

How <PubNote> supports life sciences researchers using XML for OLSPub

G. Ken Holman
XML Technology Lead
Réalta Online Publishing Solutions Ltd.
<https://RealtaOnline.com>



Brief history of PubMed by US NLM

- 1879-2004 - Index Medicus (in print form) of medical research information
- 1963 - computerization through MEDLARS (literature analysis and retrieval)
- 1968 - National Library of Medicine moved under National Institutes of Health
- 1971 - MEDLINE (MEDLARS online) metadata
- 1996 - experimental PubMed web-based interface to MEDLINE
- 1997 - official launch of free public access to all of MEDLINE via PubMed
 - functionally Anglo-centric (some articles are in other languages)
- 2002 - core system switch to direct processing of XML-based metadata:
 - PubMed In - submission XML DTD (database ingress)
 - PubMed Out - distribution XML DTD (database egress)
- 2025 - 39 million biomedical research citations (~2 million per year)
- 2026 - ???

OLSPub as an alternative to PubMed - Infrastructure made in Europe

May 2025

- ZB MED (German medical library) proposes a open-science infrastructure project aligned with European Open Science principles:
 - Title: "Open Life Science Publication database (OLSPub) - Strengthening the Biomedical Research Community by Building a Resilient and Sustainable Solution"
 - <https://zenodo.org/records/15533303>
 - <https://www.zbmed.de/en/research/current-projects/olspub>
- initialized with entire content of the PubMed database
- committed to multilingualism, openness, and transparency
- chartered for unfettered global access into the future

By end of 2025

- project expected to be approved and funded

ZB MED

Information Centre
for Life Sciences

réalta
online

20251102-1250Z

<PubNote>

Declarative Amsterdam 2025 - 2025-11-07

<PubNote>

in support of OLSPub and PubMed users

May/July 2025 - open-source GitHub project for exposition of the content of the shallow submission and distribution XML metadata instances

<https://bit.ly/ro-pubnote>

realtaonline/ PubNote

Making XML metadata approachable for OLSPub and PubMed users, <PubNote> works for you between the angle brackets.

2 Contributors 0 Issues 2 Discussions 0 Stars 0 Forks

GITHUB.COM



<PubNote>

Declarative Amsterdam 2025 - 2025-11-07

20251102-1250Z

Formats (produced using sustainable single-source publishing):



The image shows a PDF viewer interface. At the top, there is a toolbar with icons for navigation and editing. The main content area displays a PubMed article citation. The citation is formatted with bold text for the main sections and regular text for the details. The citation includes the PubMed ID, completion date, date revised, article title, journal name, ISSN, volume, issue, publication date, and title. The citation is as follows:

Medline Citation [Status="MEDLINE" in dexingMethod="Manual"]
PubMed ID: 4 [Version="1"]
Date Completed
 Year: 2000
 Month: 01
 Day: 01
Date Revised
 Year: 2000
 Month: 01
 Day: 01
Article [PubMed="Print"]
Journal
 ISSN: 0000-0000 [issn?type="Electronic"]
Journal Issue [CitedMedium="Internet"]
 Volume: 66
 Issue: 4
Publication Date
 Year: 2000
 Month: 01
 Day: 01
Title: A Study on Placeholder Data

PubMed Article

Medline Citation [Status="MEDLINE" IndexingMethod="Ma
PubMed ID 4 [Version="1"]
Date Completed
Year: 2000
Month: 01
Day: 01
Date Revised
Year: 2000
Month: 01
Day: 01
Article [PubMed="Print"]
Journal
ISSN: 0000-0000 [IssnType="Electronic"
Journal Issue [CitedMedium="Internet"]
Volume: 66
Issue: 4
Publication Date
Year: 2000
Month: 01
Day: 01
Title: A Study on Placeholder Da
ISO Abbreviation: JPlaceholderS
Article Title: AStudyonPlaceholderDa
Pagination
Medline Page Number: 100--1195
Author List [CompleteY/N="Y"]
Author [ValidY/N="Y"]
Last Name: Holmes

PubMed Article

Medline Citation [Status="MEDLINE" [AccessionMethod="Manual" Or]

PubMed ID: 4 [Version="1"]

