

IMLS Focus: Strategic Priorities 2014
National Digital Platform
April 29, 2014

Dr. Anthony W. Marx

Tony Marx welcomed attendees to NYPL and opened the convening.

IMLS Deputy Director for Libraries Maura Marx

Maura Marx asked participants to focus on identifying short- or long-term goals, priorities, and practices to hold up as models. She proposed some possible questions to be explored:

- Is it our common goal that every cultural institution, every library in the U.S. should be able to participate in a digital system that serves people across the U.S. by providing them a menu of digital services through some type of shared service platform? If so, what are the next steps?
- Is it a goal to accomplish more in actual digitization? CLIR with the support of the Mellon Foundation recently announced they will be ending the Hidden Collections cataloging program as we know it and shifting attention to the digitization of rare and unique collections. Does this present opportunities for collaboration?
- What role do e-books play in our future, and what role do we play in e-books' future?
- Are there existing local models or services that could be scaled up for greater benefit?
- What's on the horizon? What should be on our research agenda? Our policy agenda?
- Of the many such goals we'll discuss today, how should they be prioritized?

Session 1: INFRASTRUCTURE

Jim Neal (Moderator)

Dan Cohen

Brett Bobley

Elliott Shore

Dan Cohen talked about how DPLA serves as a national network model with tremendous buy-in from stakeholders, including state and regional hubs. He challenged everyone to think about how to support funding for a national network infrastructure model. He called for a dedicated national platform funding stream and asked attendees to think creatively about how to support it.

Brett Bobley posed the question of how funders can work together to ensure national infrastructure has a life, not just start-up funds. He spoke of an "All roads lead to DPLA" strategy, in which funders would no longer fund "siloed" projects but would prioritize projects that clearly fit with pieces of the national infrastructure such as DPLA.

Elliott Shore noted the additional funding stream of universities and academic libraries, which have already funded pieces of the existing digital infrastructure. With 20 years of experimentation, he suggested we can now identify the key pieces and figure out how to work together. It is a large investment, but the cost avoidance could be worthwhile.

Jim Neal summarized themes across the three presentations:

- The essential importance of the national network model

- A new way of thinking about funding, in which we assess digital projects and applications projects by how they fit with DPLA and other infrastructure
- Regional hubs, subject hubs, format hubs, service hubs, and the different functions that hubs play in this environment
- The importance of building public-private partnerships to construct and sustain this
- Narrowing our national agenda, to select the projects that support this national digital platform that we are constructing
- A systemic national strategy

Issues/Questions/Comments from group discussion:

Mackenzie Smith: Are there other national infrastructure models that are already working well? What is the relationship between sustainability and accountability?

Rachel Frick: Beyond the operational considerations, is there a research network that can watch what we do and try to abstract lessons for us to inform the way we move forward?

Ken Wiggins: Connecticut is helping to fund infrastructure because the state legislature has to sustain access to their information. Where can we create similar points of greater relevancy? The American public and the American government are not going to support whatever we come up with for infrastructure if we can't really show relevancy.

Don Waters: At certain points in technological change you need to scale up, so there's a tension in this conversation about whether to link existing parts together or build something bigger. We also need to factor in issues like succession rights, intellectual freedom, and copyright reform, which are relevant to sustainability. We need to think about how we prioritize funding one infrastructure over another, and in addition to the discovery layer, there are layers of preservation infrastructure, network infrastructure, and legal infrastructure.

Dan Cohen: There could be more coordination between various projects, between DPN, SHARE, Hathi, DPLA, and others, which perhaps becomes a meta-hub or some kind of meta-organization charged with working toward coherence.

Bernie Margolis: Unless we figure out way of diminishing competition among all the players, there will be no sustainability and conflicting possibilities of how we mount a national digital platform. Some of the other major players who own a lot of information about the digital resources in our libraries should be part of this discussion. Finally, what are we doing to guarantee access for everyone, including those who can't afford the access to these resources?

Brewster Kahle: What does it take to "finish the damn project digitization thing"? It would cost \$160 million for a collection on par with Yale's or the Boston Public Library. Storage costs for 1.6 petabytes, which is everything collected in the text area and saved for another 20 years, could be a \$1 million-\$2 million investment.

Brewster Kahle: The Public Library of Science (PLOS) is a model. Why don't we have the public library of X, of Y, of Z? It's all open-source and there for the taking, but we're not replicating it in other subject areas.

