Analytics Quarterly Review

Q1 2015
Agenda

Introduction (5 mins)

1. Research and Data + Q&A (35 mins)

Break (5 mins)

2. Engineering + Q&A (35 mins)

Conclusions & Asks + Q&A (10 mins)
Q1: Top 5 platform/process improvement priorities

4. Standardize baseline metrics & dashboards

“Enable more consistent measurement of user-impacting changes for Editors, Readers and Content.”

Commitment was for Editors in Q1; Readers and Content TBA
Q1 goals: Research and Data

Primary goal - standardize baseline metrics
- define key metrics for product teams (aka the Editor Model)
- apply them to goals and instrumentation
- define key project-level participation and content metrics
- make significant progress on reader metric definitions: page views, unique clients

Secondary goals - topical research
- anonymous acquisition (Growth)
- mobile app adoption/traffic and participation trends (Mobile)
- predictive models of editor activation/survival

(includes consulting support as priorities allow)
Q1 goals: Engineering

- Implement a dashboard to **visualize key metrics** defined by the research team and usable by product teams to drive goals and instrumentation.
  - Metrics: the Editor Model, and number of daily {Active Editors, Edits and Pages Created}
  - Broken down by all 882 projects
  - Broken down by target site (desktop website, mobile website, apps)
  - With a default view showing the Rolling Monthly Active Editors for the top 7 wikis.

**Stretch Goals**

- Page Views
- Unique Visitors
Research and Data
Outline

Q1 goals review

Q2 proposed goals

staffing outlook
Q1 goals review
Primary goal: Metric standardization

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

Community
- Monthly active editors
  - new active editors
  - surviving new active editors
  - reactivated editors
  - old active editors
- Daily unique registered editors
- Daily unique anonymous editors
- Daily unique bot editors
- Daily unique page creators
- Daily unique media creators

Content
- Daily edits
  - by registered users
  - by anonymous users
  - by bot users
- Daily pages created
- Daily media created

Curation
- Deletions
- Reverts
- Blocks

https://meta.wikimedia.org/wiki/Research:Metrics_standardization#q1-2015
Primary goal: Metric standardization

- Shipped all **11 metric definitions**
  - human-readable definition
  - parameter recommendation
  - SQL specification
  (including metrics listed as *stretch goals*)
    
    #407 #408 #415 #416 #417
    #418 #419 #420 #421 #422 #423

- Performed **background analysis**
  - bot identification method
  - activity on deleted pages

See: [https://meta.wikimedia.org/wiki/Research:Daily_unique_registered_editors](https://meta.wikimedia.org/wiki/Research:Daily_unique_registered_editors)
Primary goal: Team alignment with Editor model

Set **baselines** and **annual targets** for Product teams

- **Growth**: *completed* (desktop user acquisition/activation/retention) #501 #511
- **Mobile**: *completed* (mobile user acquisition/activation) #343 #409
- **Visual Editor**: *started* *(needs additional input from Product)*
- **Multimedia**: *on hold* *(deprioritized due to MediaViewer)*
Primary goal: Traffic metric definitions

- **Pageview definition** #288
  - Draft definition: *in progress*
  - Implement a prototype using the draft definition: *completed*
  - Set down use cases to refine the draft definition: *in progress*

- **Interim unique client definition** *(stretch goal)*
  - LUCID proposal: *started* #373
    
    *implementation on hold due to technical and privacy-related concerns*

Secondary goals: Growth research

### Anonymous editor acquisition
- experimental design for 3 tests and completed analysis for 2 tests #465 #559

### Task suggestions
- research questions for task recommendations #573
- completed analysis of performance of morelike API #563
- experimental design for test #1 in production #621
Secondary goals: Mobile research

- **Mobile app analytics:** completed
  - unique clients, PVs, session data  #379

- **Mobile traffic/participation trends:** completed
  - impact of tablet switchover  #341 #356
  - mobile activation analysis  #352
  - mobile trend report @ monthly metric  #375
  - mobile trend report on Meta  #294
Secondary goals: Editor survival models

Active editor survival models

- Prediction models for 1-month, 6-month, and 12-month survival completed #437

- Generate reverts and social interaction data: completed #344 #436

- Prediction models based on new data; identify the most important variables: delayed #438
Consulting and ad-hoc analysis

