Case Study

Manuscript Exchange Common Approach and Journal Article Tag Suite (JATS) compatibility: a case study utilizing the JATS Compatibility Meta Model

Laura Randall[1*], Sally Ubnoske[2*]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD;
[2]Aries Systems Corporation, North Andover, MA, USA

Abstract

The National Information Standards Organization (NISO) Manuscript Exchange Common Approach (MECA) project is a cross-organization industry initiative to develop a common approach to manuscript transfer that can be adopted across the scholarly publishing industry. MECA establishes a vocabulary set that includes transfer, review, and manifest models. These models are designed to work with different article XML schemas, including the latest NISO Journal Article Tag Suite (JATS) standard (v1.2). In order to avoid conflicts between these project vocabularies and the JATS, we reviewed the MECA vocabularies against the NISO JATS Compatibility Meta Model (v0.7). This paper describes the review and analysis of the MECA schemas against the JATS Meta Model, how we documented the analysis, and the recommendations we made to resolve issues revealed by the analysis. It includes the documentation we produced to communicate the results of the analysis and what actions we took to move forward with the project, including both changes to the schemas and requests for changes in the JATS. We hope sharing our experiences with this process will help others who are trying to do the same.

Keywords
Documentation; Industry; Scholarly communication; Vocabulary

*Both authors contributed equally to this article.

Introduction

Background/rationale: Academic career progression often depends on a researcher's number and quality of publications and the prestige of the journal in which the research is published [1]. It can be difficult for authors to find an initial match between the paper's content and the scope of the journal. Approximately 85% of researchers resubmit rejected manuscripts to an-

other journal, which can be a tedious and error-prone process [2]. Reviewers can become frustrated by the time wasted repeating reviews for papers that are submitted to a different publication after rejection. One study found that 20% of biomedical researchers performed between 69% and 94% of reviews [3]. In peer review, papers are often rejected but reviews are not typically shared between publications. Researchers spend 15 million person-hours a year reviewing unpublished submissions to scientific journals [4]. If 85% of researchers resubmit rejected papers to a different journal, and 20% of biomedical reviewers perform 80% of reviews, the odds are high that the same researchers will be invited to review a paper that they have previously reviewed.

The publishing ecosystem is evolving. New publication workflows require submission transfers to and from preprint servers, other submission systems, and vendors. At the same time, interest in online and collaborative authoring tools is growing. Each system must implement a separate exchange mechanism with each other system.

Objectives: It aims to describe the analysis of the Manuscript Exchange Common Approach (MECA) schemas against the Journal Article Tag Suite (JATS) Compatibility Meta Model. Specifically, the follows were included: original MECA Working Group's work, JATS Compatibility Meta Model (JCMM) analysis, and recommendation to resolves issues after analysis.

Original MECA Group and Their Work

Early in 2017, a team of producers and users of manuscript systems identified a strong need to be able to seamlessly transfer manuscripts and reviews between and among manuscript systems, such as those in use at publishers and preprint servers. An industry-wide initiative was formed to create a common mechanism for transferring submissions, based on industry standards and best practices. The original MECA group submitted a proposal to National Information Standards Organization (NISO) to form a Working Group to develop a NISO Recommended Practice. This proposal was approved by the NISO Information Creation & Curation Topic Committee and NISO voting members. The Working Group was announced in May 2018 and the group met regularly to review the use cases, transmission method, packaging, and document type definitions (DTDs). The draft MECA Recommended Practice was made available for public comment in January 2020.

The following principles guide the MECA project: (1) journals and authors set the rules on what metadata and files are transferred to another system. (2) The goal is to facilitate transfer between publications and platforms. The package may not represent a complete submission, but it should con-

tain enough data and files to create a record in the receiving system. MECA defines the minimal set of data required to initiate a submission. (3) Various systems can independently define a Minimal Viable Product to get started with submission transfers. (4) The MECA practice uses current common technology and industry standards. (5) MECA presents a set of recommendations but does not prescribe a specific code to be written. There is no hub, software, or service. Unlike Crossref or ORCID, there is no attempt to trace the path of a manuscript as it traverses the publishing ecosystem.

