Communicating the Value of Open Source Metrics

Ben Lloyd Pearson

Governor of GitHub @ Oath
Metrics are awesome!
But they aren’t inherently valuable
Focused metric sets can help solve specific problems

(Keep the end in mind)
And enable us to better communicate the value of open source efforts
We can work together to make this a reality
Speaker for the Dead
What can learn from the perspective of React Intl
A Slow Decline…

Commits

Time
A Slow Decline…

Commits

Time
Why Did this Project Die?

![Graph showing commits over time for Oath and Others.](image-url)
Why Did this Project Die?

Commits

Time

Organizations

Yahoo
7 yrs 4 mos

Engineer, Yahoo Presentation Technologies
Oct 2014 – Aug 2015 · 11 mos

Yahoo! Inc.
5 yrs 8 mos

Principal Software Engineer
Mar 2013 – Mar 2017 · 4 yrs 1 mo

Others
Why Did this Project Die?

Time

Commits

YAHOO!
7 yrs 4 mos

Yahoo
7 yrs 4 mos

Engineer
2 yrs 10 mos

Oath
7 yrs 4 mos

Oath
7 yrs 4 mos

Yahoo! Inc.
5 yrs 8 mos

Principal Software Engineer
Mar 2013 – Mar 2017
4 yrs 1 mo
Individual metrics can be misleading
We need a more holistic approach that provides better context
The goal: analytics dashboards that target specific problems
This was an Avoidable Death

1. Monitor the Development Pace
   a. Trend lines to identify long term changes
   b. Y-o-Y acceleration/deceleration to identify shorter term changes
This was an Avoidable Death

1. Monitor the Development Pace
   a. Trend lines to identify long term changes
   b. Y-o-Y acceleration/deceleration to identify shorter term changes

2. Evaluate the Contributor Distribution
   a. Commit distribution to identify sustainability of developer community
   b. Geographic distribution to identify developer communities
This was an Avoidable Death

1. Monitor the Development Pace
   a. Trend lines to identify long term changes
   b. Y-o-Y acceleration/deceleration to identify shorter term changes

2. Evaluate the Contributor Distribution
   a. Commit distribution to identify sustainability of developer community
   b. Geographic distribution to identify developer communities

3. Evaluate the Organization Distribution
   a. Contributions from other organizations to identify potential partners
   b. External code as percent of total to track success of external outreach
Real World Examples!

Moloch  Screwdriver
Development Pace
Monitor Development Pace

- **Y-o-Y Acceleration**
- **Cumulative**
Monitor Development Pace

Sudden acceleration in development pace, followed by sustained rate...

Y-o-Y Acceleration

Cumulative
Monitor Development Pace

Sudden acceleration in development pace, followed by sustained rate...

...results in a new development pace that raises the trend line
Monitor Development Pace

Y-o-Y Acceleration

Cumulative
Monitor Development Pace

A steady development pace...

Y-o-Y Acceleration

Cumulative
Monitor Development Pace

...despite regular fluctuations year to year.

A steady development pace...

Y-o-Y Acceleration

Cumulative
React Intl: Development Pace

Cumulative Commits Over Time

Year-Over-Year Percent Change in Commits

Commits

Git

386 # Commits
42 # Authors
1 # Repositories
React Intl: Development Pace

Consistent periods of decelerating development
React Intl: Development Pace

Consistent periods of decelerating development

Crossed below the trend line after more than 1 year above
React Intl - Development Pace

Consistent development deceleration across project history should have triggered alarms for further analysis

The cumulative commits dropping below the long term trend line should have triggered a much bigger alarm
Contributor Distribution
Author Distribution: Commits
Author Distribution: Commits

Extremely top heavy with a dependence on top two contributors
Author Distribution: Commits
Far more even distribution of commits per author
Author Distribution: Commits
Author Distribution: Timezone

Multiple contributors, but primarily in a single area.
Author Distribution: Timezone

- Multiple contributors, but primarily in a single area.
- Dominated by contributions from two people.
Author Distribution: Timezone

Numerous contributors in multiple locations
Author Distribution: Timezone

Numerous contributors in multiple locations

More distributed commits per author
React Intl: Contributor Distribution
React Intl: Contributor Distribution

Regular participation from community members
React Intl: Contributor Distribution

Top heavy author distribution

Regular participation from external community
React Intl: Contributor Distribution

Top heavy author distribution

Numerous contributors in diverse geographic locations!

