Qualitative and Quantitative Methods in Libraries (QQML) 10,4: 417-432,2021

Data Lake: Promoting a Home-Grown Tool for the Assessment Lifecycle

Dana Peterman[1], Sharon Shafer[1], Todd Grappone[1]

[1]University of California, Los Angeles, CA, USA

Abstract

A recently created Library Strategic Plan at the University of California, Los Angeles (UCLA) identified the need to incorporate a culture of assessment throughout its organization and to encourage data-informed decision-making. This case study encompasses the development of a home-grown tool, called UCLA Library Data Lake, which guides library staff in assessment, and acts as the central repository and documentation of UCLA Library's efforts. An appointed library team developed the UCLA Library Data Lake and employed outreach and education techniques to inculcate a culture of assessment for change. The team's interaction with library staff, management, and campus experts resulted in a flurry of discoverable library studies, collection of associated data, data applications and assets. Using the features of Data Lake allows dynamically generated reports to enter into the project management lifecycle. Library-created studies and data have improved the tool's configuration and begun to deepen the library's understanding of its mission, goals, and functions.

Keywords: Data Lake, assessment lifecycle, data-informed decisions, organizational culture, enterprise systems, knowledge management

Introduction

The UCLA Library consistently ranks among the top academic libraries in the United States serving 45,000 students in 125 majors. The Library employs approximately 100 librarians and 350 full-time staff working in more than a dozen library locations all over campus. These staff supports over 3.5 million in-person visits annually, 12 million print and electronic volumes, and more than 15 million virtual visitors via the website. Library units report to the University Librarian through four Associate University Librarians and

Received: 5.9.2019 Accepted: 17.12.2021 ISSN 2241-1925

management staff. Information technology and associated library services are highly converged as the Library operates the Digital Initiatives and Information Technology (DIIT) division. DIIT is comprised of the following units: The Data Science Center, Digital Library Program, Core Data Services, Operations and Services, and Software Development and Library Systems.

A decade ago, the Library systems department attempted to build a data warehouse to help with decision support, however it did not go beyond a pilot study. Now that the Library had an established business knowledge system and issue tracking system in Jira, it was time to re-examine the need for a data repository or the updated concept of a data warehouse – a data lake.

Data Lake

A data lake is a centralized repository of structured and unstructured data. A proposed solution to creating a centralized inventory of assessment data might be to just implement a data lake. However, merely dumping all data into a data lake without any metadata management would only lead to a *Data Swamp* (Madera & Laurent, 2016). Data by itself, even with metadata, is not enough. Staff need help learning how to ask questions of data. So, using readily available enterprise wide software, the UCLA Library created a fusion of data, abstract and index database with educational templates to create a repository. While not a data lake *per se*, test inquires led to a favorable response to naming the hybrid the UCLA Library Data Lake (Data Lake for short). The Data Lake centralizes metadata management AND assessment training templates AND tools AND reports using search and reporting macros contained within the Atlassian product, Confluence, an enterprise business knowledge and collaboration wiki platform.

A form in Data Lake guides individuals or teams through the brainstorming and planning of assessment ideas. At desired points within the assessment lifecycle (figure 1), Data Lake users can dynamically notify resource managers and stakeholders through the use of a macro that sends emails and creates notifications within the Confluence platform. The notifications encourage stakeholder feedback and participation, and may lead to decision making and resource allocation from all stakeholders. Tailored modules on the platform allow staff to abstract and index assessment ideas, data, tools, and reports. Taking advantage of Confluence's flexibility, Data Lake supports dynamic visualizations and dashboards through application programing interfaces (APIs) in all modules. Connections to service tickets can be made to Atlassian's Jira to feed into existing and potential projects. When staff have questions or need assistance of any kind, they can turn to a section of the Data Lake dedicated to consultations by appropriate individuals.

