

# Implementing Google Analytics for a Consortial Digital Repository

Greg Vogl, Yongli Zhou, Daniel Draper, and Mark Shelstad

**D**igital repositories are platforms that allow libraries to make scholarly content publically available. With a few advanced features, Google Analytics via Google Tag Manager (GTM) can provide you with the ability to gather statistics on the use of these resources and reveal how users access and navigate a repository website. This case study uses the Digital Collections of Colorado digital repository consortium to describe these essential Google Analytics and GTM features to create useful and sharable Google Analytics reports for other libraries to consider for their digital repositories.

*Google Tag Manager*

<https://tagmanager.google.com>

*Digital Collections of Colorado*

<https://dspace.library.colostate.edu>

## Why Install Google Analytics on a Digital Repository?

Many major digital repository systems come with their own internal usage reports built in. DSpace

is the digital repository platform used in this case study, and it offers several internal usage report features.<sup>1</sup> It also supports adding Google Analytics for additional site usage tracking.<sup>2</sup> Before you jump on the Google Analytics bandwagon, note that it does not replace your repository's internal usage reports, especially the reports involving file downloads. Since repository content is often discoverable outside of the repository website, Google Analytics may not be able to track that usage. For example, if a user directly opens a PDF file from your repository's collection via an external source (such as Google Scholar or another website), your repository software may load the PDF file directly without the Google Analytics tracking script. Hence, Google Analytics will not recognize that download even if it is tracked as an event.<sup>3</sup> The statistics from your repository system are more likely designed to capture and report all file downloads.

Despite its limitations, adding Google Analytics to your digital repository is recommended because it expands on the data already collected in those internal usage reports. Google Analytics is a robust and free digital analytics tool that generates a series of standard and custom usage and user reports. When used with GTM, Google Analytics can easily capture more data to enhance its reporting abilities.

*All the authors of this chapter are affiliated with Colorado State University and contribute to the Digital Collections of Colorado project. Greg Vogl is a Middleware and Web Developer in Academic Computing and Networking Services. Yongli Zhou is the Digital Repositories Librarian. Daniel Draper is the Digital Repositories Unit Coordinator. Mark Shelstad is Coordinator for Digital Collection Services.*

## Configuring Google Tag Manager and Google Analytics for DSpace: Digital Collections of Colorado Case Study

The Digital Collections of Colorado (DCC) is a consortium of seven academic institutions and one publisher in Colorado. The members include mainly public universities throughout the state with varying populations and missions: Colorado State University, Colorado State University-Pueblo, Colorado School of Mines, Colorado Mesa University, University of Colorado Anschutz Medical Campus, University of Colorado School of Law, University of Colorado Colorado Springs, and the University Press of Colorado. Colorado State University, referred to as CSU, hosts the repository and provides technical and administrative support to the other consortium members. In 2015, the DCC migrated to DSpace 5, which was configured to include a collection structure for each affiliated institution and its subsequent content collections (e.g., articles, books, reports, theses and dissertations, images, sound recordings, moving images, and datasets). Since the DCC is a consortium system, its usage reports must be available for the whole system and segmented into individual reports by the institution.

### Implementing Google Tag Manager on DSpace

GTM is a script management system that integrates with Google Analytics to enhance its data collection ability. We opted to use GTM over the traditional Google Analytics tracking method in DSpace because GTM can easily track events and custom dimensions and metrics for Google Analytics. It also requires little action from DSpace system administrators, since the only code added to the DSpace server is the GTM container snippet, and all other Google Analytics scripts are handled through the various GTM tags, triggers, and variables. We loaded the GTM container snippet via a separate JavaScript file. This injects the GTM container snippet right at the start of the body tag (<body>). This is not ideal since loading GTM through an extra file slightly reduces performance, and tracking may fail in rare cases where the file does not load before the user interacts with the page. However, we found this option still highly functional.

*JavaScript file for loading GTM container snippet*  
<https://dspace.library.colostate.edu/static/js/ga-manual.js>

Since we have two instances of the digital repository (DSpace and DSpaceTest), we created two GTM

containers so each DSpace instance has its own container snippet. All GTM customizations were tested on the test server and then exported and imported to the GTM container for the live server.

