

Technology Group

Research and Data, Design Research,
Analytics Engineering,
Special Projects, Performance
FY Q3: January - March 2015/16

Quarterly review

Research and Data

Q3 - 2015/16

Approximate team size during this quarter:

5.5 FTE, 2 research fellows, 2 collaborators

Time spent: strengthen 40%, focus 50%, experiment 10%

Q3 - Research and Data

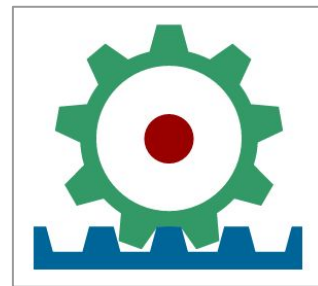
Objective: Revision scoring



Objective	Measure of success	Status
STRENGTHEN Extend revision scoring to support new models and languages <i>Team members involved: 2</i> <i>Collaborators: 3</i>	Develop and test <u>edit typing</u> model (T121715)	missed <ul style="list-style-type: none">• ran a 2nd pilot on enwiki• 1st pilot running on itwiki
	Tune damage detection model (T120138) and <u>extend</u> current models to 5 other Wikipedia language editions (T121712)	missed <ul style="list-style-type: none">• model tuning completed• 3 new languages deployed• 6 more in development
	Write and submit papers: KDD (T121718) CSCW (T121719)	KDD submission completed, CSCW on track for deadline.

Learning. productization and extension of models to other languages requires substantially more TechOps and Community Engagement support than we initially anticipated.

Revision scoring as a service



- **Performance**
 - sped up feature extraction by parallelizing it
 - profiling for expensive feature extraction
- **Productization**
 - ORES now running in the Beta cluster [<https://ores-beta.wmflabs.org>]
 - first service to adopt *scap3* for deployment (several upstream patches submitted)
- **User-facing features**
 - updated format (v2) and implemented API versioning scheme; deployed human-readable, Swagger-based documentation [<https://ores.wmflabs.org/v2/#/scoring>]
 - now showing features used when making a prediction [[example](#)]
 - users can now inject features to experiment with model behavior [[example](#)]
 - machine-readable model statistics now reflect KPIs, e.g.: "What proportion of edits can we confidently *not* review?" `filter_rate_at_recall(0.90)` [[example](#)]

Q3 - Research and Data

Objective: Article recommendations



Objective	Measure of success	Status
<p><i>EXPERIMENT</i></p> <p>Run and analyze a content creation campaign</p> <p><i>Team members involved: 3</i> <i>Collaborators: 2</i></p>	<p>Run an <u>article creation</u> campaign in the wild, teaming up with volunteers to improve the design and performance of recommendations. (<u>T121962</u>)</p>	<p>missed</p> <ul style="list-style-type: none">• Focused on design and usability improvements to the <u>tool</u> and postponed launch of a campaign


Acknowledgments. We are grateful to the Editing team for providing design support.

<http://recommend.wmflabs.org>

Wikipedia **Suggestions**


ไทย ▾ English ▾

Seed article (optional)




ธรรมศาสตร์ ซี

3185 recent views



พันท้ายนรสิงห์

3132 recent views



ผักบุ้งกับกุ้งนาง

2208 recent views

Q3 - Research and Data

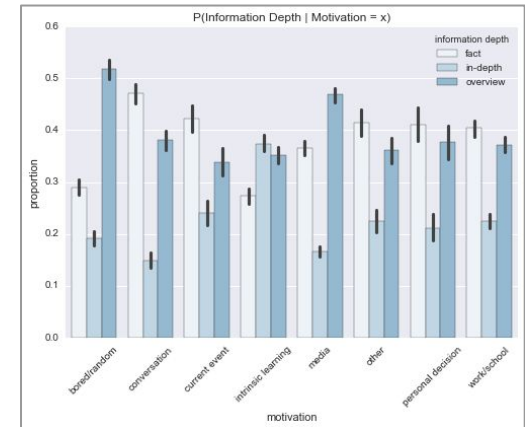
Objective: Reader segmentation



Objective	Measure of success	Status
FOCUS Reader segmentation research <i>Team members involved: 3</i> <i>Collaborators: 2</i>	Run a reader survey integrated with log analysis to identify segments in the Wikipedia reader population and their reading behavior. Analyze and present the <u>results</u> (T121727)	completed Work will be extended with new analysis in Q4

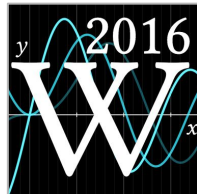
- ran 2 pilots in Spanish and Farsi and a full-scale survey in English
- analyzed the results and presented early findings to the Reading team [meta.wikimedia.org]
- we aim to conduct additional research in Q4 on data we collected and present the results more broadly

Acknowledgments. We are grateful to the Reading team for providing engineering support for this research.



