Fidelity and flexibility for clients and their products leveraging XSLT and XSL-FO

G. Ken Holman
Réalta Online Publishing Solutions Limited
https://RealtaOnline.com

JATS-Con 2023 - June 13-14, 2023

The governing business context

Fit-for-purpose publishing focus for multiple clients based on open standards

- inputs: NISO-STS, ISO-STS, JATS, OASIS DocBook, and legacy PDF
- clients have much in common but lots of exceptions/individual requirements
- clients have limited resources to address changing needs (moving to XML)
- publishing process based on XSLT 2.0 and XSL-FO 1.1 (open and fast)
- technical documents (single column), complex tables, equations
- true landscape pages interleaved with portrait pages at section level
- automatic layout of multi-part sections, graphics, and tables
- quality results of 5 - 1000 pages (or more) produced in a timely process
- arrangements: full standard, preview standard, track-changes (red/green)
- outputs: PDF, line-numbered PDF, monolithic HTML, DOCX

The governing business context (cont.)

Standards Development Organizations publish content in PDF and/or XML

- National Bodies have an obligation to reproduce that content and can add national content to provide information regarding the national context of use
  - multiple layers of adoptions of an international standard: national, regional, international
  - single layers of body content of a national standard
- a publishing manager integrates automatic content feeds from ISO, IEC, CEN, and CENELEC helping national bodies meet timely obligations
- a fast SaaS publishing service can be leveraged as part of a publishing work chain or dynamically invoked by an author during the editing process because the publishing software is not running locally in the national body environment
- increasing use of requirements markup for downstream processing

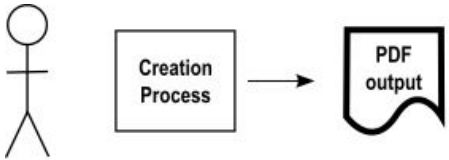
The technical context

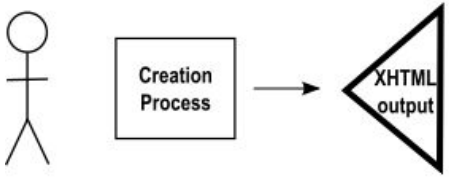
Single-source publishing to produce multiple results from input XML

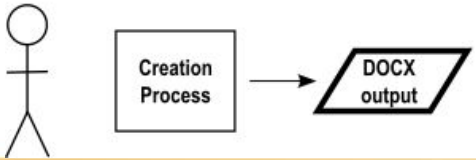
- XML also used to manage the merging and publishing of PDF content for adoptions provides a consistent approach to national bodies
- need to accommodate national content and appearance variances across national body clients
- need to accommodate variances between national body sub-organizations
- fidelity of the results is paramount across multiple output products
- semantic interpretation of input XML vocabulary governs arrangement and appearance of that content in the output
- national bodies, regional bodies, and international bodies all are using the same STS vocabulary but have nuances of appearance for a given NISO construct that need to be respected in a single adoption publication