Regular participation from external community
React Intl: Contributor Distribution

Top heavy author distribution

Numerous contributors in diverse geographic locations!

Dominated by contributions from two people

Regular participation from external community
React Intl: Contributor Distribution

React Intl has a history of contributions from a diverse developer community

However, most people never made more than 1 contribution
Organizational Distribution
Organization Commits

External contributors represent very small portion of total contributions.
Organizational Diversity Over Time

External contributions are decreasing as percent of total contributions.

Commits (% of Total)

Time
External contributors represent a **larger portion** of total contributions.
Organizational Diversity Over Time

External contributions are increasing as percent of total contributions.
React Intl: Organizational Diversity
React Intl: Organizational Diversity

External contributors represent a modest amount of total contributions.
React Intl: Organizational Diversity

External contributors represent a modest amount of total contributions

External contributions are increasing as percent of total contributions
React Intl: Organizational Diversity

React Intl had an increasing rate of external participation relative to internal participation.

However, it never reached critical mass to hit a sustainable level of external participation.
How to Save React Intl
What Works?

Numerous contributors in diverse geographic locations
What Works?

Numerous contributors in diverse geographic locations!

Regular participation from external community
What Works?

Numerous contributors in diverse geographic locations.

Increasing contributions as percent of total from external community.
How to Save React Intl

Make this

Commits

Timezone

Individual contributors
How to Save React Intl

Make this

Look more like this

Commits

Timezone

Authors

Individual contributors
How to Save React Intl

1. Evaluate the individual / organizational distributions to identify existing developer communities
How to Save React Intl

1. Evaluate the individual / organizational distributions to identify existing developer communities
2. Identify internal experts to bridge with external participants and cultivate a stronger community
How to Save React Intl

1. Evaluate the individual / organizational distributions to identify existing developer communities
2. Identify internal experts to bridge with external participants and cultivate a stronger community
3. Set alarms to notify proper authorities when certain health metric thresholds are crossed (for emergency use only!)
Three Steps for Rescue

1. Commits
2. Time
3. Contributions Per Author Over Time
Three Steps for Rescue

1. Reach out to first external contributors to build pathway to project leadership
Three Steps for Rescue

2. Leverage high internal development effort to increase involvement with external community
Three Steps for Rescue

Support the community as it attempts to take over the project leadership.

Commits

Time

3
The Plot Twist

There are still a ton of internal and external developers who use and want to contribute to this project
The Plot Twist

There are still a ton of internal and external developers who use and want to contribute to this project.

Maybe it’s not too late!
What Does this Mean for You?
Extracting More Value from Open Source Metrics

The Goal:

- Quickly gather vital information to better monitor projects for specific problems.
- Use the right data to better communicate how and where resources should be directed to better leverage open source software.
Extracting More Value from Open Source Metrics

Metric sets like these could be useful for a variety of individuals:

- Community Managers
- OSPO teams
- Development Managers
- Strategists
- Software Architects
- Etc.
Let’s Work Together!

I want to collaborate with the CHAOSS community to build more informative visualization sets that help us better communicate the value of our efforts.

I’ve published all of the dashboards from this talk at http://github.com/BenLloydPearson/grimoirelab-dashboards

P.S. I’m more than willing to grant this project to the CHAOSS community when it’s ready ;-)
Let’s Work Together!

I need help to make this awesome

● What is the best way to classify and organize these metric sets?
  ○ Profession, purpose, goal

● What other metric sets are useful for monitoring vital health metrics?
  ○ Development backlog management, dependence on individual developers

● What other metric sets could help identify and solve specific problems?
  ○ Building better organizational partnerships, identifying project risks

● How can we better combine data sources to produce more context?
  ○ Evaluating the effects of events on development pace, contributor distribution, organizational distribution.
Thanks!

http://github.com/BenLloydPearson/grimoirelab-dashboards
Check out our open source projects too!

- https://github.com/yahoo/react-intl
- https://github.com/aol/moloch
- https://github.com/screwdriver-cd