

After Launching Search and Discovery, Who Is Mission Control? A Tragicomedy in 8 or 9 Acts

William Denton <wdenton@yorku.ca>
Adam Taves <ataves@yorku.ca>
York University Libraries, Toronto

This work is licensed under a
Creative Commons Attribution-ShareAlike 3.0 Unported License

Performed on 15 October 2010 at
Access in Winnipeg MB
<http://access2010.lib.umanitoba.ca/>

Audio and slides also available: <http://hdl.handle.net/10315/6202>

Cast:

- Adam (an information literacy librarian)
- Bill (a systems librarian)

Also featuring as themselves:

- Cameron Hoffman (Concordia University)
- Sheril Hook (University of Toronto Mississauga)
- Seymour Lubetzky
- Caroline Stewart (University of Guelph-Humber)

Slide 1

[ACT 1: Introduction]

Bill: Hey Adam. So what do you think about our decision to go with this new search and discovery layer? Free software, good-sized user community, Solr on the back end? Fran in

cataloguing thinks it will be terrific.

Adam: Bill, as per usual, I have no idea about what you're talking about.

Bill: Yeah. It's awesome. Half Perl, half Python, half Ruby--

Adam [interrupts]: I think you've gone off the rails.

Bill: There's a great bunch of people working on it. The one uncool thing is they use SVN, not Git--

Adam: Git?

Bill: It's a distributed version control system. In fact I've forked it into Github and we work--

Adam: Let's start at the beginning. What is the name of it?

Bill: Correct.

Adam: What?

Bill: Exactly.

Adam: What is the name of this search and discovery layer?

Bill: You've got it, old friend.

Adam: Stop messing around. Will you tell me, what is the name of this search and discovery layer?

Bill: Wait, I believe we have a slight misunderstanding. Our name for the system is actually the word What. But when you say "What is the name of this product," you're asking a question, but I thought you were making a statement.

Adam: I'm glad we got that sorted out.

Bill: Indeed. We call it What, but really it's VuFind.

Adam: VuFind. Let's call it VuFind from now on. But this is the first I've heard of any of this.

Bill: I've only sent out about a dozen emails on it.

Adam: I didn't see them.

Bill: Stop playing with your iPhone for five minutes and have a look, then. VuFind is our new search and discovery system. We'll be implementing it over the summer.

Adam: Who from information literacy is going to be on the implementation committee?

Slide 2

Bill: IL? How much of an IL issue is this to begin with? All we're going to do is make your life a whole lot easier. It'll have everything in it... the catalogue, our e-resource database records, all that. You guys have been griping about having all this siloed information for years, but now it'll all be in one place.

Adam: Well, yeah, that's true Bill. But moving to a search and discovery system has some challenges.

Bill: Here we go.

Adam: First of all, we need to figure out how to display all this information, we need to talk about search logic and making that easier for users, and who knows what sorts of opportunities we may have here... we won't have to spend as much time teaching our users how to search and use the system... meaning we can focus on something a bit more...

Bill: I don't know why I bother talking to you. Look, this isn't going to affect how you teach. You'll still be just be talking about all your usual little things like how to search by title and use Boolean and get into databases and all that, won't you?

Adam: Well, yes of course we will be. But we have other things we'd like to discuss... how to evaluate information, issues around the production of information, I'm sure if we got the IL people around we could hash out things out a bit more clearly. I have Caroline on videophone and she may have some thoughts on the matter. Caroline, you work in instruction. What do you want from a search/discovery system?

Slide 3: VIDEO: Caroline Stewart 1: <http://www.youtube.com/watch?v=BXvs2i-M208>

[ACT 2: Cataloguers]

Bill: Caroline there may be in IL but she's making some good sense. You have all this huge big complicated mess of results you're getting back and you want to present it clearly to the user, divided up by format. Helps the user pick and choose. Do they want a book? A journal? Audio? It's all there in the MARC records. And I have to say that Fran in cataloguing may not be up on your fancy IL stuff, but she's indispensable in all this. I mean, I hack MARC, but she is all over it like maple syrup on a pancake. Couldn't have done it without her---couldn't keep on doing it without her. We hacked up VuFind majorly and added in all kinds of neat stuff based on what she knows about our records. Faceting, authority control, Kolmogorov analysis of related expressions we found in Gutenberg, and when she got into the GIS data? It was *frightening*.

