Open Scholarly Infrastructure and Community of Practice
オープンな学術情報インフラと実践コミュニティ

05 August 2018
Nobuko Miyairi

http://orcid.org/0000-0002-3229-5662
@NobukoMiyairi
COI disclosure

I am a freelance consultant/analyst, based in Tokyo, Japan, providing scholarly communications consultancy for academic societies, research institutions, scholarly publishers and solution vendors.

I have no actual or potential conflict of interest in relation to organizations and/or services referred in this presentation.

The opinions and views expressed in this presentation and on the following slides, unless otherwise credited, are solely those of the presenter.
Affiliations

University of Hawai‘i at Mānoa (2000-2002)
Thomson Scientific / Thomson Reuters (2003-2012)
ORCID, Inc. (2015-2018 April)
Freelance (2018 May – present)

University of Tsukuba / Expert Member, International Strategy Committee (2014-2016, 2018-present)
National Institute for Materials Science / Expert Member, Digital Library Committee (2016-present)
National Institute of Science and Technology Policy / Affiliated Fellow (2018-present)
infrastructure

noun  |  in-fra-struc-ture  |  \in-frə-,strək-cher, -frä-

1 : the underlying foundation or basic framework (as of a system or organization)

2 : the permanent installations required for military purposes

3 : the system of public works of a country, state, or region; also : the resources (such as personnel, buildings, or equipment) required for an activity

https://www.merriam-webster.com/dictionary/infrastructure
Elsevier to acquire Aries Systems, a best-in-class #publication workflow #solutions provider bit.ly/2vyfIrQ @AriesMarketing

8:00 AM - 2 Aug 2018

Heather Joseph
@hjoseph

Bad news for Open infrastructure.

Elsevier to Acquire Aries Systems, A Best-in-Class #publication workflow #solutions provider bit.ly/2vyfIrQ @AriesMarketing

12:06 AM - 3 Aug 2018

Margo Bargheer
@margoline

For Elsevier good news, but not for the publishing ecosystem. We need more variety and resilience, not less. Good there is OJS etc. Any commercial publisher can use open source, no matter if they do open or toll access to avoid being "customer" and competitor to Elsevier at once.

Elsevier News @ElsevierNews
Elsevier to acquire Aries Systems, a best-in-class #publication workflow #solutions provider bit.ly/2vyfIrQ @AriesMarketing

4:20 PM - 3 Aug 2018
Citation Infrastructure

- Governance
- Workflow / System / Platform
- Metadata / Conventions / Guidelines
- Contents
Setting your cites on open
The Initiative for Open Citations

Mark Patterson • Dario Taraborelli
Crossref Live • Singapore, November 2017
The Initiative for Open Citations was publicly launched one year ago! To mark the progress that's been made, and to build new momentum, we are designating April 2018 #OpenCitationsMonth - i4oc.org/news.html#April...
With 13 out of 20 biggest publishers, nearly a half of citation data is open. Looking forward to even more! @i4oc_org #CRLIVE17

This is very exciting. Who are the remaining 7?
Citation Infrastructure

Contents

Metadata / Conventions / Guidelines

Workflow / System / Platform

Governance

Crossref + publishers

WEB OF SCIENCE™
Scopus®
Clarivate Analytics

ELSEVIER

? I40C
MEMO FROM
J. LEDERBERG
GENETICS DEPARTMENT
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

May 9, 1959.

Since you first published your scheme for a "citations index" in Science about 4 years ago, I have been thinking very seriously about it, and must admit I am completely sold. In the nature of my work I have to spend a fair amount of effort reviewing the literature of collateral fields, and it is irksome how often I have been stumped in trying to update a topic, where your scheme would have been just the solution. I don’t care that your critics haven’t grasped the idea, especially the print that the author must learn to cooperate by his own choice of citations, rather than he does the critical work.

Have you tried to get this out in any adequate experiment? Would you look for support from the NSF? Of course you have to count on permission from the established outfits, which have already succeeded in blocking any program aimed at the commercial translation of the reference indexes.

Sincerely,

[Signature]
What it takes to create open scholarly infrastructures?
Example: DOI
Persistent Identifiers

A persistent identifier is a long-lasting reference to a digital resource. Typically it has two components: a unique identifier; and a service that locates the resource over time even when it's location changes. The first helps to ensure the provenance of a digital resource (that it is what it purports to be), whilst the second will ensure that the identifier resolves to the correct current location.