James Shulman: We have to recognize that IMLS or any other funder is really a venture capitalist at best, and can help address some kinks in the system, but we also have to collectively support the alleviation of other kinks inhibiting the flow of information.

Mike Furlough: Much of the national communication infrastructure is run by commercial organizations. How do we work with that side and what are the natural areas of engagement between us?

Elliott Shore: Some of the things we could do as an academic community might have real value to the commercial world. Can we keep separate the control of the intellectual side while we monetize the intellectual property in what we developed?

Alan Inouye: The 2016 election is looming and we need to think about big messages and reaching out to political champions on this type of project and its greater relevance.

Dan Cohen's final take-away: It would be wonderful to be able to say in 2016 that we have, in fact, completed a national network of hubs and that it has shown other results for the general public, researchers, and students. That is something politicians could tout, ALA could tout, and funders in this room could tout as a great end result.

Brett Bobley's final take-away: This is a major project of national importance and we need to figure it out, need to continue meeting like this and meeting more often with one another to coordinate these many efforts.

Elliott Shore's final take-away: "Who else but the library?" We are the people who collaborate best in this world. These small conversations are easier for us to lead, but we need to invite more people into the conversation.

Session 2: CONTENT

Rachel Frick (Moderator)

Sari Feldman

Katherine Skinner

Clifford Lynch

Katherine Skinner noted that content is not created or managed in a vacuum; it always happens in the context of communities. Take newspapers as a specific kind of content, which have been addressed through individual digitization projects and some networked projects. Now we need to engage across the entire stakeholder community to address current content. In terms of newspapers that includes not just the digitized copy, but online comments that represent a shift in public rhetoric.

Sari Feldman added that the shift from print to digital news created a skills gap for their library patrons, and now they are looking to patron-contributed content as part of the civil discourse. In terms of makerspaces, and audio/video in particular, how are libraries going to curate, collect, and share local content that has global potential? How does that fit with a national platform? For most of their patrons, the discoverability/usability of digital content for personal research needs is still not a very fluid process.

Clifford Lynch noted that the academic perspective adds questions of content production in an era of scholarly communication disruptions. Thinking about a national digital platform necessitates looking at

the broader cultural realm, such as the music industry over the last ten years. What happens to 25 years of YouTube videos, considering the expense and all the kitty cat content?

Rachel Frick asked whether we needed separate DPLA-style systems for the public and academic communities.

Clifford Lynch cautioned that the splitting of information is a very corrosive thing; we do the public a great disservice by fencing off academic discourse. The emergence of the Internet actually chips at that barrier a bit. It would be wonderful if research institutions, public libraries, and school libraries had a more cohesive way of working together and were more effective in providing access to information. A research alliance that bridges different cultures is a really powerful idea.

Rachel Frick asked how open access fits into the idea of prioritizing a national platform.

Katherine Skinner noted that it's about levelling the playing field so that both small and large institutions have access to the same sort of research findings. Anything government funding agencies support should strive for broad exposure, because it doesn't help for us to have a lot of small innovations that remain unknown. Figuring out how to incorporate open access is important. Libraries are constantly reevaluating the role of openness in their content, and a complicating factor is library budgets subsidizing the publishing cycle. A provocative white paper just came out about how to potentially change where the funding stream comes from. The only way we'll get a massive overhaul of our system is by aiming up several levels and by engaging across different stakeholder communities in order to make big decisions that lead to radical change.

Sari Feldman added that the library community has strong feelings about digitizing materials to provide broad access, but that is not the case for other kind of institutions. When the Cleveland Museum of Art digitized its entire collection to create virtual tours, it was difficult to pull that collection into the library for wider community use. Having opportunities to pull in partners like this is important.

Clifford Lynch commented that museums are slowly getting on board, and genuine change is taking place. It is often possible to get grant or one-off funding to build global open access resources, but the library community has had a lot of trouble getting traction around sustaining them. It gets even worse as we move beyond the purely academic world; look at Wikipedia. News is an example of something that sits between open access and genuinely private content, and we're going to need some new arrangements there. Finally, we can't overlook the things that are fundamentally content infrastructure, like authority files, which are going through incredible transition now. We're adrift in terms of how to fund those.

Rachel Frick asked whether we still needed a focused funding stream for basic digitization.