Adam: You guys did some nice work there, tweaking the relevancy rankings and digging out the useful bits of data to show in results listings. In fact, because all that basic stuff is taken care of so well, it lets us get on to those loftier goals I was talking about before, like---

Bill (ignoring him): In fact it reminds me of a Lubetzky quote.

Adam: Lubetz-who?

Bill (exasperated): Seymour Lubetzky (1898-2003), he's only the greatest cataloguer of the twentieth century. Caroline there was saying that the user should be able to enter in one simple query and then get back a large amount of information presented to them in a very carefully constructed way, based on the metadata we have. One simple query, rich presentation of results. Lubetzky was at a conference in Toronto in 1977 and I just happen to have a recording of it with me. He got asked a question about main and added entries and whether you would still need these concepts in the future when anything can be an access point because everything is online. Online! And this is 1977. Here's an edited clip of what he said, and he uses George Washington's farewell address as an example:

Slide 4: <http://www.frbr.org/2005/12/06/lubetzky-audio-clip> [edited clip]

Bill: Seymour Lubetzky. "The catalogue has to tell you more than what you ask for." And now with VuFind, we can really do that, far beyond what he was talking about. You ask it a simple question for a couple of keywords and it gives you back good relevancy rankings, faceting, book covers, recommendations, links to available full text, it's all there, and this business Caroline's talking about, presenting the results in a particular way, we can do that too. It's all in the metadata.

Adam: The metadata.

Bill: The metadata.

Slide 5

Adam: I'm looking at it in a different way. We need the metadata, true. But you're saying that's what we have and that's what we'll show. But what I would like is for the metadata to ... [grasps for the word] ... to *disappear*, so that the larger goals of information literacy can be served. Caroline isn't saying we can divide up the presentation of the results by format because it's in the metadata. She's saying we should divide it up because that's what the users need, that's what students need to know, and if we can show them that in the search results then we don't have to spend half an hour in the classroom telling them how to understand ISBD or whatever strange way our old system used to format results. The catalogue tells us more than what we ask for: it tells us what we need to know to delve into the disciplinary discourse.

Bill: Did you just say "delve into the disciplinary discourse"?

Adam: What I really like about some of the things you guys have been working on is that it allows the tools to start disappearing. So sure, I have LibX in my browser and that's all fine and good. But the great thing about it is how it underlines any ISBN it detects and just runs a search against our catalogue. It in many ways it is invisible but just does what I want. Wouldn't it be great if it could tie me into any full-text holdings from other sources like Google or Internet Archive? Zotero just quietly sits in my browser without any fuss... that's what we want. Not fiddling with complicated-to-use tools and having to teach them, but talking about what gets found.

Bill: So with VuFind, you mean we should assume that the students have been using the Internet for years and they know the basics of all of that, and you can get on to the higher level stuff. You don't need to show how to do a special title-author search because they can put a few

words into a general keyword search and it just works, like every other search engine.

[ACT 3: Information literacy]

Adam: Well, yes and no. I still need to be able to use sophisticated search techniques. If I wanted Google, I'd just use Google. But getting back to your average undergrad. They need to understand things like who wrote the book - what makes that person qualified to speak on the topic. Who published the book - is it --

Bill: Yeah, but that's really some bullshit, isn't it? I mean, come on. These loftier IL goals, isn't that all just basically stuff from a grade ten media studies course, with a bit of Neil Postman thrown in?

Adam: They had media studies back when you were in grade 10? I didn't want to make things too confusing for you. Getting back to "disciplinary discourse" - how do people in a particular subject area talk to one another?

Bill: I guess --

[Bill continues trying to interrupt]

Adam: How is publishing in high energy physics different than, say, publishing in ancient Greek history? How do psychologists communicate their research to the academic community? Or, and we hear this one all the time, why can't we just digitize all the damn books in the library and be done it with?

