Open Access Articles Reaching 50% But Their Retrieval Is Lagging

Xiaotian Chen
Electronic Services Librarian
Bradley University
Oct 30, 2014

This Article Can Be Found for Free on the Web, But Library Indexes Cannot Tell Us That

Bomar et al. (2010). “Unconventional Ubiquitin Recognition by the Ubiquitin-Binding Motif within the Y Family DNA Polymerases ι and Rev1,” *Molecular Cell*. 37(3): 408-17.

PubMed searchers get this free indicator: 

Google Scholar searchers get this free indicator:

[HTML] from sciencedirect.com

Library indexes searchers, however, will either be told that full text is not available or get a kind of document delivery link, unless their library has subscription to *Molecular Cell*, an Elsevier journal.

And the Bigger Issue Is:

There are millions of scholarly journal articles like that *Molecular Cell* article. They are free on the Web but library indexes keep that secret from library patrons.

That Is Why We Are Here Discussing: Open Access Articles Reaching 50% But Their Retrieval Is Lagging

- Chen, Xiaotian. (2014). Open Access in 2013: Reaching the 50% Milestone. *Serials Review*. 40 (1), 21-27. (Selected by the July/August 2014 issue of The INFORMED LIBRARIAN ONLINE as one of the "Featured Articles.")
- Chen, Xiaotian. (2013). Journal Articles Retrieval in an Age of Open Access. *The Journal of Web Librarianship*. 7 (3), 243-254.

The Purposes of the Above 2 Research Projects

- Assess the current status of OA;
- Assess OA retrieval by free and subscription-based indexes.

The Scope

The focus here is on scholarly journal articles.

Other kinds of OA publications, such as theses and dissertations, are not included in the discussion here.

OA Definitions/Classifications

OA can be roughly divided into Gold OA and Green OA. Besides Gold and Green, other categories may include “delayed OA”. But Gold and Green are the 2 primary classifications. The definitions of Gold and Green vary and there is some gray area.

1. Clearly Gold OA: Articles on OA journals Web sites;
2. Clearly Green OA: Articles published by traditional journals. They need subscription/purchase to get from publishers, but free from IRs, authors’ sites and other sites.
3. Gray: Like that *Molecular Cell* article. Most articles from the journal need subscription/purchase to read, but some are free from publishers.

Definitions/Classifications

That *Molecular Cell* article can be either Gold or Green, depending on different definitions.

I follow the classifications that Gold OA articles are published in journals that make **all** their articles freely available on the Internet, and other OA articles are Green OA. So that *Molecular Cell* article is Green in today's discussion.

Part I: How Many OA Articles Out There?

Human history is reaching a milestone: Either the percentage of OA articles published in the previous year is very close to 50% or it has passed that mark.

The next slides have the proof.

Empirical Study on the State of OA

- My 2013 empirical study used a 2009 study by Bjork et al. as a base to assess the status of OA articles in 2013.
- Bjork et al. found that 20.4% of the sample articles indexed by Scopus were OA articles, with 8.5% freely available on the publishers' websites and 11.9% freely available on other websites.

2013 OA Findings

- In summer of 2013, 2,655 random sample articles published during the past 12 months were generated from Scopus and were checked on the Internet for free full texts.
- 37.8% of Scopus samples had free full text on the Internet, a significant increase from 20.4% recorded 2009.
- Because Scopus indexes only about a quarter of nearly 10,000 OA journals on DOAJ, it is fairly safe to estimate that we were reaching 50% in 2013.

Findings: OA % at Scopus by Subjects

Scopus Subjects	Sample total	OA total	OA %
Health Sciences	740	341	46.1%
Life Sciences	619	243	39.3%
Physical Sciences	929	299	32.3%
Social Sciences & Humanities	367	120	32.7%
Total/Overall	2,655	1,003	37.8%

Anecdotal Data

My library spends about $8,000/month on non-mediated *Get It Now* service in a month when the semester is in session, averaging about $26/article. Most articles purchased via *Get It Now* are Elsevier and Springer articles.

My colleagues and I found that about 1/3 of the articles purchased can also be found for free.

Where are OA articles Posted?

- OA Journals. DOAJ lists nearly 10,000 OA journals.
- Traditional journals, eg, ScienceDirect journals, with some articles freely available.
- Institutional Repositories, eg, “DSpace@MIT” and “Deep Blue” (at U of Michigan).
- Other archives, such as PMC and arXiv.org
- Authors’ social network site, eg, “ResearchGate”.
- ……

How Big These Archives/Directories?

In 2014,

- PMC has 3.2 million articles;
- DOAJ has 1.7 million articles;
- arXiv.org has nearly one million e-prints;
- ResearchGate has over 3 million registered authors;
- DSpace@MIT has 60,000 items;
- And, there are about 3,400 registered IRs world wide.

Another Way to See OA Growth During the Past 5 Years

- DOAJ journals list more than doubled from 4,000 to 9,900.
- Registered IRs at Registry of Open Access Repositories more than doubled from 1,500 to 3,400.
- ResearchGate, a social networking site for authors, was founded in 2008. In 2013, 3 million authors around the world share their publications and follow research trends and other authors.

Part II: Can Library Indexes Indicate OA Availability?

