Custom Analytics with Google Tag Manager:
Assessing Usage Statistics on the MD-SOAR Platform

INTRODUCTION

Google Tag Manager (GTM) is an analytics customization interface designed to facilitate development of user-designed tags, triggers, and variables to pipe information to the familiar Google Analytics user interface. Designed in response to demands for simple, centralized control of analytics administration, this system allows for centralized control of analytics customization that can cascade to all web pages associated with a single code snippet or container. Given the reduction in manual maintenance and the prospect of customized analytics based on user- or institutionally-identified interests, GTM may provide much needed control to a metrics approach that, in the past, has been prohibitively labor intensive.

PROJECT BACKGROUND

In December 2015, David Dahl – director of the CLAS team – proposed the idea of implementing Google Tag Manager as a supplementary element of the current Google Analytics environment used to track access and usage of the MD-SOAR digital repository (https://mdsoar.org/), a DSpace repository supported by the Digital Systems & Stewardship division of the university libraries. Based on literature that supports the implementation of GTM in conjunction with DSpace (Sweeney, 2013), it was recommended that we investigate an appropriate implementation approach. In particular, we wished to find a better means of collecting aggregated usage information for downloads of MD-SOAR-hosted records.

Works Cited

Ahava, S. (2015). #GTMtips: Track files downloads in GTM V2. *Simo Ahava's Blog*. Retrieved from http://www.simoahava.com/gtm-tips/track-file-downloads-in-gtm-v2/

Conrad, S. (2015). Using Google Tag Manager and Google Analytics to track DSpace metadata fields as custom dimensions. *code{4}lib, 27*. Retrieved from http://journal.code4lib.org/articles/10311

Koberg, C. (2015). *Google Tag Manager (V2) Essential Training* [Lynda.com online course]. Retrieved from http://www.lynda.com/Google-Tag-Manager-tutorials/Welcome/168238/373593-4.html

n.a. (2016). Custom dimensions with Google Tag Manager. Retrieved from https://support.google.com/analytics/answer/6164990?hl=en

n.a. (2016). Install Google Analytics via Google Tag Manager. Retrieved from https://support.google.com/analytics/answer/6163791?hl=en

Sweeney, R. (2013). Google Tag Manager series 4: Tips and tricks when setting up a Google Analytics implementation in Google Tag Manager. Retrieved from http://www.fourthsource.com/web-development/google-tag-manager-series-4-tips-tricks-setting-google-analytics-implementation-google-tag-manager-16156

Taylor, R. (2014). Record bitstream downloads as Google Analytics events. Retrieved from https://jira.duraspace.org/browse/DS-2088?jql=text%20-%20%22Google%20Analytics%22

IMPLEMENTATION APPROACH

Review available documentation
Reviewing the general (Install Google Analytics, 2016; Custom dimensions, 2016; Koberg, 2015) and targeted (Ahava, 2015; Conrad, 2015; Sweeney, 2013) documentation, we determined the necessary customization to obtain the desired functionality.

Register Google Tag Manager account

Create custom Java Scripts to harvest DSpace metadata as GA-compliant data variables
Following Conrad's (2015) model, we created custom scripts to harvest DSpace Dublin Core metadata present in html <meta> tags. We designed scripts to harvest author, title, and university information.

```
1  function getData() {
2      var x = document.getElementsByTagName ("META");
3      var text = [];
4      for (var i=0;i<x.length;i++)
5      {
6          if (x[i].getAttribute('name') == "DC.creator") {
7              var s = x[i].getAttribute('content') + " | ";
8              text.push(s);
9          }
10     }
11     return text;
12 }
13
```

Create custom click-listener to track qualifying events
In order to track activity qualified as "download" events, we created a trigger that would fire when clicked URLs contained a RegEx string that matches a variable that contains all MD-SOAR accepted file extensions:

$\.(pdf|xml|txt|asc|doc|docx|ppt|pptx|xls|xlsx|jpeg|jpg|gif|png|tiff|tif|aiff|aif|aifc|au|snd|wav|mpeg|mpg|mpe|rtf|vsd|fm|bmp|psd|pdd|ps|eps|ai|mov|qt|mpa|abs|mpega|mpp|mpx|mpd|ma|latex|tex|dvi|sgm|sgml|wpd|ra|ram|pcd|odt|ott|oth|odm|odg|otg|odp|otp|ods|ots|odc|odf|odb|odi|oxt|sxw|stw|sxc|stc|sxd|std|sxi|sti|sxg|sxm|sdw|sgl|sdc|sda|sdd|sdp|smf|sds|sdm|rdf|zip)$