Bill: Well, that's because of copyright and intellectual property issues and --

Slide 6

Adam: Exactly. IL Standard 5, Performance Indicator 1, Section d. Wouldn't it be interesting if, when we linked to full-text, there was some little clue as to the conditions of access. Like - "No copyright", or "licensed access secured". Then the catalogue is directly supporting an IL competency. Or another issue, how do you learn from your mistakes? A particular approach is not working well, but may tell you something about how to improve it and how to look for information about a certain --

Bill: Well, I do that all the time. It's basic to systems development and programming when you're debug--

Slide 7

Adam: Standard 4, Indicator 2, Section b. That's much more interesting than talking about how to limit your search to books.

Bill: Wait ... that media studies crack ... I just figured that out! Hey!

[ACT 4: Setting priorities]

Slide 8: prioritization audio file

Slide 9: VIDEO: Caroline Stewart 2 <http://www.youtube.com/watch?v=p5Qpd0R43BY>

Bill: [pointing to screen]: See!

Adam: Okay, point taken. I should prioritize a bit better and explain the methods to my madness. But on the explanation front, my friend Sheril has a good point.

Slide 10: VIDEO: Sheril Hook: <http://www.youtube.com/watch?v=0Xxc3p3ng64>

Bill (turns away in a huff at Sheril's "just programming" remark)

[ACT 5: Staring into space]

Slide 11

Adam (checking notes): OK, so, let me see ... yeah. We were looking at all this in the Information Literacy Committee and there are a few changes we'd like. The system's working really well, but some of the things it does conflict with how we teach students about things, and what we demonstrate at the reference desk. We want to fix that up before we go into production.

Bill (curtly): Yeah, whatever ...

Adam: The advanced search is the major thing. The way it's set up, you can't use Boolean the way we teach it, which is the way it works in the other big databases. Usually you might start with two lines with three search boxes on each, with a drop-down listing all the fields you can search, and you can add a new row if you need one. Within each line you're doing an OR, and line by line you're doing an OR. That's a common pattern for advanced searches, and we'd like to reuse it. Would something like that be possible?

Bill (stares into space): ...

Adam (waiting a few beats): Bill?

Bill (types on a keyboard, ignoring Adam): ...

Adam: Umm, following up on my email from last week ...

Bill (checks phone):

Adam: Bill, will this be possible?

Bill (suddenly): No.

Adam: No?

Bill: No.

Adam: Why?

Bill: It's just technically impossible.

Adam: Why?

Bill: It's quite difficult to explain.

Adam: Why?

Bill (dissembling futilely): It's about Solr on the back end ... object-relational model ... Unicode ... it's the server. It won't run on the server. There are configuration issues.

Adam: What about them?

Bill (struggling): ... it's about how the parameters get passed in.

Adam: Can you explain a bit more? I don't know much about Solr, but I know a bit about database design. Give me the basics and if I don't understand I'll let you know. I let you in on some top-secret IL stuff... so payback time. If I can understand some of this stuff, it will make my life easier and in turn, I won't be asking you for the impossible. I just need a clear understanding from you of what's going on here. Is that OK?

Bill (somewhat cheered): Well, actually I guess it could be done, except I'd have to write a little middle layer that would transform the query. Which is actually something that's come up on the mailing list and some other libraries have been asking for it. So anyway, the problem is that when you pass any search string over to the parser---

Adam: OK, thanks ... but I need to get to a senate meeting. We can pick up on this again later.

[ACT 6: I upgraded to version 2!]

Slide 12

Adam (addressing the audience): So, it is a great pleasure to be invited to address the university senate today to talk about some of the goals of the library's information literacy program and to talk about some of the ways we're going to push this program forward for the benefit of all students here at the university. To start, I'd just like to briefly address some changes we've made to the library's search systems to contextualize some of the pedagogical goals I want to discuss. If we --

[phone rings]

Adam: I'm so sorry about this, but I'm expecting a very important call. Hello?