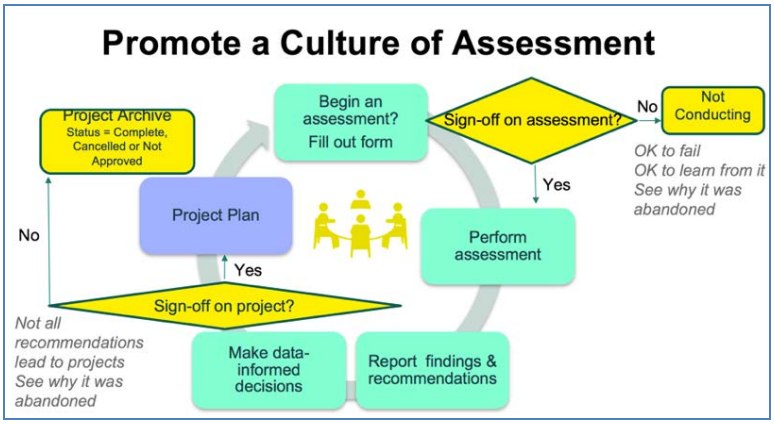

Figure 1. Assessment Lifecycle

The story of how UCLA came to create and deploy the UCLA Library Data Lake and how it informs change to help implement our strategic plan and the formation of a team; it includes the creation of an approach that combines outreach with education, transparency, and cyclical evolution that led to success.

Strategic Planning

In 2015/2016 the Library created a strategic plan guided by external consultants with the input of most library staff across the organization. Many of the mission statements and goals contained within the Library Strategic Plan made the need for assessment apparent. The identified strategic goals included:

"We use a transparent, user-centered, evidence-based approach to assess our activities and impact."

"We nurture an environment of ongoing evaluation, transparent decision-making…"

"Our research support improves continuously through assessment-based improvements."

"We improve our instructional services, focusing on effective pedagogical practices and ongoing assessment."

"Our space-related initiatives begin with a data-driven, user-centered approach."

(Strategic Plan 2016-19, n.d.)

Implementation of the plan's goals commenced in two cycles. In 2017/2018, the second cycle of the plan's implementation led to the creation of three teams, which included the Assessment for Change Team (ACT). The UCLA Library as

an organization hungered for the embrace of assessment and ACT was ready to feed it.

Assessment for Change Team (ACT): Formation and goals

Library administration solicited applications to serve as leaders based on their qualifications as researchers and trailblazers. Library administration appointed ACT leaders and members who represented a broad spectrum of skills and responsibilities across the organization. The team charge was developed by the team leaders, the Strategic Planning Implementation Coordinator, Dana Peterman, and Associate University Librarian, Todd Grappone. The charge included Specific, Measurable, Attainable, Relevant and Timely (SMART) goals in the pursuit of the following sustainable deliverables:

- Document resources and conduct in-person education of UCLA Library staff on appropriate assessment methodologies and tools for library services, products and practices.
- Oversee the development and implementation of departmental key performance indicators (KPIs) using a set of decision-making tools (dashboards, pivot tables, etc.) and instruct others on their use for the performance of liaisons, teaching activity, collection use, space, and services.
- Development of a Confluence prototype of a centralized inventory of KPIs, tools, data, reports, dashboards and user stories that are used in assessment at the UCLA Library.

The team's goals changed as ACT worked with those interested in assessment while developing enterprise-accessible tools, and interfacing with experts.

As a large and complex organization, UCLA Library has conducted business intelligence via information silos with a diversity of data sources and types making it difficult to access and use information for decision support. The library's interest in assessment as expressed in its strategic plan comes after decades of operational information gathering that staff recognized could add to its effectiveness if it were centralized and widely accessible. Library staff voiced their hunger for a culture that supported access to data for decision making. In addition, staff identified the need for the skills that broke down those business intelligence silos. ACT had to find a way to establish a set of processes that creates and shares knowledge across the organization in order to optimize and make transparent the library's use of judgement in the attainment of its mission and goals.

Key to the process was the use of extant tools in the Library arsenal with which users would feel comfortable. So, ACT chose Confluence, Jira, Box, and JasperReports, to address knowledge management within the library. Of these tools, Confluence became the most heavily implemented in the process because

it is used by more of the library staff for library documentation and because it lends itself to more user-friendly design through its macros. Jira, while not an afterthought, is not used extensively by all units because it was primarily designed for software development and project tracking and is not as user friendly as Confluence. Box became a storage solution providing space for data, reports, and any other additional information that an assessment might require. Data storage is also available through a local repository, though it is a seldom used resource. JasperReports are used by library staff to obtain or analyze data from a variety of sources, including the library catalog.

