The UserMetrics API

Measuring participation in Wikimedia projects

Dario Taraborelli
Wikimedia Foundation

Wikimania 2013, Hong Kong, 9 August 2013

Overview

Rationale

Background

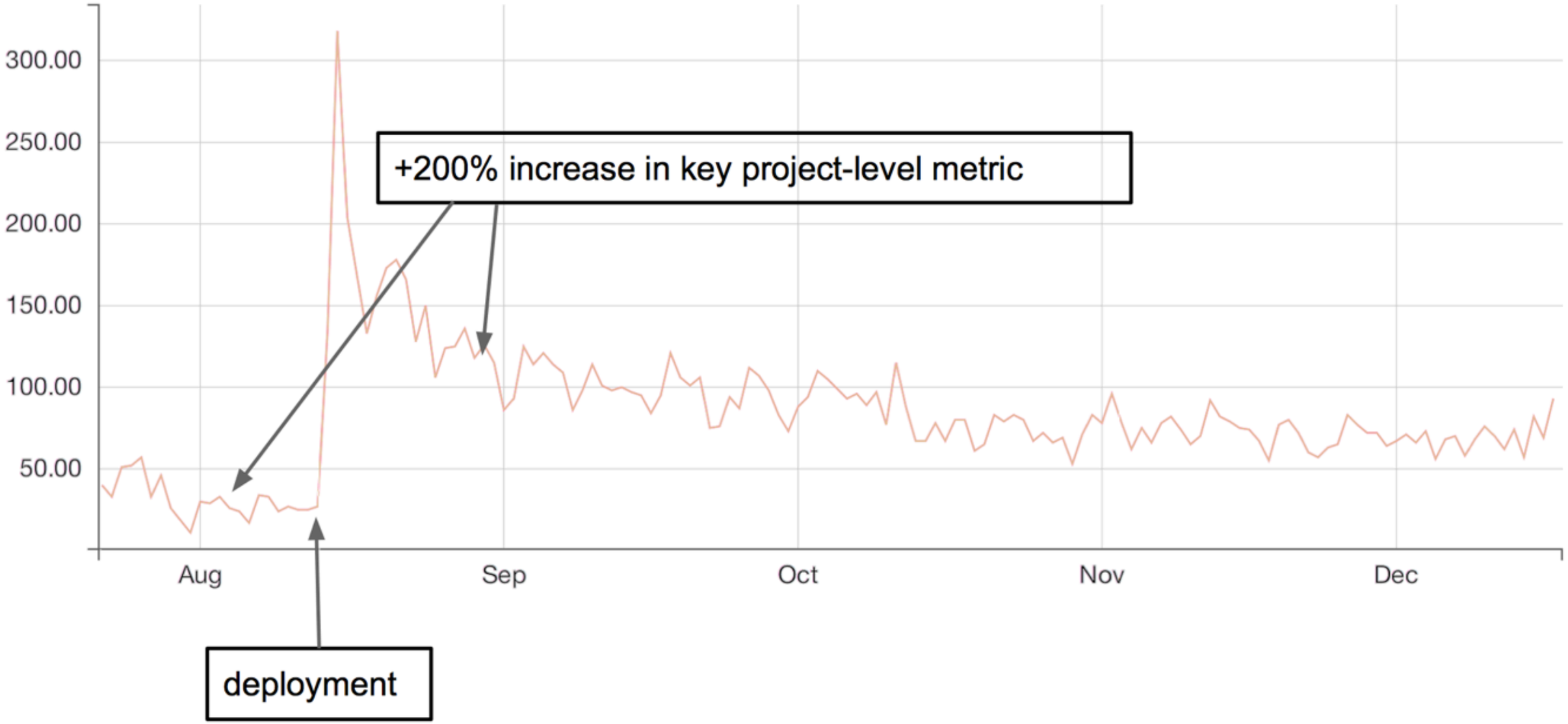
Introducing the UserMetrics API

The next steps: Wikimetrics

How do we determine if a {program, experiment, new feature} we run to engage new editors and increase participation is actually working?

How do we measure the relative success of these interventions in driving participation?

Rationale



The problem: quasi-experiments

ACTIVE EDITOR

RETAINED EDITOR

NEW WIKIPEDIAN

HIGH-QUALITY ARTICLE

SURVIVING EDIT

The problem: metric definitions

User cohorts

Research:Metrics/edit sessions

< Research:Metrics

An **edit session** represents a sequence of edits performed by an editor during a "session" of wiki-work. Assuming that editors tend to work on the encyclopedia in bursts that appear in the log as a quick succession of edits to articles and other pages, then the beginning and end of their work session can be approximated by the first and last edit recorded in such a sequence.