Bill: Hello!

Adam: Oh, it's you. Is this important?

Bill: That advanced search stuff we were talking about before? It's all working. Everything we talked about, it's all fixed up. You can add rows really easily, and your suggestion about the Boolean worked out nicely. It's all tested and in production now.

Adam: Oh, okay ... that's great ... really, I mean it. But I'm talking to university senate at the moment and we're about to look at all that stuff.

Bill: AND, I submitted the patches back to the main source tree and everyone liked them and they all got accepted. In fact they bumped it to version 2, so I downloaded that and installed it.

Adam: Version 2? We're running version 2 now? Publicly?

Bill: Yeah.

Adam: Ummmmm ... I'm about to show some of the new stuff... is this stuff I haven't even seen?

Bill (excited): Oh yeah... it's much better and completely different. It's faster. Major improvements there. We don't need to buy that third server now. CD and DVD covers display too.

Adam: Well, this is rather awkward ...

Bill: And that problem with the availability is fixed. Oh, and I loaded in 19 million articles.

Adam: You did what?

Bill: I loaded in full article-level metadata for 19 million articles. To be exact, 19,741,137. Now when you search you get back a mix of items from our local catalogue and from the set of articles that are owned by the Ontario university consortium. Cool, eh?

Adam: 19 million articles?

Bill: Yeah, isn't it awesome? Saves so much time! Everything's all in one place, people don't have to go into article databases to search, easy one-box interface for users, it's perfect, it's just what we've been after.

Adam: 19 million articles? Mixed in with our catalogue?

Bill (nodding): Yeah!

Adam (exhaling sadly): That sounds like a really impressive piece of work ...

Bill (suspecting all is not well): Mmm?

Adam (volume increasing as he goes): But from an IL point of view there's a lot going on here. There are three weeks left in the term and everyone is in an insane rush to write papers, and you've completely changed the behaviour of the catalogue and I'm standing here in front of senate looking like an idiot. Everything we've taught students, and everything they've learned by experience, doesn't work any more. Also those 19 million articles cover about 90% of what we have in the social sciences and humanities, but they're not at all right for science, engineering, medicine, or business, so it's completely misleading for them. They'll search and get bad results and never know they need to go into subject-specific databases. Oh my God...

six months of not smoking and now this! Who here has a cigarette? C'mon, not one prof on senate smokes?!!

Bill (his head has been sinking lower as Adam lists the mistakes, and now he says quietly, head in hands): I'm in my happy place. I'm in my happy place!

Slide 13: VIDEO: Cameron Hoffman <http://www.youtube.com/watch?v=yIPJwSkGNQI>

Slide 14

[ACT 7: Agile processes]

Bill: I don't think we've realized how our library has turned into a software development shop over the last while. Especially now that we're using VuFind.

Adam: It used to be that we'd install a new version of Sirsi's OPAC every two years, and had no choice about anything. It was a lot simpler. Like Henry Ford was doing OPACs: we can have any colour catalogue we want as long as it's black.

Bill: Now it isn't production line any more, it's all customized. Lots of different kinds of people working together, with different jobs and backgrounds and interests. Working on software together is hard. But you know, there's a lot we could perhaps do with an agile approach.

Adam: What's that?

Bill: Over the last few years a lot of good programmers have been thinking about ways that people can work together more closely, and more actively, so that building something is a much more productive experience, but also more pleasant, without any of the aggravations that we've been having. You've maybe heard us talking about extreme programming, or scrum?

Adam: What I don't really get is the way you freeze things at the last minute.

Bill: Freeze things? Before going into production?

Adam: Yeah. We always find lots of things that need fixing at the last minute but you say you can't do them.

Bill: It's annoying, yeah, but that's because we need time to prepare and test everything. When we're working on new things they gradually make their way from our laptops up to the development environment, then up to staging, where we do all the final testing with data copied in from the production servers. Basically at that point, before launch, we just want to kill bugs. We don't want to introduce anything new. That's where you do a lot of testing.