- Library indexes can indicate full-text availability if the articles are on OA journals and if a library's OpenURL link resolver (eg, SFX) activates the OA journals.
- Library indexes cannot indicate full-text availability if the articles are on IRs, traditional journal sites (eg, ScienceDirect), authors social network sites, or other Web sites.

Proof that Library Databases Do not Know Green OA Articles

- 471 random sample articles were generated in Dec 2012-Jan 2013 from “DSpace@MIT” and “Deep Blue” (at U of Michigan) , which are among the top IRs, based on Ranking Web Repositories-USA (Cybermetrics Lab).
- Because PubMed was used in comparison, the random samples were generated with medical keywords.
- Samples were then searched on Scopus, PubMed, and Google Scholar, to see if there is any free full text availability indication on those databases.

Findings

- 19 of the 471 samples were published in OA journals. Or, they are Gold OA articles.
- Gold OA a lot of fewer than Green OA in this sampling, because Deep Blue started coverage in the 1940s, while OA journal is a fairly new phenomenon.
- Of the 19 Gold OA articles, PubMed and Google can indicate full-text availability for all, so can Scopus (through an OpenURL link resolver).
- The difference lies in the rest of 452 samples, or, Green OA articles.

Google Scholar, PubMed, and Scopus Differ Dramatically in Green OA Indication (Total Green OA Sample Size: 452)

Index name	Green OA samples retrieved with full-text indicators	Green OA samples not found by index
Google Scholar	435 (96%)	2
PubMed	131 (29%)	26
Scopus	0	20

Examples of Google Scholar OA Indicators

- [HTML] from nih.gov
- [PDF] from stanford.edu
- [PDF] from umich.edu
- [PDF] from researchgate.net
- [PDF] from nzma.org.nz
- [PDF] from wiley.com
- [HTML] from sciencemag.org

Examples of PubMed OA Indicators

PubMed can indicate free articles on various archives.

Subscription-based Indexes, However…

are not currently able to provide users with free full-text availability indicators for Green OA articles.

Sad…

PubMed is currently able to provide not only free full-text indicators for Green OA articles archived on PMC but also Green OA articles on some publishers' Web sites, such as Wiley Online Library.

On the other hand, Scopus cannot even indicate free articles on its owner's own ScienceDirect, such as that *Molecular Cell* article, let alone free articles on Wiley, PMC, Dspace@MIT, arXiv.org, ResearchGate, and other Web sites.

Due to this Limit of Library Indexes

- Users of library indexes encounter inconveniences;
- Inter-library loan offices are processing requests of free articles;
- Libraries are paying for free stuff: About 1/3 of the articles my library buys for users via the "*Get It Now*" services can be found for free…

Even Google Naysayers Can't Deny…

Although Google Scholar may have various weaknesses, and some librarians and information professionals still resist it, Google Scholar is the best in indicating OA articles. GS and PubMed are far better than subscription-based indexes in indicating and linking to OA full text. Both GS and PubMed can be configured to link to libraries' OpenURL link resolvers; thus, they have no problem linking to libraries' full-text subscriptions.

IR Managers Know Google Is Doing Better in Retrieving OA Free Texts

Deep Blue makes this statement on why UM faculty should archive their articles to Deep Blue: "Making your work accessible via Deep Blue will ensure more of your peers can find it (in Google Scholar, for example) and will cite it."

DSpace@MIT uses this to persuade researchers to archive their publications on DSpace@MIT: "Get top search results in Google."

By the Way…

- Google Scholar and PubMed are very popular among faculty and students, whether you promote them or not.
- Hightower and Caldwell (2010) found from a survey that they are the most popular free indexes used by science researchers at UC Santa Cruz.
- In 2013, GS beat EBSCO's Academic Search Premier at Bradley as the #1 most popular starting place to search, as an SFX source, based on Bradley's SFX usage stats, even though librarians promote Academic Search Premier/Complete more than any other journal indexes.
- And PubMed is #5 at Bradley. People hardly ever use MedLine, even though MedLine is listed ahead of PubMed in alphabetical order on all library Web pages.

Extra Benefit of Using Google Scholar

- Most library indexes have limited years of coverage and do not index articles published long time ago, much less give full-text indication.
- The oldest sample from Deep Blue was published in 1943 in *Journal of Morphology*. GS could not only retrieve the citation of the article but also displayed the free full-text indicator.
- Scopus and most subscription-based indexes do not even index that 1943 article.

Some Other Interesting Thoughts

- “Open Access will lead to a disaggregation of the journal into its component articles”.
- “It is the individual article that the reader wants, not the journal.”

Lewis, David. 2012. “The Inevitability of Open Access.” *College & Research Libraries*. 73 (5):493–506.

Conclusions

About 50% journal articles published during the past 12 months are freely available on the Internet.

Nearly half of those OA articles are Green OA. There are millions of them on IRs, traditional journal Web sites, authors' social network sites, and other Web sites.

Conclusions

Subscription-based journal indexes and OpenURL link resolvers are, at present, incapable of indicating Green OA full-text availability. They were designed to work in a very closed environment.

Vendors of subscription-based indexes and OpenURL link resolvers should overcome this challenge in order to stay relevant in the age of OA. In the meantime, librarians may want to take advantage of GS and PubMed to better serve library patrons, as well as to save precious budget and staff time.

Questions

- Xiaotian Chen
- chen@bradley.edu
- 309-677-2839

Thank You.