Good morning, folks, and thanks for being here. My name is Dorothea Salo, and I teach technology as well as scholarly communication at the iSchool at the University of Wisconsin Madison, after spending several years in academic libraries working on open access and research-data stewardship. When I was asked here to NASIG, the organizers told me I was slotted into a “Vision Session,” that I needed to offer a vision relevant to this excellent and distinguished conference, something that’s not happening on the ground right now that I think should be. I pitched the organizers several ideas — which may not surprise you if you know me; I am never short of opinions — and the one that caught fire with them was the question of reader privacy with respect to electronic serials and ebooks, e-resources generally.
And that immediately brought to mind for me the unforgettable Billie Holiday singing the nineteen-twenties Grainger and Robbins blues classic “Ain’t Nobody's Business if I Do.” Which, the version of the song that Holiday sings starts out, “There ain’t nothin’ I can do nor nothin’ I can say, that folks don’t criticize me. But I’m gonna do just as I want to anyway, I don’t care if they all despise me.”

Love that. LOVE. IT. Because in my head it completely captures what’s going on with collection and exploitation of reader behavioral data. There’s a whole lotta libraries and a whole lotta content providers in the Big Data or even small-data game doing whatever they want no matter what readers think. Might as well, right? Because whether you do or you don’t, somebody'll hate you.

If you DON'T collect and exploit user data, your accountants and Big Data nerds will hate you, because you’re missing revenue opportunities — or so they think; I’m not always convinced the financial upside is what they think it is. Your usability wonks might hate you too, because they can learn useful things from snooping on how readers dink around with e-resources. And, I mean, I’m laying it on the line here: that’s snooping, y’all, I don’t care how holy the reason is. But the uproar if you tell them not to do it, well, I’m seeing some of it, and wow. Some usability wonks HATE privacy wonks like me right now.
Now, if you DO collect and exploit user data in the way it's usually done today, I tell you what, I HATE YOU RIGHT BACK. Well, okay, “hate” might be a little strong. But I am definitely NOPEing you, because if I as a reader of serials know you're doing that, I trust you less, and I trust your systems less.

And you know who else trusts you less? The American Library Association.
See, ALA has this Code of Ethics thing that first got written in nineteen-thirty-nine and has been revised a few times since then, and Article Three says that libraries will “protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted.”
Well, okay, so what? So ALA thinks reader privacy is important, so what? Hey, Facebook's Mark Zuckerberg says we should all get over the privacy thing already! So who gives a flying flip about a nearly century-old ethics code from some hysterical century-old librarians?

Did she just go there? YES, I WENT THERE, I said the h-word! I did so fully cognizant of the resonance that word has for librarians, as Section 215 of the Patriot Act teeters on the sunset bubble, and I did so because it's a pretty salient — and recent! — example of librarians taking the high road despite being called every name in the book. Which is not entirely dissimilar to what's happening to some library privacy advocates now, really!

Now, I could pound the podium at this point about abstractions like intellectual freedom, civil society, surveillance society, panopticons, blah blah blah. I could pound the podium, but there are ethicists and philosophers and legal scholars and others who are a lot better at all that abstraction than I am. I'd rather keep it concrete. Plus, the hotel probably doesn't want me to damage the podium, right?
So, concretely, let's talk for a second about what's being called the Internet of Things. I mean, things, how much more concrete can you get, right? The basic idea here, for anyone not familiar with it, is that gizmos we own that have ticked right along for ages without Internet connections can now be connected to the Internet. That lets us operate them remotely, like a thermostat in my house that I can set from my office if I happen to be coming home early that day. They can also get information from the internet, like a TV automatically knowing what's on. Networked gizmos can give us insight into how they work and how to use them more efficiently, again like the thermostat. They can also give us insights into our own behavior, as with all the fitness trackers out there, and make suggestions at least nominally aimed at helping us.
And one of the things that's happening now with the Internet of Things is that the Federal Trade Commission is scrutinizing it pretty closely — this image here is from the opening page of their January report on it, it's a good report, not too long, and I recommend reading it — because as it turns out, it is SUPER easy to cause people real tangible harm based on data coming from mundane items like a television or a fitness tracker or a thermostat.

