Technology Tuning Session Q4 FY19-20

WIKIMEDIA
FOUNDATION

Platform Evolution

We need to have a modernized, instrumented, and efficient platform to enable next generation engagement across the globe. As we engage, we need to incorporate next generation approaches around artificial intelligence to secure and enhance trustworthy content, while ensuring the safety for all our curators and readers.

Q4 Highlight Win: Better Cache Management to Address COVID Traffic Increases

A day in the life of our cache:
10B requests
90+% cache hit ratio
1.5 M edits
110M+ cache purges

- 50% reduction in purges
- Reliably delivered, easier to maintain
- Cross-team effort, zero downtime

Drill Down: Platform Evolution

The situation

We do not yet have full working definitions for updated MTP metrics on structured data and non-text, but we do have baseline counts from which metrics can be defined. We now also have two quarters worth of data so we expect to be able to make predictions on this soon.

We have also made progress on performance issues with Wikidata Query Service (WDQS). We have a working end-to-end solution for our simpler use case (updates) and are working to solve the more complex cases (deletions, visibility). Early indications on performance and reliability are positive.

The impact

For the first time ever, we now have measurements of the usage of Wikidata and Commons data in other projects! However, there is a small knowledge gap on how we *assess the impact* of this type of data.

As stated in our Q2 and Q3 tuning session, Wikidata and the Wikidata Query Service are reaching the limits of their current architecture.

We held our first ever department level Quarter-in-Review session (inspired by Product)

MTP Priority slides

Platform Evolution

Overview

- Introducing more structured data and leveraging machine learning is unlocking new capabilities in our products to support our communities.

- Improving our code quality, automation, and developer tooling increases our capacity to innovate, experiment, learn and deliver.

- Supporting our diverse technical communities results in growth and enables solutions that engage with our content and data, and leverages Wikimedia data beyond the core wiki experience.

Progress and Challenges

- We will shift the focus of new and existing programs to better recognize and support the vibrant ecosystem of technical contributors and contributions beyond MediaWiki.

- This year's developer satisfaction survey has informed our planning for FY20/21, including the evaluation of tooling for Continuous Integration (CI) and code review.

Key Deliverables

| Deliverable	Status
Content Integrity	🟡
Machine Learning (ML) Infrastructure	🟡
Technology and Product Partnerships	🟢
Improve developer productivity and efficiency to accelerate innovation	🔴

Actions

- Define growth rates for our metrics on structured data and media usage as well as Artificial Intelligence (AI) impact
- Focus on growth of tool maintainers - move reporting on independent patch submitters to department level OKRs
- Explore the correlation of contributions by community-built bots and growth of a project

Department: Technology

2019-10-01 - 2020-06-27 📅

- ☐ plwikivoyage
- ☐ plwiktionary
- ☐ pmswiki
- ☐ pmswikisource
- ☑ pnbwiki
- ☐ pnbwiktionary
- ☐ pntwiki
- ☐ pswiki
- ☐ pswikibooks
- ☐ pswikivoyage

2020-01-01

22.4% pnbwiki
1.1% pawiki
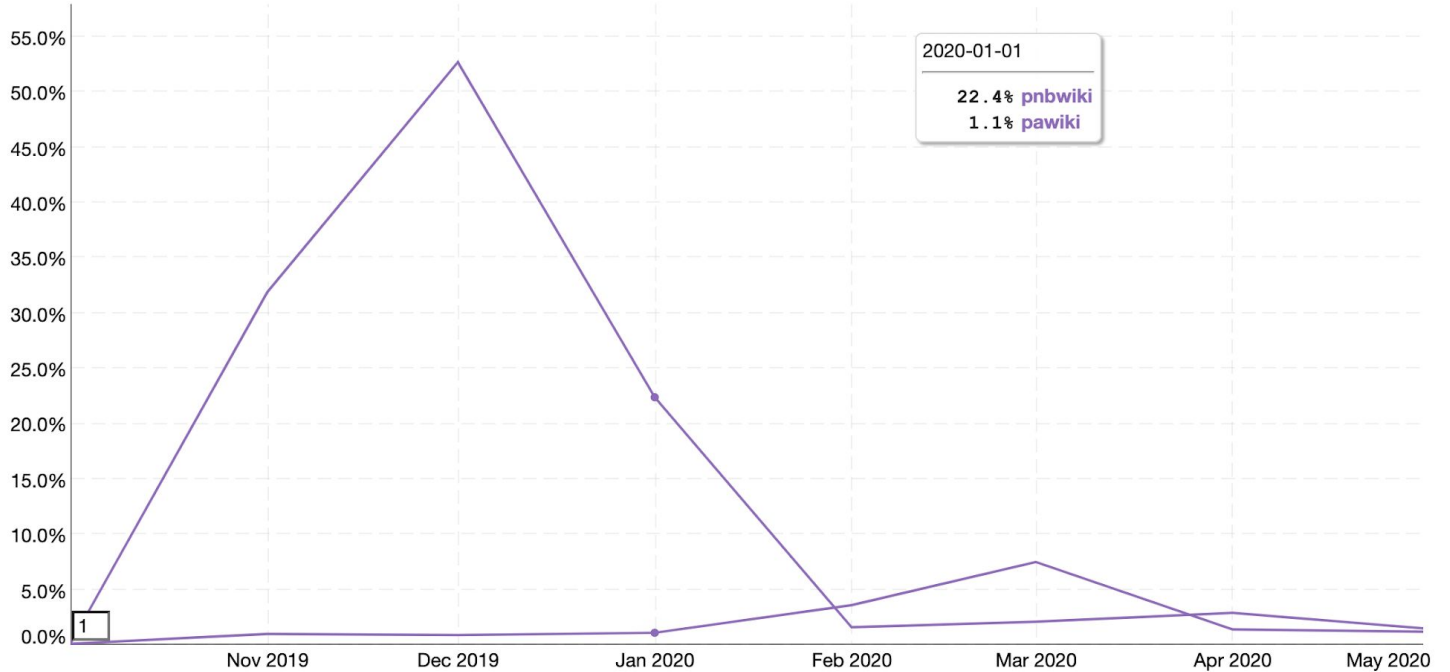
About 1/3 of the total edits to the Wikimedia projects come from tools and bots. The screenshot shows the numbers for Western Punjabi Wikipedia and Punjabi Wikipedia from the past months.