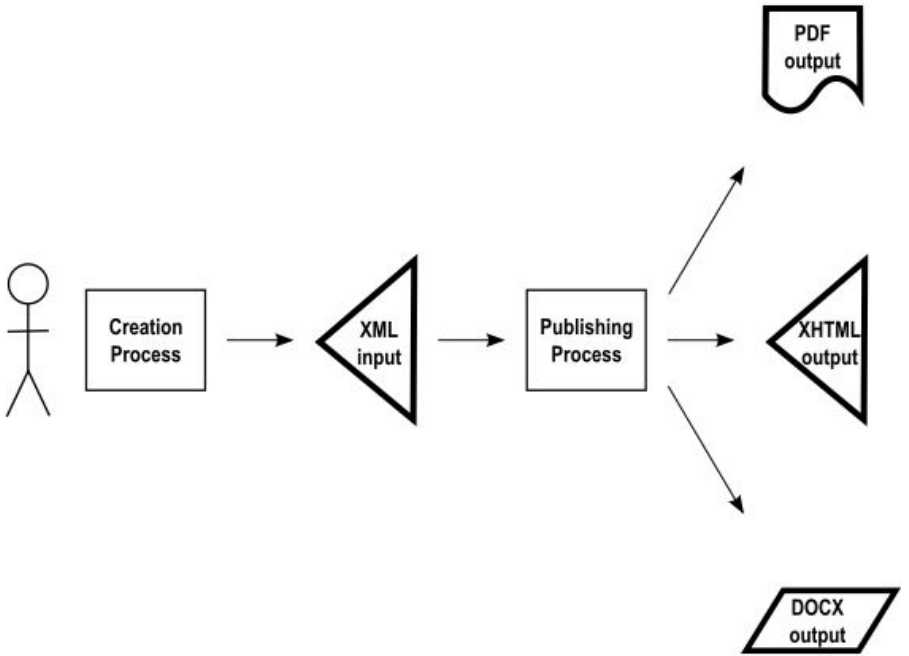
The theory of single-source publishing

Replace individual output creation with a publishing process from a single source



