

Project "Metadata" & Strategy

Sponsor: Assessment and User Experience Strategy at Duke University Libraries

- Key Sponsor Contacts:**
- Sean Aery - Digital Projects Developer
 - Thomas Crichlow - Web Experience Project Lead

Project Themes: Web Analytics, Change Management, Web Development, Academic Libraries, Information Technology, Data Privacy

Skills Used: Systems Analysis, Project Management (PM), Stakeholder Management

UNC Team Meetings: Weekly (Fall 2022, Spring 2023)

Sponsor Meetings: Biweekly (Spring 2023)

PM Tools: Basecamp, Google Suite

PM Approach: Agile/Scrum

Survey & Analysis Tools: Qualtrics, Tableau

Development Tools: GitLab, Docker

Programming Languages: Javascript, HTML, Bash (Command Line), SQL

Two high-level task categories:

- Technical implementation and testing of Matomo
- Stakeholder engagement, peer library interviews, and other research

Why an Agile/Scrum approach?

- To find synergies and overlap between two work categories
- For healthy information flows within our team and with our sponsor
- Adaptability for exploratory and sometimes uncertain nature of the pilot

Peer Library Outreach

- Identified three university libraries that have implemented Matomo in some capacity: UNC-Chapel Hill, Middle Tennessee State, and Cornell.
- Conducted interviews regarding their implementation of Matomo, inquiring about stakeholder buy-in, resource allocation, and plugin recommendations.
- Documented barriers and limitations, as well as best practices for addressing them:
 - Custom Reports, WordPress Tracking, Historical Data Retention
- Produced documentation summarizing key findings, implemented recommendations during development cycle, and conducted thorough testing.

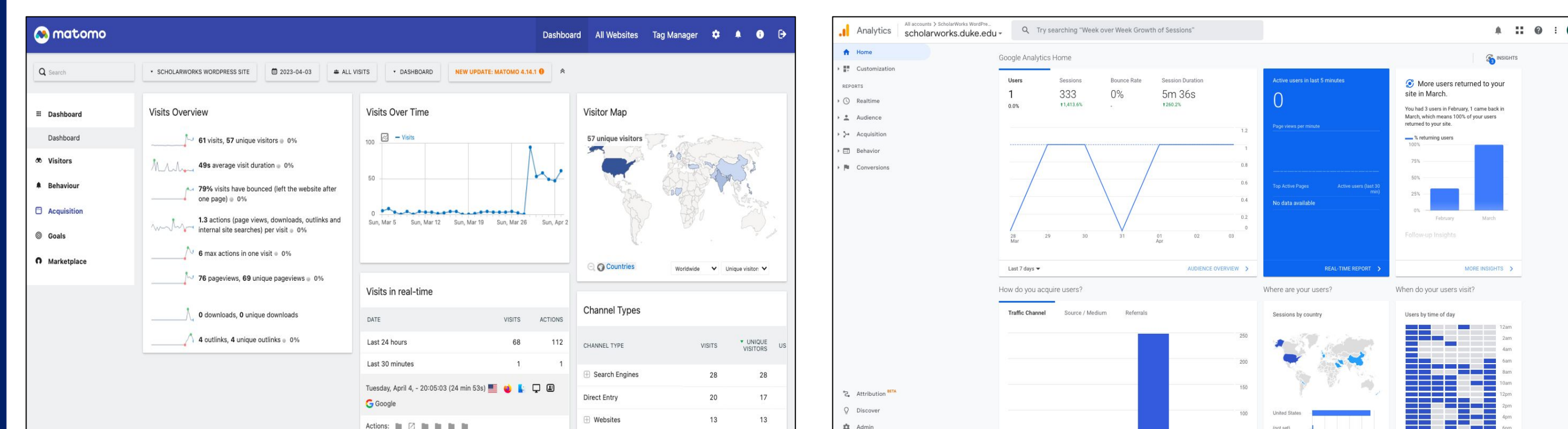


Image Source: Matomo, <https://matomo.org/>

Project Overview

Our team worked with the Assessment and User Experience Strategy (AUXS) team of Duke University Libraries (DUL) to run a pilot test of Matomo, an open-source alternative to Google Analytics. The purpose was to evaluate the feasibility of decoupling from the Google ecosystem to better meet their commitments to user privacy. Through a collaborative and agile process, we implemented Matomo on DUL servers, tested technical processes on high- and low-traffic sites, engaged stakeholders, and provided a series of recommendations on how to best transition all 110 DUL websites to Matomo.

