Thank you for the opportunity to speak with you today, and thank you to the folks at Iowa for once again including Digital Asset Management on their agenda.

My name is Alison White, and I work in Television Operations at the Corporation for Public Broadcasting. My fellow evangelists today are Grace Agnew, of Rutgers University and Dave MacCarn of WGBH, Boston.

Some of you may recall that I made a presentation about Asset Management at last year's Symposium. I'm happy to report that my Iowa speech had some pretty good legs for me – I used versions of it at conferences for months and months afterward.

Well, I recently dragged out and dusted off that old presentation, because you never know… and I must say, a year later, my own words seemed overly optimistic and my vision of the future, downright quaint.

The Asset Management environment has gotten much more serious – less fanciful and more focused. You could argue that that's a result of the many dot-com death notices that affected the Asset Management industry.

But I think it's because broadcasters – along with our counterparts in business and academia – now fully grasp the potential return-on-investment for sensibly managed data, especially rich media. We can clearly picture our satisfied clients and customers and constituents using all that data and media in productive and lucrative ways.

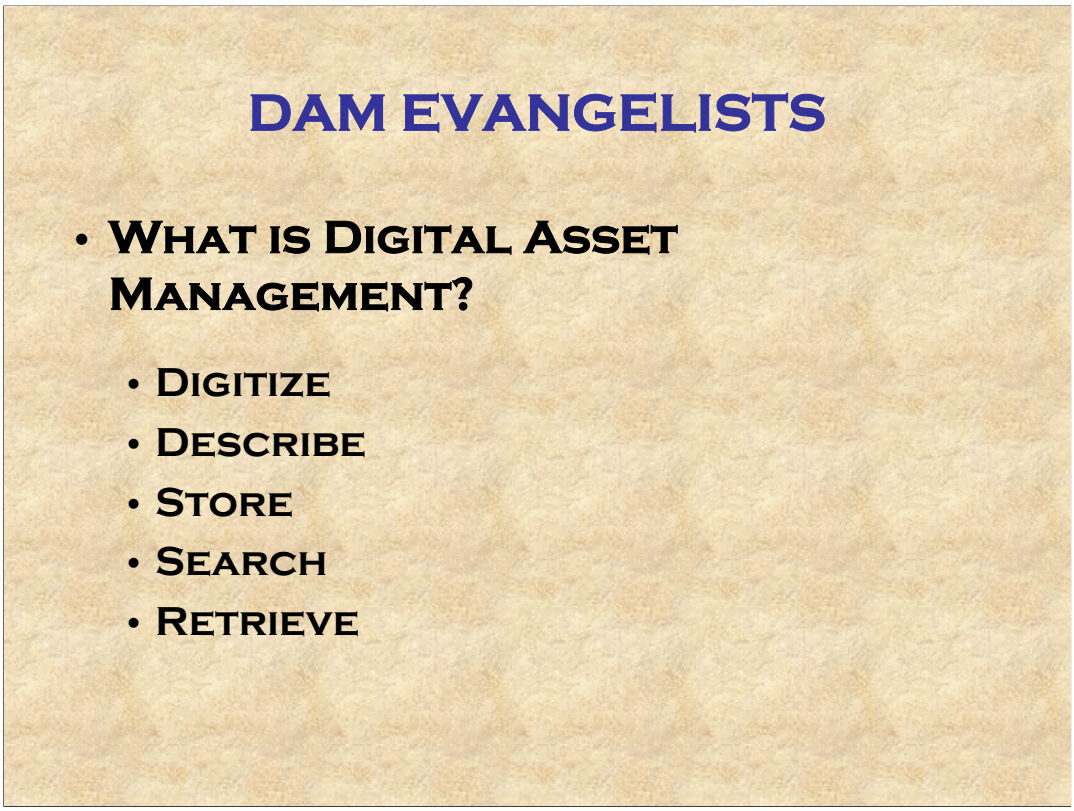
The bummer is that we also now fully grasp just how complicated and expensive and time-consuming it is to actually achieve that kind of data management.

In this session, we're going to offer what we hope are some uplifting and instructive Asset Management stories – Gospel readings, if you will.

I'm going to start by discussing public broadcasting's attempts to come to agreement on a single set of rules for labeling our content assets via a shared Metadata Dictionary.

Grace Agnew will describe a wonderful effort by the Association of Moving Image Archivists, supported by the Library of Congress, to create a single, powerful web portal for media collections around the world.