A thermostat? Really?
Kind of the paradigm case here: my spouse is home at the moment, but if that weren't so, imagine what the thermostat data would be telling a home burglar. Oh, hey, house is empty, come and get it! More subtly, though — and this became a prominent public concern when Google bought Internet of Things thermostat maker Nest — what can this thermostat tell advertisers or even law enforcement about me that actually ain't nobody's business? It ain't nobody's business when my house has people in it and when it doesn't. For one thing, that starts to indicate whether there's a stay-at-home parent, or somebody unemployed, or somebody with a disability, or somebody who works at home. And that ain't nobody's business! And it starts to be information that can be used against me, especially when correlated with all the other data coming from every corner of our lives. It could totally be used unfairly in credit and loan deliberations, rental decisions, and so on. And we're already starting to see horror stories like that, data used for redlining as well as extremely dubiously-ethical marketing.
There have already been some publicized cases of stunning Internet of Things creepery, too, like the Barbie doll that recorded whatever a kid said to it and streamed that to Mattel, where it piles up into a dossier on the kid. Yeah, totally no potential for abuse THERE.

But yeah, fine, thermostats and Barbie don’t reflect what anybody’s READING, so it’s NOT an intellectual-freedom issue, so why is Internet of Things-style creepery an issue for NASIG?
Well, thanks to our good friends at Adobe, we actually KNOW e-resource use is being snooped on and collected into dossiers, just like what kids say to Barbie. No possibility of doubt. I don’t think anybody here has been living under a rock, so we all know about this, but just a recap: Adobe’s collecting reader-behavior information from Adobe Digital Editions, including when it’s used on library-provided e-resources. Adobe got caught because they were transmitting the information in the clear, and they’ve stopped doing that — but they have NOT stopped collecting the information, as far as anybody knows.

So I’m sorry, content providers, I really am, but there’s NO benefit of the doubt possible here. Readers can’t trust you. Librarians can’t trust you. Adobe shoved its foot in it right up to the thigh for all y’all. We have to believe you’re all behaving like Adobe until and unless you state — and ideally prove — otherwise.
So what I'm ultimately saying here is, it actually makes a lot of sense to think of electronic resources as part of the Internet of Things. What's an ebook? What's an e-journal article? It's a mundane item you use for your own purposes that back in the day wasn't even Internet-connected but now is. It communicates with the Internet and leaves data about you and your behavior behind, data that can be used to dish the dirt on you, to cause you real tangible harm, individually or because your behavior happens to cluster with the behavior of others in a way that somebody with power doesn't approve of. The same privacy issues the F-T-C cares about with Internet of Things gadgetry are entirely salient to electronic resources.

*pause* Hey, F-T-C, c'mon over here and let's talk, okay? I mean, we're in Washington DC, right? If privacy's gonna be a thing for dolls and thermostats, I'd love it to be a thing for ebooks and electronic journals too. If that takes F-T-C intervention, I'm cool with that.
“protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted”

—American Library Association
Code of Ethics
Article III

Because I could go into the horror scenarios here, the real ones and the what-ifs, but I don’t see a need, they’re not all that different from what libraries have guarded against in the print era, and anyway we’ve seen examples already. Aaron Swartz, Georgia State, e-textbook and e-testing platforms collecting data about minors, you all know the score. It’s because libraries were aware of all these risks — had experience with a lot of them! — all the way back in nineteen thirty-nine that ALA took such a strong stance in favor of privacy and confidentiality. Amazingly prescient stuff. I respect ALA a lot for this.
So that's libraries' ethical stance on reader privacy... what about publishers? I mean, since y'all are here and all.
That's libraries.

What about aggregators?
That's libraries.

What about A&I services?
That's libraries.

What about content providers?

Content providers generally? What's y'all's ethical stance on reader privacy?
So I went looking for trade-level reader-privacy ethics statements, because hey, I started in publishing, I totally didn’t want to leave out all the people at NASIG who aren’t librarians, and actually I discovered something pretty interesting. For “disturbing” values of “interesting.”
I started in the obvious place, the Committee on Publication Ethics, because this is an ethical question, right?
SSP, the next likely source, plus they're right here with us — hi y'all! — so I couldn't possibly ignore them.
*crickets*
Huh. Okay, how about the STM Association?
*crickets*
And of course the question of reader privacy is just as salient for open-access journals as for anything else, if not more so, so I checked out OASPA too. I actually think it's pretty cool that to the best of my knowledge and belief, open-access journals don't seem to have turned to exploiting or selling reader data as a major revenue stream. Maybe that's because of OASPA, I thought to myself!
*crickets*

Well, so, okay, maybe OASPA left it to D-O-A-J, since they're coming up with quality criteria for inclusion and certification of open-access journals these days.
Yeah, so, as you know I'm a long-time open-access advocate, and I gotta say to my fellow open-access folks, I'm seriously not cool with this, y'all. Can open access please take the high road here?

