Response to request for public comments on ‘Open Access Metadata and Indicators’

Dr. Micah Altman
Director of Research; Head/Scientist, Program on Information Research — MIT Library, Massachusetts Institute of Technology
Non-Resident Senior Fellow, The Brookings Institution

Introduction

Thank you for the opportunity to respond to this report. Metadata and indicators for Open Access publications are an area in which standardization would benefit the scholarly community.

As a practicing social scientist, a librarian, and as MIT’s representative to NISO, I have worked extensively with (and contributed to the development of) metadata schemas, open and closed licenses, and open access publications. My contribution is made with this perspective.

Comments

Scope

The scope of the work originally approved by NISO members was to develop metadata and visual indicators that would enable a user to determine whether a specific article is openly accessible, and what other rights were available. The current draft is more limited in in two ways: First the draft does not address visual indicators. Second the metadata field proposed in the draft signals only the license available without providing information on specific usage rights.

The first limitation is a practical limitation of scope. Common, well-structure metadata is a pre-condition for systematic and reliable visual indicators. These indicators may be developed later by NISO or other communities.

The second limitation in scope is more fundamental and problematic. Metadata indicating licenses is less directly actionable. A user can take direct actions (e.g., to read, mine, disseminate, or reuse content) based on knowledge of rights, but cannot take such actions knowing only the URI of the license — unless the user has through independently determined what rights are associated with every license encountered. Moreover, different users may (correctly or incorrectly) interpret the same license as implicating different sets of rights.
This creates both additional effort and risk for users, which greatly limits the potential value of the proposed practice.

The implicit justification for this limitation of scope is not clearly argued. However it seems to be based the claims, made on page 2, that “This is a contentious area where political views on modes of access lead to differing interpretations of what constitutes “open access.” and “Considering the political, legal, and technical issues involved, the working group agreed that a simple approach for transmitting a minimal set of information would be preferred.” [1]

This is a mistake, in my judgment, since contention (or simply uncertainty) over the definition of “open access” notwithstanding, there is a well established and well-defined set of core criteria that apply to open licenses: these include the 10 criteria comprising the Open Source Initiatives Open Source Definition [2]; the 11 criteria comprising the Open Knowledge Foundation Open Definition (many of which are reused from the OSI criteria) [3]; and the four license properties defined by creative commons. [4] These criteria currently are readily applicable to dozens of existing independently-created open licenses [5], which have been applied to millions of works.

Although these properties may not be comprehensive, nor is there a universal agreement over which of these properties constitute “open access”, any plausible definition of Open Access should include at least one of these properties. Thus a metadata schema that could be used to signal these properties is feasible, and could be reliably used to indicate useful rights. These right elements should be added to complement the proposed license reference tag, which could be used to indicate other rights not covered in the schema.

**Design Issues**

The proposed draft lists nine motivating use cases. The general selection of use cases appears appropriate. However, the definition of success for each use case is not clearly defined, making the claim that use cases are satisfied arguable. Moreover, the free_to_read element does adequately address the use cases to which it is a proposed solution.

The free_to_read element is defined as meaning that “content can be read or viewed by any user without payment or authentication” (pg 4) the purpose is to “provide a very simple indication of the status of the content without making statements about any additional re-use rights or restrictions.”

No other formal definition of usage rights or conditions for this element is provided in the draft. Under the stated definition, rights other than reading could be curtailed in any variety
of ways -- including for example a restriction on the right to review, criticize or comment upon the material. Thus the rights implied by the free_to_read element are less than the minimal criteria provided by any plausible open access license.

Furthermore, this element is claimed to comprise part of the solution to compliance evaluation use cases (use cases 5.8 and 5.9 in the draft). It cannot support that purpose -- compliance auditing relies upon well-defined criteria, and the free_to_read definition is fatally ambiguous.

The free_to_read element should be removed from the draft. It should be replaced by a metadata attribute indicating readability as defined by metadata indicating ‘access’ rights, as defined by the open criteria listed above. [6]

Technical Issues

Finally, there are a number of changes to the technical implementation proposed:

- There should be a declared namespace for the proposed XML license elements, so they can be used in a structured way in XML documents without including them separately in multiple schema definitions

- Semantic markup (e.g. RDF) is required so that these elements may be used in non-XML metadata

- A schema should be supplied that formally and unambiguously defines: which elements and attributes are required; which are repeatable; what datatypes (e.g. date formats) are allowable; and any implicit default values

- Make explicit that license_ref may refer to waivers (e.g. CC0) as well as licenses.

NOTES

[1] The draft also raises the concern, on page 4, that "no legal team is going to agree to allow any specific use based on metadata unless they have agreed that the license allows." The evidence for this assertion is unclear. However, even if true, it can easily be addressed by using the criteria above to design standard rights metadata profiles for each license to complement the metadata attributes associated with individual document. A legal team can then vet the profiles associated with a license, could certify a registry responsible for maintaining such profiles, or could agree to accept profiles available from the licenses authors.


[6] Alternatively, if the definition of metadata elements is simply beyond the capacity of this project, `free_to_use` should simply be replaced by a `license_ref` instance that includes a URI to a well known license that established `free_to_read`. The latter could even be designed for this purpose. This would at least remove the ambiguity of the `free_to_read` condition, and further simplify the schema.