

---

**ACM-DL PTF *Final***  
***Report* Nov. 9, 2020**

Ron Boisvert, Stuart Feldman,  
Michael Ley, Jeanna  
Matthews, **Moshe Vardi**

---

# ACM Pubs Board's 5/2020 Resolution

“The Publications Board requests that ACM’s President, CEO, and COO assemble and charge a task force to complete a comprehensive review of the ACM Digital Library.”

## **Presidential Task Force on Digital Library, Metadata, Metrics, and Reporting Data, 6/2020**

“As a first step, the ACM President and President Elect convene a presidential task force (PTF) to :

(a) develop the scope for an independent review of the current state of metadata, metrics, and reporting data in the DL; and

(b) identify a suitable consultant or contractor to carry out the necessary analytical tasks”

## PTF Work

(a) Meetings: 7/10, 7/17, 7/31, 8/7, 8/14, 8/21, 9/4, 9/11, 9/18, 9/25, 10/2, 10/9, 10/16, 10/23, 10/30, 11/6

(b) Met with: CEO+COO (3x), Director of IS (2x), Director of Publications (2x), Publications Operations Mgr, Pubs Board Chairs (2x), Pres.+Vice Pres.

(c) Received documentation from: Pubs Board, Director of IS, and Director of Publications (Box folder)

## Key Facts We Learned

- Internal ACM DL team was spending a large percentage of time keeping up with functional requirements that DL platform companies manage well (security, upgrades, APIs, industry standards such as JATS and BITS etc).
- Migrating to a platform vendor was necessary and the process of migrating to the new platform has been largely successful.
- Key metrics needed by authors, libraries, and for the new business model have accuracy, consistency, and timeliness issues
- Error-reporting interface has usability issues.
- Under-resourcing of QA: error reporting, generalization, tracking, and resolution.
- Many sources of error are not related to platform migration, but still essential to track to their root cause and resolve.
- While the migration itself has enabled some new functionality, the process has also paused work on planning of new, needed functionality.
- Responsibility for the DL is shared between the Pubs Board, HQ IS, HQ Pubs, and coordinated by HQ COO, each as only part of their many responsibilities, leading to an unclear process for deciding and prioritizing new functionality.
- Challenge: DL as **both** an ACM content repository and a guide to computing literature

## Basic Conclusions, I

- (a) The ACM DL is the “**crown jewel**” of ACM in the sense that it is **key** to ACM’s business model and is the **official record** for the contributions of ACM members.
- (b) It is important to recognize the **accomplishments** of the ACM Headquarters IS staff, particularly, *Wayne Graves* and *Asad Ali*, under the leadership of *Pat Ryan*, in building the DL into the **amazing resource** it is today
  - (i) The **creativity and technical savvy** to implement ACM’s DL well before digital-library offerings were commonplace
  - (ii) The **wisdom** to recognize the value in moving our DL to a new platform now

## Basic Conclusions, II

- (a) The **platform-migration process** from the DLb4 to NewDL has **largely been successful**.
  - (i) **But:** Many issues related to functionality, features, and data/metadata accuracy **are yet unresolved**.
  - (ii) **And:** The world has **changed**, with new needs, new inputs, new opportunities (e.g., OA+virtual conferences)
  - (iii) **And:** Current organization and skill set now **inadequate**
- (b) **Must shift** mindset from a **DL-platform migration** process to a future **EC-level DL project** -- *integrating platform, applications, and librarianship*
- (c) **Essential to evolve** process and organization supporting the DL into a more **formalized** structure.

## What has changed?

- (a) Data/metadata accuracy have become **critical** to the ACM Open business model (AuthorID, InstID, etc.)
- (b) **Broader** sets of customers -- VPs Research, deans, chairs
- (c) **Higher** level of community emphasis on bibliometrics
- (d) **Broader** sets of possible inputs (*virtual conferences* => videos/presentations -- common rather than exceptional)
- (e) **Stricter** demands for interoperability, e.g., bibliometric standards
- (f) **Other** computing+ bibliographic resources: DBLP, Google Scholar, Semantic Scholar, etc.; also arXiv, and the like.
- (g) **Reminder**: *Greenhouse study* -- **scope and functionality**



## Structural Recommendations, I

- (a) Given the importance of the DL to ACM, its development and maintenance must be a **formal, measurable CEO accountability** (reporting to EC, incentives, etc.),
  - (i) Led by an individual with **specialized skill set** -- unique point of **responsibility** and **strong execution**
  - (ii) **Separate unit** from IS, Publications, and Vendors; working **with them**, but **not reporting to them**
  - (iii) Recognize project as an **ongoing hard effort**, not a “once and done” big push, with moderate level of **additional** investment.