Other achievements

- Hosted 2 sessions at the Wikimedia Dev Summit
- Held our first **team offsite** to draft our FY17 annual plan
- **Referrer policy** launched: recovering [Wikipedia dark traffic](#) over HTTPS
- Organized a joint **wiki workshop** to be hosted in Montreal and Cologne in Q4 [snap.stanford.edu]
- Organized (and secured funding) for **WikiCite** event in Berlin in Q4 [meta.wikimedia.org]



[Wiki Workshop 2016](#)



[WikiCite 2016](#)

Q3 - Research and Data

Category	Workflow	Comments	Type
	NDA's / endorsements	Renewed 3 NDA's for research collaborators	M
	Wikistats maintenance	Provided continuous support to the Analytics team for Wikistats, to be phased out in Q4	M
	Showcases and talks	Hosted 2 showcases and one brown bag presentation	M

Type: new, reactive, maintenance

- [Research & Data team page](#)
 - Describing goals, processes and projects.
- [Goals for Q4 FY16](#)
 - What we are planning to do in the coming quarter
- [FY16 priorities](#)
 - Top priorities for the fiscal year
- [Phabricator workboard](#)
 - What we are currently doing

Quarterly review

Design Research

Q3 - 2015/16

Approximate team size during this quarter: ... 4 FTE
Time spent: strengthen 40%, focus 50%, experiment 10%

Q3 - Design Research

Objective	Measure of success	Status
<p>FOCUS: Production Work Provide generative and evaluative design research in support of human centered product development.</p> <ul style="list-style-type: none">• Evaluative research when needed by teams• Generative interviews on going	<p>Generative</p> <ul style="list-style-type: none">• Style Guide/ UI Standardization exploratory research <p>Evaluative</p> <ul style="list-style-type: none">• iOS workflows testing (App)• Content Translation campaigns user testing• Notifications panel evaluation• VE: education for new editors testing	<p>All projects listed are completed and have been taken in and utilized by the teams they were done with/for.</p>

Q3 - Design Research

Objective	Measure of success	Status
<p>FOCUS: Production Work</p> <p>Provide generative and evaluative design research in support of human centered product development.</p> <p>Contextual Inquiry in Mexico</p>	<p>Contextual inquiry in Mexico completed: 15, 90 minute interviews, metrics presentation (starts at 20:30)</p>	<p>Mexico research completed, analysis and high level findings have been delivered. In Q4 there will be a brown bag to share deeper findings that could not be shared in a 15 minute metrics presentation.</p>

Learning: Contextual Inquiry takes a lot of planning and alignment before going into the field. It is important to have the time to do that planning. It is also important to be able to pivot when unexpected things happen to make the best of a not ideal situation so it is possible to still come back home with findings.

Q3 - Design Research

Objective	Measure of success	Status
<p>Strengthen: Mentoring</p> <p>Mentoring non design researchers to do quality design research.</p> <ul style="list-style-type: none">● Sherah Smith in Reading● UI Standardization team● Program Toolkits Team	<p>Sherah:</p> <ul style="list-style-type: none">● iOS workflows research in collaboration with Daisy● Started the prototyping labs (pilot and iOS workflows for the Reading team). <p>UI standardization team</p> <ul style="list-style-type: none">● Focus groups (in collaboration with Jonathan): research to inform the needs of community and non technical users. <p>CE Jaime Anstee and Maria Cruz, Subha Panigrahi and Edward Galvez</p> <ul style="list-style-type: none">● Series of interviews to learn about program tool kits and the needs of program, edita-a-thon, photo contest leaders	<p>2 teams collaborated with</p> <p>1 non researcher becoming a researcher.</p>

Q3 - Design Research

Objective	Measure of success	Status
<p>Strengthen: Personas Continue to iterate our set of pragmatic personas into personas informed by qualitative research.</p>	<p>Interviews were completed for Readers, but we put off analysis for Q4, because of the time we spent in Mexico.</p>	<p>We have a plan to finish analysis of, and iterate Reader and New Editor personas by end of Q4</p>

Q3 - Design Research

Objective: External Collaborations



Objective	Measure of success	Status
<p>Experiment: External Collaborations</p> <p>External Collaboration with University of Washington: Better understand university students' information-seeking behaviors. (T120760)</p>	<p>Survey has been sent out to students, ~200 responses collected</p>	<p>Ongoing successfully, expected completion June 2016.</p>

Learning: Becoming an official partner of a UW department opens up research opportunities. Collaboration so far has been low-cost and productive. Results will be relevant to multiple product verticals.

Q3 - Design Research

Other successes and misses

Other Successes:

Beginning **collaboration with Reboot** for our next contextual inquiries to learn about new readers in Nigeria (May) and India (June).

We held a **Research Team off site** (Design Research and Research and Data):

- Team bonding
- Aligning around the research we do
- Found spaces for collaboration between DR and R&D
- Planning for FY 16/17

Q3 - Design Research

Category	Workflow	Comments	Type
	Participant Recruiting	31 participants recruited for research sessions this quarter	M
	Talks and Workshops	Monthly Metrics (March) presentation: “Mexico Deep Dive” CSCW workshop on “Breaking into new data spaces” https://meta.wikimedia.org/wiki/Research:Breaking_into_new_Data-Spaces	M

- [Design Research Team Page](#)
 - Describing goals, processes and projects.
- [Goals for Q4 FY16](#)
 - Plans for Q4
- [FY16 priorities](#)
 - Top priorities for the fiscal year
- [Phabricator workboard](#)
 - What we are currently doing