(look uncomfortable, shuffle feet, let the moment go on too long…)
So, maybe looking at trade associations was the wrong route, I said to myself, maybe the privacy statements are happening at the individual-journal level. Turns out that's been looked at, in a twenty-twelve study published in College and Research Libraries.
Pretty much crickets, folks. Pretty much crickets. Yeah, there were privacy policies, but they were basically terrible.
And unfortunately, we know there are ugly issues here on both sides of the business-model fence. Eric Hellman checked twenty major research journal websites for evidence of ad-network trackers, who in case you haven't checked lately have been spreading malware as well as being generally creepy. He found trackers. Lots of trackers. In both toll-access and open-access journal websites. And yeah, this was a really tiny sample, but be my guest, expand it, do you really think the results will be MORE in favor of reader privacy? Because I don't.

This is a technology-infrastructure point, and tech infrastructure isn't really what I want to talk about today, but just this one thing, librarians: the instant we put some third-party resource on our website or in our LibGuides or in our catalog or refer patrons to it some other way, WE BECOME RESPONSIBLE for its privacy implications. If it takes some kind of systematic ethics review of content-provider websites to call out this kind of thing, hopefully make some noise toward stopping it, well, I'm in favor.
And to be fair, I’m not saying there’s a conspiracy theory here, okay? No tinfoil hats, I totally don’t believe that. It’s historical accident, this absence of ethical-responsibility statements from content providers. It happened because in the days of print, reader privacy wasn’t the content-provider’s problem; aside from the venial sin of selling subscriber lists, content providers pretty much couldn’t compromise reader privacy even if they wanted to! There was basically no way to monitor the use of a print journal or a print index or a print anything. Either it got mailed to an individual subscriber and the subscriber did whatever they wanted with it —
doodle wildly all over the pages
make an art installation for their favorite holiday
light it on fire, whatever —
— or the publication got mailed to a library, and maybe the library keeps track of how often that chunk of print leaves the shelf, but the library CERTAINLY doesn't know who picked it up, much less in what context, so it can't tell the content provider anything about that. Not that it would anyway. So content providers didn’t have to think about reader privacy.

But times have changed, folks. Times. Have. Changed. The library isn't always in the middle of the publisher-reader transaction any more, and even when we ARE, today's content provider has a lot more ways to compromise reader privacy available, so yes, content providers NEED to come up with an ethical position on reader privacy, okay?
And there's an effort underway to maybe get a handle on this, NISO is working on this thing they call a Consensus Framework to Support Patron Privacy in Digital Library and Information Systems. And I am BEGGING the NASIG community — I am begging each and every one of you here — to watch this, and to comment on it, and to make all participants in it VERY VERY CLEAR that you're watching. You are the right people, the people NISO needs to hear from!

Because I generally dig the word “consensus,” but I confess I'm a little worried that in the NISO context it’ll mean what it seems to mean in Trade Pacific Partnership negotiations, which is something like “the rich content owners are gonna set the rules in a secret smoke-filled room and the rest of the world can just lump it.” That's not consensus, that's railroading, and it NEEDS not to happen. So let's not let it, okay?
“protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted”

—American Library Association
Code of Ethics
Article III

Until that or something like it happens, though, I’ll have to rely on the ALA Code of Ethics here, which actually doesn’t bother me a bit. One thing I want you to notice about this statement is that it has ZERO qualifiers. None. Do you see an asterisk or a dagger or a footnote here? I do not.
“protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted”*

* except when that’s inconvenient

It doesn’t say “libraries protect privacy — except when that's inconvenient.”
Because sure, it's super-convenient to do usability testing or market research silently, it's super-convenient for librarians who need tenure to trawl those data, I get it, I do! And I'm not unilaterally against those things, I'm just unilaterally against doing them in the thoughtless, careless way they're often being done now. And, librarians, you get NO SMUG POINTS here, okay? I am seeing articles in the library literature right now, today, that HORRIFY me, they're so careless about reader data.
Another non-footnote goes “libraries protect reader privacy except when we’re improving our services,” which, what even IS that, that is one of the most amazing weasel phrases I have ever heard. You can hide ANYTHING behind that, no matter how creepy.
I mean, imagine that in the physical library. “We're going to follow you around the library and record what you're reading with cameras and video, and we'll keep that data indefinitely, but don't worry, we totally won't ask you your name, and we're only following you around in order to Improve Our Services!” In what world would that not be creepy? How is it any LESS creepy watching my e-resource reading trail? Just because it's immensely harder for me to figure out you're doing it, much less stop you? That's not less creepy, it's MORE creepy! We are talking sparkly vampire zombie werewolf Evil Overlord's One Ring levels of creepy here, people!