Adam: That's right, with that Bugzilla thing. I report bugs and you fix them.

Bill: Yeah. But also before we launch, we're figuring out how we can switch the production

system over from the old one to this new one, but without anything being affected or anyone noticing. No downtime, no data loss, no broken URLs, nothing like that. That takes care.

Adam: It's still annoying. In August when we're working on getting ready for all of the classes we teach in September and October, we're finding all kinds of little things here and there.

Bill: Little things we can still fix, but we just don't have the time for big things. And now that we're prioritizing everything, you know we've done all the major work. Besides, you're always telling me not to make any changes!

Adam: That's different. From the middle of August to the middle of October, we're preparing and teaching, and we need everything to stay the same. We do posters and brochures and handouts, and we want the screens to match what's on the paper. I don't use PowerPoint when I teach, but some librarians do, and they need them to match. And for all of us, when we're in a class, we need the system to work like we expect, because we have certain points we want to hit or examples we want to bring up.

Bill: But after we've put the new system into production is just when we want to get to working on new features! That's the fun stuff.

Adam: We want the new features too! And in fact, after we've had that time to teach the system, and we've been working with it at the desk, we've had a lot of experience with it and we have new ideas about how to improve the IL experience in the catalogue.

Bill: What, after two months? In mid-October?

Adam: It's crazy until after Thanksgiving. For two months, every reference and instruction librarian is busy. And then a month in January.

Bill: So that's three months of the year when we can't work together.

Adam: And then there's vacation.

Bill: And going to conferences.

Adam: And the library's understaffed anyway.

Bill: And we're tenure track at a university, so we have research and service commitments.

Adam: And with budget cuts I might have to spend half my time next year managing the student assistants.

Bill: And next time my department head is on six-month sabbatical I might fill in.

Adam: How does your agile stuff help with this?

Bill: I don't know. We're not a software development shop. We're a confused jumble of people who really only spend 40% of their time on this, at most, and that 40% hardly ever matches up in our calendars.

Adam: There is nothing to be done.

Bill: I'm beginning to come round to that opinion.

Adam: We can keep on working together as we can, keep the project going on the mailing list and in the meetings. It's on everyone's minds, it's the most visible thing we do online, it's just a matter of scheduling. If we make it a priority, it will be.

Bill: I can't go on like this.

Adam: That's what you think.

[Saddened, they pause]

[ACT 8: Adam gets excited about a stupid product]

Slide 15

Adam: Hey Bill! Guess what?

Bill: What?

Adam: I just got back from the conference of the Educational Academy of Pedagogical Teaching and Learning Schools, and it was amazingly good. Some great speakers, some really interesting research results were presented, it was all very inspiring. AND there were some vendors showing off some amazing new products. One of them was just astounding, we were all amazed by it, and we have to get it.

Bill (warily): Yeah?

Adam: It's a localized instance of Bing Scholar!

Bill: What?

Adam: It's Bing Scholar, but it's local.

Bill: When you put words together like that it makes me really worried.

Adam: What for? You know Google Scholar.

Bill: Of course.

Adam: And you know Bing.

Bill: I certainly know *of* it. I think I used it once the day it started.

Adam: Well, Bing has made their own version of Google Scholar. Microsoft bought Elsevier and Thomson Reuters and The Bay and they're loading all the articles into the system. They've got

deals with other publishers, and they're loading in all the open access articles, so it's as almost as complete as Google. Try it, it's scholar.bing.com.

Bill: But what about localization?

Adam: Yeah, it gets even better. We can license local access to Bing Scholar. They send us a computer with *megabytes* of disk space on it - so awesome - and it's automatically configured to update itself every night with fresh downloads from Microsoft. Then there's a Windows program we can give people that will give them special access to the data, and there's a web interface to it too. Hang on, I have the information sheet about it here. (PULLS PAPER FROM POCKET.) Let's see ... there's a license agreement we'll have to sign, but we can trust a reliable company like Microsoft so I don't think we need to bother reading it very closely. And there's something about their support people having access to our network. We'll have to open up a port on our firewall, but I'm sure you can configure that. Says here it's written in C sharp so any plugins you do would have to be in that too, and it runs on Windows Server. It's all stored in an Access database, so we'll need that, and we'll need to install a special Windows-only client on everyone's computers. And we'll all need to switch from iPod to Zune and Wii to Xbox. Will any of that be a problem?

