# Avoiding the Heron's Way

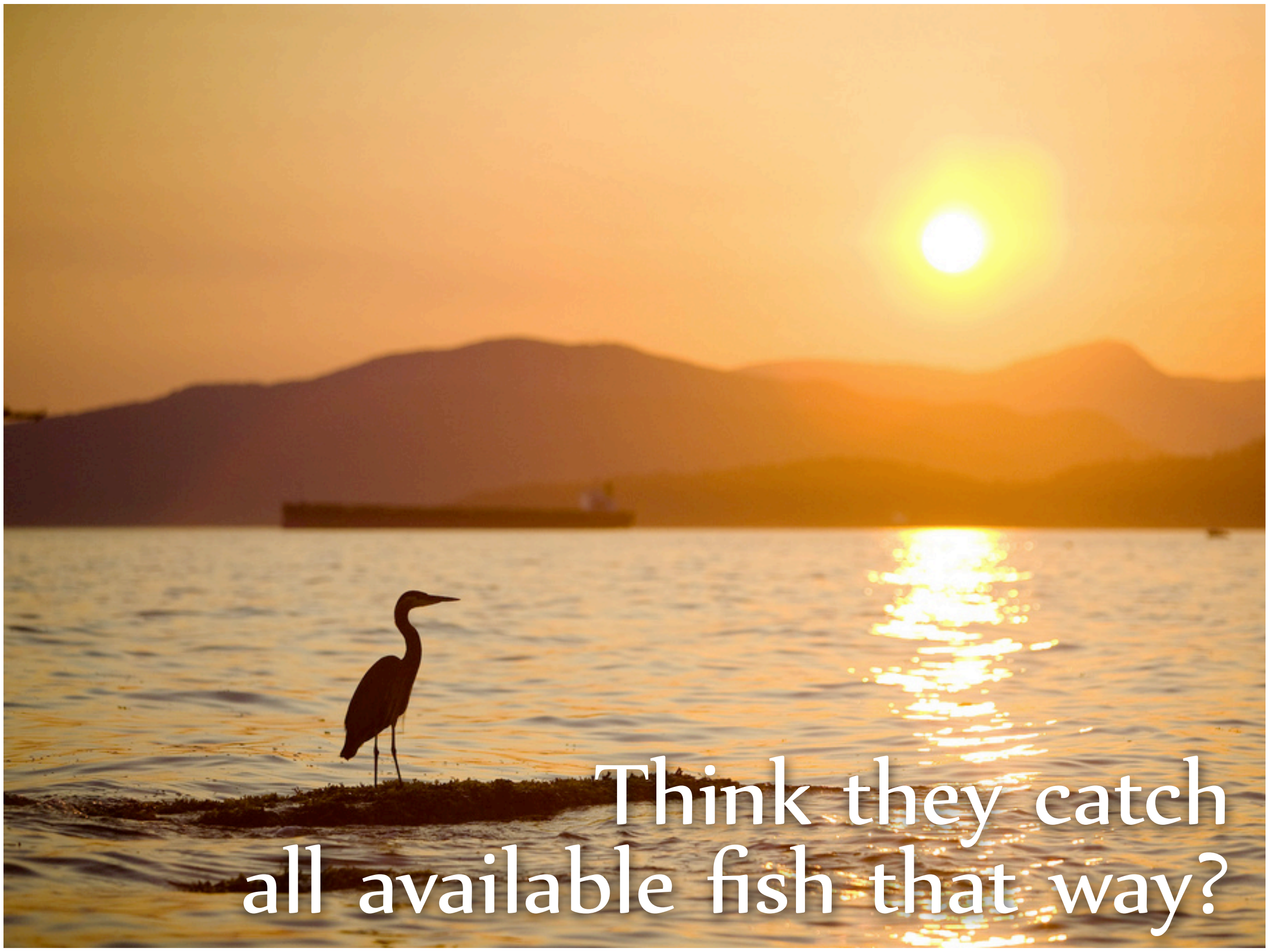
## Planning for a Practical Institutional Repository

Dorothea Salo

WiLSWorld 2011


I love herons. Great blue herons, green herons, gray herons, I love watching them. This presentation is not a slur on herons! But have you ever noticed how they hunt?

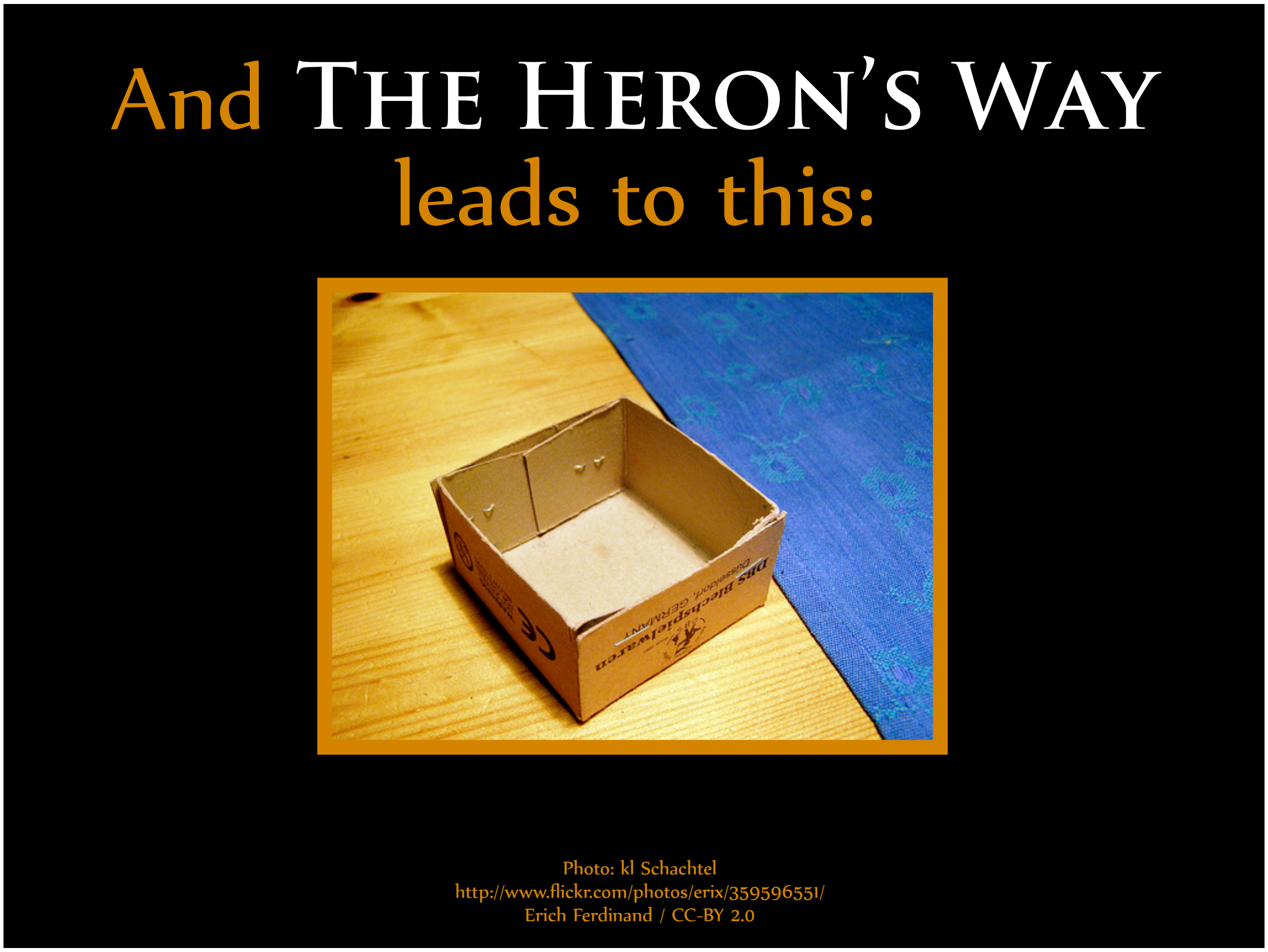

Right. They just kind of stand there. Maybe they stalk about a bit, veeeeery sloooooowly. Mostly they just stand there and wait for fish to come to them.



