Article Sharing Framework - Publisher & Platform Implementation Guide

STM STEC Working Group, Updated 3-May-2021

Dan O'Brien (ACS), Esther Heuver (Elsevier)

Table of Contents

Introduction

This document describes the process for implementing support of STM's Article Sharing Framework to enable seamless article sharing in line with publisher policies, covering both publisher and platform implementations. The overall process is best expressed in two parts:

1. establishing the article's identity, consisting of the article's unique identifier (the Digital Object Identifier, "DOI") and its version (NISO's Journal Article Version, "JAV");
2. sharing policy assertions made using Crossref's metadata services.

The article and version identifying information placed in the article PDF by publishers, combined with sharing policy identifiers placed in Crossref by publishers, can be used by platforms to determine applicable sharing rights for a given version of an article.

Note: while the term "article" is used throughout, it is not intended to specifically limit this to only journal articles; this may be applied to other published works that are distributed in PDF form including books/chapters, preprints, reference works, and etc.

Refer to resources at https://stm-assoc.org/asf for more information.

Part 1: Article identity (DOI) and version (JAV) embedded within the PDF

Publishers

During or immediately following the article generation process, publishers participating in the Article Sharing Framework encode the article identity (DOI) and version (JAV) into the PDF file. This metadata enables platforms to easily identify the article and version when they later receive an article PDF.

Two methods are defined for encoding the DOI and JAV. It is required that publishers implement at least one of these encoding methods, and highly recommended to implement both. Each encoding method has different advantages.

> **Method 1**: Encode DOI and JAV within the PDF's XMP metadata block
> **Method 2**: Encode JAV as a URI parameter within a DOI hyperlink

While Method 1 may represent a straightforward and preferred choice to implement this data encoding, publishers have found that XMP metadata may often be stripped or removed from the PDF by the time the article PDF is loaded to a repository. In addition, some older versions of PDF file formats may not support XMP. The hyperlink-based encoding approach within Method 2 has been found to be more likely to persist and remain intact during PDF file conversions or manipulations that may occur after initial distribution from the publisher, and is supported by just about all version of the PDF format.

Platforms

Conversely, upon receiving an article PDF with an intent to share, platforms can automatically scan the PDF file and look for either of the two article identify encoding methods.

> **Method 1**: Decode DOI and JAV from the PDF's XMP metadata block
> **Method 2**: Decode JAV as a URI parameter within a DOI hyperlink

As a publisher may encode the article identity using either method, the platform should attempt to locate and decode the article identity using either method, and if unsuccessful in determining the article identity, attempt the other method.

The American Chemical Society (ACS) provided a functional proof-of-concept application [link TBA] that implements both of these decoding methods, and has made the source code freely available. This proof of concept is provided to illustrate the concepts and feasibility of these decoding methods, and is not intended to be used as-is. However, platforms may find the proof of concept useful as an example as they adjust their internal technology to implement support for these decoding methods within their own technology stacks.

Method #1: Encoding/Decoding Article identity (DOI and JAV) within the PDF's XMP metadata block

Platforms will use PDF content extraction software to identify the DOI and JAV metadata fields embedded in the PDF's XMP. XMP refers to "Extensible Metadata Platform" as defined by ISO 16684-1:2012.

Open source and zero cost Java libraries such as Apache's **pdfbox** and **jempbox** are available for constructing an automated process for examining PDFs and extracting XMP data from them. Commercial tools for managing PDF content also often include these features.

The metadata elements that publishers will encode into the PDF, and platforms will identify and extract, include the following:

Metadata Element	Value to Add	Example
prism:doi	DOI of the article	`<prism:doi>10.1021/acsomega.0c03079</prism:doi>`
jav:journal_article_version	"VoR", "AM", or "AO"	`<jav:journal_article_version>VoR</jav:journal_article_version>`

See Figure 1. Article Identification and Version fields within PDF XMP metadata.

See NISO recommended practice RP-8-2008 for all allowable JAV values.

If JAV and DOI are not available in machine-readable form in XMP, platforms should attempt to identify retrieve metadata in the PDF body using Method #2.

Method #2: Encoding/Decoding JAV as a URI parameter within a DOI hyperlink

Platforms can use PDF content extraction software to look for hyperlinks within the PDF, and within those hyperlinks, which one represents a hyperlink for the article's DOI.   When this hyperlink is identified, the article's DOI and JAV values can be extracted from the URL.