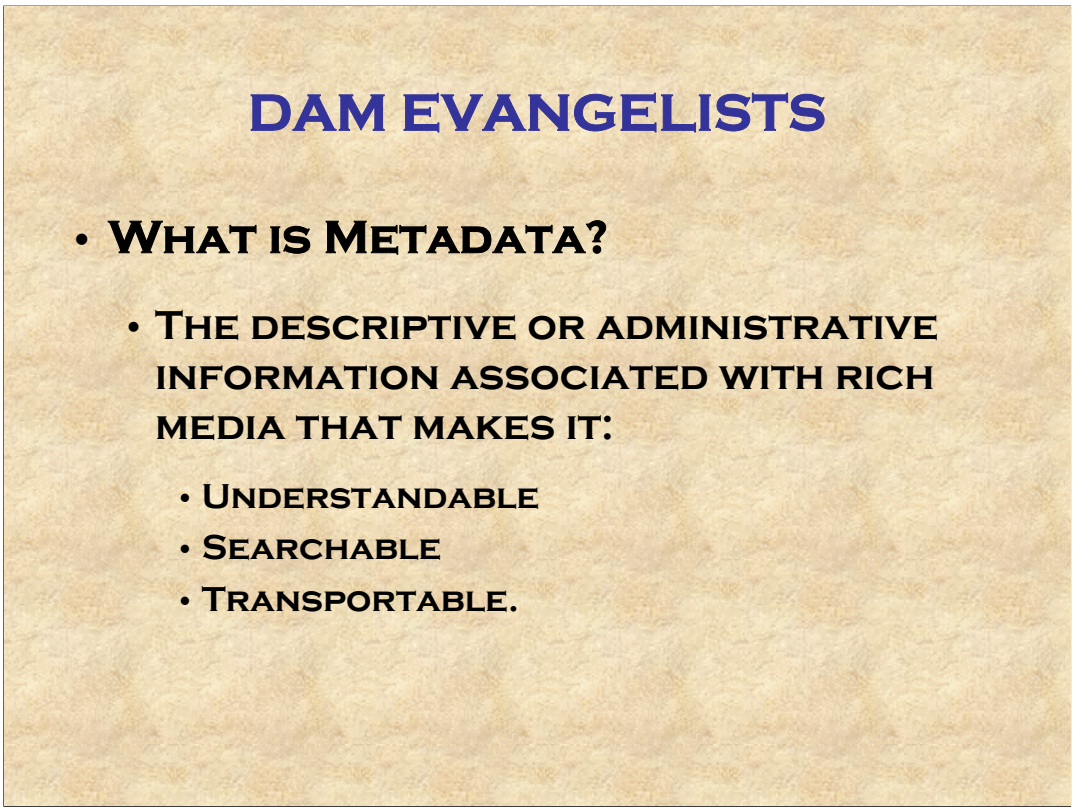
And then Dave MacCarn will show us WGBH's progress toward managing rich media assets in a complex multiplatform production environment.



First, I want to make sure that everyone is comfortable with the basic terminology.

In the context of broadcasting, "digital asset management" is a software-slash-hardware solution for managing rich media. It includes:

- ingesting or digitizing text, audio, video, stills, etc. at various qualities, or bit rates
- naming and describing the content in a structured fashion, adding what is called metadata
- storing the content and the metadata associated with it, either separately or together
- browsing and searching those databases, via an intranet or web portal
- and finally, retrieving the media as needed, via streaming, tape or other delivery.



And what is "metadata" again?

Well, it's frequently described as "data about data." That's a somewhat annoying definition but it is accurate. Once your rich media assets enter the digital environment, they become "data" – sets of ones and zeros. And when you add information that describes those assets, that's data about data.

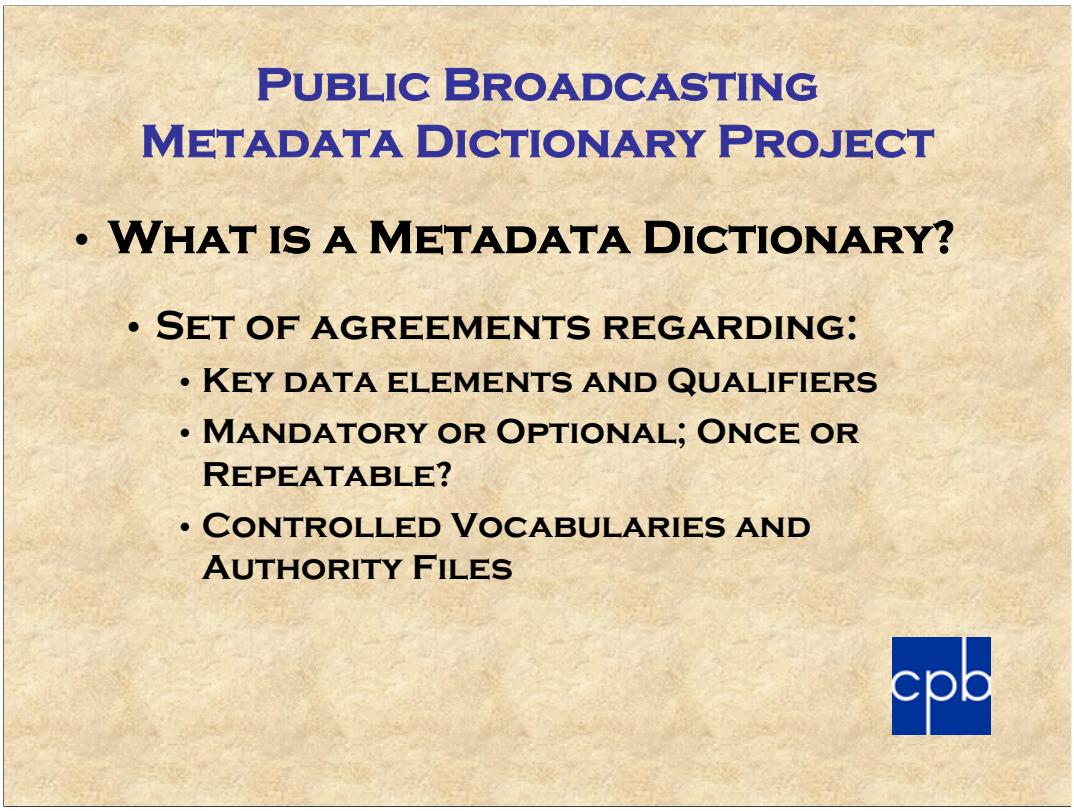
The point of adding metadata is to make the asset knowable, findable and ultimately moveable.