The team moved to create tools within Confluence to guide and centralize work, and to find ways to uncover and create business intelligence and knowledge management by educating and informing library staff about assessment and evaluation. ACT made these moves iteratively and simultaneously. ACT worked on the tools with user feedback and reached out to external parties to standardize, expand, and modify the library's approach to knowledge management.

Three-Pronged Approach

ACT has begun to instate a three-pronged approach that promotes culture change, education, and tool creation for assessment that is continuous, integrative, transparent and ongoing (figure 2).

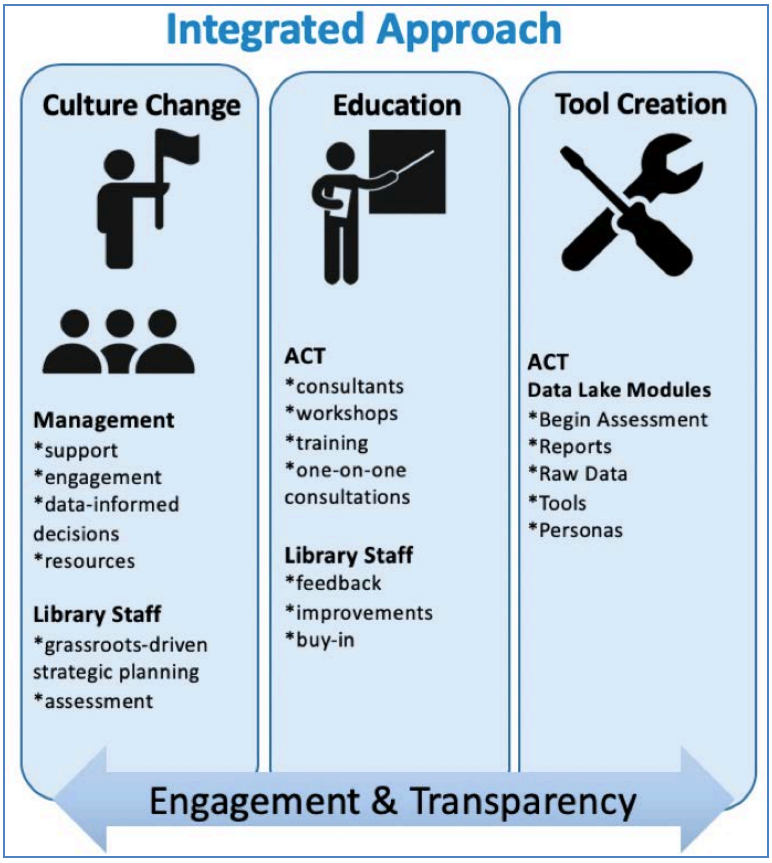

Figure 2. Three-Pronged Approach to Establishing an Assessment Culture

Approach 1: Culture Change

The work of ACT was initially advertised and promoted in an all staff meeting. University Librarian, Virginia Steel, introduced and commented on the value of ACT. The ACT leaders followed their introduction with a show of one of UCLA Library's Data Lake's initial prototypes (figure 3).

