



# Methodology of Multi-Criteria Comparison and Typology of Open Source Projects

Fedir RYKHTIK, October 22, 2018, Edinburgh, UK

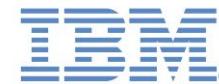
@FedirFR

# Sponsors

DIAMOND



PLATINUM



GOLD



# Fedir RYKHTIK

- Building open source web since 2007
  - Back-end developer
  - Independent researcher
  - DevOps / SA
- CTO @AgenceStratis since 2015



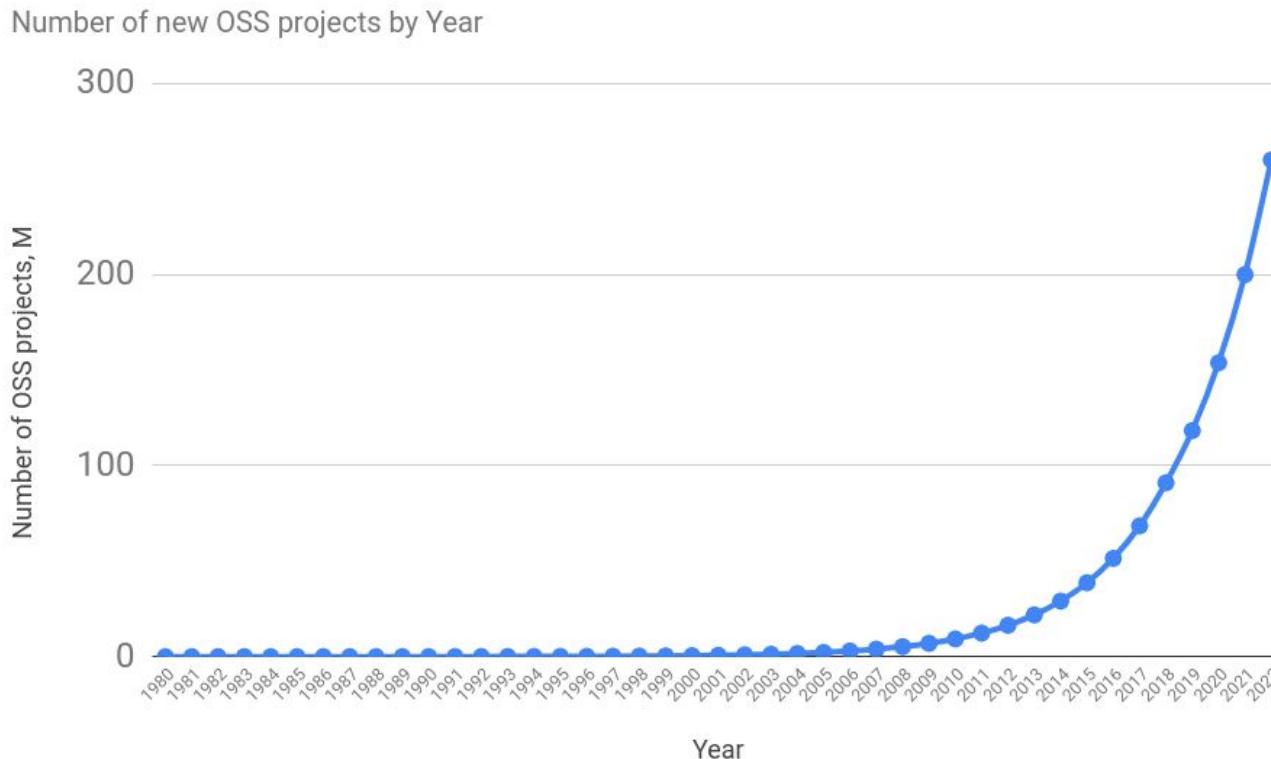
# OSS today

# OSS today

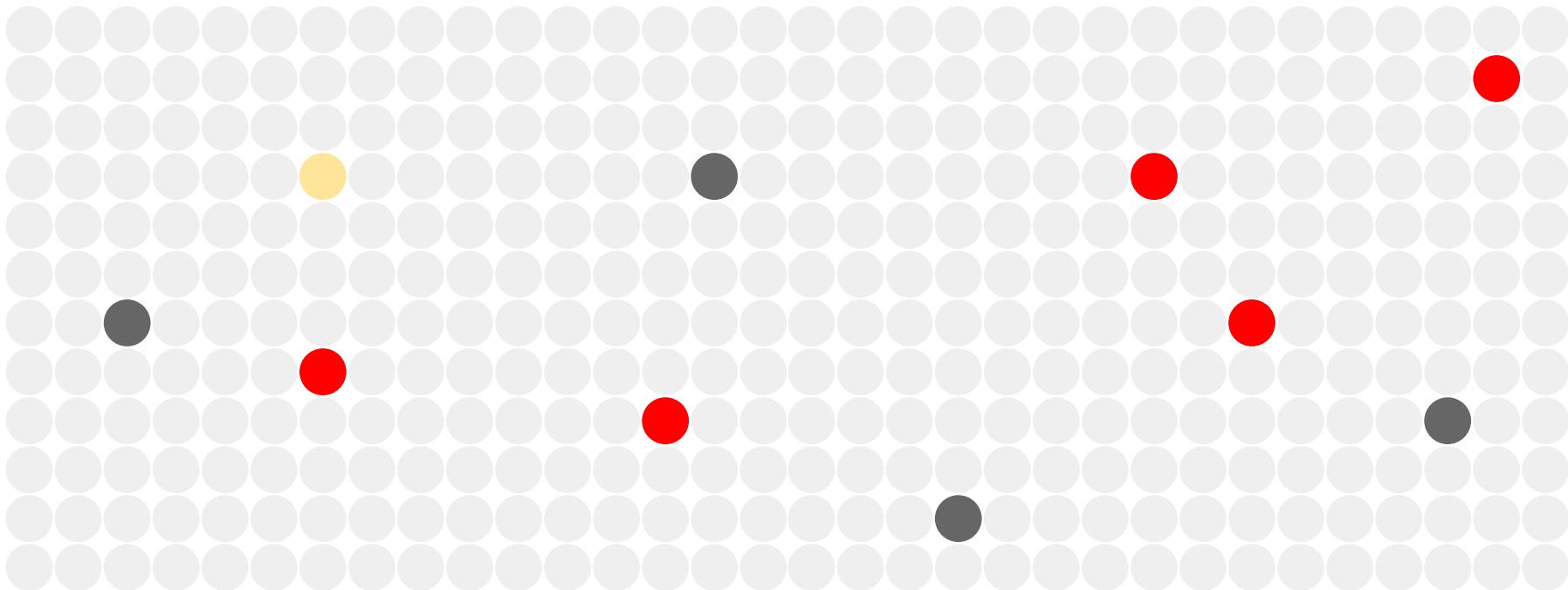
## Github only

- 96 M+ repositories
- 40% more than last year
- 31 M+ developers

# OSS today



# Problem of choice & following

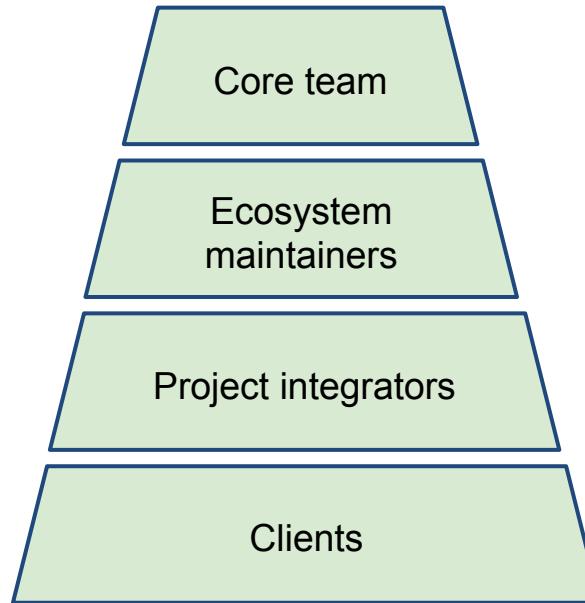