Platform Evolution Metrics

MTP Outcomes	MTP Metrics	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
We will build tooling for internal and external development and reuse of code and content	An X% increase in structured data used (uptake) across wikis. **Baseline:** 24.9% (59.5M/238.9M) of pages across Wikimedia projects use Wikidata or Structured Data on Commons in April 2020. (Note: Commons Data currently cannot be used in other Wikimedia Projects)	X% (TBD)		Changed (see appendix)	24.9% (59.5/238.9M)	**38.5% (97/251.8M)**
	An X% increase in non-text (e.g. Commons) content used across wikis. **Baseline:** 52.9% (31.2M/60M) of Commons is used across WMF projects in April 2020.	X% (TBD)		Changed (see appendix)	52.9% (31.2M of 60M files)	**52.1% (32.2M of 61.8M files)**

Platform Evolution

MTP Outcomes	MTP Metrics	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
A secure and sustainable platform that empowers a thriving developer community with the ease of software-as-a-service tooling	X% increase of independent developers who submit patches to production code **Baseline:** 146 independent devs active in Q1	X% (TBD)			Determining growth rate	**169 active independent developers**
	10% decrease in code review time **Baseline:** 19 days in June 2019	2% (18 days)	Setting baseline	-10% (21 days)	10% (17 days) Monitoring	**352% (86 days)**
	30% increase of tool maintainers **Baseline:** 1880 maintainers in Q2	5% (1974)		Setting baseline	2% (1919)	**5.5% (1984)**
	10% (4.2 / 5) increase in developer satisfaction **Baseline:** 2019 developer satisfaction: 3.8 / 5	4% (3.9)	Awaiting launch of survey	Survey after All Hands	-8% (3.4)	**No change (Qualitative analysis published)**
	20% decrease in outstanding code reviews **Baseline:** 1134 code reviews in June 2019	X% (TBD)	61% (442)	47% (601)	-1.4% (1151)	**-1.9% (1357)**
	25% increase in code quality **Baseline:** 0% in June 2019	5%	<1%	4.76%	11.9% Revisit & Recommend	**14.29%**

Drill Down: Platform Evolution

The situation

We expect the number of independent contributors to MediaWiki and related services to stay stable and see growth and great potential in the broad tooling ecosystem.

Emerging communities often lack people, skills and tooling to automate cumbersome workflows.

The impact

The focus on the ecosystem of technical contributions changes the way how we think about entry-points, supporting and growing our technical community, and opens the door for new community building initiatives.

Editors of smaller wikis struggle to keep up with content curation.

Recommendation

Strengthen services for tooling beyond MediaWiki, **focus** on the tech community as an ecosystem; **experiment** with outreach to tech communities outside of WMF.

Explore correlation of community-driven bot development and article size to to get "ahead of the problem" (is there a tipping point?).
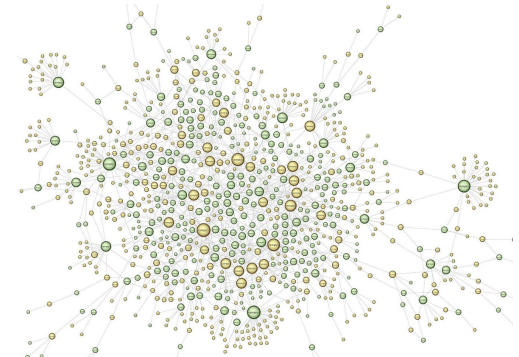
Content Integrity

Objective:

Secure and protect the platform and our communities in the free knowledge movement against the spread of disinformation and bad-actor risk

We build the **infrastructure** to assure the security of content as well as content contributors and consumers.

We build the **technologies** that empower the editor and patroller communities enforce content policies such as *verifiability* and *neutral point of view* more effectively.

Target quarter for completion: Q4 FY19/20

The OrangeMoody network of socks (ongoing)

Department: Technology

Content Integrity

Key Results	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
Create 2 security *governance* services: risk management and security awareness Create 2 security *engineering* services: application security and privacy engineering **Baseline:** 0	4	1	3	3	**4**
Develop a means to limit and disable the API access of bad actors without interrupting the access of other contributors, integrate it into our platform, and measure the effect **Baseline:** 0%	100%	20%	33%	50%	**80%**

Department: Technology

Content Integrity

Key Results	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
Build 2 sets of Formal Collaborations to expand our capacity for working on prioritized disinformation projects (by the end of Q3) **Baseline:** 0	2	n/a	In progress	1	**2**
Build a test model to address a specific type of disinformation (by the end of Q4) **Baseline:** 0	1	n/a	In progress	In progress	**3/4ths complete**