Date Completed

Year: 2000

Month: 01

Day: 01

Date Revised

Year: 2000

Month: 01

Day: 01

Article [Abstract="First"]

Journal

ISSN: 0000-0000 [Language="Electronic"]

Journal Issue [AvailableSource="Internet"]

Volume: 66

Issue: 4

Publication Date

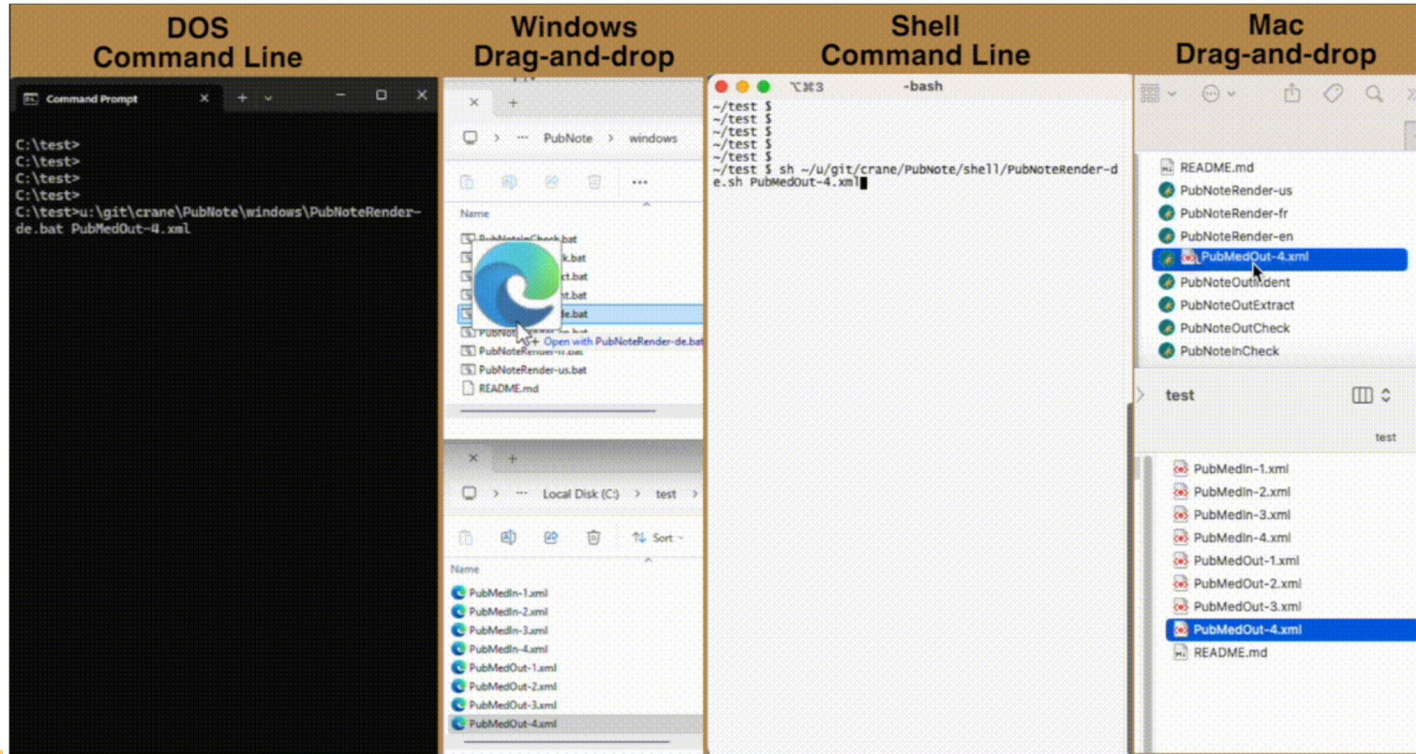
Year: 2000

Indented XML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE PubMedArticle
SYSTEM "http://dtd.nlm.nih.gov/ncbi/pubmed/out/pi
PubMedArticle">
  <MedlineCitation Status="MEDLINE" IndexingMethod=
    <PMID Version="1"4/>PMID>
    <DateCompleted>
      <Year>2000</Year>
      <Month>01</Month>
      <Day>01</Day>
    </DateCompleted>
    <DateRevised>
      <Year>2000</Year>
      <Month>01</Month>
      <Day>01</Day>
    </DateRevised>