• Digital Object Identifier (DOI)
• Handle
• Archival Resource Key (ARK)
• Persistent Uniform Resource Locator (PURL)
• Universal Resource Name (URN)

Which one(s) to use?

Crossref DOIs

Registered content

Crossref Members

Voting members

### Revenue growth by year

(Millions)

<table>
<thead>
<tr>
<th>Year</th>
<th>Revenue Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>1M</td>
</tr>
<tr>
<td>2003</td>
<td>2M</td>
</tr>
<tr>
<td>2004</td>
<td>3M</td>
</tr>
<tr>
<td>2005</td>
<td>4M</td>
</tr>
<tr>
<td>2006</td>
<td>5M</td>
</tr>
<tr>
<td>2007</td>
<td>6M</td>
</tr>
<tr>
<td>2008</td>
<td>7M</td>
</tr>
</tbody>
</table>

### Statement of activities

Years Ended December 31, 2016 and 2015

<table>
<thead>
<tr>
<th></th>
<th>2016</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unrestricted Net Assets:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revenue and support:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deposit fees</td>
<td>$4,426,246</td>
<td>$4,061,633</td>
</tr>
<tr>
<td>Member fees</td>
<td>2,684,468</td>
<td>2,588,708</td>
</tr>
<tr>
<td>Interest income</td>
<td>22,818</td>
<td>23,368</td>
</tr>
<tr>
<td>Investment return</td>
<td>36,220</td>
<td>18,690</td>
</tr>
<tr>
<td><strong>Total revenue and support</strong></td>
<td>$7,169,752</td>
<td>$6,692,399</td>
</tr>
<tr>
<td><strong>Expenses:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salaries, taxes and benefits</td>
<td>3,835,082</td>
<td>3,362,961</td>
</tr>
<tr>
<td>Travel and entertainment</td>
<td>625,968</td>
<td>626,632</td>
</tr>
<tr>
<td>Data center</td>
<td>376,679</td>
<td>366,128</td>
</tr>
<tr>
<td>Advertising and marketing</td>
<td>343,221</td>
<td>119,037</td>
</tr>
<tr>
<td>Other general and administration expenses</td>
<td>298,523</td>
<td>181,466</td>
</tr>
<tr>
<td>Registration fees, related party (Note 3)</td>
<td>263,117</td>
<td>258,260</td>
</tr>
<tr>
<td>Rent</td>
<td>184,254</td>
<td>204,798</td>
</tr>
<tr>
<td>Professional fees</td>
<td>179,875</td>
<td>261,865</td>
</tr>
<tr>
<td>Depreciation</td>
<td>160,056</td>
<td>226,482</td>
</tr>
<tr>
<td>Consulting</td>
<td>71,052</td>
<td>87,211</td>
</tr>
<tr>
<td>Dues and subscriptions</td>
<td>54,499</td>
<td>43,032</td>
</tr>
<tr>
<td>Product development</td>
<td>33,019</td>
<td>-</td>
</tr>
<tr>
<td>Insurance</td>
<td>25,642</td>
<td>29,405</td>
</tr>
<tr>
<td>Bad debt expense</td>
<td>24,353</td>
<td>24,843</td>
</tr>
<tr>
<td>Program initiatives</td>
<td>7,393</td>
<td>18,568</td>
</tr>
<tr>
<td><strong>Total expenses</strong></td>
<td>$6,482,733</td>
<td>$5,810,688</td>
</tr>
<tr>
<td><strong>Change in net assets from operations</strong></td>
<td>$687,019</td>
<td>$881,711</td>
</tr>
<tr>
<td><strong>Other expenses:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign currency exchange loss, net</td>
<td>117,453</td>
<td>35,721</td>
</tr>
<tr>
<td>Loss on disposal of property and equipment</td>
<td>15,725</td>
<td>4,602</td>
</tr>
<tr>
<td><strong>Total other expenses</strong></td>
<td>$133,178</td>
<td>$40,323</td>
</tr>
<tr>
<td><strong>Change in total net assets</strong></td>
<td>$553,841</td>
<td>$841,388</td>
</tr>
<tr>
<td><strong>Net assets, beginning of year</strong></td>
<td>$6,634,308</td>
<td>$5,792,920</td>
</tr>
<tr>
<td><strong>Net assets, end of year</strong></td>
<td><strong>$7,188,149</strong></td>
<td><strong>$6,634,308</strong></td>
</tr>
</tbody>
</table>
Sustainable infrastructures require collaborations
A radically open approach to developing infrastructure for Open Science

