International Journal for Educational Media and Technology
2018, Vol.12, No. 2, pp.42-46

Adrift In Big Data And Learner Analytics: An Exploratory Study Of A Nascent Data Governance Board

Christopher Luchs
Old Dominion University, Norfolk, VA, USA
CCCOnline, Denver, CO, USA
Christopher.Luchs@cccs.edu

Kae Novak
University of Colorado-Denver, Denver, CO, USA
Front Range Community College, Westminster, CO, USA
Kae.Novak@frontrange.edu

In a world of Big Data, today's higher educational institutions are charged with not only protecting the data of their students, faculty, and staff but also to ethically manage access and use of data. At the same time, higher education is experiencing the effects of disruptive innovation caused by technology that has allowed data to be accessed and interpreted by nonconsumers, those who traditionally did not have self-service access to data prior to new student information systems and learning management systems. This ethnographic study explores the culture of a nascent data governance board at an online consortium in the United States. Data was collected using document analysis, interviews, and a focus group. Once collected the data was analyzed using a two cycle coding method to determine categories, patterns, and themes. The results of this study provide insight for those considering starting a data governance committee, the common concerns or anxieties when it comes to disruptive innovation, and provides considerations for future research on the disruptive innovation in education.

Introduction

Zwitter (2014) points out that data is everywhere and ranges from internet search histories, social networks, text messages; even our personal fitness devices provide data. Data has long been used in businesses to shape marketing campaigns, provide reporting to regulating bodies, and meet consumer demands for more transparent production standards (Otto, 2011). Higher Education is not an exception with current and former student information systems, early alert systems, grades, student and staff applications, and the growing field of learner analytics (Coffman, 2014). In online learning, this is readily apparent as learning management systems track our student activities and provide various reports on logins, duration of visit, what links were visited, etc. (Chou & Chen, 2016).

This study explores the motivations and culture as experienced and articulated by members of a newly formed internal Data Governance committee and its stakeholders within the institution.

Data Governance

Today's higher education institutions are not only charged with protecting the information and data of students, staff and alumni, but also how to ethically manage this data when it comes to institutional marketing, student requests for information and transparency, and research efforts (Jenkins & Potter, 2007; Zwitter, 2014).

Currently, some institutions are turning to data governance as a way to ensure compliance with federal, state, and accreditation requirements for data and manage the ever increasing demands for more data by staff, students, and reporting agencies. Data governance has many definitions; however, they all agree that the term refers to the allocation of decision-making rights and related duties in the management of data in enterprises (Otto, 2011).

Disruptive Innovation

Christensen, Horn, and Johnson (2008) describe disruption as a positive force and the process by which innovation transforms a market characterized by complication and expense into one of simplicity, convenience, accessibility and affordability (p. 11). Disruptive innovation has two stages: in the first stage, an innovator develops a simpler less costly solution than the existing one, while in stage two; technology changes make it simple to build and upgrade the product (Christensen, Horn, & Johnson, 2008, p. 122-123). These two factors result in disrupting existing structures and new innovations and services replace existing ones (Christensen, Horn, & Johnson, 2008).

Methods

This study utilized an institutional ethnography method to collect data from a metropolitan community college consortium in the Midwest. One of the authors was a "complete participant" (Ali, 2008, p. 19) in the data governance process at the consortium and utilized an ethnographic methodology to "direct engagement and involvement with the world being studied" (Reeves, Kuper, & Hodges, 2008, p. 1). An ethnography methodology was selected instead of an autoethnography to focus on the voices of other participants and not just the author's experience (Hays & Singh, 2012). The selection of an ethnography was also due to the scarcity of ethnographic studies of data governance and how these committees are formed.

The target participant population of this study were administrators and staff of the consortium. These 10 individuals ranged from senior to mid-level managers and represented the five departments: Academic Technology, Student Services, Business Services, Center for Teaching Excellence, and Academic Instruction. As a member of the Governance Committee, the author disclosed his identity as a complete participant researcher to the rest of the committee and
conducted ethnographic observation at all Data Governance committee meetings. All members of the Data Governance Committee were invited to voluntarily participate in the study.

Data Collection and Fieldwork Strategies

The author employed the qualitative fieldwork methods of document analysis, focus groups, and semi-structured interviews. The focus group and interviews were recorded and transcribed and the author maintained field notes and interview notes regarding nonverbal cues provided by the subject and environmental settings. To preserve participant anonymity, all participants were assigned a pseudonym.