### Creating Custom Dimensions for DSpace

A Google Analytics dimension is a text string, such as page title or URL, while a metric is a numeric value, such as the number of pageviews or bounce rate. Google Analytics comes with many built-in dimensions and metrics, but it allows you to create custom dimensions and metrics to capture and report additional useful data points.<sup>4</sup> Sound complicated? GTM greatly simplifies this process as it captures custom dimensions and metrics via simple tags and user-defined variables. We will demonstrate how to collect custom dimensions using GTM.

For this case study, our custom dimensions are based on the various text strings within the breadcrumb navigation used on the DCC interface. This breadcrumb navigation contains several data points, including the institution and collection name on each web page. By collecting these data points as custom dimensions, we are able to create usage reports at the institution and collection level.

Creating a custom dimension is a two-step process, so we first added the custom dimensions to our DSpace's Google Analytics account. Adding custom metrics or dimensions is an administrative feature. Go to Google Analytics' Admin section to find the Custom Definitions option under the Property section and follow these steps:

1. Select Custom Dimensions and click the New Custom Dimension button.
2. Provide a name for the desired custom dimension.
3. Select the Hit option since this custom dimension should be tracked at the pageview level.
4. Save the new custom dimension.

You can ignore the series of JavaScript examples Google Analytics automatically generates. Instead, remember the assigned Index number to that custom dimension—this will be used later in GTM to sync the custom dimension to the data point found on your repository website.

Overall, we created five custom dimensions based on the different breadcrumb navigation levels: Institution, Breadcrumbs, Collection, BreadcrumbCount, and CollectionTopLevel. After adding the custom dimensions to our Google Analytics account, the next step is to build the GTM User-Defined Variables to capture the breadcrumb navigation text from our repository website as a custom dimension. To do this, go to the Variables section in your GTM container (within the GTM website). Here is the process we used:

1. Click the New button under the User-Defined Variable section.
2. Name the variable. We used the same labels as the custom dimensions added to our GA account.
3. Select the Custom JavaScript option.
4. Enter the script that identifies and captures the data related to this custom dimension. We used jQuery because DSpace already uses that JavaScript library. This script identifies the breadcrumb navigation segment and returns that segment's text or a "not set" response if the breadcrumb is not present. Here is the script we used to capture the institution level breadcrumb segment:

```
function() {
var c = jQuery('ul.breadcrumb li');
if (c.length < 2) return 'not set';
var u = [
'Auraria Library',
'Colorado Mesa University',
'Colorado School of Mines',
'Colorado State University, Fort
Collins',
'Colorado State University, Pueblo',
'Open Textbook Library',
'University of Colorado Anschutz
Medical Campus',
'University of Colorado Boulder,
William A. Wise Law Library',
'University of Colorado Colorado
Springs',
'University Press of Colorado',
'University Tester - CSU'
];
var cu = jQuery('ul.breadcrumb li:nth-child(2)').text().trim();
return jQuery.inArray(cu, u) == -1 ?
'not set' : cu;
}
```

5. Save the variable.

We created a new custom JavaScript variable for each custom dimension we previously added to Google Analytics—figure 6.1 demonstrates the institution custom dimension as a User-Defined Variable in GTM. This example highlights five different custom dimensions, but you can have as many as twenty custom dimensions per Google Analytics' property. Other potential custom dimensions include using GTM variables to collect Dublin Core metadata elements (e.g., `dc.title` and `dc.type`) to analyze repository usage based on specific items.<sup>5</sup>

To test whether a GTM variable is being set as expected, select the Preview and Debug option from the Publish menu. In the same web browser, navigate to your repository website to see the GTM debugger window load at the bottom of the page. Within the GTM debugger window, click the Page Load message



**Figure 6.1**  
User-defined variable for institution custom dimension, Google Tag Manager, Colorado State University

in the left sidebar and click the Variables tab at the top to view the values of each variable. If everything looks good, publish the changes to the GTM container.

## Using Google Tag Manager to Track Events in DSpace

GTM also greatly simplifies how Google Analytics tracks events on websites. If you are unfamiliar with Google Analytics' event tracking, a Digital Library Federation's analytics working group published a report that describes this process,<sup>6</sup> and chapter 2 of this issue of *Library Technology Reports* demonstrates event tracking using GTM. You should definitely track events on your digital repository website because there are many useful events to analyze including search and advanced searches, search facets or navigation features, any interactive features such as streaming videos, and of course, downloads.

