Technology Group
Research and Data, Design Research, Analytics Engineering, Performance & Availability
FY Q4: April - June 2015/16
## Objective: Backup Datacenter

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make codfw (Dallas data center) functional as a backup data center</td>
<td>• Serve MediaWiki application server traffic from codfw for 48 hours</td>
<td>Completed</td>
</tr>
<tr>
<td></td>
<td>• Serve Swift, ElasticSearch and RESTBase, Parsoid services from codfw</td>
<td></td>
</tr>
</tbody>
</table>

**Learning:**

We adjusted our initial plans to ensure proper coverage and a successful approach. Being flexible, iterative and thoughtful planning and communication led to a positive result. Many teams in Technology and supporting groups coordinated to make this happen. This puts us in a much better position to handle outages and support our mission. Switchback went smoothly (about 20 mins of blocked editing on second day) reading was never delayed. We will continue to improve these times.
Quarterly review
Research and Data
Q4 - 2015/16

Approximate team size during this quarter:
5 FTE, 1 PTE, 2 research fellows, 12 collaborators
## Q4 - Research and Data

**Objective:** Revscoring in production

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EXPERIMENT</strong></td>
<td>Productize ores.wikimedia.org (T106867)</td>
<td>completed</td>
</tr>
<tr>
<td><strong>Revscoring in production</strong></td>
<td>ORES extension deployed to 2+ wikis (T130210)</td>
<td>completed (wikidata, fawiki)</td>
</tr>
</tbody>
</table>

**Team members involved:** 2

**Collaborators:** 2

Wikimedia’s 1st production-level AI platform, integrated into RecentChanges as a beta feature.

**Acknowledgments.** Amir Sarabadani, Sabyasachi Ruj, Alex Kosiaris, Daniel Zahn, Jaime Crespo

[https://meta.wikimedia.org/wiki/ORES](https://meta.wikimedia.org/wiki/ORES)
Q4 - Research and Data

Revscoreing/ORES

**Wikimania session:** “Lessons learned building machine learning models for Wikidata” (T138351)

Substantial performance improvements (T134780, T134267)

14 more models added to the edit-quality classifier (github)

**ORES on Beta Cluster** (ores-beta.wmflabs.org) (T138445)

Switched to **scap3 for deployments** (in collaboration with RelEng) (T131857)

https://meta.wikimedia.org/wiki/ORES
## Q4 - Research and Data

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EXPERIMENT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion modeling</td>
<td>publicly release cleaned dataset of all user talk diffs (T133078)</td>
<td>missed</td>
</tr>
<tr>
<td>Team members involved: 1</td>
<td>analyze user_talk editing behavior of users blocked for harassment / personal attacks (T140041)</td>
<td></td>
</tr>
<tr>
<td>Collaborators: 2</td>
<td>label user talk posts for tone (friendly to aggressive) and for harassment / personal attacks (T134379)</td>
<td>completed</td>
</tr>
</tbody>
</table>

**Acknowledgments.** Nithum Thain, Lucas Dixon (Jigsaw); Patrick Earley (CE)

https://meta.wikimedia.org/wiki/Research:Detox
Q4 - Research and Data

Discussion modeling

Outreach
Community discussion at Wikimania on how to use ML-based strategies to moderate aggressive/abusive behavior: “Tools for Tackling Personal Attacks, Harassment and Toxicity” (T136763)

Demoing automated classification
Designed ML models for detecting personal attacks and aggression on user talk pages and built a proof-of-concept API: wikidetox.appspot.com

https://meta.wikimedia.org/wiki/Research:Detox
# Q4 - Research and Data

## Objective: Research outreach

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FOCUS</strong></td>
<td>host a joint research workshop <em>(Wiki Workshop 2016)</em> at WWW '16 and ICWSM '16 <em>(T125247, T125608)</em></td>
<td>completed</td>
</tr>
<tr>
<td>Research outreach</td>
<td></td>
<td>completed</td>
</tr>
<tr>
<td>Team members involved: 2 Collaborators: 8</td>
<td>host <em>WikiCite 2016</em> – a technical event/hackathon focused on data modeling and extraction of Wikipedia citations and source metadata <em>(T125186)</em></td>
<td></td>
</tr>
</tbody>
</table>

**Learning.** Substantial effort went into the organization of the 3 events, with 8 external collaborators, including logistic support for WikiCite. Need to streamline process for funding similar events in the future. Major network outage in Berlin requires improved contingency planning.

