NISO RP-36-2020

JATS4R Data Citations Version 2.0

A Recommended Practice of the
National Information Standards Organization

About NISO Recommended Practices

A NISO Recommended Practice is a recommended "best practice" or "guideline" for methods, materials, or practices in order to give guidance to the user. Such documents usually represent a leading edge, exceptional model, or proven industry practice. All elements of Recommended Practices are discretionary and may be used as stated or modified by the user to meet specific needs.

This recommended practice may be revised or withdrawn at any time. For current information on the status of this publication contact the NISO office or visit the NISO website (www.niso.org).

Published by

National Information Standards Organization (NISO)

3600 Clipper Mill Road, Suite 302

Baltimore, MD 21211

www.niso.org

Copyright © 2020 by the National Information Standards Organization

This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nd/4.0/.

All inquiries should be addressed to: NISO, 3600 Clipper Mill Road, Suite 302, Baltimore, MD 21211 or via email at nisohq@niso.org

DOI: https://doi.org/10.3789/niso-rp-36-2020

ISBN (13): 978-1-950980-09-3

https://jats4r.org/data-citations

JATS4R Data Citations

Status: *NISO Approval*
Version: 2.0
Provenance: Version 1.0 reworked by Melissa Harrison based on feedback and alignment with Data Availability Statements recommendation.
History: Working: September – October 2019
JATS4R Committee Review: November 2019 – April 29, 2020
Public Review: April 29 – May 29, 2020
JATS4R Steering Committee Review: May 30 – June 26, 2020
NISO: June 26 – July 26, 2020
PUBLISHED: 21 September 2020

Change history

Additional content:

1. As per the JATS4R recommendation on Data availability statements, there are four options for capturing references to data. However, including the references within the main <ref-list> for the article is recommended by the Force11 Publishers Early Adopters Expert Group, and only those references are recognised by Google Scholar. See recommendation 2.
2. Recommendations 7 onwards have been updated or added.

Context

<ref>, <mixed-citation>, <element-citation>, <person-group>, <data-title>, <source>, <year>, <pub-id>, <ext-link>, <version>

@publication-type, @person-group-type, @pub-id-type, @assigning-authority, @designator, @xlink:href

Description

This recommendation contains best practices for tagging citations to datasets in a reference list.

As per the JATS4R recommendation on Data Availability Statements, there are four options for capturing references to data. However, including the references within the main <ref-list> for the article is recommended by the Force11 Publishers Early Adopters Expert Group, and only those references are recognised by Google Scholar.

1. These recommendations only apply to JATS 1.1 and forward, because the tags needed to make data citations machine readable are only available from 1.1 onwards.

2. The following recommendations are specifically about citations related to data and datasets. See JATS4R's recommendations for Citations (general).

Additional reading

- Adapting JATS to support data citation
- Force 11 Joint Declaration of Data Citation Principles
- A data citation roadmap for scientific publishers

Recommendations

1. **@publication-type="data"** on **<mixed-citation>** or **<element-citation>**. Use "data" as the value of @publication-type to indicate that the citation is to a dataset, even if that dataset is the entire data repository.

   [[Validator tool result: if @publication-type not "data" and <data-title> is present ERROR]]

2. **@specific-use** on **<element-citation>** or **<mixed-citation>**. For publishers who elect to collect such granularity in their workflow, see the table below for four @specific-use attributes recommended for JATS XML. For publishers who use the Relation Type method for Crossref deposits we've provided a mapping in the table.

Data type (@specific-use)	Description	Map to this Crossref relationship type
"supporting"	Data that supports the study's findings. Use this generic value if you do not wish to further distinguish whether the supporting data were generated or analyzed	"references"
"generated"	Supporting data that were generated for the study	"isSupplementedBy"
"analyzed"	Supporting data that were analyzed (but not generated) for the study	"references"
"non-analyzed"	Referenced data that were neither generated nor analyzed for the study	"references"
–	Data type is not indicated (no @specific-use value is supplied)	"references"

[[Validator tool result: if a “Clean Value Test” on these strings in @specific-use fails. Example, warning if the value is “nonanalyzed” or “non-analysed” WARNING]]

3. **@person-group-type** on **<person-group>**. As of version 1.1, the list of values for this attribute includes “curator”, specifically to support data citations. Use “curator” whenever appropriate.

4. **<data-title>** / **<source>**. At least one of <data-title> or <source> must be present. <data-title> should hold the title of the dataset. <source> should contain the name of the holding repository. Both should be present if applicable.

   [[Validator tool result: if @publication-type on (parent::mixed-citation or parent::element-citation) is “data” and one of <data-title> or <source> is not present ERROR]]

5. **<year>**. This should contain the 4-digit year of publication. If the element contains anything other than a single 4-digit year (such as, for example, “2012A”, “2005Q1”), then use the @iso-8601-date attribute to specify the 4-digit year.

   [[Validator tool result: if the content is not a 4-digit year and there is no @iso-8601-date with content ERROR]]

6. **<pub-id>**. This element should be used to hold the repository ID for the data. The repository ID should be a DOI or similar persistent identifier. The @pub-id-type attribute must be used — see the next recommendation for details.