    <Article PubModel="Print">
      <Journal>
        <ISSN IssnType="Electronic">0000-0000</
        <JournalIssue CiteMedium="Internet">
          <Volume>66</Volume>
          <Issue>4</Issue>
          <PubDate>
            <Year>2000</Year>
            <Month>01</Month>
            <Day>01</Day>
          </PubDate>
        </JournalIssue>
        <Title>StudyonPlaceholderData</Title>
```

<PubNote> (cont.)

Interfaces (intended for ease of use by the non-XML user):

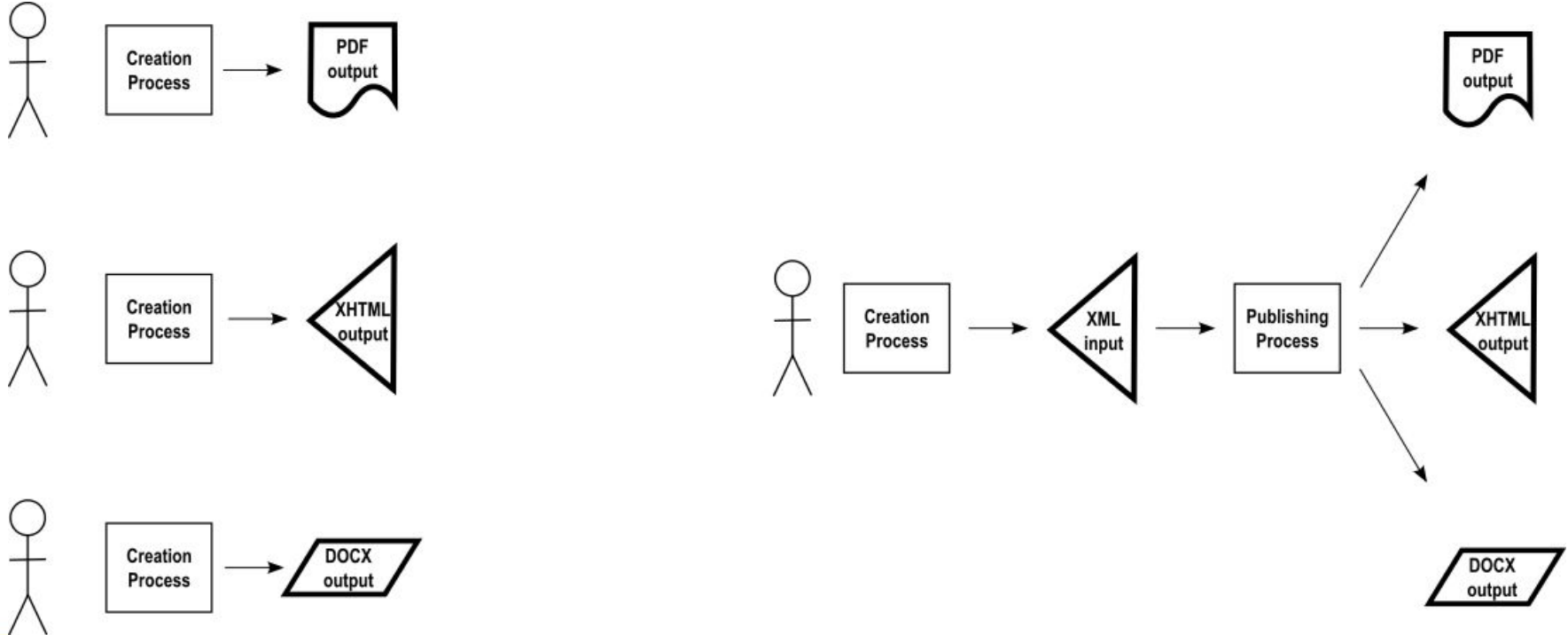


<PubNote>

Declarative Amsterdam 2025 - 2025-11-07

The theory of single-source publishing

Replace individual output creation with a publishing process from a single



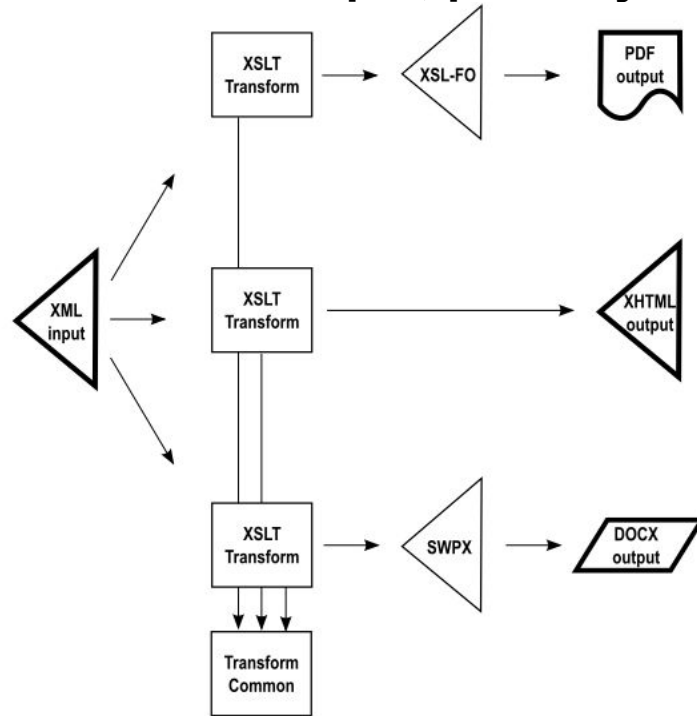
<PubNote>