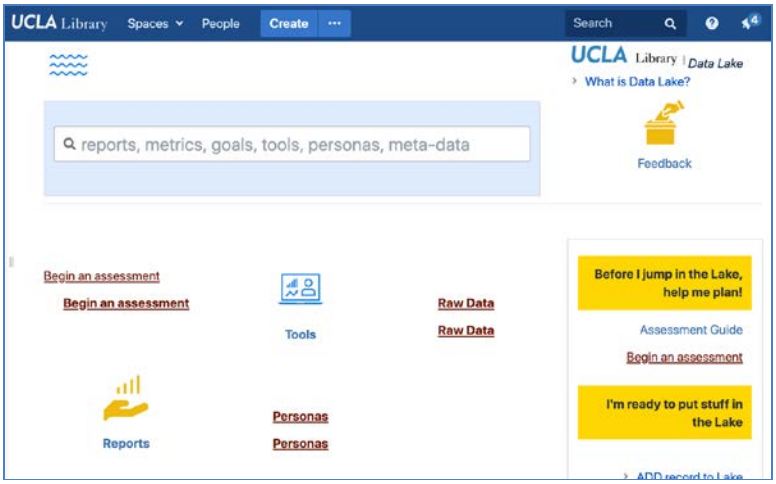

Figure 3. An Initial Prototype of the Data Lake

The all staff meeting opened the door for ACT to reach out to the Library. Though ACT had already started by searching for past assessment activities placed in Confluence, we were then able to speak with staff who had expressed frustration about locating assessment studies. ACT started to talk about how a Confluence form it was prototyping could help them. ACT employed the prototype page to test against the current studies and comments of UCLA Library staff (figure 4).

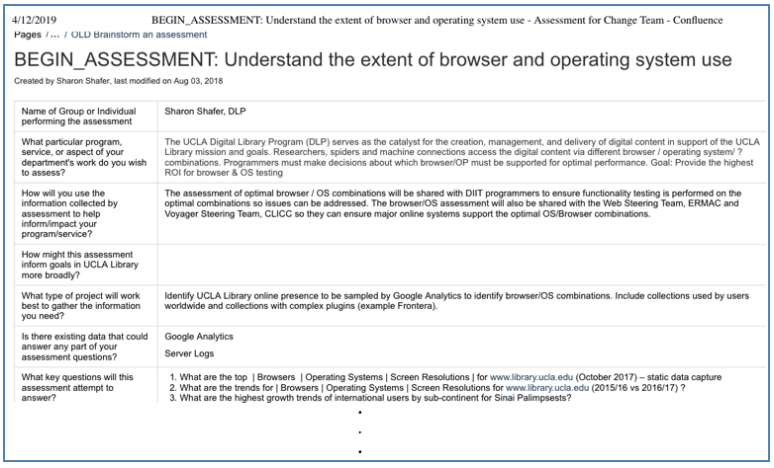

Figure 4. First Assessment Study Prototype Showing Confluence Form

During those project meetings, ACT modified Data Lake templates to meet the needs of our users. These on-the-spot changes to the templates of an enterprise-

wide tool sparked a sense of combined purpose and gradual cultural change. As an added benefit, the interactions promoted the UCLA Library Data Lake as a tool and ACT as a resource. Even better, as time progressed, ACT was sought out for consultations rather than ACT seeking out partners.

An important element that influences the formation of assessment as a cultural norm within the library is transparency. Everything in Data Lake is in draft form. Projects can be initiated or halted if users feel that data or circumstances make it impractical or out of scope for the library's mission. However, once someone has done the work of gathering information or data, the idea remains in case it is later determined to have viable elements or could be used for other projects or assessments. The value of the work is maintained because it can be found and used later.

ACT designed its Confluence templates to help lead viable projects into project management lifecycles that lent themselves to Jira tie-ins. As users developed their projects, they could be steered into more formal project management templates outside of Data Lake with more stringent requirements and resource allocation potential, though that it is rarely needed.

Approach 2: Education

ACT conducted a needs assessment by interviewing library staff and performing unstructured and diverse interviews. It became clear that we needed as an organization to establish a common core of experiences, vocabularies, and tools.

To build a shared set of experiences and vocabularies, ACT identified the UCLA Student Affairs Information & Research Office (SAIRO) as an educational resource from which to pull for a few key reasons. SAIRO offered training for free; they had access to campus-wide and system-wide University of California data; their contacts were broad, and people across campus reported their assessment efforts to them and used them as a resource. SAIRO had also just begun investing in an assessment tracking tool themselves, so ACT felt that their interactions would be mutually beneficial.

ACT engaged a guest lecturer from SAIRO to conduct a voluntary *Introduction to Assessment* workshop for Library staff. The workshop helped identify and create a cohort of staff interested in the process of assessment to ACT. In turn, the workshop helped identify ACT members to attendees so that team members could serve as consultants to any assessment processes. ACT hoped that SAIRO and they could educate staff to use resources and methods appropriately.