Bill (gapes silently)

Adam: Oh, and the four other IL librarians and I will have full administrator access, of course. You won't need access at all, we can run it all ourselves. Besides, it's cheaper that way, with one less user license. We're excited about this! The five of us have formed into six different working groups to look at the possibilities. We're going to build on some of the, uh, lessons learned from that time you loaded 19 million articles into VuFind. We really think we can do some solid IL work with this.

Bill: I'm sorry, but no.

Adam: What?

Bill: No.

Adam: But this is important for our IL program, Bill. This would be a major new initiative.

Bill: You haven't signed anything yet, have you?

Adam: No, but the reps are coming on Friday.

Bill: Well, I'm glad we're talking about this beforehand. At least we're getting better about that. If you'd signed the contract I'd still say no.

Adam: But how can you say no? This is a worthy initiative.

Bill: You're saying you want us to buy, untested, a system that would require us to hire a Windows admin, invest tens of thousands of dollars in new hardware and unreliable software, add insecure proprietary operating systems into our mix, get involved with in a language we don't use, all where Microsoft would have back-door access to our private network?

Adam: What are you, anti-Microsoft?

Bill: Yes, but that doesn't matter. Look, we run Linux servers, we use NetApps for storage, we like Postgres, we write mostly in PHP and Ruby, we use use free software wherever we can. We have absolutely no Windows software in my shop! We have a set of technologies we work with, and it's Unix and free software. That's how my department is designed, that's what we hire for. That is how we've constructed the entire way we function, and I'm afraid there is just no place for this Bing local thing.

Adam: I'm sensing some negativity here. Is this a rerun of the whole advanced search business?

Bill: Yes and no. I'll grant that your front-line experience gives you serious cred on searching and you're definitely in charge there. You design the theory of how IL part of the catalogue should work, and together we'll put it into practice. The implementation is up to me. There's a line. But--we can probably get this working another way, in an open way. In fact, let me tell you about the Knowledge for All project. That might work. But for Bing local, I'm putting my foot down. The foot has spoken, Adam. The foot has spoken.

[ACT 9: Finale, several months later]

Slide 16

Adam (checking watch): So, two minutes left to wait. You excited about coteaching this session with me?

Bill: I'm kind of nervous, but I should be OK. I've never got up in front of a class before. In fact I've never really done any public speaking before.

Adam: You'll be fine. You get used to it. I do about eight sessions a week for six weeks, so I got over any nerves pretty quickly my first year. I'll start things off and then you can talk about the authority records and how that's improved things. Then do the advanced search, then hand back to me for the Caroline view and disciplinary discourse.

Slide 17

Bill: Don't worry, I've been rehearsing all week. I'm all set. You know, I have to say, I was impressed with what all you IL people did on the new release of VuFind.

Adam: When you sent us the roadmap for the next official release it really got us thinking. There was nothing information literacy-related on it at all, it was all technical.

Bill: It's library geeks around the world hacking code, often in their spare time. That's what you get. But you really got involved, and you got IL people from all in on it too. It had a huge effect on how things turned out.

Adam: It made the whole process a lot more enjoyable, too. So, you ready?

Bill: History and Philosophy of Science 5500, Kuhn's Paradigms and Foucault's Epistemes. Sounds pretty interesting. What is it, six or eight students?

Adam: Huh? This is section fourteen of Psych 2030, Introduction to Research Methods. Didn't you get my email?

Bill: What!?

Adam: There are 500 students in there. Be sure your lapel mic is attached properly.

Bill: Oh my God.

Adam: Don't worry, if you're lucky, 300 of them will be ignoring you. If you're not, 400 will. Well? Shall we go?

Bill: Yes, let's go.

[They do not move.]

Slide 18