



To: Working Group – Public Broadcasting Metadata
Dictionary Project

From: Bill Keens

Date: April 26, 2002

Re: Facilitator's Memo – First Working Group Meeting

The purpose of this memo is to describe the principal areas of agreement and key outcomes of the first meeting of the Public Broadcasting Metadata Project Working Group, which took place on April 24-25, 2002, in Arlington, VA.

WEDNESDAY, APRIL 24

Welcome and Opening Remarks

Andy Russell, Senior Vice President for Media at the Corporation for Public Broadcasting, welcomed the Working Group members and described the meeting as “the result of years of work.” The refrain for years throughout the public broadcasting system has been the need for better, more flexible digital tools and digital services that draw on the tremendous assets of public broadcasters. The first sticking point on the way to this goal is metadata structure, a commonly agreed upon language for those assets. That’s where the work of this group comes in, and why its long-term impact is so important.

Both Alison White of CPB and Amy Rantanen of WGBH built on Russell’s welcome by describing the momentum behind this project and the purposes of this meeting.

Alison described the process of pulling together a small group of people when she arrived at CPB to start working on asset management issues and to provide some leadership around the system. Their work together gathered momentum with the CPB Asset Management Caucus held in Chicago in 2001, where everyone acknowledged the cross-station sense of urgency that transcended their individual interests. That meeting produced remarkable consensus on the to-do list. The first three items on that list were the development of shared standards; the

development of business models for asset management; and the development of prototypes.

WGBH has been part of this process, White said, and their passion has been an engine for its continued progress. Now, with the Metadata Working Group assembled today, CPB is trying to get a group together that is small enough to make decisions but big enough to be representative of the public broadcasting system, both radio and television. The people in the Working Group all have a passion for this work – an “inner librarian.” A lot of information will be developed and shared over the next two days, but the purpose of this meeting is not just to inform the participants. “We need to figure out whether public broadcasters can and should develop a shared metadata standard, and if so, how to begin to move toward that outcome.”

Amy Rantanen echoed Alison White’s remarks. “We want these two days to be productive, to build consensus and gain momentum,” she said. WGBH is working on asset management, but so are others, and CPB realizes the importance of inter-operability across all of these projects, including those underway and those still on the horizon. We have a responsibility to work together and with CPB, and to think of the field’s needs and benefits, not just our own. At its best, the Working Group will come up not only with a metadata dictionary but also with guidance as to how to deal with the projects already in the pipeline or in planning. We need to ensure success across the system, where increasingly we are all in what are almost forced collaborations, triggered by the growing urgency of our asset management needs and our resource constraints. To that end, WGBH wants to share what it knows, and is highly committed, like CPB and others, to this level of collaboration. “Let’s articulate and address our concerns together, document the issues and challenges before us, and decide as a group how best to proceed. WGBH doesn’t have all the answers, but we are making headway answering our many questions one at a time.”

Overview and Discussion

Discussion began with a comprehensive overview of the elements related to metadata and the creation of a “common language,” and included a glossary of metadata terms as a handout. (Note: this glossary will help the Working Group use consistent terminology in its efforts to educate the public broadcasting system about the project.) Working Group participants then identified issues that need to be addressed in the course of developing a single metadata dictionary. Those issues were classified as either “high priority” or “moderate priority” by the participants; not surprisingly, almost every issue had a sense of urgency about it if the aim is to create a metadata dictionary in the near-term future.

High Priority Issues

The high priority issues and the six “clusters” into which participants placed them were as follows:

- Limit the scope of the metadata dictionary project – don't get paralyzed by getting beyond our scope.
- Why Dublin Core? Let's review other standards and decide which is best for broadcast media.
- Keep track of the changes in the industry and their effect on the public broadcasting metadata dictionary.
- Create authorities for public broadcasting.
- Use standards that work inside the broadcast industry but are compatible with print and still media as well as other content or essence.
- Make specific recommendations (to the system) regarding minimum metadata standards that all stations, producers, and distributors should adhere to.
- Agree on setting a standard dictionary.
- Ideally, create a metadata dictionary that allows all licensees to use just the parts they need.
- Adopt Qualified Dublin Core – adopt standards and processes for industry-specific extensions.
- Ensure inter-operability with system projects, both those underway and those in the pipeline.
- Have the ability to share content seamlessly across multiple platforms.
- Focus more on metadata exchange among stations and other users rather than on internal full metadata record definition.
- Model the community – scenarios include administrative, educational, and other uses: who are the primary user-groups of the metadata?
- Develop a core schema.
- Work on specific audio-video schemas.
- Integrate LOM into our schema (IEEE, Adept, etc.)
- Agree on a basic set of inter-operable terms in a flexible schema.

