OAI-PMH and the Peer-Review Process

Marko Antonio Rodriguez
Digital Library Research & Prototyping Team
Los Alamos National Laboratory
&
Center for Evolution, Complexity, and Cognition
Vrije Universiteit Brussel
&
Computer Science Department
University of California at Santa Cruz

email: marko@lanl.gov  . URL: http://www.soe.ucsc.edu/~okram
Purpose of this Talk

- Describe a peer-review model that revolves around OAI repositories.
- The model removes the need for editors and publishers in scholarly communication.
- The only human components are authors and referees.
- The model can be implemented as a OAI service-provider.
- The peer-review service is able to solicit referees, aggregate referee evaluations, and generate peer-review metadata for the resource’s metadata record. (the editor’s role)
- The OAI repository provides the information dissemination infrastructure. (the publisher’s role)
Overview of the Current Peer-Review Model

- Researcher writes a manuscript they feel is worthy of publishing.
- Researcher submits manuscript to a journal editor.
- Journal editor pre-filters the manuscript (within scope of journal, well written, etc.)
- Journal editor locates experts in the domain to review the manuscript.
- Referees accept/reject/comment on the manuscript and return reviews to the journal editor.
- Editor accepts/rejects the manuscript (or revision loop).
- Accepted manuscript is published in journal.
Overview of the Proposed OAI Peer-Review Model

• Individual writes a manuscript they feel is worthy of publishing.

• Individual submits manuscript to OAI repository.

• Peer-review service-provider harvests those e-manuscripts that are worthy of review (i.e. high usage stats, high citation stats, no Journal-Ref, within a certain ACM classification, author requested, community requested, etc.).

• Peer-review service locates experts in the domain to review the manuscript.

• Referees review the manuscript and provide an evaluation by way of an online interface.

• Peer-review service aggregates referee scores and generates the manuscripts peer-review metadata.

• OAI repository provides the manuscript and its associated peer-review metadata to the public.
Overview of the OAI Peer-Review Architecture

1. Harvest Pre-Print Record Metadata
2. Solicit Reviewers
3. Submit Reviews
4. Store Review Metadata
5. Augment Pre-Print Metadata
Co-Authorship Networks as a Model of Expertise

- Agent-based Models
- Mathematical Ecology
- Statistical Physics
- Structure of RNA

- Herbert Van de Sompel
- Richard Luce
- Xioming Liu
- Somasekhar Vemulapeli
- Linn Marks
- Johan Bollen
- Michael L. Nelson
- Jeremy Hussell
- Xionming Liu
Automatic Solicitation of Referees

Harvested Pre-Print (or citation metadata if available)

References
S1 & S3, “APaper”, *AConference*, 2005
DEMO

- [http://127.0.0.1:8080/peerper/](http://127.0.0.1:8080/peerper/)
Referee Influence for this talks associated Manuscript

Rodriguez, M.A., Bollen, J., Van de Sompel, H.,

<table>
<thead>
<tr>
<th>Referee Name</th>
<th>Influence</th>
<th>Recent Interests Related to Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sompel, HV</td>
<td>0.09844</td>
<td>OAI-PMH and Co-Authorship Networks</td>
</tr>
<tr>
<td>Bollen, J.</td>
<td>0.08594</td>
<td>Digital-Libraries and Network-Based Impact Metrics</td>
</tr>
<tr>
<td>Carr, L.</td>
<td>0.08516</td>
<td>Digital-Libraries and Open Archive Services</td>
</tr>
<tr>
<td>Hall, W.</td>
<td>0.08066</td>
<td>Knowledge Management and Digital-Libraries</td>
</tr>
<tr>
<td>Rocha, L.M.</td>
<td>0.07892</td>
<td>Document Recommendation Systems</td>
</tr>
<tr>
<td>Lagoze, C.</td>
<td>0.05328</td>
<td>OAI-PMH and Digital-Library Architectures</td>
</tr>
<tr>
<td>Harnad, S.</td>
<td>0.04883</td>
<td>Open Citation Linking and Digital-Library Architectures</td>
</tr>
<tr>
<td>Hitchcock, S.</td>
<td>0.04177</td>
<td>Electronic Journals and Citation Linking</td>
</tr>
<tr>
<td>Blake, M.</td>
<td>0.04156</td>
<td>OAI Repositories and Citation Linking</td>
</tr>
<tr>
<td>Jiao, Z.</td>
<td>0.03386</td>
<td>E-Print Services</td>
</tr>
<tr>
<td>Bergmark, D.</td>
<td>0.03262</td>
<td>Digital-Libraries and OAI-PMH</td>
</tr>
<tr>
<td>Miles-Board, T.</td>
<td>0.02049</td>
<td>Digital-Libraries</td>
</tr>
<tr>
<td>Davis, H.C.</td>
<td>0.01211</td>
<td>Digital-Libraries and Adaptive Linking</td>
</tr>
<tr>
<td>Roure, D.D.</td>
<td>0.01125</td>
<td>Dissemination of Scientific Information Services</td>
</tr>
<tr>
<td>French, J.C.</td>
<td>0.01081</td>
<td>Digital-Library Distributed Searching and Interfaces</td>
</tr>
<tr>
<td>Bailey, C.</td>
<td>0.01043</td>
<td>Digital-Libraries and Distributed Media</td>
</tr>
<tr>
<td>Brody, T.</td>
<td>0.00986</td>
<td>CiteBase and Open Citation Linking</td>
</tr>
</tbody>
</table>
Using Real Peer-Review Bid Data to Validate Algorithm