Declarative Amsterdam 2025 - 2025-11-07

20251102-1250Z

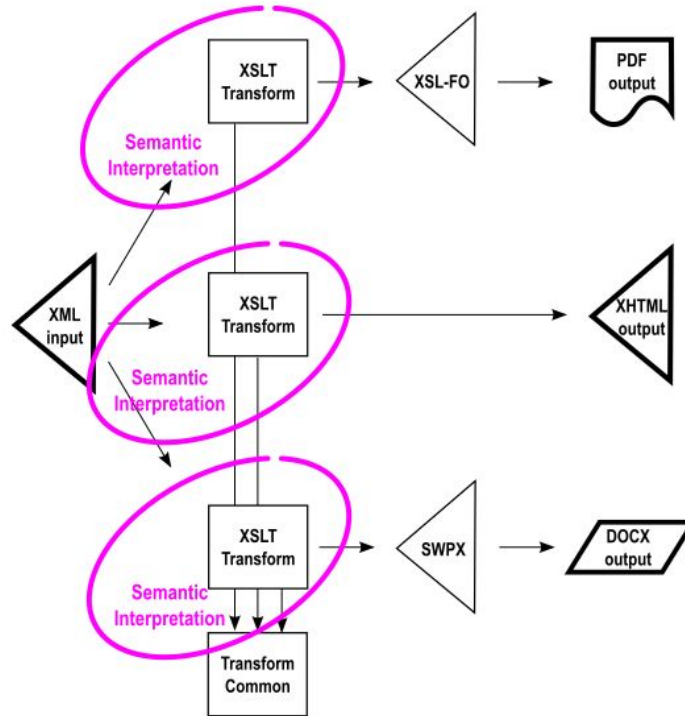
Unsustainable single-source publishing

Parallel stylesheets one for each output, possibly with shared common logic



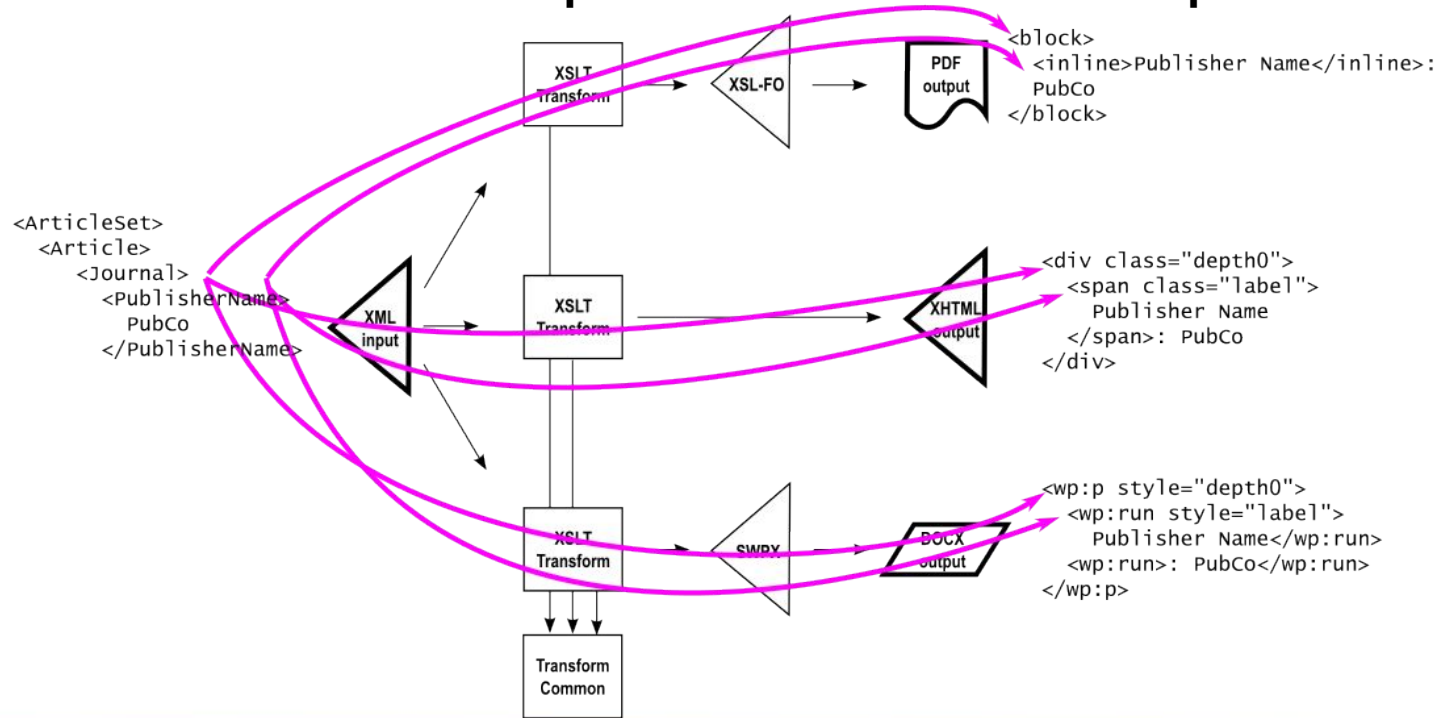
Unsustainable single-source publishing (cont.)

Parallel stylesheets each need separate maintenance for same interpretation



Unsustainable single-source publishing (cont.)

Each transformation has to interpret the semantics of the input vocabulary



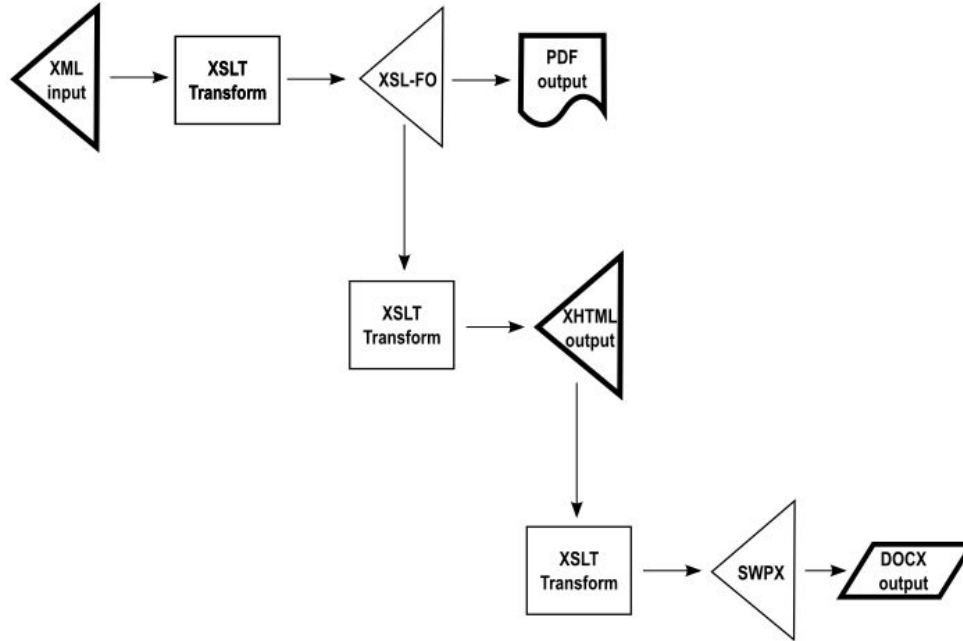
<PubNote>

Declarative Amsterdam 2025 - 2025-11-07

20251102-1250Z

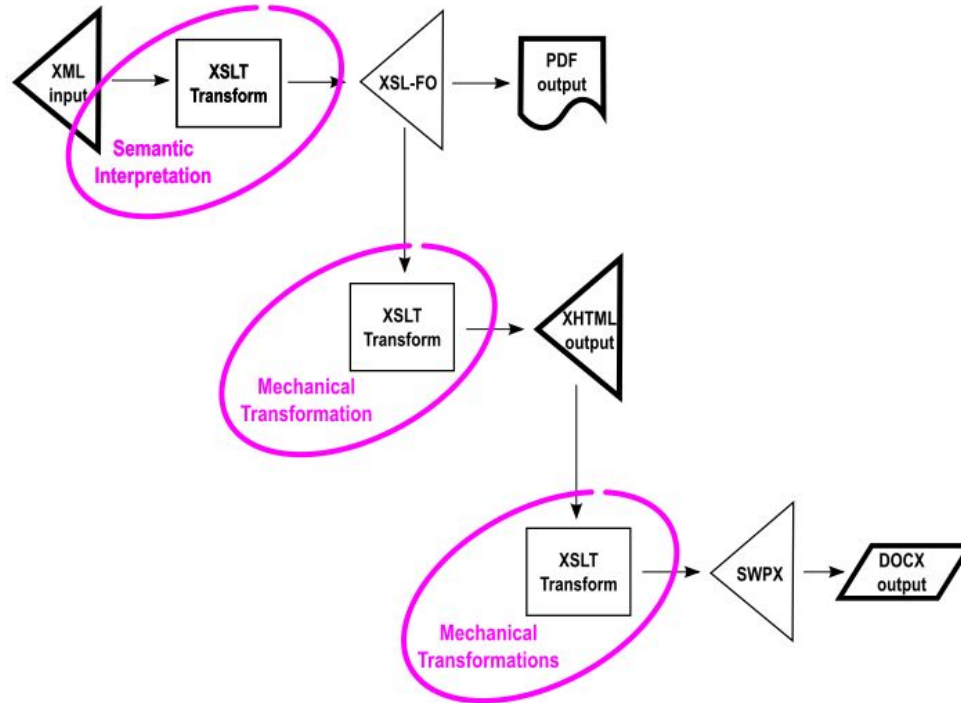
Sustainable single-source publishing

Focus downstream requirements in a single semantic interpretation



