CNDLS
THOUGHT PAPERS

A Case for Open & Interconnected Systems in Universities

BY MARIE SELVANADIN

Abstract

The use of the internet continues to grow globally, and people around the world both contribute to and have access to a massive amount of data. How can we harness data to benefit the communities that produce it? In this paper, I will discuss why universities should think about allowing open access to some of their course data, like course descriptions and required readings. I will also discuss the advantages and disadvantages of spending the time and resources to develop Application Programming Interfaces (APIs) at universities.

About the Author

***Marie Selvanadin** is a Senior Software Engineer at CNDLS. She holds an MS in Computer Science from The Catholic University of America. At CNDLS, she designs, develops and researches web applications that meet the pedagogical needs of faculty members and students.*

The Pew internet research[1] on internet adoption shows us that 78% of adults in the United States use the internet. One can agree that the number of global internet users is only going to continue to increase, and corollary to any increase in number of users is an increase in data for analysis and interpretation. The topic of data analysis and interpretation is becoming a huge topic of discussion in all sectors, from business and the sciences to education and the government. Until very recently, it was assumed that the data would only be available to the entity that "owned" it, but recently there has been a call to action to make these data sets available for public consumption. The idea that certain types of data, such as public interest data and general health data, need to be open for interpretation by any interested person is at the core of the "open data" movement. Wikipedia's article on open data[2] provides a more comprehensive list of the types of data that would be useful contributions to the open data movement. Several governments, including the United States government, have started contributing data and have an open data initiative. For example, personnel data based on gender and race and local area unemployment statistics are some of the datasets that you will find on data.gov, which houses the data sets released by the U.S. government. One of the first things that President Obama did after taking office was to establish an open government initiative, stating that "openness will strengthen our democracy and promote efficiency and effectiveness in government."[3]

In this article, I hope to make a case for open data in universities and to explore some open data use cases for teaching and learning. Of course, some types of teaching and learning data cannot be opened to the general public, but other types can; historically, though, university data has only been accessible by a university's senior administrators and information technology (IT) department. In this article, I will frame the university community as a microcosm of the open data movement discussed above and discuss how open data within a university can be beneficial to its students, faculty, and staff. Furthermore, addressing the open data issue within the frame of a smaller community (in this case, the university) provides a model to then think about how to share data responsibly and

meaningfully with a larger community.

Empowering users to tell their stories. Universities use a wide range of computer and web applications to store data, and this data includes everything from personal information, course descriptions, and readings to grades and meal plans. Just as there is increased interest in "big data" in private industries, there is also a lot of interest in this data in the higher education community. "Big data" refers to vast amounts of data that cannot be easily analyzed with simple software programs and databases. It requires sophisticated analysis and complicated software written specifically to be able to handle such large quantities of data. The higher education community is also keen to learn more about the topic of big data, particularly as it relates to another university interest, learning analytics. Learning analytics is defined as "the interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues."[4] Most of the discussion, at least for now, seems to revolve around the idea that "someone" will do an analysis on student progress (as captured through grades, "time on task," and more) and present the findings to the teacher or administrator, who then in turn will "intervene" to help the student in question. While there are a range of views on how effective or appropriate this type of intervention may be (see, for example, the CNDLS Thought Paper by Kruse & Pongsajapan, "Student-Centered Learning Analytics"), it is important to point out that this approach represents a narrow application of data.

"Open data" refers to the idea that data is available to any interested party and, importantly, is open to interpretation by any interested member of the community. In his presentation at the 2008 An Event Apart conference, Jeffrey Veen presented the idea that developers and designers should "enable people to find their stories."[5] While learning analytics certainly offers benefits to administrators and faculty users, we should also consider the productive and creative uses of their data that the students themselves could bring to the table. At CNDLS, we are in the process of building an ePortfolio system (called Pegasus) that would allow a user to easily collect any of her own data that she wants and be able to present that data in meaningful and interesting ways. As part of our Pegasus design discussions, we have also thought about what sorts of data visualizations would help faculty and students evaluate and peer assess ePortfolios. Basic visualizations are crucial to any open data initiative because they provide users with a starting point for understanding the data. Then users can interpret the data with the assistance of that visual analysis and draw their own conclusions. This is where an Application Programming Interface (API) becomes very useful. An API[6] is a mechanism for connecting different systems that would allow developers

to access university data as appropriate and then be able to use that data to help students find connections.

Linked Data. In his TED talk, Sir Tim Berners-Lee talks about the notion of "linked data."[7] Linked data is defined as a "method of publishing structured data so that it can be interlinked and become more useful."[8] Open data is at the very core of this concept because it is the key piece that would enable the discovery of these "links." What would those connections potentially look like in a university setting? The simplest case would be to open up the course descriptions offered by a university and then find some connections between the courses taught so that students could discover the interconnectedness of their learning.

