Descending Mount Everest
Steps towards applied Wikipedia research

Dario Taraborelli
Wikimedia Foundation

WikiSym 2013, Hong Kong, 6 August 2013
Your recent editing history at Jonathan Frid shows that you are currently engaged in an edit war. **Being involved in an edit war can result in your being blocked from editing**—especially if you violate the three-revert rule, which states that an editor must not perform more than three reverts on a single page within a 24-hour period. Undoing another editor's work—whether in whole or in part, whether involving the same or different material each time—counts as a revert. Also keep in mind that while violating the three-revert rule often leads to a block, you can still be blocked for edit warring—even if you don't violate the three-revert rule—should your behavior indicate that you intend to continue reverting repeatedly.
Editing Conservation in Italy

You are not logged in. Your IP address will be publicly visible if you make any edits. Please log in or create an account to have your edits associated with a user name, among other benefits.

Content that violates any copyrights will be deleted. Encyclopedic content must be verifiable. Work submitted to Wikipedia can be edited, used, and redistributed—by anyone—subject to certain terms and conditions.

{{TOCright}}
"Conservation in Italy"

*[[List of extinct and endangered species of Italy]]

==Protected areas==
*[[List of National Parks of Italy]]

==Conservation organisation==
*[[LIPU|Lega italiana protezione uccelli (LIPU)]] (Italian League for Bird Protection)
*[[Italia Nostra]]
This is your **last warning**. The next time you **vandalize** Wikipedia, you may be **blocked from editing** without further notice.
extensible
Scaling our Everest: Wikipedia Studies I

- Technology-Mediated Contributions: Editing Behaviors Among Judd Antin, Coye Cheshire, Oded Nov

- Conflict, Criticism, or Confidence: An Empirical Examination of Benjamin Collier, Julia Bear

- From Individual Minds to Social Structures: The Structuring of Process Yiftach Nagar

- Classroom Wikipedia Participation Effects on Future Intentions Cliff Lampe, Jonathan Obar, Elif Ozkaya, Paul Zube, Alcides Velasque
descending Mount Everest
Unique Visitors per Region (comScore)

popularity
participation

- New users
  - Wikipedia: 10,000
  - Github: 15,000

- New entries created
  - Wikipedia: 30,000
  - Github: 150,000

- Entries modified
  - Wikipedia: 75,000
  - Github: 150,000

Legend:
- Red: Github
Wikipedia in scholarly research
shifting attention

New research areas

gender gap, breaking news collaboration

Fast-growing topics

Wikipedia and higher education

Steady growth

Wikipedia as a corpus


N. Jullien (2012) What we know about Wikipedia. A review of the literature analyzing the project(s).
John Riedl
building artifacts to change the world
Task recommendation

**SuggestBot**

Vandalism detection

**Stiki, ClueBot**

Trending topic analysis

**Wikipedia Live Monitor**

Good-faith newbie detection

**Snuggle**
“Hello, I need a sample of 500 complete responses from active English Wikipedia administrators by the end of this week”
policy change is hard
software integration is even harder
Equipment
TODAY

Mary Dunlap and 1 other posted on your talk page.
"Great article on Tam Valley!: new section"
Wednesday at 09:40

Mary Dunlap thanked you for your edit on Talk:Tamalpais Valley
"Does anyone have references about Coyote Ridge?: new section"
Wednesday at 09:39

Fabrice Florin mentioned you on Talk:Tamalpais Valley
"Here are some tips on Coyote Ridge, to improve your Tam Valley Page."
Wednesday at 09:31

Kaldari reviewed a page you started: Tamalpais Valley
Wednesday at 09:36

YESTERDAY

Wikipedia is a free encyclopedia written by people like you. Get started by making your first edit!
30 April

Welcome to Wikipedia, Jane Finch! We're glad you're here.
30 April

Experimental toolkit 1. notifications
Experimental toolkit: 2. guiders
Create account

Enter your information below.