I actually think “would we do this in the physical library, the physical bookstore, the NASIG exhibit floor?” is a fairly decent heuristic for assessing something's creep factor. It's not perfect, absolutely not, but it's useful, because our sense of what we will and won't do in physical spaces is pretty strong, pretty sophisticated, pretty well thought through. It also keeps our patron base from being divided into physical-library users and digital-library users, and one group having better privacy protections than the other, which, that just ain’t right. Throwing that out there for people to take home.
Here's another one, there's no asterisk in Article Three saying that libraries protect privacy except when sharing data with partners — whoever THEY are. And librarians, “partners” doesn’t just mean “content providers,” so some of us need to be way more nervous about what we’ve got on our websites than we are. We saw that with Hellman’s quick look at research journals. Google Analytics, anybody? Facebook Like button? Yeah. Yeah. We need to be nervous about that.
Our good buddy Google? Totally shoved its foot in it up to the thigh on privacy. If you're not in K-twelve circles you might have missed this, so the story is that schools using Google Apps for Education suddenly found out that Google was assembling data and profiling students based on their email to use for advertising, despite many public protestations that Apps for Education respected privacy!

What can I even say about that? Except that I don’t trust Google with behavior data as far as I could THROW Google. I don’t think any of us should. Yeah, I know Google Analytics’ terms of service says it respects privacy. I just don’t believe what Google says about that. Why should I? Why should you? Why should anybody?
“protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted”*

* except learning analytics

Speaking of education, Article Three has no exception for learning analytics either, whatever THOSE even are. Librarians generally don’t rat out our students to their professors, even when students are being stunningly unwise. They're learning, right? We know we have to leave them a private space for the various kinds of unwisdom that happen during the learning process. Digital doesn’t change that! It’s not any more okay to rat students out now just because we have lots more detailed ways to do it.
And I’ll tell a story on myself for this one. Our course-management system at U-W Madison, like many, tracks what students do on their course websites and how long they spend doing it. So for one online course I taught, I noticed that students weren’t spending hardly any time on the main lesson pages where the video content was, and they weren’t clicking on links to readings. And I got pretty upset about that, and I made a huge angry fuss — only to find out that students were downloading video rather than streaming it because the streaming didn’t work real great, and they were clicking on links from the PDF syllabus instead of the course pages.

They were doing the work. They were! They just weren’t doing the work in the way that the course-management system was able to capture. And my poor students were sincerely hurt and scared, and they had every right to be, and I’m sorry about it to this day. And since then, I’ve been super-skeptical of whether learning analytics tell us much that’s useful, and super-aware that they break trust bonds between student and instructor that I for one absolutely need to do effective work in the classroom. So I’ve learned my lesson — I don’t want to surveil my students. I don’t want anybody ELSE surveilling them either — and that absolutely includes the library and e-resource content providers.

But at this point it doesn’t even look like I can say no! How do I say no, people? How? How do I tell all y’all to leave my students alone? Speaking of Georgia State, I want publishers OUT OUT OUT of course-management systems, ‘cos y’all creepy.
And finally, there isn’t an exception to library protection of privacy based on whether the patron knows or cares about what’s going on. Libraries pretty much assume that patrons usually don’t know or care. Safe assumption, right? But not enough to let us do whatever we want, not ever. We also know, for example, that some privacy violators go to great lengths to KEEP people from knowing their privacy is being systematically trashed, cough-cough-Patriot-Act, cough-cough-Snowden.

We also know that some of our patrons absolutely vitally need their privacy respected to be safe and to feel safe. Even if some of our patrons don’t particularly have to care about privacy, others absolutely DO based on their research interests, their life circumstances, whatever — and y’all, I gotta say here, I do NOT think it coincidence that practically all the librarians and other pundits I’ve seen saying libraries go too far in protecting privacy have been white men. Check yourself before you wreck yourself like Google Buzz and Google Plus, folks.

