

Meeting of NDIIPP Public Television Project
WGBH, Boston
August 19, 2005

Attending for the whole meeting: Nan Rubin (Thirteen), Dave MacCarn (WGBH), Mike Boeglin (Thirteen), Ken Devine (Thirteen), Jim Kutzner (PBS), Jeff Bary (NYU), Carl Fleischhauer (LC, drafted these notes); attending for about one hour: David B. Liroff (WGBH Vice President and Chief Technology Officer)

This meeting was intended to assess the project's status, at the end of year one, regarding the investigative track called "Formats and Wrappers." There was a round of intros; the new face was Jeff Bary, part of the Arts & Technology group at NYU (may have the unit's name wrong). The larger group of which this is a part has been reorganized and, as of a few days ago, Bary works for David Ackerman in the library, asked by David to join "since I [Jeff] know something about video." It sounded like Bary was a placeholder with our NDIIPP project, attending so as to update Ackerman. The hiring of someone to replace Jerry McDonough is expected soon and Bary is filling in.

Nan started the status discussion noting that there have been delays because of circumstances at NYU, both Jerry's departure and also the reorg. Dave MacCarn talked about progress thus far on formats and wrappers: discussion, not decisions. He had hoped for NYU to contribute more by now, especially re: packaging and related matters, this would have been Jerry's contribution. Later in the meeting, Nan extracted a promise from Dave to write a report on findings to this point re: formats and wrappers. This would represent the project deliverable that marks the conclusion of this set of facilitated meetings. Nan talked in terms of "by the end of the year" but Carl interjected a hope for earlier, and Dave said he would try to do something sooner than end-of-the-year. In the end, Nan's formula for this was "by about Thanksgiving." Dave agreed that he could do this.

Later in the day, Nan also observed that she foresaw an early November meeting with the NYU people--after the reorganization had been completed and new people hired and the school year rolling forward--to catch up and restart the NYU elements in the project.

There was some talk about the concept of *packaging*, somewhat aimed at filling Jeff in. Dave explained about television production and things at WGBH. Carl talked about the OAIS model and how there is a succession of packages in that concept. Also about the hope that NYU participation would bring digital library people (and OAIS thinking) into a joint effort with people in the TV community, where digital asset management systems sort of fill the niche of the repository. Ken concurred that this meeting of communities was a desired outcome of the project.

Ken talked about his own interest in finding the points of synergy with the LC Culpeper planning effort. He described the meeting with the LC group, and subsequently with Sam Gustman. He said that he and Sam shared the idea that there was a natural role for LC to exercise leadership for formats and packaging, on the national stage. Some of Ken's recent conversations, however,

suggested that as the Culpeper group proceeds to think thru these issues, and makes or defers decisions about them, this might be moving slowly. Thus Ken was uncertain about how and to what degree working with the Library in the near term would play out. He and Greg Lukow will meet at the FIAT conference in NY in September to discuss this, and also discuss more general matters of synergy between this NDIIPP project and Culpeper development.

This led to a discussion about “ultimate” content handoffs to LC, by the project but also (extra-project) by PBS in their continuing relationship with LC. In an offhand comment, Ken wondered if PBS might not be advised to continue to go to videotape while the LC process of decision-making moves forward. Carl said no, that LC will be ready to receive file based content (although the Public TV folks might have to determine the specifications on their own, consulting with LC as they do so). Jim also said no: PBS will discontinue the use of videotape, period.

Ken said that Sam Gustman had expressed an interest in having a meeting to develop an approach for formats and wrappers, but Ken was uncertain if that was still in the cards. Such meeting would include folks like Brad Gilmer (AAF format), Merrill Weiss (SMPTE), and others to talk about packages and wrappers.

Carl wondered about the option of defining SIPs in OAIS terms, which could be done independently from sorting out a repository design. In effect, keep is simple, and only define an exchange package, like the idea that motivated MXF development. Dave MacCarn said that he could understand this simple outcome of package design but that he felt the discussion of formats and wrappers should connect to repository development. But there was no necessity, he said, that the project’s investigation of this had to connect to LC. Mike Boeglin, as an IT specialist, recommended keeping the investigation of repositories and formats separate, they need to be independent of each other. All of these comments were, in fact, much more in step with each other than is conveyed by this typed rendering.

This discussion continued, highlighting various pros and cons, and talking about who will do what. The trend of the conversation ended up with a bit of a retreat from the idea of the project connecting to LC’s Culpeper development in a significant way.

There was some talk about the nature of a “package.” Carl mentioned preservation metadata (e.g., PREMIS), and Dave talked about making a UPF-like package, noting that the Universal Preservation Format called for the inclusion of such things as the source code for codecs to permit future rendering. [“I wanted to avoid dealing with a registry at all costs,” Dave added.] What was alluded to here, of course, be a submission package. The pace of the NDIIPP project may be fast enough to mean that the Public TV group develops a structure for a formatted package and presents it to LC as a recommendation for incoming television content.