Moderate Priority Issues

Issues labeled as less urgent (but, in the long-term, no less demanding of attention) were grouped in two clusters, as follows:

- Address the management of data integrity, beyond the creation of a metadata dictionary.

- Balance comprehensiveness and ease of access.
- Address the extensibility of the metadata dictionary to state education standards across all states.
- Think about the usability of the metadata dictionary: from the perspective of those who enter metadata, regardless of scheme; from the perspective of those who must manage the integrity of the databases; from the perspective of the end-user, who must successfully search for and retrieve wanted assets.

Success Indicators (2002)

The Working Group identified several success indicators that could be realized this year and that can be used to benchmark the progress of the metadata dictionary project. Participants also indicated whether they believed these various indicators were within the power of the Working Group itself to deliver.

- System managers would know early in the process whether their systems are compatible with the metadata dictionary being developed, and if not, what they need to do to become compliant. (*Working Group*)
- We have a process for how metadata work is done at the local level. (*Working Group*, to the extent that it may recommend an approach to training or provide other guidance)
- There is an actual exchange of data to test and confirm the metadata dictionary. (*Working Group*)
- We have tested the compatibility of the metadata dictionary with systems either in place or in the pipeline. (*Working Group*)
- The metadata dictionary has achieved vendor recognition and acceptance.
- Public broadcasting has emerged as an active, major player in the development and dissemination of metadata standards. (*Working Group*)
- There is no question about who the “owner” of the metadata dictionary is – who has responsibility and authority for it – and the process for maintaining that dictionary is well-defined. (*Working Group*)
- The number of users of the new public broadcasting dictionary is substantial and growing.
- New asset management business models are being developed as critical mass in the dissemination and use of the metadata dictionary is achieved.

The Foundation for Consensus

The balance of this session was devoted to recapping key points of agreement achieved during the discussion, on which further consensus might be built:

- *Agreed* – that a public broadcasting metadata dictionary is needed, and that public broadcasters will have to be compelled to use it in some way in order for it to become commonplace.
- *Agreed* – that controlled vocabularies are essential.
- *Agreed* – that the case for a common metadata dictionary will be stronger if it is framed by a problem statement that explains *why* it is needed.
- *Agreed* – that the public broadcasting metadata dictionary has to reflect the various constituencies who will use it.
- *Agreed* – that we need a dictionary that is modular and layered, and that supports inter-operability and extensibility.
- *Agreed* – that in developing a metadata dictionary we should consider what already exists and “take the best and leave the rest.”
- *Agreed* – that we need to map against existing standards.
- *Agreed* – that we should focus on the exchange of metadata rather than internal system keeping as we develop our dictionary.
- *Agreed* – that the dictionary has to help access, use, and reuse material.
- *Agreed* – that we need a central resource for training in the use of the metadata dictionary and maintenance of metadata.

All of these lists are essentially preliminary inventories of the issues, attributes and expectations that people have regarding the development of a shared metadata dictionary. As such, each of these lists will likely be expanded and amended as the thinking of the Working Group and the field unfolds.

