



Data Sources

Program Evaluation and Design Workshop

Sunday June, 22, 2013

Budapest, HU

Data Sources

Existing Data (Secondary)

- Attendance records
- Intake forms: demographics, etc.
- Staff or participant journals or observations
- Existing participant satisfaction data
- Meeting minutes
- Previous reports
- Other program documentation

Data Sources

Primary Data Collection

- Program participants
- Program staff
- Program administrators
- Community members
- Program collaborators/partners
- Others who interact with participants or program in some way

Types of Data

Qualitative data

- Data described in terms of quality, as opposed to "quantity"
- Often obtained through asking open-ended questions
- Most useful when:
 - you would like information in people's own words
 - the questions you are asking have too many possible answers for you to know/list

Types of Data

Quantitative data

- Data described in terms of a quantity or number
- Quantitative data is collected through closed-ended questions
- Most suitable for collecting numeric data such as:
 - Age, hours, staff size, number of users, etc.
 - Self-reported behaviors collected quantitatively on a scale
 - Self-rated attitudes and judgments constrained to a set of choices

Data Collection Methods

- Surveys
 - Satisfaction with program
 - Pre-post tests (Knowledge, behavioral self-report, specific instruments, etc.)
- Individual Interviews
- Focus Groups and Facilitated Dialogues
- Journals and Reflections
- Physiological measures
- Observations, Tracking logs, and Records

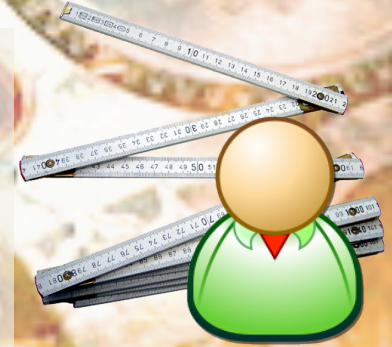
User Metrics API (UMAPI): An Introductory Overview

- UserMetrics API is the name of a platform developed by the Wikimedia Foundation to measure user activity based on a set of standardized *metrics*.
- Using this platform, a set of key *metrics* can be selected and applied to a *cohort* of users to measure their overall productivity.

Cohort Metrics?

Metric - well-defined values or sets of values that can be computed for any user registered in Wikimedia projects

Cohort - set of users sharing one or more property or attribute—the time of account creation, for example, or participation in an outreach event or experimental group.



Cohort Metrics?

Cohort Metrics help us understand activity and behavior of user groups.

From quality, quantity and type of user contribution, to how well our editors are retained.

UMAPI Purpose

- **Standardization** Standard set of metrics to:
 - permit everyone to have the same understanding
 - provide comparability across efforts within the organization to examine impacts.
- **Workflow (Access)** Streamlined process to obtain and analyze data needed to evaluate programs, projects, and initiatives.

Metrics Standardization

- Can be applied to any Wikimedia project
- Unambiguous and consistent measurement
- Project level comparability
(to measure the success of an experimental treatment or outreach initiative)
- Organizational level comparability
(to compare the impact of programs and projects across the organization)

Workflow

- **Define cohorts**

Cohorts can be defined by specifying custom lists of usernames/userIds.

- **Measure**

The quality, productivity, or retention of these cohorts can be assessed via a standard set of metrics.

- **Compare**

Cohorts can be compared against each other or against a baseline.

Adding Custom Cohorts

- **Prepare UserIds list:**

Username, project
Frank Schulenburg, en
Sarah Stierch, en
Janstee, en

Note: Most Users Do Not Have Underscores In Their Names
If_This_Fails_try_replacing_underscores_with_spaces,en

- **Save as csv file**
- **Follow “Upload a new cohort” link from “Cohort” selection screen**

Adding Custom Cohorts

User Metrics API

Cohorts

Metrics

Jobs

About

Contact

Cohort Name

(automatically checks availability)

Wiki Project

(select one or specify in file)

CSV File

No file chosen

Upload CSV

****See Guide 1: Adding Cohorts****



Demonstration

FORMER SLIDES REMOVED

*NEW DEMO SLIDES TBD with release of
“WikiMetrics” late July*

Volume Metrics

bytes_added: indicates the amount of content an editor has added, removed, and modified within a given time period.

The system returns the net bytes contributed, bytes added (positive), absolute bytes contributed, bytes removed (negative), and edit count (e.g., [100, 150, 200, -50, 2]) for your cohort in the specified time period.

Other Potential Metrics

- **Retention metrics**

Help measure whether an editor is active and how active an editor is during or after a given timespan: **threshold**, **survival**, **live_account**.

- **Volume metrics**

Help measure the quantity of an editor's wiki work: **edit_rate**, **bytes_added**, **time_to_threshold**, **pages_created**.

- **Quality of contribution metrics**

Measure the LACK of quality in an editor's work: **revert-rate**, **blocks**.

- **Type of contribution metrics**

Measure the diversity of an editor's wiki work: **namespace_edits**

Aggregators and associated metrics

- **sum** Sum of the metric values returned for every user in the cohort.
- **mean** Mean of the metric values returned for a cohort.
- **std** Standard deviation of the metric values returned for a cohort.
- **median** Median of the metric values returned for a cohort.
- **min** Minimum metric value returned for a cohort.
- **max** Maximum metric value returned for a cohort.
- **dist** Distribution values for sum, mean, std, median, min, max, returned as an array for each cohort.



SLIDES REDACTED

Details changing and new demonstration slides will be created for the new “Wikimetrics” to be released at the end of July

DOs and DON'Ts for UMAPI

DO:

- Get consent to an opt-in before collecting any information. ***See Example Sign-in Sheet for consent***
- Tell participants:
 - what you are collecting (i.e., User tag)
 - what you plan to use it for (i.e., tracking user activity metric),
 - what level of confidentiality they should expect, and
 - you will be transferring this data internationally to WMF for processing in the US.
- Keep all cohorts above a minimum size (i.e., No cohorts of 1)

DOs and DON'Ts for UMAPI

DO NOT collect information/base cohort membership on categories that are protected under the EU Data Protection Directive. You should NOT collect:

- Real Names
- Date of birth (nor age if it can be tied to a specific user)
- Gender, sexual orientation, racial or ethnic origins (if it can be tied to a specific user)
- Medical conditions or disabilities
- Addresses or Phone Numbers
- Passwords
- Social security or other governmental identification numbers
- Political affiliations or religion

DOs and DON'Ts for UMAPI

DO NOT:

- Release or publish per-user information - only publish aggregated numbers
 - Publish the membership of the cohorts (i.e., id tags of who is in what cohort)
 - Use/inspect/publish the cohorts, or data about the cohorts, created by other groups
 - Use for personal use – Access to the UMAPI is for movement or organization-related purposes only
- 