Here are some examples of metadata that might typically be associated with a television program, or pieces of a television program. As you can see, it's all pretty familiar stuff.

Keep in mind, however, that there are innumerable ways to express what looks like basic information, especially when you factor in the many organizations and individuals it takes to produce and distribute a single piece of television content.

As we contemplate a future with even more viable distribution platforms then we already have, with smaller and smaller audiences, and with personalized content delivered on demand, it becomes clear that we'll need to know more about our assets and to recognize what might make them attractive and useful in their various forms. We can't stay in business if we don't think this way.

So lots of people in the media industry, and many in the educational community have taken cues from their local library, and have started developing standards for expressing this basic, important information about their content – a sort of Dewey Decimal System.



In public broadcasting, we've decided to engage in a similar effort, and CPB is supporting the Public Broadcasting Metadata Dictionary Project.

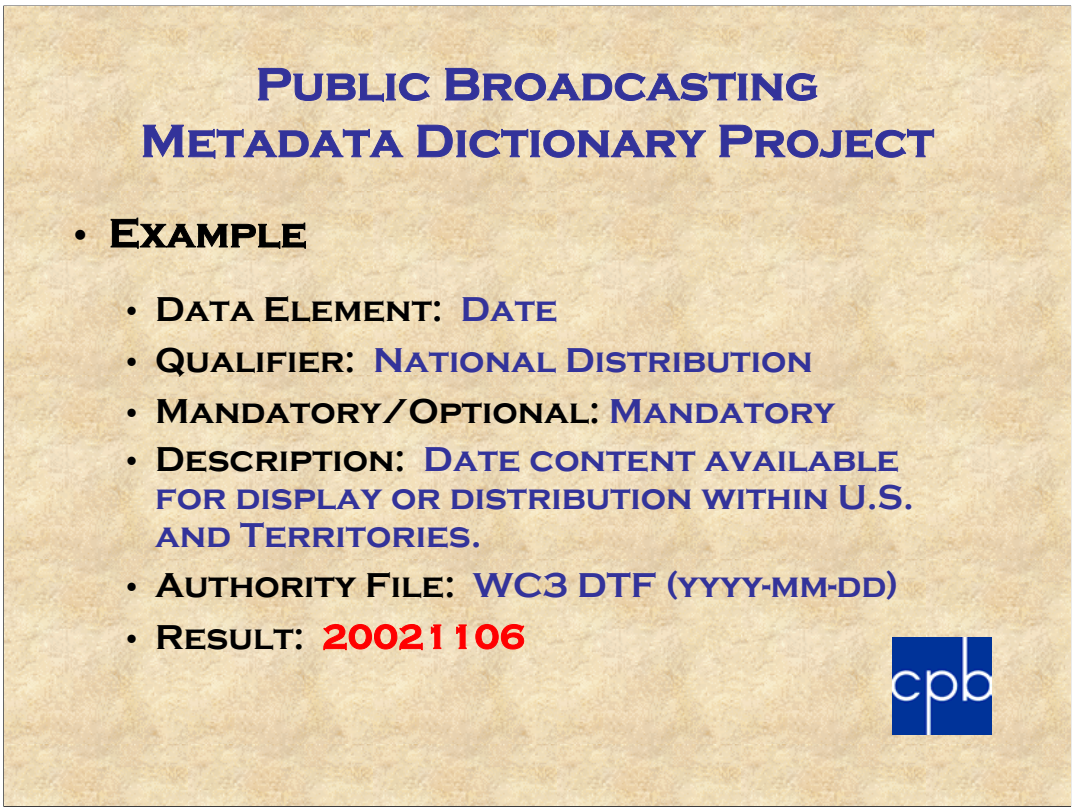
What is a metadata dictionary? Simply put, it's a set of protocols – human readable and machine readable.

First, it stipulates which pieces of data about an asset are most important to us and to our customers. What are the key data elements?

Then, it addresses whether these pieces of data need to be "qualified" or made more specific. (For example, when we say title, do we mean series or episode?)

A dictionary also stipulates whether the information is mandatory or optional within a given system.