# Multi-Criteria Comparison

# What is a good OSS project ?

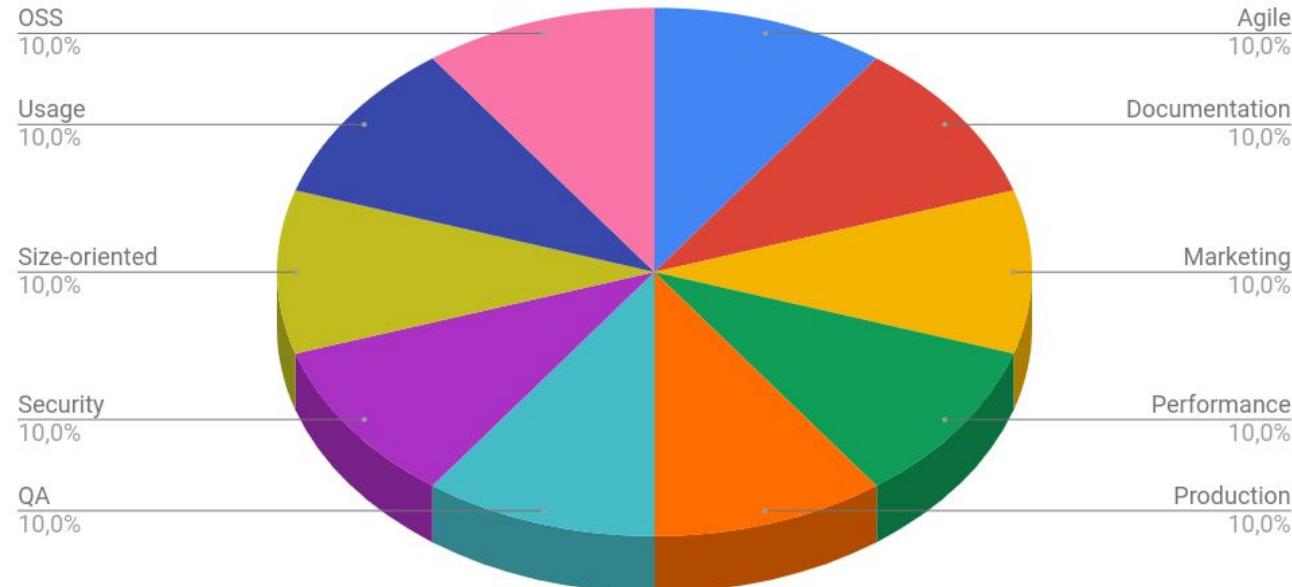
Reliable  
useful  
Maintained  
Secure

# Different layers have own metrics



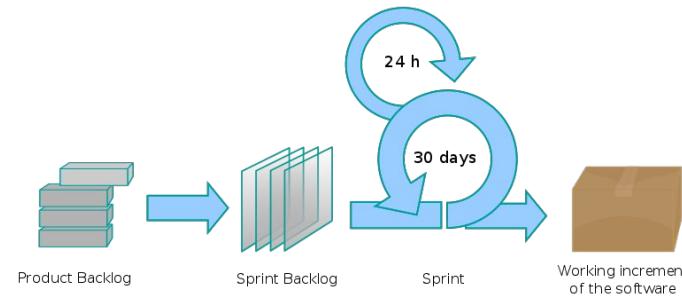
# Software groups of metrics (A-Z)

# Metrics groups



# Metrics > Agile

- Lead time (period of time between ticket creation and resolving)
- Open / closed rate