Paul Peters  October 23rd, 2017

https://about.hindawi.com/opinion/a-radically-open-approach-to-developing-infrastructure-for-open-science/
Providing open scholarly infrastructure is likely to be more challenging than providing Open Access to scholarly articles. The primary challenge in moving towards Open Access publication models has been a reorganization of how publication costs are paid, rather than a fundamentally new approach to scholarly publication. In contrast, open scholarly infrastructure will require completely new models of interaction between commercial companies (publishers, technology providers, data aggregators), non-profit organizations, and the research community. Developing open infrastructure for the creation, dissemination, and assessment of scholarly outputs will require parties with disparate incentives to work together to overcome difficult challenges.
Example: ORCID
…an author ID system proposed earlier this month and backed by 23 organizations, including Thomson Reuters, Nature Publishing Group, Elsevier, ProQuest, Springer, CrossRef, the British Library and the Wellcome Trust.
ORCID live iDs

https://support.orcid.org/knowledgebase/articles/150557-number-of-orcid-ids
Community Distribution

- Europe 48.4%
- USA/Canada 26.68%
- Asia Pacific 19.33%
- M East & Africa 3.51%
- Latin America 2.08%

Member organizations by region

- Research institute 73%
- Publisher 8.63%
- Repository/profile 6.07%
- Association 4.47%
- Funder 3.67%
- Other 4.36%

Member organizations by sector

ORCID Annual Report 2017. ORCID.
https://doi.org/10.23640/07243.5950318.v1
ORCID publisher mandates

More and more publishers are providing options to link authors’ and reviewers’ ORCID in manuscript submission systems. More than 1,800 journals already require ORCID.

http://orcid.org/content/requiring-orcid-publication-workflows-open-letter
Things – DOI
People – ORCID
Organizations - ???
Organization ID?

Content identifiers:
- DataCite
- Crossref

Organisation identifiers:
- ISNI?
- Ringgold?
- FundRef?
- GRID?
- OrgRef?
- Open Corporates?
- LEI?

Contributor identifiers:
- ORCID

Adapted from: Geoffrey Bilder. The OI Project. [https://www.slideshare.net/CrossRef/the-oi-project-geoffrey-bilder](https://www.slideshare.net/CrossRef/the-oi-project-geoffrey-bilder)
Org ID: Framing Principles

1. The registry will support reliable, open, permanent, and unambiguous identification of organizations with whom researchers are affiliated.
2. The registry will not be limited by institutional, stakeholder, corporate entity type, business, geographic, or national boundaries.
3. Participation in the registry and access to its services will be based on transparent and non-discriminatory terms of use.
4. The identifier and metadata supporting the disambiguation of organizations will be available as Open Data using a CC-0 license.
5. The registry will have a sustainable business model based on revenue generation consistent with its mission to ensure its long-term openness, persistence, and reliability.
6. All software developed for the registry will be publicly released under an Open Source Software license approved by the Open Source Initiative.
7. For the software it adopts, the initiative will prefer Open Source.
8. The registry will be governed by representatives from a broad cross-section of stakeholders drawn from the community it serves, the majority of whom represent not-for-profit organizations, and will strive for maximal transparency by publicly posting summaries of all meetings and provide financial reports. The registry will have an explicit living will and, if acquired, it will continue to operate under these principles.

https://orcid.org/content/organization-identifier-working-group
Org ID: a recap and a hint of things to come

Cross-posted on the blogs of University of California (UC3), ORCID, and DataCite: https://doi.org/10.5438/67sj-4y05. Over the past couple of years, a group of organizations with a shared purpose—California Digital Library, Crossref, DataCite, and ORCID—invested our time and energy into launching the Org ID initiative, with the goal of defining requirements for an open, community-led organization identifier registry. The goal of our initiative has been to offer a transparent, accessible process that builds a better system for all of our communities.

https://doi.org/10.5438/67sj-4y05
Membership

Working Group members were selected to ensure a broad range of experience and perspectives, including expertise in research data discovery, data management, persistent identifiers, economics research, funding, archiving, non-profit membership organizations, academia, publishing, and metadata development.