Finally, it establishes protocols for exactly how the information is expressed. Do you use an alpha-numeric in the field? Is the term you enter chosen from a pull-down list, or controlled vocabulary? Is there an authority file that established the format of the information?

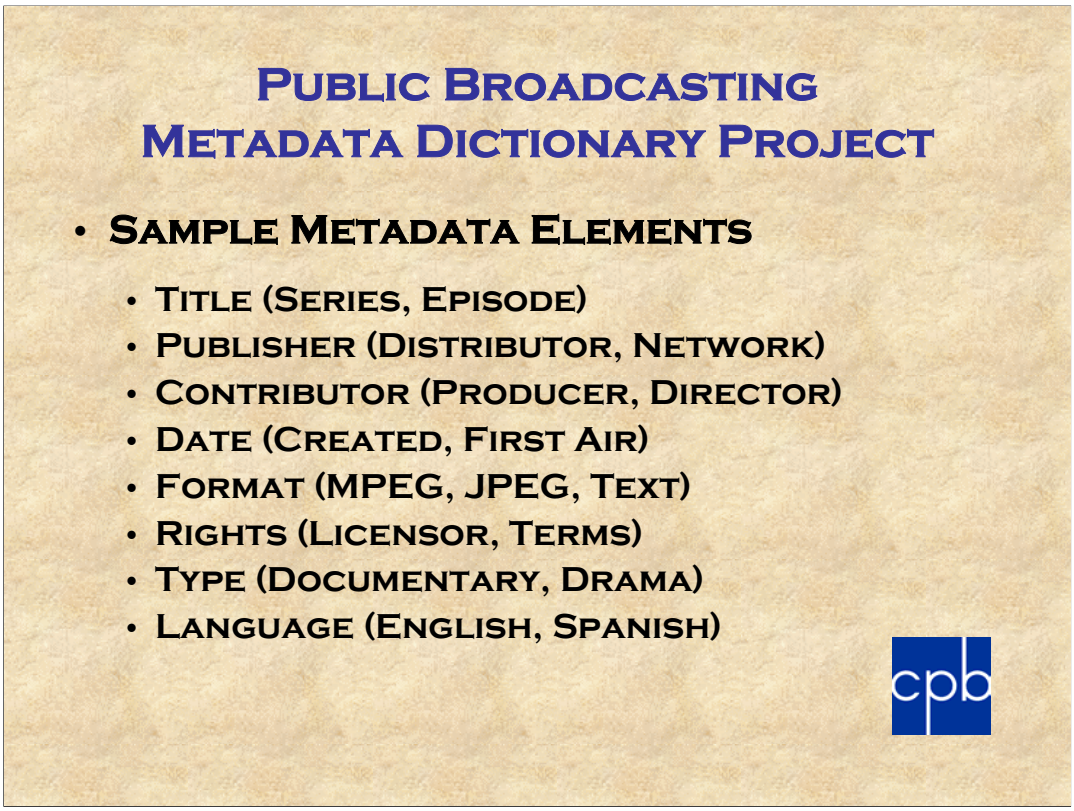


Let me quickly run through an example of one metadata element…in this case DATE.

If you've worked with producers from Great Britain, you've probably noticed that things can get confused regarding date. Today's date, November 6, would be expressed on a BBC tape label as 6 slash 11, or the 6th of November. As any good American producer knows, 6 slash 11 is June 11th, not November 6th!

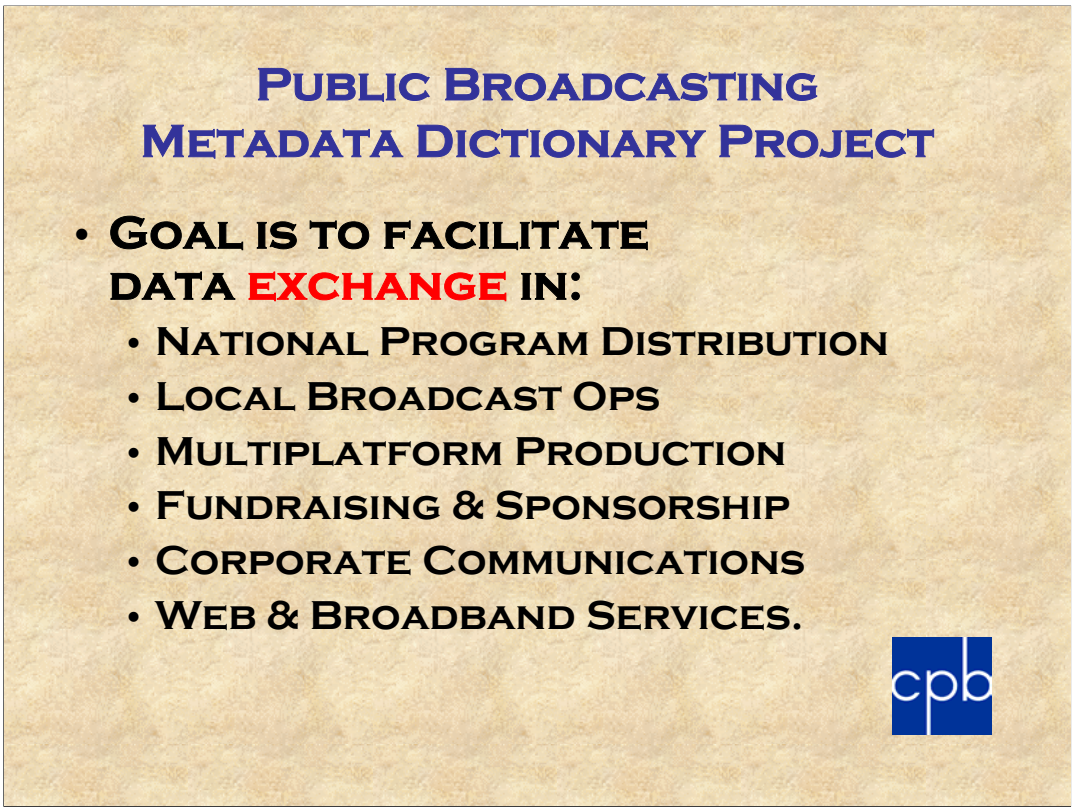
On the slide you can see how this critical piece of information would be entered into an asset management system, if one were using an metadata dictionary or shared protocol.