- Received Bid Data and Submission Archive

- Referees bid on papers according to this scale:
  - 0: did not provide bid data
  - 1: expert in domain and wants to review paper
  - 2: expert in domain and doesn’t care to review paper
  - 3: non-expert
  - 4: conflict of interest

<table>
<thead>
<tr>
<th>Sub</th>
<th>Ref1</th>
<th>Ref2</th>
<th>Ref3</th>
<th>Ref4</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Optimal Referee Inclusion Value

\[ 0.0 = \text{G4} \sim \text{G3} \ll \text{G2} \sim \text{G1} = 1.0 \]
Co-Authorship Relative Rank vs. Referee Similarity Matrix

Degrees of freedom = 2399
p < 2.2^{-16}
Pearson Correlation of 0.383

• Therefore DBLP Co-Authorship Network is correlated with the bidding behavior of the referee similarity matrix.

• Both represent a similar aspect of the scientific community: namely the relative expertise of scientists.

Results of a Energy Distribution within the DBLP Co-Authorship Network

Total Energy for each Group

Normalized by population
Automatic Solicitation of Referees

Harvested Pre-Print

Author

References

S1 & S3, “APaper”, AConference, 2005

S1
S2
S3
S4
S5
S6
S7
S8
The Inclusion of Negative Energy to Curtail Conflict-of-Interest Scenarios
The Inclusion of Negative Energy to Curtail Conflict-of-Interest Scenarios
Energy Distribution Amongst the 4 Groups (k=0)

(1) expert wanting to review (k=0)

(2) expert (k=0)

(3) non-expert (k=0)

(4) conflict of interest (k=0)
Energy Distribution Amongst the 4 Groups (k=2)

(1) expert wanting to review (k=2)
(2) expert (k=2)
(3) non-expert (k=2)
(4) conflict of interest (k=2)
Overview of the OAI Peer-Review Architecture

1. Harvest Pre-Print Record Metadata
2. Solicit Reviewers
3. Submit Reviews
4. Store Review Metadata
5. Augment Pre-Print Metadata
3 Proposed Metadata Tags for the Pre-Print’s Metadata Record

• <pr:review>
  o <pr:referee>
    - <pr:comment>
Peer-Review Metadata

Determined by co-authorship network algorithm

Subjective evaluation specified by referee

<pr:referee name="Heylighen, Francis" influence="0.076" evaluation="0.65" />
<pr:comment date="2005-11-30">
  Your description of the ‘particle-swarm’ algorithm is not well explained. Your math formalisms are not clear and the overall subsection is poorly organized.
</pr:comment>

Referee’s comments on the author’s manuscript
Peer-Review Metadata

http://peer.review.service.org/oai2?
verb=GetRecord&identifier=oai:arXiv.org:cs/0504084&metadataPrefix=pr

<record>
  <header>
    <identifier>oai:arXiv.org:cs/0504084</identifier>
    <datestamp>2005-04-24</datestamp>
    <setSpec>cs</setSpec>
  </header>
  <metadata>
    <pr:review evaluation="0.755" stability="0.50">
      <pr:referee name="Heylighen, Francis" influence="0.076" evaluation="0.65" />
      <pr:comment date="2005-11-30">
        Your description of the ‘particle-swarm’ algorithm is not well explained.
        Your math formalisms are not clear and the overall subsection is poorly organized.
      </pr:comment>
    </pr:referee>
    <pr:referee>
      ...
    </pr:referee>
  </metadata>
</record>
Evaluation and Stability Metadata Scores

-Evaluation:

\[ E = \frac{\sum_{x=0}^{\mid A \mid} \text{inf}(n_x) \cdot \text{eval}(n_x)}{\sum_{x=0}^{\mid A \mid} \text{inf}(n_x)} \]

* Simple average of the evaluations of all participating referees.

-Stability:

\[ S = \sum_{x=0}^{\mid A \mid} \text{inf}(n_x) \]

* Stability allows the community to know how much of the reviewer influence has been associated with an evaluation.
So?

• The separation between certification and dissemination.

• Scholarly communication process solely mediated by the scholarly community. No third part intervention.

• A quantitative representation of the peer-review process. Therefore, the peer-review process can become the object of scientific inquiry.

• In combination with OAI repositories, a publication model that has limited monetary overhead.
Questions?

Refer to heading.