

Daqing He^{*#}, Dan Wu, Wayne Graves, Martin Klein

Creation of a Digital Library by the Communities and for the Communities

<https://doi.org/10.2478/dim-2020-0001>

received September 18, 2019; accepted November 28, 2019.

Keywords: digital library, ACM Digital Library, information access, information organization, personalized information service, user interface

1 Introduction

Association for Computing Machinery Digital Library (ACM DL) is a “research, discovery, and networking” platform developed by ACM for hosting a variety of ACM and related publications that include journals, conference proceedings, technical magazines, newsletters, and books, most of which are full text. It serves as a repository for high-quality computing literature and provides rich interconnecting relationships among authors, publications, institutions, and ACM special interest groups. As one of the oldest and most authoritative web archives for computing literature, ACM DL has greatly benefited authors, readers, and researchers of the computing community.

Last decade has been an exciting time for digital content publishing. Technologies for more powerful archive and access have been developed, and artifacts with much more diverse nature are published by academic and professional communities. Some digital libraries, such as the ACM DL, are also developing and designing their new digital libraries. Therefore, it is important for the digital

library community to work with ACM to identify critical existing barriers and potentially important directions for further development of ACM DL, and to provide more user-centered digital library services.

At the ACM/IEEE Joint Conference on Digital Library (JCDL) 2019 held in Urbana-Champaign, Illinois, USA, four researchers organized a panel named “Creation of a Digital Library by the Communities and for the Communities.” The goal of this panel was to initiate a collaborative relationship between the DL community and ACM DL. The panelists understood that the collaboration can happen on a much wide range of topics, including publication policy, open access models, curation of published artifacts, etc. Therefore, this panel focused its discussion around the tools and functions which the community wants to see in ACM DL, and which they could help to develop.

This panel consisted of two parts. The first part included four presentations: Wayne Graves, from ACM, discussed “ACM DL visions, goals and new roadmap”; Daqing He from University of Pittsburgh discussed “ACM DL users’ views on access and organization” barriers obtained from an online survey conducted for this panel; Dan Wu from Wuhan University presented the results from the same online survey with a focus “ACM DL users’ views on personalization and notification”; and Martin Klein from Los Alamos National Laboratory discussed “Piloting a ResourceSync interface for the ACM DL.”

Following is the summary of their presentations and discussions.

2 Presentations

2.1 Wayne Graves: “ACM DL Visions, Goals and New Roadmap”

Wayne’s presentation started with a brief history of ACM DL. ACM DL started in 1998 as initially an in-house development for enabling the digital library part of the publishing exercise. Over time, ACM DL has been

***Corresponding author: Daqing He**, School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, USA.

E-mail: dah44@pitt.edu

Dan Wu, School of Information Management, Wuhan University, Wuhan, China

Wayne Graves, Information Systems, Association for Computing Machinery (ACM), New York City, NY, USA

Martin Klein, Research Library, Los Alamos National Laboratory, Santa Fe, NM, USA

[#]Corresponding author Daqing He was responsible for recording, text conversion, revision and proofreading of this panel paper.

integrated closer and closer with the actual publishing. In addition, ACM DL also has the ACM Guide to computing literature as the other cool asset. This literature evolved into the foundation for computing literatures, not simply around publications, but also computing space. Through a few iterations, various parts of the ACM DL have been well integrated to provide the readers with the current digital library experience.

About 2 years ago, underneath the direction of ACM's publication board, ACM DL started a new round of improvement. This is to answer the tremendous development of the publishing industry since 1998. The focus of the new improvement is around the scalability of the DL and the diversity of the artifacts that the DL has to handle. ACM is working with a platform provider called Atypon for building the new DL site. This is to take advantage of an existing platform rather than to reinvent the wheel.

Wayne also presented a roadmap for further improvement of ACM DL. Some of the activities have been under the way, and some are on the table for prioritization. This is why he came to JCDL to engage the digital library committee to develop the right set of core features with the right capabilities. Some of the new features and capabilities mentioned in his communication:

- Conference is really important to this community. It should have more visibility as a conference itself as opposed to the collection of the artifacts that it produces.
- Some artifacts in the DL are actually the people who involved in the conference. The DL designers really feel like this is a strong message.
- User's engagement and feeling of ownership is a key feature to be developed in the DL, so are the features around personalization.
- Exploration will be on the right kind of metrics to evaluate and engage with content, people, institutions and event. There is a core set of metrics right now in the DL, but new exciting metrics will be developed with the community feedback.

