

Implementing web analytics on www.ietf.org

2019-05-21

Introduction

Website analytic packages are commonly used to understand and improve how visitors make use of web pages with the aim of improving the effectiveness of a site in serving its intended purposes. This information includes:

- how the site is being used,
- through which channels visitors are finding the site,
- whether specific content is providing the intended benefit, and
- what difficulties visitors encounter in completing tasks.

There is currently no accurate data reported about visits to the www.ietf.org website.¹ Therefore, it is not possible currently to understand things such as: which pages are most commonly visited, which paths visitors travel to find IETF meeting registration pages, or whether introductory information, such as tutorials, leads to further exploration of website content. The result is a lack of information upon which to base improvements to the www.ietf.org website.

Existing IETF participants are able to use knowledge about existing channels, such as mailing lists, meetings, or other IETF channels, to provide website feedback. However, others, particularly *potential* IETF participants who may visit the website but are not otherwise engaged in the IETF, likely do not have ready access to those feedback channels.

This document outlines a proposal for implementing website analytics on the portions of the www.ietf.org website administered via the Wagtail CMS. Once it has been reviewed and approved, an implementation plan will be developed. This plan is explicitly scoped to www.ietf.org. Experience gained via www.ietf.org may be applied when considering whether or not to implement web analytics on sites such as datatracker.ietf.org and mailarchive.ietf.org, while taking into account different expectations and requirements that accompany those websites.

¹Information had been previously published at <https://www.ietf.org/usedata/> but, because of the use of a CDN, it no longer reflected actual visits to www.ietf.org.

Considerations

To be consistent with the existing practices of the IETF and the specific requirements developed for the latest redesign of www.ietf.org, website analytics must be implemented to:

- limit data being collected to that needed to serve specific identified purposes,
- not require the use of web cookies, and
- not impede the use of the website via browsers that do not have javascript enabled.

Furthermore, the details of how analytics are implemented and how that data is used to improve the website should be available for the IETF community to review and comment on before implementation, and available for review by any visitor to understand after implementation.

Proposal

This proposal is explicitly scoped to include only content on www.ietf.org and maintained in the Wagtail CMS. Any parts of www.ietf.org which are not maintained in Wagtail, and other IETF websites such as datatracker.ietf.org, mailarchive.ietf.org, and tools.ietf.org, are not included in the scope of this proposal. The set of specific data to be reported is:

- overall number of visitors;
- views per webpage;
- time spent on pages;
- traffic sources; and
- aggregated visitor profiles (including OS, browser, and primary languages)
- visitors' paths through the site (including time spent on webpages, as well as entry and exit pages).

After considering several options for implementing analytics, the choices were narrowed to self-hosted analytics packages. After consulting with the IETF Tools Team, the proposal is to deploy “[django-analytical](#)” application in conjunction with the [Matomo On-Premise](#) analytics package.

[Matomo](#) provides a broad range of “out of the box” functionality, with [Matomo \(Piwik\) JavaScript Tracker](#) tracking more information than is needed for the purpose of improving the www.ietf.org website. Matomo thus will be configured to

collect only the following information:

- **IP addresses**
IP addresses are used to associate particular actions within a visit. Matomo's IP address anonymization feature will be used to disassociate visits from individuals.
- **Date and time of requests**
The timing of requests are used to understand when particular pages and resources are requested, and to sequence requests made as part of a visit.
- **Title of the page being viewed (Page Title)**
Web page title information aids in reporting web page views. No page titles in the www.ietf.org website include information that can be used as personal identifiers.
- **URL of the page being viewed (Page URL)**
URLs indicate which pages or resources are viewed, and how many times they are viewed. No URLs in the www.ietf.org website include information that can be used as personal identifiers.
- **URL of the page that was viewed prior to the current page (Referrer URL)**
Referrer URLs help understand the path used to navigate through the site during a particular visit.
- **Files that were clicked and downloaded ([Download](#))**
This feature allows Matomo to track requests for website resources that are not web pages. An example of this is when a visitor downloads documents such as the [IETF Hackathon Sponsorship prospectus](#).
- **Links to an outside domain that were clicked ([Outlink](#))**
This feature tracks to which sites visitors navigate when they follow links that are not to other resources on www.ietf.org. Methods such as using `navigator.sendBeacon()` will be employed to mitigate decreases in responsiveness for link clicks.
- **Page generation time ([Page speed](#))**
This feature tracks the time it takes for web pages to be generated by the webserver and then downloaded by the user. It will be used to evaluate website performance.

- **Main Language of the browser being used** (`Accept-Language` header)
This information will be used to understand what the main languages of visitors are, and whether there is a difference in visits for visitors who indicated different languages.
- **User Agent of the browser being used** (`User-Agent` header)
This information will be used to help understand what kinds of browsers and platforms are used to visit the www.ietf.org website.

From the User-Agent information provided, Matomo uses its [Universal Device Detection library](#) to detect the browser, operating system, device used (desktop, tablet, mobile, tv, cars, console, etc.), brand, and model.

Matomo allows configuration to scope measurement and tune reporting to meet IETF requirements. In general, the aim is to collect the minimum amount of data required to provide the desired information. Matomo provides guidance on [how to configure privacy settings](#). Expected steps to minimize data collection include:

- [disabling tracking cookies](#),
- deactivating [Matomo's User ID feature](#)
- [anonymizing](#) visitors' IP addresses, and
- [deleting visitor logs](#) after 13 months.

Steps to ensure data security will be reviewed and implemented based on the recommendation of the IETF Tools Team in coordination with the IETF Secretariat team and consistent with current data security practices for existing IETF data stores. The analytics data dashboard will be made available to the IETF Secretariat, communications staff, and the IESG. Providing a publicly-available summary of analytics data will be explored.

Implementation

In addition to installing and running the Matomo On-premise package, a modest amount of development would be needed to integrate Matomo into the IETF's Wagtail installation. A specific implementation plan will be developed after this outline proposal is finalized and approved. Significant review by the IETF community is expected before approval. A draft process for this review is provided below.

Following finalization and implementation of the proposal, and beyond adjustments that become immediately apparent to ensure expected operation, the web analytics

and reports will be reviewed by the IETF Tools Team after one-year to confirm they are delivering anticipated results.