Supporting a Passion for New Ideas through Open APIs

Carl Grant

*Associate Dean, Knowledge Services and Chief Technology Officer, University of Oklahoma Libraries,
401 W. Brooks St. Rm 212 NW, Norman, OK 73019, USA*
E-mail: carl.grant@ou.edu

Abstract. Despite the many technological advances that the information industry has witnessed over the past decades, barriers to the full use of information remain. Part of this problem is due to the volume of content and data that is available today. But the information silos created by proprietary system interfaces and the resultant lack of open use of content are far more serious factors inhibiting the full use of information and ultimately the generation of new ideas. This paper will take a look at the current situation and offer some solutions to the current problem.

Keywords: API, Application Programming Interfaces, Knowledge Creation Platforms, Licensed Digital Content, Databases, Proprietary Interfaces, Library Assessment Data, NISO, National Information Standards Organization, LMS, Learning Management System, CMS, Course Management System, Value Triangle, Open Data, Open Access

1. Introduction

When I presented a paper on this topic at the 2016 NFAIS Annual Conference it was just shortly after Valentine's Day, and I used that day of romance to begin my opening remarks. I asked the members of the audience to close their eyes for ten seconds. During those moments, I asked them to think about the time when they first realized they were in love with another person and to recall that moment. It probably involved feelings of weakness, neediness, palpitations, and dizziness. Yet that moment was most likely accompanied by an incredible feeling of completeness, incredible happiness, and, most of all, incredible passion. Passion is described as one of the best feelings in the world. Many believe it's something only we humans can experience. I asked them to hang onto the memory of that feeling. I now ask you to do the same.

Think about this – how many of us can say that we feel that way about our jobs? Now be honest! Let me share something that you might not realize. Librarians often do feel that way. They feel passionate about helping people empower themselves to create a better world and life through knowledge and the creation of new ideas.

I'm an incredibly lucky person because I'm one of those people. I've worked in or with libraries since I was sixteen. One of the reasons I've invested my life in libraries is that two of my bosses in that environment took note of me as a young worker and took me under their wing. They did an assessment of some number of variables and determined I had skills and talents that deserved nurturing and could be grown in order to make me a respected librarian. They did an assessment of some data points, arrived at a conclusion, and acted on it. Of course, I didn't realize at the time how valuable that was and what an investment they were making in me. But I do now. And I've often wondered what it was that they

saw. Was it that I sprouted ideas like a fountain? That I was well read as a result of being a person who consumed books at a voracious rate? When did they see it, and what prompted them to invest in me, just another normal, fairly quiet kid?

I realize there are a lot of kids out there today that are no different than I was, but who won't be as lucky. Some undoubtedly have the raw talent, but won't have the same level of access to knowledge that I had, or the same type of education, or the same supportive family, or the same kind of mentors. But this doesn't mean that they don't deserve the same chance. We have to find the methodology, the tools, and the data to identify them and help them be successful. As Steven Miller said yesterday, we have to improve the confidence of people making these decisions; or perhaps we need a Data Scientist, as he suggested! (See Miller's paper elsewhere in this issue.) Now, I know we can't fix all the problems that stand in the way of creativity. But here's the thing, all of us work in the knowledge business, so there *are* some problems that we can fix.

I am a former businessperson and understand the importance of setting goals. I ask you to join me in setting a goal, a goal toward which I want you to apply that same level of passion you felt about your first love. What is that goal? To help me, to help *us*, create an environment in which every student, every instructor, every faculty and staff member can have access to the existing knowledge and tools that they need in order to create and incubate four new, viable ideas per year. Four is not a big goal, but it's a good start.

At the University of Oklahoma, we're right at twenty-five thousand students and faculty. Applying that goal to this group would result in over one hundred thousand new ideas per year, *just* from my university. That's BIG. Why is this so important? Because these ideas may be big or small, but they may also cure diseases, provide new energy sources, protect people from weather, or help save our planet. As stated in Roberta B. Ness' 2015 book, *The Creativity Crisis: Reinventing Science to Unleash Possibilities* (Oxford University Press), we live in an age where "American productivity has tapered off, the speed of travel is slower now than a generation ago, life expectancy has lost speed..." We need to change this.