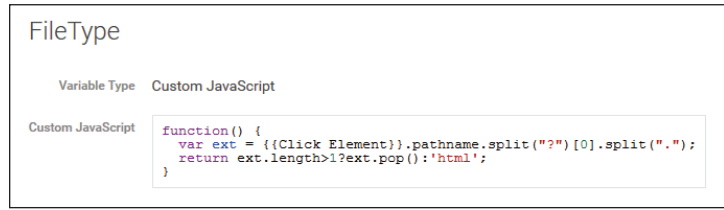
### Creating a Download Event Tag

Understanding how and when users download content from a digital repository is extremely useful especially if you associate all those great custom dimensions with that download event. This provides you with more data to analyze your digital repository within Google Analytics. The following section describes creating a GTM event tag to track file downloads along with the custom dimensions we previously made.

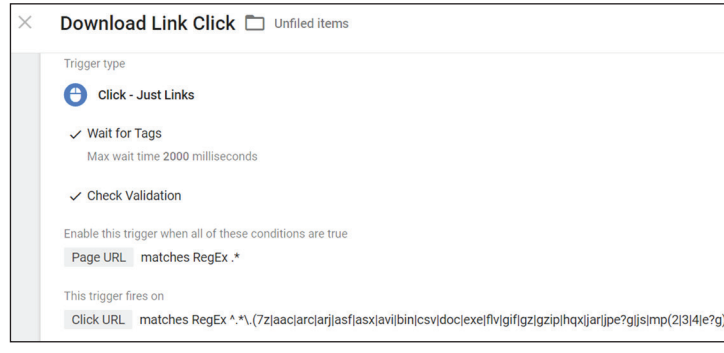
1. Click on the New button in the Tag section of your GTM container.
2. Name the new tag. We called it Downloads Event.
3. Under the Choose a Product section, select Google Analytics and then select Universal Analytics. Click the Continue button.
4. Add your website's Universal Analytics' Tracking ID number. We used a User-Defined Variable

called `{{gaPropertyID}}` to store our tracking ID number. This process is documented on the GTM Help website.<sup>7</sup>

5. Select Event as the Track Type since this is an event tag.
6. Add the category, action, and label information for this event. Our category was simply called downloads, while the action and label uses GTM variables. Action uses a User-Defined Variable we created called `{{FileType}}`, which is a custom JavaScript variable to retrieve the file extension from the URL of the clicked element (figure 6.2). Label uses the GTM system variable `{{Click URL}}` to record the link the user clicks.
7. Expand the More Settings option to add the custom dimensions. You can add as many custom dimensions you wish to associate with this event activity. In the Index text box, add the Index number Google Analytics assigned to that specific dimension. In the Dimension Value text box, add the matching GTM User-Defined Variable you created to capture the custom dimension. Click the Continue button when you are done adding your custom dimensions.
8. Under the Fire On Options, select the Click option to create a click trigger so the downloads event occurs only when a user clicks on a file link.
9. This will pop open the Create Trigger box where you name the new click trigger. We named ours Download Link Click.
10. Select Just Links as the Targets option.
11. If you want GTM to provide more time for the tag to fire, leave the Wait for Tags and Check Validation options checked. This is completely optional, but if you do use these features it may slow down your repository website's response time since all necessary GTM tags must load on the web page before the file downloads. Additionally, you must complete an additional step to define when GTM should watch for these clicks. Under the Enable When area, select Page URL and matches RegEx. In the textbox add the RegEx (regular expression): `.*`. This tells GTM to monitor all links on your repository website for a file extension.
12. Click the Continue Button and under the Fire On section, select the Some Clicks options.
13. Set the trigger conditions by selecting the GTM system variable `{{Click URL}}`. Next, select matches RegEx. In the text box, add the RegEx string to match the various file extensions your digital repository contains. We included the most common file extensions so our RegEx string was very comprehensive as shown in figure 6.3.



**Figure 6.2**  
FileType user-defined variable, Google Tag Manager, Colorado State University



**Figure 6.3**  
Download link click trigger, Google Tag Manager, Colorado State University

14. Save the new click trigger.
15. Save the download event tag.

Test your new tag in the Preview and Debug Mode and publish it when you are ready. Within twelve to twenty-four hours of publishing, your Google Analytics Event reports will show this new data. Additionally, you will be able to use any of those added custom dimensions as a secondary dimension in your Events reports to further analyze the data—figure 6.4 demonstrates Google Analytics Event report with the `CollectionTopLevel` custom dimension as a secondary dimension.