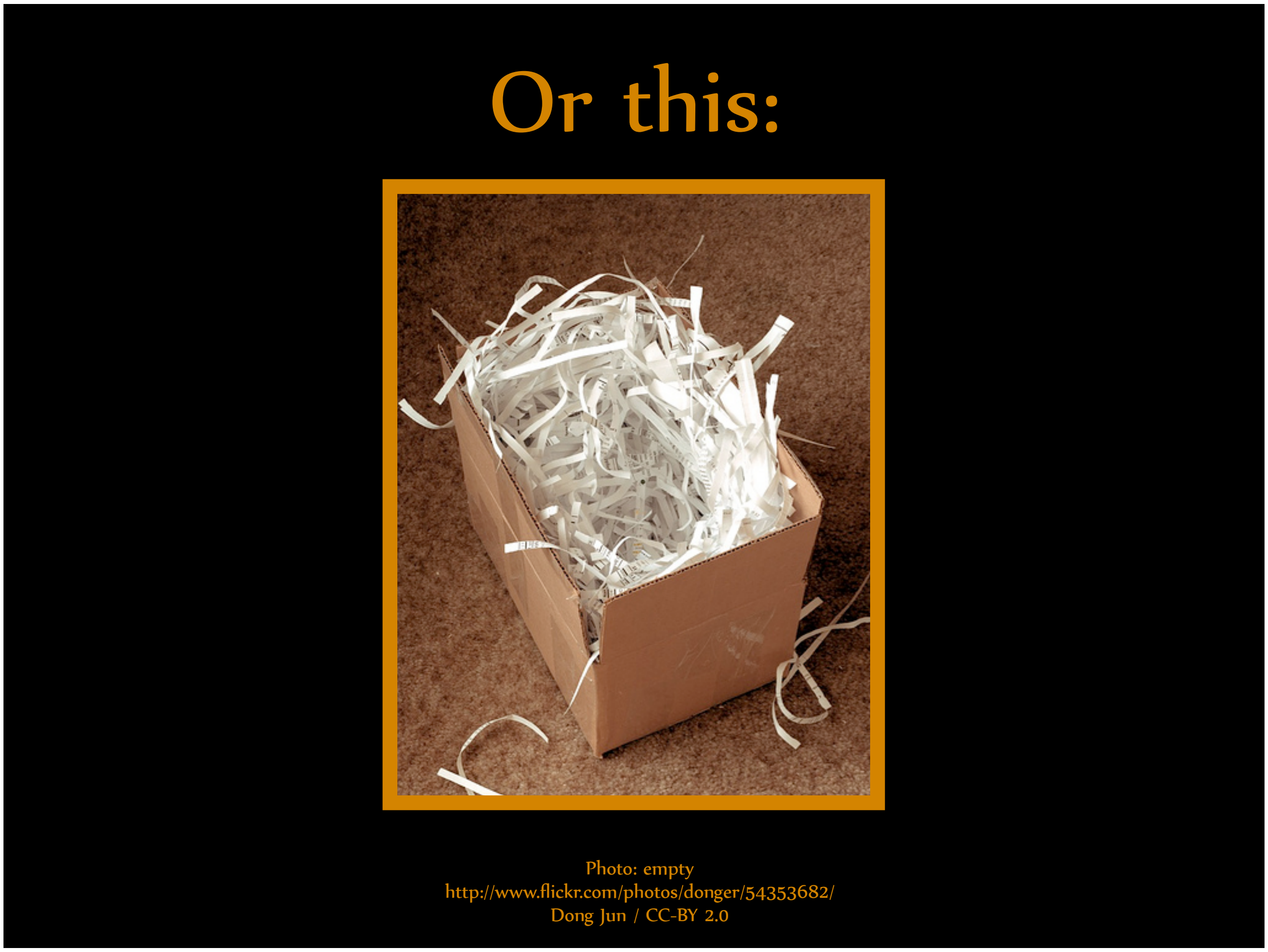

They've got an entire body of water to hunt in! Lots of fish in it! Are they really going to catch them all, or even just the BEST ones, by standing there and waiting?


Yeah, I don’t think so either.


Yet that's the way a lot of institutional repositories are planned. We're going to get some software running, and then we're going to... stand there and wait for stuff to show up in it. Maybe do a little marketing and outreach, maybe not. But mostly, just stand there. This passive style of IR management is what I call the Heron's Way, and it's still MUCH too common.




What The Heron's Way usually leads to is the digital equivalent of an empty cardboard box. How much good digital stuff is going to end up in that box, do you think?




Speaking as a long-time IR manager, if what I've got is an empty cardboard box, my instinct is to fill it, right? With any old thing. Including dubious scraps that I'm not totally sure anybody wants, but I take them because it's what I can get, or because I think it's important, regardless of the judgment of others.

Have I caught flak for some of these decisions? Are you kidding me? I've caught flak for ALL OF THEM. With the exception of electronic theses and dissertations -- anything I collect, anything at all, SOMEBODY thinks I shouldn't be collecting it, and thinks so strongly enough to tell me so.

# Or this:

Photo: cat in a box
http://www.flickr.com/photos/29278394@N00/5303091893/
Anne Norman / CC-BY 2.0

Or I may end up with things that are really nice, like this gorgeous cat, but that just don't fit in the box I happen to have. In non-metaphorical terms: things end up in institutional repositories that IR software isn't very well-suited to. Like institutional records, or image collections, or multimedia that you want to stream, or anything that you want people to interact with... things like that.

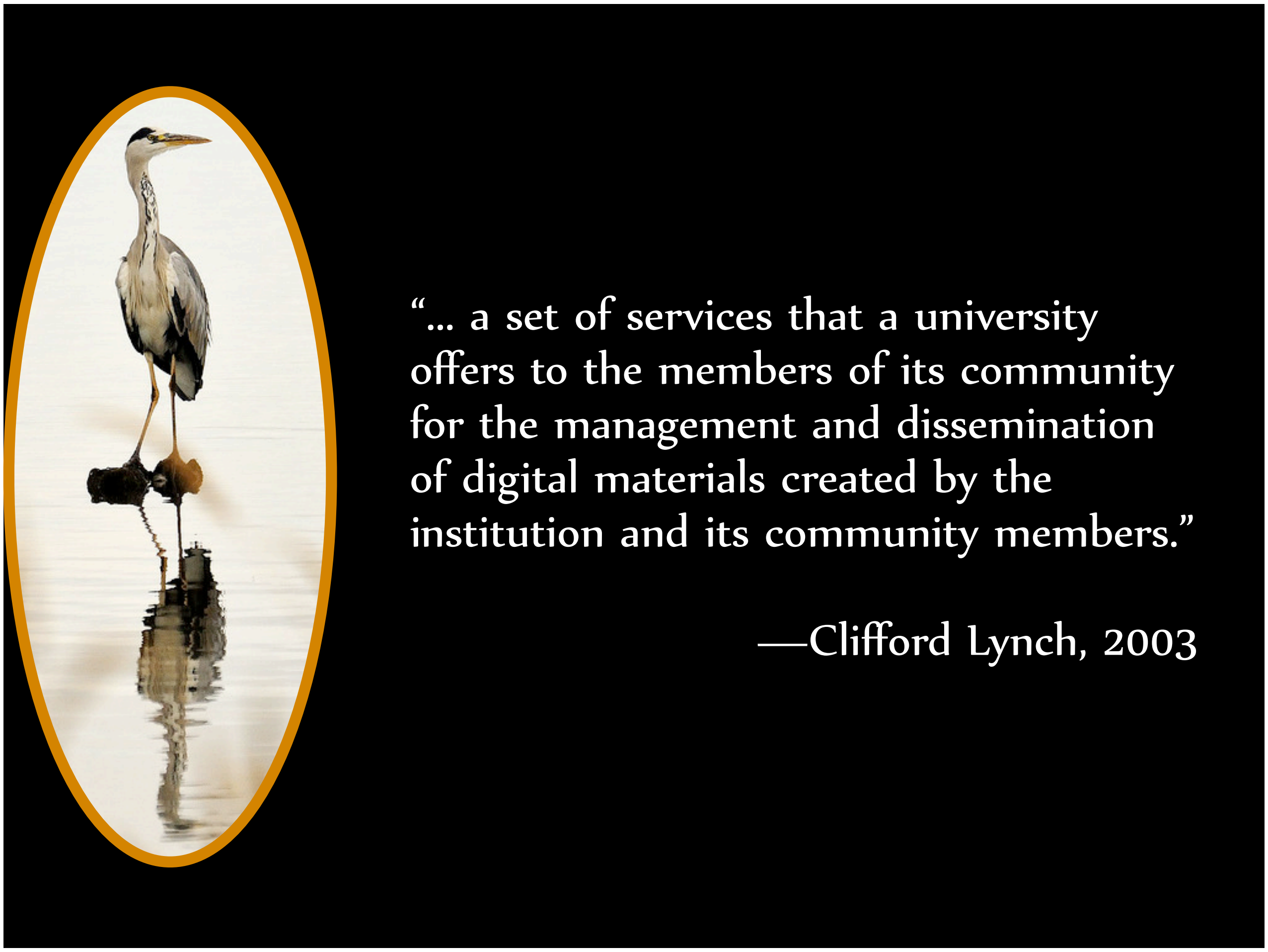

These aren't what we have in mind when we plan for institutional repositories, are they?

(CLICK) Right. They're really not. So we need not to take the Heron's Way. We can't open up an IR and just stand there.

What is an
INSTITUTIONAL
REPOSITORY
anyway?

(besides an ugly, vague bit of jargon)

So let's step back and regroup here. (read question, CLICK)




This is the classic definition that everybody uses, so I'll use it too. Cliff, if you're in this room, I apologize to you, man, and I am your biggest fan EVER -- but I'm also going to suggest that this definition of yours is, and has always been, REALLY INCOMPLETE from a planning and implementation perspective. Let's unpack it and see why I think that.