Sari Feldman shared that in the public library world people still need funding for basic digitization, as well as help in accomplishing it.

Katherine Skinner noted that before putting more funds toward digitization, we need to look at infrastructure. We lack the fundamental preservation infrastructure that is needed in order to make sure what we digitize survives.

Clifford Lynch added that continued funding for digitization is important, but we need it in a context that emphasizes the deliberate building of digitization capabilities in a few centers, particularly around media-based material. Until we fix the relative lack of access and curation services, it is dangerous to go

out of balance in the funding for digitization. DPN keeps coming up, but it is not a magic bullet for preservation. DPN is trying to deal with very high value, very long-term digital preservation, and it needs to be complimented by services in lighter-weight preservation environments. We don't have those things nicely packaged as services yet, so there is a need for investment in that area.

Issues/Questions/Comments from discussion:

Tim DeLauro: We need to be careful when we say that a tool like DPN is going to preserve something, because preservation includes things like format migration for long-term use. DPN is going to archive a certain set of content, so we need to be careful around terminology.

Doron Weber: To address the earlier example of Wikipedia having funding issues, they are, in fact, so successful that they cut their recent fundraiser short. They have deliberately chosen for the moment not to go to the endowment model. As for the earlier reference to corporations, they are engines of innovation and we want to work with them. But the values need to be our values, and the way we think about endurance, longevity, quality, and what is right for society is very different than maximizing shareholder value.

Suanne Wones: Is it time for some broad collaborative part of the library world to build its own version of a product like Overdrive, creating an e-book platform we can control and ensure that users' needs are met?

(A lot of clapping and amens from people on the live stream).

Jack Martin: This conversation is great and we should have had it 25 years ago. Librarians love to talk about infrastructure and rules, but easy-to-use, low-barrier tools like Twitter offer something we could learn from.

Sari Feldman's final take-away: Agrees that we should have had this conversation in the past, and now one of the big issues is that we have to move faster. There have to be concrete outcomes from conversations like this because we can't wait any longer.

Katherine Skinner's final take-away: Worries about constant money complaints, because there is a lot of it in this sector. The problem is where we direct the money. Over the last 20 years, a lot of innovative projects on two- to three-year cycles have been seeded, but now the funders need to figure out where to concentrate finances across these initiatives. Instead of funding projects, how do we fund cohorts that are forced to interact with one another?

Clifford Lynch's final take-away: Agrees that we're trying to move from scattered institutional investment to genuinely national programs where progress is measured in very different ways. Secondly, many of our funding strategies are state-based and community-based, but we ought to think about communities of interest that scatter in odd ways across institutions. Finally, we have an interesting set of opportunities in content that stands between the genuinely hard-core academic content and the very mainstream commercial content.

Session 3: USE

Susan Hildreth (Moderator)

Susan Gibbons
Bernie Margolis

Susan Gibbons noted that with digitization, we immediately leap over recent content, because it's safer to work with content from the distant past. But we know our users are incredibly interested in that recent content and we cannot continue to ignore the complications around it. We're starting to build a community of practice around copyright and implement take-down notices, so it's not that. Content can be restricted in other ways; we have personal information, Social Security numbers, and other information we cannot digitize. We have the archives of anthropologists and psychologists that did their work before the 1974 National Research Act and therefore before IRB (Institutional Review Board). Right now you must physically come to the reading room, show your ID, be verified, and agree to the use policy, all of which controls your use of the collection. It means all of this content cannot participate yet in this exciting world that DPLA and others are creating. So what can we do to bring that content forward? What's at stake if we don't figure this out is that those who have the luxury and the money to travel the world for access to this content will be the only ones who can work with it, and that really narrows the scope of what scholarship will be. Today's humanists cannot easily work with modern content, and so we're forced to go back a century. As we build this infrastructure, we also need to build protocols that would allow us to make available securely online this type of restricted access. It has to be something more sophisticated than firewalls and passwords. We can create a temporary VPN or secure tunnel to provide access and create a network between the library and other cultural institutions where you can go to a local reading room and be verified. We also have to think about how we modify the IRB processes for content that we already have, because the default will be "no, you can't work with this." We don't know what is safe and tend to have a risk-averse university council and others guiding us. We risk leaving behind this content that our users are clamoring for.