The new ACM DL is at the design phase, and the URL is dlnext.acm.org. It is a beta site that is completely functional. The users can sign in with their accounts, and search for artifacts in the DL. A cross-linking will be added into the current ACM DL site so that users can be guided to the new site too. Feedback from users will be collected for obtaining great ideas. A formal usability examination in the design will be conducted too.

2.2 Daqing He: “ACM DL Users’ Views on Access and Organization”

Daqing's presentation focused around the results from an online survey regarding the access and organization of information inside ACM DL. The survey asked respondents to look at existing ACM DL as well as future ones for their functionalities. The survey was conducted using Wuhan University's resources, and the responses were collected between May 12, 2019 and May 22, 2019.

In total, 146 responses were collected from 63 male and 80 female respondents. The majority of them were in the age range of 19 to 40 years, and lots of them were students. Their disciplines range from computer science, information science, library science, and other kind of engineering and science area. Majority respondents came from East Asia, mainly China (about 69.18%), but we do get people from North America, Europe, and other places.

The results show two important messages. The first one was that different users had different motivations and different tasks when they engage in ACM DL. Majority of them aimed for obtaining updates on a specific topic, and looking for the more recent publications. However, there is indication that users from East Asia focused more on getting familiar with the topic, and there were lots of activities related to searching for an author. There are differences between students and non-student users too. Students often looked more on getting familiar with a specific topic/subject area, whereas non-students aimed for getting updates on recent publications. Similarly, academic users wanted more on getting familiar with a specific subject area. Non-academia users wanted more to “search for an author” with less emphasis on “update on recent publications.”

All of these reminded Daqing of Gary Marchionini's 1997 study of Library of Congress Digital Library's interfaces. This study showed that digital library in this scale needs to consider the role and the task that its users perform, and designs different entrance points in the digital library for users with different roles and tasks. Each role, such as students, with a task, such as getting familiar with a topic, can have a specific path to gain access to the DL.

The second important message is that even though search has been very important in ACM DL, the further improvement of the DL should be around the subject areas of ACM DL. When users want to get updates in some areas, they look for particular subject areas. They also examine publications on individual conferences and journals in certain subject areas as a way to access information. Even when people look for authors inside ACM DL, they may

also want to know other authors within the same subject areas.

Another evidence to support the second message is that when the major access barriers for accessing ACM DL are asked, search was mentioned by respondents, but majority complaints about the barriers centered around browsing using different subject areas, browsing through special interest group, and browsing using ACM subject CCS (Computing Classification Scheme). All of these are related to the subject areas too. Therefore, it is great to know from Wayne's presentation that ACM DL has been working on improving its subject areas.

Users expressed that they wanted more inside the papers, not just paper themselves, but also datasets, figures, tables, and supplementary materials. They also did not just want papers in PDF format. The majority of those people want HTML5 too.

Finally, Daqing's presentation showed that ACM DL is part of the ecosystems for people to access. Google search engine, Google Scholar, library catalog systems, and various conference sites could all be the initial entry points for users' access.

2.3 Dan Wu: "ACM DL Users' Views on Personalization and Notification"

Dan's presentation started with a comparison between the old and new version of ACM DL. The comparison focused on the aspects related to the homepage design, the search, and browsing functions. The old version of ACM DL has excessive sublinks on the homepage, which resembles a list structure. The search in this version is also keyword-based queries without any intelligent support. There is no classification on the returned articles. The browsing function of the old ACM DL uses excessive texts without sufficient preview capability.

In the redesign for the new version of ACM DL, top menu tabs are added, and more dynamic information is presented, which includes award winners, preview of books, leaderboard of articles, magazine cover, details of conference, and hotspots of proceedings. The search function of the new ACM DL has article classification and filtering option. Users can also choose to save their search history. More support features in the search, such as query suggestion and auto completion, as well as cross-language search, are also added. The new version also has recommendation for new articles and books, and allows users to recommend or share their articles. The article page of the new version has clearer layout and provides supplemental materials such as the related videos.

Dan's presentation then moved to discuss the survey results on ACM DL users' attitude about personalization and notification in the DL. The results showed that most users hold a positive attitude to create a personal profile inside the DL using their email. They are willing to provide their research interests in the personal profile. Most users preferred the recommendations on articles and journals, and interested to see the DL having social platforms for sharing individual research outcome.

Users also want more personalized support in the DL. In search, query suggestion is the most needed support, followed by personalized ranking of the results based on the user's search history, browsing history, and research interests. They also want the navigation components in the interface to be customizable. Around the function of intelligent notification from ACM DL, most users hold a positive attitude to get the notifications for latest updates related to individual research.