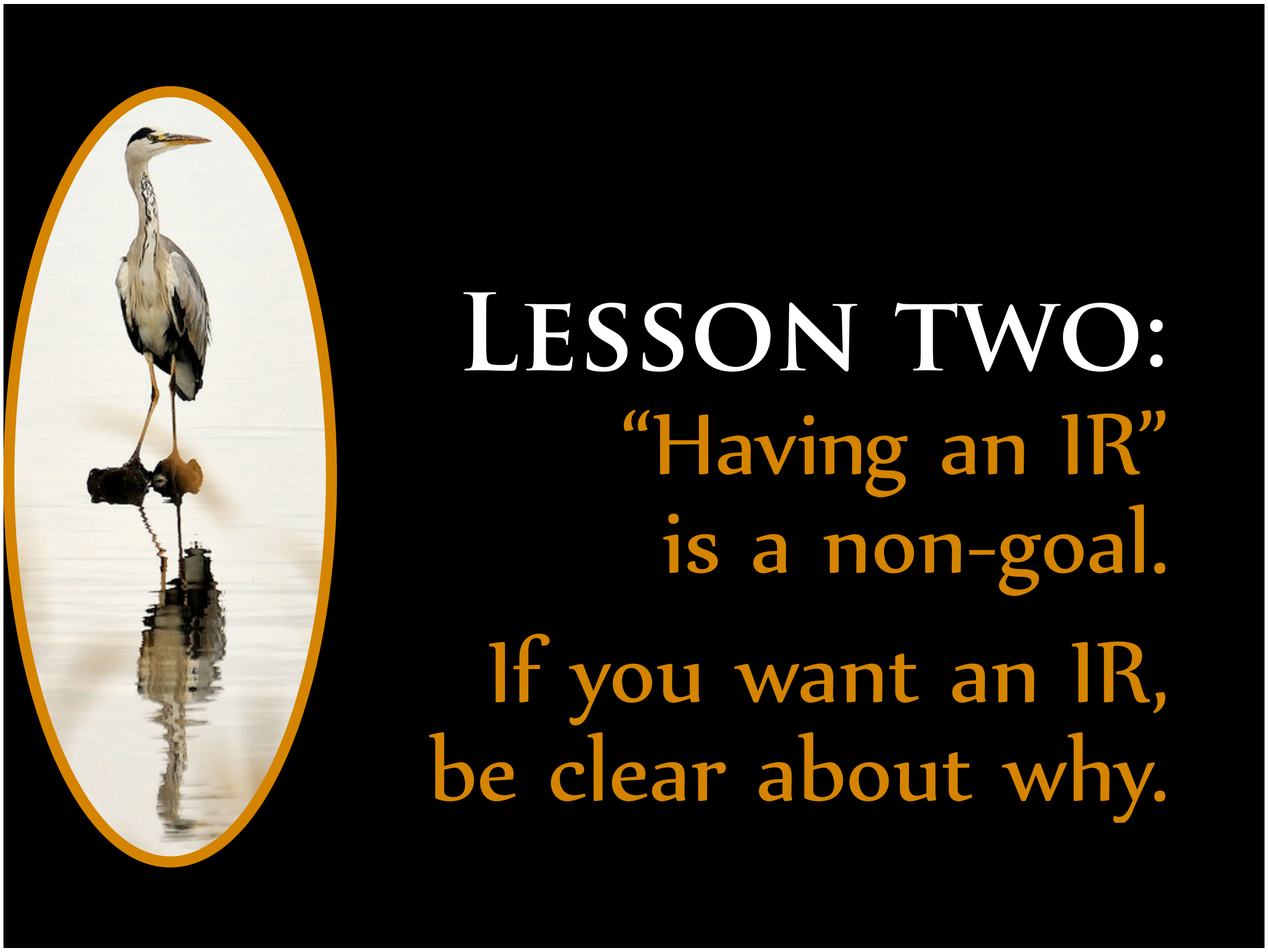
# Unpacking that…

| | |
|---|---|
| "… a set of services | What services? |
| that a university offers | University or library? |
| to the members of its community | Which members? |
| for the management and dissemination | Both of those? Really? |
| of digital materials created | Which materials? |
| by the institution | What does "the institution" create? |
| and its community members." | Which members, again? |

Without answers to these questions,
can we do anything
but stand there LIKE A HERON?

"A set of services:" what services might those be, then?
(CLICK) "That a university offers:" Does the university know this thing exists? Are its priorities the same as the library's? If they're not, whose priorities win out?
(CLICK) "To the members of its community:" which members? Do undergraduates count? What about records managers?
(CLICK) "For the management or dissemination:" Management? Really? So an IR is a content-management system? Did anybody tell IR software developers that? And what about materials that we can't actually disseminate, because of copyright or whatever reason? Do they count?
(CLICK) "Of digital materials created:" what materials do we actually want? We obviously don't want every single bit and pixel from everybody on campus!
(CLICK) "By the institution:" what does "the institution" create? Records, sure; do those belong in an IR?
(CLICK) "And its community members:" and we're back to which members again. (CLICK, read text)


So, our first two lessons. Lesson one: Cliff Lynch once said that an IR is "essential infrastructure for the digital age." I love you again, Cliff, but you were wrong on this one. Not everyone can or should have an IR. And there are no IR police! Nobody is going to revoke your MLS or your consortium membership or anything like that because your library doesn't have an IR. If you can't come up with answers you can believe in to my questions about Cliff's definition of an IR, don't have an IR! It's fine. Really.




Especially don’t open an IR just to “have an IR.” That’s just silly. Do you open a library building just for the sake of a building? Do you open a library building just so you can have some nifty shelving? I didn’t think so. You want the stuff that’s ON the shelves IN the building, not the shelves or the building per se. “Having an IR” is a complete non-goal for the same reason; nobody cares about an empty digital cardboard box! And nobody should! If you want an IR, you owe it to everybody in your library, especially the poor schmucks who’ll have to run the thing, to be clear about WHY you want it, what you want in that digital cardboard box.


So why might you want one, then? From where I'm sitting, I see three basic reasons libraries open IRs. (THREE CLICKS)

I'll discuss each of them in turn, and pull out some more planning lessons.


Early on in IRs' history, libraries opened them because they considered IRs their contribution to the open-access movement. We've opened an IR, so we're fixing scholarly communication! Go us! Of course, it hasn't turned out to be that simple.

Let me be completely clear about one thing: open-access advocacy is HARD. It takes patience, guts, and political capital. There's a ton of misinformation out there that you'll have to counteract. It's risky, too; faculty AND librarians are accustomed to the current system. I have scars aplenty from trying to do advocacy with faculty AND librarians who feel invested in toll-access publishing and the systems and structures around it.


How good an advocacy tool is an institutional repository? Well, in the early days of IR adoption, the thinking seemed to go something like this: we'll offer people an empty cardboard box, they'll put their stuff in it for some not-adequately-explored reason, and then magic will happen and the whole campus will go open-access.

(CLICK) Really? No, seriously, really? How does that work, exactly? I wish I knew. It's never happened to me.

Look, a lone IR all by its lonesome is a LOUSY advocacy tool. Terrible. And if open-access advocacy is the only reason you have an IR, it's in all likelihood a LOUSY reason. An IR can definitely be a backstop for a good advocacy effort, but all by itself, it's not advocacy.


Now, I am all for open-access advocacy! We need more of it, not less. But we'll do advocacy better without empty cardboard boxes confusing the issue. Here are some options:

National-level lobbying: The NIH Public Access Policy has done pretty well for itself, but it hasn't expanded beyond NIH's borders. If we want that to happen, we will need to lobby for it, because toll-access publishers are already lobbying against it.

Locally, there's plenty to consider: OA author funds. Memberships with reputable open-access publishers. Serials licensing agreements that help local authors publish OA. Library-internal open-access mandates, for libraries whose librarians publish in the literature. Plenty of examples of all of these; talk to me after if you'd like some.

# Example

"[T]his innovative and energetic individual… will work collaboratively with colleagues on the faculty and in the libraries to develop and implement an open access repository in support of the Open Access Policy, including policies, procedures, workflows, metadata, recruiting and harvesting content, and marketing and outreach to the University community advocating for best practices in open access."

—job advertisement

When I shared a draft of this presentation, a friend of mine said, "come on, surely nobody thinks an IR is only or mostly for advocacy?" By pure coincidence, two days later I came across an ad for a repository manager that made it VERY clear that advocacy is a major part of the job. So no, I'm not making this up.

Interestingly, this ad is from an institution whose faculty just put an institutional Open Access mandate in place, in the footsteps of Harvard. Which is very cool, so go them! But you notice something about that? The library didn't plan for, much less open, an IR until AFTER they had the policy mandate from faculty to do so! So let's talk about institutional open-access mandates a moment, because they're actually a REALLY GOOD reason to have a repository.

# Want a mandate? Find faculty champions.

Photo: Two Soldiers take fourth at U.S. National Boxing Championships 090617
http://www.flickr.com/photos/familymwr/4999667562/
Tim Hipps, FMWRC Public Affairs, US Army / CC-BY 2.0

If you have your eye on a Harvard-style open-access mandate: the first step is to find local faculty champions! We know that we librarians can't tell faculty what to do. We also know that faculty are intensely suspicious of college and university administrations, so we can't just go to the provost or chancellor and ask for a mandate. It's NOT coincidence that every single institutional open-access mandate I know of comes through shared-governance arrangements like faculty senates and library councils!

If you want this to happen where you are, find, cultivate, or BUILD an army of faculty champions who will fight for it. There is no other way. But you can do that without an IR! In fact, Harvard did. DASH didn't open until well AFTER the first mandates passed their faculty-governance bodies. Frankly, I suspect it's easier that way.

# Want a mandate?
# Eat your own dog food.

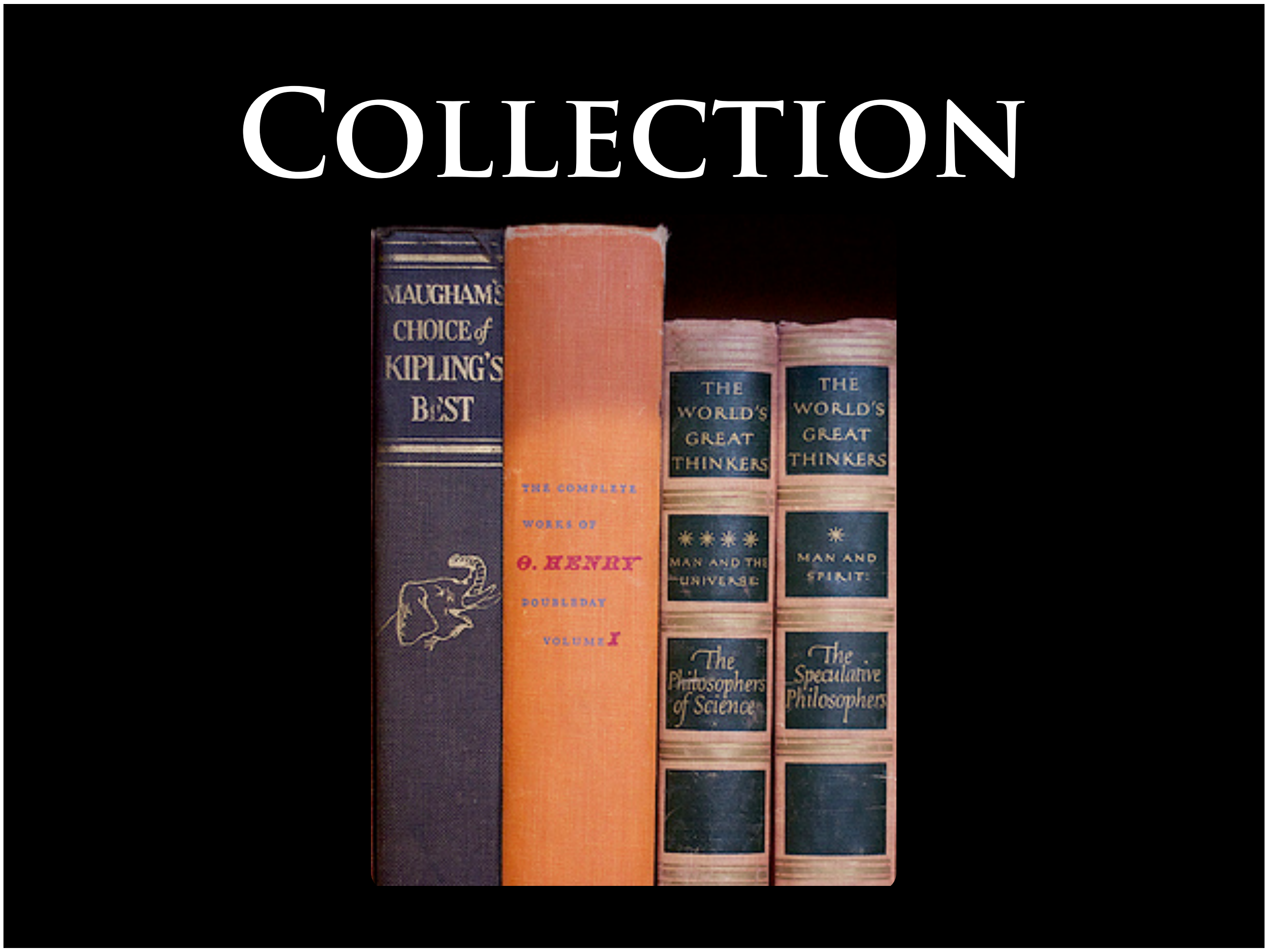
Photo: Dog food
http://www.flickr.com/photos/mshades/2132689629/
Chris Gladis / CC-BY 2.0

The other thing you can and should do if you're after a faculty mandate is SET THE EXAMPLE INSIDE THE LIBRARY. Several libraries, like Michigan and Oregon State, have instituted library-internal policies that materials that librarians present or publish must end up in the IR. I mean, how obvious is this? If we want faculty to do it, let's do it ourselves!