Bernie Margolis noted that Susan touched on the academic or scholarly user perspective. What about Sally, the nine-year old? How do we look at the content resources of a digital platform and design for her benefit? What is our plan, and what does the map of digital resources look like? What is the corporate material that is not digitized now? And if we had the money that Brewster Kahle referenced, how would we prioritize digitizing this content? It's probably not wise to digitize everything, even if we had that needed investment, and we should let the market decide some of that going forward. We are part of a world economy, and our resources are being used by people way beyond our boundaries. What is our obligation with regard to that broader world perspective? Those who practice reference work know that not just anyone can do a successful Google search. How do we create the tools for effective access? This is the domain of librarians, which leads to greater use of digital resources. Finally, we sell ourselves way too low when we use the term "digital literacy" or "information literacy." That is several levels below where the target should be, which is "digital fluency" and "information fluency." He doesn't want his neighbors to just speak the language; he wants them to speak it well so that when we communicate, we can do it with an enormous amount of effectiveness. One of the goals for IMLS should in fact be that by 2020, 98 percent of Americans are digitally fluent. Let's be bold.

Susan Hildreth asked about opportunities for libraries to help users with expanded access to digital content, and whether they are aligned with services/activities.

Susan Gibbons felt that services could certainly be provided, and within the growing field of digital humanities scholars need assistance with activities like topical mapping and data visualization. The parallel in science is the laboratory where staff and tools come together to work on problems. Libraries can become the laboratory for humanists and bring skills and tools to the table that humanists haven't been exposed to because their own training is text based.

Bernie Margolis returned to Sally the nine-year old. If she goes to DPLA and does a search that doesn't return relevant results, or she can't figure out what to do with the results, what happens? If we really were there to provide for her, there would be a place for Sally to connect to 24 hours a day. But today Sally would probably call her girlfriend and rely on peer counseling to help her out. That's not enough. If we were to do real metrics about who uses digital content, we would probably find that children under the age of 16 are among the highest users of the entire population. School drives that kind of use as more and more states are adopting research papers as a requirement for graduation, and Common Core is pushing more in-depth use. Yet have we designed systems that are geared for nine-year old Sally? There's a big gap, which represents older folks as well, who didn't grow up with digital technologies. The opportunities are immense, but there's probably not a magic bullet terms of the solution.

Susan Hildreth asked about challenges or opportunities for our users to be able to remix or pick and choose content from all kinds of different sources and create their own special content package.

Bernie Margolis posed the scenario of somebody who wants to find all of the photorealists between a set of dates represented in every public museum collection in the world. Today there is no tool to enable you to do that.

Susan Gibbons commented that there is growing potential for user-driven digitization, because too often librarians and archivists are acting as proxies to decide what gets digitized and made available. There are other ways this could be done that puts it within the hands of the users.

Issues/Questions/Comments from discussion:

Jim Neal: Quality going forward is going to be equal to content plus functionality. The one mass arena we're beginning to wrap our mind around is web content, which is multi-media, dynamic content. Twenty years from now our users are going to be angry with us because we did not take responsibility for the deep content that needs to be captured and organized.

Kathleen Williams: There's an opportunity here to work more closely with librarians, archivists, and educators to help us broker students' needs around researching digital content.

Katherine Skinner: Sally is really important, but we also need to look at people who are of age and can vote, because they're driving where we're going. Do we have information that helps raise the literacy of our people as a whole, because the universe is getting dumbed down in a lot of ways. One thing libraries specialize in is knowledge transmission, which can affect adults who are making decisions that determine what the world will look like for Sally 20 years from now.

Ken Wiggin: If the context is not there, people aren't going to understand how to use the data, and that understanding is as important as understanding the structure of the data or having access to the data.

Post-Lunch Update

Maura Marx gave an update on impending changes to the National Leadership Grants program, one of the IMLS discretionary programs. In order to be more impactful, the program will become more focused:

- Move away from wide-open call for innovation

- Look for proposals that move the needle on shared goals and community priorities, which will be named in the guidelines
- Shift evaluation practice from looking only at the individual project level up to the national/shared-goal level
- Enable IMLS to work better, in a more timely way, with partners and funders
- Help IMLS tell a more compelling and coherent story
- Add a second annual deadline in October 2014 (existing deadline in February)
- Publish new guidelines in July 2014

Session 4: Tools

Mary Lee Kennedy (Moderator)