Department: Technology

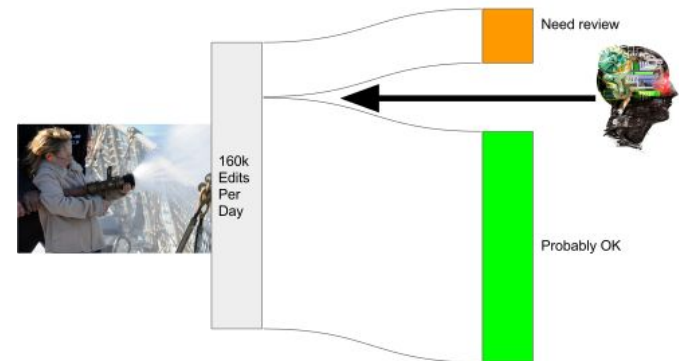# Content Integrity

Objective:

Expand quality control Artificial Intelligence (AI) tooling to underserved communities, to make fundamental services available for consumption by those tools that support the growth of high quality content and the maintenance of quality standards

Portuguese Wikipedia volunteers can now measure the progress of their work and to support review of new articles.

Ukrainian Wikipedia volunteers can now use models to fight vandalism and manage content issues.

Improved English Wikipedia's article quality model.

Target quarter for completion: Q4 FY19/20

Need review

160k
Edits
Per
Day

Probably OK

Content Integrity

Key Results	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
Deploy 8 new quality control AI tools to Wikimedia Projects, increasing availability of AIs for tooling **Baseline:** 0	8	4	6	11	**15**
Recruit 4 new campaign coordinators to advertise the availability of the new AI tools, increasing rate of consumption of those tools **Baseline:** 0	4	0	Blocked	2	**4**
Improve 3 AI tools in a statistically significant way based on community feedback, ensuring utility of AIs for quality control work **Baseline:** ongoing	3	4	7	9	**12**

Department: Technology

Machine Learning Infrastructure ⊕

Objective:

Enable our communities to more easily detect the hidden algorithmic biases in current Machine Learning (ML) solutions

We have completed building out Jade MVP functionality.

Last quarter, we were slowed by the complexity of building front-end components in MediaWiki *(shout out to FAWG!)*

We have recruited collaborators from 2 wikis for our pilot deployment.

Pilot deployment is likely to go out in the beginning of Q1 FY20/21

Target quarter for completion: Q4 FY20/21

Edit quality represents the productivity and intent of an edit.
describes the quality and the apparent intent of the edit.

⦾ ✅ **productive** 💙 **goodfaith**

This edit was productive and was saved with good inten

⦾ ❗ **damaging** 💙 **goodfaith**

This edit was unproductive but it appears as though it w

⦾ ❗ **damaging** 🕵 **badfaith**

This edit is vandalism. It is unproductive and it appears t

Notes

(optional) Summarize why you chose this label. Consider d

Machine Learning Infrastructure

Key Results	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
Build, deploy, and establish baseline metrics for infrastructure that enables Wikipedians to correct the algorithmic predictions around quality of content to 4 wikis **Baseline:** 0%	4	33%	50% of the MVP complete	90% of the MVP complete	**100% of MVP complete**
Increase the rate of community-based false-positive reporting in damage detection models by 100 times **Baseline:** 1 report per day	100%	0%	0% Blocked on deployment of the MVP	0% Blocked on deployment of the MVP	**0% MVP in production but measurement pending**

Drill Down: ML Infrastructure

The situation

Front-end engineering in MediaWiki was a bigger hurdle than we expected.

Still, we managed to get the MVP functionality together and we're actively consulting with wiki communities about deployments

The impact

Jade is poised to dramatically increase our ability to train and evaluation machine learning models in production.
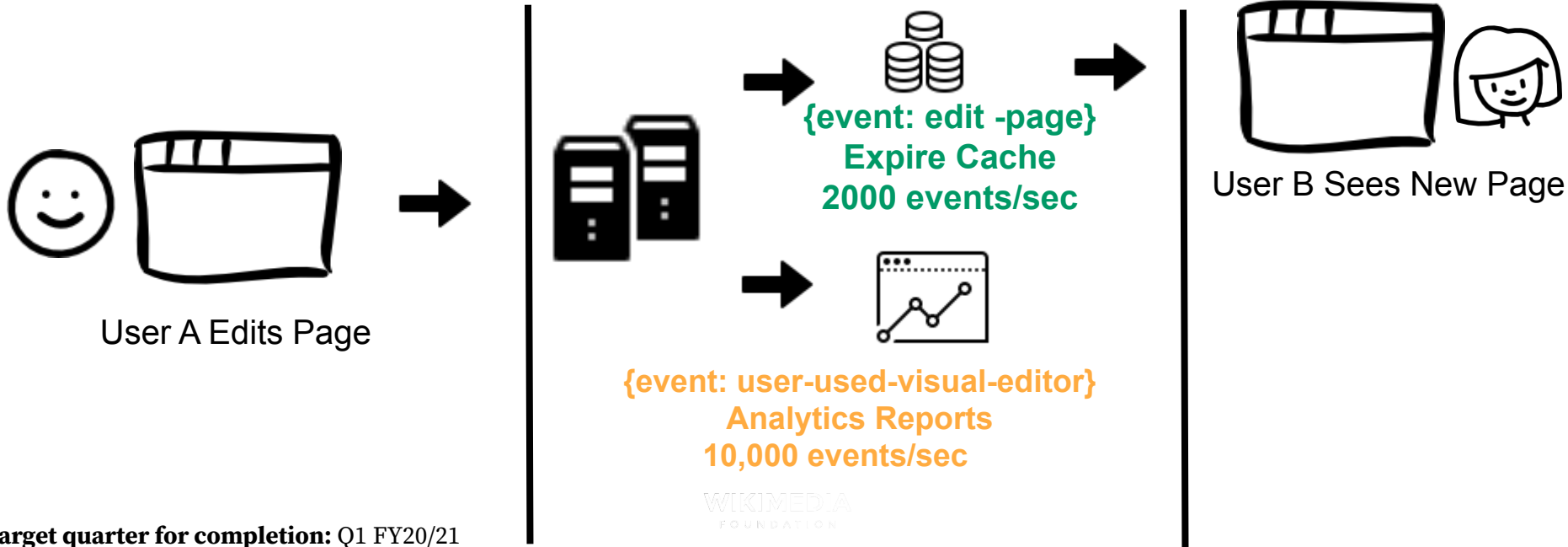
The recommendation