And if we can't make ourselves do it, on what ethical basis can we tell faculty that they should? Come on, folks, if we're serious about mandates it's hypocritical not to have one for ourselves. Especially if we're tenure-track.


So that brings me to Lesson Three...

You don’t get a get-out-of-advocacy-free card just because you have an IR. Sorry. Frankly, if all you have is an IR, you do not actually HAVE an open-access advocacy program. And if you call somebody a scholarly-communications or digital-repository librarian, and you expect them to create this huge cultural shift on campus all by themselves, and all you give them to work with is a lousy IR? You’re delusional. And you’re not a great planner and leader, either, because you’re sending that poor person into a gunfight with a butter knife. Think again. Think HARD.




So let's say you think of your IR as a container for well-chosen digital collections, the same way your buildings and shelves are containers for your well-chosen print collection.

# COLLECTION

## What do you want?

(n.b. not "what you will accept!"
That is THE HERON'S WAY.)

## How will you collect it?

(n.b. not "they'll just put it in!"
That is THE HERON'S WAY.)

You have two basic questions to ask yourself...


Here's some stuff you might want. I can't tell you whether you want it or not; that's for YOU to decide, just like any collection-development policy. But when you're deciding whether you want these things, you need to keep my second question -- "how will you get it?" -- in mind. Because part of deciding whether you want something, as any good collection developer will tell you, is deciding whether it's worth the effort and expense to collect -- and saying "I want something!" without figuring out a plan to collect it is a classic Heron's Way error. Trust me, what you want ain't just gonna show up.

So how WILL you collect these things?

## How you won't get stuff

# SELF-ARCHIVING

Nobody self-archives in IRs.
Not even LIBRARIANS.
We KNOW this.

Here's how you won't...

If the planning for your IR assumes that people will just magically self-archive, you have just set sail for the failboat.

# How you won't get stuff

# MARKETING

If the IR doesn't help anybody,
marketing won't help the IR.
We KNOW this.

This time I need to apologize to Ann Hanlon, because she's going to be talking about this, so I'm sorry, Ann, and I hope with all my heart you prove me wrong. But there have been ENDLESS webinars and conference sessions and articles about IR marketing and outreach. I haven't seen too much result from the techniques they suggest. If your fundamental stance toward the IR is The Heron's Way, completely passive except for some marketing, you're sailing for the failboat again. Marketing is just not enough. If you don't look at an IR from a faculty member's point of view and make sure THEY will get something tangible and useful out of participation in the IR, no marketing plan will help you.

Now, if you're sneaky, you can engineer a success or two, THEN start marketing. I hear that sometimes works. So does having a library-internal mandate, sometimes. But don't kid yourself, engineering a visible success is the hard part here!

## Your inevitable bête noire

# COPYRIGHT

People will put in stuff they shouldn't.
People are afraid to put in stuff they can.
We KNOW this.

Here's a trap that caught a lot of early IRs. Nobody planned to help anybody out with copyright issues! People, copyright is a MORASS. It touches almost everything you might want to put in an IR, from dissertations and theses, to faculty who think they own copyright in articles when they don't, to really stupid faculty who want to put in somebody else's work, to third-party copyright questions, to people who are terrified to put in a lesson plan with a two-line quotation from a book because they think they might get sued.

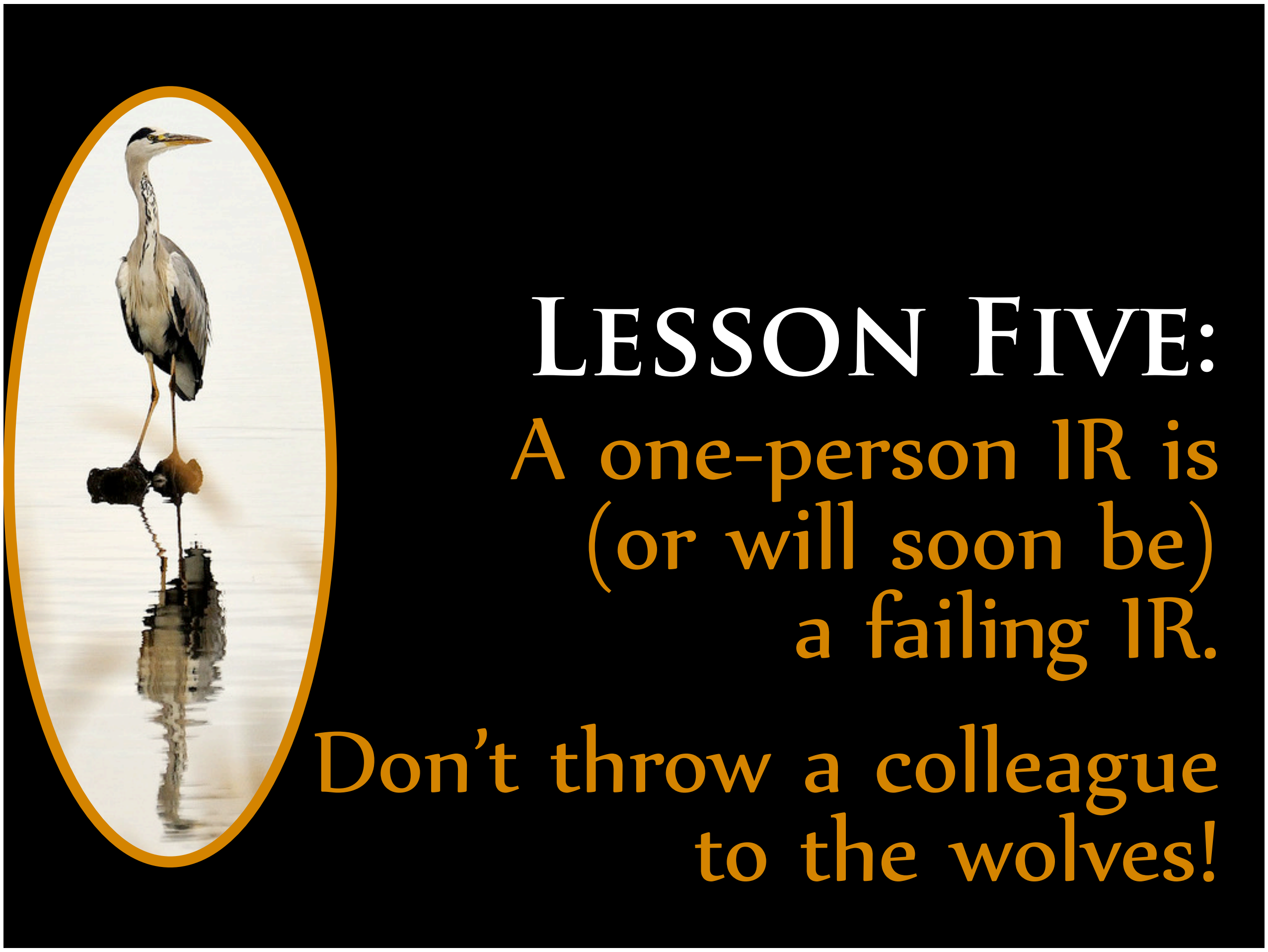
I hate this too. I hate it a lot. But it's a reality. Your IR plan had better plan to deal with it!


After all that gloom and doom, here’s how you MIGHT catch some fish, as this heron has. (go thru list)

These strategies often work together. You might, for example, crawl your local faculty’s web pages and find -- HALF a working-paper series. The other half only exists on paper. If you can digitize it, chances are they’ll give you the whole thing! If you can’t, they won’t be interested.


You might have noticed that all those strategies involve time and effort, well beyond the technical effort involved in setting up and maintaining IR software, well beyond marketing. Yep. That's right. But how many IRs plan for that time and effort?

If you don't WANT to put any such effort into an IR... remember Lesson One. You don't have to have an IR in the first place! But wimping out on the work is the Heron's Way. Don't go there. You won't catch fish.