Username
Enter your username

Password
Enter a password

Confirm password
Enter password again

Email address (optional)
Enter your email address

Security check

Wikipedia is made by people like you.

630,432,841 edits

4,297,352 articles

125,702 recent contributors

Create your account

Experimental toolkit: 3. campaigns
**Experimental toolkit**

4. instrumentation

<table>
<thead>
<tr>
<th>description</th>
<th>Logs generic events related to editing activity. All events are logged client-side.</th>
</tr>
</thead>
</table>
| version     |**version**

<table>
<thead>
<tr>
<th>type</th>
<th>&quot;integer&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>required</td>
<td>true</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>description</th>
<th>Version number. This field can be used to track changes in instrumentation that affect the nature of the data collected.</th>
</tr>
</thead>
</table>
| enum        | 0  "edit-link-click"

1  "edit-source-link-click"

2  "section-edit-link-click"

3  "section-edit-source-link-click"
Wikimedia Tool Labs
### Data Dumps

**Homepage**
Dumps of all WMF projects for backup, offline use, research, etc.
- Wiki content, revisions, metadata, and page-to-page and outside links
- XML and SQL format
- once/twice a month
- large file sizes

### API

**Homepage**
The API provides direct, high-level access to the data contained in MediaWiki databases through HTTP requests to the web service.
- Meta info about the wiki and logged-in user, properties of pages (revisions, content, etc.) and lists of pages based on criteria
- JSON, WDDX, XML, YAML and PHP’s native serialization format

### Toolserver

**Homepage**
Toolserver allows you to query a copy of the database (with some lag).
- acts as a standard web server hosting web-based tools
- command line tools
- account required

### IRC Feeds

**Homepage**
Live Recent changes feeds on IRC channels.
- Changes shown automatically as they happen.
- Feeds for each wiki in a separate channel.
- Filtered feeds available with cloak

### Pageview Stats

**Homepage**
Raw pageview dumps (not unique hits) based on squid server logs.
- Project, title of the page, number of requests, size of the content
- Delimited and JSON
- Aggregated hourly

### WikiStats

**Homepage**
Reports in 25+ languages based on data dumps and server log files.
- Unique visits, page views, active editors and more
- Intermediate CSV files available.
- Graphical presentation.
- Monthly

### DBpedia

**Homepage**
DBpedia extracts structured data from Wikipedia, allows users to run complex queries and link Wikipedia data to other data sets.
- RDF, N-triplets, SPARQL endpoint, Linked Data
- billions of triplets of info in a consistent Ontology

### DataHub

**Homepage**
A collection of various Wikimedia-related datasets.
- smaller (usually one-time) surveys/studies
- dbpedia lite, DBpedia-Live and others
- EPIC/Oxford quality assessment

---

**Data sources and APIs**
Research dissemination
WMF Product

**VisualEditor**
Complete the rollout. Make it faster, more efficient, more intuitive. Fix bugs. Get adoption to 100%. Support real-time collaboration.

**Growth**
Increase the first-edit rate of new registered users. Improve conversion along the IP->registration funnel.

**Flow**
Redesign the UX of collaboration. Fix the powerful but deeply flawed talk page system.

**Mobile**
Support mobile editing. Design microtasks optimized for mobile devices.

**Media**
Build a new media contribution workflow, integrated with VisualEditor.

**Language**
Content translation as a new participation vector.
A new social contract for Wikipedia research?
thank you

Image credits
Mount Everest from Kalapatthar, by Pavel Novak (CC BY)
John T. Riedl, by Nicole Holdorph (CC BY SA)
Argentina - Mt Tronador Ascent - 33 - descending among the crevasses, by McKay Savage (CC BY)
Le Tarbesou, by Jean-Baptiste Bellet (CC BY)

Dario Taraborelli
Descending Mount Everest: Steps towards applied Wikipedia research

dario@wikimedia.org
@readermeter