The original MECA group consisted of Aries Systems, Clarivate, eJournalPress, HighWire Press, and PLOS. The expanded working group includes the original MECA members, plus representatives of the American Chemical Society, the American Physical Society, Cold Spring Harbor, eLife, IEEE, Green Fifteen, Jisc, the Journal of Clinical Investigation, the US National Library of Medicine, Springer Nature, and Taylor and Francis.

JATS Compatibility Meta Model Analysis

For a project like MECA—whose primary goals include the interchange of JATS documents—it is very important that the methods and models respect the existing systems that are intended to utilize the MECA data. The JCMM was created to ensure that data creators can "extend the reach of the JATS vocabularies without conflicting with current JATS vocabularies" [5]. In order to ensure that the MECA models did not conflict with the current JATS vocabularies, we evaluated those models against v0.7 of the JCMM [5].

Compatibility Analysis of MECA to JATS

Performing the compatibility analysis began by listing the structures in the MECA schemas that shared names with structures in JATS. The shared structures needed to match JATS on all the points of compatibility: semantic match; element or attribute; the structure is a section-like model; structure contains alternatives; the ID/IDREF definition in attributes; and whitespace handling.

To check for a semantic match, we utilized the non-normative JATS Tag Libraries for version 1.2 [6] and compared the definitions of the JATS elements against the intended use of the MECA structures. For example, an early draft of the MECA manifest DTD included a *name* element in a structure to capture metadata.

```
<!ELEMENT  metadata (name, value)>
<!ELEMENT  name (#PCDATA)>
<!ELEMENT  value (#PCDATA)>
```

In this structure, the semantic definition of the name can be

stated as the name component of a name-value pair describing a piece of metadata not otherwise enumerated in the model.

JATS also includes a *name* element, which is defined as, "Container element for the component elements of personal names, such as a *<surname>*" [7].

The use in MECA, to define metadata, is not a semantic match to the element in JATS which is used to capture the name of a person. This semantic mismatch was recorded as a point of incompatibility between MECA and JATS in the analysis.

In order to make the MECA models JATS-compliant, they need to be compliant on each of the aspects defined in the JCMM. So even once we identified one point of incompatibility, we continued to evaluate the models. All points of incompatibility need to be resolved for MECA to be compatible with the JATS. Continuing with the analysis of the name element, evaluating the remaining points of compatibility involved comparing definitions in the different DTDs. The element model in NISO JATS v1.2 is an element-only model:

<!ELEMENT name (((surname, given-names?) | given-names), prefix?, suffix?)>

Both the MECA manifest and JATS models define the name as an element, so on the element vs. attribute compatibility requirement, MECA is compatible. Neither the MECA manifest DTD nor the JATS definitions of the name have a section-like model or contain alternatives, so on those two points of compatibility, MECA is compliant. Since the name is an element in both models, the ID/IDREF definition isn't a consideration.

The last point of compatibility is how an element handles whitespace—whether it handles it as data (significant whitespace) or element-only (insignificant whitespace). For the *#PCDATA* model of the name in MECA's manifest DTD, the whitespace is significant. For the element-only name defined in JATS, the whitespace is insignificant. On this compatibility requirement, the name defined in the early MECA manifest DTD is incompatible with JATS. Once the analysis of all the shared structures was complete, the results of the analysis need to be recorded in a way that could be communicated with the entire MECA Working Group.

Communication of the Compatibility Requirements

In addition to the description of the compatibility requirements, the JCMM includes a Suppl. 1 that communicates the compatibility requirements in a table format (Table 1). Table 1 provides a clear and concise look at the compatibility requirements, so we utilized this format to communicate the results of the compatibility analysis, expanding it to include columns for recording semantic match and which model's properties were being recorded (Table 2).

In addition to the table recording compatibility model properties, the analysis document contained details of all the recorded conflicts and suggested solutions to those conflicts. An example of conflicts noted in the summary table is as follows:

<aff>
JATS name: Affiliation
JATS definition: Name of an institution or organization (for

Table 1. JATS compatibility properties catalog

JATS structure name	Element or Attribute property	Alternatives property[a]	Section-like property[a]	Whitespace handling property[b]	Attribute ID or IDREF property
abbr	Attribute				
abbrev	Element			D	

JATS, Journal Article Tag Suite.
[a]In these columns, "X" means "yes", and no value means "no"; [b]In this column, "E" means "element-like whitespace", "D" means "data-like whitespace", and "P" means "preserve whitespace." Table 1 do not contain X, E, and P since this is a snippet copied from the supplementary data rather than a complete table.

