

# Event Logging

**Analytics for Wikimedians**  
**User:Strainu**

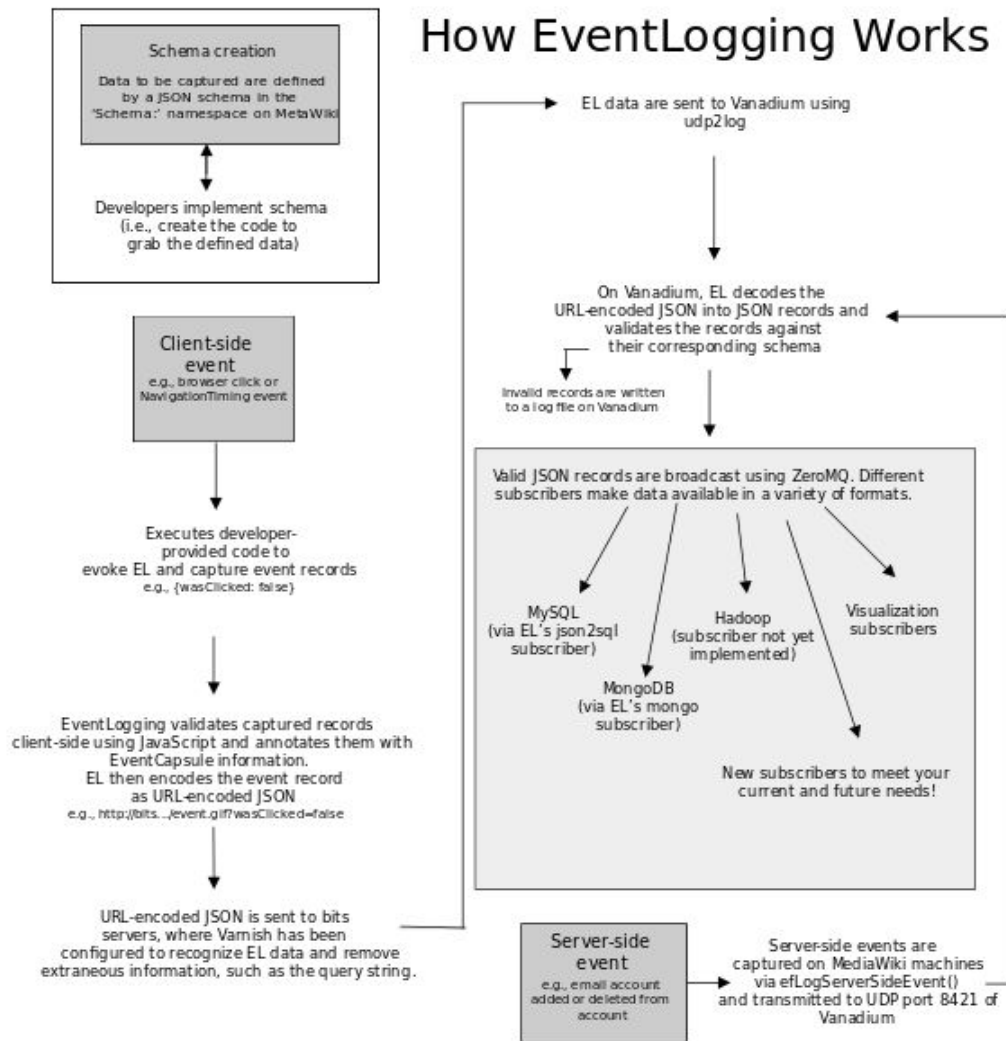
# What does it do?

X1	2	6
X2	3	4
X3	3	8
X4	4	7
X5	6	2
X6	6	4
X7	7	3
X8	7	4
X9	8	5
X10	7	6

# But, how?

Follow the grey boxes:

- create the schema
- log the event
- consume the data



A red square with a white border containing the text "Let's dig into it!".

**Let's dig into it!**

# Create the schema

Create a json schema file on **meta**, in the *Schema*: namespace

- Only pages on meta are accepted; no Vagrant page, no Wikipedia
- The revision id is important! Each revision has a different table in the MySQL backend
- Example is part of `[[m:Schema:GettingStarted]]`