User Interface Comparison



Matomo UI

vs.

Google Analytics UI

- Used input from sponsors and stakeholder engagement to identify important features within Google Analytics for DUL.
- Performed a screen by screen comparison of important features in GA and identified the equivalent features in Matomo.
- While there are differences in visual appearance and terminology between Matomo and GA, Matomo offers comparable functionality to satisfy DUL's web analytics needs.

Tracking User Data with Matomo

```

35 <!-- Matomo -->
36 <script>
37 var _paq = window._paq = window._paq || [];
38 /* tracker methods like "setCustomDimension" should be called before "trackPageView" */
39 //track sub domains
40 _paq.push(['trackPageView']);
41 _paq.push(['enableLinkTracking']);
42 // accurately measure the time spent in the visit
43 _paq.push(['enableHeartBeatTimer']);
44 (function() {
45     var u="//analytics-dev.lib.duke.edu/";
46     _paq.push(['setTrackerUrl', u+'matomo.php']);
47     // site ID = 3 is the analytics dev home site, this is the most important variable.
48     _paq.push(['setSiteId', '3']);
49     var d=document, g=document.createElement('script'), s=document.getElementsByTagName('script')[0];
50     g.async=true; g.src=u+'matomo.js'; s.parentNode.insertBefore(g,s);
51 })();
52 </script>
53 <!-- End Matomo Code -->
    
```

Figure 3: The JavaScript snippet that is embedded on the DUL Archives website that sends user data to the Matomo instance.

Data Comparison

Events, Users, Site Performance, and Acquisition are roughly equivalent in Matomo and GA. PageViews is aggregated in GA and dispersed in Matomo; DUL will need to account for this in future analysis.

- Installed tracking code on a complex application and a WordPress site to gather user data and evaluate the technical process.
- Worked in DUL's DevOps environment to recreate GA tracking of custom web events (searches, clicks, interactions) in Matomo, ensure changes meet DUL's deployment standards, and push the code to live production sites.
- Analyzed web analytics data tracked by Matomo and GA at the micro and macro level to determine usability and did rounds of testing to identify how live tracking affects data migration efforts.

Migration of Google Analytics Data to Matomo

- Desired to migrate historical GA data into Matomo, merging it with live Matomo tracking data.
- Migrated data using the public Matomo GA Data Import Plugin for three websites all varying in web traffic and historical data.
 - GA data was merged with data generated by Matomo tracking.
- Limitations include GA set API call limits (50,000 per day and 10,000 per site/view).
- Created a Matomo Support issue due to incorrect plugin behavior.