2. How can we do that?

Over the years, I've read countless biographies of leaders. A common thread between what I've found in those works and what librarians do is the "wow" moment. You know the "wow" moment; you've seen it on the faces of others. It's that moment when a person gets it- when it all comes together for him or her. It's an incredible, powerful, and sometimes, very brief moment. They could be a student or faculty member, or an administrator trying to figure out why students are failing a course, not graduating, or dropping out entirely, but it's that moment when, because we brought them data, access to existing knowledge, and tools that could help them, they smash through a barrier of understanding on their way to a new idea.

But, you know something? There are a lot of barriers on the pathway to those "wow" moments, far more today than in the past. This is in part because we have so much more data and information available to us from so many disparate sources. IDC has issued a report telling us: "Like the physical universe, the digital universe is large – by 2020 containing nearly as many digital bits as there are stars in the universe. It is doubling in size every two years, and by 2020 the digital universe – the data we create and copy annually – will reach forty-four zettabytes, or forty-four trillion gigabytes."[1]

---

[1]http://www.emc.co/leadership/digital-universe/2014iview/executive-summary.htm Accessed on April 25, 2016.
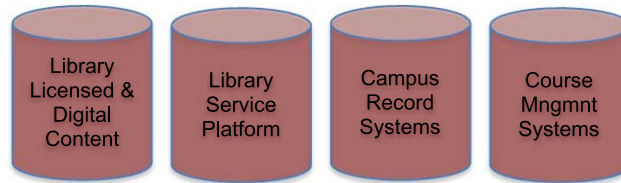
Fig. 1. Typical University data silos.

It's scary because, as you know, the data it represents goes into so many separate silos that frequently can be quite inaccessible and are often called the "dark web." We have the same problem in our Universities and Colleges. At the University of Oklahoma we have silos that (see Fig. 1) include:

- Library Licensed and Digital Content;
- Library Service Platform (LSP);
- Campus Record System(s);
- Course Management System(s).

For further example, at the University of Oklahoma Libraries we have around four hundred databases of content. We have three hundred databases that are licensed content from vendors. Some of these silos are the Library's digital content that we've created. Then there are the University of Oklahoma's own course management systems and campus record systems. All contain data and knowledge. But most of these silos operate independent of each other. Somehow, for us to create "wow" moments and provide valid assessments, we have to knit all of this information together in a coherent way. We need to provide discovery and access of this data, as well as the tools to utilize and *analyze* it, at the same time and in one interface. But there are barriers that prevent us from doing this well, and that get in the way of the creation of new ideas.

Let's look at those a little closer. As I mentioned, at the University of Oklahoma we have about three hundred databases of licensed content. Most of these have a proprietary interface on top of them, which users must learn to search and utilize.

Access to many of these silos is locked into proprietary interfaces, so we can't provide uniform and easy discovery access to that content. This makes it very difficult to track who is using what content, for how long, or with what results. But there is a solution.

We need what are called Application Programming Interfaces, or APIs. For those of you who don't know, these are similar to the electrical sockets in the wall, only they're software related. They are a standardized way to connect to a database through another interface, like a Discovery system, and to query a database and get results as data in a standardized manner. These can be powerful tools that allow value creation for end-users, and new knowledge and idea creation in powerful and exciting new ways that I will mention later. But first, we need to provide access through those APIs. Here's the difference that could make. We've done the analysis (see Fig. 2) at our university and out of those three hundred licensed databases, which are from eighty-seven different organizations, there are seventeen organizations that give us an API interface. Out of those, there are two that provide us with what I consider to be full-function APIs; that is, I can do anything I can do in the proprietary interface through the API.