properties	task	<table><tr><td>type</td><td>"string"</td></tr><tr><td>enum</td><td><table><tr><td>0</td><td>"gettingstarted"</td></tr><tr><td>1</td><td>"returnto"</td></tr></table><p>OBJECT, 2 PROPERTIES</p></td></tr><tr><td>description</td><td>"To delineate the two funnels for edit, submit, and save actions"</td></tr></table> <p>OBJECT, 3 PROPERTIES</p>	type	"string"	enum	<table><tr><td>0</td><td>"gettingstarted"</td></tr><tr><td>1</td><td>"returnto"</td></tr></table> <p>OBJECT, 2 PROPERTIES</p>	0	"gettingstarted"	1	"returnto"	description	"To delineate the two funnels for edit, submit, and save actions"
	type	"string"										
	enum	<table><tr><td>0</td><td>"gettingstarted"</td></tr><tr><td>1</td><td>"returnto"</td></tr></table> <p>OBJECT, 2 PROPERTIES</p>	0	"gettingstarted"	1	"returnto"						
0	"gettingstarted"											
1	"returnto"											
description	"To delineate the two funnels for edit, submit, and save actions"											
	articleId	<table><tr><td>type</td><td>"integer"</td></tr><tr><td>description</td><td>"Internal ID (wgArticleId) of the article with which the user is tinkering ('edit', 'complete', etc. actions)"</td></tr></table> <p>OBJECT, 2 PROPERTIES</p>	type	"integer"	description	"Internal ID (wgArticleId) of the article with which the user is tinkering ('edit', 'complete', etc. actions)"						
type	"integer"											
description	"Internal ID (wgArticleId) of the article with which the user is tinkering ('edit', 'complete', etc. actions)"											
	revId	<table><tr><td>type</td><td>"integer"</td></tr><tr><td>description</td><td>"Revision ID of the article with which the user is tinkering ('edit', 'complete', etc. actions)"</td></tr></table> <p>OBJECT, 2 PROPERTIES</p>	type	"integer"	description	"Revision ID of the article with which the user is tinkering ('edit', 'complete', etc. actions)"						
type	"integer"											
description	"Revision ID of the article with which the user is tinkering ('edit', 'complete', etc. actions)"											

# Log the data

- You can log data from wherever you can access the beacon
  - Theoretically, there are APIs for JS and PHP
  - In reality, I had to write my own function
- More data is added silently, see [\[\[m:Schema:EventCapsule\]\]](#)
  - All the “identifiable” data is there: wiki, user-agent etc.
- Once you know what you want to log, a new event can be added in under an hour

```
var logEvent = function(schema, revision, event) {
    var payload = {
        schema: schema,
        revision: revision,
        wiki: 'metawiki',
        event: event
    };

    var url = 'https://meta.wikimedia.org/beacon/event'
+ '?' + encodeURIComponent(JSON.stringify(payload));

    document.createElement( 'img' ).src = url;
};

logEvent( 'DiacriticsVisibility', 15606222, {"issues":
issues, "country": window.Geo.country} );
```

# Profit!

Well, at least if you write your own consumer.

If not, you will probably have a database table which only Analytics people can access due to privacy concerns.

To get the data, you need to sign an NDA or use [\[\[wikitech:Analytics/Reportupdater\]\]](#)



# Why haven't I heard of EL sooner?

- Perhaps you have: **EL is not a secret.**
- Blame the NDA: Some people might be discouraged by the lack of access to the results without support from an employee
- Easy input, hard output: it's relatively easy to configure a probe based on available examples with minimal programming knowledge, but you're on your own when interpreting the results



# Data privacy for EL

- We need to follow the [WMF Privacy Policy](#), but others don't
- The more data we collect, the easier it is to reconstruct user behavior if someone subpoenas (or steals) that data
  
- Checks were put in place to limit access to the data
- There is also a retaining policy - most raw data is deleted after 3 months

**Thank you!**



A large red square with a white border, centered on a white background. The text "Bonus slides" is written in white inside the square.

**Bonus slides**



# Documentation

- <https://www.mediawiki.org/wiki/Extension:EventLogging#Documentation> (incomplete and confusing)
- [#wikimedia-analytics](#) (friendly and knowledgeable :)
- <https://ro.wikipedia.org/wiki/MediaWiki:DiacriticsPoll.js> (my implementation; the magic is in the *logEvent* function)

# Image attribution

- <https://commons.wikimedia.org/wiki/File:Kmedoidt1.jpg> by User:Shrikantnangare, PD
- <https://commons.wikimedia.org/wiki/File:EventLogging2.svg> by User:Kmenger, CC-BY-SA-3.0
- [https://commons.wikimedia.org/wiki/File:EventLogging\\_Schema.png](https://commons.wikimedia.org/wiki/File:EventLogging_Schema.png) by DarTar, CC-BY-SA-3.0
- [https://commons.wikimedia.org/wiki/File:RomanianTrafficSign\\_C01-NoEntry\\_2011.svg](https://commons.wikimedia.org/wiki/File:RomanianTrafficSign_C01-NoEntry_2011.svg) by Alex:D, CC-BY-SA-3.0
- Youtube movie is available on commons:  
[https://commons.wikimedia.org/wiki/File:EventLogging\\_Workshop.webmhd.webm](https://commons.wikimedia.org/wiki/File:EventLogging_Workshop.webmhd.webm) by Chip Deubner, CC-BY-SA-3.0