Martin Kalfatovic

Tom Scheinfeldt

Ben Vershbow

Martin Kalfatovic observed that we've created a new set of tools: the information on the Internet, the interconnectivity, and the network pieces. How do we build tools on top of this meta-tool that we have in the form of the networked information? Three principles for tool building are that they are ubiquitous, transparent, and integrated. Ubiquity covers tools that are everywhere and usable at all times. Transparency is related to tools' invisibility – in the way that people today don't think much about where electricity comes from. The integrated principle is the most important part. If our tools don't work with all of the content that we have across all of these great systems that we're building, then we just have new silos that don't get used. The ecosystem of these elements is also important and explains why e-books finally took off. Lastly, the whole concept of tool building needs to be iterative; we can't just build a tool, walk away from it, and think it's going to be there forever. Iteration also relates to the part of the design process where we cut things out; we're going to build a lot more tools than we can actually use. How can we bring all of types of content that we're building into a place where the tool makers aren't just the people in our libraries, museums, and archives but the whole world of tool makers?

Tom Scheinfeldt opened with the question of what it means to be innovative. Innovation can't just mean new, it has to be somehow useful and implemented, especially by others. His criticism of the National Leadership Grants program is that it has created many leaders and maybe not as many followers. A model without someone implementing the model is not really a model at all, and that's especially true in the digital tools realm. There should be room for experimentation and failure, but not enough tools have been funded that have had a serious uptake within the community. There are a couple of things that might change that ratio of leaders to followers:

- More emphasis on uptake, and go-to-market strategies in the guidelines (there's currently a lot about sustainability, but when things are used, they are ultimately sustained)
- More emphasis on platforms than projects (this means picking a couple of winning tools/platforms and “doubling down” on them)
- More emphasis on outreach in proposals and in the guidelines (marketing to whoever the user base is, and articulating strong outreach plans)

Ben Vershbow would like to see IMLS and others collectively look at the library sector and

identify areas that are driving costs into the way we work, like the lack of a good open source ILS system. We've been justly celebrating DPLA and its strides in the discovery layer of the digital platform, but what does the rest of that end-to-end, publicly-supported technology platform look like for libraries and the cultural sector? How could IMLS and others fund groups of projects that fit together in the right way, without being overly prescriptive? For example, an emerging solution for e-books in libraries is developing in different pieces. If we pooled our collective efforts maybe we could be self-reliant in terms of e-book acquisition, lending, and reading in a matter of a few years. Software projects are like pets and take life cycle management. We're hatching a lot of new projects in Petri dishes, but often they die and some of them should have been picked up. Every project should have a fully documented archive so that it can be reconstituted at any moment, and so opportunities can be identified by other actors.

Mary Lee Kennedy summarized the discussion thus far and noted PubMed as an example of something happening in the broader ecosystem where there is a shared knowledgebase. She asked each panelist for a take-away.

Ben Vershbow responded to Tom's perspective on the importance of followers. There's an adage that the most courageous person in the beginning of a new movement is that second person who follows.

Tom Scheinfeldt noted that one of the strengths of the library community is that it's got a really strong bench of inventors. Sometimes the discussions get a little depressing, because we have so much to do, but we need to be mindful of how much room for invention this field allows.

Martin Kalfatovic offered a reminder that the tool makers are not necessarily within our own institutions, environments, and areas of discipline. We need to build our communities and open them up to accept tools that are built outside of our own, sometimes narrow, field of focus.

Issues/Questions/Comments from discussion:

Brewster Kahle: Would it make sense to have IMLS ask how grantees are going to address the tools that they've built in the digital realm and take them open source? Will it have an API? Is it Creative Commons? Will it be publicly archived?

Ben Vershbow: In terms of federal funding, the values that are encoded into the grant requirements include encouragement of open source, data management plans, and ways of dealing with data storage. What is missing is more the structure, the social frameworks to attract those followers and to raise the awareness, the community element. We also need to look at networking the review process.

Tom Scheinfeldt: Just as we think about our tools as executables, we could think about content projects that are in some way executable and can interoperate with other platforms and projects.

Katherine Skinner: There is a need for better distillation of the types of tools that are out there. What role might funding play in generating documentation?

Martin Kalfatovic: One of the things funders could look at going forward (and a number of them are already doing this) is the role of a technical writer and a community manager in a project. It would be somebody to manage the user community and to create documentation.