Another major barrier we face is the legal language that imposes severe restrictions on the use of the APIs or the data delivered through the APIs. Trust me, I understand intellectual property and trade secrets, and I also understand that vendors want to protect data from extraction and reuse, particularly
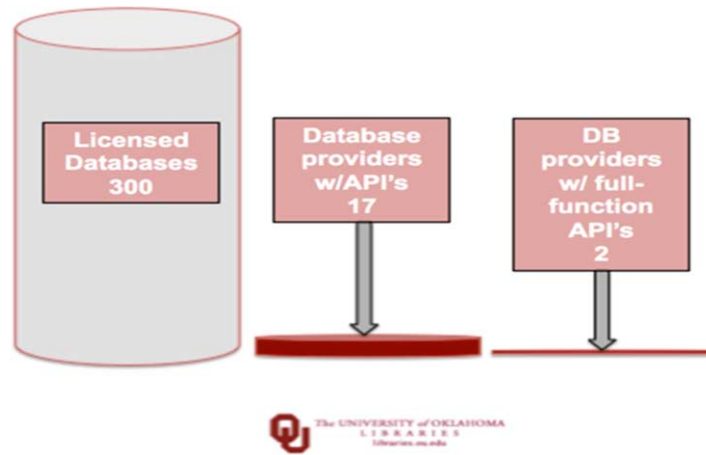
Fig. 2. Analysis of API availability/type for OU's licensed content databases.

when they've manipulated and created that data using the IP within their systems. (Although, let's remember, in many cases the *starting* point data was ours!) Ok, I somewhat get this; but please, let's make this simple. Librarians are reasonable people. Put reasonable protections in our contract and we'll sign it. But, as Nike says, "Just do it"!

But librarians do NOT want to hear that because a vendor considers another vendor to be a competitor, they can't have access to the APIs. Remember, please, that other company is also my supplier and they have to be able to use those APIs on my behalf! (I have an example where we've had an agreement ready for months. Both sides agree no more changes are needed; yet neither will actually sign it! It's insane!) Please, please make it happen! Let me describe the absurdity of this situation in another way.

Let's say that you are asked to speak at a conference and the organizers say to you, "There are approximately two hundred people registered for the conference." You think, "Wonderful, this will be a great opportunity to present my point of view to those two hundred people, and to do so in a very easy way." And, of course, you are right. But then, what if the conference organizers say to you, "Yes, we'll provide you with the opportunity to go from room to room. All sixty-six rooms will have three attendees in them, and you can give each group your talk and do Q&A." Now you've got a thirty-minute slot in sixty-six different rooms, so. . . . it's going to take four solid days of your time to do it.

Is that OK? Probably not. I'm pretty certain if you heard that, your response would be to tell them (politely) that they had lost their minds. Yet, is this really any different than what content vendors are doing to libraries? In my mind, it isn't.

Oh sure, I've heard the responses here: "Those that need to use our content already know about it and come straight to it." Based on my years of experience, there are a lot of users who might benefit from awareness and access to vendor content, but that aren't even exposed to it, and certainly aren't connecting to it as a result.

I've also heard the following: "The librarians know our databases, they'll connect the users to them." True, librarians are the experts at leaping over information barriers, but there aren't enough librarians to meet the need (at least, not till we can virtualize them and distribute them across the web). Plus, with the financial pressures today, this approach isn't very realistic or efficient. There are some other barriers we need to solve:

- Authentication/Authorization systems that don't easily talk to each other and, as a result, lead to pay walls, especially for remote users.
- A failure to share metadata and/or full text for indexing in library Discovery systems, or to load that metadata quickly rather than loading it weekly or even monthly.
- Finally, there is imprecise or totally incorrect metadata, which happens all too frequently.

If we can't get these solved, then I need to ask this question:

> How do we even begin to think about getting valid assessment data on the value of our library collections for our students, faculty, and staff, when we can't easily connect them to that content, track their usage, or analyze what they're doing?

Well, let's pause and think about passion again. I'd say those barriers just mentioned pose some rocky ground in our vendor/librarian relationships, and are areas where we mutually face significant challenges. Of course, at the core, if we share the passion for idea creation I've described, we can meet these challenges.

Let's start by building a "value triangle" (see Fig. 3) on top of and connecting these silos in order to accomplish that goal.

It would be a place where we can jointly create new, sustaining value in order to offer our mutual users real differentiation and access to the knowledge needed to create those "wow" moments, and thus all those new ideas. In addition, we'll be able to offer our administrators the tools they need to assess the value of our collections, services, and libraries.

