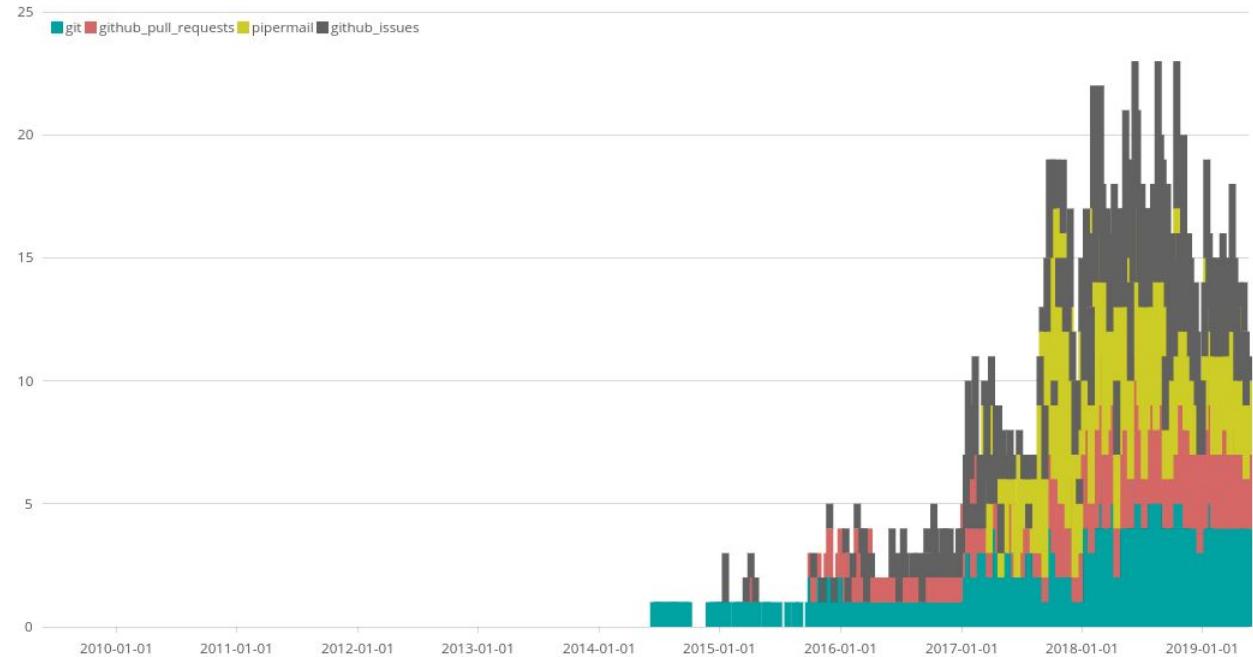
**How many contributors  
do we have ?**