Data Analysis

This study utilized three methods to triangulate the data. The document analysis consisted of six months of agendas and minutes. The interviews and focus group were transcribed and member checked. All data collected was processed through two cycles of coding. The initial cycle was an open coding process to identify topics, categories, and patterns. The rationalization for open coding is that the data collection is still in the beginning stages (Saldana, 2016) and consists of relatively few data points. Once the first phase of coding was completed and member checked, the author applied a second cycle theoretical coding process based on disruptive innovation theory framework (Christensen & Johnson, 2008) to pull out additional patterns and themes.

Findings

Several themes emerged from the initial open coding cycle and the secondary theoretical coding cycle. Table 1 provides a summary of the themes that emerged from the three data collection methods and the four theoretical themes that emerged from the second cycle coding.

Table 1
Summary of open coding themes and second cycle theoretical coding themes

Open Coding			Theoretical Coding
Document Analysis	Semi-Structured Interviews	Focus Group	Disruptive Innovation Themes Present
Focus on past, present, and future constraints	Minimal experience with data governance	Fragmentation of data	Anxiety due to Disruption
Anxiety over incomplete information	Experienced a tipping point	Concern about external access to data and lack of context	Increased external demands for new and faster data

Anxiety over effectiveness and efficiency	Data / Information Silos	Consistency and structure	Desire for simple solutions
Need for a plan/ Something must be done	Anxiety over stalled initiatives	Culture of data	Technology as a driver of demand for data
Reinforcing constraints			
Lack of affordances and benefits			

Discussion

Theme One: Anxiety due to Disruption: Several categories of anxiety were recorded across the three data collection methods. These categories were: Anxiety over lack of context, lack of involvement, isolation, being forced to be reactionary, insufficient information, and attribution without involvement. While the source of the anxiety comes from many sources, one overarching theme is the committee's minimal experience with data governance. This is to be expected in a committee populated by traditional nonconsumers. The participants indicated that there is a desire to use data more in decision making; however, due to the challenges of getting data, many do not wait for data before making decisions based on more anecdotal or "gut" feelings.

Theme Two: Increase External Demand for Data: Another theme that permeated the discussion and contributed to some concerns and anxiety was the demand for data from external stakeholders such as the systems office, member colleges, state and local legislators, and regional accreditors. The new student information system allows other super users within the consortia system to pull data on member colleges and online learning departments. The reason for these concerns were external users generating reports that do not match up with the reports from the online learning department. The reason for this disparity was the use of different data cleaning protocols that break out data by state and institutional stakeholder. The external users are retrieving all data from the consortium, which includes data from a member college that is not in the same state as the rest of the member colleges. This error had led to several tense conversations about report disparity between what was internally reported and what external user generated reports show. There is also a strong concern that these external generated reports lack the context that internal reports provide and contributes to miscommunications and different reports from the same data set.

Theme Three: Desire for Simple Solution: The desire for a simplified solution also emerged from the desire for consistency in regards to terminology, reporting, and analytics. PA2 stated:

> I want structure, again kind of with the consistency. If you want data, you have to fill out the form again. So everyone knows this is what you do and also having consistent reports. That's why I came up with the report card idea. So that way everybody's talking about the same data the same way.

PA2 expressed a common desire for order and simplicity when it came to reporting and data requests. PA2 had recently introduced a standardized data request form to reduce the constant back and forth to answer questions and clarify issues that was common in previous data requests. The creation of simple solutions is one of the key pieces of Disruptive Innovation theory (Christensen, Horn, & Johnson, 2008).

The desire for simple solution stems from the lack of an organized and accessible comprehensive data set that could be accessed by all. The current state of data at the college is fragmented and "siloed" which results in inconsistency of data. Fragmented data and tracking is cumbersome and requires the user to manually query the data and search many different locations to find the correct data set. The following excerpt illustrates this issue:

> PA4: … I think everyone has their own little data sets, or silo, or they know the person that has one little piece they might be not realizing that there are other data that someone else might have that could help.
> PA4: … I mean a lot of the things we track on spreadsheets. We have all these spreadsheets that are all shared. It's not easy to find what you needs. And they are all saved in different places and named differently.