Look. if libraries don’t respect privacy, patrons who desperately need privacy won’t trust libraries, and sometimes these are the very patrons who need libraries the most. So libraries default to privacy, and I believe with all my heart and soul that’s the correct default.

Notice that Article Three doesn’t say “when the patron KNOWINGLY CONSENTED,” which, yes, I agree that’s a whole different ball game. What I’m seeing — again, pretty much from white men — is some kind of sense that the library can do whatever it wants with patron web-behavior and reading-behavior data because supposedly patrons don’t care. And I don’t know where that sense comes from, but it sure ain’t the ALA Code of Ethics.
Oh, and nobody cares, you say? WELL, I CARE. Y’all don’t get to just say “patrons don’t care, readers don’t care.” I AM a library patron, as well as a reader of e-journals and other electronic resources, and I care a WHOLE LOT about my personal privacy. If your whiz-bang technology rig, whatever it is, doesn’t account for me, and for other people with the smarts and the grit to believe in privacy and to WANT privacy, there is something pretty seriously wrong with your technology rig, and you might want to check your ethics too.
And this is where I get back to the song made famous by Bessie Smith and Lady Day, because if you listen to it — and I honestly didn’t remember this until I’d already chosen my talk title — if you listen to the song Ain't Nobody's Business If I Do, you find out that it’s about the singer allowing other people to walk all over her, to hurt her and exploit her — and I’m being a little vague here, because the song is really painful and hard-hitting, so I’m warning people, only look at the lyrics if you’re okay with that. And the song insists that it's her right to let those awful things happen and nobody should interfere with that. And the way Lady Day sings it, it's really clear to me that for her the song comes from a place of deep despair and helplessness. Don’t interfere with my self-destruction, she sings, because if I can’t even do myself any good here, what good do you think you can do me?
And I’m guessing that sounds familiar to some folks here, who feel helpless faced with ubiquitous incessant onslauts on reader privacy. It's super-easy for any information seeker to throw their privacy down the drain, it’s super-easy for any library to enable that, it's super-easy for any web service that libraries use or that uses library interfaces to enable that, it's super-easy for any content provider to enable that, and yes, patrons do sometimes tell us “screw privacy, I want what I want!”

And what I’m saying here is, just because it’s easy and convenient to screw privacy doesn’t make it right, and it doesn’t mean we have to lie down and take it. Especially at Internet of Things, Big Data scale.
“protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted”

NO EXCEPTIONS.
Because seriously, it ain't nobody's business — it ain't the webmistress's business, it ain't my wonderful departmental librarians' business, it ain't no publisher's or aggregator's or A-and-I provider's business, it ain't the NSA's business, it ain't YOUR business — it ain't NOBODY'S business if I do read serials!
So what can we do, besides providing input to the NISO process I mentioned earlier, which I totally hope y’all will all do, to bring this vision closer to reality?
Wow. I wish I knew. I wish to the bottom of my heart I had a pat answer for you today. All this is enormously complicated, right? I've just scratched the surface here today. But I think I know where to start. I do. It's practically my classroom go-to for all kinds of situations, from privacy considerations in donor agreements to copyright and digitization to digital preservation planning.
First we understand the risks, as best we can.
Then we mitigate the risks, again as best we can, certainly acknowledging that there are some things, like organizations obsessed enough to dive to the bottom of the ocean in order to copy traffic off fiber optic cable, that we just don’t control.

I mean, how is this the world I live in, I don’t even know.
The first question I think we need to get a handle on is exactly what kind of information about our patrons causes risks to them.
And for convenience, I’m dividing that into three buckets of information: P-I-I, what I call “long-tail information,” and behavior-trail information.
And, I mean, I think we're all clear on personally-identifying information being really scary, so I don’t need to elaborate much on that. All I want to say is that it is NOT the only category of data we need to be concerned about, and that sometimes privacy policies use P-I-I as a smokescreen for abuse of other classes of data — they proclaim very loudly that P-I-I is either not collected at all or very carefully protected, and don’t say ANYTHING about anything else. Not okay, people. Any ethical framework we build around data needs to consider more than just P-I-I.
The next class of information I want us to be concerned about is what I’m calling “long tail information,” by which I mean data collected about patrons that’s a serious outlier in privacy-problematic ways. There’s data about people, for example, that isn’t strictly speaking P-I-I but is still uncommon enough to identify specific individuals — this happened in the A-O-L search-log release fiasco, it happened with the Netflix prize fiasco, it happened with the Hauser Facebook case, it’s how browser fingerprinting works, it’s really pretty common! What’s more, classic anonymization techniques don’t fix this.