Quarterly review

Analytics Engineering

Q3 - 2015/16

Approximate team size during this quarter: 6FTE (2 of them devops), 1 almost FT manager

Key performance indicator: **Velocity**

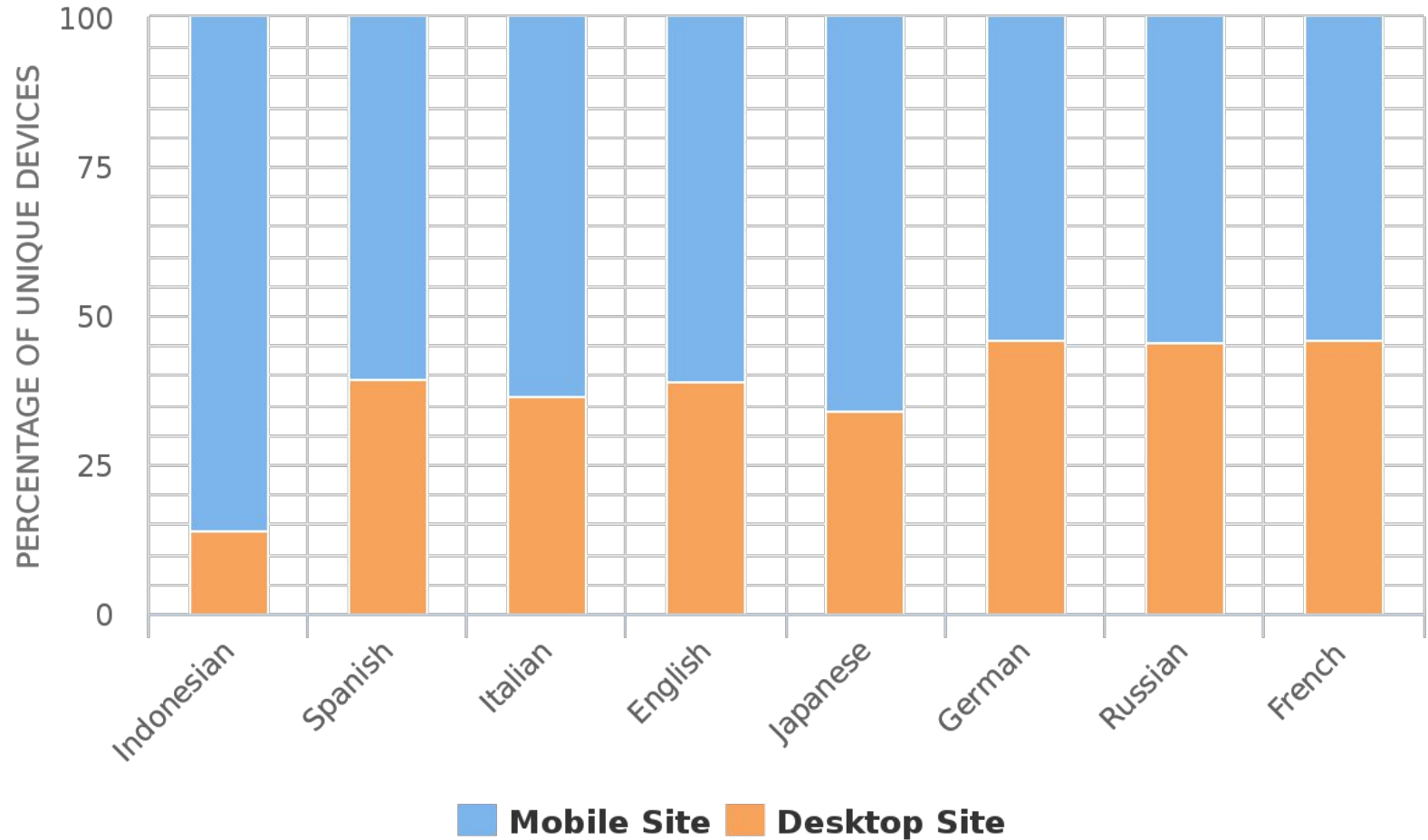
Quarter: 871 (~910 last quarter)	234 (-10%MoM)	315 (+26%MoM)	322 (+3% MoM)
----------------------------------	---------------	---------------	---------------

Q3 - Analytics Engineering

Objective	Measure of success	Status
<p>Calculate Unique Devices for all wiki projects monthly and daily using a privacy conscious methodology.</p> <p><i>Team members involved: 3 plus collaboration from research and ops</i></p>	<p>Besides delivery of a long awaited feature that gives us insight to the reach of wikimedia projects...</p> <p>Being able to compute numbers internally per country.</p> <p>Write blogpost on our calculation methodology.</p>	<p>Data available both internally in hive and externally via downloads.</p> <p>Country data only for internal use.</p>

Learning: Took the better part of a year to deliver this project, the hardest part was to remove the effect of bot traffic in our calculations, an issue we did not anticipate.