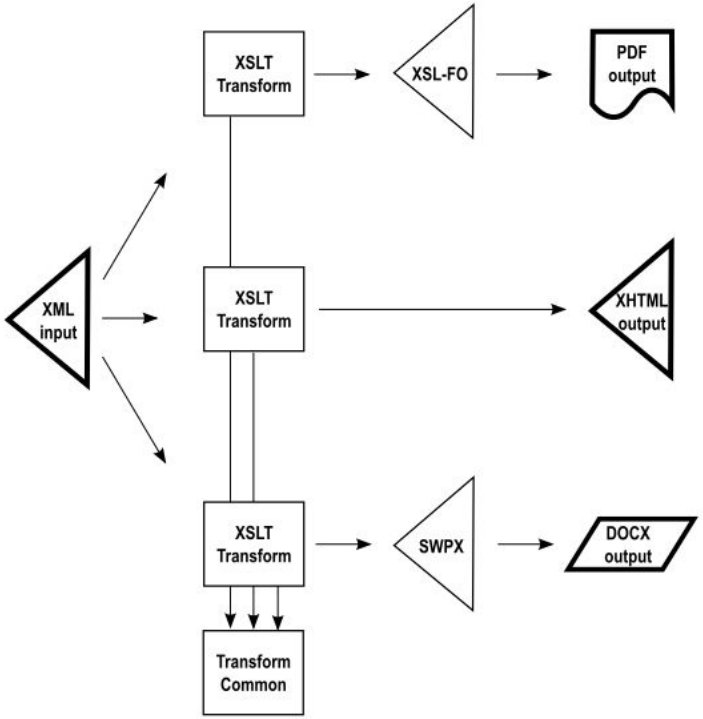






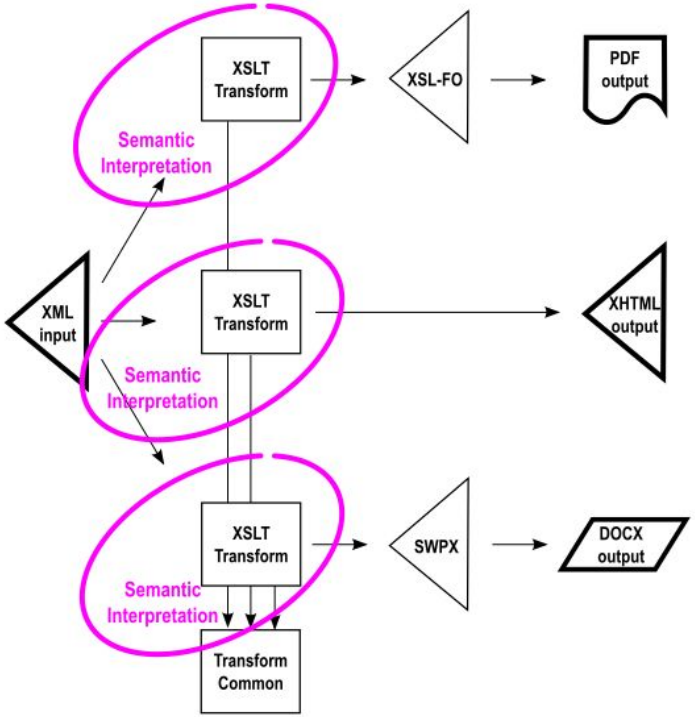
Unsustainable single-source publishing

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



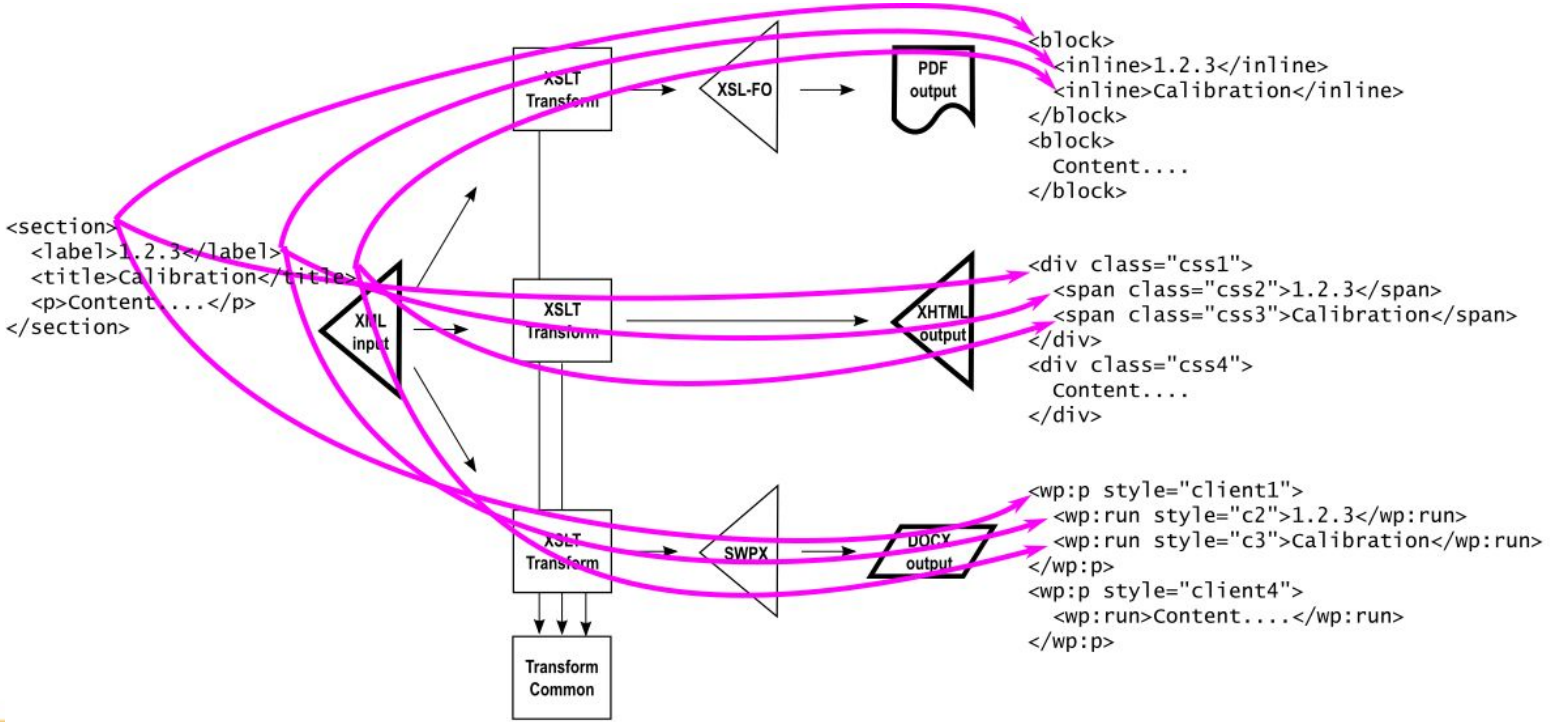
Unsustainable single-source publishing (cont.)