And the thing is, we’re all outliers in some way or other. So even if somebody doesn’t stick out in one dataset, combine a bunch of datasets that contain information about them — and this is exactly what data brokers and web trackers and ad networks DO — and more and more people become individually identifiable, P-I-I or no P-I-I.

What people read? Totally long-tail information that can be tracked back to us. For some of us even more than others. I mean, if y’all could get your hands on journal-reading data from U-W Madison, you could correlate my reading with my public Pinboard bookmarks in like two point five SECONDS and know it was me, because believe you me, I regularly read stuff nobody else on my campus does. And anything in my journal reading that’s unexpected, an outlier? You’d know I read it and you’d be able to start guessing why. And just as another paradigm example, four years ago? When my mother was dying of cancer of unknown origin? My outlier journal reading would have been EXTREMELY sensitive information, you get me?
That leads me to behavior trails. Just one reading transaction probably isn’t super-re-identifiable, unless what’s being read is a serious, serious outlier. An individual visit to an individual web page, likewise. Where it starts being problematic is where you track a whole bunch of reads and a whole bunch of visits from the same person, even when you supposedly de-identify them. Or even just when you keep highly specific timestamps along with the interactions; that’s sometimes enough to let somebody reconstruct a behavior trail. And the more behavior trail data you have and the longer you keep it, the worse the privacy problem gets, because the easier it is to correlate interactions, and the more likely it is that you capture outlier reads that patrons would rather you didn’t associate with them.
And one great way libraries totally bypass this problem is by tracking uses of stuff without tracking people, and without trying to chain together or even correlate uses. Going back to the physical library, again, you see a bunch of stuff on the cart, you scan the barcodes and it goes into a database, and who the heck knows who used it? And correlating use is dubious at best because you don’t have any idea how many people put stuff on that cart, so nobody tries it. As data-collection practices go, this is pretty respectful of reader privacy.

Libraries also watch out for proxy server logs, because those are FULL of behavior-trail information. We do have to collect them, unfortunately, to deal with the would-be data miners appropriately, but we don’t usually keep them very long because we understand there's a privacy issue there. More of that, please, more intentional discarding of data. Data is a hot potato! Drop it whenever you can! This is records-manager wisdom, y’all, LISTEN to your records managers, okay?
The next question we need to get a grip on is “who wants to know?” And a lot of times people answer this question by occupation, you know? You got your spooks, your marketers, your academic researchers, your usability wonks, your black-hat hackers, and so on.
I’m actually going to divvy it up a different way, by how and why people approach data about other people, and the techniques they’re likely to use to get hold of it and analyze it. Because I think that gets at the risks better, and is more helpful at suggesting ways to mitigate those risks.

So this is only a first approximation, don’t hold me to it, but I think there are data omnivores, data opportunists, and data paparazzi.
The N-S-A is an omnivore. Google, Facebook, Amazon, commercial data brokers, they're omnivores. Black-hat hackers are omnivores, typically. If there's data, they want it, and they want to match your data TO you. The only way to prevent that is to keep data out of their greedy paws, even when they're actively lying to you and trying to subvert any effort you make to kick them out of your systems. How to do that is a bit beyond the scope of this talk, but for what it's worth most of the fixes I know of are partial at best, and they're technical in nature.
Opportunists are trying to do cool and useful things with data. They're academic researchers, and data collectors trying to be nice to academic researchers. They're web and social-media developers, usability wonks. They're hackathonners. They're open-data advocates and assessment experts. They're what Ann Arbor District Library used to call "superpatrons." They're people with their hearts in the right place — but that doesn't mean they've thought things through. Data opportunists have made some pretty big privacy messes!

This is actually also where I'd place patrons who want to reuse their own data, or who want access to a family member's data for reasonable reasons, things like that. There's nothing wrong with what they want to do necessarily, they just don't understand the broader implications or are lucky enough not to have to care.