Dan's survey results show different attitudes in different regions and user groups. North American users showed more interest in using email address to create personal profiles, whereas European users were more interested in personalized search and social platform. Users from East Asia were more concerned about the font and navigation buttons in personalized user interface, and they were not interested to provide information to establish personal profiles. The results are also different in the student and non-student groups. The students focused on the font size, color, and navigation components. They had higher need for a social platform inside ACM DL to communicate with others. Non-student users concentrated on issues such as notifications related to recent work and new citations, sharing individual outcome, and making comments. They also preferred the personalized rankings and cross-language search services in DL. In addition, academia group users would be interested to provide more information about creating personal profiles and to accept publications recommended by ACM DL, but non-academia users had uncertain attitude for creating a personal profile and had less need for intelligent notifications.

In the last part of her presentation, Dan proposed suggestions for ACM DL:

- The DL can display interested ACM publications and Special Interest Groups (SIGs) to users, and allow creating tags for publications and discussion groups.
- For better communication, the DL can provide the functions for users to leave message publicly or privately.
- The DL can recommend query fields based on real-time search hotspots and track user's usage behavior for personalized search results.

- Cross-language search is very important.
- The DL can allow users to customize the font, size, and color of the interface.
- The DL can enhance intelligent notification with messages on new citations, special issues, and upcoming conferences.

2.4 Martin Klein: “Piloting a ResourceSync Interface for the ACM DL”

Martin’s presentation first introduced his work on using sitemaps to enable the programming interface to ACM DL’s resources. His goal is to use the pull-based approach to provide a framework for better synchronizing and understanding resources on the web.

ACM DL currently provides access to metadata files, PDF format, and HTML format of its published work. The access is granted individually via an FTP server. So, the presented work is to provide a standard-based machine accessible interface to ACM DL resources. By taking advantage of the fact that current ACM DL is a site with organized books, journals, magazines, newsletters, and proceedings, machines can follow the directories in the ZIP file to locate the metadata in XML and the published works in PDF.

As part of the presentation, Martin released 12 resource lists that contain description of ACM DL resource. The lists also contain pieces of metadata proceedings in the capability and resource lists, and the links that connect metadata resources with published PDF files. By conveying links with the linear relation type, machines can interpret scenarios within the XML file for describing the PDF files. This is a huge advantage over other approaches.

Martin also wished to implement something called Signposting which is another approach to foster interoperability between systems and across systems for machines. Signposting uses HTTP links to explain the relationship between link resources and machine.

However, there are still some important questions to be answered. On the result page of ACM DL, users can identify a whole bunch of resources that are linked from landing page, such as the link to the PDF document, authors with their affiliations, the digital object identifier (DOI), as well as the citation information. Human users are smart enough to find this information. But how would a machine approach this? If a machine references the DOI by following the link, how does it identify what is this DOI? How do users disambiguate between the first author and the second author? People can identify the fuzzy

concept like names, but machines cannot do any of these interpretations as human.

By using HTTP links, we can convey relationships. HTTP links basically cost nothing, but they have the potential to really make a huge step forward in terms of interoperability of systems.

Finally, Martin wished to get more communications and discussions about the feedback of use cases which will try to be better stewards for humans and machines.

3 Questions and Discussions

After the panel presentations, there was a question and answering session. Following is the summary of the questions and answers.

Question: *If a random individual users use the ACM DL, the first step is to identify them as individuals rather than institutions. Or as the university subscribes, they have to log in to university library? What is the process of the DL to make individual’s account free? Or they have to need to be an ACM member?*

Wayne: Yes. There are some advantages to ACM membership. Although ACM DL system identifies you coming from your institution, ACM DL encourages user to sign in and register. The system can recognize users now even when they’re not on campus. In this scenario, users can sign in their mobile phones and continue to use the university’s library or research library. There are some benefits that users still have those access rights, especially about personalization and customization.

Question: *Would there be ways to opt-out of certain features if some of them are not interesting to me?*

Daqing: Through the survey, we indeed identify that users want to personalize the ACM DL capabilities. So that’s opt-out function should definitely be available.

Wayne: I will say there’s a somewhat of an exception to that case. For example, when users browse the author page, they can see the aggregation of the published work, which is just another view of publications. In this case, users didn’t opt into that. And now users do have the option of opting in to provide more personal type of information on that page. And we could have taken the approach of starting to pace that stuff on there. But we really felt that that was sort of crossing the line and it should be on you to do it.

Question: *Some of the things people might want to opt-out. These things might open up just like any social network bullying or harassment. How are you going to make ACM DL a safe space for people? Can you block some people? Are*

you going to be able to control who can comment, who can do those sorts of things?