ACCESS TO CONTENT IN WIKIPEDIA



Q3 - Analytics Engineering

Objective: Wikistats 2.0

Objective	Measure of success	Status
<p>Use Hadoop infrastructure to replace computation of Wikistats browser reports</p> <p><i>Team members involved: 3 plus Erik Z. from Research</i></p>	<p>Usage of reports by WMF and community.</p> <p>Data used in Fundraising and bug triage decisions.</p>	<p>Data and visualization available internally and externally at: https://browser-reports.wmflabs.org</p> <p>We will be doing UI tweaks and provisioning a prod domain to deploy tool and announce externally.</p>

Learning: We knew this but here it is again... the UI is NEVER finished. We need a designer to be more efficient in UI projects.

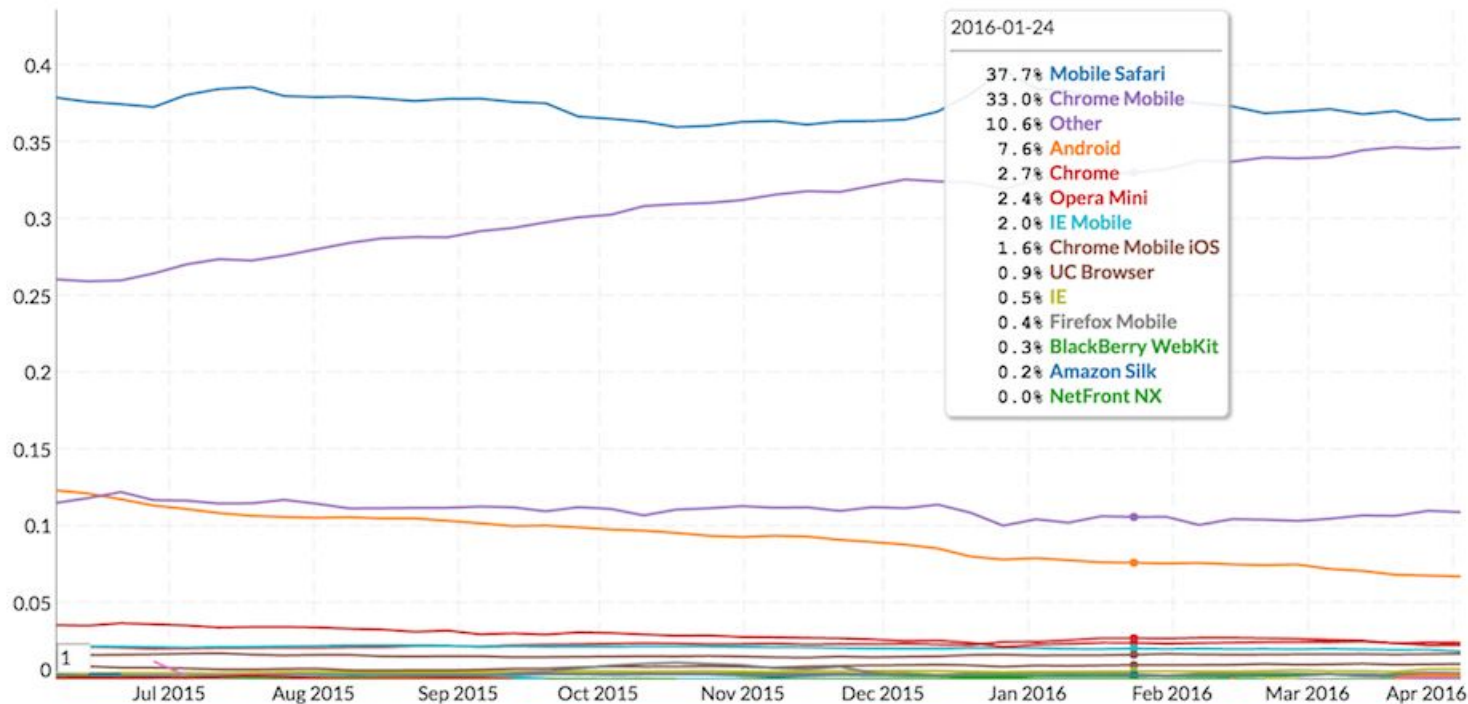
User Agent Breakdowns Wikimedia Foundation

All Sites by OS ▾ All Sites by Browser ▾ All Sites by OS and Browser ▾ Mobile Site by OS ▾ **Mobile Site by Browser ▾**

Browser Family Timeseries

filter by date 

- Amazon Silk
- Android
- BlackBerry WebKit
- Chrome
- Chrome Mobile
- Chrome Mobile iOS
- Edge Mobile
- Firefox Mobile
- IE
- IE Mobile
- Mobile Safari



Q3 - Analytics Engineering

Objective: **Sanitization**

Objective	Measure of success	Status
<p>Sanitization of Pageview Data</p> <p><i>Team members involved: 2</i></p>	<p>Dataset cannot be used for identity reconstruction. Enough information is kept for data to be useful.</p>	<p>Review of methodology done by research, still pending security review.</p>

Learning: Sanitization is a hard problem computationally, if we want to commit to release all our data publically we are going to have to devote some resources in this area.