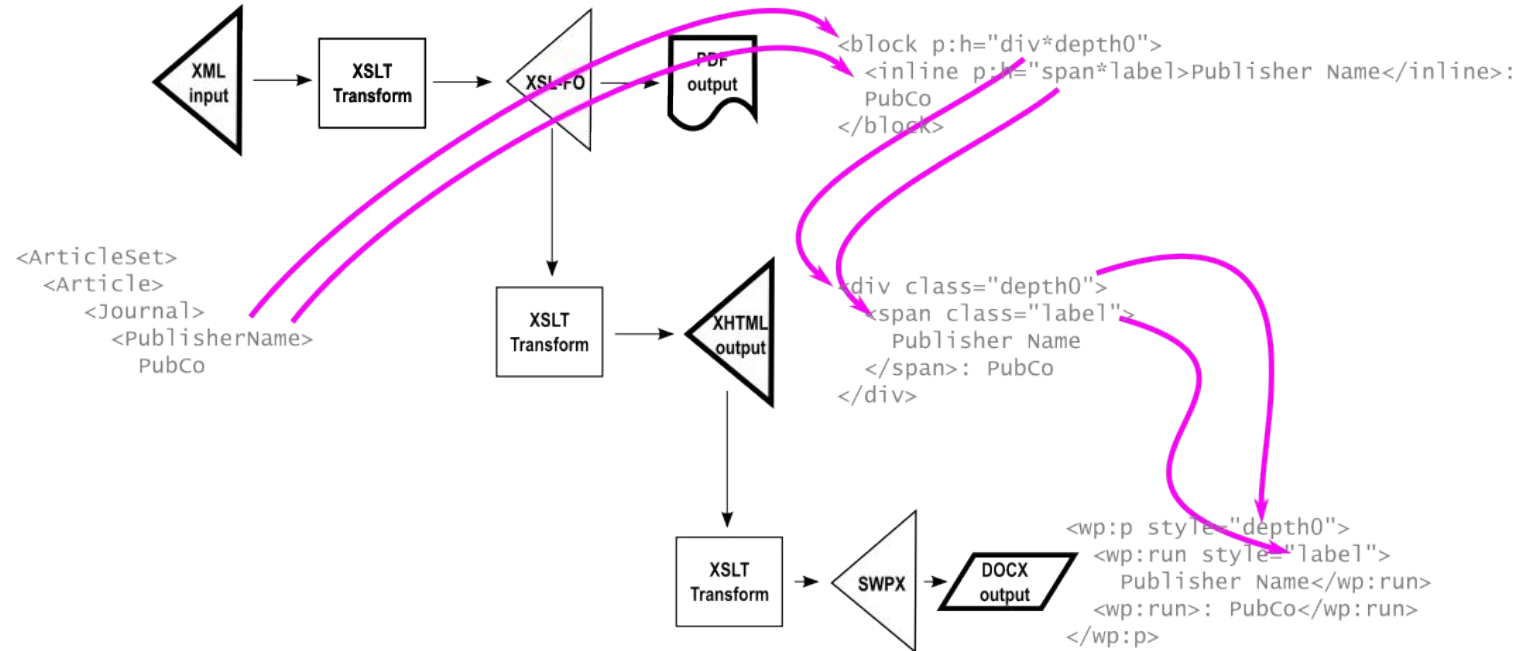
Sustainable single-source publishing (cont.)

Only one of the transforms focuses on the interpretation of the input semantics



Sustainable single-source publishing (cont.)

XSL-FO tolerates foreign namespaces containing downstream signals



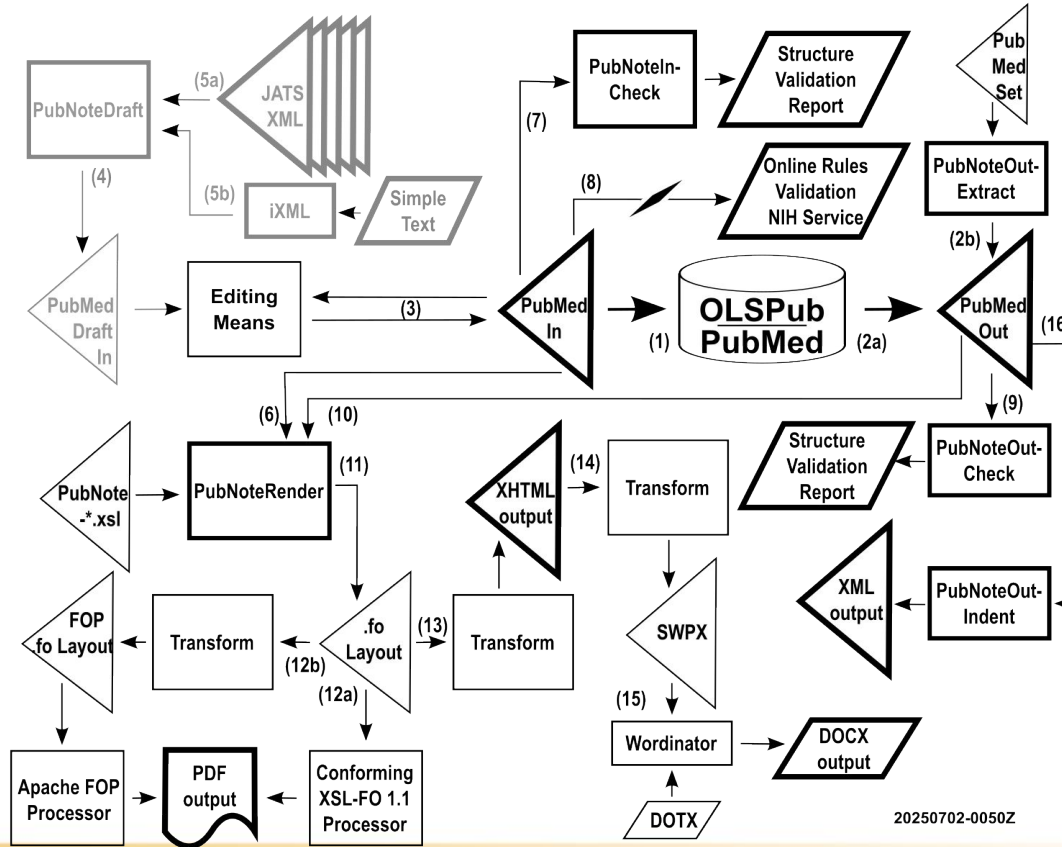
Sustainable single-source publishing (cont.)

Complete JATS-Con 2023 paper at:

<https://www.ncbi.nlm.nih.gov/books/NBK591971/>

Principles framed for <PubNote> in Declarative Amsterdam 2025 paper

<PubNote> data flows



20250702-0050Z

<PubNote> data flows (cont)

1. A valid **PubMedIn** XML submission document is forwarded for inclusion in the OLSPub and/or PubMed database