# Time series (bar chart)



# Time series

Active Organizations over Time by Data Source

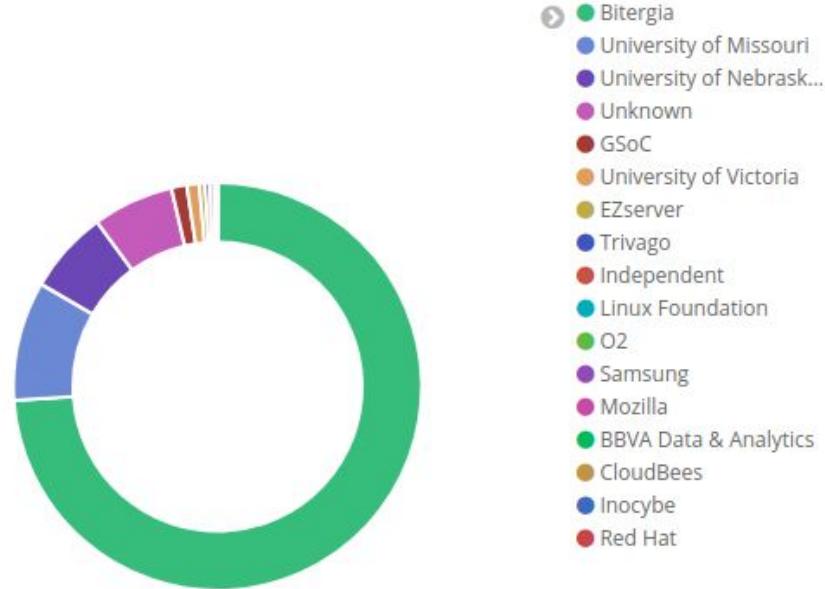




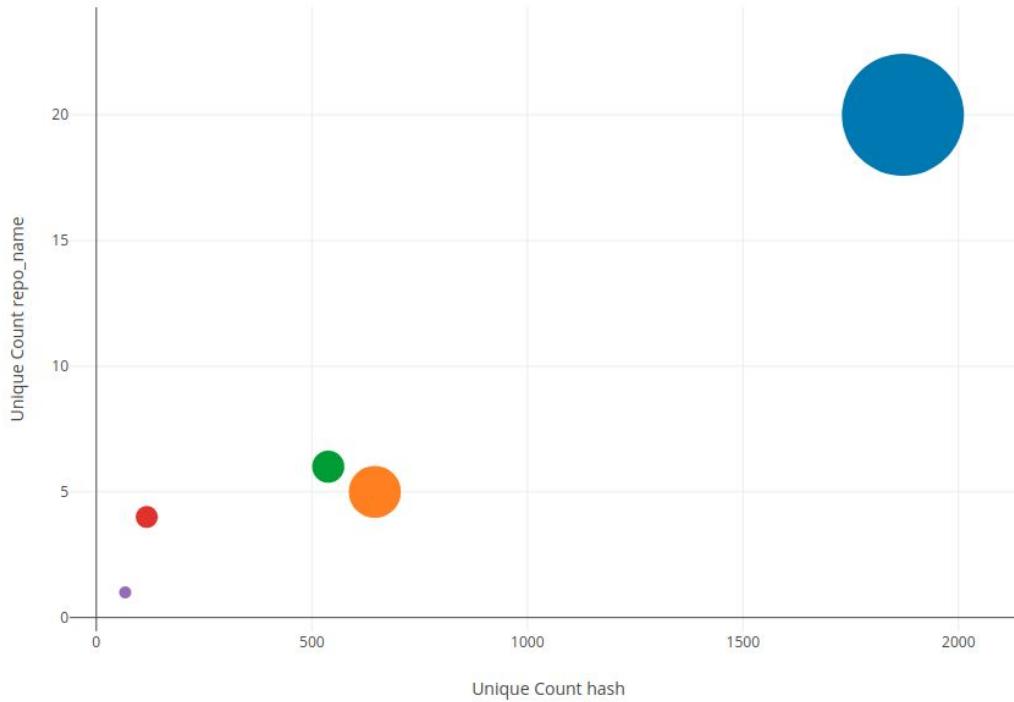
**How many companies are contributing to my project?**

# Pie chart

Organizations



# Dot plot





How good we are in handling  
issues?