# Job Description for DL Leader - *Key Points*

- Specialized Experience - 10+ years
- Librarianship Expertise: Curation, Digital presentation, Metadata and ontologies, Scientific publication, user interaction
- Management experience: working with and organizing teams, including direct reports, peers and other contributors
- Experience with vendor management and
- Experience with volunteers and experts - getting value, maintaining relations
- Experience with process definition, management, monitoring
- Experience with software-system development and use
- Execution and analytic and skills
- ***Conduct a broad search (internal/external) using a search firm***

## Organization *Strawman* Proposal

- New Digital Library ACM HQ unit
- Head as described - librarianship, management, etc
- Should either include or have direct control of resources for: Metadata, Quality control, User experience
- **Formalized responsibility agreements** with IS, Publications, multiple vendors

## Structural Recommendations, II

(b) Develop a **formal DL scope and functionality requirements** document (not just a list of high-level goals or vendor-contract specs) and also **metrics, performance, and quality expectations**.

(i) **Consultation and broad buy-in**: E.g., EC, Pubs Board, SGB, authors, librarians, VPs of Research: (1) Needs **active** volunteer oversight, **independent** of current boards; (2) Needs **External Advisory Board** of customers

(ii) **Formal process** for periodical revision of requirements

(iii) **Formal process** for regular monitoring of accuracy, quality, and performance

(iv) **Regular reporting to EC and Council** on monitored performance and quality metrics

## Structural Recommendations, III

- (c) Develop robust ongoing **quality-assurance process**:
  - (i) PTF came across **many instances** of data/metadata incompleteness, inconsistency, incorrectness -- examples that are often symptomatic of ***critical underlying issues***.
  - (ii) **Robust QA loop**: error reporting and deep root-cause analysis for errors that cuts across multiple instances
    - (1) **Easy-to-use** error reporting mechanism
  - (iv) Develop appropriate processes, e.g., *author identity*
  - (v) Requires investment in **QA tools** (collab. w. **DBLP**)
  - (vi) Requires investment in **QA personnel**

# Metrics, performance, and quality expectations, I

## 1. Examples of data issues:

- a. Different order of authors on articles and in DL metadata  
<https://dl.acm.org/doi/abs/10.1145/2046556.2046580>
- b. Articles incorrectly dated: <https://dl.acm.org/doi/10.1145/3410470>
- c. Delay in articles showing up in the DL
- d. Missing DOIs due to registration failure
- e. Incorrect page numbering
- f. Author-name misspelling (inconsistencies, accent marks, etc.)
- g. Author-name conflation

1. Compliance with identification standards: ORCID, Crossref, Ror, etc.

2. Develop consistent policies for

- a. Person profiles (see below)
- b. Author-name changes (disparate impact)
- c. Conference-name changes
- d. Naming editors of conference proceedings

## Metrics, performance, and quality expectations, II

**DL person profiles** should summarize **all public information** (known to ACM) about an individual - Examples of issues:

1. DL page of [Vinton Cerf](#) does not mention that he was ACM President
2. [Vivek Sarkar](#) got the 2020 [Ken Kennedy Award](#) - these ACM web pages are not connected
3. The “People of the ACM” entries are not linked to the DL - e.g. Koji Yatani : [DL](#) - [PACM](#)
4. DL PPs do not mention any activity as editor, PC chair/member, or many other roles - e.g. [JACM](#) EiC [Éva Tardos](#)
5. Homonyms and synonyms are a difficult matter, but there are many ways to improve - e.g. [M. Dunham](#) and [M. Eich](#) is the same person

In addition DL PPs should contain more links to other sites: homepages, ORCID, Wikipedia, IEEE, Google Scholar, dblp, ...