# Metrics > Documentation

- Technical documentation coverage
- Articles & manuals
- Books



# Metrics > Marketing

- Social networks marketing
- Search engine optimization



# Metrics > Performance

- Volume of servers/CPU/... required
- Execution speed
- Supported charge



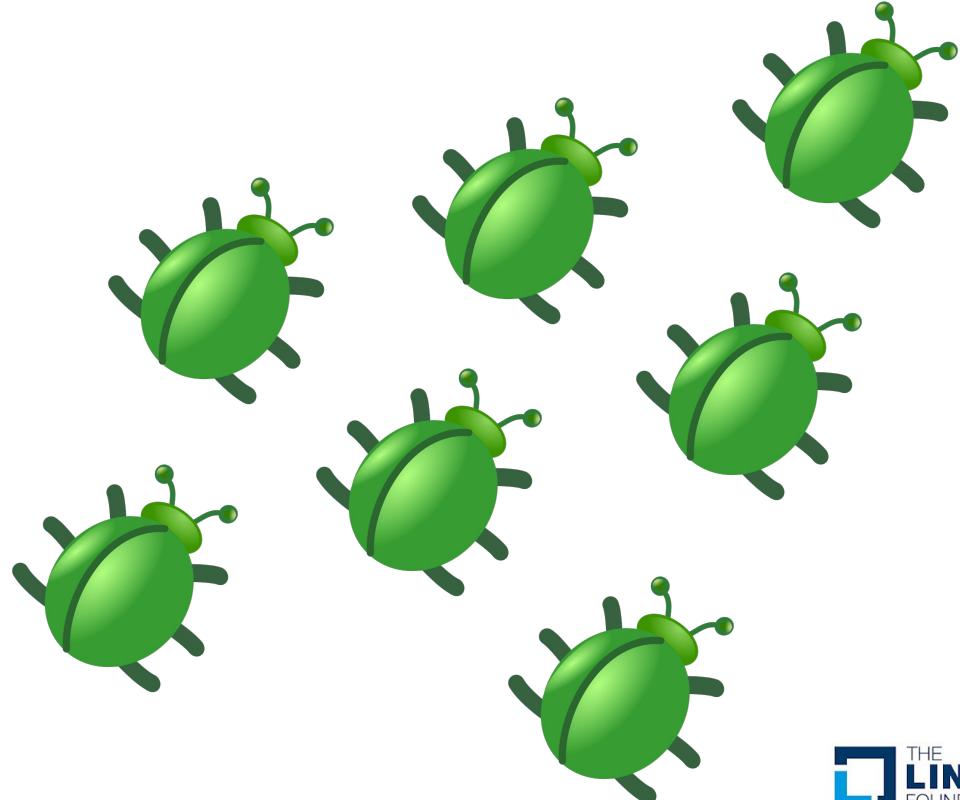
# Metrics > Production

- Active days
- Tasks scope
- Code churn
- Apps crash rate



# Metrics > QA

- Number of bugs
- Frequency of bugs
- Returning bugs



# Metrics > Security

- Time / ressources to find a security bug
- Time / ressources to fix it



# Metrics > Size-oriented

- Number of code lines
- Number of bugs per 1000 code lines
- Number of classes and interfaces
- Number of commits