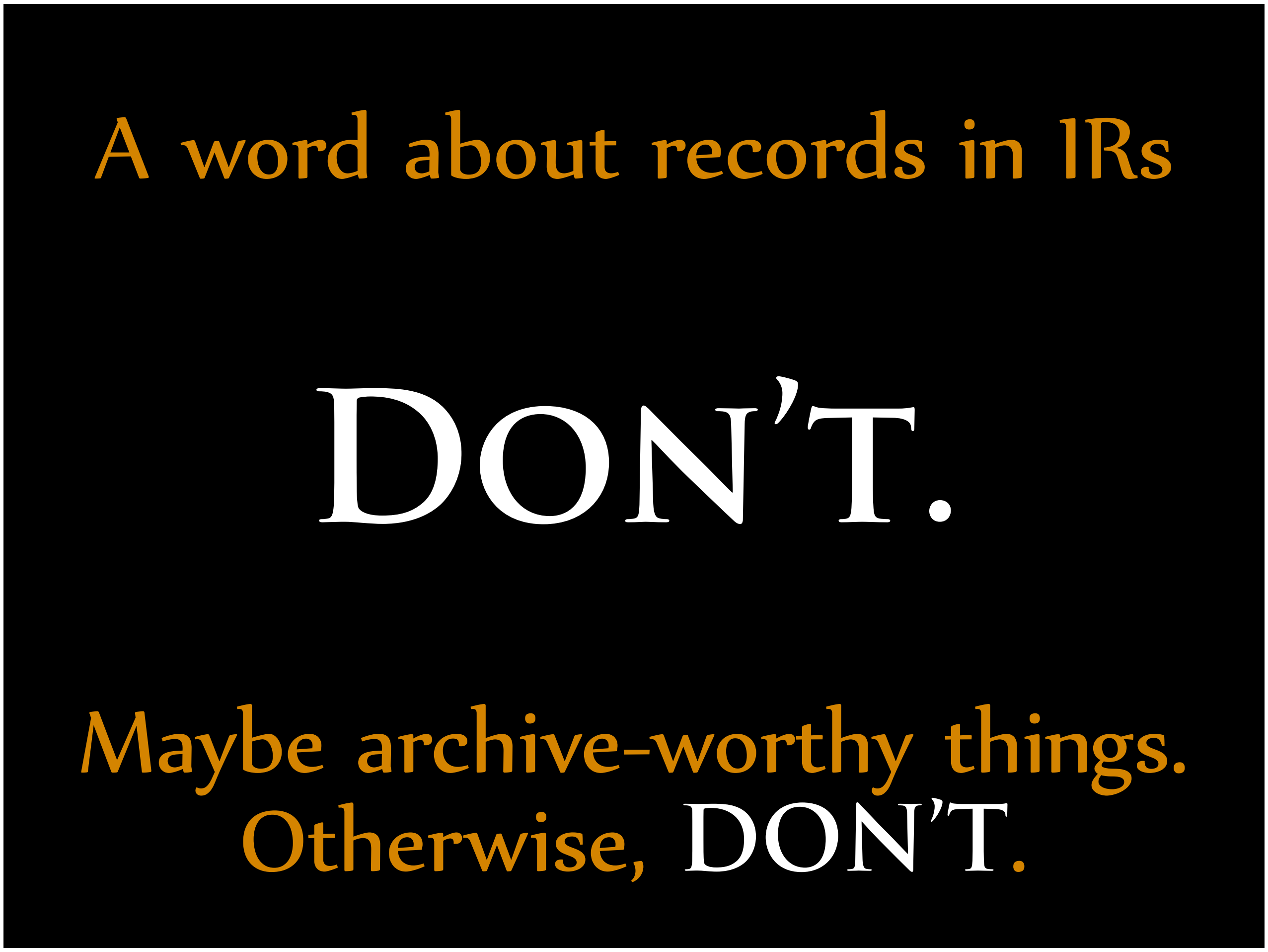



The corollary to "it takes work" is "it takes work from more than one person."

Now, obviously this depends on the size of your campus and your library. If you're a six-person library, feel free to ignore me. But if one person can't do regular collection development for your whole library, why would you think one person can "do the IR"?

In my six years doing this, I've seen plenty of libraries hire a single IR librarian -- often a brand-new librarian, at that -- and think they're done. More often than not they then completely ignore that person, except to blame them when the IR isn't what they think they want it to be. This is a horrible, reprehensible thing to do to a colleague, especially a colleague with a young career, as often happens. If the IR is NOT a whole-library priority and a whole-library initiative, why do you have one? Remember Lesson One!

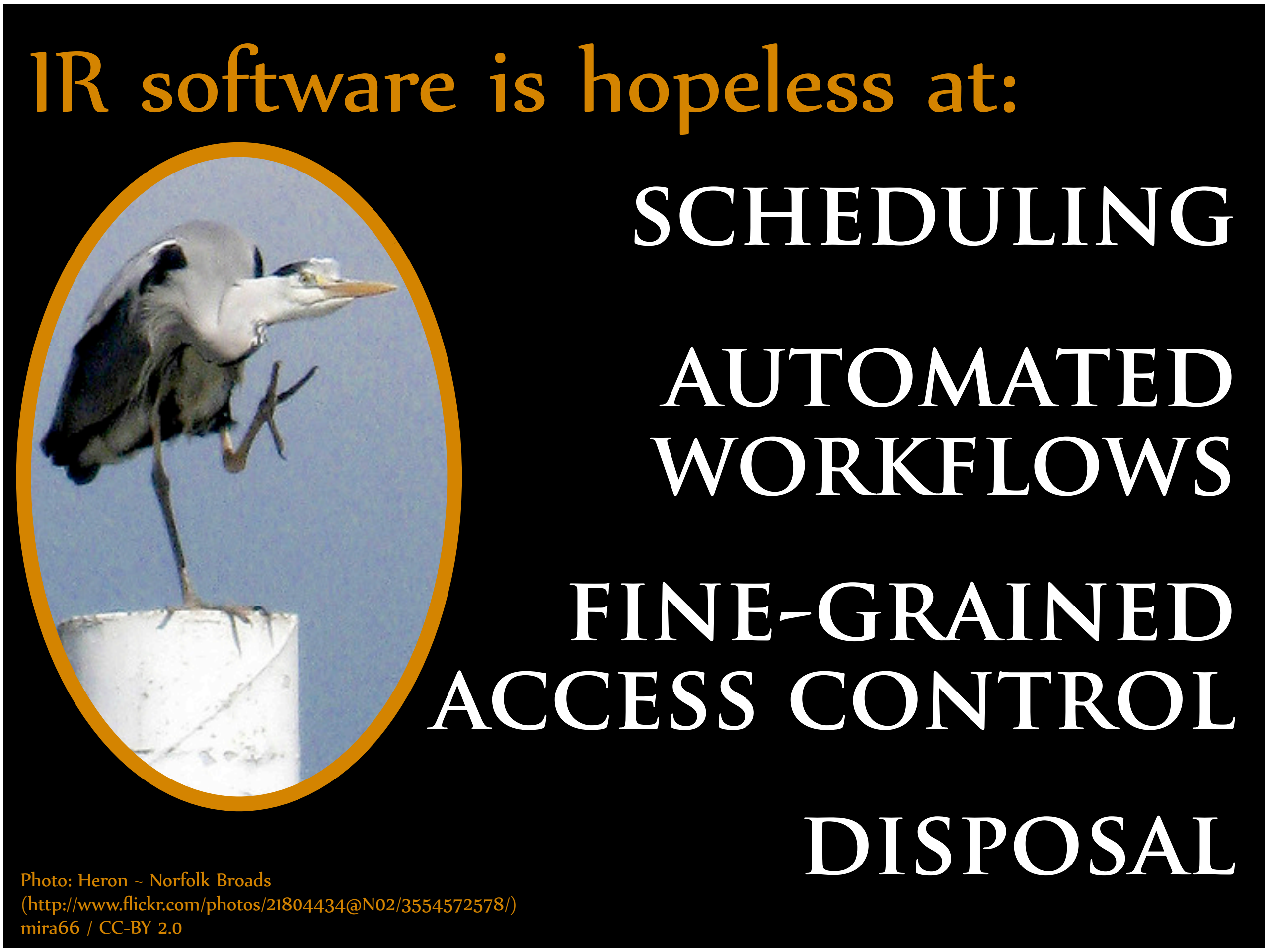
Successful IRs have more than one person behind them. Ohio State knows this. So does Harvard; they've got a whole office! So does Illinois. Plenty of other examples.




A quick word about records in IRs...

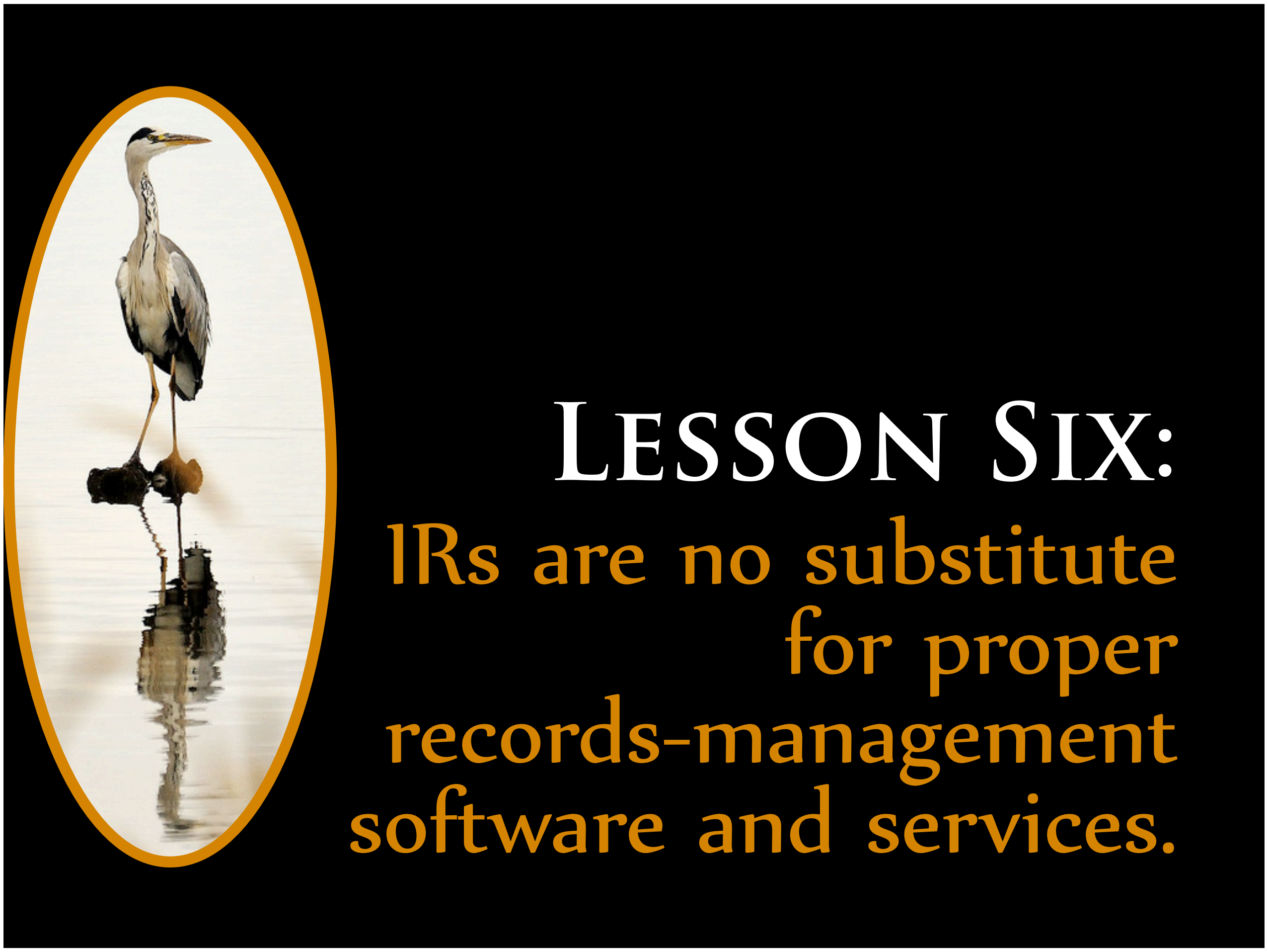
I admit I've changed my thinking on this over time. I used to think no-harm-no-foul, but honestly, I think IRs have to some extent gotten in the way of building usable e-records-management systems, and THAT is AWFUL. So I'm serious. Don't use IRs for records management. I see small schools particularly doing this, and I think they're making a big mistake, adopting the wrong software for the job at hand.