The structure of the DOI hyperlink with encoded JAV is as follows:

Base DOI URL	`https://doi.org/`
Base DOI URI RegEx	`https?:\/\/.*doi\.org\/.*`
DOI path	`10.5555/12345678`
JAV URL parameter	`jav=VoR` or `jav=AM`
Target URL parameter	`rel=cite-as`

Note: the URL parameter "`rel=cite-as`", also added to the hyperlink by the publisher, allows the platform to distinguish which DOI hyperlink within the PDF refers to the <u>current</u> article. In many cases, there could be multiple DOI hyperlinks within an article PDF, for example when reference lists contain DOI hyperlinks to their sources.

Examples
Displayed: **https://doi.org/10.5555/12345678** Actual link URI: `https://doi.org/10.5555/12345678?rel=cite-as&jav=VoR`
Displayed: **https://doi.org/10.5555/12345678** Actual link URI: `https://doi.org/10.5555/12345678?jav=VoR&rel=cite-as`
Displayed: **https://doi.org/10.5555/12345678** Actual link URI: `https://doi.org/10.5555/12345678?ref=PDF&jav=VoR&rel=cite-as`

The URL parameters are intended to serve as targets for automated machine-based processing. However, a user may view this info by hovering their mouse pointer over the "displayed" link to see the "actual" link.

See Figure 2 as an illustration. In this example from ACS, a hyperlink exists on the DOI URL displayed in the lower right corner of the first page of the article PDF. The DOI URL that is displayed on the page does not need to match the actual hyperlink URI applied.

Case insensitivity: The JAV encoding implementation by publishers is case insensitive.   Thus, platforms should also search using case-insensitive JAV values.   For example, VOR, vor, and VoR should all be treated as a match for "version of record".

Not found: If JAV and DOI are not found or available in machine-readable form within the PDF, platforms should attempt to identify the article DOI and version through manual inspection and verification.

Part 2: Article sharing rights assertions within Crossref

To keep matters simple, publishers agree to a defined set of common sharing policies. Each sharing policy has been assigned with a unique and persistent identifier expressed in the form of a URI. These identifiers point to a human-readable description of the policy on the STM's website.

A standard method is defined on how to express these identifiers in Crossref's XML deposit scheme for a given article, and how to retrieve them for a given article.

Sharing Contexts

The Article Sharing Framework has defined four (4) types of sharing contexts, and within those, eleven specific sharing contexts.   Each sharing policy represents a combination of one sharing context of each of the four types. Altogether, this equates to up to 48 different permutations of sharing contexts, and thus 48 sharing policies have been defined, each identified with a persistent URI.   A publisher would likely assert only a small subset of these policies for a given article.

Context Type	Context code	Context Name	Article Sharing Policy text
Platform Type	pns	Platform not signed	Any platform, regardless whether it has signed complies with the STM Voluntary Principles for Article Sharing, can: …
	ps	Platform signed	A platform that has signed and is compliant with the STM Voluntary Principles for Article Sharing can: …
Displayable Elements	ft	Full text	… allow the sharing of the Full-Text, including Abstract, References, and Citation Metadata …
	ab	Abstract	… allow the sharing of the Abstract and Citation Metadata …
	ref	References	… allow the sharing of the References and Citation Metadata …
	cm	Citation Metadata	… allow only the sharing of the Citation Metadata …
JAV Article Version	vor	Version of Record	… of the Version of Record …
	am	Accepted Manuscript	… of the Accepted Manuscript …
	ao	Author Original	… of the Author Original …
Audience Scope	ga	General Access (public)	… for General Access, including any Research Collaboration Groups.
	rcg	Research Collaboration Group	… in Research Collaboration Groups.

A list of the full set of Article Sharing Framework policies and their assigned URI indicators are listed in Table 1 below.   Refer to resources at https://stm-assoc.org/asf for more information.

Sharing Policy Indicators in Crossref

Each sharing policy is identified by a unique, persistent URI. In fact, it is expressed as a DOI – not a DOI referring to the article intended to be shared or its publisher – but rather referring to the sharing policy document and STM as the publisher/steward of that policy.