**Acknowledgments.** Advancement team for support with funders. WMDE for logistics support with WikiCite. Jure Leskovec and Bob West for co-organizing Wiki Workshop.
Q4 - Research and Data

Research outreach

Wiki Workshop 2016
- workshop hosted jointly at two conferences, WWW ‘16 and ICWSM ‘16
- 28 accepted papers, 10 invited talks, over 70 attendees combined

WikiCite 2016
- 55 participants from ~50 groups and organizations
- 8 distinct tracks ([draft report](#) + [mailing list](#))
- collaborated with a team at CMU to run [post-event survey](#)
- post-event outreach ([Open Science Radio](#), ELAG ‘16, Wikimania ‘16)
- complete report to be presented in Q1-FY17
Q4 - Research and Data

Objective: Reader segmentation

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FOCUS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader segmentation research</td>
<td>Deepen our understanding of Wikipedia readers by a quarter-long focused research on the data collected via the large scale English survey (carry-over goal T121727)</td>
<td>missed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Learning.** Other projects took priority and we couldn’t complete this work, despite additional analyses being published on Meta. Will need to reassess next steps for the project.

https://meta.wikimedia.org/wiki/Research:Characterizing_Wikipedia_Reader_Behaviour
Q4 - Research and Data

Other achievements in Q4

Wikistats 2.0
Handed off wikistats reports to Analytics (infodisiac) (T128870)

Wikipedia Navigation Vectors data release
Released session navigation vector data (docs) (T134210)

Article recs in the news
Wikimedia Blog; VentureBeat; Stanford News (T130334)

Research FAQ
Rewritten and published extensive FAQ, per FDC request (T137095)
## Q4 - Research and Data

<table>
<thead>
<tr>
<th>Category</th>
<th>Workflow</th>
<th>Comments</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NDAs / MOUs</td>
<td>Created 3 new MOUs for research collaborators (T129582, T130324, T135922)</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Showcases and talks</td>
<td>Hosted 0 research showcases</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Review / PC work</td>
<td>Contributed paper reviews for OpenSym ‘16 (T133704)</td>
<td>M</td>
</tr>
</tbody>
</table>

**Learning.** We put research showcases completely on hold during the quarter, due to 3 major other outreach activities (Wiki Workshop 2016, WikiCite 2016, Wikimania 2016). We will resume in July 2016.

*Type: new, reactive, maintenance*
Q4 - Research and Data

- **Research & Data team page**
  - Describing goals, processes and projects.
- **Goals for Q1 FY17**
  - What we are planning to do in the coming quarter
- **FY17 priorities (Annual Plan)**
  - Top priorities for the fiscal year
- **Phabricator workboard**
  - What we are currently doing (see also our dedicated project boards)

**Q4 collaborators**
Amir Sarabadani, Sabyasachi Ruj (ORES); Lucas Dixon, Nithum Thain (Discussion modeling); Jonathan Dugan, Anna Filippova, Daniel Mietchen, Cameron Neylon, Lydia Pintcher, Erik Trainer (WikiCite); Jure Leskovec, Robert West (Wiki Workshop)
Quarterly review
Design Research
Q4 - 2015/16

Approximate team size during this quarter: 4 FTE
# Q4 - Design Research

**Objective:** Personas

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use personas in product development and iterate pragmatic personas.</td>
<td>Support the use of personas in product development. Analysis on the collected reader, new editor interviews and addition of deep dive (New Readers) personas.</td>
<td>Not completed - Did some work in supporting the use of personas. Because of changes in design team and lack of awareness of personas and how to use them, we deprioritized iterating the personas for now. We will try collaboration and education about how to use personas and after that move forward with iterating personas if needed.</td>
</tr>
</tbody>
</table>