(CLICK) Now, once something has gone through the records mill and become genuinely archival, it might be worth posting. That's your call. But for day-to-day records management, seriously, DON'T.




(go thru list)

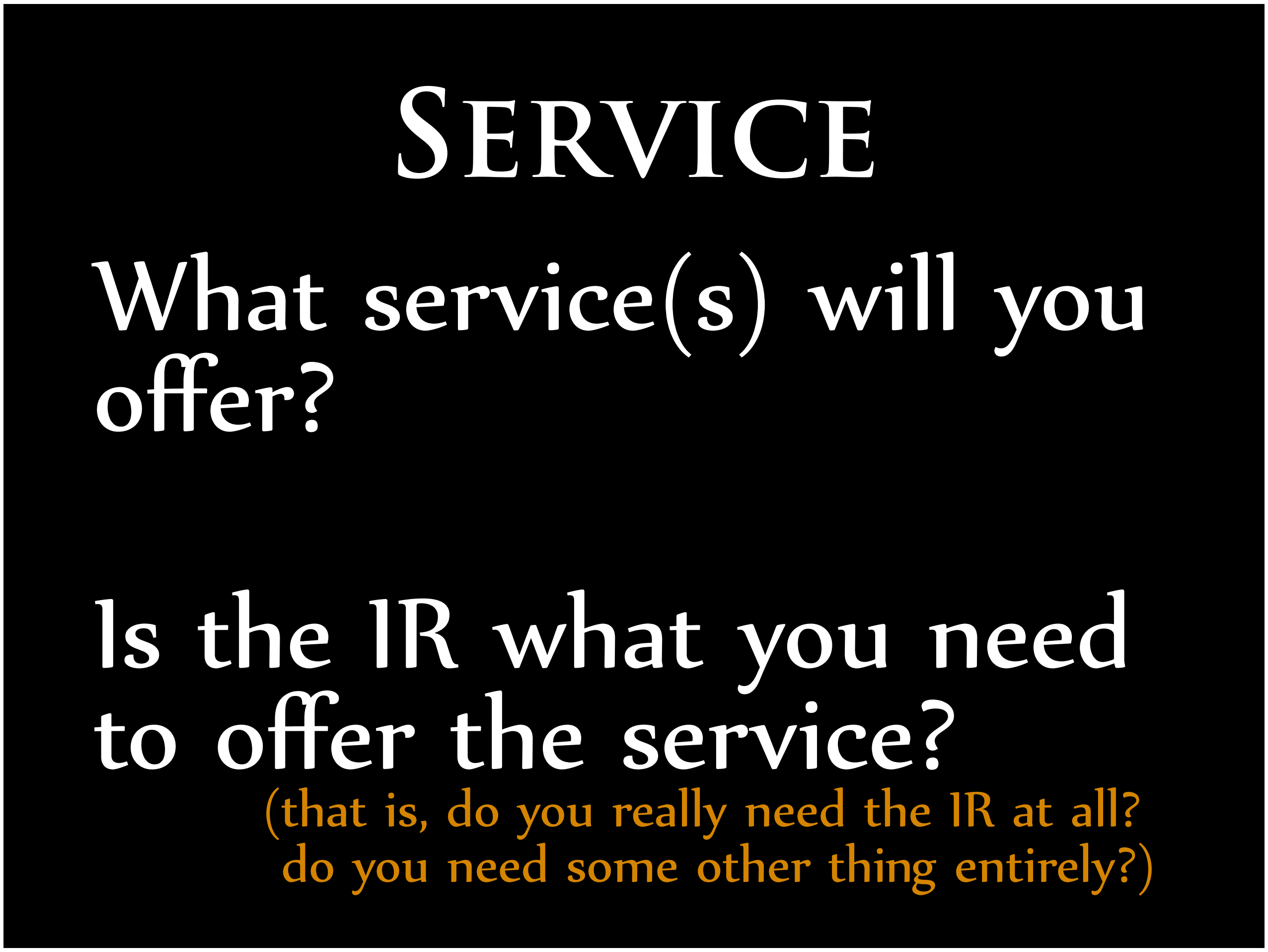
IRs also have bells and whistles you don't even WANT if you're a records manager. Do you want WorldCat harvesting your meeting minutes through OAI-PMH? Yeah, no, probably not. So your records don't even NEED OAI-PMH, but with an IR you'll get it even though you don't need it, and you'll look as goofy as this heron here does.




If you want records-management software, get it. Look at Archivematica and iRODS and Kuali and whathaveyou. Don't use an IR. The software is just hopelessly inadequate for what you need.

# SERVICE

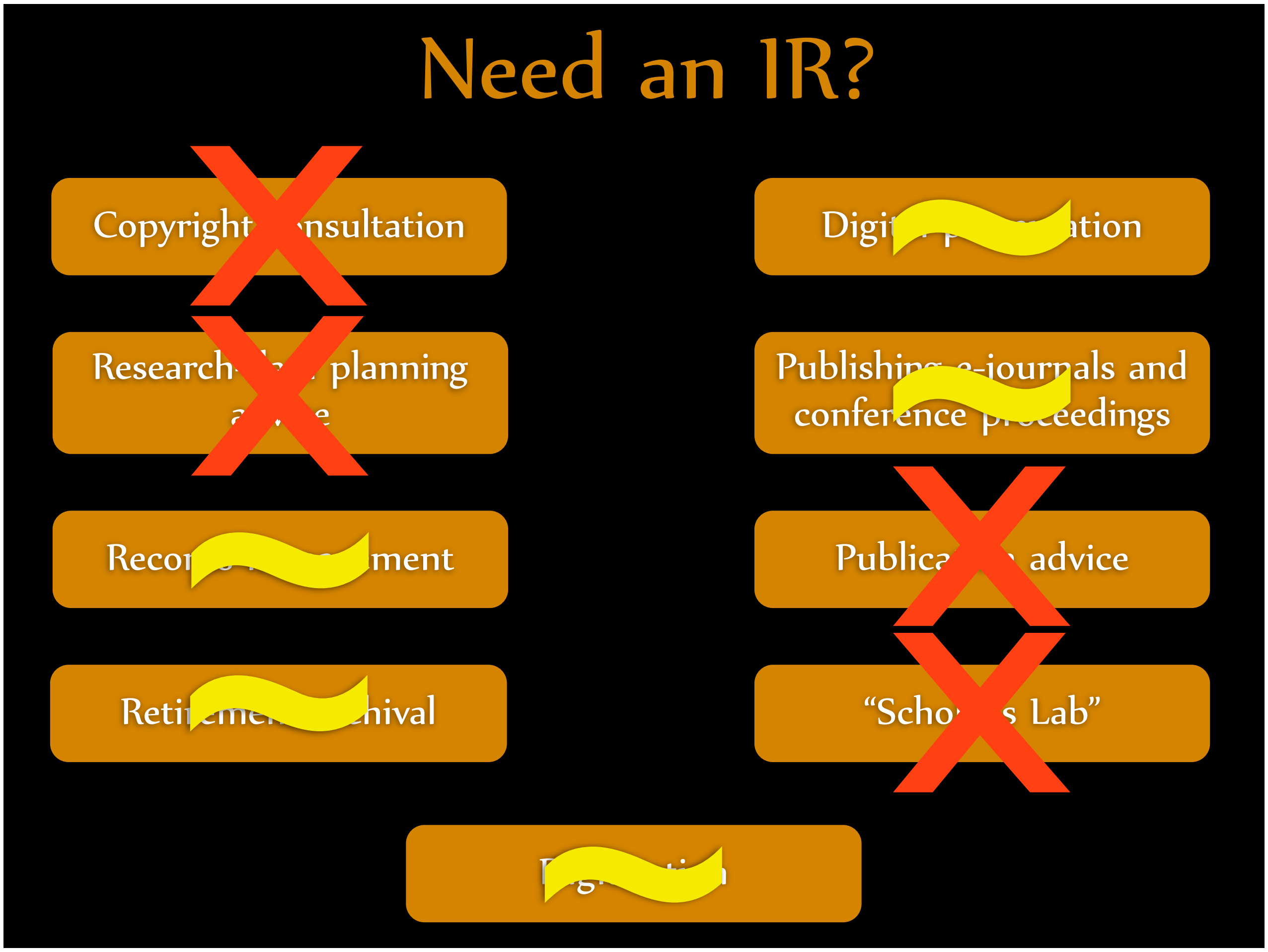
So you think your IR is a service, the way Cliff Lynch said...




Again, I think you have two questions to ask yourself...