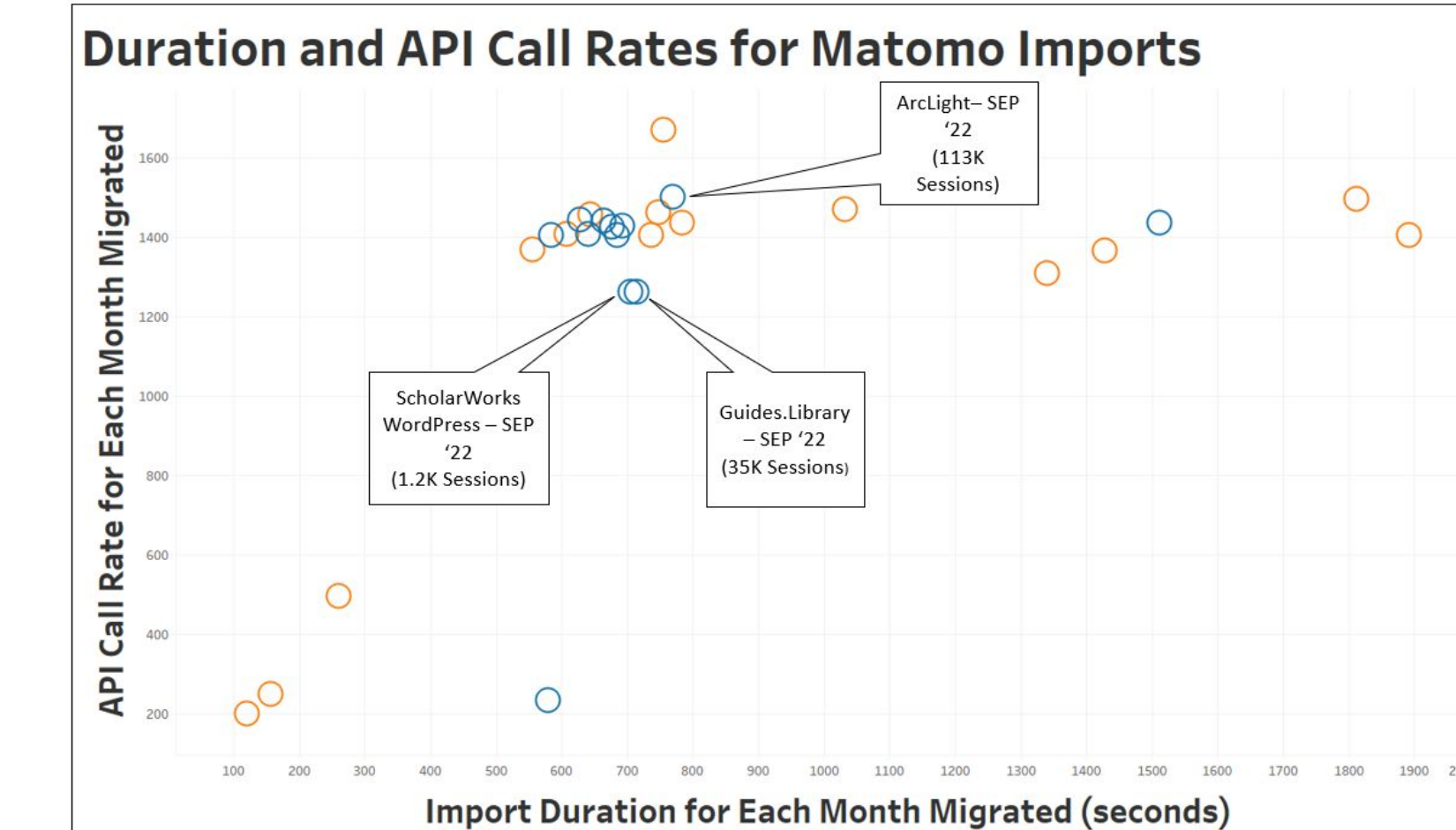


Figure 4: By analyzing server logs we estimated import duration per month of data migrated and confirmed Google API call rate limitations while taking into account the varying size, complexity, and time in existence of DUL sites.

Stakeholder Outreach & Analysis

Goals: Engage with internal Duke stakeholders to measure their current use of GA and identify important features. Assess DUL's stakeholders opinions regarding privacy considerations for web analytics data and the Matomo transition.

Part 1: Qualtrics Survey

- Methodology**
- Collaborated with sponsors to identify 37 stakeholders who have access to GA at DUL.
 - Distributed a 15 question survey to stakeholders.
 - Achieved 78% completion rate of survey.

- Findings**
- Strong stakeholder support for the transition (see Figure 2).
 - Divided stakeholders into high, medium, and low priority users based on importance of web analytics to their job.
 - Identified key metrics, including pageviews and referrals.

Part 2: Follow-Up Interviews

- Methodology**
- From initial survey results, identified three "power users" to follow-up with for more in-depth information.
 - Conducted three 30-minute interviews with power users.

