Caught in the middle: Scholars, publishers, librarians and information revolutions today and tomorrow

David Shumaker

Clinical Associate Professor, Department of Library and Information Science, Catholic University of America, Washington, DC 20064, USA
Tel.: +1 202 319 5551; Fax: +1 202 319 5574; E-mail: shumaker@cua.edu

Abstract. All of us – scholars, publishers, librarians – are being pulled apart by the forces of two concurrent revolutions in information access and processing. Every day, we are reminded that Stewart Brand was right: "Information wants to be free, and information wants to be expensive". Meanwhile, smart appliances and self-driving cars show the way to the Internet of Things in which information is not just for human consumption, while people crowdsource, crowdfund, and share everything from lolcats to regime change. These twin revolutions have already driven fundamental changes in publishing and in library services, and more changes will be coming. Meanwhile, the scholar's limited time and attention are overwhelmed by the rapid changes taking place. By shifting their focus to the scholar's personal information management needs, librarians and publishers can collaborate to deploy new solutions and improve the scholar's information environment.

Keywords: Personal information management, embedded librarianship, scholarly work

All of us, authors, publishers, librarians and others, find ourselves today caught in the middle of an information revolution. This revolution has been underway for over twenty years. We all know this, in fact, it has become a cliché to speak of the information revolution.

Still, we cannot avoid it. It is disrupting our traditional business models, our processes and our strategies – everything about what we do, in higher education, in publishing and in librarianship. This notion of disruption, while sometimes contested, has also become something of a truism, thanks to Clayton Christensen [2] and the many writers who have taken the idea from him.

What I want to add is that not only is there a revolution underway; not only is it disrupting our traditional models, but that in fact there are at least two revolutions, and in each case they are pulling us in opposite directions at once.

The first example of this is Steward Brand's famous saying, now almost 30 years old, that "information wants to be free, and information wants to be expensive [1]". He was right of course and we have seen the principle in operation for a long time now. The vast growth in freely-accessible content on the web has supplanted commercial publications for many purposes. Wikipedia is perhaps the poster child for this. Yet at the same time, content vendors have developed ingenious content licensing systems and rights management regimes to ensure that they get paid for content not just once, as in the old "doctrine of first sale" model, but every time it is used. So in our current society, the content-consumer is dependent on both: the free and the expensive. Personally, in my work, I use both free and commercial information. I would be hard pressed to do without either one.

Another example of an information revolution that is now underway occurred to me as I was reflecting on the news regarding autonomous vehicles, the Internet of Things, and so forth. At the same time

that pundits are touting the emergence of this new mode, we are also witnessing the proliferation of social information. We now have more ways to share our facts, news and opinions with friends and total strangers, anywhere and everywhere, than we could have dreamed a generation ago. So, I propose that we are simultaneously moving in the opposite directions of "information is for machines" and "information is for people". This all makes for a pretty dynamic and bewildering environment. As we think about these grand developments all of us, scholars, librarians, publishers and vendors, may see ourselves as being caught in the middle of them.

There is another sense in which we are all "in the middle". Each of us is the center of our own information universe. So, maybe it would be more useful to stop approaching our topic from the point of view that we are pawns in the cosmic struggles of expensive vs. free and machine vs. social. Maybe it would be better to start from the point of view that each of us is the center of our own information world, and think about how we are functioning, and how we would like to function.

This line of thought leads directly to the idea of personal information management. Each of us has a personal information management system (PIMS). It is the sum total of our efforts to discover and manage the content that is relevant to our work and our daily life. It may not be very systematic, but it is a system nonetheless.

Figure 1 provides a representation of a personal information system. Our personal information systems include several functions, and fundamentally they are the same function across different professions and disciplines and modes of life.

The first is discovery. The discovery subsystem delivers new information to us. I think my discovery subsystem is typical. It includes an eclectic array of sources and contributors. For scholarly and professional articles, I subscribe to alerting services from several commercial databases and from Google Scholar. I also subscribe to email distribution lists that are relevant to my work, and to individual journals. I monitor a few blogs. I participate in webinars, conferences and other presentations, most of which I find out about through my email subscriptions and alerting services. I could be following various Twitter streams, though in fact I am not doing that. Occasionally I will seek out new information by doing *ad hoc* searches in the very same databases that are providing me with alerts. But it is interesting that
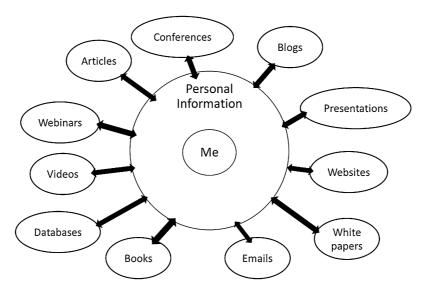

Fig. 1. A personal information system.

most of what I put into it comes by way of "push" services, not by my "pulling" content through *ad hoc* searching.