Five Presentations – and Issues Revisited

In order to ensure that the Working Group was fully aware of existing efforts in the development and management of public broadcasting metadata, five members of the group were asked to talk about the metadata needs their organizations have and the steps they have taken to address them. Each presentation was illustrated by computer-generated slide shows, database screens, and other visuals, and each prompted questions and discussion. The organizations making presentations were:

- Public Broadcasting Service “Orion Project”
- National Public Radio “Content Depot” and other distribution initiatives
- KUED/University of Utah “Content Asset Management System”
- Minnesota Public Radio “Digital Archive Initiative”
- WGBH Educational Foundation

The Working Group agreed that the substance of these presentations was impressive and the time spent reviewing them worthwhile. Participants also raised a few final questions that helped set the stage for the next day:

- How can we make our metadata dictionary even more “robust” than those presented here? This robustness is likely to require more attention to detail and more steps to ensure that the dictionary doesn’t fall short of expectations a few years down the road.
- How can we add logic into the *search* as users access our metadata dictionary, rather than expecting the dictionary itself to carry all the burden?
- How can we learn about and draw on additional metadata dictionaries, such as the one used by the United States military (based on IEEE) and others?
- Our notion of “core” seems to be getting smaller and smaller. What should we do about that, if anything?
- How do we anticipate the different metadata demands of three agents that will act on our metadata dictionary: 1) the end users; 2) producers, stations, and other intermediaries; and 3) change itself – because data is not static?

THURSDAY, APRIL 25

Review and Comment

Day two began with a brief review of the agreed-upon issues, success indicators, and points of consensus from the prior day’s discussions. In addition to the items above, the Working Group added the following:

- The enforcement of intellectual property rights is beyond the immediate scope of this Working Group, but the sharing of rights-related information is within bounds.

- Our consensus needs to emphasize that what we do is ultimately all about access by others within and outside of the public broadcasting system – to the extent that the dictionary we develop allows access to the metadata (but not the appropriation of content).
- The public broadcasting metadata dictionary has to be in harmony with knowledge management systems, which are all merging; we don't want to be left out of that movement.
- Our task is to make the *content* of the metadata dictionary enticing, while leaving the ease-of-use of the interface to others.
- Station buy-in will be essential to the widespread use of the metadata dictionary – and the centralized training resource described above will be a key part of teaching station staff to use the dictionary in order to achieve persuasive efficiencies and return on investment.

Charge to the Breakout Groups

The Working Group was divided into three smaller breakout groups of about seven people each to tackle the following questions, listed here in priority order:

- What tasks need to be completed – and by when – to establish a common metadata dictionary?
- What can the Working Group do – as individuals, teams, or all together – to accomplish those tasks?
- What must others be recruited to do?

Breakout Group Reports

Group One

Marty Bloss, NPR

Paul Burrows, KUED

Efthimis Efthimidias, University of Washington

Bea Morse, PBS

Amy Rantanen, WGBH

James Steinbach, WPT

Alison White, CPB

Group One constructed a planning grid that divided the work to be done into four phases: 1) Requirements; 2) Analysis; 3) Implementation; and 4) Testing, Modification, and Deployment. Each of these phases were listed vertically down the left side of the grid, and each was broken into subtasks. Across the top of the grid were categories for the Due Date for each phase of work, as well as who is

tasked with doing that work – the full group, subject matter experts who are a subset of the full group, subcommittees, project managers (CPB and WGBH, presumably), and other. The most extended due date, for phase 4 above, was April 30, 2003 – with the metadata dictionary actually built or adopted by December 31, 2002. The grid itself follows:

TASKS	DUE DATE	WHO				
REQUIREMENTS		FULL GROUP	SME'S *	SUBJECT COMMITTEE	PROJECT MANAGERS	OTHER
Mission/Chart/Scope <ul style="list-style-type: none"> ➤ Input ➤ Compile/Suggest ➤ Review/Comment ➤ Finalize/Publish 		X X		X	X	
Define Constituencies		X				
Gather and Review Existing Use Cases			X			
Gather and Review Lifecycle of Broadcast Assets and Workflow Processes		X REVIEW	X GATHER	X CREATE REPORT		
Generate Requirements <ul style="list-style-type: none"> ➤ Determine elements, layers, modules 	AUG 1	X REVIEW		X CREATE		
ANALYSIS						
Review Existing Metadata Dictionaries and Determine Applicability		X REVIEW	X ANALYZE & PRESENT			
Choose: <ul style="list-style-type: none"> ➤ What to Adopt ➤ What to Build 		X REVIEW & APPROVE		X RECOMMEND		
Reassess <ul style="list-style-type: none"> ➤ Scope, Subcommittees, Outside Experts 	OCT 1					
IMPLEMENTATION						
Adopt/Build Determine: <ul style="list-style-type: none"> ➤ Elements ➤ Qualifiers ➤ Values Controlled Vocabularies	DEC 31, 2002	X REVIEW	X ADVISE	X		MAYBE
Test Modify Deploy	APR 30, 2003	X	X	X	X	