In this case, the chosen authority file, which dictates how the information is to be expressed, is the World Wide Web Consortium's Date-Time-Format. It stipulates that a date is year-year-year-year-month-month-day-day. Leaving us with the very exact expression of date that you see at the bottom of the screen. No more cross-Atlantic confusion.



Hopefully, you can imagine from this example what a Metadata Dictionary would consist of – 20 or 30 key data elements and qualifiers, and the rules for entering that data.

Luckily, most of you, unless you're building your own Digital Asset Management system or Archival database from scratch, won't ever have to convert a Metadata Dictionary into computer code or into a user interface. As you'll see in both Grace's and Dave's presentations, it's likely that someone else will do that that for you.

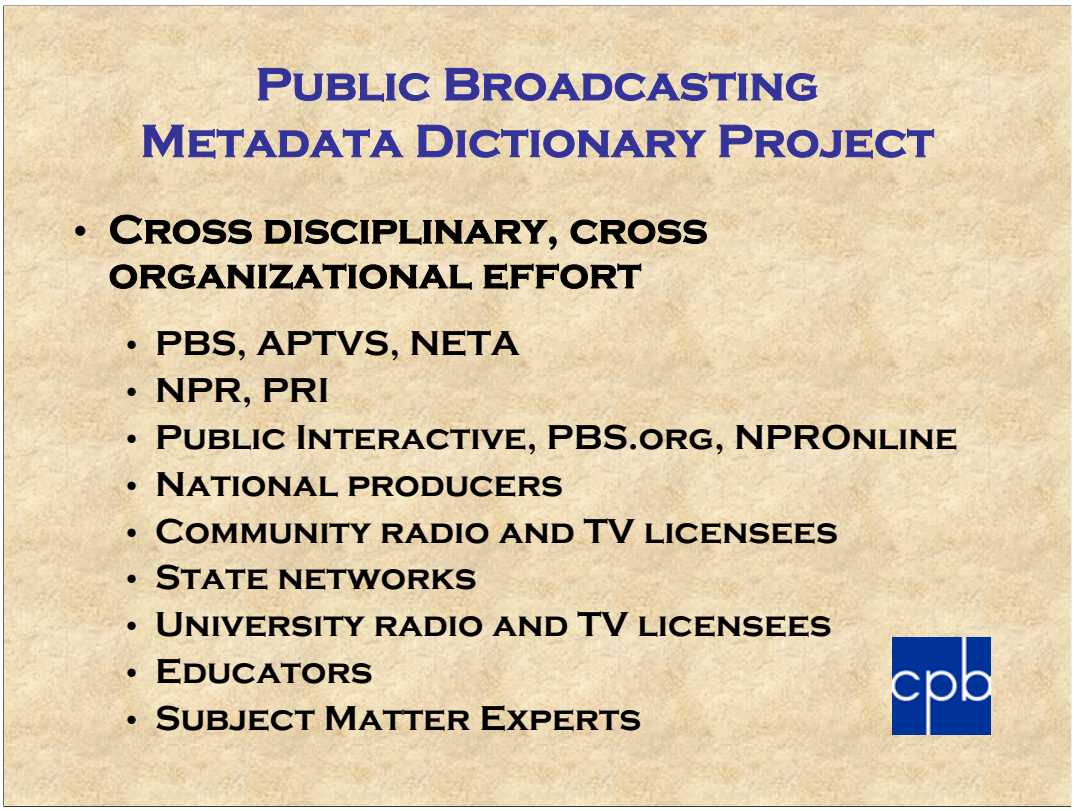


You do need to understand the point of it all – how you and your organizations are helped by complying with this kind of standard.