The value triangle would allow us do a much better job of making possible the integration, mash-up, analysis, and use of content from those silos. It will create new revenue streams for content providers and more "wow" moments for librarians.

We see this happen in other fields all the time. Here's an extremely well known and well-remembered instance when Steve Jobs made the original announcement about the iPhone. (Watch this YouTube video: https://www.youtube.com/watch?v=9hUIxyE2Ns8 from time marker 2:23 through 3:06).

Note what happened there. Three silos – a phone, a music player, and an Internet communication device – all now incorporated into one device. Of course, the rest is history. Apple ran away with the marketplace for years, and it wasn't because any of those three devices was the best in its field. It was the *integration* that transformed our thinking and gave us the device that we have in our pockets today. Now, I'm passionate about providing that kind of value triangle to our universities' libraries' users. But
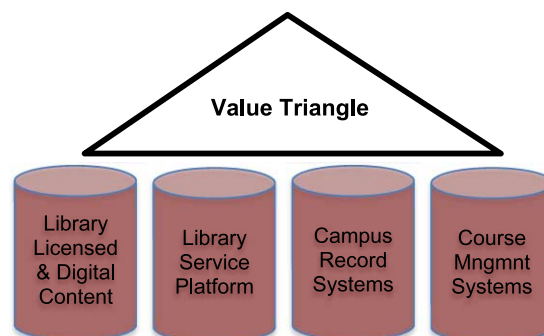

Fig. 3. Value Triangle on top of and connecting Data Silos.

I can't do it alone. Let's all apply that passion for innovating, enabling, or creating this value triangle so our users can create new knowledge and ideas!

Here's how content providers can help make this happen.

1. Enable tools and interfaces beyond their own to sit atop the content silos. Ensure that APIs exist, and that they are full-function; in other words, they should deliver ALL of the functionality of a product's standard interface, including APIs, using standard calls.
   When I was in the vendor shoes I made my own development team use the same APIs to build the product that they gave to the customers for their use. It's called "eating your own dog food," and it's sorely overdue in the content business. If vendors would do this, they would, by default, develop full-function APIs (and documentation) that are good and that provide real value. In addition, it would position them to support the value triangle.
2. Vendors need to change their approach from "making customers pay for access" to "making customers *want* to pay for access." I know what I'm asking requires investment. But by delivering superior value through ideas like we've been discussing, content providers will have customers like me banging on their doors to use their products. Not because I have to, but because I *want* to. Plus, I believe other new revenue sources would arise for the content vendor offering reasonable subscription pricing for those APIs. For instance, librarians, as customers, would also be willing to pay for online courses that we can take to understand the utilization of the API's.
3. Replacing old revenue streams with these new ones will make it possible for vendors to support more open access. Open access isn't going away. Financially, librarians have little choice but to move in that direction. We all have to learn to deal with that fact, and creating new revenue streams now, while there is time, would be smart.
4. Vendors need to join and fully SUPPORT standards and best practices. They can do this through NISO, or similar organizations.
5. Vendors need to know that when I'm looking at assessment data I'm not trying to simply measure usage counts, I'm trying to measure *impact*. And I understand that, in doing so, I need to respect privacy laws, i.e., HIPAA, FERPA, and other national, state, and organizational privacy laws.

But for me to accurately assess impact I need to:

a. Be able to link usage counts to details of the usage, such as:

   - What percentage of the article was read
   - At what time of day
   - Over what amount of time
   - If a citation tool was used to indicate the article was cited

b. Know if the content was linked to from our Course Management System;
c. Know if a text analysis tool was run against the content;
d. Know in some anonymous but meaningful way what group of users this represents (I know I can't get granular detail here!).

Then I need all this data linked back into our other information silos on campus so I can analyze the impact on that class of students; i.e., Do they get a better grade than those who don't do these things? Does this contribute to their graduating in four years instead of six?