# Services you might offer

- Copyright consultation
- Digital preservation
- Research-data planning advice
- Publishing e-journal and conference proceedings
- Records management
- Publication advice
- Retirement archival
- “Scholar’s Lab”
- Digitization

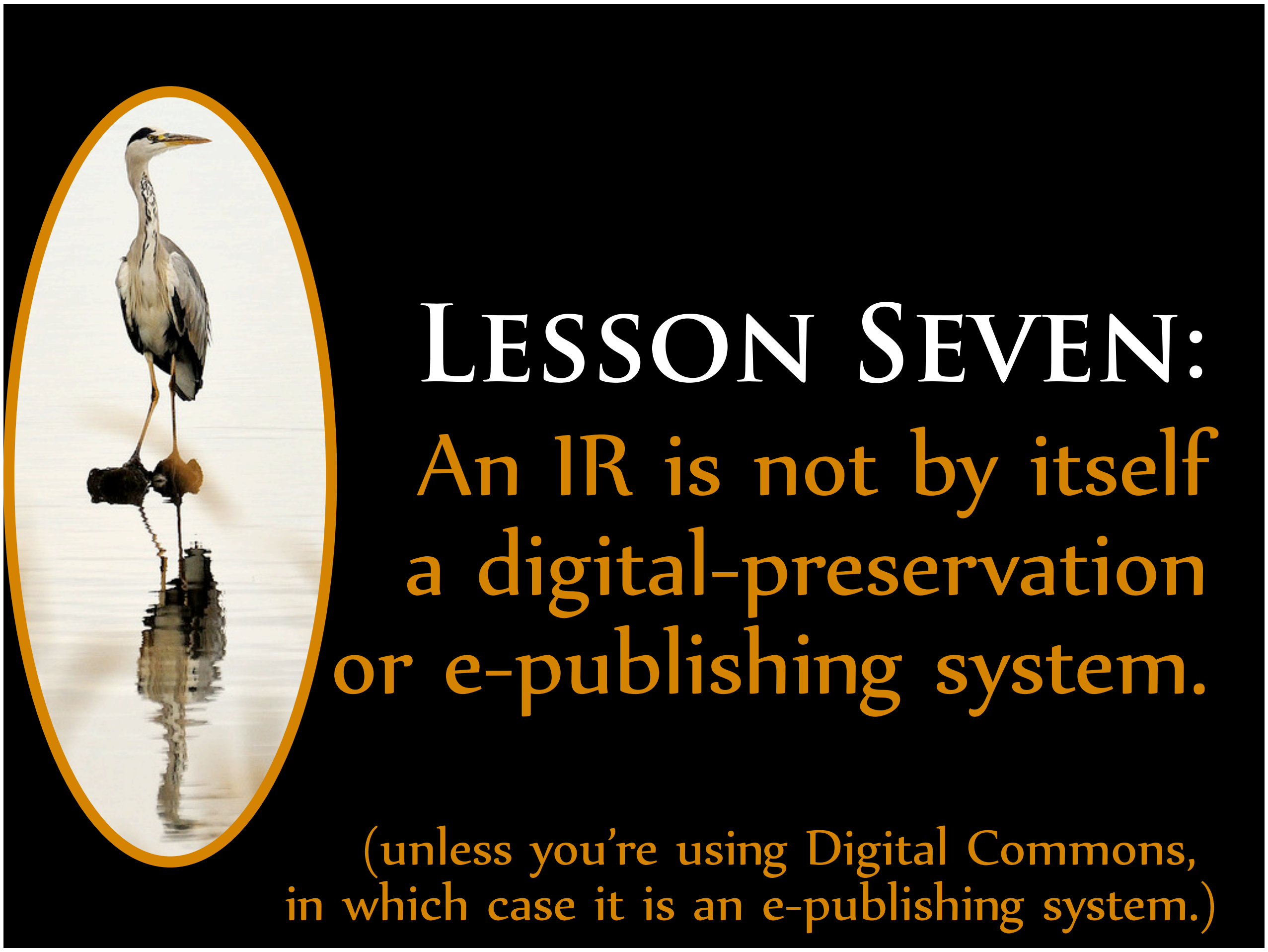



So do you need an IR to do these things?

You don’t need an IR just to give advice or consult. Whether it’s copyright, research-data planning, publishing, or the refdesk-y kind of function you’d find in a Scholar’s Lab, you DO NOT need an IR to do it! Just do it!

(CLICK) For some other things, maybe you want an IR, maybe you don’t. Think hard before you assume that of course the IR is perfect for whatever you have in mind.

Notice what you’re not seeing here? Green checkmarks! It’s amazing how often IR software is just plain not necessary to do something useful. To be clear, if you’re still bound and determined to have an IR, some of these are services you SHOULD be offering alongside it. Never underestimate the copyright monster, or the appetite for digitization services!




As I mentioned before, libraries commonly think that with IRs they're offering digital-preservation services. Well, they're not, not really; they may be offering pieces of a digital preservation service, but not the whole hog.

If you're interested in digital preservation, use the Trusted Repository Audit Checklist or DRAMBORA as planning tools, because digital preservation goes far beyond software. Just as IRs aren't get-out-of-open-access-advocacy-free cards, they aren't get-out-of-digital-preservation-free cards either. There are LOTS of services you might want to offer in the scholarly-communication and digital-preservation spaces; don't make the mistake of thinking that opening an IR is all you need to do to offer them.


So right now you might be thinking that you want it all. You want an open-access advocacy tool AND a quality digital-materials collection AND a quality service!

(CLICK) Well, gee, I want a pony. Can I have a pony?

(CLICK) What you need to understand is that these goals, fine as they all are, conflict with each other in practice. While you're planning, you need to be aware of the tradeoffs.


If you're doing open-access advocacy, classically what you want is the peer-reviewed journal literature, usually in preprint or postprint form. Well, I'm afraid that's the hardest thing to pry out of people in the absence of a mandate. (CLICK) So what you end up with is the empty box. If you're ALSO assessing the IR by the size of its collection, you're going to be disappointed.

(CLICK) The same is true if you think you're doing advocacy plus service. Nobody actually WANTS an open-access service, so you wind up right at the empty box again. This is the mistake early IR adopters made in droves.

This is why I said earlier that advocacy needs to be a separate thing. It's harder to do, harder to assess, and doesn't fit well with other common goals for IRs.


So what happens when you want both a great collection and a great service? Let's assume for the moment that the service your IR claims to offer is digital preservation, because that's fairly typical. (CLICK) Well, honestly, your collection will probably turn into a bit of a scrapheap. People want to preserve the weirdest and most useless things! People aren't good at selection and weeding the way librarians are! Are you ready to do that for them? If not, are you offering the service they need?

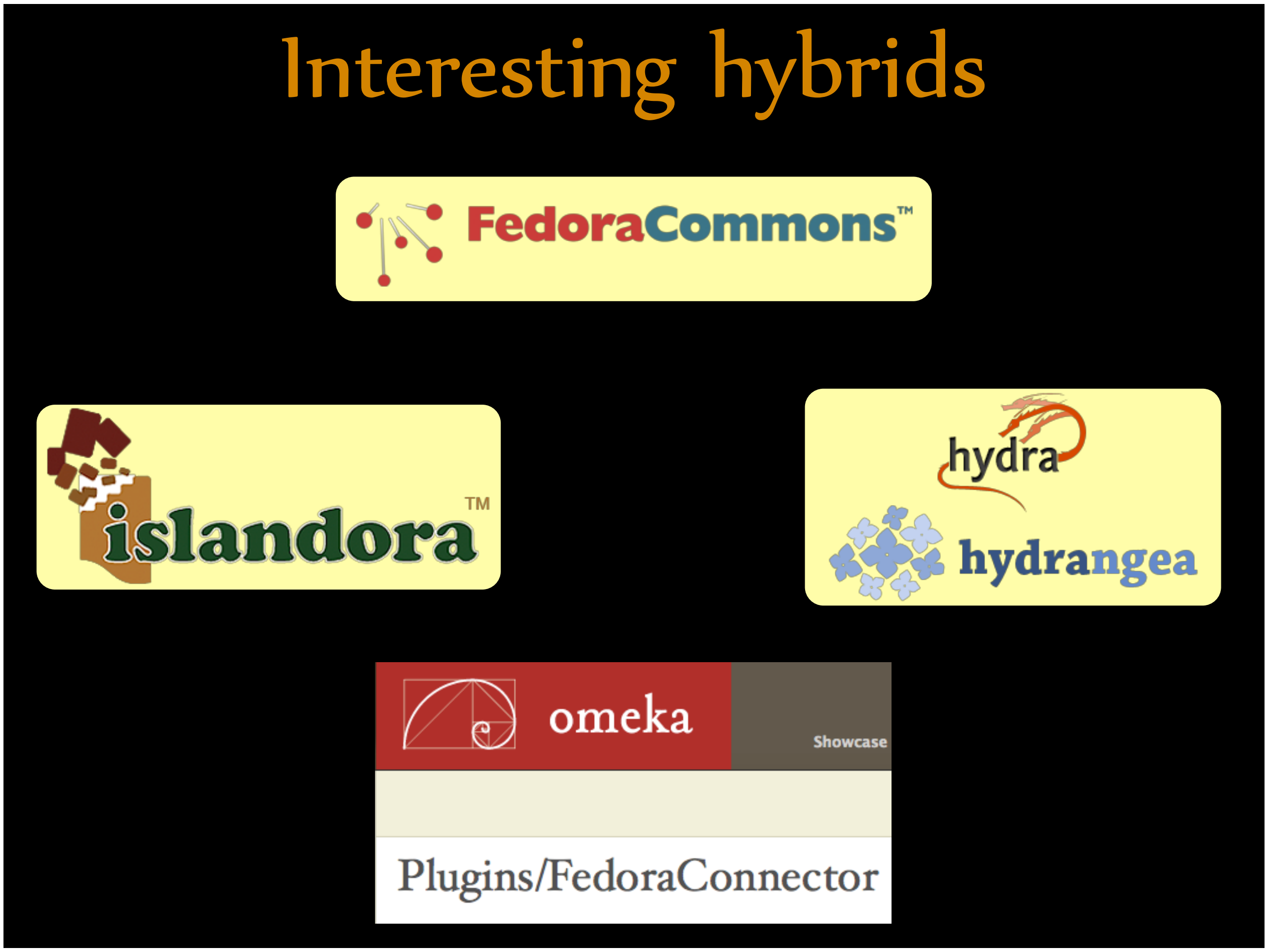
And if you bait-and-switch, tell people you're a generic digital-preservation service and THEN tell them no, you DON'T want fifty years of their department's faculty-meeting minutes, or terabytes of badly-digitized video from their local speaker series, they turn around and never darken your door again, even with stuff you DO want, and they'll tell all their faculty friends to avoid you too. Another mistake a lot of early IRs made was writing these incredibly strict collection policies, which led them straight to the empty box. If you really want to be a service, then taking in stuff that makes you hold your nose is part of the cost of doing business.