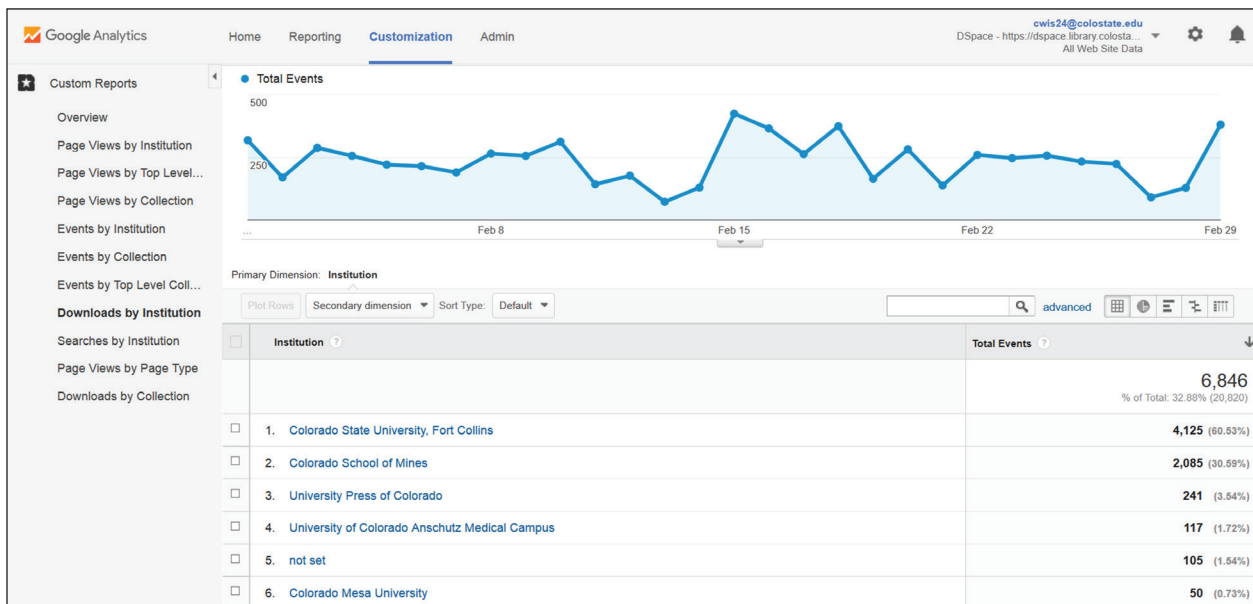
## Using Google Analytics Custom Reports for Creating and Sharing Reports

Custom dimensions and metrics are useful Google Analytics enhancements when combined with Google Analytics custom reports. Custom reports allow you to use custom dimensions as primary dimensions and then select a handful of other metrics to create useful reports. For example, we were able to create a report for the number of download events for the different institutions, as shown in figure 6.5.

Creating a custom report is simple: just click on the Customization tab within Google Analytics. Select the metrics and dimensions you wish to view. Add any

Event Category ?	CollectionTopLevel ?	Total Events ?	Unique Events ?	Event Value ?	Avg. Value ?
		20,820 % of Total: 100.00% (20,820)	6,498 % of Total: 67.55% (9,620)	0 % of Total: 0.00% (0)	0.00 Avg for View: 0.00 (0.00%)
1. Downloads	CSU Archives and Special Collections	1,376 (6.62%)	428 (5.75%)	0 (0.00%)	0.00
2. Community Collection	CSU Archives and Special Collections	1,345 (6.47%)	195 (2.62%)	0 (0.00%)	0.00
3. Downloads	CSU Theses and Dissertations	1,231 (5.92%)	832 (11.18%)	0 (0.00%)	0.00
4. Downloads	CSU Departments and Schools	1,155 (5.56%)	641 (8.61%)	0 (0.00%)	0.00
5. Community Collection	not set	973 (4.68%)	467 (6.28%)	0 (0.00%)	0.00
6. Downloads	Theses & Dissertations	888 (4.27%)	572 (7.69%)	0 (0.00%)	0.00
7. Community Collection	CSU Theses and Dissertations	873 (4.20%)	128 (1.72%)	0 (0.00%)	0.00
8. Downloads	Library Special Collections, Archives	867 (4.17%)	129 (1.73%)	0 (0.00%)	0.00