Jim and Dave and Ken all talked about PBS handoff issues. Jim was concerned to determine which version would actually go to LC, at which quality, packaged how, with what metadata. Knowing the desired outcome for LC, Jim said, would make it easier for PBS to do a good job with designing the NGIS system (Next Generation Interconnection System). Carl said that he

would not be surprised if LC favors the higher quality 50 Mb/s Production Files discussed by PBS (as compared to the 8 Mb/s Distribution Files).

Carl went on to ask if there was and value in LC having a copy of the Production Files, would this serve as a safety backup for PBS and support its own preservation goals? Jim was uncertain as to the answer. He mentioned a backup system in Nebraska, but this seemed to be for Distribution Files. He also reminded Carl that the Production Files are not the most fully packaged final version, and may miss some of the underwriting credits that are added for the initial broadcast (and changed for subsequent re-broadcasts). Ken quoted Andre Mendes (PBS's Chief Technology Integration Officer, described as combining CTO and CIO duties) as saying that the "keeper" for PBS is the Distribution File. Thus, Ken opined, there was possible value for having LC keep the Production File for backup. But then Jim said, "We will keep the Production Files for a certain period, for as long as we have the rights to broadcast the programs."

At this point Dave joined in, reminding the group of the hoped-for distributed storage possibility (the collective repository): "Ought not the producers join with PBS to keep the Production Files," he asked. Ken added that this is where the LambdaRail high-speed network comes in, because it supports this kind distributed storage. Jim added that, very soon, DS3 is going to be in place from the major producers to PBS.

[After the meeting, Carl looked up the "DS" business on the Web and found this for himself and other lay people: DS stands for Digital Signal or Data Service level. It was originally an AT&T classification of transmitting one or more voice conversations in one digital data stream. The best known DS levels are DS0 (a single conversation), DS1 (24 conversations multiplexed), DS1C, DS2, and DS3. By extension, the DS level can refer to the raw data rate necessary for transmission:

- DS0 64 Kb/s
- DS1 1.544 Mb/s
- DS1C 3.15 Mb/s
- DS2 6.31 Mb/s
- DS3 44.736 Mb/s
- DS4 274.1 Mb/s]

Ken said that he liked the idea of the distributed Public TV repository. He was thinking of one copy of the content at four locations: Thirteen, WGBH, PBS, and LC. Dave noted that this storage capability would manage both source material and finished programs. Ken wondered if there was a way for LC to be a node in the distributed network

The rhetoric tended to be stated "as if" each program was duplicated at four locations. Jeff reminded us that the overall resource would be managed in a manner analogous to RAID storage, where what is going on is more complicated but ultimately "safer" than storing four copies of an item. "You are storing bits not programs," Jeff said. "You don't just store bits but also data like checksums, think of the four units as a storage area. The question is how you use the four units to manage the preservation of the bits." Dave agreed, noting that our talk about

copies of programs is an abstraction. Jim added the thought that if the four systems are used by their owners for other purposes, then you may need have enough intelligence to manage files or programs as such.

Nan asked if, as the project moved forward, would we propose ideas about a Public TV trusted repository to LC. Ken said he would worry this might seem scary: "I would rather go for lesser push in this project, maybe we should focus on the Production Files from PBS." He would like to consider seeing that as the distributed repository for Pub TV, this may be a better investment for us than trying to sync everything with NAVCC.at this time.

At this point there was a side conversation on a possible meeting of LC and Public TV folks at FIAT (conference in New York in September). Although Ken had once hoped for a bigger meeting with several principals, he would be happy just to have a focused meeting with Greg Lukow. Emails sent directly from Boston started the ball rolling for this meeting.

There was intermittent talk about metadata and other matters. Before the lunch break, Ken said that he wanted to study up on DRM. He said he was interested in Rights Expression Languages, what should we be doing here? In a distributed repository, security is an issue, we need to manage loaner versions. Ken and Nan made reference to discussions with Bill Rosenblatt, described as a clever fellow on this and other topics. (Carl recently forwarded Rosenblatt's paper on Enterprise Content Integration, downloaded from the consultant's Web site: <http://www.giantstepsmts.com/ECIwhitepaper.pdf>.) The Public TV folks met him at a conference, reporting that Rosenblatt founded <http://www.drmwatch.com/>, and although it seems to have been recently sold, remains active with it. Someone mentioned <http://www.fathom.com/> but Carl did not catch whether Rosenblatt had been connected to this as well: Ken said that Fathom was an effort by several universities to monetize their digital library assets.