The second subsystem consists of evaluation and selection. When I am notified of new content, I have to evaluate it, and decide if it merits retention in my information system or not. Sometimes I look at the item, sometimes I delegate the decision to my student research assistant.

Then, if a particular item is reviewed and deemed worthy of retention in my personal collection, I have to create a metadata record for it. I have to keep track of its provenance: where did I obtain it? What was the URL? Are there any limitations on its use? Metadata management is a job I generally delegate to my student assistant.

I also have to save the item and link it to its metadata record. Often I retain both print and digital copies of the documents in my personal information management system, which complicates this task. Then, when I am updating course content, or preparing a presentation, or an article, I have to be able to find what is relevant. For many needs, it is the first place I look. Only if I am pretty sure that I am in a new area and have not collected the information I need am I going to start by pulling information from outside sources.

Finally, my personal information management system supports my publishing my own intellectual work. For that function, I use many of the same mechanisms that provide me with input in the first place. I write articles. I contribute to email lists. I give presentations and webinars, and participate in conferences. I write a blog.

So that is my personal information management system, and I speculate that it is not so very different from yours. The next question is, how well is it meeting my needs? I cannot speak for anyone else, but I wish I could say that my personal information system ran like a sports car, powerful, perfectly tuned, beautiful, and above all fun to drive. But that is not the case. The truth is quite different. The reality is that my personal information system is more like an old jalopy than a high-performance sports car. It barely provides me with the basic functionality I need. The subsystems are not well aligned and it requires continual repairs just to avoid a major breakdown. Here are some of my problems:

I cannot track all of my sources. I do not even know what alerts and subscriptions I am running.

Some of the commercial database alerts create roadblocks for me. They send me links that only work if I am connected to my campus network. If I am reading the alert from home, and I work at home a good deal, the links do not work.

And then, integrating new content is far from seamless. I am using a personal bibliography management tool. It supports importing metadata from a variety of databases, but the process is never seamless. Entries always have to be cleaned up. Not every field from the database populates the correct field in the bibliography manager. Some data populates into fields where it does not belong. Subject terminology that is useless to me comes along, and clutters up my database, and I do not have a way to exclude it, except by manually deleting it after import. Meanwhile, it is cumbersome to add the terms that reflect my interests.

For other documents, metadata has to be entered by hand. If there is a whitepaper, or a conference or webinar presentation, the source – that is, the URL, does not come along with the document, and I must remember to record it so that it can later be entered into the metadata record.

I am also hampered by the fact that I do not store my documents, which are almost always in PDF format, in my personal bibliography manager. They are in a separate cloud-based repository. This means it can be tricky to ensure that everything in my repository has been cataloged in the bibliography manager, and everything in the bibliography manager has been stored in the cloud. The workflow is complicated and error-prone. Of course, some media types do not integrate with my cloud storage at all. Email is the

worst example. I use Microsoft Outlook. Unless I take the cumbersome step of converting the text of an email message to a PDF document, it can only be retained in a usable format in my Outlook account.

Moving to the retrieval subsystem, the search and retrieval functions of my bibliography manager are primitive. Searching is tedious and cumbersome. And since the document store is not integrated, searching the full documents is beyond hope. This is especially annoying on those occasions when I know I have seen an important comment or insight somewhere, and I really need to recall it right now, and I do not remember when or where I saw it. The search function is of no use, and I am reduced to combing through my document collection, physically or digitally, or trying to reconstruct the process by which I found the document in the first place.

And finally, support for assembling publications and outputs as I create them, while integrated with my authoring tools, is complex enough that I have to re-learn it every time I need to use it.

The reader may ask, as I ask myself, why do I allow this situation to continue? The perfect solution to all these problems may not exist, but there are almost certainly better tools than I am using at present – tools that could ameliorate my predicament. But I do not know what they are, and the reasons why illustrate a well-known point about human information behavior. Over the past generation, information scientists have proposed a number of models of human information behavior. For an overview of these models, see T.D. Wilson's article, "Information behavior models", in the *Encyclopedia of Library and Information Science*, 3rd edition [3]. One feature these models tend to have in common is that they show information seeking, management, and use as taking place within a context of other activity. In other words, information behavior is almost always a means to an end. We need information for sense-making; we need it to resolve challenges in life and work. Once we have the information we need, we focus on the challenge in our work or daily lives, applying the information as we need it. Once the challenge is met, the anomaly is resolved, we focus elsewhere.