As a librarian, here are some things that I'm being passionate about on my end to help enable the idea creation I'm talking about:

1. I'm helping our campus put into place Data Governance policies. This won't be true of all colleges or universities, but I'm betting many will not have data governance in place. This task is huge and involves participants from all over campus. But as we open up access to data, we have to establish the ground rules.
2. We are:

   - Installing "infrastructure" that will allow the interconnection of all the databases and information silos developed on campus through a common set of APIs, as these will help enforce that Data Governance policy. We've gone to RFP on this step and are in the process of procuring and installing this.
   - Supporting NISO through our organizational membership and participation.
   - Looking at every product we buy to determine if it has a suite of full-function APIs that we can utilize.
   - Building multiple open access repositories that we're filling with Open Access content and Open Educational Resources, as well as Open Data. Open Access matters. We believe in it, and we are supporting it.

I attended the ACRL Conference last year and was able to hear Larry Lessig's closing keynote. It was an incredibly powerful talk, but there was one moment that really underscored how important open access can be.

(Watch this video: http://acrl.learningtimesevents.org/keynote-lessig/ from time marker 47:08 through 48:50.) Now, maybe the reason that video punches me so hard is that I'm a diabetic, and I, too, have a disease caused by the pancreas. Not as bad as the cancer described in this video, but *definitely not fun*. Not a day goes by that I don't hope someone will find a cure for this disease. *Not a day*. And maybe it will come from one of those one hundred thousand plus new ideas that could originate from my university. Or anyone else's. I want to improve the odds of that happening.

We're also creating a "Knowledge Creation Platform." It's open access, which I wrote about in an article published in ab *Open Access* journal which is accessible here: http://0277.ch/ojs/index.php/cdrs_0277/article/view/32/78.

So, let me briefly describe what a Knowledge Creation Platform is. It's a tool that sits firmly inside that "Value Triangle" I've been talking about, because it provides new functionality for bridging across our content silos, offers users all the tools needed to analyze and draw correlations from existing knowledge located in many silos, and will be a tool for creating new knowledge. Existing knowledge might be a data set or a journal article, or it might be an interactive, online research collaborative. The platform will include search, extensive analysis tools, office-like tools, data management tools, programming support, and citation, publishing, and repository tools- in ONE platform. The analysis tools that exist now are amazing.

Let me share a story. I was working with the CEO of the company developing the software we're using to build our Knowledge Creation Platform in that "value triangle." He described work he was doing for the Gates Foundation analyzing disease and its causes. He noted that they use their software to look at data variables from a wide range of digital sources, and use all types of analytic devices to search for connections and correlations that could predict which children would ultimately develop a disease in a foreign country.

As I listened to him, I commented that this is really no different than what we're trying to do in predicting what students will drop out or fail to graduate. Or which researchers will fail to win their grant applications. I told him, "We need to use those same tools with different data to arrive at a different set

of answers." That is exactly what this company does. In fact, the company's name, Exaptive, means "a feature having a function for which it was not originally adapted or selected." In other words, I can pick up the tools that researchers are using, that serve a similar function with other types of data, and apply them to mine.

Behind the scenes is this powerful tool kit that even the average researcher can use to create and run all different types of analysis. The researcher can then simply plug modules together, move data between them, and instantly see the results on the right. Even better, if they agree, you can see who the authors of those modules are, and contact and collaborate with them in doing your research.

Finally, we're hoping, as a later step, to combine what we're doing with our Knowledge Creation Platform with what we're doing in Virtual Reality Systems. You can see some of this by watching this YouTube video, in its entirety: https://www.youtube.com/watch?v=tmL3T28Ud1k.

3. Summary

Now, are there big challenges in what I've described? You bet there are. There are new revenue models to be created, new license agreements to be written, and new tools to be created. But I also believe there are big thinkers and innovators among content providers and at my university; people who, provided access to the right knowledge and right tools, could create those one hundred thousand-plus new ideas per year. I passionately want to see more of "Wow!" expressions in my library. I need the Information Community to join me so we can make this happen, not only at my university, but also all around the world.

About the author

Carl Grant is Associate Dean for Knowledge Services & Chief Technology Officer at the University of Oklahoma Libraries. Previously he was Chief Librarian and President of Ex Libris North America. He's active in ALA, ACRL, LITA and NISO. Mr. Grant holds an MLS from the University of Missouri at Columbia.