Q3 - Analytics Engineering

Objective: Operational Excellence

Objective	Measure of success	Status
<p>Operational Excellence</p> <p>Pageview API</p> <p>EventLogging</p> <p>Cluster</p> <p><i>Team members involved: 1\2 always FT plus 1 as need arises.</i></p>	<p>Reduce outages on Eventlogging (Q2 versus Q3)</p> <p>Sustain traffic on pageview API with good response times.</p>	<p>Doing pretty well on all fronts after having hired a second ops person. Better alarms on Kafka and Eventlogging.</p> <p>Pageview API doing well but we need SSD hardware to do better</p>


Learning: Our effort in ops is paying off. Cluster needs some love when it comes to deployment. Beta Eventlogging storage on MySQL is the most brittle part of the system, like other teams we suffer from having too few DBA resources.

Q3 - Analytics Engineering





Other successes and misses 

Piwik, small site analytics stats available for website usage for transparency report and wikipedia15.

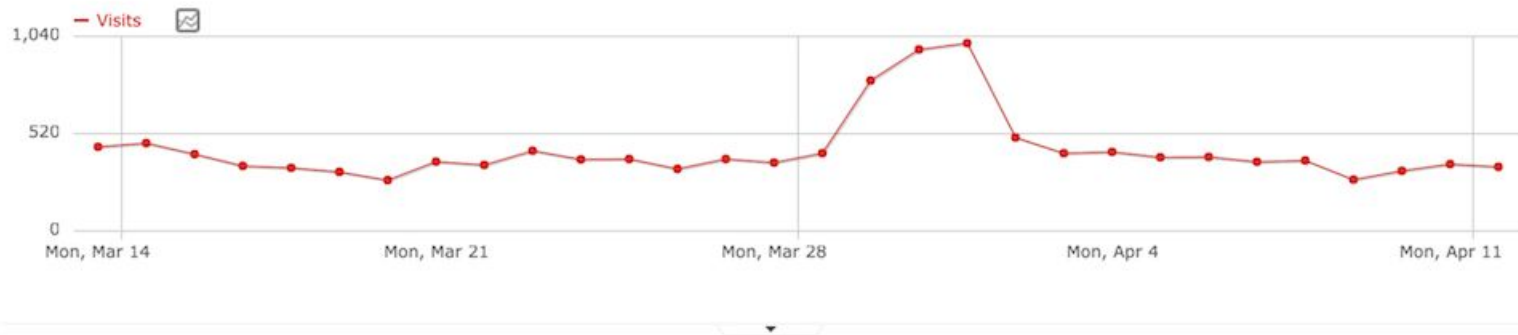
WEBSITE: WIKIPEDIA15 ▾

2016-04-12 

ALL VISITS 

-  **Dashboard**
-  **Visitors**
- Overview
- Visitor Log
- Custom Variables
- Devices
- Software
- Locations
- Engagement
- Times
- Real-time Map
-  **Actions**
-  **Referrers**

Evolution over the period



Report

 339 visits, 309 unique visitors	 532 pageviews, 426 unique pageviews
 1 min 14s average visit duration	 0 total searches on your website, 0 unique keywords
 81% visits have bounced (left the website after one page)	 0 downloads, 0 unique downloads
	 24 outlinks, 20 unique outlinks

EventBus

, collaboration with services team, mediawiki publishing edit events, we hope to use those next quarter to bootstrap our edit data pipeline in hdfs.

Public By Default

, removed IP capturing from EventLogging so data we capture has as little Personally Identifiable Information as possible.

Quarterly review

SPECIAL PROJECTS

Q3 - 2015/16

Approximate team size during this quarter: 1 FTE

Time spent: experiment 100%

Q3 - Special Projects

Objective: <Begin a Community Consultation>



Objective	Measure of success	Status
Consult with WMF and the Community on Engagement Metrics <i>Team members involved: 7</i>	<ul style="list-style-type: none">• Host internal brown bag• Run external survey• Begin community consultation	Postponed: to be revisited in 6 months to a year depending on organizational needs and executive direction.

What was accomplished:

- **Shed light on engagement metrics and unique tokens** by hosting internal discussions via a brown bag lunch, office wiki talk pages, and meetings with stakeholders.
- Achieved consensus to **not use unique tokens** at this time.
- **Avoided** a potentially costly **community consultation**.

Learnings

- **Unique device counts** using Last-Access tokens **satisfies an immediate need** for more engagement metrics and it will take time to fully interpret and integrate this new metric into the Foundation's operations.
- Other engagement metrics such as sessions may be **computed using other technologies**, such as EventLogging.
- We would need significantly **more internal support for unique tokens** before consulting with the community.
- Unique tokens opens up additional **privacy risks**, but does provide some **data retention advantages**.