Multimedia
- MV pref elicitation test design #391
- impact of MV on edit volume #370

VE
- new VE usage data #361

Strategy
- edit conflict data analysis #325

Legal
- ACLU data request #349

UX
- thumbnail size usage #363

Platform

Flow
- usage of PDF renderer #368
- impact of SSL changes on NavTiming #340

Zero
- watchlist size distribution #367

Research
- impact of 7/9 deployment #364

Community
- local edit activity in Wikipedia #381
- decline of Italian Wikipedia #431
- usage of Aragonese Wikipedia #376
Team process

- **Horizontal integration with other researchers: completed**
  - set up joint standups and weekly research group

- **Finetune operating model: in progress**
  - revised and socialized new engagement model for R&D
  - worked on a proposal to re-scope R&D

- **Uplevel team process: on hold**
  - suffering from lack of scrum master / project management support
Formal collaborations / hiring

- Onboard F/T Fundraising Research Analyst
  - completed

- Morten Warncke-Wang on task recommendations
  - started (volunteer agreement/NDA filed)

- Los Alamos National Laboratory on PV aggregation/anonymization
  - on hold because of legal and operational issues

- Opening of traffic research analyst position
  - on hold, conditional on review of Product needs; no capacity for hiring in Q1
Outreach

**Wikimania ‘14**
- 8 sessions in the main conference track
- WikiResearch hackathon
  - *WikiProject Medicine assessment, Public event standardization, Live demos, Quarry querying service*
- spin-off projects *Medicine Translation Taskforce, Wikipedia Stub Bot*

**Talks**
- Oxford Internet Institute (UK)
- INRIA (FR)
- WMDE (DE)
- ISI Foundation (IT)
- DensityDesign Lab (IT)
- University of Minnesota (US)
- University of Michigan (US)

**Showcase**
- hosted 2 research showcases with 4 talks

**Paper collaborations**
- presented *OpenSym ‘14 paper (WP article review)*
- revised and accepted *CSCW ‘15 paper (Moodbar)*
- submitted 2 *CHI ‘15 papers (geodata analysis)*
Q2 goals
Proposed Q2 goals: **Overview**

**Strategic research**
- conduct groundwork on strategic research questions, subject to prioritization
- prototype APIs / data services to support internal and external research

**Product support**
- support Product teams via consulting and self-service infrastructure
- discontinue “embedded” model

**Collaborations**
- scale up collaborations with community/external researchers to increase research scope and throughput
Proposed Q2 goals: **Strategic research**

- Understanding reader behavior
- Detecting knowledge gaps
- Value-based measurements
- Mobile micro-contributions
- Fundraising experimental strategy

*(see Appendix)*
Proposed Q2 goals: Data services

**A/B testing infrastructure** (primary goal)
provide functional requirements

**Public data sources** (secondary goal)
revert metadata, page creation log

**Prototyped labs tools** (stretch goal)
article quality classifier, revision scoring, topic classifier
Proposed Q2 goals: **Product support**

**Design and test consulting model** (primary goal)
- allow R&D team to allocate resources more flexibly
- identify high impact research opportunities from under-resourced teams
- socialize self-service / reporting infrastructure with Analytics Dev

**Hire** (primary goal)
- hire traffic research analyst with industry experience
- hire and onboard scrum master
Proposed Q2 goals: Collaborations and outreach

**Formal collaborations** (primary goal)
unblock formal collaborations stalled in Q1

**IEG advising** (secondary goal)
advise IEG for grants focused on research and technology

*wikibrain*, *notability modeling*, *editor interaction*, *revscoring*

**Outreach** (stretch goal)
Kick off research collaborations with Oxford, Stanford, Princeton, GroupLens, Northeastern, INRIA

Wiki quality hackathon @ GROUP’14; Wiki research panel @ NCA’14
Staffing outlook
Team growth

Erik Zachte
Aaron Halfaker
Oliver Keyes
Dario Taraborelli
Leila Zia
Ellery Wulczyn