**Figure 6.4** Event report with CollectionTopLevel (custom dimension) as a secondary dimension, Google Analytics, Colorado State University



**Figure 6.5** Custom report, Google Analytics, Colorado State University

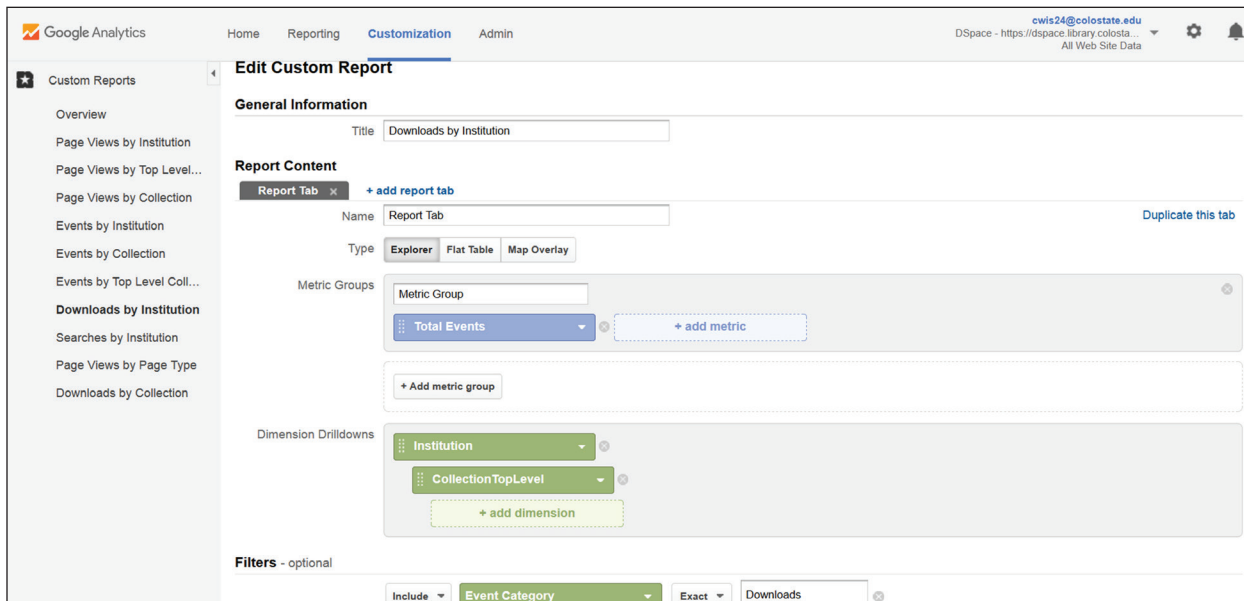
filters that are necessary to drill down to the relevant data. Save the report, and your new custom report is stored within Google Analytics. Figure 6.6 on the following page depicts the configurations for the Downloads by Institution custom report that used Total Events as the metric and two custom dimensions: Institution and CollectionTopLevel. Then we added a filter to look only at events that used the download event category.

The final step is to get these reports to the different institutions that need them. Google Analytics provides several export and sharing options, such as