- Findings**
- Determined how analytics data informs business decisions within university libraries.
 - Identified key concerns for Matomo transition from power users and relayed findings to sponsors.

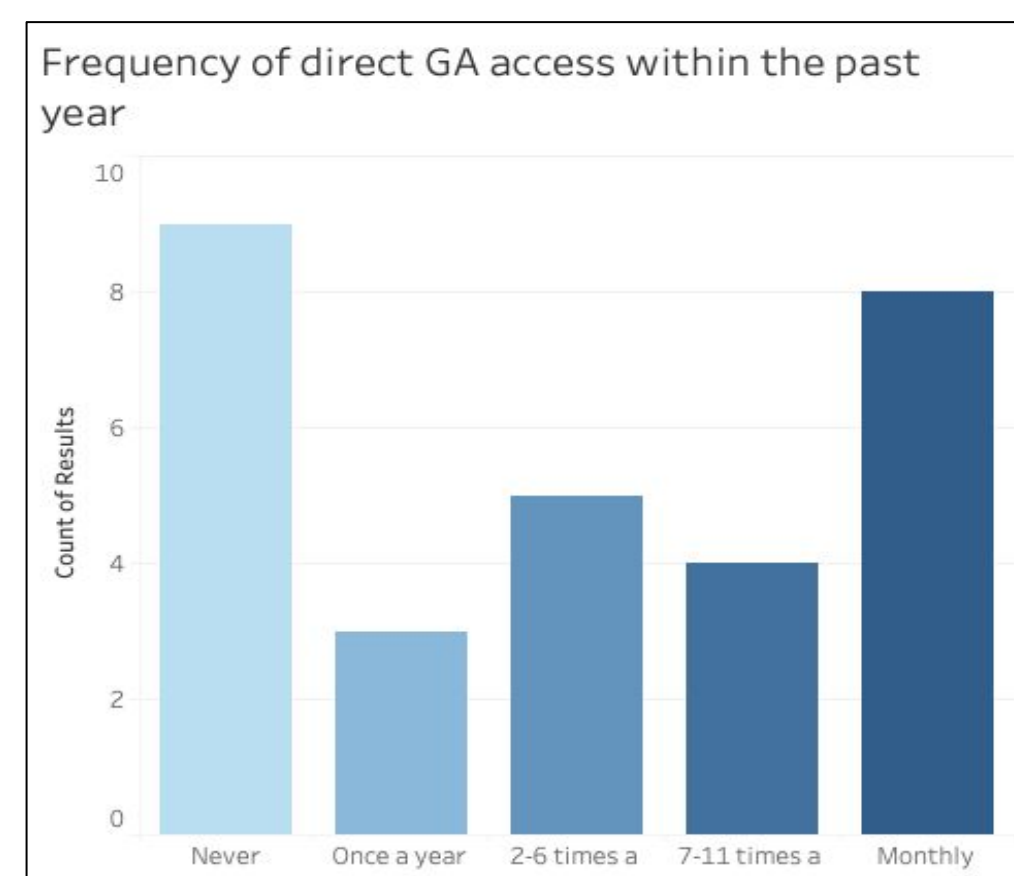


Figure 1: Survey results showed varying levels of direct engagement with GA among respondents.