# Metrics > Usage

- Accessibility
- Number of features
- Simplicity of usage
- Unique features



# OSS specific metrics

# Metrics > OSS > Author

- Notoriety / Experience
- Involvement



## Metrics > OSS > Community of contributors / integrators

- Social ranking (stars)
- Downloads



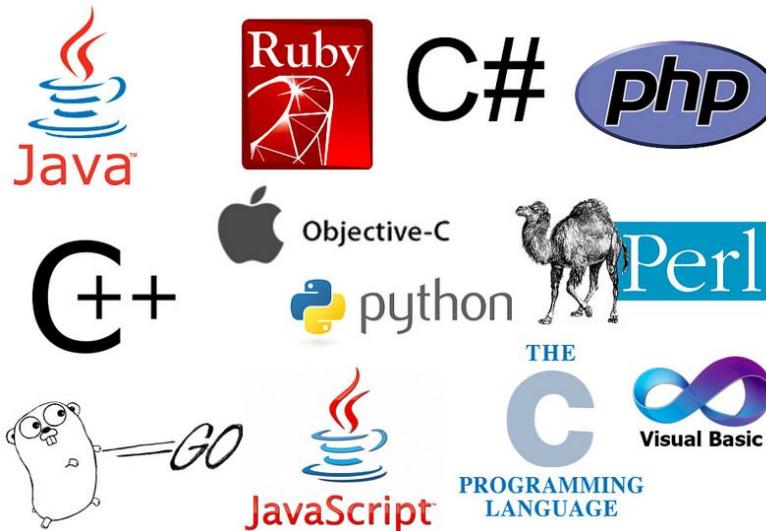
# Metrics > OSS > Community of developers

- Community size
- Active forks
- Notoriety / Experience
- Returning contributors
- Medium contribution period



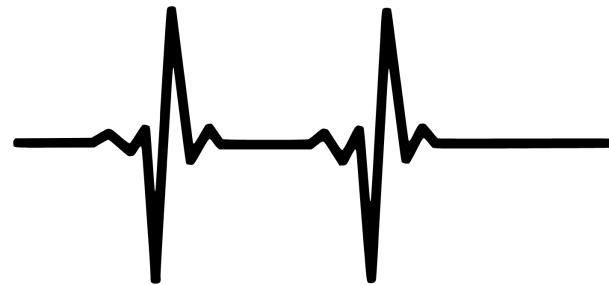
# Metrics > OSS > Languages

- Number of used languages
- Popularity of the language



# Metrics > OSS > Rhythm

- Last commits
- Regular maintenance
- New versions



# How do we collect it ?

# Data collection

- Code analysis (LoC, coding conventions, coupling, deps)
- Unit testing (code coverage, number of scenarios)
- Versioning systems (code, contributors, branches, tags)
- Using application (confirming functional perimeter)
- Social networks (community, feedback)
- Search engines (buzz, books, materials)
- Bugtracker statistics (bugs, maintainers activity)
- Benchmarking (load, endurance, stress, limits)
- Pentesting (automated, manual)



# Project quality index

# Counting project quality index

$$Q_{total} = \sum_m Q_m$$

$Q_m$  - quality index of current metric

# Counting project quality (custom)

$$Q_{total} = \sum_m (Q_m * C_m)$$

$Q_m$  - quality index of current metric

$C_m$  - optional custom coefficient for current metric

# Analyzer prototype

# fedir/ghstat (WIP)

- <https://github.com/fedir/ghstat>
  - Statistical multi-criteria comparator for Github's projects
- What does it do
  - Collects statistics from Github
  - Calculate additional metrics
  - Gives points and ranks projects
- Statistics: Name, URL, Author, Author's location, Main language, All used languages, Number of languages, Description, Total code size, License, Author's followers, Top 10 contributors followers, Created at, Age in days, Total commits, Total additions, Total deletions, Total code changes, Last commit date, Commits/day, Average contribution period by contributor in days, Medium commit size, Total releases, Stargazers, Forks, Contributors, Active forkers(%), Returning contributors (more than 4 weeks), Open issues, Closed issues, Total issues, Issue/day, Closed issues (%)