The workshop that SAIRO provided led to a definition of assessment that put the process into perspective. SAIRO's lecturer, Kevin Cleland, defined assessment for us as, "a way to make decisions that guide our future, not to validate decisions of the past." This singular conception of assessment helped prioritize what to pursue. Counting reference statistics just for the sake of reporting them to the Association of Research Libraries was not enough.

Instead, trying to instigate a positive change in reference services related to our library's mission that could be measured by a change in the statistics and other metrics was the way to go. The need to create forward thinking data-informed decisions becomes clearer when focused on change. In addition, SAIRO introduced to the Library staff another conceptual tool used in librarianship and other fields called the logic model in which one looks at inputs, outputs, and outcomes to form an assessment for the purpose of change.

The ACT team conducted a second workshop that further solidified the definition of assessment as a planning tool for the Library's aspirations. The new cohort used the second workshop to begin a flurry of new assessments and to guide staff interested in making change. We followed up with the introductory assessments by having ACT members meet with interested staff project-by-project.

Approach 3: Tools

Tools: Requirements and specifications for an assessment tool

Recognizing that knowledge management in large organizations tends to be diverse, discrete, and decentralized (Townley, 2001), ACT attempted to accommodate diverse and discrete knowledge in a centralized organizational structure within Confluence as an enterprise business knowledge system.

As ACT investigated assessment needs within the UCLA Library, the following requirements and specifications for the Data Lake began to emerge.

- Discover via search and browse
- Add and modify all assessment related records
- Connect to information (Box, Jira, JasperReports, Google Analytics - Studio-Google Tag Manager, and more)
- Include dynamic data visualization
- Create an inventory of data assets and applications
- Facilitate assessment consultation and guidance
- Facilitate communication with stakeholders

Tools: Determine scope of the Data Lake

ACT determined that the scope of the Data Lake would include the initial modules for assessment related to raw data, reports, personas, tools, and educational templates to help staff begin assessment studies (figure 5). The Data Lake modules were designed to describe and point to external sources like Box, JasperReports, Google Data Studio, and many more. Discovery included keyword searching and module browsing based on tasks and user experience of members. The interface allowed for both retrieval of information and adding of information which is similar to a repository known by ACT members as eScholarship.

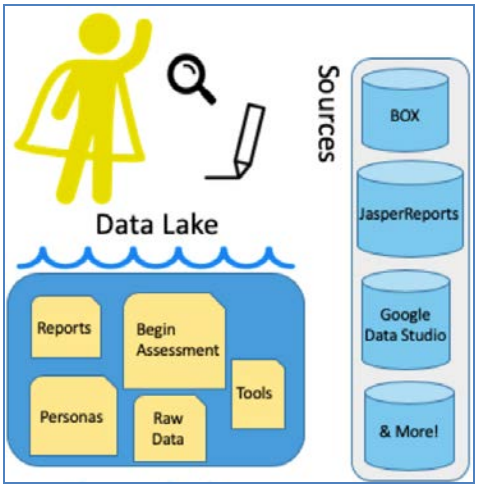

Figure 5. Data Lake Modules and Data Sources

Tools: Modules are the recipe for success

Many factors contributed to the recipe required to assemble a Data Lake based on Confluence. ACT based its modular design on social science research tools, and the recent implementation of a UCLA Library Service Catalog in which two of the authors had played a significant role. The modules employ a simplified modular design using macros. Records in every module were seeded with information to encourage participation. Every module contains an index of records that may be browsed by title or by searching the text of the entire record. Additionally, the entire corpus of Data Lake may be searched. The *Assessment* module explicitly ties Library goals to the project, and both the *Assessment* and *Report* modules emphasize the ability to notify key library stakeholders.

Some terms need to be present in multiple modules

Many examples show that terms may be used as both records in modules and facets of modules as they play different roles within Data Lake. Google Analytics is Data (figure 6) as well as a Tool (figure 7), but can also be a Report facet (figure 8).