Continue pilot deployment and iteration in FY20/21.

Expand deployment to 7 communities.

Department: Technology

Platform Evolution

Modern Event Platform: Build a reliable, scalable, and comprehensive platform for building services, tools and user facing features that produce and consume event data.

User A Edits Page

{event: edit -page}
Expire Cache
2000 events/sec

User B Sees New Page

{event: user-used-visual-editor}
Analytics Reports
10,000 events/sec

Target quarter for completion: Q1 FY20/21

Platform Evolution

Key Results	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
5% of analytics events and 100% of production events migrated to the new event platform. *The percentage of production and analytics events will increase every quarter until older systems can be fully deprecated.* Baseline:0	5%	100% of production events and 5% of analytics events migrated	100% of production events and 5% of analytics events migrated	Completed	Completed
Client Error Logging is deployed to 1 wiki and error stats are displayed on our operation dashboards. **Baseline:** 0	1			Completed	Completed
By June 2020, all production and consumption of **new event data** originated in our websites is flowing through this new event platform. **Baseline:** 0%	100%	0%	0%	Slightly delayed but still on track	**Delayed until end of Q1 FY20/21**

Drill Down: Platform Evolution

The situation

The adoption of Modern Event Platform (MEP) by Product Teams instrumenting analytics events has been delayed by about one quarter -

to Q1 FY20/21

The impact

Not much, teams should start adding new events in the new platform next quarter.

It will make possible, for example, to change sampling rates without deploying any code

The recommendation

Next year we are working on wider adoption of MEP with Product Teams and a **wider adoption of MEP in our Architecture work.**

Department: Technology

Improve Developer Productivity

Objective:

Maintain and evolve developer tooling, testing infrastructure, validation environments, deployment infrastructure, and supporting processes.

After extensive feedback, including in the Developer Satisfaction Survey, we have decided to evaluate replacement tooling for for Continuous Integration (CI) and Code Review, beginning in FY20/21. This is a Big Deal^{TM}.

Our current CI stack is still serving our developers feedback needs until we migrate to the new system; and our standard deviation of feedback delay continues to be within bounds.

Target quarter for completion: Q4 FY19/20

Improve Developer Productivity

Key Results	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
Release Engineering and SRE teams create a plan to implement a Deployment Pipeline compliant CI system. **Baseline:** 0%	100%	10%	60%	70%	**100%**
Maintain and improve the Continuous Integration and Testing services. **Baseline:** Standard deviation: 12 minutes	<13.2 mins	On track	4.62 mins	9.31 mins	**6.73 mins**
Developers have a consistent and dependable deployment service. **Baseline:** 1 issue per quarter	Address new reports within 1 month	On track	0 relevant incidents this past quarter	4 incidents	**0 relevant follow-ups this past quarter**
Reduce infrastructure gaps in the areas of backups, disaster preparedness, observability, infrastructure automation and team structure & support. **Baseline:** TBD	TBD	TBD	85%	92%	97%

Department: Technology

Drill Down: Improve Developer Productivity

The situation

We are undertaking a very large project to evaluate moving all development to a new platform. This will require collective buy-in and coordination across all engineering teams and the community.

The Engineering Productivity team will lead this effort and take the brunt of the socializing and training effort and impact.

The impact

We are addressing a long standing complaint within our engineering community: the dislike of Gerrit.

This change also provides a more robust and modern Continuous Integration (CI) and testing workflow which allows us more flexibility in our technical evolution.

The recommendation

Working together on this is the most important part.

Department: Technology

Improve Developer Productivity

Objective:

We will improve developer efficiency for all developers: new and experienced, internal and external.

COVID and other holidays have been slowing down our Cycle Time metric. The big slowdown happened when we had a Foundation-wide 5 day weekend at the end of April.

While the numbers are down from last year, the Developer Satisfaction survey informed much of our plans for the next fiscal year (see also: GitLab). We expect to see improvements based on these changes.

Q4 has been a key quarter for our efforts around developing formats to increase technical capacity in emerging communities. While we were still developing the Starter Kit and conducting a virtual workshop series jointly with Indic TechCom, lessons learned from FY 19/20 already played a key role in defining the technical engagement strategy for the years ahead.

