



Research and Data

Q3 2015

Q3 - Research & Data

Quarterly Objectives Summary

Objective	Impact on Goal	Dependency	ETA	Status
Data services <ul style="list-style-type: none"> • Prototype revscoring service 	Have a working prototype up on Labs of a public, query-able APIs of machine classified scores of Wikipedia revisions.	Labs: hosting the service Grantmaking: access to grantees	End of Q3	Released libraries the service depends on, tools for handcoding and project documentation. API up on Labs
Traffic reporting <ul style="list-style-type: none"> • Finalize Unique Clients requirements • Support AnEng with the productization of Page Views reports • Continued responsibility for ad hoc traffic analysis and reports 	Clear all research dependencies for productizing traffic reports by Analytics Eng	Analytics Eng Legal	End of Q3	<p>Completed QA of data based on new pageview definition and supported AnEng with the design of UDFs. PV reports as of Q4 will be automatically generated from Hadoop.</p> <p>Provided methodology and requirements for counting Unique Client, implementation is pending.</p>
Mobile research <ul style="list-style-type: none"> • Continue research support to WikiGrok and microcontributions • Research support for Project Gather (aka Collections) 	Deliver analysis of an AB test for WikiGrok on stable, perform all supportive analysis needed by the team to decide on next iteration	Mobile team: product goals and engineering support	End of Q3	<p>Completed reports on WikiGrok test.</p> <p>Provided ad-hoc support to Gather team prior to launch.</p>

Q3 - Research & Data

Quarterly Objectives Summary

Objective	Impact on Goal	Dependency	ETA	Status
Fundraising research • Develop proposal to run FR as a minimization problem [Better AB testing methods, Characterizing low value impressions]	Deliver and publish AB test methodological proposals	Fundraising	End of Q3	Completed analysis of Big English. Released tool with traffic projections. Published methodological reports: 1 - 2 ; #1 is now used as part of ongoing FR analysis
Formal collaborations Wrap up <i>Improving linking structure</i> Kick off <i>Increasing article coverage</i>	Complete report on improving link structure	External collaborators	5/8	Improving linking structure report in progress. Algorithm implementation being tested Work on article creation recommenders started.

Q3 - *Research & Data*

Quarterly Objectives
Success

Achievements

The team showed strong leadership in solving complex problems:

- Aaron took the lead of a team of grantees to do R&D work on first AI-as-a-service prototype.
- Leila performed research on microcontributions, led research with Stanford's InfoLab team, and started a new collaboration with Stanford's DeepDive team.
- Ellery redefined our testing methodology for FR and started work on big data services.
- Oliver completed work towards a new, robust page view metric

Other Q3 achievements beyond stated goals

- Completed [4 major open data releases](#)
- Led and released [org-wide Open Access policy](#) for all WMF research collaborations
- Hosted a workshop on cross-platform research and a Wikipedia meetup at [CSCW '15](#)
- Developed with UX team a vision for [content remix and reuse](#)
- Supported Legal and ACLU with data for the [NSA lawsuit](#)
- Organized a Wikipedia research workshop at [ICWSM '15](#)

(received great support from Team Practices, Legal, Analytics Eng, Ops, Mobile for all these projects)

Q3 - *Research & Data*

Quarterly Objectives
Miss

The completion of **link recommendation** is delayed due to the need (identified through research in Q3) for an improved algorithm. There are edge cases in which the algorithm doesn't perform well. We are 1) assessing the quality of the algorithm, 2) doing experiments with improved versions. We commit to having a working version of the algorithm by 5/8, along with its quality assessment, a research page report and a paper submission and a release of the underlying data (aggregate browsing traces) for the community.

The team provided continuous responsibility for **ad hoc reporting and querying** (both at team- and org- level), lacking self-service reporting infrastructure for key metrics: while we substantially met our stated goals, ad hoc reporting slowed down our progress.