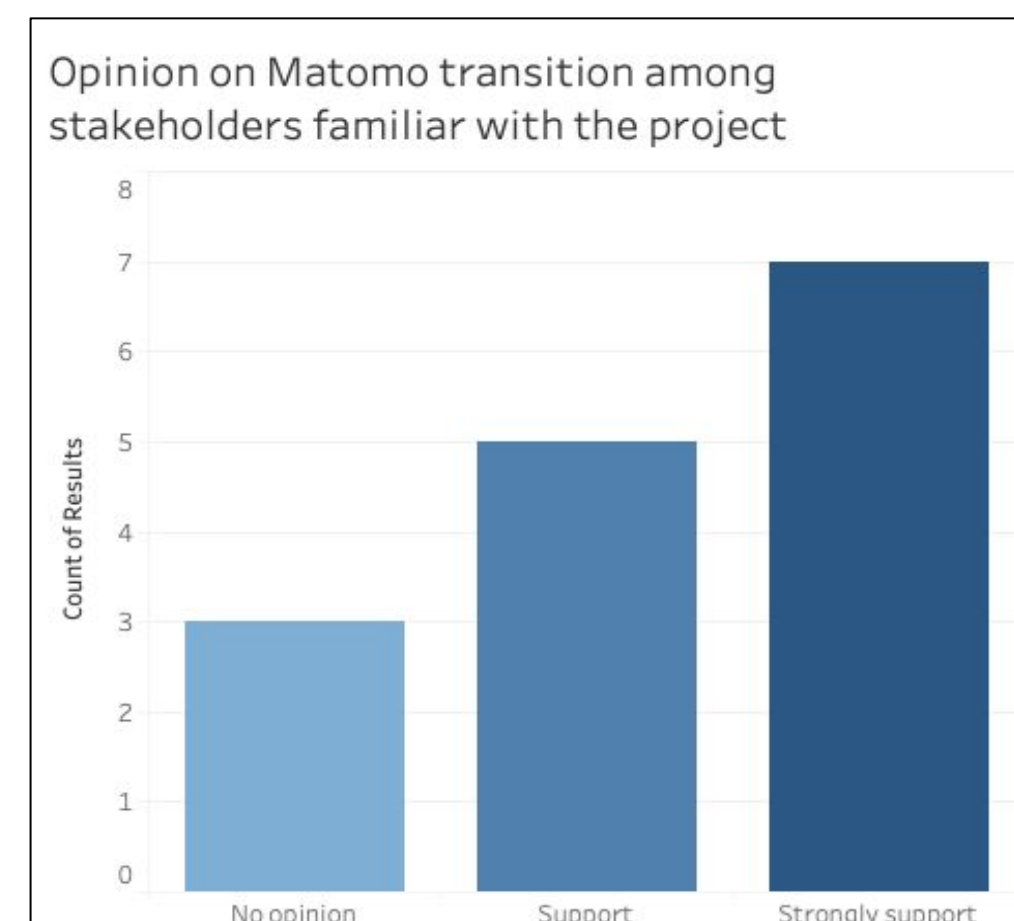


Figure 2: Strong support in survey results for the Matomo transition with 0 respondents opposing.

SWOT Analysis

Strengths

- Strong organizational/staff buy-in for Matomo
- Dedicated Dev-Ops and IT teams for implementation
- System resources appear sufficient

Weaknesses

- Historical GA data for 110 websites/apps presents challenges, e.g., reaching consensus on data policies
- Staff that use GA may need training to transition/adapt to Matomo.

Opportunities

- Engaging peer university library network
- Active Matomo forum and development community
- Tracking code consistent with GA, straightforward to implement

Threats

- GA Data Importer Plugin inefficiencies: API limits
- Standard open-source disadvantages: premium add-ons/costs, technical requirements, limited customer support