# Metric and Gauge

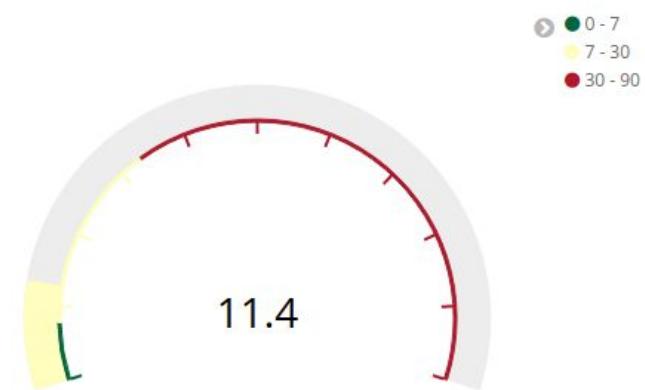
**575**  
# Issues

**72**  
# Submitters

**21**  
# Assignees

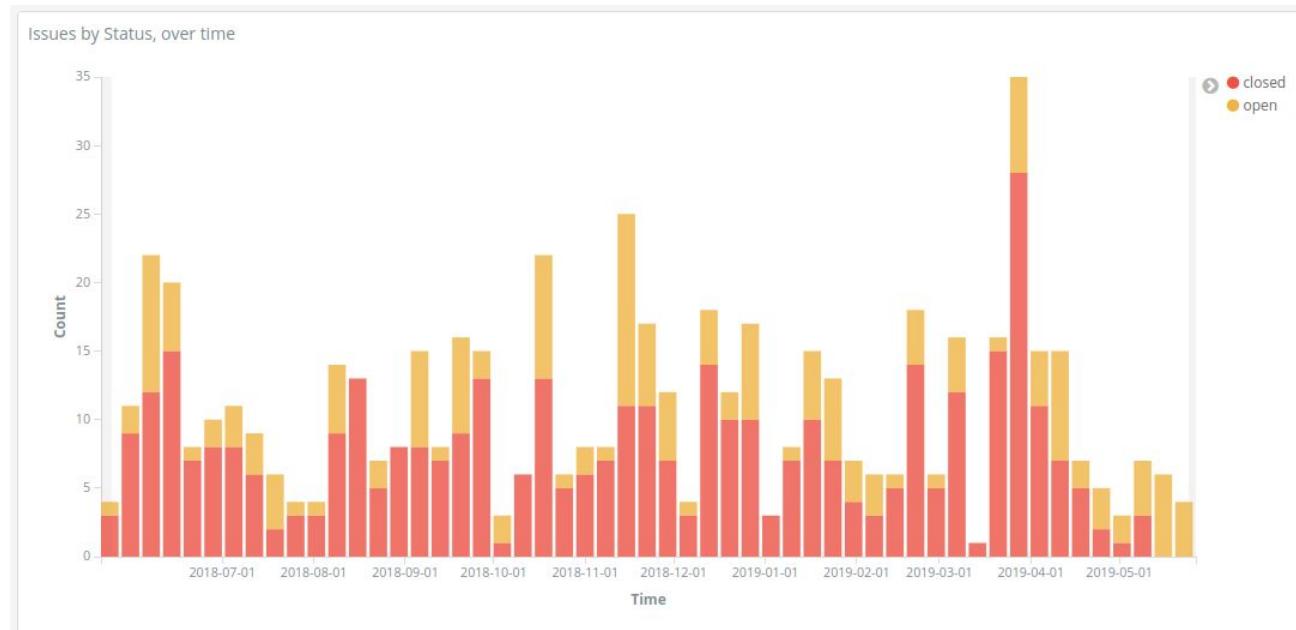
**21.879**  
Median Time Open (Days)