For example, a publisher may register three sharing policies within Crossref for a given article:

Platform Type		JAV Article Version			Audience Scope		Displayable Elements				Policy URI
pns	ps	vor	am	ao	ga	rcg	ft	ab	ref	cm	
x		x			x			x			https://doi.org/10.15223/policy-002
	x	x				x	x				https://doi.org/10.15223/policy-029
	x		x		x		x				https://doi.org/10.15223/policy-033

Publishers:

During the article publication process, publishers participating in the Article Sharing Framework add sharing policy identifiers to the (existing) metadata deposits that they – or their delivery platforms – make to Crossref. These sharing policy identifiers, expressed as URIs, enable platforms to easily identify specific article sharing rights.   Multiple sharing policy identifiers may be expressed per article, covering differing possible sharing contexts and platforms.

Using NISO ALI within the Crossref metadata scheme, the policies indicated by context permutations above would be provided by the publisher to Crossref as the following:

```
<license_ref
            start_date="..."
            applies-to="stm-asf"
        >https://doi.org/10.15223/policy-002</license_ref>

<license_ref
            start_date="..."
            applies-to="stm-asf"
        >https://doi.org/10.15223/policy-029</license_ref>
<license_ref
            start_date="..."
            applies-to="stm-asf"
        >https://doi.org/10.15223/policy-033</license_ref>
```

The new `applies-to="stm-asf"` tag indicates that this license tag is a part of the STM article sharing framework.

Platforms:

After a platform receives an article PDF file intended for sharing, and has determined the article's identity by decoding the article DOI and JAV from the PDF (see Part 1 above), the platform then uses this article identify information to query the article's Crossref metadata record for any sharing policy indicators that were asserted by the publisher for the article in question.   If a sharing policy indicator is present within Crossref's record for the article, and sharing policy indicator matches the

context of the platform's type, audience, and sharing intent, then the article may be shared within the indicated context.

To aid an automated implementation, platform may find it useful to predefine an "accept list" of compatible sharing policy indicators that align with the platform's sharing context. A match within Crossref for any sharing policy indictor from the platform's accept list for a given article would indicate that the article may be shared.

Crossref metadata access

Crossref offers several methods and APIs for accessing the metadata record for a given article. These services are listed at https://www.crossref.org/services/metadata-retrieval/#00356, and many of them are free to use.

For example, the Crossref REST API allows retrieval of an article metadata record using this URI syntax:

```
https://api.crossref.org/works/{doi}
```

where **{doi}** is the DOI value that was decoded from the article PDF in Part 1 above. The results from this API call, returned in JSON format, may contain one or more "license" sections like this:

```
{
        "status": "ok",
        "message-type": "work",
        "message-version":"1.0.0", "message": {
                …
                "license": [{"URL": "https://doi.org/10.15223/policy-002" }],
                "license": [{"URL": "https://doi.org/10.15223/policy-029" }],
                "license": [{"URL": "https://doi.org/10.15223/policy-033" }],
                …
        }
}
```

In this case, the URL of **https://doi.org/10.15223/policy-029** within the license element represents an STM Article Sharing Framework sharing policy that conveys

Example of successful policy match

To continue the prior example from the Publisher section above, a platform that may want to share an article in the following context:

- Platform has signed and is compliant with the STM Voluntary Principles for Article Sharing
- The Full-Text of the article, including Abstract, References, and Citation Metadata
- The Version of Record of the article
- Share within a Research Collaboration Group.

Platform Type		JAV Article Version			Audience Scope		Displayable Elements				Policy URI
pns	ps	vor	am	ao	ga	rcg	ft	ab	ref	cm	
	x	x				x	x				https://doi.org/10.15223/policy-029

If the sharing policy indicator of https://doi.org/10.15223/policy-029 is present in the CrossRef record for the given article, then the platform may conclude that article may be shared in this context.

The text of policy #29 states:

> *"A platform that has signed and is compliant with the STM Voluntary Principles for Article Sharing can: allow the sharing of the Full-Text, including Abstract, References, and Citation Metadata of the Version of Record in Research Collaboration Groups."*

Example of unsuccessful policy match

To again continue the prior example from the Publisher section above, a platform may instead want to share an article in this context:

- Platform has not signed or is not compliant with the STM Voluntary Principles for Article Sharing
- The Full-Text of the article, including Abstract, References, and Citation Metadata
- The Version of Record of the article
- Share with a Global Audience.