Target quarter for completion: Q4 FY19/20

Improve Developer Productivity

Key Results	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
Determine a baseline set of metrics to assess internal developer efficiency, including time to first merge (new devs), time to first review (new devs), average time to merge (fully ramped devs), and average time to review (fully ramped devs), by end of Q2. **Baseline:** 0	4 baselines	0	Completed	Completed	Completed
Improve all baseline developer efficiency metrics by 10% by the end of the year. **Baseline:** 3.8/5 or 76%	4.18/10% increase	No change	No change, update in Feb 2020	3.4 (-8%)	**No change**
Improve Cycle Time by 10% year over year. **Baseline:** 11.6 days	10% decrease	No change	11% decrease (10.3 days)	36% increase (14 days)	**39% increase (16 days)**

Department: Technology

Improve Developer Productivity ✓

Key Results	Y1 Goal	Q1 Status	Q2 Status	Q3 Status	Q4 Status
Successfully run Wikimedia's technical internship and outreach programs — Google Summer of Code (GSOC), Google Season of Docs (GSOD), Outreachy and Google Code-In (GCI) — that are measured by number of completed projects in GSOC, GSOD, Outreachy and the number of completed task instances in GCI.	20 projects completed, 700 tasks completed, 14 projects promoted for the 2020 rounds	n/a	15 projects completed in GSOC, GSOD, Outreachy. 600 task instances completed in GCI	**21** projects completed in GSOC, GSOD, Outreachy round 19 and 20, **715** tasks instances completed in GCI, **17** projects promoted for the 2020 rounds	**Completed in Q3** **During Q4: 17 students started, more diversity than usual!**
Develop, test and evaluate different formats for building technical capacity in emerging communities.	At least 3 formats are developed and tested, year 1 evaluation	n/a	2 formats developed, ongoing testing	Started to work on format 3; improved related documentation and drafted a landing page on Meta	**Completed**

Department: Technology

Home-made Small Wiki Toolkits banner used at Wikimania 2019. *Image by Birgit Müller, cc-by SA 4.0*

"Group photo" with some participants of the Indic tech workshop series in June 2020. *Image by Satdeep Gill, cc-by sa 4.0.*
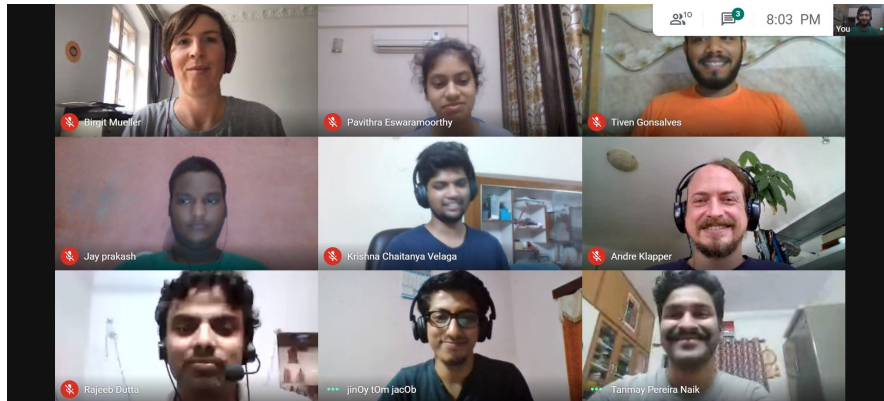
How can we increase the technical capacity of emerging communities?
By exploring technical challenges of smaller wikis, learning by doing and iterating, starting a network, and raising awareness.

Formats:
- **Brainstorming sessions** to explore challenges and gather actionable ideas
- **Technical workshops** & introductions in key areas
- A **Starter Kit** for smaller wikis

About 200 people participated in sessions and conversations related to technical capacity building in smaller wikis throughout the FY19/20 year.

Drill Down: Improve Developer Productivity

The situation

The Starter Kit covers key areas, but has limits where official tools should replace current cumbersome workflows (i.e. templates, modules).

We had to course-correct our event plans due to COVID-19. Instead of 2 days of onsite training in Hyderabad, we conducted a virtual workshop series jointly with the local group in India at the end of June. First feedback has been very positive.

The impact

Insights and lessons learned on technical challenges of smaller wikis informed the development of the FY20/21 roadmap.

Activities throughout the year led to a growing awareness of Small Wiki Toolkits as a resource and as a network; notable in Q4.

The collaboration with Indic TechCom serves as a best practise model for future regional technical capacity building.

The recommendation

Continue to work with wiki communities around the globe to increase technical capacity.

Gather more insights on the role of community-built tools and bots to raise awareness and find better ways to support onwiki tooling.