## Metrics, performance, and quality expectations, III

### Examples:

1. Uptime of platform
2. Dashboard for error reports and resolution
3. Response time for error reports
4. Time for metadata to appear in DL after publication
5. Time for various forms of data to appear in DL after publication
6. Time for corrections to be applied
7. Measures of correctness / completeness of metadata, conference information, journal references



# Volunteer Oversight

- This is a **major project reboot**.
- Current structures haven't worked well enough - **so start afresh**.
- **Proposal: Create a new DL Board**, separate from SGB, Pubs, Ed though, perhaps, sharing members
- Need an **activist energetic respected moderately senior chair**
  - Someone who cares about the DL passionately and personally
  - Someone who can help recruit a Board of similar people
  - Someone who can spend time contributing and working with Boards and HQ
- Board responsibilities
  - Provide input on functionality, features, and deliverable priorities
  - Help the EC monitor progress and quality
  - Advocate for end users (authors and other stakeholders)
  - Provide guidance on new features, new tools, new needs and possibilities

## Project Requirements I

ACM needs to find a way to shift from its intense focus on **platform migration** (which was necessary) to creating a fresh vision for the **future ACM DL**.

### Need to answer:

- Who are the current customers and how can they be served better?
- Who are potential new **customers** and what capabilities would attract them to the DL?
- What new functionality should be supported?
- What new content/data/metadata need to be supported?
- What new statistics and other derived data should be supported?
- What **should not** be supported any more?

## Project Requirements II

- Formal documentation and definition of the lifecycle of **content, data and metadata, workflow, and architecture**
- **Tracking of entities** - authors, organizations, funders, papers and associated objects (talks, slides, demos, etc.), conference/journal, SIG, etc.
- Focus on **data/metadata accuracy** and iterative improvement (tracking, reporting, generalizing and fixing problems)
- Focus on **specific statistics and data** that help different stakeholders - authors, libraries, VPs/Dirs of Research, funders
- Develop **management metrics** relating to accuracy, usability, productivity, etc.; **summary reports to CEO, DL Board, EC**

## Project Requirements III

- Easy way for users of all kinds to **report** inaccuracies and receive feedback on resolution (beyond platform vendor's bug tracking)
- **Prioritize efforts** that are most relevant to stakeholders
- Staff with **algorithmic mindset** to generalize from each error report to other errors of the same kind
- Staff to work with both the platform vendor and within ACM to **track generalized errors to resolution**
- Thorough **tracking and reporting** of error categories and resolutions including a summary dashboard that is widely visible

## Project Requirements IV

1. Compare *ACM Guide to the Computing Literature* (AGCL) to Google Scholar and Microsoft Academic Academic
2. Compare AGCL to DBLP
3. Questions to answer:
  - a. What is the competitive advantage of AGCL?
  - b. What is the unique value of AGCL?
  - c. What should be the strategic focus of AGCL?

## Path Forward

*The ACM-DL PTF recommends that ACM EC asks ACM's CEO and COO to promptly develop and submit to the consideration of EC a formal proposal for a DL 2.0 project/organization consistent with the findings and conclusions of this report.*

## Immediate Step

- It will take time to implement the sort of changes we recommend
  - Hiring a qualified individual
  - Restructuring HQ and volunteer organizations
  - Management and accountability decisions
  - This could take 6-12 months of decision making and recruiting
- **Top immediate priority:** with CEO-focal accountability with metrics of success - QA
  - Clear mechanisms for feedback (bugs, errors, etc.)
  - Closed-loop process (feedback, measurements, improvement)
  - Metadata accuracy
  - Unique names (authors, institutions, etc.) and their management
  - Handling of new forms of content in quantity - needing storage and metadata management

## Another Immediate Step: ACM/DBLP working group

- **Initial goal:** understand and compare details of data modeling and processes of both systems -- prerequisite to transfer algorithmic ideas between DBLP and ACM DL.
- **Suggested Working Group:** ACM -- Wayne Graves and Asad Ali, DBLP -- Michael Ley and Florian Reitz and/or Marcel Ackermann (expand later as appropriate).
- **Longer-term Goal:** Develop ideas how to **improve processes** for higher data quality in **both** ACM-DL and DBLP; explore **strategic relationship** between ACM DL and DBLP



## Important Step: A New DL Strategic Plan

One of the first responsibilities of the new head of the DL is to formulate a DL **strategic plan** that addresses Project Requirements I-IV, and get formal agreement on that plan from ACM Council.

PTF offers to assist in that effort.