Contents [hide]

1 Definition
2 Construction
  2.1 Session duration
3 Limitations
  3.1 The arbitrary value of
4 Usage
5 References

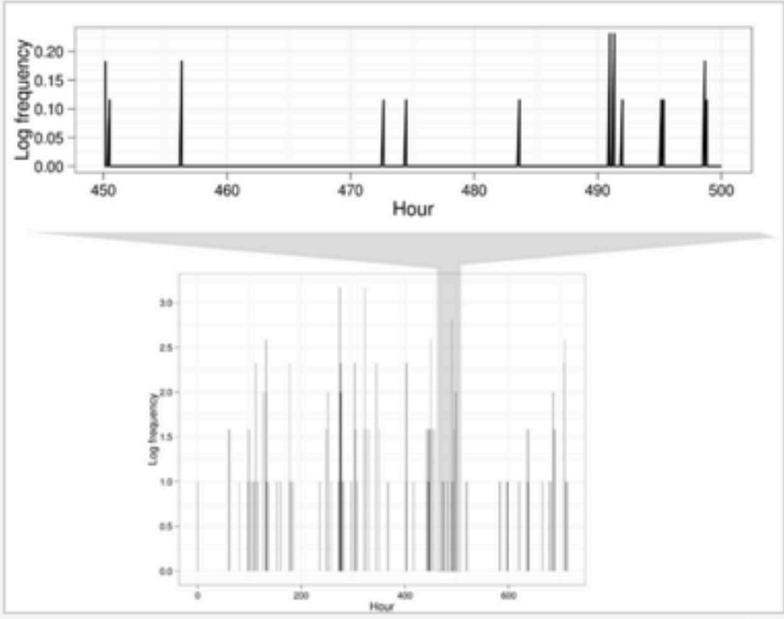


Figure 1. Empirical observations of edit sessions. Bursts of edit activity over time (March 2012) are plotted for en:User:Jimbo_Wales.

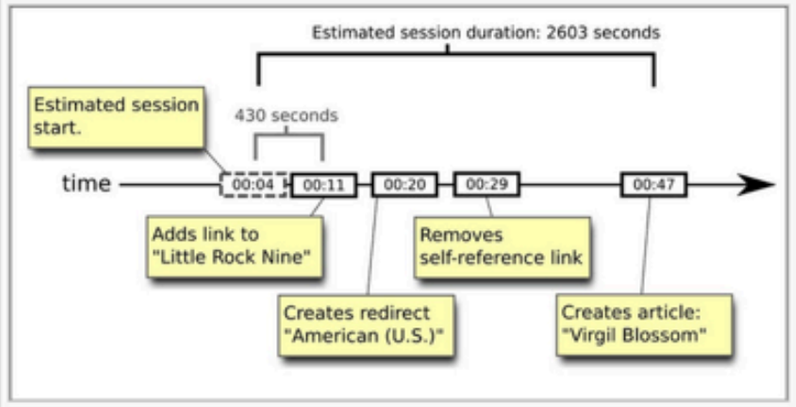


Figure 2. Example: Toby Bartels' edit session. Edits made by "Toby Bartels" are plotted and annotated over time for a session he completed on September 4th, 2010. The estimated session start time is plotted at 430 seconds before the user's first edit.

Definition [edit]

An edit session is a sequence of edits made by an editor where the difference between the time at which any two sequential edits occurred is less than t. In other words, a set of edits S is an edit session if:

$$\forall e_1, e_2 \in S : I(e_1) = I(e_2) - 1 \rightarrow T(e_2) - T(e_1) \leq t$$

Where:

- $I(e)$ = the index of edit e in a sequence of edits
- $T(e)$ = the time at which edit e occurred in seconds
- t = the maximum time between edits (commonly set to one hour)

[[m:Research:Metrics/edit sessions]]

Metric research

Cohort analysis - 1

Define a **cohort** from a list of editors sharing a meaningful attribute

Examples

- participants in an experimental condition of an A/B test
- students enrolled in a Global Education course
- members of a Wikiproject
- new editors reaching 10+ edits in their first month since registration
- users registered on mobile devices
- participants in an editathon

Cohort analysis - 2

Measure the {quality, productivity, retention} of these users via a standard set of metrics

survival(t)