Syllabi Visualization. At CNDLS, we ran a small experiment with exactly that in mind. The data source we used—a sampling of course syllabi collected online—is only open to visitors formally associated with Georgetown University, so we developed the project behind a user authentication system. To narrow our dataset, we decided to focus on the syllabi of courses that were recognized as options for various general education requirements. We manually recreated a subset of the dataset on which to perform some initial analysis and groupings; we created the groupings by tagging courses. (These tags were created manually by reading the course descriptions and identifying keywords.) Tags could represent themes, texts, topics, or other information in the syllabus that would help us visualize connections among courses. This growing collection of tags allowed us to quickly get an idea of the connections that could be formed between courses in entirely different departments. The screenshot below (Figure 1) gives us a list of general education courses at Georgetown that directly or indirectly refer to "ethnicity" in their syllabi:

Figure 1. Results for the tag "ethnicity" on the Syllabi Visualization project

As Figure 1 shows, the results span humanities, sociology, anthropology, and government courses, and that is just the beginning. A student could easily explore themes that he is interested in and track those themes across disciplines, uncovering cross-disciplinary connections. Students could use this tool to discover several courses with the same assigned reading (or author), helping them understand how their course assignments are relevant in varied contexts. A tool like this could be useful for students as they consider their course schedules each semester, and integrative searches like those possible with this tool could quickly lead to serendipitous discoveries and new interests. There are plenty of possible uses for faculty as well; being aware of the many other courses across the university that are discussing particular topics or using particular readings can help faculty situate their courses in a larger academic conversation.

This was, of course, only an initial experiment to try to understand what some of the benefits and limitations of a syllabus visualization tool might be. How might such a system be useful to a student or a faculty member? For one, the connections uncovered through a simple search on this tool could prove to be important catalysts for teaching and learning. From a faculty member's perspective, these connections might provide an opportunity to forge future collaborations across disciplines, and from a student's perspective, the possibility of creating a more meaningful, cohesive lineup of courses is thought-provoking. A good next step for the project would be to provide mechanisms so that the user (a student or faculty member) could manipulate the data to find her own connections. This step would mean two things—we would need to provide an API (the connective mechanism) so that application developers could tap into the data, and we would need to provide different types of filtering mechanisms that would give the user wide-ranging possibilities for mining the data. But for now, the basic visualization provides the user with a good foundation. This is just a small example of some of the benefits of investing resources in setting up the infrastructure for open data and an API in a university, and the advantages would definitely be worth the time, effort, and money spent building it. One of the tasks that needed to be done in setting this project up was data entry. That time-consuming step could have been avoided if there had been a central space where we could have accessed the data without necessarily duplicating it.

Linked Data and High-Impact Practices. In their article on open data, Dennis McAuley and colleagues discuss linked data and assert that linked data provides the necessary connection between sets of open data that would allow educational data to be "liberated from its traditional silos."[9] Learning Management Systems (LMSs) and other applications that are used for course administration already hold massive quantities of student learning data. Using this data only for an "intervention" type of

analysis would be a very limited way of looking at the big picture.

In his article on "The Problem of Learning in Higher Education," Randy Bass discusses high-impact practices, which include learning communities and collaborative assignments and projects, among others.[10] Imagine a scenario where a student is asked to do homework in two entirely separate systems: the Blackboard LMS and a course wiki hosted on another platform. Normally there would be no easy way to bring data from the two systems together, but there are plenty of useful, generative connections to be made if the content could be brought together. It would be ideal if the student could log in to one central site and access their data—course work, grades, course description, and more—and then be able to present any or all parts of it in whatever intentional way the student wants. The ePortfolio tool, Pegasus, that I briefly discussed earlier could be one space to use this open data and provide the necessary tools that a student could use to advance his learning through that data. Another option is to create a space where a student user could directly manipulate her data and then be able to export that data in different formats. Websites such as ManyEyes and GapMinder already provide this type of functionality, and the possibilities for educational use are compelling. Making use of meaning-generating technologies is becoming even more crucial in academia as we look to the future and recognize the value of forging interdisciplinary connections.

Issues for Further Research. I am not suggesting that we need one big system that handles every possible scenario; in fact, as a developer, I would argue against having a monolithic system. However, I would argue that if there was a standard for providing data that these systems agreed upon, then student and faculty users could feel more empowered to use learning data meaningfully. The benefits of having learning data opened up for analysis and also providing a mechanism (an API) to analyze and access the data are many, as we have explored. As we continue our discussions regarding learning analytics, open data, and APIs, there are certainly some major issues that we would need to address. Some of the major issues that come to mind are:

- *Data privacy*: How do we satisfy the privacy requirements? How do we stay compliant with the Family Educational Rights and Privacy Act (FERPA)[11] while allowing users to access some of the data?
- *Amount of data*: How much data do we release? Where will this data be stored?
- *Maintenance*: How do we support and maintain such an infrastructure?

