

Publishing Notebooks Using JATS and MECA

JJ Allaire Charles Teague

12/9/22

Scope

JATS

- Referencing notebooks used as source for entire article
- Referencing notebooks used to produce figures, tables, etc.

MECA

- Bundling notebooks into MECA archives
- Referencing notebook source code in MECA

Notebooks in JATS XML

Notebooks in Article Metadata

Include notebooks associated with the article as supplementary material in `<article-metadata>`

```
1 <article-meta>
2   <article-id pub-id-type="pmid">10092260</article-id>
3   ...
4   <supplementary-material id="nb-sup-1"
5     xlink:href="notebooks/source.ipynb"
6     specific-use="document"
7     mimetype="application/x-ipynb+json"/>
8 </article-meta>
```

- Use `<supplementary-material>` to reference
- Include all notebooks associated with article (the document source notebook as well as source notebooks)
- Use `specific-use` attribute to differentiate document, supporting, or output notebooks (`document`, `source`, `supporting`)

<alternatives/>

Use the `<alternatives>` element to provide notebook references for figures, tables, etc.

```
1 <figure>
2   ...
3   <alternatives></alternatives>
4 </figure>
5 <table-wrap>
6   ..
7   <alternatives></alternatives>
8 </table>
```

- Contains processing alternatives for a given element
- Referenceable via `<xref>`
- Include notebook as `<media>` or `<supplementary-material>`

<media>

Approach 1: Use `media` to provide notebooks as `alternatives` for `body`, `fig`, and `table-wrap` elements.

```
1 <fig id="f3" position="float">
2   <alternatives>
3     <graphic ...>...</graphic>
4     <media id="nb-cell-id"
5       mimetype="application/x-ipynb+json"
6       specific-use="source-notebook"
7       xlink:href="notebooks/source.ipynb#cell-id"/>
8   </alternatives>
9 </fig>
10 <table-wrap>
11   <alternatives>
12     <table>...</table>
13     <media id="nb-cell-id-tbl"
14       mimetype="application/x-ipynb+json"
15       specific-use="source-notebook"
16       xlink:href="notebooks/source.ipynb#cell-id-tb
17   </alternatives>
18 </table-wrap>
```

- Pointer to non-XML object
- Object is considered integral to article
- `href` includes cell identifier

<supplementary-material>

Approach 2: Use `supplementary-material` to provide notebooks as alternatives for `body`, `fig`, and `table-wrap` elements.

```
1 <fig id="f3" position="float">
2   <alternatives>
3     <graphic ...>...</graphic>
4     <supplementary-material id="nb-sup-1"
5       xlink:href="notebooks/source.ipynb#cell-id"
6       mimetype="application/x-ipynb+json">
7       <caption>
8         <p>source notebook for this figure</p>
9       </caption>
10    </supplementary-material>
11  </alternatives>
12 </fig>
13 <table-wrap>
14   <alternatives>
15     <table>...</table>
16     <supplementary-material id="nb-sup-2"
17       xlink:href="notebooks/source.ipynb#cell-id2"
18       mimetype="application/x-ipynb+json">
19       <caption>
20         <p>source notebook for this table</p>
21       </caption>
22     </supplementary-material>
23   </alternatives>
24 </table-wrap>
```

- External resource that supports the article but which is not integral
- Used to add detail, background, or context to an article
- Can appear in `<article-meta>` or in the flow of content

Notebook href

When referencing a notebook, hrefs may optionally include a fragment which addresses a specific cell within a notebook

For example:

```
1 <supplementary-material  
2   ...  
3   xlink:href="notebooks/plot-src.ipynb#cell-id-1"  
4 />
```

Fragments are resolved against notebooks using the following, in order:

1. **Cell Id**: the cell having the identifier
2. **Cell Tags**: the first cell tagged with the identifier

MECA and Notebooks

V1: Basic Notebook Support

- Include notebooks within MECA bundles
- Notebooks properly referenced from `manifest.xml`
- Notebooks referenced from within `article.xml` file via JATS XML
- By convention, place referenced notebooks within a `notebooks` subfolder

V2: Notebook Execution Support

- Add a `notebooks.xml` (notebook manifest) file to MECA bundle
- Optionally contains url to live interactive version
- Optionally includes a `<source/>` which provides ability to recreate execution environment

```
1 <notebook url='notebooks/document.ipynb'  
2         deploy='(uri-to-live-interactive-version) '>  
3   <source path='(local-REES-path)|(remote-repo-path)' type='REES' />  
4 </notebook>
```

- `source` points to a local or remote path which can be used to recreate the execution environment for the notebook
 - **REES Compliant** local or remote path
 - Path to local container or remote container registry
- By convention, place local sources in a `source` folder

Open Question: Review

- Reviewers likely need a Word or PDF copy of the article for review
- Bundling rendered copies within MECA is straightforward, however links to notebooks are an open question

Appendix - Complete Example

Sample JATS

Simple article created using a note and include a figure generated using a notebook.

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <!DOCTYPE article PUBLIC "-//NLM//DTD JATS (Z39.96) Journal Archiving and Interchange DTD
3 v1.2 20190208//EN" "https://jats.nlm.nih.gov/archiving/1.2/JATS-archivearticle1.dtd">
4 <article-meta>
5   <supplementary-material id="nb-sup-1"
6     xlink:href="notebooks/document.ipynb"
7     specific-use="document"
8     mimetype="application/x-ipynb+json"/>
9   <supplementary-material id="nb-sup-1"
10     xlink:href="notebooks/source.ipynb"
11     specific-use="source"
12     mimetype="application/x-ipynb+json"/>
13 </article-meta>
14 <article>
15   <body>
16     <fig id="f3" position="float">
17       <alternatives>
18         <graphic ...>...</graphic>
19         <supplementary-material id="nb-sup-1"
20           xlink:href="notebooks/source.ipynb#cell-id"
21           mimetype="application/x-ipynb+json">
22         </supplementary-material>
23       </alternatives>
24     </fig>
25   </body>
26 </article>
```

Sample MECA Zip Contents

Notebook and environment data included to allow execution of notebooks

File Name	File Type
123e4567-e89b-12d3-a456-426655440000-MECA.zip	Archive file
manifest.xml	Manifest metadata file
demoarticle.xml	Article metadata file
demoarticle_files/fig-jats/plot.jpg	File uploaded by author
demoarticle_files/fig-jats/plto2.jpg	File uploaded by author
demoarticle.pdf	System file
notebooks.xml	Notebook metadata file
notebooks/source.ipynb	Notebook file (source of plot)
notebooks/article.ipynb	Notebook file (document)
source/requirements.txt	REES compliant folder

Sample notebooks.xml

Notebooks manifest

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE manifest PUBLIC "-//MECA//DTD Manifest v2.0//en" "MECA_manifest.dtd">
3
4 <notebooks
5   xmlns="https://www.manuscriptexchange.org/schema/notebook"
6   notebook-version="1.0">
7
8   <notebook url='notebooks/document.ipynb'
9     deploy='https://colab.research.google.com/drive/doc'>
10     <source path='source/' type='REES'/>
11   </notebook>
12
13   <notebook url='notebooks/source.ipynb'
14     deploy='https://colab.research.google.com/drive/source'>
15     <source path='source/' type='REES'/>
16   </notebook>
17
18 </notebooks>
```