Tom Scheinfeldt: Sometimes there just needs to be a person on the project whose job is equivalent to what in the corporate world would amount to sales. You would never launch a product in the corporate world without a salesman, but we do that all the time. Some of it is having a line in the budget for a person, whether it's a community outreach specialist or technical role, on the project team.

Katherine Skinner: Is there a different project that might be funded that would select the most applicable tools and document them at a different stage, not necessarily as the project is taking place, but as a meta-project?

Ben Vershbow: Something at a meta level is needed. Some meta aggregation, with standards for documentation, so we're not always have having to drive attention to our projects on our own would be incredibly valuable.

Mary Lee Kennedy: What comes to mind is all the information done to document Open Access Repository. It's not complicated, but it does at least create a catalog of what's going on.

John Voss: Looking at European Commission funding around Europeana is a great guidance for us on multi-country projects, in which they're bringing together national libraries, nonprofit organizations, and enterprises. That's required in their funding to make sure they're building collaboratives as part of the funding process.

Session 5: Access at Scale

Josh Greenberg (Moderator)

Mackenzie Smith

Jason Kucsma

Dan Chudnov

Mackenzie Smith hoped to address three things: digital research collections at a typical research library, the parallel universes observed here today, and some research challenges. Analog research libraries like hers were historically very efficient, but digital research is so fundamentally different that our concept of collections doesn't scale anymore. The parallel universe is that there are a plethora of national digital platforms in existence, and we've been focused on the library-led ones. Other platforms, such as NDIIPP, LOCKSS, the San Diego Supercomputer Center's Chronopolis, DRIADE (Data Repository for Evolutionary Biology), CTSA (Clinical and Translational Science Awards), and DOE National Laboratories, are oblivious to ours, so we have some work to do in defining how these are supposed to work together. Research is also evolving so fast that even those platforms are becoming obsolete pretty quickly in favor of cloud technology. Publishers, quite frighteningly, are starting to move into the research collaboration platform space, which we haven't talked about. The other platforms have economies of scale and influence that we can only dream about, so we really need to work with them. What is the right relationship with a Microsoft or Amazon or Google? Research is spanning national boundaries now, and if we cannot figure out how to do this outside of U.S. boundaries, researchers are just going to walk away and find something else that works better for them. The infrastructure we're designing here

and trying to foster needs to work at web scale and that means global. Research challenges include preservation as a technical challenge; legal issues, including the huge tensions we see between creators, the institutions that hire and manage them, and the funders; privacy and security, because a lot of research data is human subject data; economics and sustainability; collaborations and good models for crowd-sourced funding; competencies and expertise; and governance models of scale. None of this is going to work if we can't figure out how to trust each other.

Dan Chudnov noted that when you collect something like all of Twitter, there are constraints you have to live with if you wish to succeed. It's reassuring that the Library of Congress has a copy of all of the public tweets, but there are limitations on providing access to that scale of data. How do you connect one-on-one with individuals, to allow them to take slices out of that? Our humble project that lets us collect some Twitter data and puts us in classrooms with journalism professors, teaching their undergraduates data tools. We have a push button on our app that every student can use to download all their elected representatives' tweets, which they can use to write a story about that. We're working with political science researchers to see how campaigns in the upcoming election talk about gender. There are a couple of gaps worth mentioning, though. Some of this is available through commercial sources. We're failing to capture the records of the communities we traditionally captured in other forms. The tool I hope we're building is giving us a view into that window, but it would be prohibitive for us to get into if we had to purchase that access. It's important to remember that some people don't even have a direct way to work with the scale of data we're collecting. Through our tool you can collect parts of the web that represent underserved communities. And it's important to be able to turn that around into something individuals or small groups can run themselves. Finally, there are a lot of opportunities for collaboration among different scales. When we turn this stuff into a special collection, which we're starting to do, we need to figure out where the overlap with MARC and EAD (Encoded Archival Description) is because that's what we know how to do. We need to partner with much larger organizations and much smaller organizations alike to figure out what modalities will fit.