The goal of a Metadata Dictionary is to facilitate the exchange of data, both between organizations, like you and PBS, or you and your network, and within organizations.

The simple fact is that the more alike your data is across the various systems that use it, whether it's an Avid, or a PSIP inserter, or a sponsorship client manager, the more efficient and successful you'll become. Operationally speaking, "translating" data, especially manually, by re-keying it into different software systems, wastes time and costs money.

From a business standpoint, a badly managed flow of information about your editorial assets can cause you to miss opportunities to derive value from them. If a consumer, for example, is searching for a particular kind of program on their local Video on Demand service, or EPG, and you haven't figured out how to attach or associate genre metadata to your shows, you have failed – you've lost a customer or annoyed a potential donor.

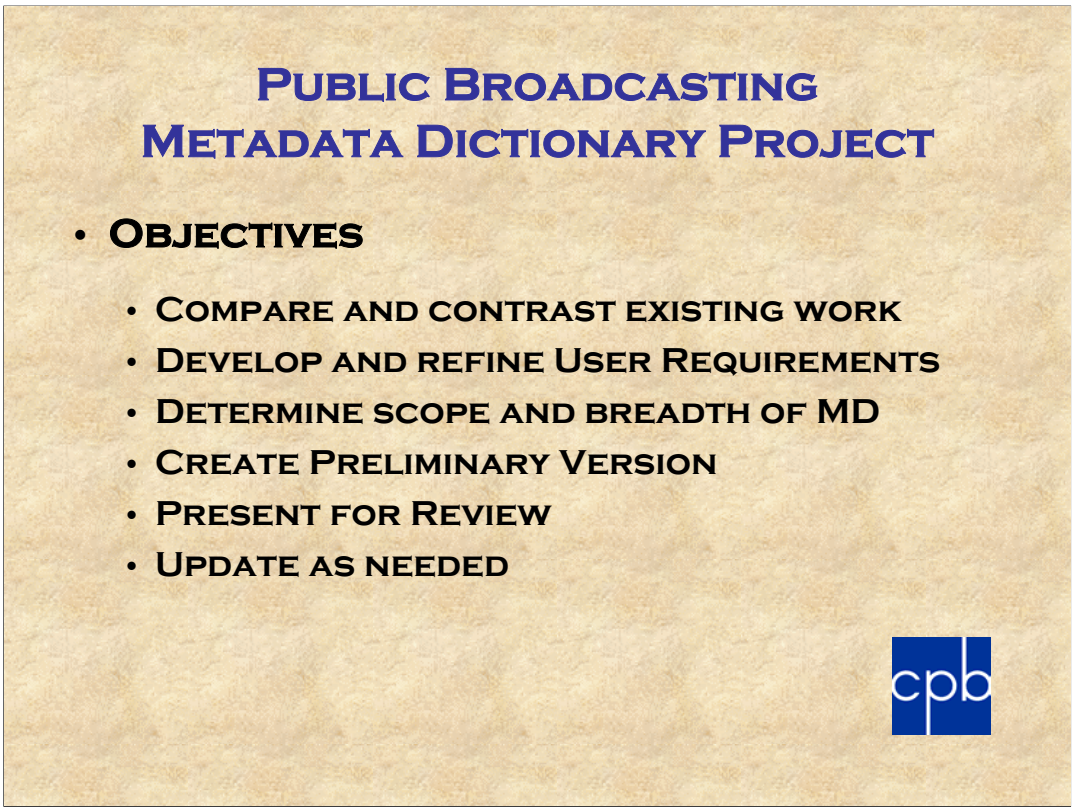


So here we are, public broadcasters, believing that agreement on metadata will make us better at serving our constituents, and much more efficient in executing our internal operations. Now what?

Well, if you're in public broadcasting, you know how complicated and democratic our system is, and how difficult it is for us to come to agreement on anything.