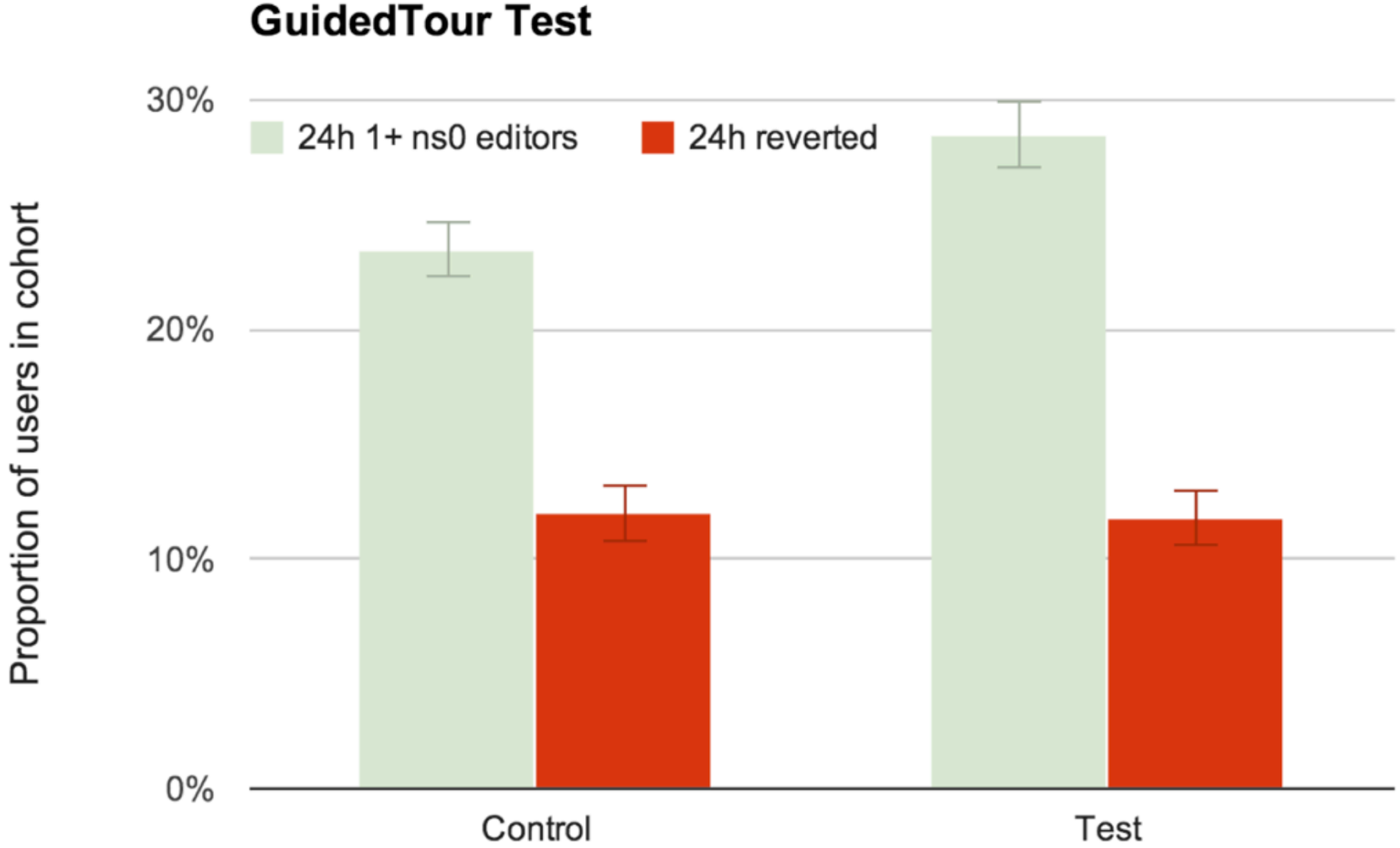
Boolean measure of an editor retention. Editors are considered as "retained" or "surviving" if they continue to edit after a time t since a reference event (typically: account registration time).

threshold(t,n)

Boolean measure of an editor's level of activity. This metric measures if an editor reaches some threshold of activity (e.g. n edits, words added, pages created) within time t of a given reference event (typically: account registration time).

Cohort analysis - 3

Compare a cohort to a (meaningful) baseline using a standard set of metrics



User Metrics API

Welcome to the Wikimedia Foundation's user metrics API homepage. This API allows you to select a set of users, also known as a "cohort" (for example, all users who signed up via the Thank You campaign), select a metric to be computed for each of these users (for example, their 7-day revert rate) with optional parameters (for example, a registration time range) and retrieve the response in JSON format.

You can also compute a single, aggregate value for the cohort (like the mean revert rate for these users)

Learn more

List of cohorts

dan_test Next

List of available metrics

namespace_edits Next

Additional parameters

- Date start - date_start
- Date end - date_end
- Aggregator function - aggregator

Choose a Cohort.
Check the Job Queue.
See all generated requests.