* SME's – Subject Matter Experts

Group Two

Judy Brown, ADL CoLab+SCORM

Dennis Haarsager, CPB

Robin Mudge, OnCourse

Marilyn Pierce, PBS

Richard Ruotolo, PRI

Thom Shepard, WGBH

Tracy Vosburgh, WPSX

Group Two reported the tasks to be accomplished as a sequence of steps extending from the present through December 2002, as follows:

1. Internal communication – both virtual and face-to-face (by 5/31/02)
2. Basic assumptions – including a problem statement and explaining what we are doing and why (by 6/10/02)
3. Ongoing research – of other schemes and legacy data, informing...
4. ...Defining the end-users (by 7/15/02)
5. Information to date collated and presented – including projected impact and next steps, leading to...
6. ...Potential metadata dictionary(s) – including discussion and determination of the preferred dictionary (by 11/1/02)
7. Test the preferred dictionary – and exchange data (by 12/31/02)
8. Report on results and next steps

Throughout this sequence, two additional steps would be taken as early as possible and as often as necessary:

- Establish adequate funding
- Share lessons learned

What can the Working Group itself do? This breakout group's report defined a number of specific tasks:

- Capture what we already know
- Determine next steps
- Transfer knowledge to the station level
- Dedicate resources

- Report out – both to the field and to others, to build awareness both horizontally and vertically, to seek buy-in, and to identify expertise

What can others outside the Working Group provide? The group made the following suggestions:

- Fund the work to be done, including underwriting the people needed to do it
- Provide hardware, as vendors might
- Provide additional expertise

Group Three

Grace Agnew, AMIA

Alan Baker, MPR

Marcia Brooks, WGBH

Dave MacCarn, WGBH

Tim Olson, KCTS

Cate Twohill, PBS

Steven Vedro, WPT

Group Three began by asserting agreement within the group that communication with constituencies on all tasks, missions, objectives, and reports *are the responsibility of CPB*. The group then identified a number of the tasks needed to begin to establish a common metadata dictionary, including:

- Review existing public broadcasting systems and schemas
- Decide the scope and granularity of the common metadata dictionary project
- Create a “table” of priorities showing the big picture, then pick a subset on which to focus, including issues of content, intellectual property rights, instantiation; and addressing users like CPB, PBS, K-12 educators, and others
- Define who does the work – including physical meetings, information-sharing, website development, in-system and in-group communications, and review of existing research (controlled vocabularies, authority lists)
- Collaborate with other metadata bodies, including SCORM, the Library of Congress, AMIA
- Create a list of resources
- Build consensus among those who have established metadata systems, stations of various sizes and types, and producers

The breakout group also described key elements of implementing a metadata dictionary (taking as it’s motto, “pick the best and leave the rest”):

- Define the core working group

- Define the core terms and propose a list of fields
- Meet to draft the fields list and review with others its schema architecture, structural elements, and attributes
- Test the fields list with real data
- Justify field requests
- Compare against existing systems (Connect, Traffict, OnCourse)
- Implement the fields list and authorities list
- Designate data managers – the controlling enforcer (e.g., steering committee) – which should also be...
- ...the ongoing maintenance body, charged with registry maintenance, staffing, etc.