Create custom tag to deliver an 'Event' to Google Analytics
Using our established Google Analytics dashboard, we created a tag that – once triggered – sends GA an event with a custom category (*bitstream*), action (*download*), and label (downloaded file name). We also created a custom dimension data layer that forwards the author, title, and university variables to the dashboard.
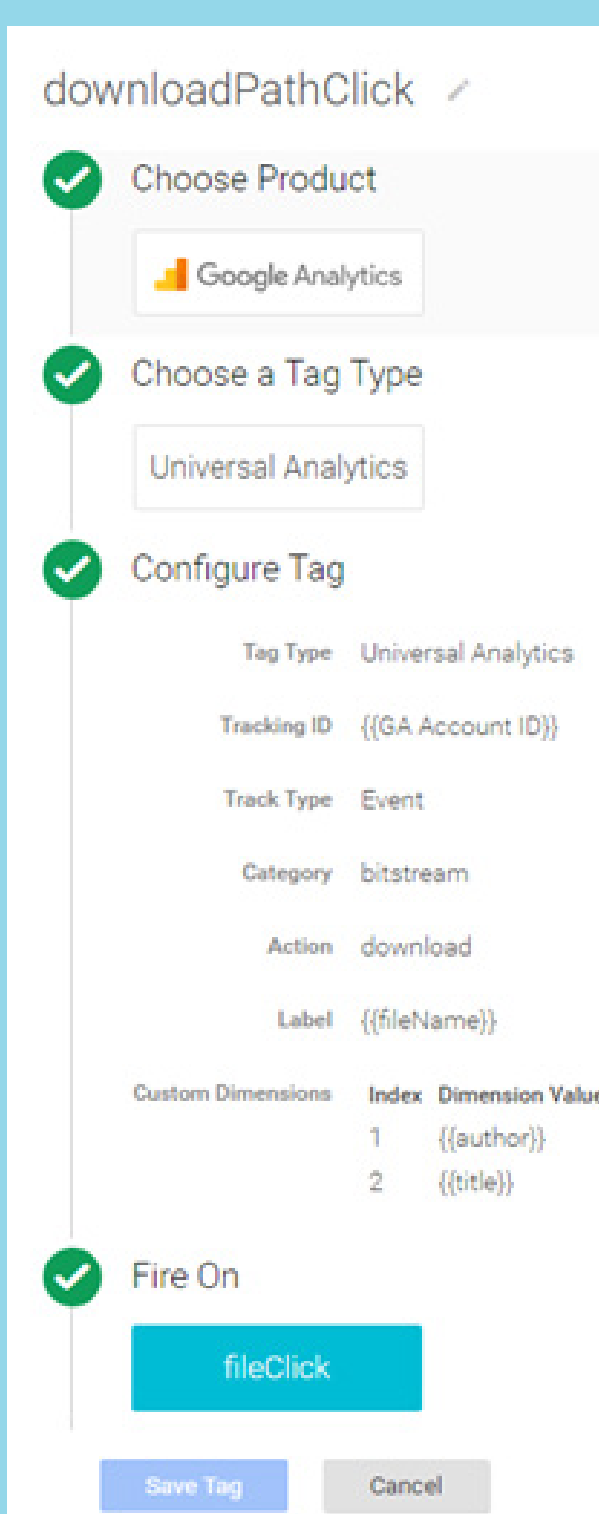
INITIAL RESULTS

1) After initial configuration in a testing environment, we determined that we would be able to successfully implement a custom download tracking tag that could provide data layer variables for author and title metadata for a selected work. Due to inconsistent use of the <dc.relation> element, capturing the university of origin was determined to be outside of initial implementation scope.

2) Collaborating with the SSDR group, we successfully introduced the GTM code snippet in to the page template HTML code, thereby activating our updated tracking approach.

3) While the code snippet has been successfully activated, we are currently within two weeks of go live and therefore have only limited results.

DISCUSSION & RECOMMENDATIONS

Despite the lack of reportable results, there are some actionable recommendations that can be made based on our implementation experience:

Metadata Alignment: While the aim of MD-SOAR – and all digital collections – is to have consistent metadata that presents parallel information across all elements, the reality of such collections is that non-conformance and outliers can confound systemic harvesting of information. As we have seen with the inability to use certain DC elements due to data variance, lack of metadata alignment can prevent programmatic approaches to data layer harvesting. This issue is compounded in consortial environments in which record submitters are expected to supply object metadata.

Migration Considerations: When migrating to GTM, be sure to review all relevant documentation regarding the move away from Google Analytics. Of primary importance is the need to replicate traditional GA metrics through GTM. Documentation is available through the article "Install Google Analytics via Google Tag Manager" (2016). Be sure to also capture any custom GA metrics and their user source. This will help when comparing pre- and post-migration metrics.

DSpace & Limitations: While GTM is an excellent approach to capturing metrics for activity on all pertinent web pages, the DSpace built-in document viewer creates some limitations on what access can be captured as these pages do not include the GTM code snippet. In-bound links from third-party sites and search engines will return no event activity, creating an artificially low record of activity. This is due to loss of custom event tracking code that was included as a de facto element of DSpace implementations since 2014 (Taylor, 2014). CLAS is currently investigating alternate means of capturing in-bound download data.