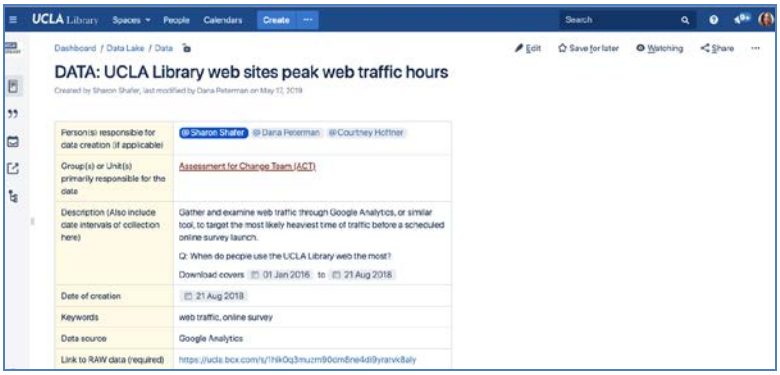

Figure 6. Google Analytics as a Data

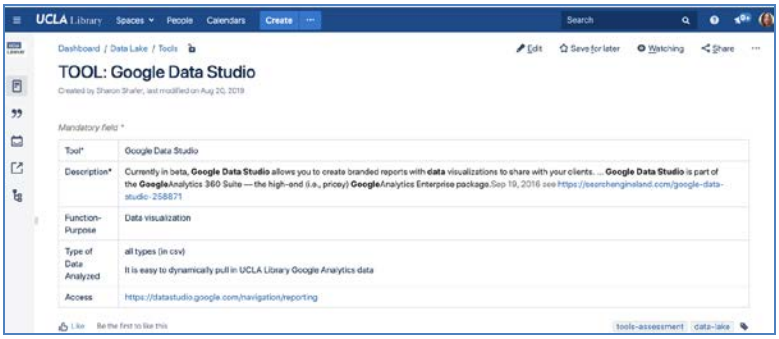

Figure 7. Google Analytics as a Record in the Tool Module

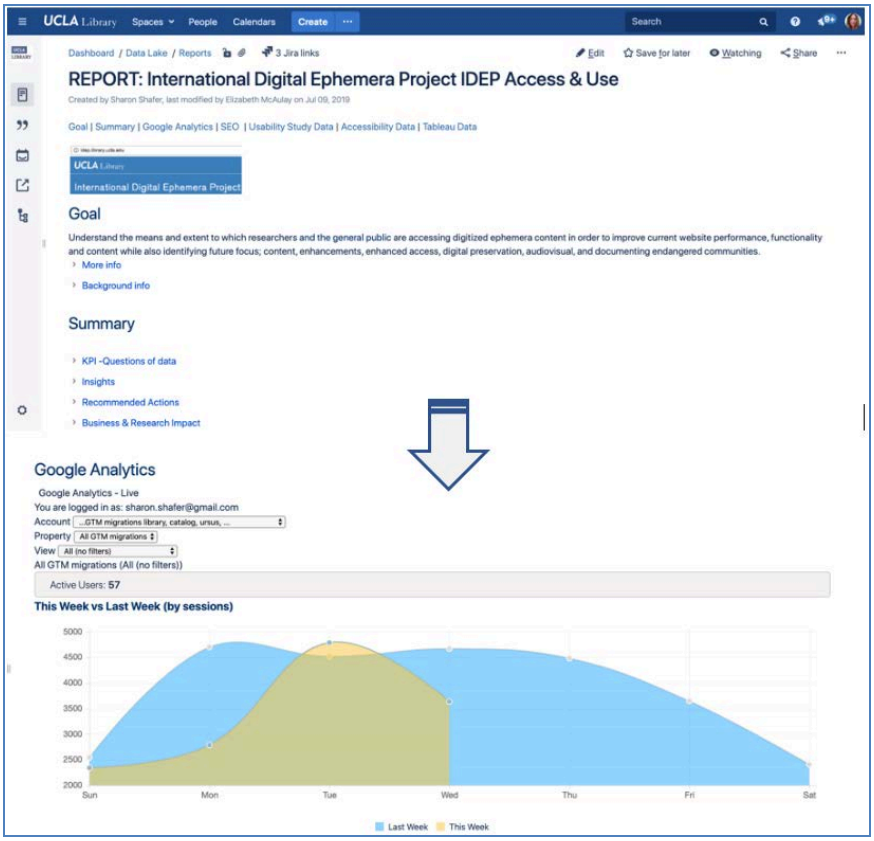

Figure 8. Google Analytics Data Visualization as a Report facet

ACT experimented with many different modules until arriving at the final four Data Lake modules:

Assessment Module

The KPI brainstorm evolved to an *Assessment* module based on changes to guide the library's future, not to validate its past decisions. The fill-in-the-form module emphasizes the importance of data and its link to change and to the library's mission and goals. It is important to note that no field is mandatory so that users feel free to experiment and not be judged. Some fields are somewhat repetitive in order to elicit a more complete picture of how key elements, such as the role of data, apply to a proposed change. A logic model worksheet in the form places the focus on results so users can reverse engineer how best to create change.

Tools Module

This module was initiated as a way to discover, share, or request tools that assist with assessment. Using a template, it is possible to catalog a tool record by providing a title, description, function and access mechanism/link. Of note, this module entailed the most content seeding by ACT members because it was discovered that many library staff were unaware of the variety of currently available tools and educational resources within the UCLA Library Data Center. The Head of the UCLA Library Data Center, Tim Dennis, helped ACT members identify important tools as well as facilitating access to Data Carpentry workshops. Data Carpentry develops and teaches workshops on the fundamental data skills needed to conduct research.

Reports Module

Assessment reports existed in the UCLA Library Confluence database prior to the implementation of Data Lake, but they often lacked qualities of a planned assessment tied to data informed decisions and transparent change. The *Reports* module was configured as a way to offer discovery and sharing of reports that assist with assessment in a centralized repository. Using a template, it is possible for library staff to create a metadata description of a report. Staff can either point to the full text report located in another space or include the full text in the report record. It is possible to include report text, images and associated API-driven data visualizations.

Data Module

UCLA Library collects various types of data from hundreds of sources. The *Data* module was created as a centralized way to discover or share data. This module is configured as a repository to search, browse or share data that can help someone conducting an assessment. Using a template, it is possible for library staff to create a metadata description of data. It is possible to either point to the full data files located in another space or include the full data in the *Data* module record. We hope to explore dynamic connections to various data sources in the near future.

Data Lake Prototypes: configure, test, enhance, and repeat

ACT produced many Data Lake prototypes and held library-wide assessment workshops to help staff populate the Data Lake. As library staff worked with the Data Lake they gave feedback and helped drive enhancements to the design, functionality and content. ACT members also employed the Data Lake while conducting the needs assessment for assessment in the UCLA Library. ACT assessment studies served as skin in the game and as a model for other library staff to learn from. It is possible to see the evolution of the Data Lake during these prototype sprints by sampling two prototypes (figures 9, 10).

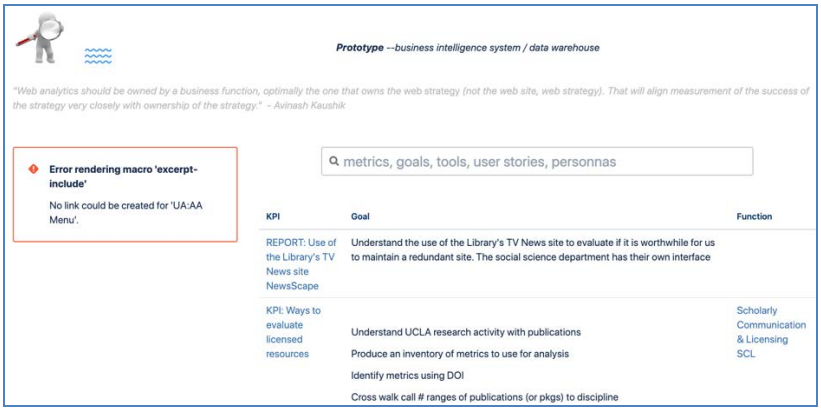

Figure 9: Data Lake Prototype Sprint 1 (2018)

Figure 10: Data Lake current prototype (2019)

Findings

Over time, ACT came to interpret assessment and documentation as knowledge management and creation for the purpose of change and decision support. Whereas before, ACT members thought of assessment as validation of prior

decisions or as a way to simply document what was required by some outside agency. The team recognized the need to invest in a wide spectrum of areas related to knowledge capture, storage, value addition, distribution and finally the need to educate ourselves about the benefits of knowledge creation and sharing (Davenport, 2000).