Build funnels to other teams in the Foundation to raise awareness for the unmet needs of our technical contributors where they can't solve them in an easy way.
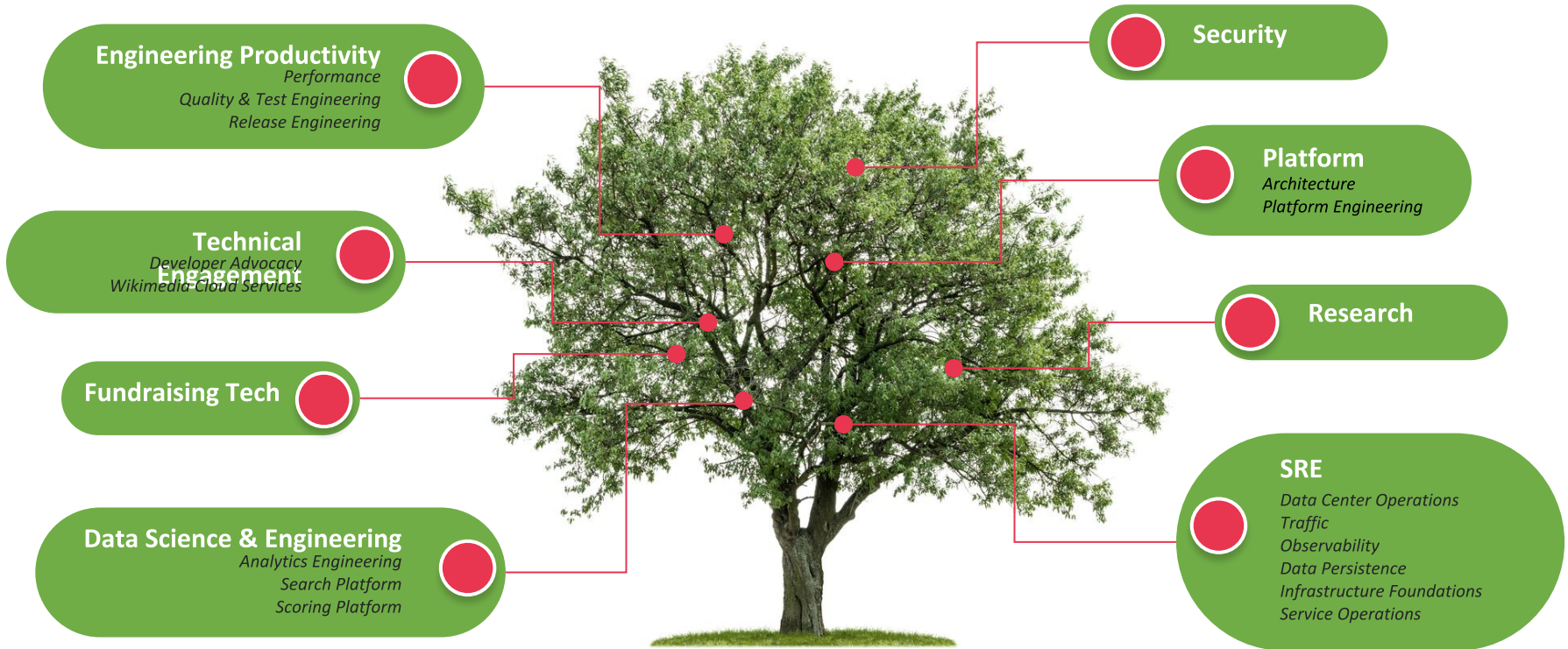
Department: Technology

Supporting Work

WIKIMEDIA

FOUNDATION

Who are we?

Photo by Brandon Green on Unsplash

Technology Department

Engineering Productivity
Performance
Quality & Test Engineering
Release Engineering

Technical Engagement
Developer Advocacy
Wikimedia Cloud Services

Fundraising Tech

Data Science & Engineering
Analytics Engineering
Search Platform
Scoring Platform

Security

Platform
Architecture
Platform Engineering

Research

SRE
Data Center Operations
Traffic
Observability
Data Persistence
Infrastructure Foundations
Service Operations

New Hires

SRE

Stephen Shirley
Leo Mata

Engineering Productivity

Ahmon Dancy

Platform

Naike Nembetwa Nzali

Technical Engagement

Nicholas Skaggs

Scoring

Chris Albon

Search Platform

Ryan Kemper

Anniversaries *(April - June)*

14 years
Tim Starling

13 years
Robert Halsell

9 years
Timo Tijhof

8 years
Andrew Otto
Faidon Liambotis

7 years
Erik Bernhardson
Monte Hurd
Brandon Black
Alexandros Kosiaris

6 years
Sarah Rodlund
Giuseppe Lavagetto
Rummana Yasmeen
Dan Duvall
Mukunda Modell
Filippo Giunchedi
Elliot Eggleston
Papaul Tshibamba