7. **@pub-id-type** on **<pub-id>**.
   In contrast to what is stated in the Tag Library (“Type of publication identifier or the organization or system that defined the identifier”) this attribute should *only* be used to state the type of identifier, and not to specify the organisation or system that defined the identifier.

   [[Validator tool result: Defer result pending the discussion of the attribute value registry]]

8. **@assigning-authority** on **<pub-id>**. When the given type of identifier can be assigned by more than one organisation (e.g. DOIs minted by CrossRef or DataCite) and the organisation registering the identifier is known, include the @assigning-authority attribute on the <pub-id> element. For example, a DOI that is assigned by CrossRef should have “doi” as the @pub-id-type, and “crossref” as the @assigning-authority. For many types of identifiers, there is only one assigning authority. For example, PubMed IDs are always assigned by the National Library of Medicine. In these cases @assigning-authority is not necessary.

[[Validator tool result: Info for now; may change pending attribute value registry implementation]]

9. **@xlink:href** on **<pub-id>**. Optional. Including an @xlink:href with a fully resolved URI can improve the near term reusability of content. However, as identifying authorities can change their base URIs over time, omitting @xlink:href can improve the long term reusability of content. The decision to include @xlink:href should be evaluated by the interested parties on a case by case basis.

10. **<ext-link>**. Use <ext-link> to provide a link directly to the data citation. If there is no pub-id, ext-link should be included.

    [[Validator tool result: if no pub-id or ext-link ERROR]]

11. **<version>** Use this element to display the human-readable version number of the dataset.

12. **@designator** on **<version>** (1.1 and above). Use this attribute to contain the machine-readable version number of the dataset. The element contents can be a more human-readable note (see the example).

    [[Validator tool result: if <version> present and no @designator present ERROR]]

Examples

```
<ref id="d1">
   <element-citation publication-type="data" specific-use="analyzed">
      <person-group person-group-type="author">
         <collab>The Concerto Consortium</collab>
         <name>
            <surname>van Beethoven</surname>
            <given-names>Ludwig</given-names>
         </name>
         <name>
            <surname>Liszt</surname>
            <given-names>F</given-names>
       </name>
    </person-group>
    <person-group person-group-type="curator">
       <name>
          <surname>Bach</surname>
          <given-names>JS</given-names>
      </name>
```

```
   </person-group>
   <data-title>Title of dataset</data-title>
   <year iso-8601-date="2014">2014</year>
   <source>Repository Name</source>
   <pub-id pub-id-type="doi" assigning-authority="datacite"
>10.1234/1234321</pub-id>
   <version designator="16.2">16th version, second release</version>
   </element-citation>
</ref>
```