Theme Four: Technology as a Driver of Demand for Data: This final theme arose by looking at the context of the discussion. While never mentioned by the participants, the presence of technology permeated this inquiry. The data requests, retrieval, and external retrieval were all made possible by the changes in technology, the new student information system and the Learning Management System used by the consortium. While big data and analytics are main stream for business (Coffman, 2014; Otto, 2011), they are still relatively new concepts for education. The ubiquity of the technology behind big data and analytics in the business sector, has led to external stakeholders to pressure higher education institutions to provide similar metrics in their reporting. As PA1 stated in the interview, one of the reasons for the formation of the Data Governance Committee was due to "executive recommendations and hints that there needed to be more application of analytics and data analysis."

Potential Implications

While the small sample size hinders the generalizability of the findings of this study, it does allow for practical and scholarly applications to be drawn from the study. In a practical sense, this study can be used to create an action plan for forming new data governance committees and highlight common issues to disruptive innovation like anxiety, fragmentation, and uncertainty inherent to nonconsumers adopting new services and products while they find how to effectively and efficiently incorporate the new solution into their day to day. Another practical application of the study would be to identify common anxieties and concerns when it comes to disruptive innovation and develop ways to alleviate or less the personal and institutional stress disruption causes for nonconsumers.

From the scholarly perspective, this study indicates that more research needs to be conducted to determine how disruptive innovation is affecting higher education particularly in the area of learner analytics and data governance. While limited, the literature review illustrates the small amount of education literature that currently exists. There is ample room to expand and build new knowledge on how higher educational institutions and leaders can effectively and efficiently leverage their data to better support their students and serve stakeholders. Additionally, the population studied was a nascent initiative and more research need to be conducted to determine how leaders can successfully navigate the organizational, personal, and external issues generated by disruptive innovation.

Conclusion

Despite the small sample size, this study does shed some light on the present state of data, data governance, and culture at a consortium of community colleges. The results of the data collection allow us to draw conclusions for each of the research questions being pursued. In regards to what motivates an organization to develop a data governance committee, this study found that external pressure and intrinsic motivation by a few employees were the impetus for the formation of the Data Governance Committee at the consortium studied.

In response to questions about how does the current culture react to the new influx of data, the respondents agreed that the current culture was characterized by the terms "poorly, haphazardly, frustrated by lack of access, and overwhelmed at times." These sentiments echoed throughout the data collected and were responsible for the high levels of anxiety and frustration in the minutes, interviews, and focus group.

When questioned, the respondents desired that the committee efforts lead to increasing consistency of data definitions, use of data, and understanding of the context of the data. These desires were shared by all participants and there was a strong feeling that consistency would allow them the stability to move the organization forward and better incorporate data into decision making and in internal and external reports.

References

Ali, J. (2008). The utility of participant observation in applied sociological research. *Nexus, 20*(2), 18-20. Retrieved at https://tasa.org.au/wp-content/uploads/2010/10/nexus-202-enexus.pdf

Chou, H.L., & Chen, C.H. (2016). Beyond identifying privacy issues in e-learning settings: Implications for instructional designers. *Computers & Education. 103*, 124-133. doi: 10.1016/j.compedu.2016.10.002

Christensen, C. M., Horn, M. B., & Johnson, C. W. (2008) *Disrupting class: How disruptive innovation will change the way the world learns*. New York, NY: McGraw-Hill.

Coffman, D. (2014). Managing data protection in higher education. *Risk Management*, 33-35. Retrieved at http://www.rmmagazine.com/2014/09/01/managing-data-protection-in-higher-education

Hays, D. G. & Singh, A. A. (2012). *Qualitative inquiry in clinical and educational settings*. New York, NY: The Guilford Press.

Jenkins, P., & Potter, S. (2007). No more 'personal notes'? Data protection policy and practices in higher education counseling services in the UK. *British Journal of Guidance & Counseling, 35*(1). 131-146. doi: 10.1080/03069880701219849

Otto, B. (2011). Data governance. *Business & Information Systems Engineering. 3*(4), 241-244. doi: 10.1007/s1259901101628

Reeves, S., Kuper, A., & Hodges, B.D. (2008). Qualitative research methodologies: Ethnography. *BMJ, 337*(7668), 512-514. doi: 10.1136/bmj.a1020

Saldana, J. (2016). *The coding manual for qualitative researchers*. (3rd ed.) Thousand Oaks, CA: Sage Publications.

Zwitter, A., (2014). Big data ethics. *Big Data & Society. 1(2)*. doi: 10.1177/2053951714559253