Accordingly, managing my Personal Information Management System is not a big part of my life. I do not have a lot of time to feed this beast. I can either keep up with my field and the demands of my work, or worry about personal information management. My own priorities, and the incentives presented to me, encourage the former, not the latter. Not surprisingly, the former is what I choose. In fact, I estimate that I allocate my time as follows:

- 40% teaching;
- 28% each service and research;
- 4% feeding the PIMS.

Figure 2 graphically represents my time allocation. I think I am fairly typical of an academic in my position, and that the proportion of time devoted to managing the personal information management system is likely similar for other knowledge workers.

This assertion does not dispute the studies that find that knowledge workers spend great amounts of time looking for and processing information. I am processing info all day long in one form or another. But it is consistent with the information behavior models referred to previously – we seek information as a means to some other end. For me that is the teaching, service or scholarship work I am engaged in. I just do not spend a lot of time engaged with information for information's sake, and certainly not with the systems to manage it. So undoubtedly I am missing things.

In sum, I do not have the resources, in terms of time and attention, to identify, implement, and maintain a new and improved personal information management system for myself. This is a shame, because I know there are information systems vendors and content vendors out there who have good products and services to offer. Some of them have tried to reach me, and I have even tried to respond. But my
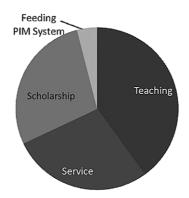
Fig. 2. My time and attention.

responses fall short, because I cannot devote the necessary time and attention, and nothing happens. The picture seems bleak, with the individual knowledge worker – me, in this case, caught in the middle of a barely functioning information environment, yet surrounded by vast information capabilities. Yet the situation is not hopeless, and I would like to suggest a pathway out of the predicament.

The situation can be improved by forging better collaboration among vendors of the content and tools, the knowledge workers who need these tools, and librarians. The vendors have the tools and are developing better ones. The knowledge workers need these tools, but their time and attention are focused elsewhere.

This is where librarians can play an important role as team information and knowledge manager. One of the things they can do is help the team deploy the best tools available. In fact, I know that there are some librarians who are playing this role today. It requires them to:

- Keep up with the market and technology innovations;
- Have a deep understanding and personal connection to the members of the team;
- Make it easy to implement and sustain the best tools.

As a teacher/scholar/member of a professional community, I would find it valuable to have a librarian playing this role. The librarian would relieve me of the need to evaluate many different tools myself and to try to understand their strengths and weaknesses, and to pick the one I think will work. The librarian would be someone sitting on my side of the table who knows me, knows how I work, knows the market-place, who will give me good advice, and whose advice I can trust. I would be able to concentrate on my teaching and scholarship, and at the same time have a more effective personal information management system, with the confidence that when further innovations become available, I will be getting effective help to implement them as painlessly as possible.

As a scholar whose specialty is embedded librarianship, I see this as a natural development in the field, where librarians have an opportunity to be change agents and add great value to their institutions.

To summarize this vision, the ongoing information revolutions of our time are pulling knowledge workers apart and disrupting our old models and processes. Vendors are developing ever more robust and effective processes and services to help knowledge workers strengthen their personal information management systems, but have great difficulty in connecting with their potential customers because the latter have insufficient time to focus on the marketplace for these solutions. They feel caught in the middle as they know they need better tools to manage the multiplicity of channels they have and the re-sources they collect. If librarians develop strong, collaborative relationships with knowledge workers in

their organizations on the one hand, and vendors on the other, they can step up to the role of information and knowledge manager, evaluating, implementing and sustaining the best solutions to meet the needs of the knowledge workers – and provide the insightful feedback that will enable vendors to be responsive to real end user needs.

About the author

David Shumaker is Clinical Associate Professor, Department of Library and Information Science, Catholic University of America. Previously, he was Manager of Information Services at the MITRE Corporation. His book, *The Embedded Librarian*, was termed "a must read for every librarian" by Maureen Sullivan, former president of the American Library Association. He blogs at http://www. embeddedlibrarian.com.

References

[1] S. Brand, *The Media Lab*, Penguin, New York, 1987.
[2] C.M. Christensen, *The Innovator's Dilemma*, Harvard Business School Press, Boston, 1997.
[3] T.D. Wilson, Information behavior models, in: *Encyclopedia of Library and Information Sciences*, 3rd edn, Taylor & Francis, London, 2010, pp. 2392–2400.