As to the role of the Working Group and others in accomplishing the work that lies ahead, the breakout group's suggestions included:

- Identify current projects as resources – such as MARS for its vocabulary and existing research; Connect for its information architecture and CMS; and Content Depot for its application to radio.
- Identify external funders – such as the NSF, NEH and DOE
- Identify internal funders – including CPB and PBS, who might also provide dedicated staff to the work ahead

Having reviewed these elements of Group Three's thinking, the group proposed both a structure and timeline for the work to be done:

- Form a framing and research subcommittee of the Working Group to review existing systems and research – the work of this subcommittee should get underway asap, and should be completed before the third and fourth subcommittees begin their work.
- Form a consensus-building and communications subcommittee to work over the course of the project, including clarifying basic assumptions.
- Form a subcommittee to draft a field list and create schema architecture, to develop vocabularies and a registry, and to vet and test its work.
- Form a subcommittee to take care of data maintenance and management.

With respect to the timeline for this work, the breakout group first proposed a schedule that would have concluded in May 2003, but subsequently revised this schedule to reflect the following deadlines:

- May 1 – steering committee gets underway; subcommittees assigned
- September 1 – complete the work of the framing and research subcommittee
- October 1 – complete the draft fields and schema architecture

- April 1 – vocabulary developed
- June 1 – vetting and testing completed
- November 1 – registry completed

Ongoing throughout this timeframe and beyond is the work of the consensus-building and communications subcommittee, and the work of the data maintenance and management subcommittee.

Implementation

The final session of the Working Group meeting was dedicated to the implementation process to come, with discussion organized around three questions:

- What timelines are other metadata projects on?
- When should this working group meet again, and what is its primary agenda?
- Do we need a steering committee? Who should serve on it and/or organize it? How will the project coordinating/management work get done?

The discussion roamed over all of these questions, but began with an inventory of other metadata projects currently underway and the dates by which completion is expected. This catalogue of initiatives – and the news that about 18 stations are ready to begin asset management projects – lent weight to the argument that the metadata dictionary project has to move ahead as quickly as is reasonably possible. The projects underway include:

- WGBH/Artesia Project – ongoing
- OnCourse – September 2002, but can adapt from the metadata dictionary project once it is complete (based on SCORM – uses IEEE/LOM)
- PRI – which will roll out its DAMI prototypes this spring
- Minnesota Public Radio – goes live in July 2002, but hopes to make modifications to come into compliance with a metadata dictionary (based on Dublin Core)
- NPR – Content Depot, with definitions in the summer of 2002 and implementation in early 2003; also, a Media Asset Management project is on the same timeline
- MARS (KCTS, et al) – has a similar timeline, also with early 2003 implementation
- Rich Media Utility Project (Wisconsin PTV, et al) – a test of workflow and use with start-up in September 2002
- Ridgway – a project local to WPSX/Penn State that rolls out in May 2002
- PBS – has several projects underway, including: Orion, which goes live on May 28, 2002; a NOLA evaluation thereafter, to bring NOLA into Orion by September 2002; Connect, which is implemented in January