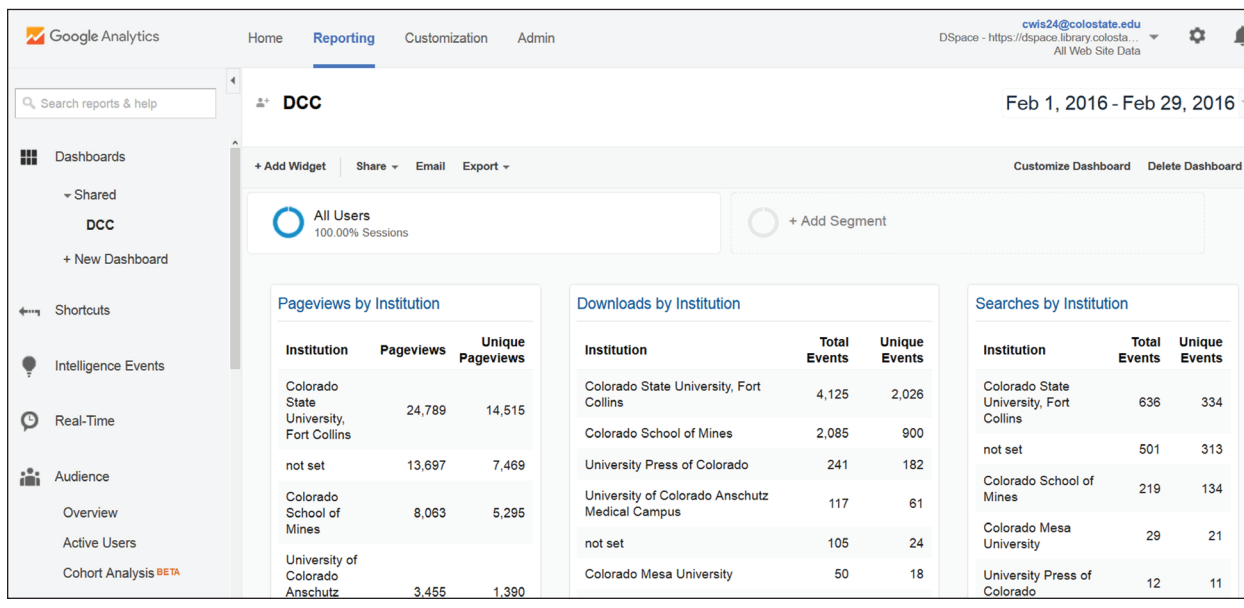
scheduled e-mails and shortcut links.<sup>8</sup> You can even use these options to create custom dashboards within your Google Analytics account. Custom dashboards allow you to add several standard or custom reports to one location as shown in figure 6.7. You can share a lot more data in one scheduled e-mail or shortcut link using a custom dashboard report.

## Conclusion

Google Analytics and GTM provide great flexibility to



**Figure 6.6**  
Custom report configuration, Google Analytics, Colorado State University



**Figure 6.7**  
Custom dashboard, Google Analytics, Colorado State University

collect and report additional usage data through their customization options and advanced features, but libraries need to be strategic on how they implement these tools on their digital repositories. It is important to create an assessment plan before implementation and thoroughly test your customizations to ensure you are getting useful data. Once you have that data, share it with the people that need it.

## Notes

1. Bram Luyten, "Statistics and Metrics," "DSpace 5.x Documentation," DuraSpace wiki, created October 26, 2014, <https://wiki.duraspace.org/display/DSDOC5x/Statistics+and+Metrics>.
2. Robin Taylor, "DSpace Google Analytics Statistics," "DSpace 5.x Documentation," DuraSpace wiki, last modified by Peter Dietz May 26, 2015, <https://>

- wiki.duraspace.org/display/DSDOC5x/DSpace+Google+Analytics+Statistics.
3. “About Events,” Google Analytics Help, accessed February 29, 2016, <https://support.google.com/analytics/answer/1033068?hl=en>.
  4. “Custom Dimensions and Metrics,” Google Analytics Help, accessed March 8, 2016, <https://support.google.com/analytics/answer/2709828>.
  5. Suzanna Conrad, “Using Google Tag Manager and Google Analytics to track DSpace metadata fields as custom dimensions,” *Code4Lib Journal*, no. 27 (January 2015), <http://journal.code4lib.org/articles/10311>.
  6. Molly Bragg, Joyce Chapman, Jody DeRidder, Rita Johnston, Ranti Junus, Martha Kyrillidou, and Eric Stedfeld, *Best Practices for Google Analytics in Digital Libraries* (Washington, DC: Digital Library Federation, September 2015), [https://docs.google.com/document/d/1QmiLJEZYGAY-s7BG\\_nyF6EUAqcyH0mhQ7j2VPpLpxCQ/edit](https://docs.google.com/document/d/1QmiLJEZYGAY-s7BG_nyF6EUAqcyH0mhQ7j2VPpLpxCQ/edit).
  7. “Use Constant String User-Defined Variables,” Google Tag Manager Help, accessed March 15, 2015, <https://support.google.com/tagmanager/answer/6164390?hl=en>.
  8. “Export and Email Options,” Google Analytics Help, accessed on March 15, 2016, <https://support.google.com/analytics/answer/1038573?hl=en>; “Report Shortcuts,” Google Analytics Help, accessed on March 15, 2016, <https://support.google.com/analytics/answer/2676996?hl=en>.

## Additional Resource

“Website Analytics,” Colorado State University Libraries, accessed March 9, 2016, <http://libguides.colostate.edu/analytics>. This online guide includes extensive documentation of the Google Tag Manager and Google Analytics configurations used for the Digital Collections of Colorado in DSpace.