**Team members involved:** 2

**Learning:**

People need training and collaboration for using personas in our evolving product processes. If there is interest in using them, we will move forward.
Q4 - Design Research

We have a set of 6 pragmatic personas and 6 personas from the New Readers project: 3 from Nigeria and 3 from India.
## Q4 - Design Research

**Objective: Evaluative Design Research**

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintain <em>evaluative design research projects</em> with Editing, Reading, Community Engagement and Discovery</td>
<td>Continue to mature collaboration between design research and product teams (make our system more lean). Continue to mentor and coach teams that do not have support to do design research (PC&amp;L) Continue to provide evaluative research to official product teams (Reading Editing and Discovery)</td>
<td>Completed (6 projects)</td>
</tr>
</tbody>
</table>

**Team members involved:** 3

**Learning:**
- Communication is key: regular meeting cadence with vertical heads and PMs keeps us clued-in
- On-demand product support requires innovation/iteration of our recruiting, testing methods
Q4 - Design Research

Completed projects

- Notification user survey (T127257)
- Notification page user study (T135564)
- Single edit tab / NUE
- Reading lists heuristics (Android)
- Text sizing accessibility (Android & iOS)
- Language switching (Android & Mobile Web)
## Q4 - Design Research

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deep dive / contextual inquiries in Nigeria and India.</strong></td>
<td>Forward our deep dive / CI process and method to be a repeatable practice implemented by WMF. Bring data and stories about New Readers in Nigeria and India back to WMF and the movement. Start to look across deep dives (South Africa, Mexico, Nigeria and India)</td>
<td><strong>Completed:</strong> Participated in data collection, analysis and synthesis of contextual inquiries in Nigeria and India. Workshop to deliver findings July 13 and 14, 2016. All content (data, observations, images, video, fanalysis, synthesis, findings, opportunities, concepts and future plans) will be on wiki within Q1.</td>
</tr>
</tbody>
</table>

*Team members involved: 1*
Q4 - Design Research

Nigeria:
70 interviews
Lagos
Epe
Benin City

Team:
4 Reboot staff
2-3 WMF staff
4 local researchers
2 weeks

India:
60 interviews
Delhi
Jaipur
Chennai

Team:
2 Reboot
1-2 WMF staff
3 local researchers
2 weeks

These contextual inquiries are part of the work collected from across teams working on the New Readers project (Global Partnerships, Communications, Reading and community Engagement and Design Research) to better understand and serve potential new readers where access to the internet is quickly growing. Workshop to deliver findings and plan concept development is July 13 and 14, 2016. All info about this work will be available on wiki during Q1.
## Objective: Collaboration with UW on Survey

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
</table>
| Continue, complete **Collaboration with UW on survey.**<br><br>*Team members involved: 1 Collaborators: 11* | • Deploy survey to 200 or more students at University of Washington (**T127251**)  
• Analyze and publish the results on Wiki (**T135563**) | Completed |

### Learning:
- Leveraging pre-existing academic ties leads to efficient research partnerships
- We gained insight into the behaviors and needs of a valuable set of readers: college students
Q4 - Design Research

UW survey

- Learned where on the web students turn to for homework, research help
- Results shared at June 2016 Metric meeting
- UW blog post: http://www.hcde.washington.edu/news/directed-research-group-examines-free-online-resources
# Q4 - Design Research

## Objective: “Benchmarking” / Tooling

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
</table>
| Create a project plan and understand dependencies and resource needs for benchmarking.  
*Team members involved: 5* | - Pick up where we left off, using the work completed before to create a stable testing environment  
- Make a plan to create a stable testing environment.  
- Prepare to run pilots with userzoom and loop11 (and other possible tools) | Completed: We focused on tooling instead of benchmarking - and secured a contract with UserZoom |

## Learning:

- Finding an affordable solution that addressed diverse needs, technical and legal constraints was a challenge
- We needed to re-scope the project mid-quarter to make sure we had a user testing solution in place by Q1 2016-17
Q4 - Design Research