Parallel stylesheets each need separate maintenance for same interpretation



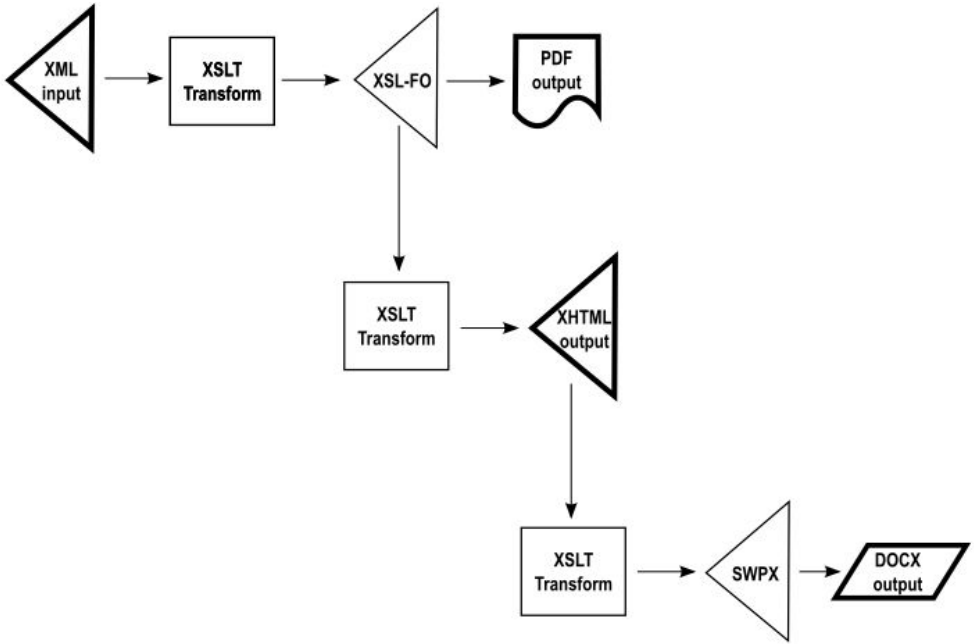
Unsustainable single-source publishing (cont.)