Jason Kucsma talked about how the Metropolitan New York Library Council (METRO) works to scale up smaller and medium-sized institutions to have the kind of concerns and worries addressed thus far. His first project example, Linked Open Data New York City, illustrates how a hub like METRO can serve as a link between the expertise in larger institutions and smaller institutions. The project creates tools, including infrastructure and a discovery interface, for smaller cultural heritage institutions to publish linked open data and provide access to their collections at a scale never seen before. A second project, the Empire State Digital Network, connects the digital collections of different library councils' member institutions with each other through a statewide service hub. METRO is also administering an IMLS grant for the National Digital Stewardship Residency Program, which is part of the program's second iteration and speaks to the "followers" strand that came up earlier. It provides paid residencies, and the focus on digital stewardship gives host institutions expertise and builds on institutional capacity that has not been realized before. There are two through-points encountered between each of these three projects. One is that each leverages local relationships for a global impact, and enables these institutions to be able to be part of a conversation on a global scale. The other through-point that is important to think about with funding priorities is the idea of distributed risks and shared rewards. Some of their projects might not be fully realized, but the risks are distributed, and when other projects take off there are shared rewards.

Josh Greenberg asked fellow panelists for any immediate reactions.

MacKenzie Smith noted that the regional networks approach Jason talked about sounded like the regional networks that worked with OCLC. They became a thing of the past when OCLC was able to disintermediate the relationships, and she wondered about repeating patterns with these new platforms.

Jason Kucsma talked about METRO's funding, the majority of which comes from the state, and noted that there will always be a need to build relationships.

Issues/Questions/Comments from discussion:

Mike Furlough: Scale can imply a sort of comprehensiveness that we know is not there. As we try to collect more and more automatically, how can we make more transparent to the end users the biases that are inherent in our collecting methods and algorithms?

MacKenzie Smith: The idea of demand-driven or patron-driven acquisition/digitization/collection is one response, and we don't have good ways of doing this at scale. Hathi Trust and the Internet Archive have potential, and this is something we could do collectively.

Dan Chudnov: On the research side, the people who study social media are coming to their own realizations about what the scope of bias they're aware of allows them to do or not. When it comes to a Twitter sample stream, which is all we can afford to collect at GW, we're at least making guesses based on the direct requests we've received from members of our community.

Bernie Margolis: Why would we want to collect everything? Our training in terms of professional librarianship is to apply judgment and discernment on all the material available, and decide what's important to collect.

Dan Chudnov: Dan Cohen's post four years ago, on why it was valuable for the Library of Congress to collect all of Twitter's record, was a really nice guidepost.

MacKenzie Smith: The truth is the collection is already out there. It's the web. So we don't really have to worry anymore about what we're *collecting*. It's what we're going to *preserve* into the future.

Josh Greenberg: Over the next decade or two as we have more devices streaming their own data, heuristics for triage are going to be ever more important.

Brewster Kahle: As an industry we need to declare our stand as to what we are going to collect. The web, television, books, and digitizing are all relatively easy, but e-books have been a nightmare, Twitter is a real problem, and phones are becoming more closed at every level, including the apps and the network. Librarians' role could be to break every one of those digital rights management systems, to preserve the materials and make them as accessible as we can, but we need some kind of proclamation from our community.

Dan Cohen: It is essential to have a community of practice around data. When you have a group of practitioners then you get people understanding why we need the materials.

Josh Greenberg: A theme we're hearing is how do you deal with industry, the private sector? Are there concerns, opportunities, or thoughts?

MacKenzie Smith: One of the things libraries have used in the past to threaten companies who wanted to lock stuff up is that we have a fair use mandate. Threats of involving the federal government in demands to escrow proprietary software can be effective. Maybe we find out what the public collectively wants and connect with people in these companies who can work with us on a reasonable, non-commercially threatening approach to getting this data at the right moment. It would be very expensive and different for every company, but it would be a way of being proactive.

Dan Chudnov: The commercial marketplace is wonderful for certain services we depend upon. However it can change so quickly that we do have to hedge against buy-outs. It's a lot easier for us to partner with other institutions like those represented in the room, because we know we can build lasting relationships with each other.

Bernie Margolis: There is value to having a really robust national digital platform as a counter balance to the potential commercialization of the data resources that we have. However, is it robust enough to hold off somebody with great marketing skills and financial resources that would try to brand it?