Table 2. Summary table of JATS compatibility

Structure name	Semantic match	Model	Element or attribute	Alternative	Section-like	Whitespace handling (D or E)[a]	ID/IDRE
addr-line	Yes	JATS	Element			D	
		MECA	Element			D	
aff[b]	Yes	JATS	Element			D	
		MECA	Element			E	

JATS, Journal Article Tag Suite; MECA, Manuscript Exchange Common Approach.
[a]In this column, "E" means "element-like whitespace", "D" means "data-like whitespace"; [b]JATS-compatibility problem.

example, university, corporation) with which a contributor is affiliated.

MECA usage: Element-only content

Conflict: JATS defines <aff> a mixed-content model, MECA defines it as element only. This affects whitespace handling.

Possible solution: Allow mixed-content

After detailing the analysis of the specific element and attribute model compatibilities, conflicts, and possible solutions, the analysis document also made note of several items defined in the MECA schemas that, while not direct conflicts with the JATS models, are close enough to existing JATS structures that they were worth review. This category included, for example, the use of an *href* attribute on external links in the MECA models that is identical to the *xlink:href* attribute usage in JATS.

Methods to Resolve Compatibility Conflicts

The Working Group used three different methods to resolve compatibility conflicts: (1) changing the MECA model to match the model in JATS. This was used to resolve incompatibilities in whitespace handling or to bring the models into alignment. (2) Changing the name of the element or attribute. This implementation typically resulted in element and attribute name that were more specific. (3) Submit a change request to the NISO JATS Standing Committee. The Working Group submitted two requests to the JATS Standing Committee requesting that definitions be revised.

In the previously discussed example regarding *name*, the Working Group opted to restructure the way the metadata was captured, using a single element (*metadata*) and putting the name of the metadata into an attribute:

<!ELEMENT metadata (#PCDATA)>

<!ATTLIST metadata metadata-name CDATA #REQUIRED>

By creating an attribute named *metadata-name* instead of *name*, this updated model avoids compatibility conflicts with the JATS.

In two cases of the semantic mismatch, *date-type* and *version* attributes, the recommended action for resolving the conflicts was to suggest to the NISO JATS Standing Committee that they redefine the JATS structure.

The JATS v1.2 definition of the *date-type* attribute includes the word "article"

"Event in the lifecycle of an article that this date is marking..." [8]

In performing the review for MECA, however, we noted that the JATS uses the attribute on the *date* element which is used to capture dates of non-article content. JATS includes the *date* element in the citation models (*element-citation* and *mixed-citation*) and on *related-object*, all of which describe both article and non-article content. Because JATS was already using the

attribute in ways beyond its own semantic definition, we requested that the semantic definition in JATS be broadened [9].

This requested was granted by the NISO JATS SC and was implemented in v1.3d1 of the Tag Suite (emphasis added): "Event in the lifecycle of an **object** that this date is marking..."

https://jats.nlm.nih.gov/archiving/tag-library/1.3d1/attribute/date-type.html

The second request made to the NISO JATS Standing Committee was to revise the definition of the version *attribute*. The NISO JATS currently has a *version* attribute used exclusively on the tex-math element. Its semantic definition limits the use to just that element:

"Version of TeX or LaTeX used to produce the mathematics."[10]

Early drafts of MECA used a more generic *version* attribute on several different elements to capture the appropriate version information of the element, without creating more specific attributes for those elements. Because of the wide-ranging applicability of an attribute to capture an element's version information, we requested the NISO JATS SC consider broadening the definition of *version* so it could be applied more widely [11]. This request acknowledged that this is a non-trivial change, but one that would offer a broad solution to capture version information. Unlike the request for *date-type*, however, this request was denied by the JATS Standing Committee.

Since this request was denied, the MECA use of the *version* attribute was still incompatible with the JATS per the JCMM guidelines. To resolve this, the MECA Working Group chose to change the attributes in the models to make them more specific. Ultimately the Working Group implemented several element specific **-version* attributes (*item-version*, *manifest-version*, *review-version*, *transfer-version*) to be used on the respective elements. The NISO MECA Working Group made revisions to the models to make them compliant with the JATS as described by the JCMM.