Tooling

- **Core features:** remote, unmoderated user testing for desktop, mobile web, mobile apps*
- 3 ‘seats’ for design researchers supporting Reading, Editing, Discovery
- Flexibility to recruit our own participants, or use professional testers through UserZoom

*requires integration with UserZoom SDK (not yet completed)
# Q4 - Design Research

## Core workflows and metrics

<table>
<thead>
<tr>
<th>Category</th>
<th>Workflow</th>
<th>Comments</th>
<th>Type</th>
</tr>
</thead>
</table>
|          | Participant recruiting | Metrics  
  - 2,000 participants contacted in Q4  
  - 70 participants recruited in Q4  
  - 9 projects recruited for in Q4 | M |
|          | Participant recruiting | Social media outreach to grow our participant database  
  - Recruitment posts on Facebook and Twitter every two weeks  
  - Initial result: 109 new participant signups in 7 days | N |
|          | Participant recruiting | User study participation incentives to improve recruiting success  
  - t-shirt or Wikipedia water bottle from the Wikimedia Store  
  - first-time participants only, 45-minute minimum session | N |
|          | IdeaLab/Inspire campaign infrastructure maintenance | Phasing out: Jonathan training Chris Schilling (Resources) to take over legacy responsibilities (gadgets, templates, bot). No additional work anticipated in 2016-17. | M |

*Type: new, reactive, maintenance*
Quarterly review
Analytics Engineering
Q4 - 2015/16

Approximate team size during this quarter: 6 FTE (2 devops) and 1 PT

Key performance indicator: Velocity

<table>
<thead>
<tr>
<th>Quarter</th>
<th>April</th>
<th>May</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>929</td>
<td>522</td>
<td>341</td>
<td>66</td>
</tr>
<tr>
<td>Objective</td>
<td>Measure of success</td>
<td>Status</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Make Unique Devices dataset public <strong>task</strong> T126767</td>
<td>API Endpoint where Unique Devices Data is available.</td>
<td>Done.</td>
<td></td>
</tr>
</tbody>
</table>

**Learning:**

Great community contributions towards clients of our APIs. Our tools also benefit from public endpoints that are easier to integrate than internal-only data.
## Q4 - Analytics Engineering

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wikistats 2.0 (ongoing goal).</strong> Remove dependency of dumps as sources of edit data to be able to replace wikistats edit reports.</td>
<td>Proof of concept code that proves we can get most (if not all) data we need to reconstruct edit history from mediawiki databases.</td>
<td>Done.</td>
</tr>
</tbody>
</table>

**Learning:**

We can reconstruct most edit history from data in our dbs, but in some cases details like page titles at the time of an edit are not available. We can sometimes parse existing dumps for this data, and our approach is extensible enough to do this at a later time.

**A byproduct of this work is the ability to generate dumps from data on hdfs.**
<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Public Event Stream</strong> task T130651</td>
<td>POC that makes arbitrary JSON events available for public consumption.</td>
<td><strong>Not Done. We had to prioritize efforts around Druid Pipeline.</strong></td>
</tr>
<tr>
<td>Remove dependency of dumps as sources of edit data to be able to replace wikistats edit reports.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
View Count: 13.0 m

Belgium: 40.4 k
## Q4 - Analytics Engineering

### Objective: Better Data Access

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prototype <strong>Pageview Data Pipeline On Druid</strong> task T130258</td>
<td>POC UI available to look at Pageview Data without the need of the command line. Backend puppetized, will puppetize UI next quarter.</td>
<td>Done.</td>
</tr>
</tbody>
</table>

**Learning:**

**Much faster pageview queries than hive (seconds versus minutes)**

Users love NOT having to use the command line :)
## Objective: Operational Excellence

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational Excellence</td>
<td>Drop rate limit on article endpoint, right now we can sustain only 30 requests per sec, we are aiming for 300.</td>
<td>Not Done. Biggest priority for next quarter.</td>
</tr>
<tr>
<td><strong>Better response times</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>pageview API</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>task</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>T124314</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Learning:**

Cassandra is been quite a bit of work and with our volume of data per node has significant scaling issues. Having hardware with SSDs will help but issues loading and compaction of data remain, still working on those. New cluster is in place we are performance testing.
Significant operational time on varnishkafka and varnish4 upgrade (ops goal)

Launched http://analytics.wikimedia.org
Dashboards and Data Downloads for Wikimedia Projects

Data compiled by community and staff, from projects hosted by the Wikimedia Foundation.