Implementing the UCLA Library Data Lake as a tool for supporting the assessment lifecycle and promoting a culture of assessment within the UCLA Library resulted in library staff conducting studies and contributing to a centralized assessment related information repository. During the course of using the Data Lake with library staff, ACT members learned that even assessment for change is difficult and requires ample mentoring and follow-up from on-site consultants. Assessment forms need to be Sherpa-like and simple in order for library staff to use them effectively, but work best with human intervention. So, assessment workshops and in-person consultations must be continuously offered.

For the assessment lifecycle to result in actual data-informed decisions and projects to make positive changes, it is imperative to have management support and engagement throughout the process. In particular, the UCLA Library Data Lake reporting mechanisms highlighted orphaned projects which were brought to the attention of managers and decisions were made to resolve project orphan status. ACT realized that we needed to teach both assessment and advocacy. Assessment for change is not a matter of simply documenting the need for change, but actively engaging in communication with those who have the needed resources to make that change happen.

As library staff, we found out a few surprising things about ourselves in relation to assessment. The logic models employed in many disciplines typically start with a listing of inputs, outputs, and stakeholders leading up to the expression of an outcome or outcomes, but we often know what outcome(s) we want first. There was a strong desire to focus on gathering data without justifying the need to collect that data. While we had data that could be used in its place, we had a strong bias toward using surveys. Perhaps less surprising as library staff who organize and standardize knowledge, we viewed the Confluence template pages as internally bounded. We did not see the scratch space available on every page to make it their own beyond the form - only one person has so far added a sub-page to their assessment. It is difficult to say if this illustrates ACT's success in making Data Lake look and behave like a real database, or if this hesitancy to claim Confluence page real estate requires more experience with using the tool. Much like a new cook following a recipe for the first time, each user has to learn how to make it their own.

As with any custom programming within a subscription platform, ACT encountered technical debt when Confluence upgrades were enacted which temporarily broke some Data Lake functionalities. ACT found it to be a

limitation that it could not treat the Data Lake like a true database with all of its attendant capabilities. In the case that we may wish to transfer the Data Lake to another platform, we would need to go through the trouble of replicating the interface.

The differences among the Data Lake modules (*Data*, *Reports*, *Tools*, *Assessment*) were not always as clear as ACT originally thought. For example, JasperReports and Google Analytics acted as report, tool, and data. Users had to make multiple records, particularly for data and tools.

Perhaps, the most important, but not surprising, finding was that library staff need to be encouraged and rewarded for applying useful knowledge to achieve organizational goals on a regular basis. Our consultations with SAIRO indicated the need to follow up frequently and help was in alignment with their experiences with campus departments. ACT felt the need for reinforcement and assistance might require an ongoing commitment to either a position and/or committee charged with assessment.

Recommendations

We see our greatest challenge as managing and maintaining one solution as our organization's assessment systems and data holdings expand. As much data is sensitive and subject to privacy issues and access restrictions, it is necessary to ensure a data governance policy is in place and being followed. Management and governance of data assets requires oversight and maintenance of permissions and data retention and data governance policies as well as addressing the technical debt of maintaining connections and policies with external systems such as Box for use as a true large data repository.

We also see the need to continue mentoring assessment projects as they appear in the Data Lake in addition to continuing to offer staff training workshops and illustrating tangible examples where UCLA Library has enacted data-informed change. In any event, assessment for the purpose of change in support of the UCLA Library's mission and goals must continue.

References

Madera, C., & Laurent, A. (2016). The next information architecture evolution. In R. Chbeir (ed.), *Proceedings of the 8th International Conference on Management of Digital Ecosystems*. (pp. 174-180). New York: ACM. doi: 10.1145/3012071.3012077

Townley, C. T. (2001). Knowledge management and academic libraries. *College & Research Libraries, 62*(2), 44-55. doi:https://doi.org/10.5860/crl.62.1.44

UCLA Library (n.d.). Strategic plan 2016-19. Access date 07.25.2019 available at https://www.library.ucla.edu/about/administration-organization/strategic-plan-2015/goals