Platform Type		JAV Article Version			Audience Scope		Displayable Elements				Policy URI
pns	ps	vor	am	ao	ga	rcg	ft	ab	ref	cm	
x		x			x		x				https://doi.org/10.15223/policy-001

In this example from the Publisher section above, the sharing policy indicator of https://doi.org/10.15223/policy-001 is not present in the CrossRef record for the given article, thus the platform must conclude that article may be not shared in this context.

The text of policy #1 states:

"Any platform, regardless whether it has signed and complies with the STM Voluntary Principles for Article Sharing, can: allow the sharing of the Full-Text Version of Record, including Abstract, References, and Citation Metadata for General Access, including any Research Collaboration Groups."

Figures

Figure 1. Article Identification and Version fields within PDF XMP metadata
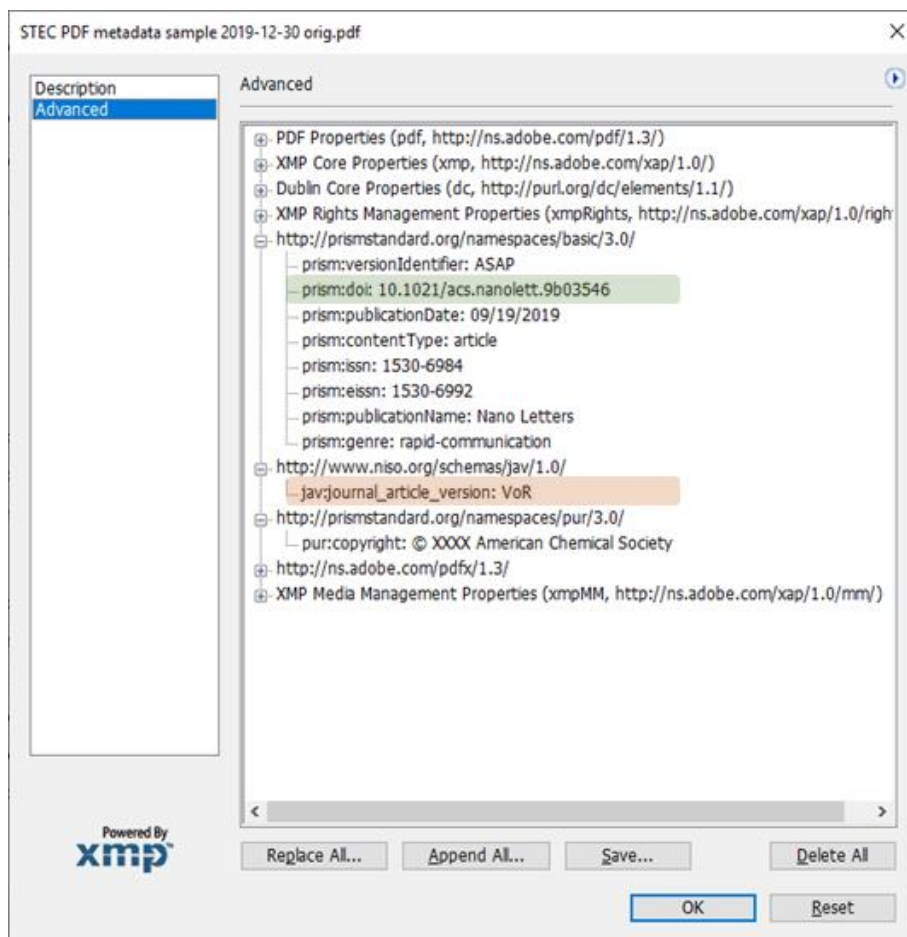
Figure 2. DOI hyperlink contains JAV and REL parameters

1000 mg/L and even higher, and the microbe community

https://doi.org/10.1021/acsomega.0c03079

A   https://doi.org/10.1021/acsomega.0c03079?ref=pdf&rel=cite-as&jav=VoR

Text displayed on the PDF page	https://doi.org/10.1021/acsomega.0c03079
Actual URI of the hyperlink	https://doi.org/10.1021/acsomega.0c03079?ref=PDF&jav=VoR&rel=cite-as

Table 1. List of Article Sharing Framework Policies

Key:
  x - permission is granted for this context.
  * - permission is inferred and also granted for this context.