Scrum Master
Traffic research analyst

2014
Q1 Q2 Q3 Q4

2015
Q1 Q2 Q3
Appendix
Strategic research: **Understanding reader behavior**

**Reader behavior metrics**
- Vetted page view definition applicable to all access methods
- Work out a proposal and privacy implications of unique client implementation
- A robust and standardised way of identifying reader sessions

**Research on traffic sources**
- Analyze where readers come from (referral analysis)

**Reader rhythms and behavioural patterns**
- Understand how people access Wikimedia projects as a function of {connection classes, device types, times-of-day}. 
Strategic research: Mobile micro-contributions

WikiGrok
Rollout and test strategy for WikiGrok

Hybrid contributions
Use cases, impact theory and technical requirements for other types of micro- or hybrid contribution
Strategic research: Value-added measurements

Definitions of value-added metrics

Vetted metric definitions of productivity, quality and importance of contribution

Fundamental research

“Who writes Wikipedia?” has been a persistent question. We’ll aim to answer it.

<table>
<thead>
<tr>
<th>Revisions</th>
<th>PWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Apples are red.</td>
<td>6</td>
</tr>
<tr>
<td>2: Apples are blue.</td>
<td>0</td>
</tr>
<tr>
<td>3: Apples are red.</td>
<td>0</td>
</tr>
<tr>
<td>4: Apples are tasty and red.</td>
<td>1</td>
</tr>
<tr>
<td>5: Apples are tasty and blue.</td>
<td>0</td>
</tr>
</tbody>
</table>
Strategic research: Fundraising experimental strategy

Robust statistical testing A/B
Incorporate a statistical test for which banner is the winner in an A/B test. Let user control the probability of making the wrong decision as a function of effect size.

Fraud detection
Learn weights and threshold for current fraud score to optimize cost from false blocking of a valid donation and charge-backs
Strategic research: Detecting knowledge gaps

Build a hierarchical topic classification of WP articles

Identify asymmetries in coverage between specific Wikipedia language editions

Identifying knowledge gaps based on user need analysis (search volume, redlink traffic, inbound links)
Outline

Q1 goals review
Systems Overview
Staff & Process
Q2
Q1 goals review
Goals: Vital Signs

- Implement a **dashboard to visualize key metrics** defined by the research team and usable by product teams to drive goals and instrumentation.

  - Metrics: the Editor Model, and number of daily Edits and Pages Created
  - Broken down by all 882 projects
  - Broken down by target site (desktop, mobile or apps)
  - With a default view showing the top 7 wikis
Stretch Goals:

- Add the metric for each project to Vital Signs dashboards
  - Page Views
    Close to a pageview definition and ability to count them in Hadoop. Still need to implement moving that data into Wikimetrics
  
- Unique Visitors
  Blocked on developing a technical proposal that honors opting out.
Systems Overview

- Vital Signs
- Wikimetrics
- EventLogging
- Refinery
## Value Proposition: Vital Signs

<table>
<thead>
<tr>
<th>For</th>
<th>Decision Makers, PMs, Grantmaking and the community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who are dissatisfied with</td>
<td>building their custom dashboards that display custom metrics</td>
</tr>
<tr>
<td>Vital Signs is a</td>
<td>dashboard</td>
</tr>
<tr>
<td>That provides</td>
<td>standardized metrics on every wiki in an easy to navigate way</td>
</tr>
<tr>
<td>Unlike</td>
<td>calling on busy researchers</td>
</tr>
</tbody>
</table>
Vital Signs: Demo
Vital Signs: Backlog

UI improvements
Setting Date Range
Interactivity
More Standard Metrics
Mobile Breakdown (by target site)
Annotations
## Wikimetrics: Value Proposition

<table>
<thead>
<tr>
<th>For</th>
<th>Researchers, Product Manager and Grantmaking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who are dissatisfied with</td>
<td>writing their custom queries, or going to researchers for data</td>
</tr>
<tr>
<td>Wikimetrics is a</td>
<td>website and platform</td>
</tr>
<tr>
<td>That provides</td>
<td>reports on cohorts and wikis using standardized metrics</td>
</tr>
<tr>
<td>Unlike</td>
<td>cron jobs on various servers</td>
</tr>
</tbody>
</table>
Wikimetrics: Accomplishments

Wikimetrics compiles the data, Vital Signs displays it.