Code for this page can be seen here: 🔄
Quarterly review

PERFORMANCE TEAM

Q4 - 2015/16
KPI: First paint time

Time in milliseconds it takes before user sees any content.

<table>
<thead>
<tr>
<th>%ile</th>
<th>Q3</th>
<th>Q4</th>
<th>+/- %</th>
</tr>
</thead>
<tbody>
<tr>
<td>90th</td>
<td>3,063</td>
<td>3,285</td>
<td>+7.25%</td>
</tr>
<tr>
<td>75th</td>
<td>1,761</td>
<td>1,877</td>
<td>+6.59%</td>
</tr>
<tr>
<td>50th</td>
<td>987</td>
<td>1,058</td>
<td>+7.19%</td>
</tr>
<tr>
<td>25th</td>
<td>575</td>
<td>616</td>
<td>+7.13%</td>
</tr>
<tr>
<td>10th</td>
<td>330</td>
<td>352</td>
<td>+6.67%</td>
</tr>
</tbody>
</table>
KPI: Page save time

Time in milliseconds between the user clicking [Save] and the edited article starting to load.
KPI: Page save time

Time in milliseconds between the user clicking [Save] and the edited article starting to load.
### Q4 - Performance

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deploy Thumbor</td>
<td>Thumbor deployed in production.</td>
<td>Production integration of Thumbor has been mimicked on a virtual machine. Objective pushed to next quarter.</td>
</tr>
</tbody>
</table>

*Team members involved: 1*

The target was missed because we decided to make Thumbor available on Debian Linux, which is a time-consuming task. It was a learning experience for Gilles, who had never done Debian packaging before.

**Learning:** Debian packaging is time-consuming but not difficult per se. It’s worth doing because it gives more exposure to the open source tools we use, which make them more likely to receive bugfixes and improvements from the wider open source community.

Filippo from the Operations team was instrumental in having Thumbor move forward, providing code review and assistance through the whole project.
### Q4 - Performance

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Inspector in a deployable state.</td>
<td>Performance Inspector is ready to be deployed for Wikipedia users.</td>
<td>The inspector is not releasable, we want to integrate more functionality before the release. Objective pushed to next quarter.</td>
</tr>
</tbody>
</table>

**Team members involved: 1**

The target was missed because Peter also focused on helping the reading team and getting WebPageTest metrics more consistent and working with HTTP/2. It was also a learning experience for Peter who had never built an extension before and it took longer time than expected.

**Learning:** Focus on 1 thing.
# Q4 - Performance

<table>
<thead>
<tr>
<th>Objective</th>
<th>Measure of success</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support CODFW switchover.</td>
<td>Switchover without incident.</td>
<td>Switchover occurred. Fixed bugs with Configuration management, Database abstraction, and more. Second switchover was smooth.</td>
</tr>
<tr>
<td><em>Team members involved: 2</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Learning:** The main problem on the first attempt was a misspelled server name in the master hostname configuration file for MediaWiki ("ProductionServices.php"). Several people did not notice this in code review. A simple unit test will avoid this kind of incident in the future.
Q4 - Performance

- Slave lag <5s
- Improved WebPageTest infrastructure. Metrics are more consistent, making it easy to spot regressions now and in the future.
- Made edit stashing work for the majority of edits.
- Identified SPDY/HTTP2 problems on slow connections for large pages.
- Navigation Timing metrics were significantly improved. More reliable gathering. Increased visibility surface area (by geolocation, browser, mobile/desktop/beta modes)
- Advocate, collaborate and educate: Discovery (Wikipedia Portal), Language Engineering (Interlanguage links), Fundraising Tech (CentralNotice storage), Reading Web (Loading of citations and images for mobile).