Time to Close

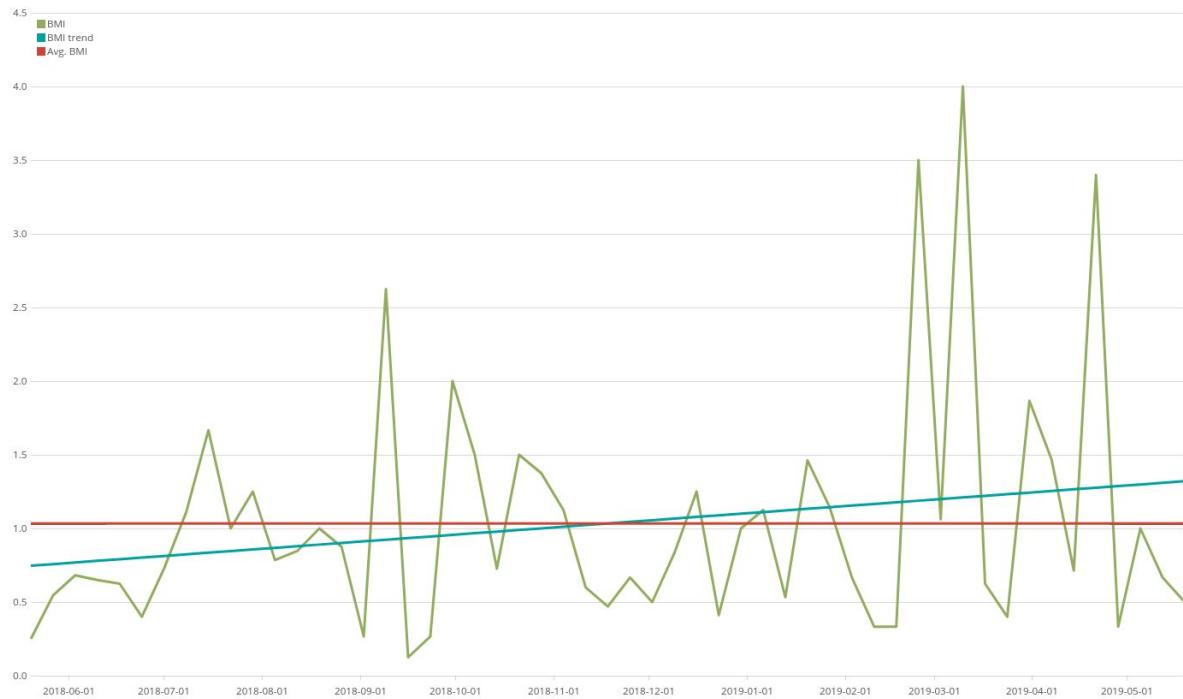


Median Time to Close (Days)

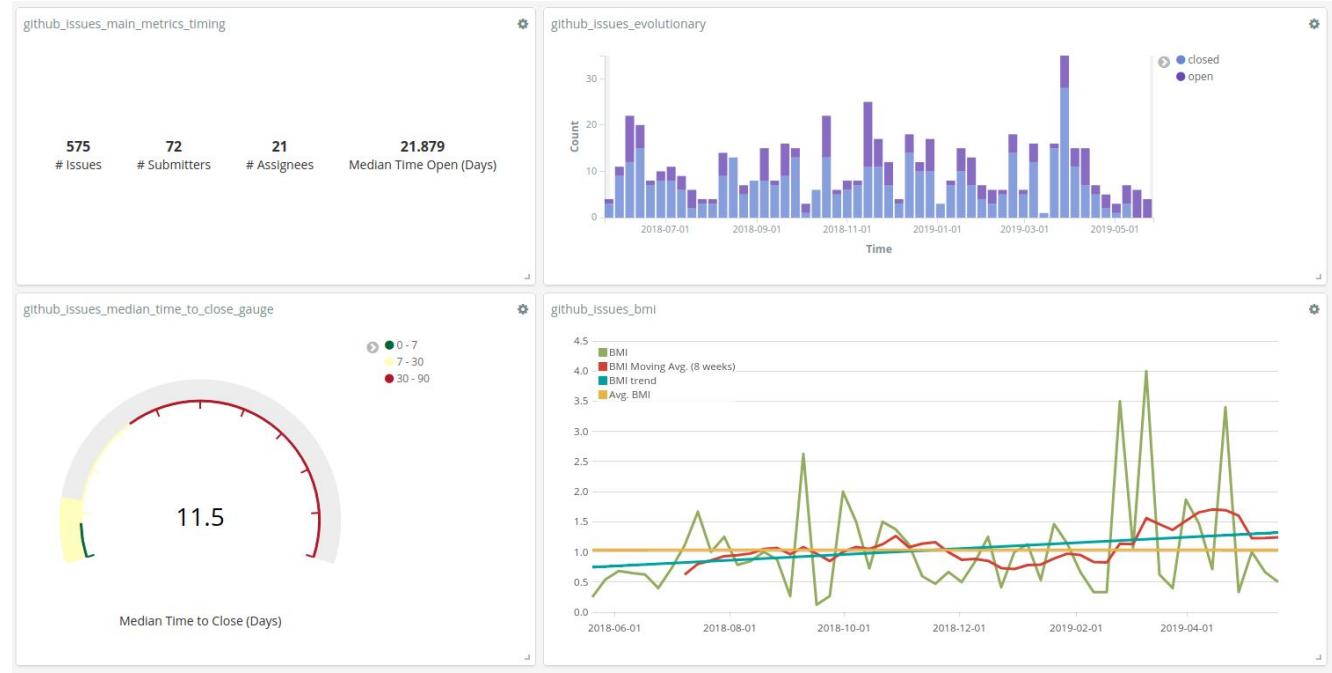
# Bar chart



# BMI (Backlog Management Index)



# Efficiency panel (custom)



# Let's go for questions!

Miguel Ángel Fernández, Valerio Cosentino  
{mafesan, valcos}@bitergia.com

@Mghfdez, @\_valcos\_



@grimoirelab  
@bitergia

@CHAOSSproj  
@crossminer

Thanks

