Open access in the spotlight

UKSG eNews 407

20 Oct 2017

Nettie Lagace, NISO, and Christine Stohn, Ex Libris

Open access and nobody knows about it?

In the current scholarly communication environment where there is a burgeoning amount of material available for free under a variety of licence models, it is increasingly important that metadata transmissions to end user discovery interfaces include clear open access information. In the past, open access status could be indicated by the journal in which the article was published, but now – with many hybrid models – it moves to article level and the journal metadata is no longer always a reliable signal. In addition, the discovery and access landscape is complex and diverse with multiple pathways for data to be transmitted. Users search and find the material they need in many different places; in web search engines such as Google or Google Scholar, publisher webpages, databases or library discovery systems, to name just a few. This assortment of discovery and access mechanisms also means that there are many different stakeholders engaged in the exchange of metadata to make it discoverable to users, or to create visibility and increase usage on behalf of authors. Machine readable formats and automated processes are essential for this exchange to be successful, as are industrywide agreed formats, standards and indicators.

Lack of consistent indication means lack of visibility

Open access material is available from a vast array of sources, typically with metadata and licence indicators which may vary in accuracy across these sources. An article may be published as part of a standard subscription journal, where the author paid an article processing charge (APC) for it to be published in open access form. It may be part of an open access journal, or it may be available in an institutional or discipline specific repository such as arXiv. The metadata – often including the link to the full text – is harvested and distributed by search engines such as Google and Google Scholar, by aggregates such as CORE and Bielefeld Academic Search Engine (BASE), and by library discovery systems. Ideally, the result lists from searches in these places indicate whether access to any particular article is available or not, and – unless the users or their institutions have access to a subscription – they link to the free version. In some cases, especially in the case of library discovery systems, the first result list users see is filtered by items available to them. Therefore, items which have incomplete or inaccurate metadata will be flagged incorrectly, and remain hidden. In this scenario, the user is deprived of what could have been a useful search result, and additionally, if the author of the 'hidden' article depends on usage data for career advancement, she or he is deprived of 'hits'. Usage drives citations and other metrics that are important for researchers' careers: visibility and discoverability of research output regardless of where a user searches for material has to be a key concern for all stakeholders. It may not seem that it would be so difficult for a discovery

interface to achieve correct open access indication, but in fact, often there is no indicator available in the harvested metadata at all or the data is incorrect or inconsistent and thus unclear. Where article level data is harvested from a variety of worldwide institutional repositories or where various publishers use different definitions for access, it can become confusing and difficult to handle in the aggregate. Search engines and discovery systems harvest millions of metadata items on a regular basis and are constantly updating their information to ensure user satisfaction; they necessarily rely on accuracy and consistency from the 'upstream' provider, put into machine readable terms for the most efficient processing of this huge amount of available data.

NISO ALI recommendations are closing a gap

The NISO Recommended Practice, 'ALI' [Access and License Indicators (NISO RP-22-2015)](#), can help to resolve these issues. Published in 2015, it addresses issues with confusing and inconsistent metadata for open access indications, and thus also helps to effectively highlight and promote material that is freely available across the many different interfaces where users seek their information. ALI provides and promotes consistent article level descriptors in machine readable format across the industry to allow their use across different systems. A key indicator is the so called 'free-to-read' tag. At the time when it was discussed in the NISO ALI working group, it was decided to not call this an 'open access' indicator since it addresses one specific need: the ability of the user to read the article for free, regardless of any other licence restriction that may apply, for example, reuse rights. ALI does not seek to cover all such restrictions but provides instead two additional tags – one for providing a short licence statement and one for providing a link to the actual licence terms. Overall the reaction to the ALI recommendation was positive, with many stakeholders welcoming the availability of the tags and inspecting them for possible incorporation into their publishing workflows. The ALI tags have since been incorporated into the ANSI/NISO Z39.96 JATS: Journal Article Tag Suite standard for the exchange of journal based content, widely used in the medical and scientific publishing community, and Crossref has begun communicating the licence tag in its metadata, propagated across the industry in many varied datasets and other tools.

Where are we now and what needs still to be done?

Some technical issues remain in the definitions of the tags, especially regarding the date restrictions and the schema definition. Unfortunately, these issues have hampered some take up. Library discovery providers continue to receive open access data indicators in proprietary formats. While this information can still be used to make the content discoverable and link to the open access version, it also causes additional work because the normalisation rules for the data ingestion must be adjusted to every provider. The hope for a unified indicator is to enable all parties involved in the data exchange to automate processes and therefore increase accuracy and coverage of such material. NISO hopes to combine a second stage industry survey to provide input with an updated version of the ALI Recommended Practice. Consideration of future steps would be done by the NISO Information Discovery and Interchange Topic Committee, a leadership group consisting of librarians, publishers, and system providers, in whose 'portfolio' of responsibility the ALI Recommended Practice lies. Feedback is always welcome at nisohq@niso.org.