(CLICK) You also run into the stuff-that-doesn't-fit problem. A lot. This is largely because most IR software is so overoptimized for the single PDF article. IR software won't stream audio and video. It doesn't have an internal pageturner for page-scanned books and articles. It doesn't have image galleries or nifty Flash embeds. It's really a big ugly mostly-useless silo. You need to plan around these limitations, at least for now.


Because IR software is what it is. It isn't necessarily what you WANT it to be...

Worst of all, IR software does NOT play well with others. It won't play with your courseware, or your content-management systems, or your catalog, or really almost anything you can imagine you'd want it to play with. IR software is profoundly antisocial. This is a HUGE problem, but you can't just wish it away -- you have to plan around it!

You notice how often "management" comes up in these descriptions of what IR software is not. This is something else a lot of early IR adopters didn't understand: most IR software assumes that what you're putting in is the final, immutable version of something. Funny thing about that: most people don't think they need a whole service just to manage the final version of something digital! To manage digital things while work is ongoing: THAT they need. But IRs won't do it.



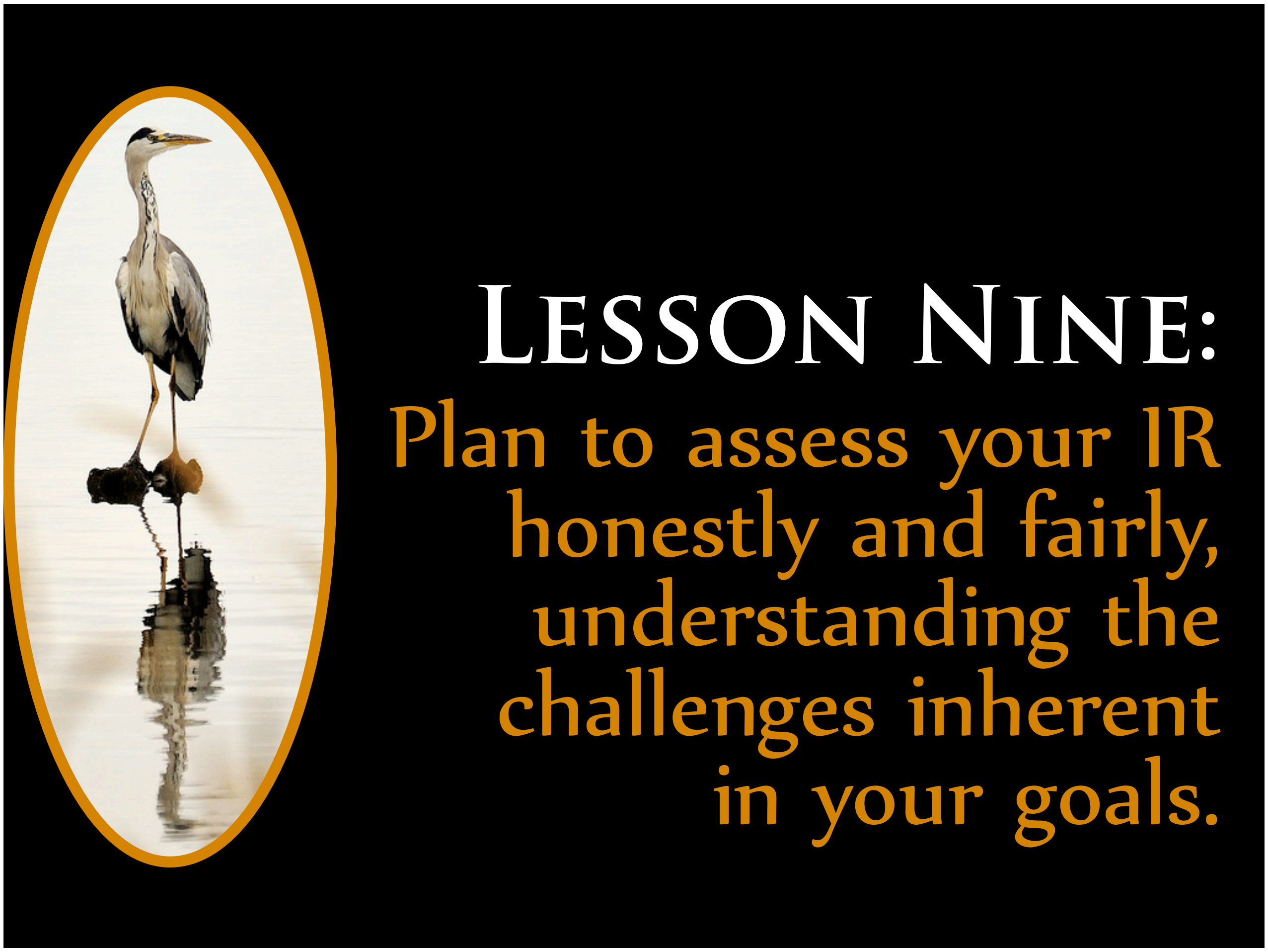

Now that said, there are some interesting software hybrids out there now. I already mentioned that Digital Commons is an IR/e-journal publishing system hybrid. But in my book, the better hybrids are coming out of Fedora Commons land. Fedora Commons is an open-source software package that provides a lot of the underlying engineering you need for a useful, reliable repository. What Fedora DOESN'T provide is all the user-interface chrome on top. So that opens up the intriguing possibility of putting all kinds of DIFFERENT user interfaces on top, both for getting stuff INTO the repo and for using stuff once it's there.

And that's what these hybrid systems are doing. Islandora, for example, marries Fedora with the open-source content-management system Drupal, which is incredibly flexible and powerful. Hydra is an attempt to build a content-management stack from the ground up atop Fedora; "hydrangea" is its first demo system. And the University of Virginia Scholar's Lab is connecting up the online exhibition software Omeka to Fedora, via an Omeka plugin they're calling FedoraConnector. There are lots more of these hybrids; RUCore from Rutgers is another great one to look at.

So fairly soon, I think, you might be able to have and eat a few more cakes. I hope, anyway. I do think this style of hybridization is the right way to go, whether your underlying base is Fedora Commons or curation microservices or iRODS or something homegrown. Silos are awful.


“What software (or service) will we use?” is often the FIRST question someone asks in an IR-planning process. This is completely counterproductive. PLEASE don’t do it. You MUST know what you’ll be collecting first, and how, and how you’re going to assess success, so that you can gauge how different platforms fit what you’re trying to do. If you don’t choose the platform to fit the content and the workflows, the content and the workflows will be CONSTRAINED by your choice of platform! This is not what you want!

There are a lot of DSpace-based IRs in the States. I’ve been running DSpace IRs for six years, and I honestly don’t understand its broad adoption. I have to figure that planning processes chose software first instead of last, and I also have to guess that “nobody ever got fired for choosing DSpace.” Well, maybe somebody should have been!




So all these equations and all this talk about software affordances, it's really about assessment, which is an important thing to build explicitly into your planning. People have asked me for YEARS, "where are the successful IRs?" "What does a successful IR look like?" I ask back, and oddly enough, nobody ever has an answer.

Look, you have to establish your success conditions up-front. I can't tell you how many IRs never do that. I CAN tell you what not doing that leads to: a lot of confusion among librarians and potential IR users about what the IR is there for, IRs that aren't as successful as they could be if it was clearer what they're supposed to accomplish, and a lot of hideously painful uncertainty for IR managers -- unclear success conditions are a GREAT way to abuse your IR manager.

So once you have your success conditions, figure out when and how you'll assess them, and think about what you'll do if your goals aren't met. And be fair about it! If you decide you're running a preservation service, it's NOT FAIR to assess on collection quality. If you're running an advocacy effort with an IR as backstop, it's NOT FAIR to assess solely on collection size (though hey, if you have a big collection of quality stuff, clearly your advocacy is getting somewhere). If you're focused on building a quality collection, whatever your definition of "quality" happens to be, it's NOT FAIR to complain that the IR isn't solving everybody's digital-preservation problems.

# Lessons

1. It's fine NOT to have an IR.
2. "Having an IR" is a non-goal. If you want an IR, be clear about why.
3. Open-access advocacy is a separate question from IR planning. Do justice to the intricacies of both.
4. IRs take work. Anybody who says otherwise is selling something. Don't buy.
5. A one-person IR is (or will soon be) a failing IR. Don't throw a colleague to the wolves!

So, time for a recap.

# Lessons

6. IRs are no substitute for proper records-management software and services.
7. An IR is not by itself a digital-preservation or e-publishing system.
8. Choose your platform LAST. Choose what fits your goals and workflows, current and future.
9. Plan to assess your IR honestly and fairly, understanding the challenges inherent in your goals.


All this leads to my final lesson. I didn't come out of library school in two-thousand-five to run my first IR knowing all this. Nobody in libraries knew all this then; we were still experimenting and learning. So these lessons and others like them weren't even in the library literature then! But they are now, so don't stick your head under the water and ignore them! READ the literature that's out there now. ASK people who have been there and done that and have the scars to show for it. If you're not reading Carol Hixson and Denise Troll Covey, what is wrong with you?

And then go do better. My own example is a very low bar. VERY low. I did just about nothing but fail for six very long and frankly agonizing years of this running-IRs business. You can beat me! All of you! So go plan your IR right, and show me up right and proper.


Have you ever seen an osprey hunt? It’s fabulous. They survey the whole lake from on high until they see something they KNOW they want, and then they hurtle down and nab it with a huge splash. I think IRs need to be more like ospreys than herons. Decide what you want with your eye on the big picture, and seek it out, wherever it’s hiding. Make the effort to dive for it. And grab it -- don’t wait for it to come to you.


Thank you!
This presentation is available under a Creative Commons Attribution 3.0 United States license.
BY
Photo: Pops with a snack (http://www.flickr.com/photos/kim/3483559778/) / Kim Taylor / CC-BY 2.0