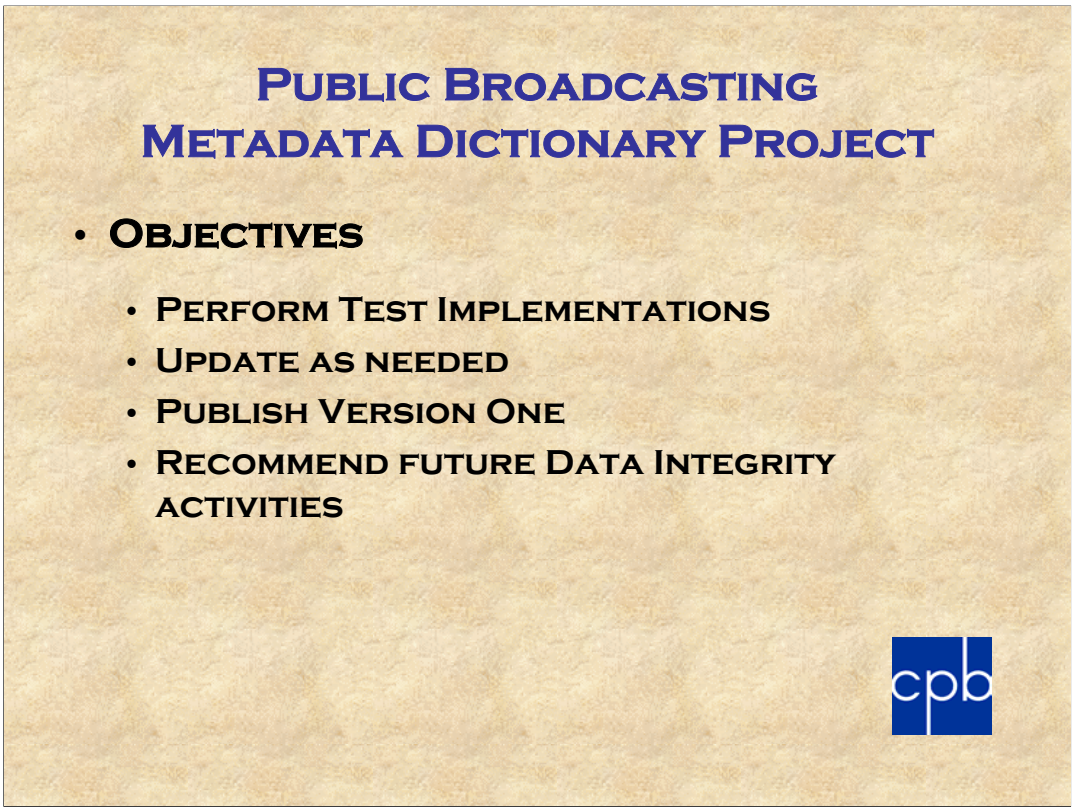
We knew that we'd have to get lots of representation from the right people if we were ever going to come to any kind of consensus.

So CPB, working with project administrator WGBH, assembled a working group made up of individuals representing multiple organizations and disciplines. Perhaps more importantly, these individuals really got down with this metadata stuff – they already understood it for the most part, and were very much in touch with their inner librarians.



When we got this group of metadata mavens together last April, and again in September, here's what they established as their to-do list:

- First, compare and contrast applicable metadata work already done by other standards groups, such MPEG-7 and Dublin Core, and by public broadcasters themselves. (KUED, WGBH, MPR, PBS and NPR have all done extensive metadata work, and we wanted to take advantage of it.)
- Examine who uses our content, how, and why.
- Determine how simple or complicated the Dictionary should be – how extensive.
- Distill (arm wrestle, really) all this thinking into the creation of a Preliminary Version of the Dictionary.
- Present that for broad review by public broadcasters and others, including vendors, and update it as required.



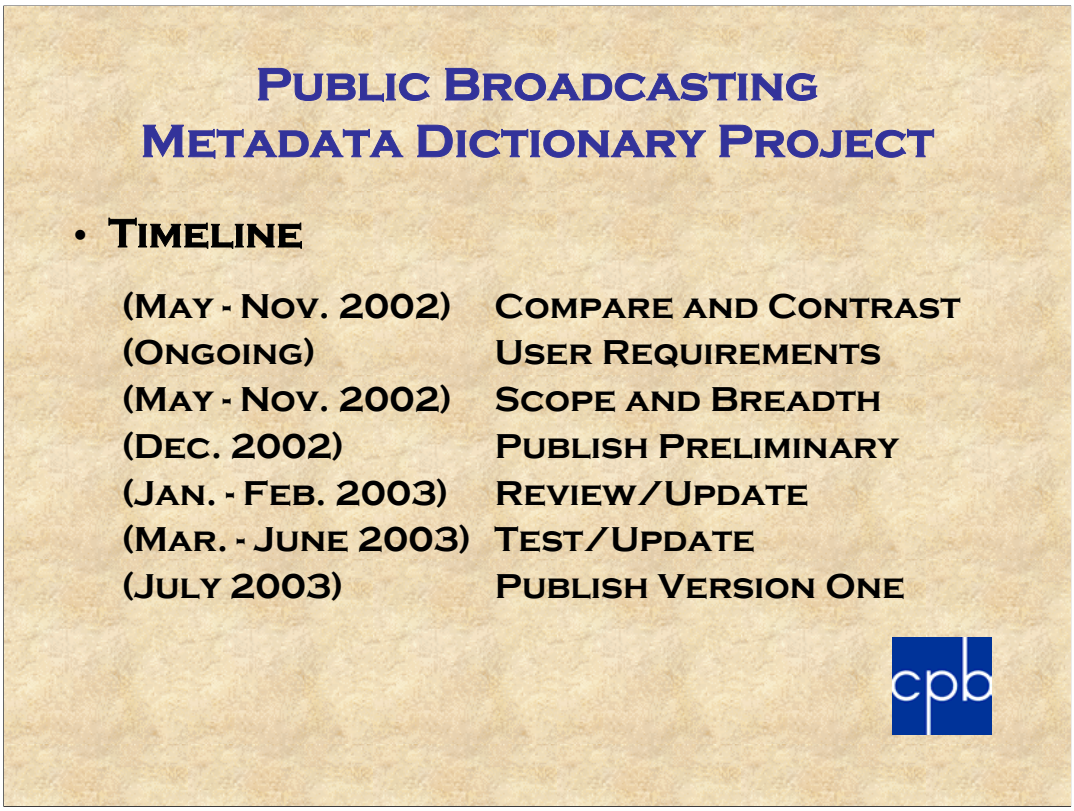
Then, actually test the Dictionary in real life settings, using it, for example, in a new project to create educational assets, or to update a previously archived collection.