Quarterly review

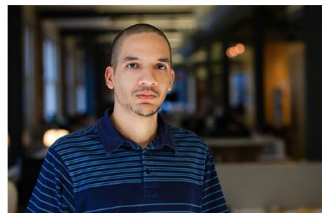
PERFORMANCE TEAM

Q3 - 2015/16

Approximate team size during this quarter: 5 FTE



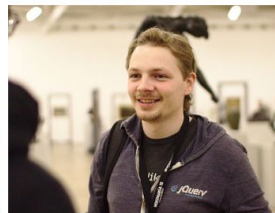
Ori Livneh



Aaron Schulz



Gilles Dubuc



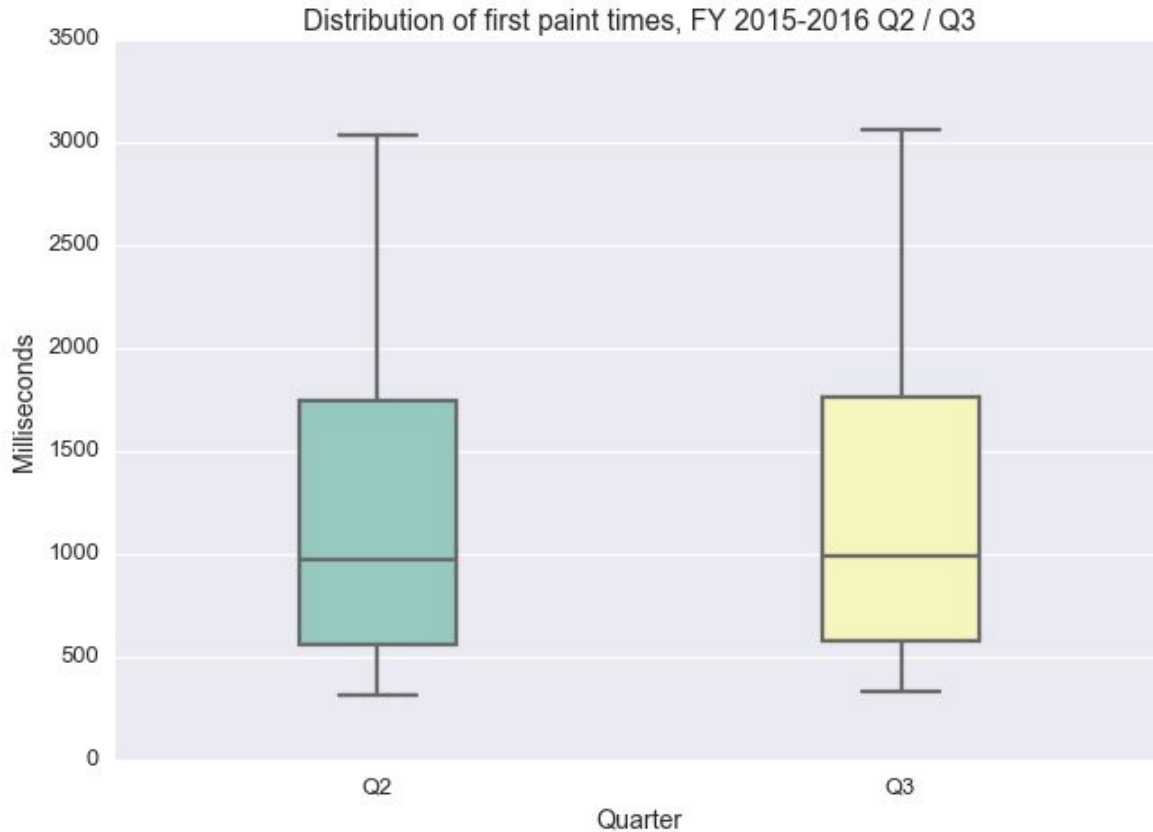
Timo Tijhof



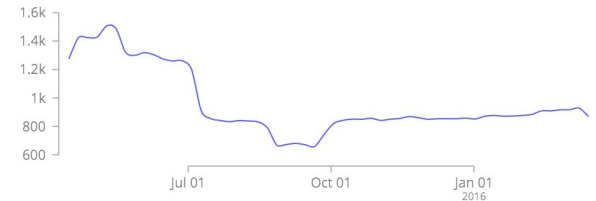
Peter Hedenskog

KPI: First paint time

Time in milliseconds it takes before user sees any content.



%ile	Q2	Q3	+/- %
90 th	3,033	3,063	+0.98%
75 th	1,744	1,761	+0.97%
50 th	970	987	+1.72%
25 th	558	575	+2.96%
10 th	316	330	+4.24%

