Tableau Online Pilot (2019-2020) Report to COLD

January 2021

I.   Tableau Server Pilot Overview

Background

In December 2018, the CSU's Data Visualization Task Force reviewed functionality available through Tableau's free Public version of its software and recommended that a paid Server license be acquired to test the ability to more effectively share data and visualizations across the CSU.[1] This report details features tested during the pilot and documents features and limitations of the paid Server license. The Server license was hosted by Tableau (Tableau refers to this as Tableau Online).

Costs

Table 1. Summary of Costs for Phased CSU-Wide Tableau Adoption, 2019-2020			
	Item	Unit Cost	Year 1 Cost
Year 1 (2019-2020)	Two (2) Tableau Desktop Creator Licenses (Tableau Online)	$630*	$1260
	Five (5) Tableau Desktop Explorer Licenses (Tableau Online)	$375 per license	$1875
		TOTAL, Year 1	$3135

A second year of the Tableau pilot for 2020-2021 was initially proposed but put on hold due to the rise of other priorities and lack of capacity to continue testing Tableau due to the COVID-19 pandemic.

---

[1]Data Visualization Task Force Report To COLD (2018).
https://docs.google.com/document/d/13vakA3MwT2lBtOxv1qJKSILKtvQDaJwk9zfaG4oc_-M/edit?usp=sharing

II.   Alma Analytics Tableau Integration

One goal of the Tableau pilot involved investigating functionality for creating automatic or "live" updating dashboards, whereby a Tableau visualization would update with fresh data as it became available (data from Alma Analytics would therefore automatically refresh daily). While it is possible to create a Web Data Connector (WDC) to retrieve data from Alma Analytics for manipulation and visualization in Tableau, it was not possible at the time of our testing to configure Tableau Online reports to automatically refresh data via the Alma WDC. It is possible to manually update the visualization on Tableau Desktop and push the updated data to the Tableau Online visualization. Tableau works well for static analyses (e.g., based on data for a single year, or manually created analyses comparing multiple years) but currently requires manual intervention to maintain visualizations that are frequently updated on a scheduled or automated basis. It was also not possible to share the WDC connection as a data source for other Creator accounts on our account to use via Tableau Online; each user with a Tableau Desktop license wishing to retrieve Alma Analytics data via the Analytics WDC must enter their connector information and API key each time they create an analysis with the Alma Analytics WDC.

Some Tableau and Alma Analytics users have indicated that the optimal workflow for creating dynamic, automatically updating visualization is to directly utilize data from the Alma Analytics API to output data as XML for scheduled harvesting by Tableau instead of utilizing the Web Data connector. This would require custom development but would also allow for greater flexibility and transparency between the Alma Analytics data and Tableau, and could also facilitate using the same data output from Alma Analytics for other visualization tools (e.g., Power BI, Google Charts).

In terms of the Alma Analytics integration, the paid Tableau Online service did not offer improved functionality over the 'free' Tableau Desktop/Public features apart from the ability to create and share 'internal' visualizations with other paid users on the account (the Tableau Explorer licenses). While this functionality could be useful for creating visualizations of library data not intended for public sharing (e.g., data on library expenditures or aggregated/anonymized data about library user behavior) the benefits of the paid / shared Tableau Online service are limited.

III. Limitations of Tableau Online

The following limitations were identified in the course of testing the Tableau Online service:

- *Lack of scheduling/automation of data harvesting from Alma Analytics*
  - In our testing we could not get scheduled data updating from the Alma Web Data Connector to update automatically, but rather had to manually retrieve data from the Tableau Desktop client in order to fetch new data, and then re-publish the visualization to Tableau Online.
- *No ability to join or 'mash' multiple Alma reports or automatically create trend analysis over time from the Alma Web Data Connector*
  - Tableau is not intended to serve as a data repository or directory. Some visualizations that might require joining together multiple Alma Analytics reports from different subject areas cannot be done using the built-in Alma Web Data Connector. Reports can be exported from Alma Analytics in spreadsheet form and loaded as shared data sources in Tableau Online, but the organization of these shared reports quickly becomes unwieldy and difficult to manage using Tableau's shared data source directory.
- *Publishing limitations*
  - One reason for exploring the paid Tableau licensing model was an interest in having the ability to share visualizations not intended for public viewing, but accessible to internal Library stakeholders. While this is possible with Tableau Online, these 'private' published visualizations require a Tableau "Viewer" license in order to view, meaning that every CSU-library stakeholder would require a license to view internal dashboards (at the time of this writing, a Tableau Viewer license is $12/month per user, with a minimum of 100 users).[2] Visualizations created with the paid Tableau licenses can, of course, be published to Tableau Public, but this negates any benefit of the paid license in terms of more privately sharing visualizations.

IV. Recommendations

The benefits of a paid CSU-wide subscription to Tableau Online (or any data visualization application) are minimal if data sources are not organized, findable, and flexible. The most impactful resource limitation with regard to data visualization and

---

[2] https://www.tableau.com/pricing/teams-orgs

CSU-wide Library visual data assessment is not access to powerful data visualization tools, but rather the lack of an organized data library and the lack of "clean" data that is understandable and ready to be analyzed using available tools. Data directly from Alma Analytics often has to be extensively manipulated and cleaned up in order to be usable in data visualization systems, and relying on built-in direct connections (such as with the Alma Tableau Web Data Connector) require a significant investment in time to manipulate extracted Alma Analytics data into a meaningful Tableau data stories. Without a searchable directory of data sources (whether data sources be manipulated spreadsheets, reports created and available in the Alma Analytics shared catalog or network zone instance, or other data sources such as vendor-supplied usage data reports or ILLiad data) it is difficult to envision a CSU-wide data visualization strategy or software tool implementation.

It is recommended that the next steps of the Data Visualization Task Force and/or the ULMS Assessment and Analytics Committee involve the creation of a Data Library and/or a Data Lake[3] to centralize data sources across the CSU. This would bring together data sources from a variety of CSU library systems (Alma, Primo, ILLiad, vendor-supplied usage reports for SDLC-negotiated and ECC resources, etc.) and organize them in a single repository, including relevant metadata regarding the context of the data creation and its known limitations. Given the changing landscape of data visualization and assessment tools for libraries and the many possible vendors, tools and products available, this work would allow the CSU libraries to connect data from its Data Lake to a variety of different tools and more carefully and comprehensively assess the usefulness of available products.

---

[3] One example of a data library/data lake hybrid for an academic library, created using Confluence, is described by Shafer et al. (2019). Data Lake: Promoting a Mega-Tool for the Assessment Lifecycle. http://qqml.org/wp-content/uploads/2017/09/Shafer-QQML-2019-20190527_Data-Lake.pdf