Each transformation has to interpret the semantics of the input vocabulary



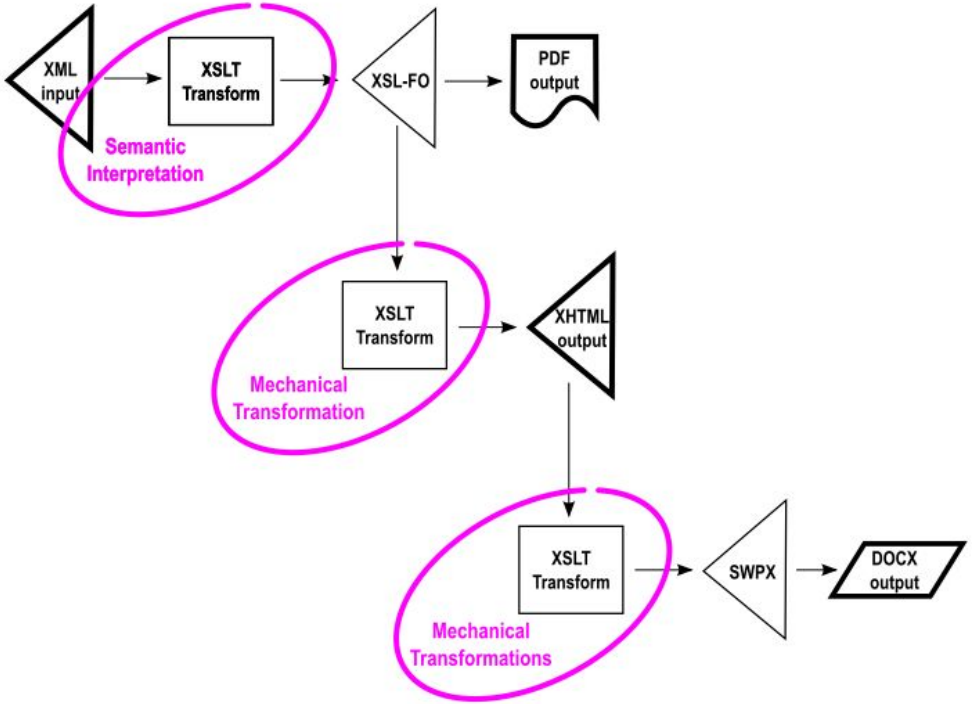
Sustainable single-source publishing

Focus downstream requirements in a single semantic interpretation



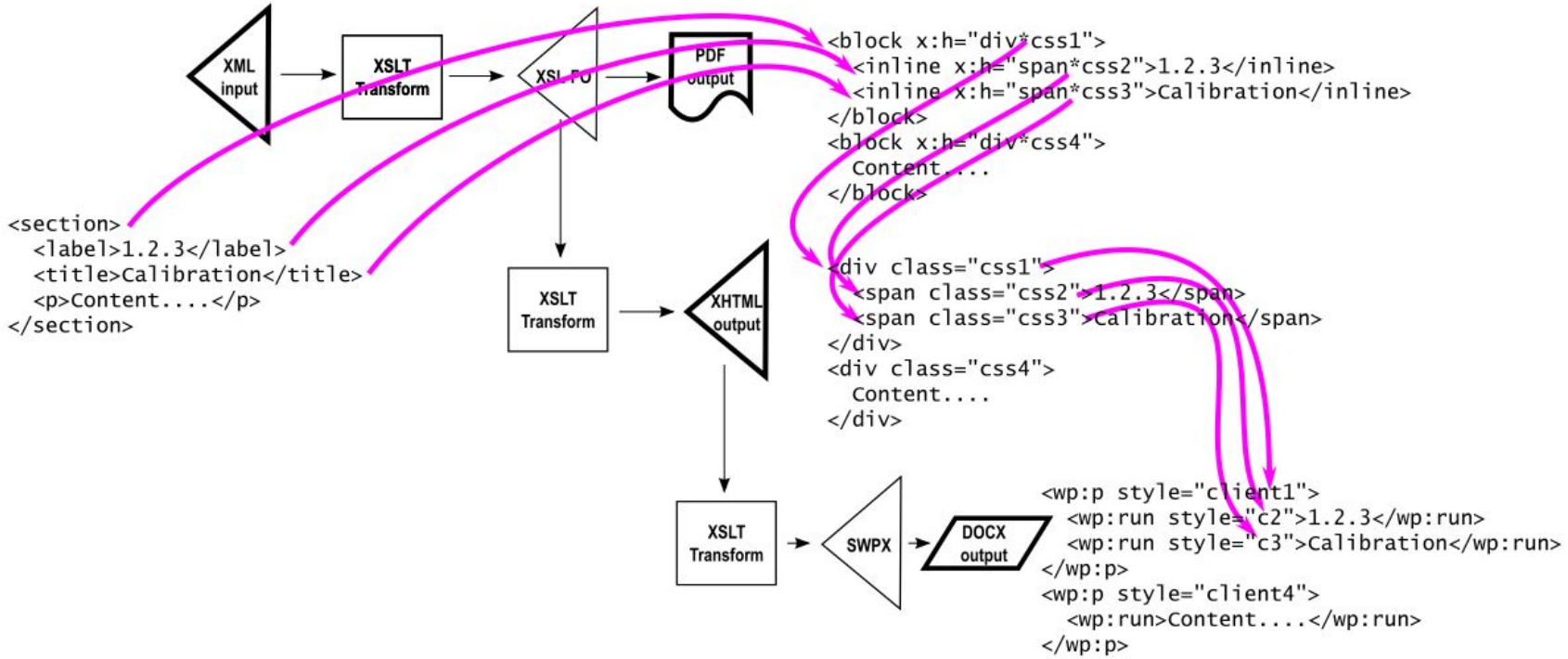
Sustainable single-source publishing (cont.)

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



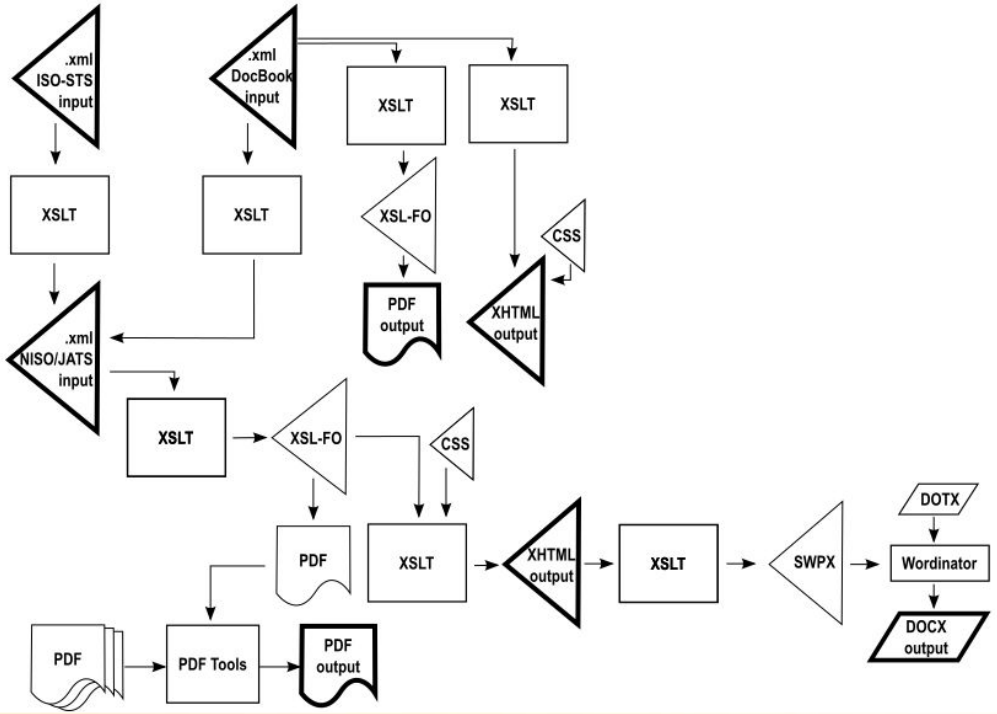
Sustainable single-source publishing (cont.)