Matomo On-Premise Storage Demands

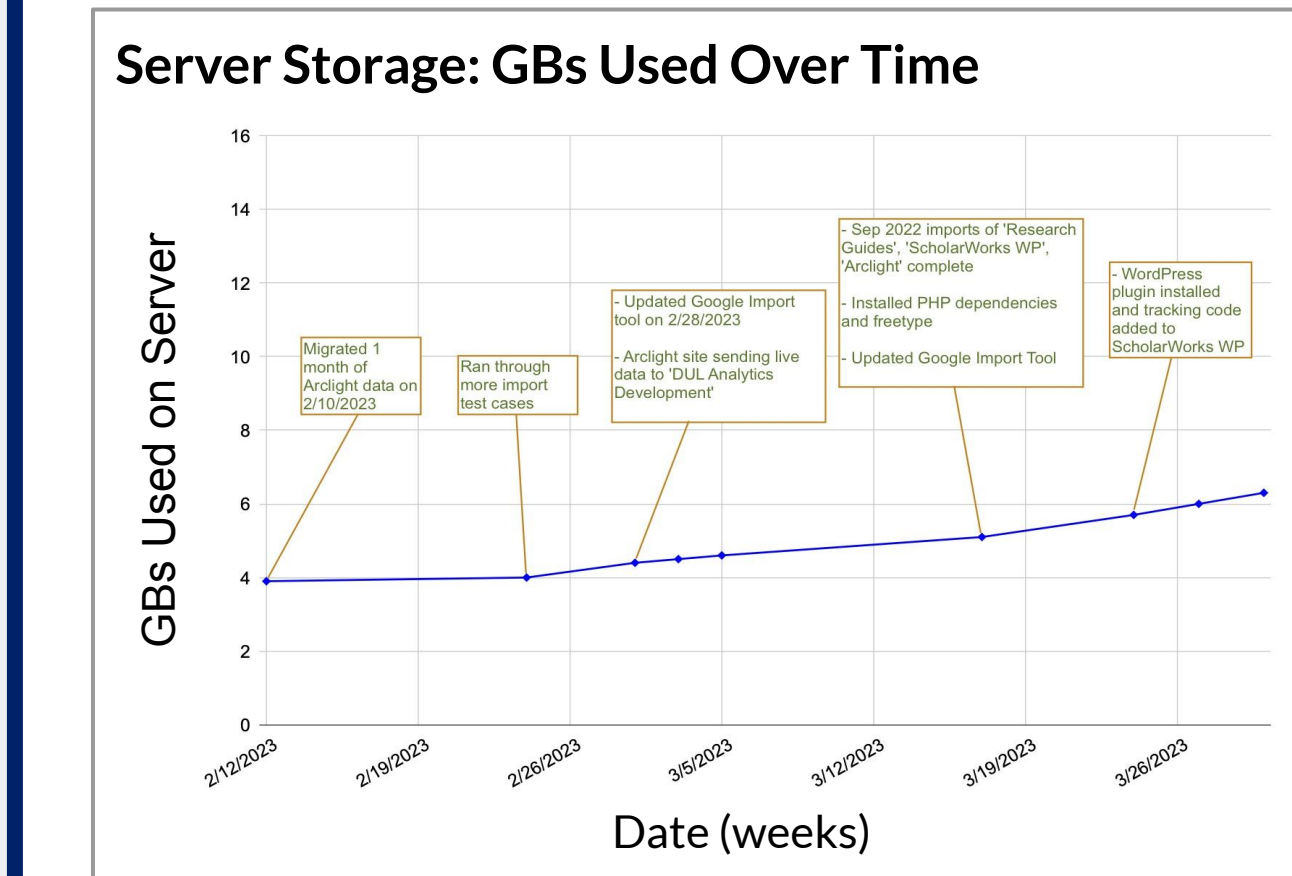


Figure 5: Data use on Matomo DB and application server, with annotations of prior actions taken.

- Executed command line statements via SSH, ran SQL queries, and referenced the Matomo UI to record how much disk-space Matomo was consuming.
- Used Months of Matomo site/view data as a dimension to make a handful of projections for minimum storage requirements based on two possible data retention policies.

Storage projections with 2-year data retention policy
 59 active sites at 8.5 MB/month = 12,036 MB (~12 GB)
 110 total sites at 8.5 MB/month = 22,440 MB (~22.4 GB)

Project Outcomes & Next Steps

- Through our testing and development efforts, we helped DUL reach the conclusion that Matomo is a viable application for their web analytics strategy assuming proper user training, appropriate allocation of network resources, and disciplined data migration effort.
- In conducting our stakeholder outreach, we identified mission critical processes across departments that depend on web analytics, with the interviews and surveys serving as a messaging platform regarding the proposal to move to Matomo.
- DUL aims to fully transition to Matomo by July 1st, 2023. Throughout the process, they will rely on a volume of technical documentation and user feedback generated by our project.