Again, revise the dictionary as indicated by the testing results.

Then, publish Version One for the free use of all public broadcasters and their partners.

Finally, the group was convinced early on by one of our subject matter experts, Grace Agnew, that all this work wouldn't be worth much if we didn't have some kind of plan for maintaining the Dictionary over time.

The working group's last act will be to make recommendations about how public broadcasting should handle the matter of data, and who's the boss of it. It promises to be another difficult and complex conversation.



To give you a sense of what's been involved, and what's planned, here's the timeline for the metadata project and the working group's activities.

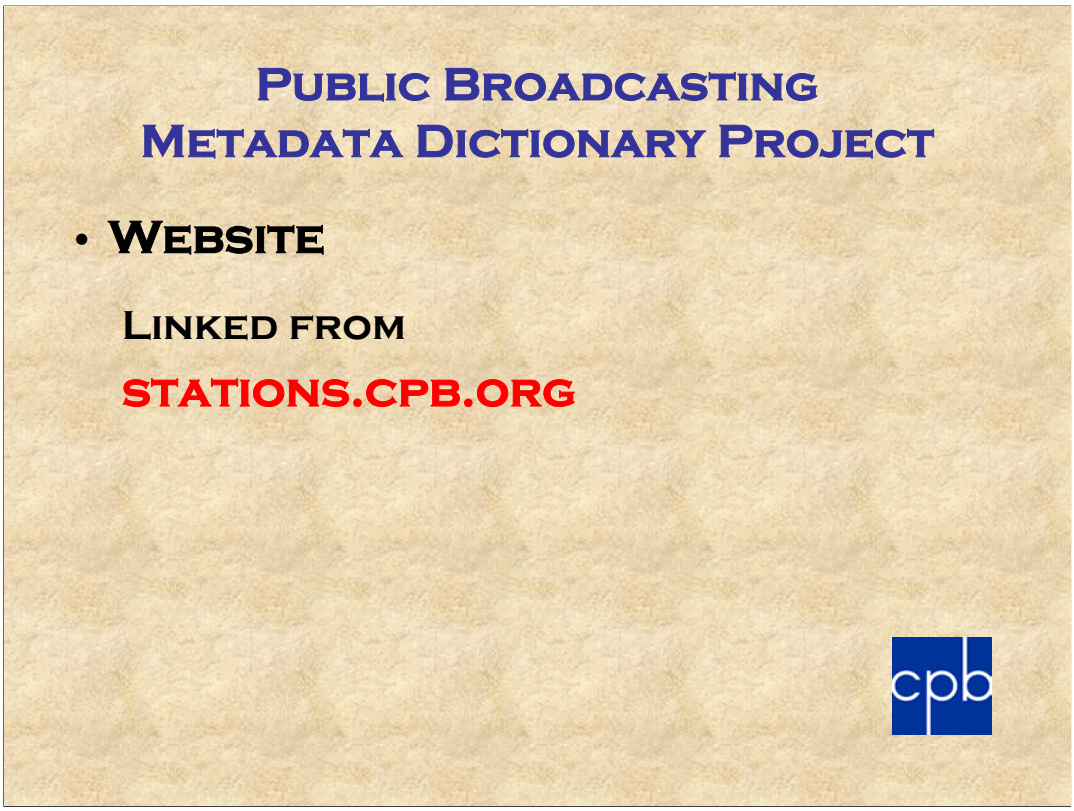
•May through November, the group was performing the difficult task of comparing and contrasting all the previous metadata work, and determining how deep and wide the Dictionary had to me. This effort culminated in a meeting in Boston in late October, which the participants dubbed the "Beantown Smackdown" because of all the wrestling over these labels.

•What we call things is very emotional and very contextual; it varies from individual to individual and organization to organization.

•The user requirements process is ongoing, and will continue throughout the process.

•The Dictionary will be reviewed early next year, and updated, then tested and revised between March and June.

•Finally, we hope to publish Version One in July.



If you're a public broadcaster, and want to follow the progress of the group, or contribute to the review or testing process, we have a website where you can read all the public documents, and make comments.

It's most easily reached via the homepage of stations.cpb.org.

Now, I'm going to turn to Grace Agnew, of Rutgers University.

She's going to describe a project that I believe will give you a sense of the glorious possibilities of managed assets and applied metadata protocols.

DAM EVANGELISTS:
The Gospel of Digital Asset Management

IOWA DTV SYMPOSIUM 2002

Alison M. White, CPB
Grace Agnew, Rutgers University
Dave MacCarn, WGBH