- Christopher Brown, Jisc
- John Chodacki, California Digital Library (Chair)
- Laura Cox, Ringgold
- Trisha Cruse, DataCite (Steering Committee)
- Tim Devenport, Editeur
- Mike Frame, U.S. Geological Survey
- Danny Goroff, Alfred P. Sloan Foundation
- Laure Haak, ORCID (Steering Committee)
- Christina Hoppermann, Springer Nature
- Andres Mori, Digital Science
- Ed Pentz, Crossref (Steering Committee)
- Paul Peters, Hindawi Publishing
- Andrew Pitts, The IP Registry
- Erin Robinson, Foundation for Earth Sciences
- Arthur Smith, American Physical Society
- Thomas Vestdam, Elsevier
- Simeon Warner, Cornell University

https://orcid.org/content/organization-identifier-working-group
#UKSG18 your second and last chance to join breakout group A on #Onyar the upcoming #Org_ID registry from Crossref, @datacite, & @ORCID_Org. @epentz leading the session, starting after the break. Don't say we didn't warn you.

6:45 PM - 10 Apr 2018
PIDapalooza will bring together creators and users of persistent identifiers (PIDs) from around the world to shape the future research information landscape.

Conference website: https://pidapalooza.org
Presentations repository: https://pidapalooza.figshare.com/
Who pays for scholarly infrastructures?
Principles for Open Scholarly Infrastructures


infrastructure (noun) – the basic physical and organizational structures and facilities (e.g. buildings, roads, power supplies) needed for the operation of a society or enterprise. – New Oxford American Dictionary

Everything we have gained by opening content and data will be under threat if we allow the enclosure of scholarly infrastructures. We propose a set of principles by which Open Infrastructures to support the research community could be run and sustained. – Geoffrey Bilder, Jennifer Lin, Cameron Neylon

https://doi.org/10.6084/m9.figshare.1314859.v1
When we wrote the Infrastructure Principles we published some weeks back, ORCID was at the centre of our thinking, both as one of the best examples of good governance practice and as an infrastructure that needs sustaining... Publishers get a lot of stick when it comes to demanding money, but when it comes to community initiatives it is generally publisher that put up the initial funding. This has definitely been the case with ORCID, with funders and institutions falling visibly behind, apparently assuming others will get things moving.

Communities pay and support scholarly infrastructures
Principles for Open Scholarly Infrastructure

Design principles for successful shared infrastructure

“What should a shared infrastructure look like? Infrastructure at its best is invisible… If successful, it is stable and sustainable. Above all, it is trusted and relied on by the broad community it serves.”

✓ Governance
✓ Sustainability
✓ Insurance

https://doi.org/10.6084/m9.figshare.1314859.v1
<table>
<thead>
<tr>
<th>Governance</th>
<th>Sustainability</th>
<th>Insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Coverage across the research enterprise</td>
<td>• Time-limited funds are used only for time-limited activities</td>
<td>• Open source</td>
</tr>
<tr>
<td>• Stakeholder governed</td>
<td>• Goal to generate surplus</td>
<td>• Open data (within constraints of privacy laws)</td>
</tr>
<tr>
<td>• Non-discriminatory membership</td>
<td>• Goal to create contingency fund to support operations for 12 months</td>
<td>• Available data (within constraints of privacy laws)</td>
</tr>
<tr>
<td>• Transparent operations</td>
<td>• Mission-consistent revenue generation</td>
<td>• Patent non-assertion</td>
</tr>
<tr>
<td>• Cannot lobby</td>
<td>• Revenue based on services, not data</td>
<td></td>
</tr>
<tr>
<td>• Living will</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Formal incentives to fulfil mission &amp; wind-down</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Community of Practice

...a community of practice is a group of individuals participating in communal activity, and experiencing/continuously creating their shared identity through engaging in and contributing to the practices of their communities.

**Domain:** A domain of knowledge creates common ground, inspires members to participate, guides their learning and gives meaning to their actions.

**Community:** The notion of a community creates the social fabric for that learning. A strong community fosters interactions and encourages a willingness to share ideas.

**Practice:** While the domain provides the general area of interest for the community, the practice is the specific focus around which the community develops, shares and maintains its core of knowledge.

実践コミュニティ

あるテーマについて関心や問題、熱意などを共有し、その分野の知識や技能を、持続的な相互交流を通じて深めていく人々の集団を「実践コミュニティ（Community of Practice）」と言う。ウェンガー, E（米）が1991年、徒弟制度の観察から導き出した概念。

要件として以下の3つが重要とされる。
①＜領域＞熱意をもって取り組む専門知識の分野があること
②＜コミュニティ＞人が同じ関心や熱心さでつながれる場
③＜実践（practice）＞かかわり合いの中でなされた活動
Community of Practice (not only Interest)
THANKS!

Nobuko Miyairi
nobuko@miyairi.info