The thing about data opportunists is, they generally don't want to hurt anybody, and they absolutely don't want the backlash that happens if they mess up on privacy. We can, I believe, help teach them how not to, not to mention WHY not to, and we should — they're often good privacy allies once they understand the issues.
So, data paparazzi. Data paparazzi have a TARGET, a specific person they want to track, and they are gonna pursue that specific target through whatever data they can find. They are people on political crusades, speaking of Washington DC. They are doxxers. They are kidnappers, perpetrators of violence, other people who hate and who harm. And paparazzi are terrifying, because they are obsessed, they are amoral, and they stop at NOTHING — they will social-engineer you, they will hack your systems and try to use them against their target, they will take over a target’s account to impersonate them or ruin them, they will correlate whatever they find out from you about their target with anything else they can find anywhere. Don’t be thinking “well, they don’t want the data WE have” — yes, yes they do!

Some people, even some security researchers, will tell you “don’t worry, be happy” about behavior trails and other non-P-I-I data, because reidentification attacks don’t have a real high success probability. Ed Felten and his research crew argue — and I agree with them — that this notion is based on the assumption that the attackers are data omnivores or opportunists, not data paparazzi. Attacks carried out by paparazzi, because they’re so tightly targeted and so relentless, have a much higher chance of being successful and of causing somebody harm.

So I’m telling you, yes, worry about privacy. Worry about the patrons you have, the readers you have, who have paparazzi on their trail. I am absolutely sure you have at least one such patron or reader, and probably you have lots more.
What ought we (not) do?

The last question is the big one, right? What should we do? What should we not do?
First thing is, NO OSTRICHING, okay? Heads OUT of the sand.
And to that end, I want you to know about the Library Freedom Project if you don't already. This is Knight Foundation funded, it's run by the amazingly badass Alison Macrina (ma-KREE-na), and it is all about libraries protecting reader privacy in all the ways we can find to do that. It's at libraryfreedomproject dot org.
And on that same theme, profession-level and industry-level advocacy and policy work are super-important here. Without them, librarians don’t know what to do, what to negotiate for, or even what to hope for, and content providers get stuck in a nasty prisoner’s-dilemma cat fight because anybody who takes the high road on privacy has to be afraid they’ll be outcompeted by somebody else taking the road to hell. Trade associations, this is your job; please do it. No more crickets, okay?
Next thing is, DON'T GIVE UP. Forget the song. We are NOT helpless, we CAN take concrete action to protect reader privacy, and it's ABSOLUTELY worth doing.
Next, librarians, let us use our benjamins wisely. License-negotiation time is the time we can ask the hard questions about privacy, and nail content providers down to real concrete answers. I know, I know, world plus dog is trying to use e-resource licensing as a policy tool, but I’m asking you to consider doing it one more time, okay? Because as I said earlier, once we add something to our website or our catalog, once we’re pointing patrons at it, WE ARE RESPONSIBLE if it compromises their privacy.

Content providers, give us privacy policies we can feel good about, please. I can’t put it any more simply than that.
And I know assessment is a thing and it's not going away, but can we please, please assess mindfully, conscious of potential data leakage and data abuse scenarios? Right now in too many libraries and at too many content providers, assessment is so compelling that it's utterly obliterating privacy considerations. That's NOT OKAY. It's actually really scary, I mean, I-R-Bs exist because scientists decided their work was too important to bother about whether they were harming people or lying to them about what was happening to them. Are we gonna revisit those days now? Please let's not.

When we see this, we need to call it out, refuse to participate, refuse to publish or otherwise countenance this kind of work, insist on confronting the privacy issues openly and conscientiously.
One final thought to take home with you: not even the greediest data omnivore, the most clueless data opportunist, or the most evil of data paparazzi can misuse data that isn’t there. Right now, collectively, our reader-data default is “collect it! unless there’s a reason not to!” That’s backwards. The correct default is “don’t collect reader data unless there’s a CLEAR reason we should. Just don’t.” The shoe needs to be on the foot of reasonable and transparent — not just transparent to us, but transparent to our readers — justification for any data we collect and use.

Because that’s another useful decision heuristic, right? If you dread explaining your data collection and use to your readers, if you start weasel-wording all over it because you fear backlash, maybe-just-maybe whatever you’re doing doesn’t pass the sniff test.
Article III. ALA Code of Ethics. That's my vision of where we need to be.

Please help me make this vision real.
Because one more time, say it with me — it ain't NOBODY'S business if I do read serials!
Thank you!

This presentation is available under a Creative Commons Attribution 4.0 International license. Please respect licenses on included images.

Dorothea Salo
the iSchool at UW-Madison
salo@wisc.edu
http://dsalo.info/