These are just some of the issues that need to be discussed, but we do not need to hold off piloting innovative tools like the Syllabi Visualiza-

tion project until conclusions have been reached. We can begin operating within the current constraints while holding important conversations on how to address the key issues and move forward.

Conclusion. While there has been a lot of development toward open data models in other sectors, there does not seem to be a lot of progress in universities in this particular area. So much interesting research and teaching happens in a university, and to first release some of this data from its current tight constraints for exploration within a university community seems a good first step. A next step might be to think about how to make that data accessible to the outside community responsibly. As we move forward, we should continue to think about the types of integrative connections that would be useful for student learning—and how universities can help enable those connections to be made.

Endnotes

[1]"Internet Adoption, 1995-2011," Pew Internet, accessed April 16, 2012, http://pewinternet.org/Trend-Data/Internet-Adoption.aspx.

[2]*Wikipedia*, s.v. "Open Data," last modified April 10, 2102, http://en.wikipedia.org/wiki/Open_data.

[3]"Open Government Initiative," White House, accessed April 16, 2012, http://www.whitehouse.gov/Open.

[4]New Media Consortium, *Horizon Report 2012,* 22, http://net.educause.edu/ir/library/pdf/HR2012.pdf.

[5]Jeffrey Veen, "Designing the Next Generation of Web Apps" (session, Chicago, IL, 2008), http://aneventapart.com/events/2008/chicago/.

[6]*Wikipedia*, s.v. "Application Programming Interface," last modified April 26, 2012, http://en.wikipedia.org/wiki/Application_programming_interface.

[7] Tim Berners-Lee, "The Next Web of Open, Linked Data," TED Talk, http://blog.ted.com/2009/03/13/tim_berners_lee_web/.

[8]*Wikipedia*, s.v. "Linked Data," last modified April 27, 2012, http://en.wikipedia.org/wiki/Linked_data.

[9]Dennis McAuley, Hanif Rahemtulla, James Goulding, and Catherine Souch, "How Open Data, Data Literacy, and Linked Data Will Revolutionize Higher Education," http://pearsonblueskies.com/wp-content/uploads/2011/05/21-pp_088-093.pdf.

[10]Randy Bass, "Disrupting Ourselves: The Problem of Learning in Higher Education," *EDUCAUSE Review* 47.2 (2012), http://www.educause.edu/ero/article/disrupting-ourselves-problem-learning-higher-education.

[11]Family Educational Rights and Privacy Act, United States Department of Education, http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html.

Bibliography

Bass, Randy. "Disrupting Ourselves: The Problem of Learning in Higher Education." *EDUCAUSE Review* 47.2 (2012). http://www.edu cause.edu/ero/article/disrupting-ourselves-problem-learning-higher-education.

Berners-Lee, Tim. "The Next Web of Open, Linked Data." TED Talk. http://blog.ted.com/2009/03/13/tim_berners_lee_web/

Gapminder. http://www.gapminder.org/.

Kruse, Anna and Rob Pongsajapan. "Student-Centered Learning Analytics." CNDLS Thought Papers, Center for New Designs in Learning and Scholarship, Georgetown University, Washington, DC, 2012. https://cndls.georgetown.edu/m/documents/thought paper-krusepongsajapan.pdf.

ManyEyes. http://www-958.ibm.com/software/data/cognos/manyeyes/.

McAuley, Dennis, et al. "How Open Data, Data Literacy, and Linked Data Will Revolutionize Higher Education." Pearson Blue Skies. http://pearsonblueskies.com/wp-content/uploads/2011/05/21-pp_088-093.pdf.

New Media Consortium. *Horizon Report 2012*. http://net.educause.edu/ir/library/pdf/HR2012.pdf.

Pew Internet. "Internet Adoption, 1995-2011." Last modified April 16, 2012. http://pewinternet.org/Trend-Data/Internet-Adoption.aspx.

Veen, Jeffrey. "Designing the Next Generation of Web Apps." Session at An Event Apart, Chicago, IL, 2008. http://aneventapart.com/events/2008/chicago/.

White House Website. "Open Government Initiative." http://www.whitehouse.gov/Open.

About CNDLS Thought Papers

CNDLS Thought Papers, written by CNDLS staff members and affiliates, are short opinion pieces that address forward-looking technologies and trends that are (or will be) impacting the classroom. They reflect the opinions of their authors and are meant to serve as the beginning of a conversation on the pedagogical value of the tools and approaches they address. Find more Thought Papers at cndls.georgetown.edu/publications/.