[Carl googled this later: The Fathom Knowledge Network, Inc., was founded by Columbia in 1999 and launched its website in 2000 with the goal of providing high quality educational resources to a global audience through the Internet. The Fathom academic consortium grew to include 14 leading educational and cultural institutions dedicated to that goal. The Fathom website ceased operations in March 2003, as part of a reorganization of Columbia University's digital media activities Although visitors are no longer be able to purchase courses through Fathom, Columbia will provide this Fathom archive so that interested users may continue to access to the complete range of free content developed for Fathom by its member institutions. Columbia encourages you to browse this archive of online learning resources, including lectures, articles, interviews, exhibits and free seminars.]

Dave remarked on a DRM-related paper that he had really enjoyed reading recently. It is the "Microsoft Research DRM talk" by Cory Doctorow, dated June 2004, and available at <http://www.craphound.com/msftdrm.txt>.

After lunch, Jim passed on the very interesting news that an MXF application specification (AS) was being developed for PBS, for the exchange of content. Called the AS/PBS for short. He explained that, in order to implement the specification, PBS was working with various

specialized video server manufacturers. PBS will have a test file, and server makers will have to demo their systems with the specified file. AS/PBS is getting a lot of design help from Oliver Morgan. Dave said that Metaglue is Oliver's company. Morgan is a former SONY and AVID worker, and he shepherded MXF thru SMPTE.

[From the Metaglue Web site: "Metaglue Corporation was founded to provide software, support and services to media companies who are taking advantage of the new file transfer and metadata standards, and need to build interoperable systems using best-of-breed products of their choice." One of their products is MXFixer, described as a "tool to manipulate MXF Files. Like AAFixer (another Metaglue product), it works with both binary and XML representations of the file, allowing users to connect MXF-based media systems with AAF-based Postproduction systems and XML-based business and media systems. MXFixer plugs in to the Diffuser system for integration of standard and private metadata. MXFixer is a powerful integration tool for MXF- and XML-based systems."]

Dave noted that he had lobbied for "non-relevant data" in the MXF in his work with Morgan. Dave was working to bring to life his UPF ideas, to have codec and codec tools in the format, this could be the source code for the codec." Carl asked about XML versus KLV encoding. Dave said that in MXF all of the data is encoded as KLV, adding, "It's just an encoding."

Ken asked if AS/PBS might be used by this NDIIPP project? Is there a package idea in here that fits our needs? Dave asked Jim, "Will you share the AS/PBS with the stations?" Jim said that he thought they would, after the specification is final.

Jim observed that server manufacturers have all tried to be unique in how they format video data within their systems, but now it is harder to protect your uniqueness. The only remaining unique feature is "how you store data, and that is what MXF attacks." PBS wanted to tailor instances of MXF to our needs. Ken said he would like to talk to Oliver Morgan about adding UPF data (i.e., preservation metadata and more, like codec source code). Dave agreed that this would be a good conversation to have, even it is too late to specifically deal with modifying AS/PBS. Jim said that we could have the next meeting here, since Morgan lives in Lexington, Massachusetts.

The discussion of metadata proceeded with Ken explaining that, in connection with PB Core, they had been talking to Lisa Carter, formerly of KET (Kentucky Educational Television), now of the University of Kentucky, and very active with PB Core. Ken planned to explore with Lisa if she would be interested in this NDIIPP project, to connect this project to the emerging PB Core users group.

There was sidebar on Jerry Field and WGBH's NCAM (National Center for Accessible Media; <http://ncam.wgbh.org/>) project. They may have just gotten funding for the development of a tool for PB Core; how might this connect to the NDIIPP project in terms of toolmaking? Dave noted that WGBH, years ago, created a Dublin Core toolset. But it was not taken up by the community, not understood by the community at the time, and he would like to try again.

Ken said we will try to engage Lisa. Nan added that they have talked to her, and found that she

may have some time in the near future. David asked about Merrill Weiss and a connection to SMPTE in this metadata matter. Ken said that he did not see the need to engage a formal SMPTE process, the desired specification was just for ourselves and for Public TV. But he noted that MXF is a brand, and therefore is useful in the broadcast community. Someone said that hope for a “block diagram” regarding metadata [NDIIPP project metadata? all PB Core metadata?] by the next NAB-Public TV meeting.

Ken noted the possible interest in mapping PB Core to LC’s structure in MAVIS, maybe Lisa could help with this. The hope was to find a direct pathway to put content into LC, which will save everyone money: can we use PB Core and do this? Nan added that maybe we should (also or instead) talk to Grace Agnew about data mapping.

There was then some discussion of the repository, how to jump start this at NYU? Ken said that we just have to wait for Jerry’s successor. This was to be an implementation of DSpace.

There was a brief discussion of publicizing the NDIIPP project, with presentations at mtgs like Public TV technical conference, a project Web site, etc. There was talk of an article or special issue in *Current* (a Public Broadcasting Web magazine, <http://www.current.org/>).