# fedir/ghstat statistics examples

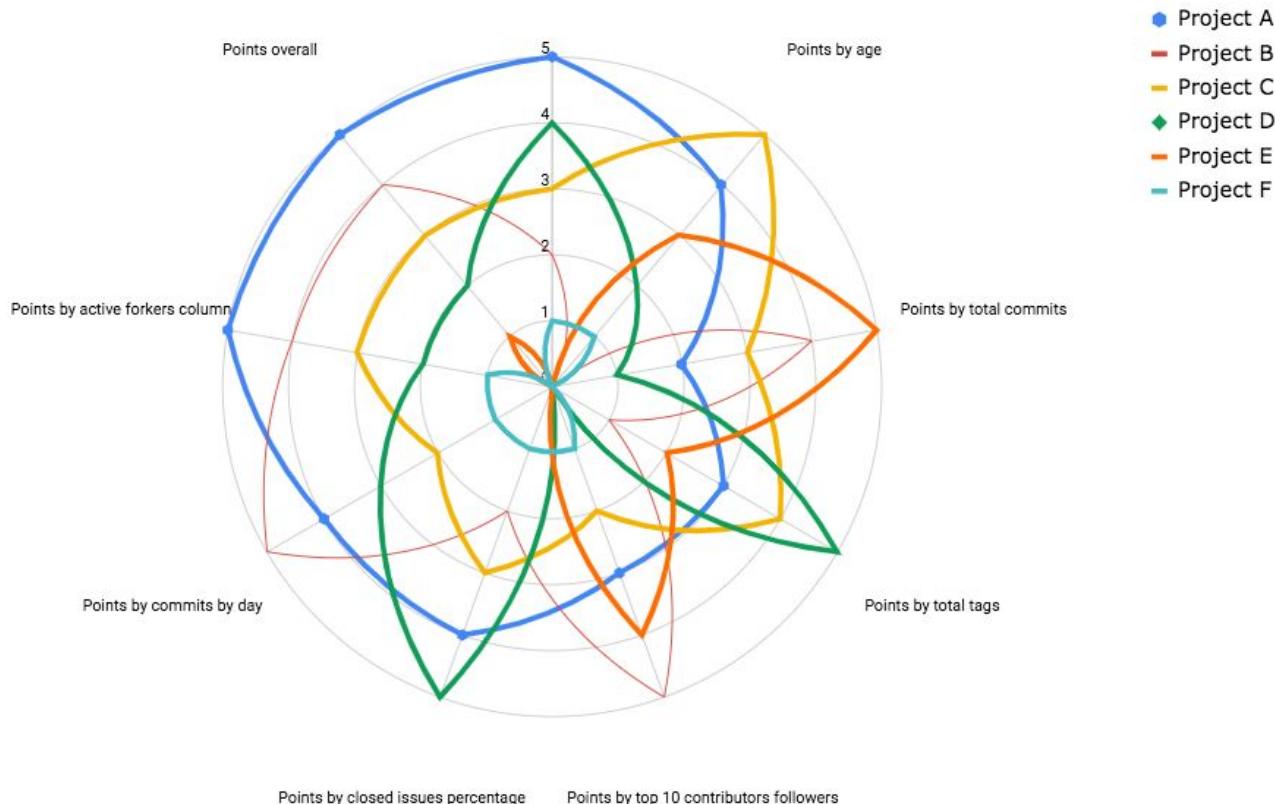
<https://github.com/fedir/ghstat/tree/master/stats>

- Open source programming languages
- Web frameworks
- Content management systems
- ...

# fedir/ghstat - placements

- Placement by popularity
- Placement by age
- Placement by total commits
- Placement by total tags
- Placement by top 10 contributors followers
- Placement by closed issues percentage
- Placement by commits by day
- Placement by active forkers column

# fedir/ghstat results sample



# fedir/ghstat > future features

- offline git repository scanner
- social connectors (gitlab, bitbucket, gitea...)
- more metrics
  - lots of stuff to do

# Live demo

```
$ ./ghstat -r Microsoft/MS-DOS,freebsd/freebsd,redox-os/redox -f os.csv  
$ ./ghstat -r torvalds/linux
```

# Summary



# Results

- OSS has additional domain specific metrics
- Using multi-criteria comparison methods we can choose / control states of packages health by our needs
- Using statistical analyzers gives You more possibilities
  - You could identify interesting packages faster
  - As developer, You could identify, which project needs Your help
  - As contributor, You could identify, which project follows Your needs