5 years
Moritz Muhlenhoff
Jaime Crespo
Dylan Kozlowski
David Causse
David Strine

3 years
Maggie Epps
Arzhel Younsi
Keith Herron

1 year
Willy Pao
James Fishback
Will Doran
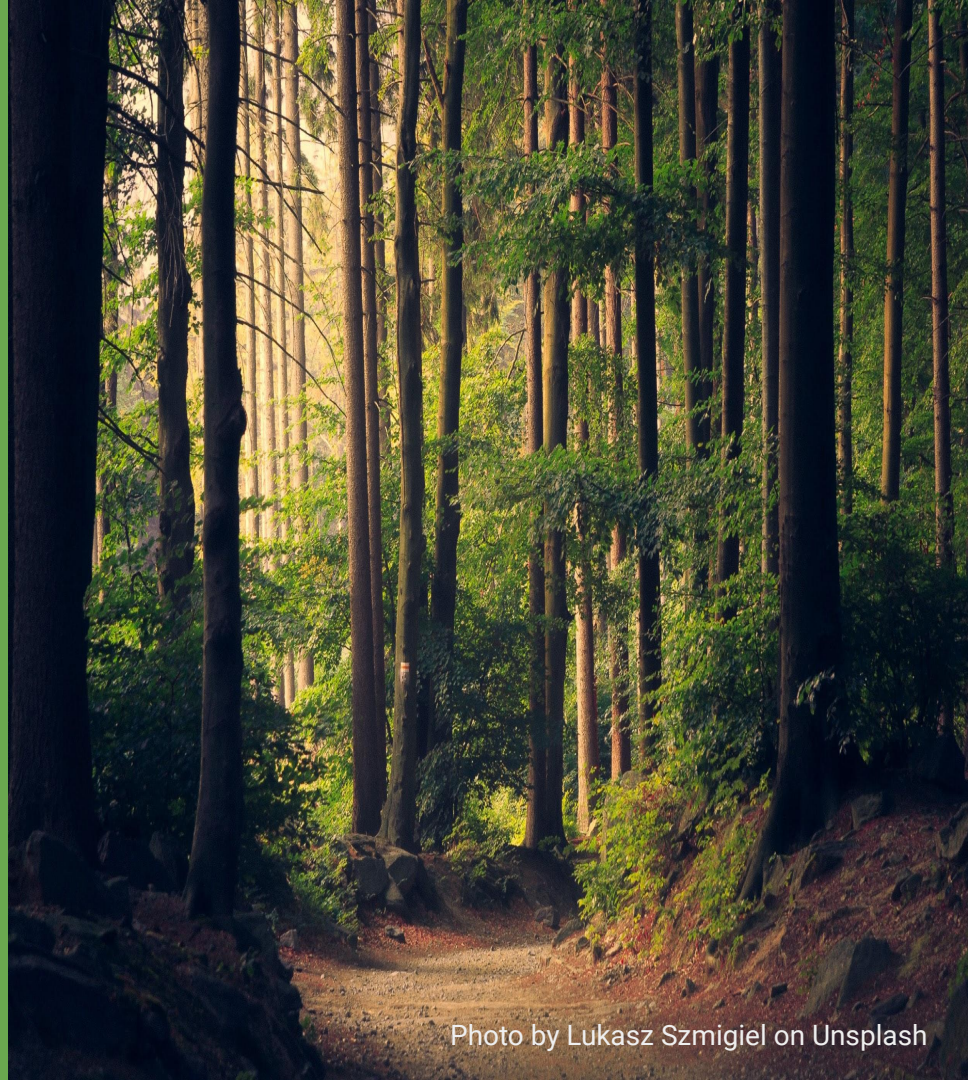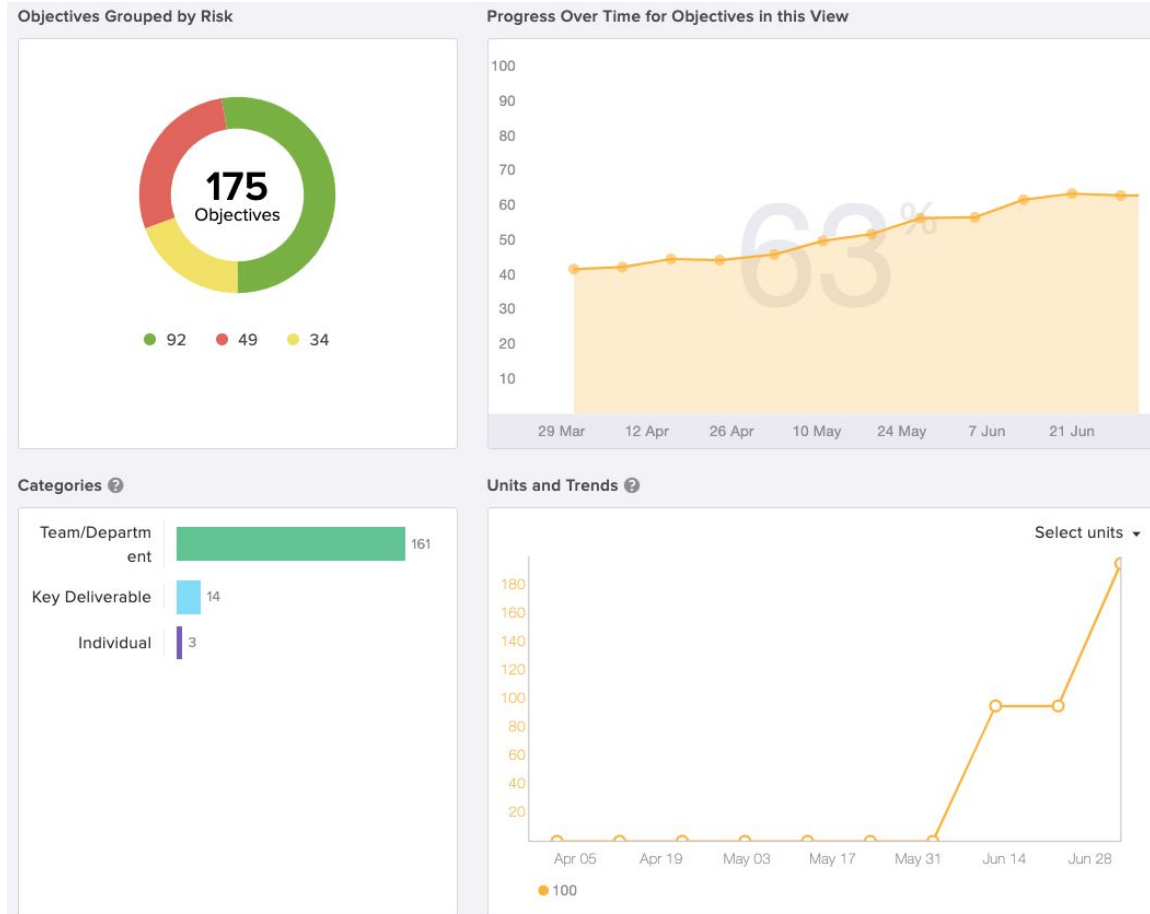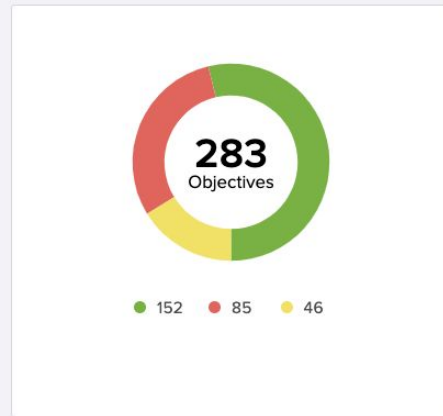Andy Craze
Jose Pita