XSL-FO tolerates foreign namespaces containing downstream signals



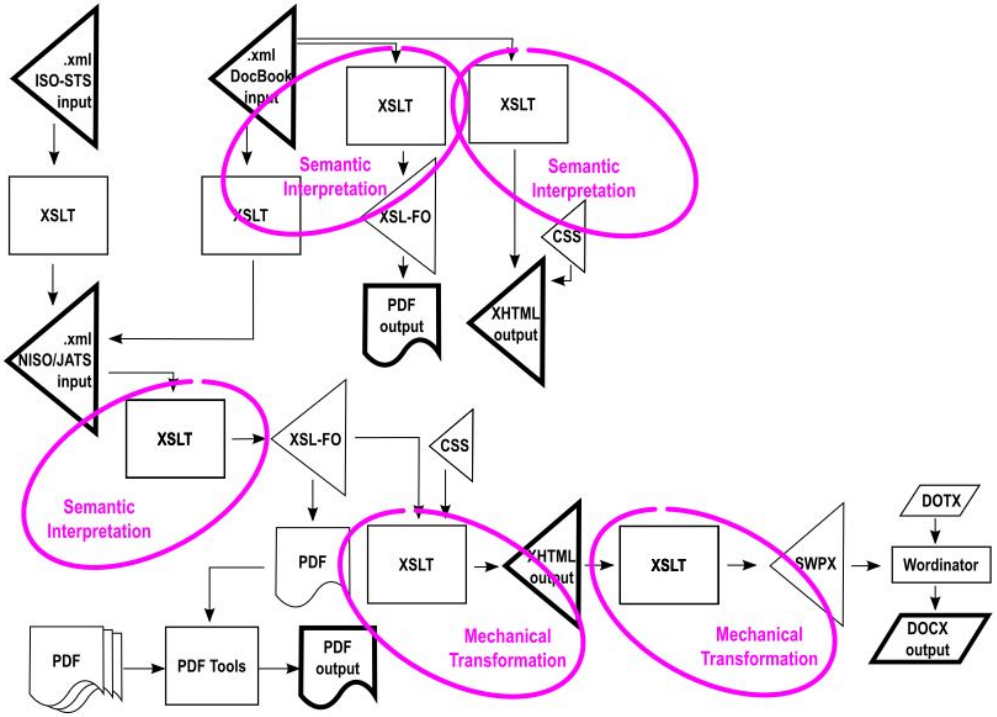
Réalta system inputs and outputs

DocBook and ISO-STS are transformed into NISO XML for JATS processing



Réalta system inputs and outputs (cont.)

JATS layout semantics interpreted into XSL-FO, mechanically producing others



DocBook stylesheet structure

The DocBook semantic stylesheets are well-engineered to maintain in parallel

xsl/fo/docbook.xsl

```
<xsl:include href="autotoc.xsl"/>
<xsl:include href="autoidx.xsl"/>
<xsl:include href="lists.xsl"/>
<xsl:include href="callout.xsl"/>
...
<xsl:include href="glossary.xsl"/>
<xsl:include href="block.xsl"/>
<xsl:include href="task.xsl"/>
<xsl:include href="qandaset.xsl"/>
<xsl:include href="synop.xsl"/>
...
...
<xsl:include href="../lib/lib.xsl"/>
<xsl:include href="../common/l10n.xsl"/>
<xsl:include href="../common/common.xsl"/>
<xsl:include href="../common/utility.xsl"/>
<xsl:include href="../common/labels.xsl"/>
<xsl:include href="../common/titles.xsl"/>
<xsl:include href="../common/subtitles.xsl"/>
<xsl:include href="../common/gentext.xsl"/>
<xsl:include href="../common/olink.xsl"/>
<xsl:include href="../common/targets.xsl"/>
<xsl:include href="../common/pi.xsl"/>
```