# Q&A

## Any questions ?

# Related materials

- <https://octoverse.github.com/>
- [https://en.wikipedia.org/wiki/Multiple-criteria\\_decision\\_analysis](https://en.wikipedia.org/wiki/Multiple-criteria_decision_analysis)
- [https://en.wikipedia.org/wiki/Analytic\\_hierarchy\\_process](https://en.wikipedia.org/wiki/Analytic_hierarchy_process)
- [https://en.wikipedia.org/wiki/Group\\_decision-making](https://en.wikipedia.org/wiki/Group_decision-making)
- [https://en.wikipedia.org/wiki/Analytic\\_network\\_process](https://en.wikipedia.org/wiki/Analytic_network_process)
- <https://wackowiki.org/doc/org/articles/5typesopensourceprojects>
- <https://techbeacon.com/top-5-software-quality-metrics-matter-right-now>
- <https://diceus.com/top-7-software-quality-metrics-matter/>
- [https://en.wikipedia.org/wiki/Programming\\_complexity](https://en.wikipedia.org/wiki/Programming_complexity)
- [https://en.wikipedia.org/wiki/Maintainability#Software\\_engineering](https://en.wikipedia.org/wiki/Maintainability#Software_engineering)
- <https://github.com/fedir/ghstat>

# Used media ressources

- [https://commons.wikimedia.org/wiki/File:Big\\_%26\\_Small\\_Pumkins.JPG](https://commons.wikimedia.org/wiki/File:Big_%26_Small_Pumkins.JPG)
- [https://commons.wikimedia.org/wiki/File:Singapore\\_Road\\_Signs\\_-\\_Restrictive\\_Sign\\_-\\_Stop\\_-\\_Security\\_Check.svg](https://commons.wikimedia.org/wiki/File:Singapore_Road_Signs_-_Restrictive_Sign_-_Stop_-_Security_Check.svg)
- [https://commons.wikimedia.org/wiki/File:Green\\_bug.svg](https://commons.wikimedia.org/wiki/File:Green_bug.svg)
- [https://commons.wikimedia.org/wiki/File:Virgin\\_Voyage\\_-\\_Land\\_Rover\\_Celebrates\\_Production\\_of\\_First\\_New\\_Discovery\\_Sport\\_\(15572646535\).jpg](https://commons.wikimedia.org/wiki/File:Virgin_Voyage_-_Land_Rover_Celebrates_Production_of_First_New_Discovery_Sport_(15572646535).jpg)
- <https://pixabay.com/en/ferrari-formula-1-fernand-alonso-f1-490617/>
- <https://pixabay.com/en/social-media-marketing-seo-social-3216077/>
- [https://commons.wikimedia.org/wiki/File:Scrum\\_process.svg](https://commons.wikimedia.org/wiki/File:Scrum_process.svg)
- [https://fr.m.wikipedia.org/wiki/Fichier:Question\\_book-4.svg](https://fr.m.wikipedia.org/wiki/Fichier:Question_book-4.svg)
- <https://pixabay.com/en/user-male-happy-smiling-smile-37448/>
- <https://pixabay.com/en/writer-shadow-man-1129708/>
- <https://pixabay.com/en/community-friends-globe-continents-909149/>
- <https://www.goodfreephotos.com/vector-images/three-developers-character-set-vector-clipart.png.php>
- <https://svgsilh.com/image/459225.html>
- [https://www.flickr.com/photos/daniel\\_iverson/15090961835](https://www.flickr.com/photos/daniel_iverson/15090961835)
- <https://pixabay.com/en/facebook-analytics-graphs-2265786/>



# Thank You !



OPEN SOURCE  
EUROPE

THE LINUX FOUNDATION

OPEN SOURCE SUMMIT

Send me feedback

<https://fedir.github.io/feedback.html>



@FedirFR