Research and Data

Q4-2014 quarterly review
Outline

who we are / what we do
Q4-2014 retrospective
Q1-2015 goals
H1-2015 staffing outlook
Who we are

Erik Zachte

Aaron Halfaker

Oliver Keyes

Dario Taraborelli

Leila Zia

2013

Q4

2014

Q1

Q2

Q3

Q4

2015

Q1
What we do

We apply a range of research methods to produce knowledge on our users and our projects and support decision-making, product evaluation and strategy at the Foundation and within the movement.

DATA MINING  BEHAVIORAL ANALYSIS  DATA MODELING
PREDICTIVE MODELING  CONTROLLED EXPERIMENTATION, A/B TESTING
EXPLORATORY RESEARCH  RESEARCH CONSULTING
Q4-2014 retrospective

Metrics standardization

Topical research

Team support / community outreach
Q4 goals

A. Deliver stage-2 metrics

B. Topical research

C. Support focus areas (growth, mobile, fundraising)

D. Ad-hoc consulting for other teams
Metrics standardization
Metrics Standardization: Editor Model

activation

short-term retention

reactivation

long-term retention

Monthly Active Editors

https://meta.wikimedia.org/wiki/Research:Rolling_monthly_active_editor
Metrics Standardization: Editor Model

Research: Rolling monthly active editor

This page in a nutshell: This page describes an active editor metric computed on a daily basis using a 30-day rolling window.

About

The Wikimedia Foundation has been using monthly active editors as the key metric to track the size of the registered user population of contributors to Wikimedia projects. This page documents a new metric tracking monthly active editors and:

- computed using a rolling 30-day window instead of calendar months
- broken down into different segments of the registered user population

Generally speaking, in any given month, three categories of users are included in the active editor definition:

- New users who became active for the first time in the current month;
- Recently registered users who remained active (or survived) in the current month;
- Previously registered users who either remained active or were reactivated in the current month.

These informal categories can be operationalized as follows:

Acquisition and short-term retention

The first two categories measure the acquisition and short-term retention of active editors.

New active editors

A rolling new active editor \( (T, u, n) \) is a newly registered user who both registered and completed \( n \) edits to pages in any namespace of a Wikimedia project between \( T - u \) and \( T \).

- \( n = 5 \) edits
- \( u = 30 \) days

https://commons.wikimedia.org/wiki/File:Editor_Model_review.pdf
Topical research
Topical research: Mobile reach
How many visitors can we reach on mobile?

https://meta.wikimedia.org/wiki/Research:Mobile_Traffic
Topical research: Mobile acquisition

How many users can we acquire on mobile compared to desktop?

https://meta.wikimedia.org/wiki/Research:Mobile_editor_engagement/Newly_registered_users
Topical research: Editor activation

Where do new (active) editors come from?

https://commons.wikimedia.org/wiki/File:A_bird%27s_eye_view_of_editor_activation.pdf
Topical research: Active editor retention / trajectories
What are the top predictors of short or long-term active editor survival?

https://commons.wikimedia.org/wiki/File:Active_editor_survival.pdf
Topical research: Active editor migration

Do active editors *leave* or *migrate* to other projects?

This report is about a subset of editors who contribute to only those with at least 50 edits in a year and at least 5

| Migrations per path (= from project -> to project) |
|---------------------------------|--------|--------|--------|--------|--------|--------|--------|
| total             | 786    | 588    | 509    | 461    | 373    | 356    |         |
| 2002              | 3      | 6      | 2      | 1      | 3      |        |         |
| 2003              |        |        |        |        |        |        |         |
| 2004              | 6      | 6      | 2      | 5      | 3      |        |         |
| 2005              | 32     | 26     | 14     | 10     | 31     |        |         |
| 2006              | 67     | 53     | 42     | 13     | 51     |        |         |
| 2007              | 100    | 64     | 50     | 32     | 63     |        |         |
| 2008              | 94     | 58     | 65     | 25     | 57     |        |         |
| 2009              | 130    | 63     | 49     | 34     | 41     |        |         |
| 2010              | 84     | 80     | 51     | 31     | 37     |        |         |
| 2011              | 74     | 76     | 48     | 49     | 21     |        |         |
| 2012              | 86     | 69     | 53     | 51     | 50     | 20     |         |
| 2013              | 74     | 60     | 359    | 59     | 58     | 19     |         |

https://trello.com/c/3ecjp9aM/237-master-monthly-editor-activity-data
Topical research: Anonymous editor acquisition