xsl/html/docbook.xsl

```
<xsl:include href="autotoc.xsl"/>
<xsl:include href="autoidx.xsl"/>
<xsl:include href="lists.xsl"/>
<xsl:include href="callout.xsl"/>
...
<xsl:include href="glossary.xsl"/>
<xsl:include href="block.xsl"/>
<xsl:include href="task.xsl"/>
<xsl:include href="qandaset.xsl"/>
<xsl:include href="synop.xsl"/>
...
...
<xsl:include href="../lib/lib.xsl"/>
<xsl:include href="../common/l10n.xsl"/>
<xsl:include href="../common/common.xsl"/>
<xsl:include href="../common/utility.xsl"/>
<xsl:include href="../common/labels.xsl"/>
<xsl:include href="../common/titles.xsl"/>
<xsl:include href="../common/subtitles.xsl"/>
<xsl:include href="../common/gentext.xsl"/>
<xsl:include href="../common/olink.xsl"/>
<xsl:include href="../common/targets.xsl"/>
<xsl:include href="../common/pi.xsl"/>
```

DocBook list processing

Each DocBook fragment interprets the DocBook semantic for the given output

xsl/fo/lists.xsl

```
<fo:list-block id="{$id}" xsl:use-attribute-sets="itemizedlist.properties">
  <xsl:attribute name="provisional-distance-between-starts">
    <xsl:value-of select="$label-width"/>
  </xsl:attribute>
  ...
  <xsl:copy-of select="$content"/>
</fo:list-block>
```

xsl/html/lists.xsl

```
<ul>
  <xsl:call-template name="generate.class.attribute">
    <xsl:with-param name="class" select="$default.class"/>
  </xsl:call-template>
  ...
  <xsl:apply-templates
        select="listitem
                |comment()[preceding-sibling::listitem]
                |processing-instruction()[preceding-sibling::listitem]"/>
</ul>
```

Réalta stylesheet structure

The Réalta semantic stylesheets are layered to accommodate client needs

Client adoptions of international standards:

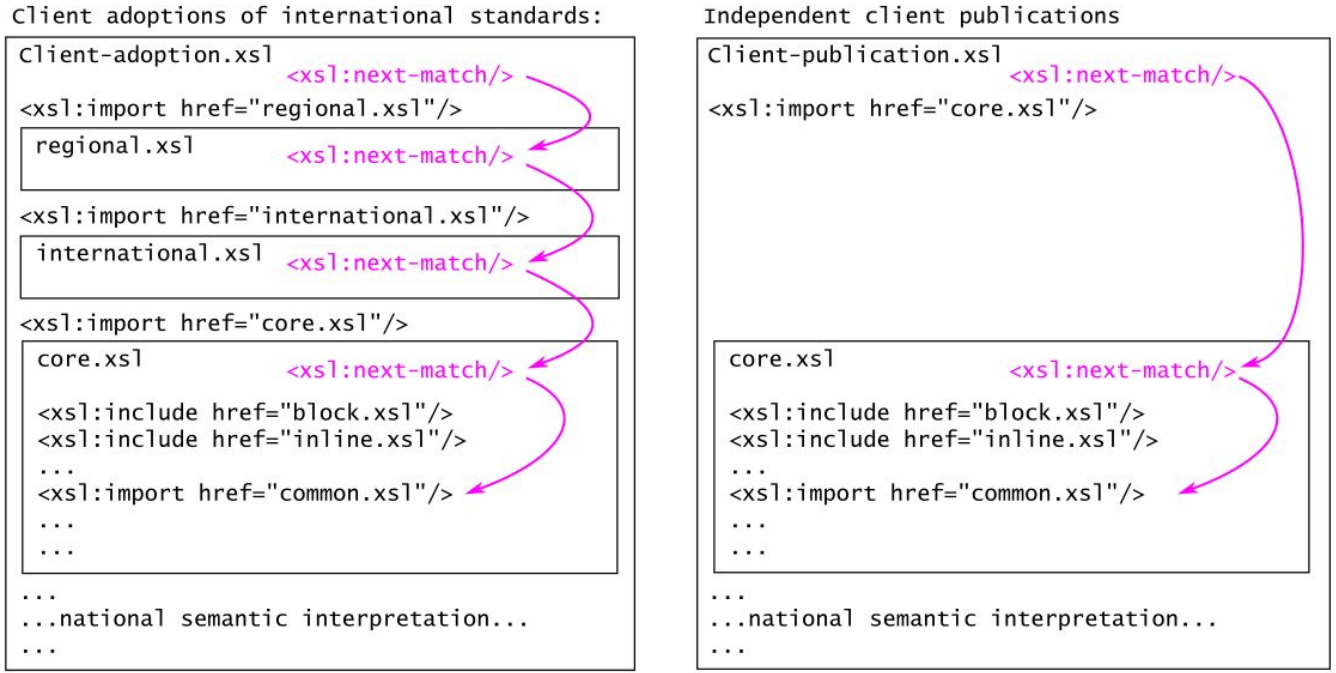
```
Client-adoption.xsl

<xsl:import href="regional.xsl"/>
  regional.xsl

<xsl:import href="international.xsl"/>
  international.xsl

<xsl:import href="core.xsl"/>
  core.xsl

  <xsl:include href="block.xsl"/>
  <xsl:include href="inline.xsl"/>
  ...
  <xsl:import href="common.xsl"/>
  ...
  ...

...
...national semantic interpretation...
...
```