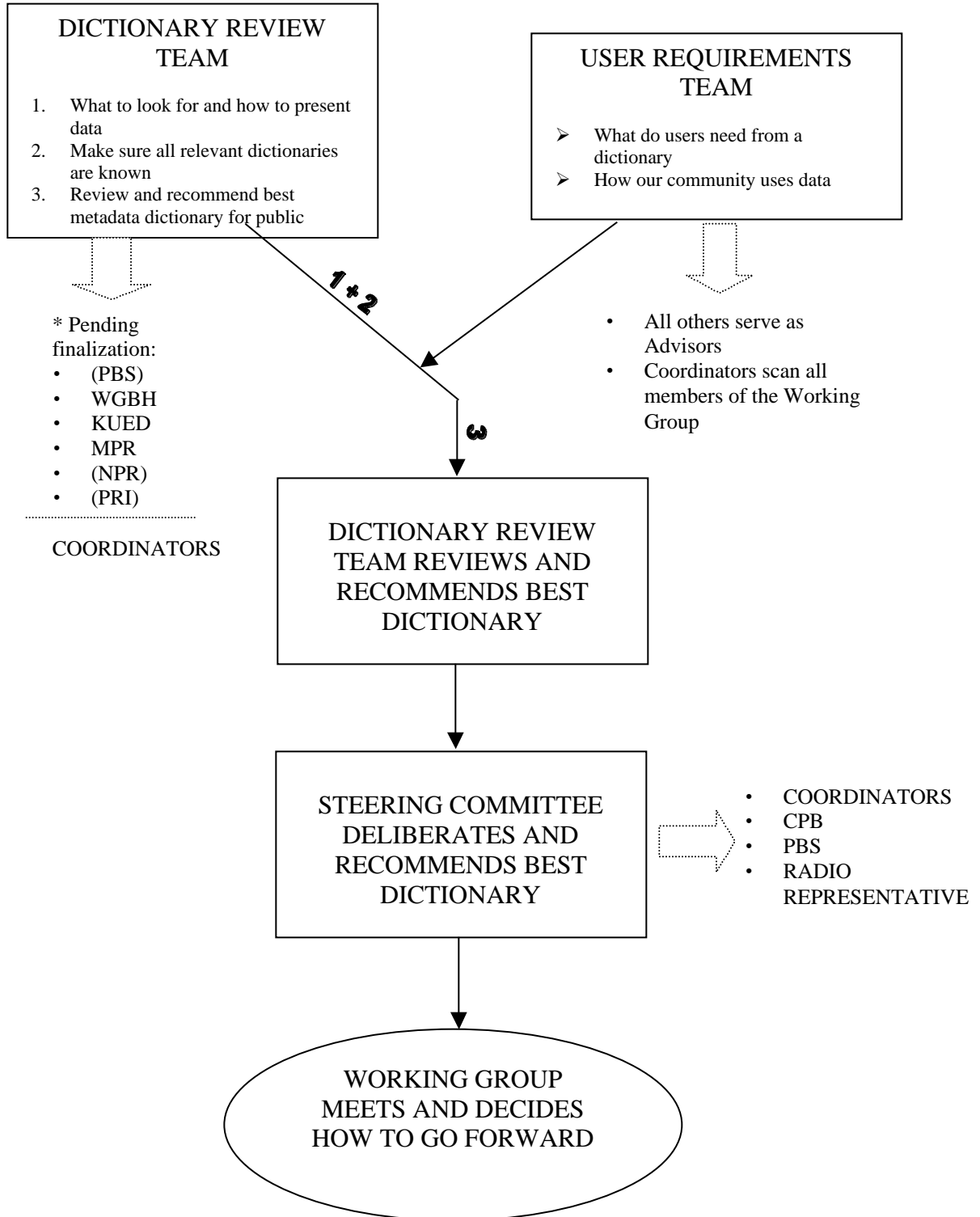
2003; and the Archives Preservation Project, which can be adapted to match the timeline for the metadata dictionary project

In light of the deadlines for these various metadata projects, the Working Group grappled with the development of a “reasonable but ambitious” timeline, and concluded that:

- There is a great deal of existing work in the field to build on (see the list above); the metadata dictionary project does not have to start from scratch.
- The critical work now is twofold: a) review current dictionaries in place, first establishing *what* we want to look for and then making sure all significant dictionaries are taken into account; and b) get a real feeling for the data dictionary needs of users, so that when a dictionary is ultimately developed, it won’t backfire.
- Based on that research and work, the Working Group should then be in a position to “pick the best and leave the rest” – essentially adopting and adapting a dictionary that is likely to have “a high percentage of what we need” at the outset.
- In the near term, if the Working Group hopes to influence and benefit the many other projects that are rolling out around this fall, it should develop and disseminate a “provisional core” by Labor Day of this year.

With Labor Day as the target date for the provisional core, participants were able to work backwards and, in the process, to construct a flow chart for key steps and tasks that concludes with a second meeting of the Working Group in late summer.

*Note: the teams indicated in the following flow chart are pending finalization; phone calls will ensue to elicit a better sense of each group’s project status.



As this flow chart makes clear, the work of two nearly parallel teams – each supported by a paid coordinator/facilitator – will culminate in a preliminary recommendation as to the best dictionary to adopt/adapt. That recommendation will be made to a still-to-be-constituted steering committee. The steering committee will deliberate and, in light of what it has learned, will make its recommendation to the Working Group as a whole. The Working Group will convene to discuss the steering committee’s recommendation and, finally, to map out the implementation steps that should follow, including the roles and responsibilities of all involved.