UserMetrics

namespace_edits

revert_rate

time_to_threshold

blocks

threshold

bytes_added

survival

edit_rate

pages_created

live_account

UserMetrics

UserMetrics

Formally defined metrics, backed by extensive research and crash-tested in the WMF Editor Engagement program.

Support for cohorts from any Wikimedia project.

Fast extraction of computationally intensive metrics for cohorts of thousands of users.

Individual responses or aggregate data using various aggregator methods. Supports time series.

Extensible and fully parametrized (metrics can be fine-tuned, default parameters overridden as needed)

```
{
 "time_of_response": "2013-04-16 17:18:32",
 "cohort_last_generated": "2013-04-16 17:18:26",
  "metric": "threshold",
  "cohort": "test2",
  "type": "raw",
  "group": "REGISTRATION",
  "aggregator": "None",
  "namespace": [
    0
  ],
  "n": 1,
  "t": 24,
  "project": "enwiki",
  "slice_size": 24,
  "datetime_start": "2010-01-01 00:00:00",
  "datetime_end": "2020-01-01 00:00:00",
  "header": [
    "user_id",
    "has_reached_threshold"
  ],
  "data": {
    "13234584": [
      0
    ],
    "15972135": [
      0
    ],
    "15972203": [
      1
    ]
  },
}
```


raw results

```
{
  "time_of_response": "2013-04-16 22:19:31",
  "cohort_last_generated": "2013-04-16 22:19:12",
  "metric": "threshold",
  "cohort": "test2",
  "type": "aggregator",
  "group": "REGISTRATION",
  "aggregator": "proportion",
  "namespace": [
    0
  ],
  "n": 1,
  "t": 24,
  "project": "enwiki",
  "slice_size": 24,
  "datetime_start": "2010-01-01 00:00:00",
  "datetime_end": "2020-01-01 00:00:00",
  "header": [
    "total_users",
    "threshold_reached",
    "rate"
  ],
  "data": [
    3,
    1,
    0.3333333333333333
  ],
}
```


aggregate results

```
{
  "time_of_response": "2013-04-23 17:44:30",
  "cohort_last_generated": "2013-04-23 17:44:24",
  "metric": "threshold",
  "cohort": "test2",
  "type": "time_series",
  "group": "REGISTRATION",
  "aggregator": "proportion",
  "namespace": [
    0
  ],
  "n": 1,
  "project": "enwiki",
  "t": 24,
  "slice_size": 720.0,
  "datetime_start": "2011-12-01 00:00:00",
  "datetime_end": "2012-03-01 00:00:00",
  "header": [
    "timestamp",
    "total_users",
    "threshold_reached",
    "rate"
  ],
  "data": {
    "2011-12-01 00:00:00": [
      2,
      1,
      0.5
    ],
    "2011-12-31 00:00:00": [
      0,
      0,
      0.0
    ],
    "2012-01-30 00:00:00": [
      0,
      0,
      0.0
```


time series

Wiki Metrics

Welcome to the Wikimedia Foundation's Wikimetrics homepage. This API allows you to select a set of users, also known as a "cohort" (for example, **all users who signed up via the Thank You campaign**) , select a metric to be computed for each of these users (for example, how many bytes they've added) with optional parameters (for example, a **time range**) and retrieve the response in JSON or CSV format.

You can also compute a single, aggregate value for the cohort (like the **mean** revert rate) .

Learn More Analyze

Your cohorts

- dartar
- dartar_test

List of metrics

- Bytes Added
- Edits

Your reports

- **07/25/2013** - Bytes Added - dartar
- **07/25/2013** - Edits - dartar
- **08/01/2013** - Bytes Added - dartar
- **08/01/2013** - Edits - dartar
- **08/01/2013** - Bytes Added - dartar_test
- **08/01/2013** - Edits - dartar_test

WikiMetrics

Functionality ported from UserMetrics

Improved UX

Redesigned backend and job management

Access control and user authentication

Abstraction

Get involved

UserMetrics (private)

end-user documentation

https://www.mediawiki.org/wiki/UserMetrics/Guide

source code

http://github.com/wikimedia/analytics-user-metrics

WikiMetrics (public, under development)

home

http://metrics.wmflabs.org

source code

http://github.com/wikimedia/analytics-wikimetrics

mailing list

https://lists.wikimedia.org/mailman/listinfo/wikimetrics

thank you

Dario Taraborelli

The UserMetrics API: Measuring participation in Wikimedia projects

dario@wikimedia.org
@readermeter