How effectively can we acquire anonymous editors?

https://meta.wikimedia.org/wiki/Research:Anonymous_editor_acquisition/Signup_CTA_experiment
Topical research: Article survival

How article creation workflows (drafts, AfC) impact content growth

Team support, ad-hoc consulting and community outreach
Teams supported: Focus areas

Q1
- Growth
- VE

Q2
- Growth
- Mobile

Q3
- Growth
- Mobile

Q4
- Growth
- Mobile
- Zero
- Fundraising
Teams supported: Consulting and ad-hoc analysis

- **Multimedia**
  - instrumentation and dashboarding
  - longitudinal data for all VE-enabled projects

- **VE**
  - data support for editor trends review;
  - longitudinal active editor retention data;
  - daily logged in users estimation

- **Strategy**
  - changes to retention guidelines;
  - log deletion/aggregation strategy

- **Legal**
  - browser support data;
  - non-default search options usage

- **Platform**
  - breakdown of API requests by client;
  - performance of SSL deployment;
  - page volume and length for Search;
  - IE6 support for logged-in users

- **Flow**
  - most active talk pages

- **Language**
  - tofu hits by language and UA

- **Grantmaking**
  - education portal traffic;
  - Brazil traffic data;
  - metrics support for IEG;
  - AE country-level geolocation
Community support

Editor trends in Portuguese Wikipedia

Portuguese Wikibooks edit dashboards

GLAM analytics (NARA pilot)

Wikistats overhaul

Various community requests @ Zurich hackathon
Outreach

**WMF Research Showcase** (3 showcases, 6 presentations) + reading groups

2 conference papers submitted
- 1 accepted at WikiSym ‘15 (AfC process & productivity)
- 1 submitted to CSCW ‘15 (MoodBar)

8 proposals submitted and accepted at Wikimania ‘15
- 5 driven by Research & Data
- 3 co-authored

Mozilla UX / Research summit; Mozilla Science Labs talk
Toolkits and documentation

- **wikipediR**: R wrapper for the MediaWiki API
- **WMUtils**: Utilities for geolookup and UA parsing of request logs
- **mediawiki-utilities**: General data processing utilities in python
- **mwoauth**: MediaWiki OAuth support for python tools
- **wikiclass**: Automated quality assessment of Wikipedia articles

Analytics infrastructure documentation on Wikitech
(data access - geolocation - hive queries)
Q1-2015 goals

R&D process
Metrics standardization
Topical research
Team support / community outreach
Process

Uplevel team process (prioritization sprints, project management)

Finetune operating model (consulting vs embedded)

Push on horizontal integration with other research teams (UX, grantmaking)
Metrics standardization: Organizational alignment

Socialize the Editor Model and Vital Signs metrics

Consistent target setting across teams

Provide metrics training
Metrics standardization: Definitions and analysis

New users
- Newly registered users
- New editors
- Productive new editors
- Surviving new editors

Community
- New active editors
- Surviving new active editors
- Reactivated editors
- Old active editors

Content
- Edits
- Uploads
- Pages
- (+ Funnel Metrics)

Curation
- Deletions
- Reverts
- Blocks

Stage 1
- 2014-Q3

Stage 2
- 2014-Q4

Stage 3
- 2015-Q1

Vital Signs MVP

https://meta.wikimedia.org/wiki/Research:Metrics_standardization
Topical research, continued

Anonymous acquisition research

Mobile apps adoption/mobile activation trends

Predictive models of editor activation/retention

Unique visitors analysis; Readership metric definitions; mobile traffic trends

Cross-wiki migration
Formal collaborations

Knowledge graph and recommender systems
(GroupLens, UMN)

Traffic data anonymization and aggregation
(Los Alamos National Laboratory)
Teams supported: Focus areas

Q1-2015

- Growth
- Mobile
- Fundraising
H1-2015

Staffing outlook