Session 6 – SKILLS

Bob Horton (Moderator)

Nancy McGovern

Jack Martin

Nancy McGovern noted that despite some great recent efforts in the field, people keep asking what skills we need in libraries. We have a much better idea of the "what," but we also need to address the "what then," the "how," and the "how much," and think about who's asking, and whether it's an individual, team, organization, or community. There's still kind of a bridge that needs to be built for people coming out of programs, and even if we develop the best skills list we could imagine, that list will evolve. "Organization one" in the diagram on the slide, might be a professional organization, such as ARL or SAA. "Organization two" might have a bunch of people at different stages of their careers, so how do we hire, evolve, keep, and bring on specific skills for short periods of time through consulting and term appointments? Then there are teams, through which most organizations work to achieve their near-term and long-term goals. The person level on the slide might look as though it's wandering around, and it's because that person is trying to figure out which organization or team is best for them. We have a continuum of academic continuing and training that looks very tidy. We're getting to the point of collaborations in places where 10-20 years ago, those were silos, so how do we get strategic about skills? The aggregate capabilities listed on her slide operate at a higher level and aren't done by a single person. Lots of job descriptions include very finite things, but in practice people need to write policies and do

preservation planning, which are combinations of things. Another way of looking at it is what kind of characteristics are needed, which is a different approach than some of the skills we've had in the past. The hardest thing for people to do is self-assess, because they are very inclined to overestimate or underestimate their own skills. Could we get some tools for that? What about development plans for individuals that last across careers and multiple organizations? How do we fit this all together?

Jack Martin shared that someone like him brings new skills to the traditional sense of librarianship that have to do more with fabulous programming or enhancing educational experiences through digital tools. One of the skills that librarians need to figure out is the idea of a translator, and what librarian issues, like "infrastructure," mean for the user. His library is shifting from the idea of a traditional library model to more of an informal educational model, because that's where you gain a lot of the skills that were just discussed. It's also important that we recognize the folks who use our collections are learners rather than readers, which is a fundamental shift. The library is shifting from a grocery store model to more of a kitchen where you make, build, and learn how to hack stuff. The people who are working in libraries may not have a lot of those skills, and it's not necessarily a matter of hiring. Library schools may not be teaching the kind of skills that are needed to do the jobs we're describing. How do we find the right people to take on this educator role? You don't necessarily have to be an expert at using every digital tool, you just have to be really curious and unafraid of getting your hands dirty and applying it toward learning and libraries. You also need to be willing to share what you've learned, whether you're tweeting, blogging, etc., so that the community knows more about it. You need to be an advocate and speak up about issues that surfaced in today's discussion. You also need to be design-minded, by recognizing information problems and not necessarily solving them yourself, but reaching out to the right folks in the community who might be able to help. Finally, don't forget connectivity and all the communities outside of the library community.

Issues/Questions/Comments from discussion:

Tom Scheinfeldt: A lot of students seem to have a kind of fundamentally different relationship to information and citation. To what extent do we need to be training librarians to do the job that we think they should be doing in this new environment and to what extent should we be training librarians to react to what the new practices are whether we necessarily buy them or not?

Nancy McGovern: In terms of the characteristics, they're things you're not necessarily born with, but build up with a lot of different experiences. We have to reach out and figure out what the users are talking about. It has to be very responsive and it's going to evolve and also depends upon what setting you're in.

Jack Martin: There's an example of a learning environment for students where the instructor asks the kids to plan the course for the semester and then walks out. The result is a moderated conversation with peer feedback. That gives people a chance to investigate the 360 degrees around whatever issue they might be looking at, but also provides opportunities for other folks in the classroom to give feedback.

Jim Neal. The issue of the people who are not here and the diversity of that community is worrisome. We need to have representation from underrepresented cultural and ethnic

minorities as part of this conversation, and we should continue to invest in that problem. Data shows that in spite of the investments we've made, there's been no progress in terms of the percentage of individuals in our profession who are from African-American and Hispanic-American communities.

Nancy McGovern: Holistic kinds of training and professional development can help with all of these things, including diversity.

Ken Wiggin: I think there's some opportunity for IMLS and others to work with schools of education. We focused a lot on schools of library science, but I'm still seeing a lot of teachers, administrators, and schools who don't understand how to use all of this digital content and allow it in their classrooms. We can build these great tools, but until the students and teachers have an opportunity to really work with it, we're not going to gain that relevance we need in the schools.

Brewster Kahle: The people library schools are graduating are completely inappropriate for my hiring needs. It's a digital library, and library people have no clue how to write a script to be able to deal with something that doesn't have a point-and-click interface to it already. It would be great to see something really targeted to be able to produce hundreds of people that could work in our new libraries.