2. A valid **PubMedOut** XML distribution document is obtained for processing
 - 2a. exported directly from the OLSPub and/or PubMed database according to some criteria, or
 - 2b. extracted (with or without added indentation) using the **PubNoteOutExtract** process against one or more valid PubMedOut XML distribution documents, such as from an ~200MB XML single article set document of 30,000 article documents unpacked from a ~20MB ZIP file from the [PubMed FTP site](#)
3. Presumably, there is some editing process for the user's **PubMedIn** submission document
4. Optionally, the user uses **PubNoteDraft** to create a **PubMedDraftIn** document as a starting point (*Phase 2*)
5. Inputs used to create draft are expected to include (among others):
 - 5a. – a distillation of multiple JATS XML documents into a draft, or
 - 5b. – a simple text file with no angle brackets converted to XML using iXML (Invisible XML)
6. For human review, **PubNoteRender** formats the information found in a **PubMedIn** document
7. Optionally, **PubNoteInCheck** can validate a **PubMedIn** document against the [PubMed.dtd](#)
8. Optionally, a **PubMedIn** document can be checked using the NIH online business rules validator
9. Optionally, **PubNoteOutCheck** can validate a **PubMedOut** document against the [pubmed_250101.dtd](#)
10. For human review, **PubNoteRender** formats the information found in a **PubMedOut** document
11. XSL-FO is created with all of the content to be presented downstream in PDF, HTML, and DOCX
12. XSL-FO is transformed into **PDF** for print purposes
 - 12a. – a conforming XSL-FO 1.1 processor directly produces the output (*not part of this repository*)
 - 12b. – a non-conforming XSL-FO 1.1 process requires a transformation to make a suitable subset XSL-FO
13. XSL-FO is transformed into **XHTML** for browser display
14. XHTML is transformed into **SWPX** (Simple Word Processing XML) for word processing purposes
15. The **Wordinator** tool combines SWPX and a Microsoft **DOTX** template to produce a **DOCX** file
16. Optionally, **PubNoteOutIndent** processes a **PubMedOut** document into an indented XML output

Réalta Online Publishing Solutions Ltd.

<https://RealtaOnline.com>



Declarative Amsterdam 2025 - 2025-11-07

20251102-1250Z