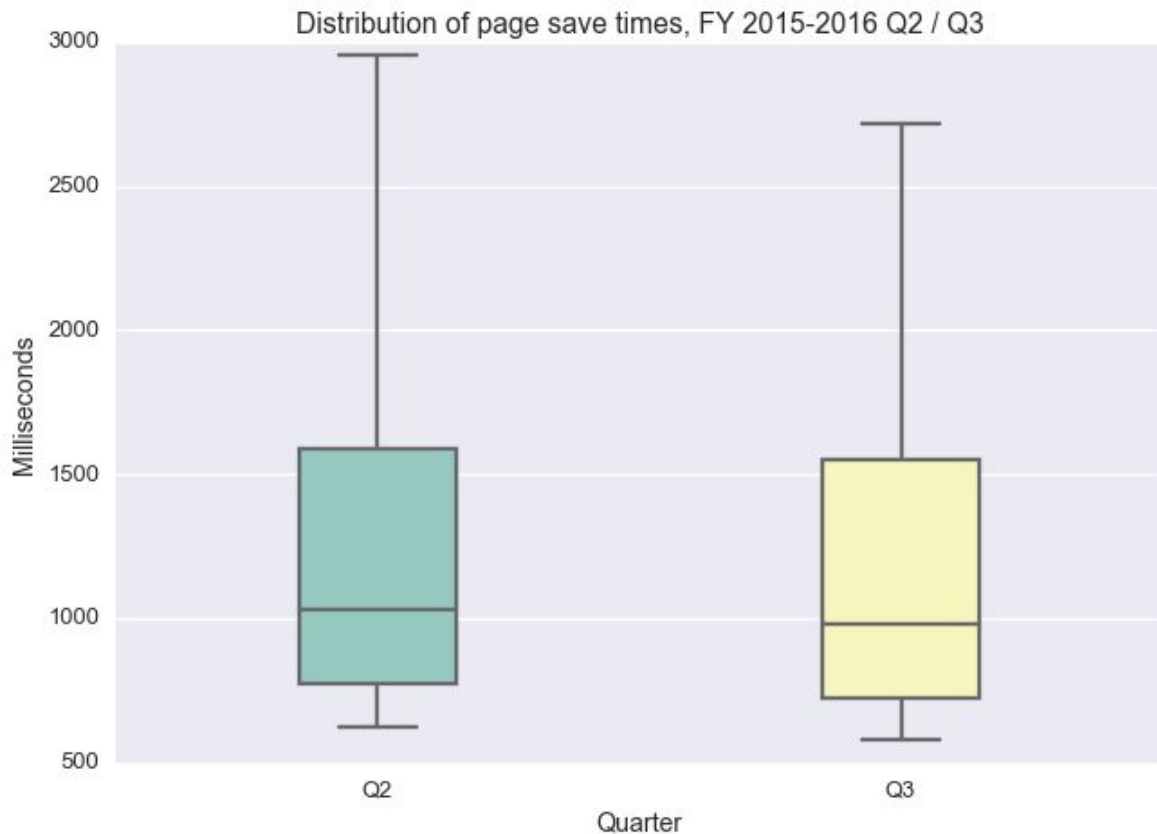


SPDY usage vs. time to first paint, 1-Dec-2015 - 10-Apr-2016

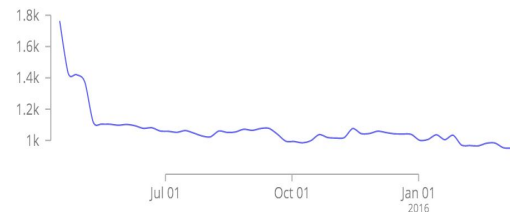


KPI: Page save time

Time in milliseconds between the user clicking **[Save]** and the edited article starting to load.



%ile	Q2	Q3	+/- %
90 th	2,953	2,718	-8.65%
75 th	1,590	1,550	-2.58%
50 th	1,032	983	-4.98%
25 th	772	725	-6.48%
10 th	623	580	-7.41%



Q3 - Performance

Objective: Performance inspector (T117411)



Objective	Measure of success	Status
Performance inspector tool <i>Team members involved: 1</i>	<ul style="list-style-type: none">Share performance data with editors, so they know how the article they are on will load on different connection profiles.Make performance team obsolete.	<ul style="list-style-type: none">Completed initial prototype.Team not yet obsolete :(Carried over to Q4.

The screenshot shows a MediaWiki page for 'Figure of Speech'. The page content includes a definition of a figure of speech or rhetorical figure as figurative language, listing examples like simile, metaphor, hyperbole, personification, and synecdoche. A featured image shows a 'WHITEHALL SW1 CITY OF WESTMINSTER' sign. The sidebar on the left contains navigation links such as 'Main page', 'Recent changes', and 'Tools'. The top navigation bar includes 'Admin', 'Talk', 'Preferences', 'Watchlist', 'Contributions', and 'Log out'. A search box is visible in the top right.