Conclusion

It can be challenging to work in a group comprised of competitive systems. The original MECA group was formed with a common goal of defining a method of manuscript exchange that could be adopted industry-wide. This meant that each organization had to compromise on some aspects of the solution. Having a common goal and an outlook of compromise in favor of the common good enabled the group to work collaboratively to define the MECA practice. Members of the original MECA group developed reference implementations based on the initial MECA definition. Each organization understood that the initial method would be modified by the NISO Working Group and committed to updating its implementation to conform with the NISO Recommended Prac-

tice. The NISO MECA Working Group benefited from having the initial MECA recommendation as a starting point. Yet it still took about a year to complete the project. Commitment to the project from each member of the group was essential to developing the NISO Recommended Practice.

Supplementary Material

Supplementary file is available from https://doi.org/10.6087/kcse.235

Suppl. 1. MECA model review for JATS compatibility

References

1. Hoffman AJ. In praise of 'B' journals [Internet]. Washington, DC: Inside Higher Ed; 2017 [cited 2020 Feb 14]. Available from: https://www.insidehighered.com/views/2017/03/28/academics-shouldnt-focus-only-prestigious-journals-essay

2. Belcher WL. When a journal says no [Internet]. Washington, DC: Inside Higher Ed; 2009 [cited 2020 Feb 14]. Available from: https://www.insidehighered.com/advice/2009/04/27/when-journal-says-no

3. Kovanis M, Porcher R, Ravaud P, Trinquart L. The global burden of journal peer review in the biomedical literature: strong imbalance in the collective enterprise. PLoS One 2016; 11:e0166387. https:/doi.org/10.1371/journal.pone.0166387

4. The AJE Team. Peer review: how we found 15 million hours of lost time [Internet]. Durham, NC: AJE Scholar [cited 2019 Aug 6]. Available from: https://www.aje.com/arc/peer-review-process-15-million-hours-lost-time/

5. Usdin BT, Lapeyre DA, Randall L, Beck J. JATS compatibility meta-model description draft 0.7 [Internet]. Baltimore, MD: National Information Standards Organization; 2016 [cited 2020 Feb 14]. Available from: https://groups.niso.org/apps/group_public/document.php?document_id=16764&wg_abbrev=jats-sc

6. Mulberry Technologies. Journal Archiving and Interchange Tag Library NISO JATS version 1.2 (ANSI/NISO Z39.96-2019) [Internet]. Bethesda, MD: National Center for Biotechnology Information; 2019 [cited 2020 Feb 14]. Available from: https://jats.nlm.nih.gov/archiving/tag-library/1.2/

7. Mulberry Technologies. Element: Name of Person [Internet]. Bethesda, MD: National Center for Biotechnology Information; 2019 [cited 2020 Feb 14]. Available from: https://jats.nlm.nih.gov/archiving/tag-library/1.2/element/name.html

8. Mulberry Technologies. Attribute: Type of Date [Internet]. Bethesda, MD: National Center for Biotechnology Information; 2019 [cited 2020 Feb 14]. Available from: https://jats.nlm.nih.gov/archiving/tag-library/1.2/attribute/date-type.html

9. Randall L. Comment #00788-Broaden definition of @date-type-z39_96-2015.pdf (revision #1) [Internet]. Baltimore, MD: National Information Standards Organization; 2019 [cited 2020 Feb 14]. Available from: https://groups.niso.org/apps/group_public/view_comment.php?comment_id=788

10. Mulberry Technologies. Attribute: Version of TeX or LaTeX [Internet]. Bethesda, MD: National Center for Biotechnology Information; 2019 [cited 2020 Feb 14]. Available from: https://jats.nlm.nih.gov/archiving/tag-library/1.2/attribute/version.html

11. Randall L. Comment #00789-Rename and broaden definition of version attribute-z39_96-2015.pdf (revision #1) [Internet]. Baltimore, MD: National Information Standards Organization; 2019 [cited 2020 Feb 14]. Available from: https://groups.niso.org/apps/group_public/view_comment.php?comment_id=789