What we've done and, what we will do

Photo by Lukasz Szmigiel on Unsplash

Department OKR Status *Q4*

Objectives Grouped by Risk

175 Objectives

- ● 92
- ● 49
- ● 34

Progress Over Time for Objectives in this View

100
90
80
70
60
50
40
30
20
10

29 Mar | 12 Apr | 26 Apr | 10 May | 24 May | 7 Jun | 21 Jun

63%

Categories ⑦

- Team/Department ▬▬▬▬ 161
- Key Deliverable ▬ 14
- Individual | 3

Units and Trends ⑦

Select units ▾

180
160
140
120
100
80
60
40
20

Apr 05 | Apr 19 | May 03 | May 17 | May 31 | Jun 14 | Jun 28

● 100

Department OKR Status *FY19/20*
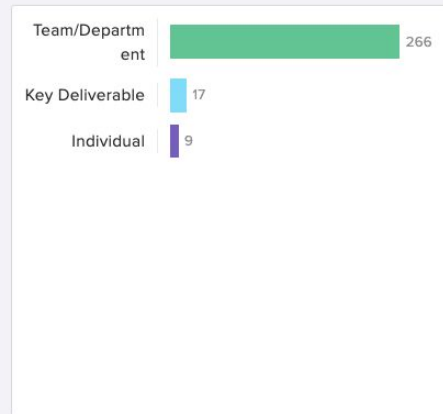
Supporting wins

Collaboration and Communication

Scoring: deploying Jade MVP for control over AIs and manage labeled data

Performance: trained 23 staff on 16 teams on frontend web performance

Observability: migrated 100% of SRE to a robust and automated paging tool for better incident management

Research: created a comprehensive framework for understanding knowledge gaps and supported covid-19 article creation

Security: published privacy roadmap and threat assessment for risk modeling

QTE: kicked off cross training within Tech and Product departments

Better Infrastructure

DC Ops: decreased amount of tech debt by 33%, while lowering costs by committing to longer contracts

Service Ops: successfully implemented (and was tested against) a stabilized memcached gutter pool

WM Cloud Services: launched toolforge.org DNS domain with individualized sub-domains

Analytics: upgraded pageview data to highlight bot vandalism and bot spam *(removing spikes that aren't real pageviews)*

Infra Foundations: automated IP allocation & DNS generation, deployed 42 security updates

Search: prioritized paying down Wikidata Query Service and Commons Query Service tech debt, also collaborated on SDAW work

Growth and Improved Capabilities

DevAdvocacy: blog posts exploded - 13 were published in Q4

FR-Tech: eliminated SPoF *(single point of failure)* in the donation work flow

Architecture: began exploration with Structured Data Across Wikipedia, and a proof of concept for structured content storage

Platform Eng: finished 5 key initiatives that helped enable high throughput for active articles, and serving all wiki sessions from new multi-dc aware data storage

RelEng: automated train branching, added Gerrit and Phabricator features

Challenges

We're all figuring out our "new normal"

DC Ops: being physically in data centers exposed some folks to covid-19

Security and other teams: dealing with some ongoing security related incidents that are requiring more mitigation than usual

- ❖ dealing with global pandemic challenges

- ❖ shortage of time available for work due to family obligations and illnesses

- ❖ trying to minimize single points of failures *(personnel where only they know how to do xyz)*

- ❖ lack of normal planning cycles due to no in-person collaboration

- ❖ uncertainty about the future

- ❖ widespread societal unrest

COVID-19 👎

WIKIMEDIA
FOUNDATION
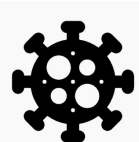
Department:
Technology

Supporting themes

❖ flexibility is key
❖ showing compassion and understanding for what others might be going through
❖ assuming good faith under heavy workloads
❖ understanding that it's ok to not be as productive as normal during a global pandemic
❖ maintaining good connections to our network of community researchers and developers
❖ good, easy to understand documentation is critical to have
❖ collaboration across teams and departments is essential
❖ cross-training is in high demand and needs to be prioritized
❖ turning planned onsite trainings and workshops into amazing virtual events

Department:
Technology

Upcoming

❖ Understanding and reducing gender disparity for women readers of Wikipedia
❖ Real time performance monitoring via new device lab (Kobiton)
❖ Planning for next iteration of ORES
❖ Technical outreach targeting folks that are busy 'filling in the gaps' and creating solutions
❖ Support Advancement and annual fundraising goals
❖ Resolving single points of failures in WMCS and creating easier workflows for volunteer contributions
❖ Continuous Integration & Code Review evaluation
❖ Code Review evaluation
❖ OKAPI proof of concept
❖ Data center switchover testing
❖ Establishing Service Level Objectives (SLO's) with Product Dept
❖ Publish and socialize architecture process
❖ Encouraging more system-scoped testing

Questions?