The Editor Model metrics

Run daily & ability to backfill

Robust
Wikimetrics: Backlog

Grantmaking

Add Documentation to the website
Expand cohorts with accounts across wikis (CentralAuth)
Remove a user from a cohort
Tag cohorts

Analytics

Optimize the reporting speed
Implement remaining metrics defined by Research Team
**EventLogging: Value Proposition**

<table>
<thead>
<tr>
<th>For</th>
<th>development teams and researchers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who are dissatisfied with</td>
<td>reinventing APIs and services</td>
</tr>
<tr>
<td>EventLogging is a</td>
<td>logging platform and main the source of metrics around features and user behavior</td>
</tr>
<tr>
<td>That provides</td>
<td>curated data</td>
</tr>
<tr>
<td>Unlike</td>
<td>unstructured data sets</td>
</tr>
</tbody>
</table>

EventLogging is used by Multimedia, Growth, Mobile teams etc.
EventLogging: Backlog

Maintenance (outstanding tasks)
Sanitize User Agent in logs
Move sanitized Mobile data into LabsDB
Purging of old data automatically (per privacy policy)
Aggregate / Anonymize data to preserve

A/B testing depends on EventLogging
## Refinery: Value Proposition

<table>
<thead>
<tr>
<th>For</th>
<th>researchers and analytics developers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who are dissatisfied with</td>
<td>sampled logs, packet losses, processing limitations</td>
</tr>
<tr>
<td>Refinery is a</td>
<td>data processing ecosystems (including Hadoop)</td>
</tr>
<tr>
<td>That provides</td>
<td>scalable &amp; reliable traffic log delivery and analysis</td>
</tr>
<tr>
<td>Unlike</td>
<td>our old systems</td>
</tr>
</tbody>
</table>
Refinery: Accomplishments

25 Node Cluster is fully operational (upgraded to CDH5 & new nodes added)

Data pipeline (Kafka) is monitored for delivery and integrity
   Operational since late July (160K RPS)

Collaboration between Analytics Development and Operations teams

Turned off legacy data feeds to third parties

Current source of page view data (WebStatsCollector) is being ported to Hadoop
   First production service; eliminates 3-6% errors + other benefits

The Research team is using Hadoop regularly
Refinery: Backlog

Work jointly with Research Team on Unique Clients

Make PageViews per project available to Vital Signs (in LabsDB)

Implement ETL with the new PageView definition

Compile PageViews per page and make the data public
Staff & Process
Who we are

Dan Andreescu
Nuria Ruiz
Kevin Leduc
Christian Aistleitner
Andrew Otto
Jeff Gage
Marcel Ruiz Forns
Scrum Master

2014
Q3
Q4

2015
Q1
Q2

2 Open Reqs
Process & Metrics: Agile Scrum & Velocity on Vital Signs

2 week sprints

Weekly tasking & estimating meeting

Retrospectives

Backlog managed with Scrumbugs & Etherpad

Release Planning - reaching our goal
Q2 Goals
Preliminary Q2 Goals: Enabling Infrastructure

Depends on selection of top engineering priorities

**Wikimetrics**

Release Wikimetrics with the following features: CentralAuth Cohorts, “Delete User”, and Tagging (not sharing)

**EventLogging**

Fulfill Privacy Policy Commitments (Sanitize User Agent in Logs/Purge old data)
Support AB Testing (?)

**Refinery**

Make PageView data public for both projects (wikis) and individual pages
Q2 Goals: Out of Scope

Vital Signs
   Major UI work
   Mobile Breakdowns

Refinery
   Putting PageView data in a public database or exposing an API
Q2 Goals: Visualization

What is the future of Limn?
- Still maintained and supported by Analytics Team
- No new features/development

Vital Signs taught us Limn tried to do too much

Dual Strategy
- Vital Signs stack for public dashboards
- Development spike on Open Source exploration/visualization tool (Saiku)
Appendix
Unique Clients: Proposed Milestones

1. Technical solution is designed
2. Community has given feedback (RfC)
3. Mobile and Desktop teams have implemented opt-out
4. Hadoop starts counting
Q2 Asks

Continue with space and support

Recruiting/staffing

Work with us to bring data informed perspectives into our goals and activities