Q3 - Performance

Objective: High-availability for MediaWiki



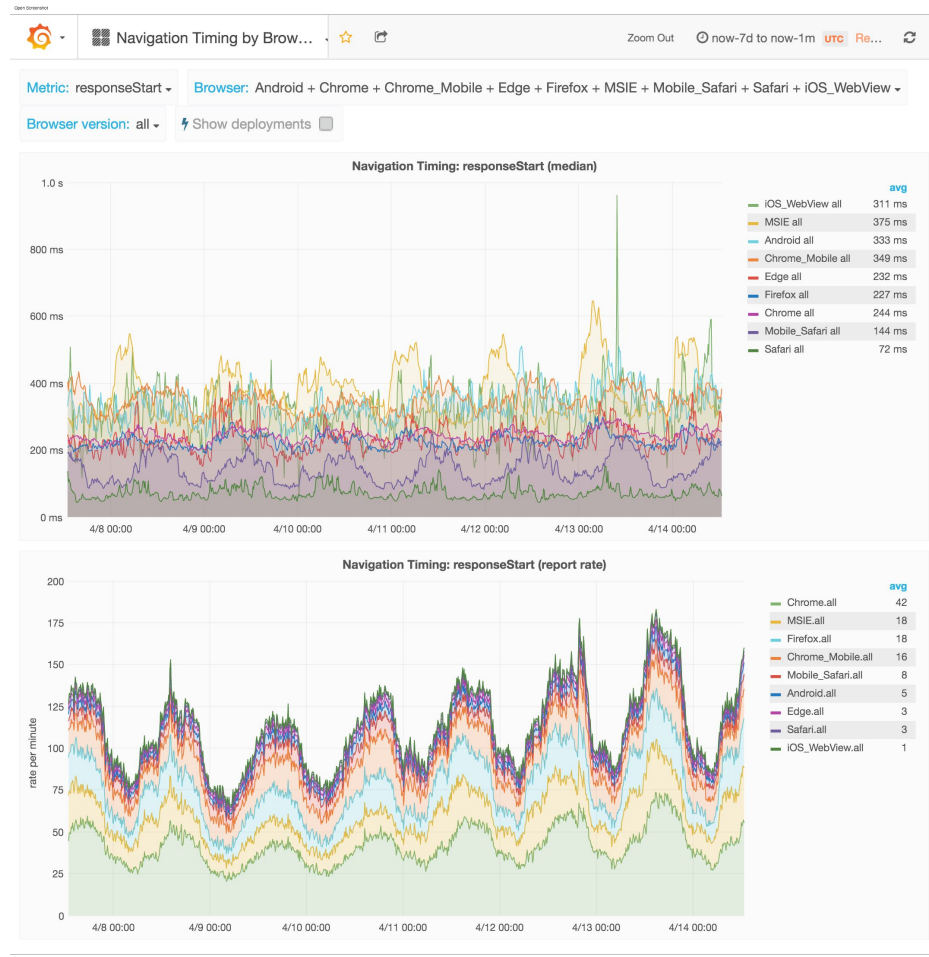
Objective	Measure of success	Status
<p>Make MediaWiki resilient to failure in our primary DC.</p> <p><i>Team members involved: 1</i></p>	<p>Successful switchover to CODFW.</p>	<ul style="list-style-type: none">• Read-only test completed successfully on March 15.• CODFW switchover scheduled for April 19.

Objective: Leaner mobile web



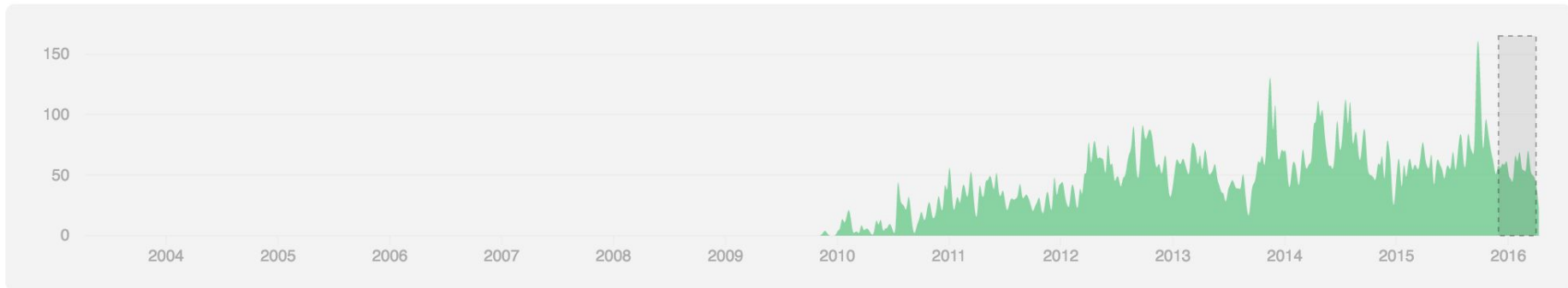
Objective	Measure of success	Status
<p>Make the mobile web site leaner.</p> <p><i>Team members involved: 2</i></p>	<p>Segment the mobile web site into a high-bandwidth and a low-bandwidth variant. Route readers to the best variant based on the connection speed.</p>	<p>Feedback from designers was that high-pixel-density images did not provide enough value even on fast connections. Disabled globally.</p>

Granular Performance Dashboards



Jan 1, 2016 – Apr 1, 2016

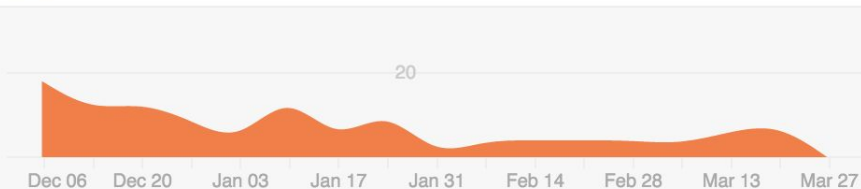
Contributions to master, excluding merge commits



AaronSchulz

118 commits / 2,961 ++ / 2,329 --

#1



Krinkle

85 commits / 9,789 ++ / 9,773 --

#3



Problems and Prospects

- The model of planning each quarter as a small set of discrete projects is a poor fit for the team:
 - We follow the numbers.
 - We respond to failures.
 - We provide guidance.
 - Team members motivated and attracted by low process overhead and local autonomy.
- Slides represent a small fraction of team activity.
- Effective reporting is still a challenge.
- In the future: improve delivery and resilience.
- Still missing: impact on editor engagement.
- Lack of clarity on global delivery strategy for logged-in users.

Granular Performance Dashboards

<https://grafana.wikimedia.org/dashboard/db/navigation-timing-by-geolocation>