Here are some additional examples:

```
 <!--Data reference: Dryad dataset -->
            <ref id="bib8">
                <element-citation publication-type="data">
                    <person-group person-group-type="author">
                        <name>
                            <surname>Kok</surname>
                            <given-names>K</given-names>
                        </name>
                        <name>
                            <surname>Ay</surname>
                            <given-names>A</given-names>
                        </name>
                        <name>
                            <surname>Li</surname>
                            <given-names>L</given-names>
                        </name>
                        <name>
                            <surname>Arnosti</surname>
                            <given-names>DN</given-names>
                        </name>
                    </person-group>
                    <year iso-8601-date="2015">2015</year>
                    <data-title>Data from: Genome-wide errant targeting by
Hairy</data-title>
                    <source>Dryad Digital Repository</source>
                    <pub-id pub-id-type="doi"
>10.5061/dryad.cv323</pub-id>
                </element-citation>
            </ref>

            <!--Data reference: RCSB Protein Data Bank -->
            <ref id="bib9">
                <element-citation publication-type="data">
                    <person-group person-group-type="author">
                        <name>
```

```
                                <surname>Du</surname>
                                <given-names>J</given-names>
                            </name>
                            <name>
                                <surname>Johnson</surname>
                                <given-names>LM</given-names>
                            </name>
                            <name>
                                <surname>Groth</surname>
                                <given-names>M</given-names>
                            </name>
                            <name>
                                <surname>Feng</surname>
                                <given-names>S</given-names>
                            </name>
                            <name>
                                <surname>Hale</surname>
                                <given-names>CJ</given-names>
                            </name>
                            <name>
                                <surname>Li</surname>
                                <given-names>S</given-names>
                            </name>
                            <name>
                                <surname>Vashisht</surname>
                                <given-names>AA</given-names>
                            </name>
                            <name>
                                <surname>Gallego-Bartolome</surname>
                                <given-names>J</given-names>
                            </name>
                            <name>
                                <surname>Wohlschlegel</surname>
                                <given-names>JA</given-names>
                            </name>
                            <name>
                                <surname>Patel</surname>
                                <given-names>DJ</given-names>
                            </name>
                            <name>
                                <surname>Jacobsen</surname>
                                <given-names>SE</given-names>
                            </name>
                        </person-group>
                        <year iso-8601-date="2014">2014</year>
                        <data-title>Crystal structure of KRYPTONITE in complex with
mCHH DNA and
                            SAH</data-title>
```

```
                    <source>RCSB Protein Data Bank</source>
                    <pub-id pub-id-type="doi">10.2210/pdb4qen/pdb</pub-id>
                </element-citation>
            </ref>


            <!--Data reference: ArrayExpress: pub-id-type="accession"-->
            <ref id="bib11">
                <element-citation publication-type="data">
                    <person-group person-group-type="author">
                        <name>
                            <surname>Radoshevich</surname>
                            <given-names>L</given-names>
                        </name>
                        <name>
                            <surname>Impens</surname>
                            <given-names>F</given-names>
                        </name>
                        <name>
                            <surname>Ribet</surname>
                            <given-names>D</given-names>
                        </name>
                        <name>
                            <surname>Quereda</surname>
                            <given-names>JJ</given-names>
                        </name>
                        <name>
                            <surname>Nam Tham</surname>
                            <given-names>T</given-names>
                        </name>
                        <name>
                            <surname>Nahori</surname>
                            <given-names>MA</given-names>
                        </name>
                        <name>
                            <surname>Bierne</surname>
                            <given-names>H</given-names>
                        </name>
                        <name>
                            <surname>Dussurget</surname>
                            <given-names>O</given-names>
                        </name>
                        <name>
                            <surname>Pizarro-Cerdá</surname>
                            <given-names>J</given-names>
                        </name>
                        <name>
                            <surname>Knobeloch</surname>
```

```
                            <given-names>KP</given-names>
                        </name>
                        <name>
                            <surname>Cossart</surname>
                            <given-names>P</given-names>
                        </name>
                    </person-group>
                    <year iso-8601-date="2015">2015b</year>
                    <data-title>Transcription profiling by high throughput
sequencing of LoVo
                        cells infected with Listeria for 24 hr compared to
uninfected
                        cells</data-title>
                    <source>ArrayExpress</source>
                    <pub-id pub-id-type="accession"

xlink:href="https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3649/"
                        >E-MTAB-3649</pub-id>
                </element-citation>
            </ref>


<ref>
    <mixed-citation publication-type="data" id="msb188202-cit-0063">
        <person-group person-group-type="author">
            <string-name>
                <surname>Vaarala</surname>
                <given-names>MH</given-names>
            </string-name>, <string-name>
                <surname>Hirvikoski</surname>
                <given-names>P</given-names>
            </string-name>, <string-name>
                <surname>Kauppila</surname>
                <given-names>S</given-names>
            </string-name>, <string-name>
                <surname>Vuoristo</surname>
                <given-names>JT</given-names>
            </string-name>, <string-name>
                <surname>Paavonen</surname>
                <given-names>TK</given-names>
            </string-name>
        </person-group>
        (<year>2012</year>)
        <data-title>Androgen regulated gene expression in human
prostate</data-title>,
        <source>Gene Expression Omnibus</source>, <pub-id
pub-id-type="accession"
```

```
xlink:href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32982">GSE329
82</pub-id>.
    </mixed-citation>
</ref>


<ref id="ref1">
    <mixed-citation publication-type="data" id="data1" specific-use="analyzed">
        <person-group person-group-type="author">
            <string-name>
                <surname>Traka</surname>
                <given-names>M</given-names>
            </string-name>, <string-name>
                <surname>Gasper</surname>
                <given-names>AV</given-names>
            </string-name>, <string-name>
                <surname>Melchini</surname>
                <given-names>A</given-names>
            </string-name>, <string-name>
                <surname>Bacon</surname>
                <given-names>JR</given-names>
            </string-name>, <string-name>
                <surname>Needs</surname>
                <given-names>PW</given-names>
            </string-name>, <string-name>
                <surname>Frost</surname>
                <given-names>V</given-names>
            </string-name>, <string-name>
                <surname>Chantry</surname>
                <given-names>A</given-names>
            </string-name>, <string-name>
                <surname>Jones</surname>
                <given-names>AM</given-names>
            </string-name>, <string-name>
                <surname>Ortori</surname>
                <given-names>CA</given-names>
            </string-name>, <string-name>
                <surname>Barrett</surname>
                <given-names>DA</given-names>
            </string-name>, <string-name>
                <surname>Ball</surname>
                <given-names>RY</given-names>
            </string-name>, <string-name>
                <surname>Mills</surname>
                <given-names>RD</given-names>
            </string-name>, <string-name>
                <surname>Mithen</surname>
                <given-names>RF</given-names>
```

```
            </string-name>
        </person-group>
        (<year>2008</year>)
        <data-title>Transcription profiling by array of human prostate from
patients with a previous diagnosis of Prostatic Intraepithelial Neoplasia and
following consumption of high glucosinolate broccoli or peas to investigate
interactions with the GSTM1 genotype</data-title>.
        <source>ArrayExpress</source>.
        <pub-id pub-id-type="accession"
assigning-authority="EBI:arrayexpress">E-MEXP-1243</pub-id>
            (<ext-link extlink-type="uri"
xlink:href="https://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-1243">https:/
/www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-1243</ext-link>).
    </mixed-citation>
</ref>
```