Wayne: ACM DL is not just a software platform. ACM, as an association, a society, and a community, has done huge amounts of work on this code of conduct stuff and code of ethics. We don't really moderate, but the posts would be looked at by the community, and a flag can be raised to take down a post. The whole community are dealing with this.

Question: *How the new library, like communities in other collections and academic libraries, help administrators evaluate organizations and individuals within communities or within universities?*

Wayne: You know, there have been some conversations about how can we get these author page representations, what are those pages really need to have, how can they best convey to these use cases, where someone's doing specific kind of evaluation, what are the stats that are missing, and what are the connections that need to be done? This question is very helpful. We need people to participate in these kinds of discussions and these are the things that are going to get float back to the digital library committee, and how do we get the community to give us feedback on it?

Daqing: As I mentioned early. This panel is really just a subgroup of the bigger ACM DL committee. I think this metrics evaluation certainly is the topics in the bigger committee. As a panel, we did not specifically look at that. That's why we did not mention any of this in our presentation.

Question: *I was looking through many places to find the JCDL, because I want to find out if there was something going on. And I know it's a little complicated. In the end, I didn't find anything I could spend forever looking. But I think that this notion, at a minimum, goes back to the conference proceeding. I don't know where are the documents, but you have a link to the JCDL's conference, the upcoming conference, and the past conference. I understand some people wouldn't be interested in. But a lot of young people, like a Ph.D. student, bringing a connection between the proceedings and being able to find the people who run the conference, and information about conference planning, these are very important for them.*

Daqing: Through our survey, we find lots of users want to access the digital library via conference sites, and also go to conference sites via digital library. We already start to see that people don't think ACM DL is just a paper collection. It wants to become some kind of knowledge repository and organized repository.

Wayne: This is very much in line. We didn't get to spend a lot of time on some of those like new pages. We

have the first attempt in this beta site, which I encourage you take a look at and find JCDL when it first comes up.

Question: *I'm wondering if we can come up with some ways to allow some of these experiments to take place. How did you get access to the raw materials that people can try out their vision out there?*

Wayne: All these great ideas are not going to be implemented by me. We do need to implement some functions to make this core product fly. But there will be things, particularly those around visualization, that are well-tailored for exposing to the community in a simple clean standardized way for contribution. Once received a community product for extending the DL, I can potentially promote it into the product and make it available. This idea is one of the priorities for this committee to figure out.

Daqing: We can see this as a potential to develop ACM DL challenges for a collaboration between researchers. A theme of priority and topics can be developed every year to let people to explore. Once the winner is selected, there can be negotiations between ACM and the researchers to figure out how to put it inside ACM DL.

Question: *We're both consumers of the digital library. We produce the content for publishing, the DL is the publishing mechanism. Right? In our life cycle, we're generating papers, pushing or more evolution of trite templates, whatever the things that we generated in the ADM template. Imagining what our publication work cycle is in relation now to the digital library. Do you have any comments about that?*

Wayne: I will say a lot of this. This platform is your space. We're trying to reimagine the sort of the workflow or the review cycle for publishing papers. So, I think there are a whole learning curve and sort of cultural shift. I think it really is a collaboration. And the further upstream were involved, the more benefits and automation we get in.

Question: *I have a question about the recommendation system. Is there a risk that people use it as the only entry point? Maybe they will ignore some other sources that are important for researches. There will be an issue for the fairness of citations, the recommendations would likely be highly cited, and it will increase their probability to be showing up for our users within this library. And another concern is about the privacy issue. When you talk about personalization, there is always something about the behavior data stored. Have you got any policy and mechanism to keep our private data safe?*

Wayne: For the first concern, the recommendation scope is concerned all the time. We always try to expand the recommendation boundary without necessarily stopping this from being a specific kind of repository. There is a balance between the recommendation result and the

recommendation boundary. References recommendation are now resolved through cross-reference. It's no longer that just a recommendation between two articles, it also can be designed at the digital library scope to consider articles being interlinked and the citations being counted. This expands the scope pretty wide and starts to increase the universe of things that could be potentially recommended. The second concern is about privacy. Frankly speaking, we don't have a lot of experience with worrying about behavioral data, because we haven't made much of a push there. But one thing is for sure, the ACM DL didn't and won't sell anything about user behavior data. We're not in that kind of business.

Question: *My question is about ACM DL's search function. When I searched for information and collaboration, I used Google and it did well enough. Maybe we don't really want the digital library to do everything well by itself. How much can you improve on the search part that you can do better than Google?*

Wayne: The search function is a work in progress, and it will be improved. It is a totally different approach to this new platform. We will try to do better on academic information search. You will be the judge on whether it is better than Google.