#	STM ASF Context Attributes											STM ASF Policy DOI
	Platform Type		JAV Article Version			Audience Scope		Displayable Elements				
	pns	ps	vor	am	ao	ga	rcg	ft	ab	ref	cm	
1	x	*	x			x	*	x	*	*	*	https://doi.org/10.15223/policy-001
2	x	*	x			x	*		x		*	https://doi.org/10.15223/policy-002
3	x	*	x			x	*			x	*	https://doi.org/10.15223/policy-003
4	x	*	x			x	*				x	https://doi.org/10.15223/policy-004
5	x	*	x				x	x	*	*	*	https://doi.org/10.15223/policy-005
6	x	*	x				x		x		*	https://doi.org/10.15223/policy-006
7	x	*	x				x			x	*	https://doi.org/10.15223/policy-007
8	x	*	x				x				x	https://doi.org/10.15223/policy-008
9	x	*		x		x	*	x	*	*	*	https://doi.org/10.15223/policy-009
10	x	*		x		x	*		x		*	https://doi.org/10.15223/policy-010
11	x	*		x		x	*			x	*	https://doi.org/10.15223/policy-011
12	x	*		x		x	*				x	https://doi.org/10.15223/policy-012
13	x	*		x			x	x	*	*	*	https://doi.org/10.15223/policy-013
14	x	*		x			x		x		*	https://doi.org/10.15223/policy-014
15	x	*		x			x			x	*	https://doi.org/10.15223/policy-015
16	x	*		x			x				x	https://doi.org/10.15223/policy-016
17	x	*			x	x	*	x	*	*	*	https://doi.org/10.15223/policy-017
18	x	*			x	x	*		x		*	https://doi.org/10.15223/policy-018
19	x	*			x	x	*			x	*	https://doi.org/10.15223/policy-019
20	x	*			x	x	*				x	https://doi.org/10.15223/policy-020
21	x	*			x		x	x	*	*	*	https://doi.org/10.15223/policy-021
22	x	*			x		x		x		*	https://doi.org/10.15223/policy-022
23	x	*			x		x			x	*	https://doi.org/10.15223/policy-023
24	x	*			x		x				x	https://doi.org/10.15223/policy-024
25		x	x			x	*	x	*	*	*	https://doi.org/10.15223/policy-025
26		x	x			x	*		x		*	https://doi.org/10.15223/policy-026
27		x	x			x	*			x	*	https://doi.org/10.15223/policy-027
28		x	x			x	*				x	https://doi.org/10.15223/policy-028
29		x	x				x	x	*	*	*	https://doi.org/10.15223/policy-029
30		x	x				x		x		*	https://doi.org/10.15223/policy-030
31		x	x				x			x	*	https://doi.org/10.15223/policy-031
32		x	x				x				x	https://doi.org/10.15223/policy-032
33		x		x		x	*	x	*	*	*	https://doi.org/10.15223/policy-033
34		x		x		x	*		x		*	https://doi.org/10.15223/policy-034
35		x		x		x	*			x	*	https://doi.org/10.15223/policy-035
36		x		x		x	*				x	https://doi.org/10.15223/policy-036
37		x		x			x	x	*	*	*	https://doi.org/10.15223/policy-037
38		x		x			x		x		*	https://doi.org/10.15223/policy-038
39		x		x			x			x	*	https://doi.org/10.15223/policy-039
40		x		x			x				x	https://doi.org/10.15223/policy-040

41		x			x	x	*	x	*	*	*	https://doi.org/10.15223/policy-041
42		x			x	x	*		x		*	https://doi.org/10.15223/policy-042
43		x			x	x	*			x	*	https://doi.org/10.15223/policy-043
44		x			x	x	*				x	https://doi.org/10.15223/policy-044
45		x			x		x	x	*	*	*	https://doi.org/10.15223/policy-045
46		x			x		x		x		*	https://doi.org/10.15223/policy-046
47		x			x		x			x	*	https://doi.org/10.15223/policy-047
48		x			x		x				x	https://doi.org/10.15223/policy-048

For more information, refer to resources available at https://stm-assoc.org/asf