Independent client publications

```
Client-publication.xsl

<xsl:import href="core.xsl"/>

  core.xsl

  <xsl:include href="block.xsl"/>
  <xsl:include href="inline.xsl"/>
  ...
  <xsl:import href="common.xsl"/>
  ...
  ...

...
...national semantic interpretation...
...
```

Réalta stylesheet structure (cont.)

The `<xsl:next-match/>` directive gives every layer the opportunity to produce



Réalta list processing

The list semantic handling is written once with information for downstream use

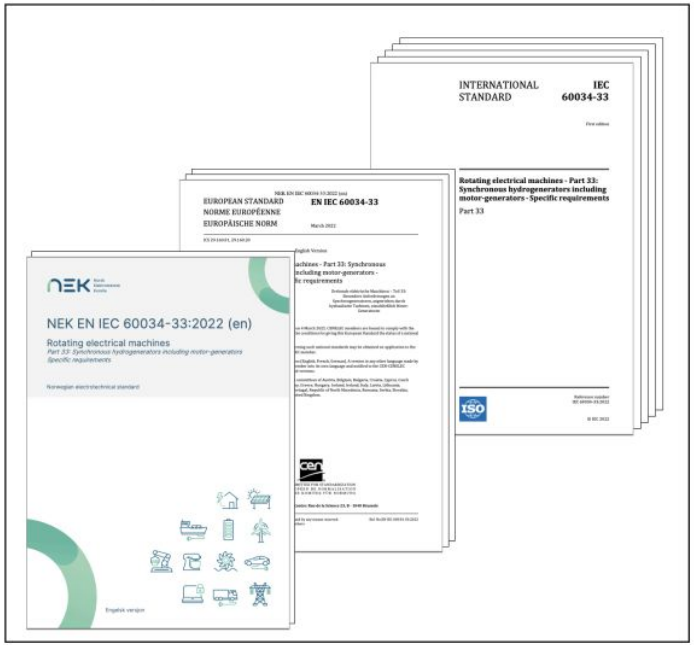
- `x:h="{HTML-element}*{CSS-class}"` accommodates type semantic for later use

```
<xsl:template match="list-item" priority="1">
  <list-item x:h="div*{name(.)}">
    <list-item-label x:h="div*item-label" end-indent="label-end()">
      <block text-align="start">
        <xsl:call-template name="c:listLabel"/>
      </block>
    </list-item-label>
    <list-item-body x:h="div*item-body" start-indent="body-start()">
      <block-container>
        <block start-indent="0pt">
          <xsl:apply-templates select="*[not(name(.)='label')]"/>
        </block>
      </block-container>
    </list-item-body>
  </list-item>
</xsl:template>
```

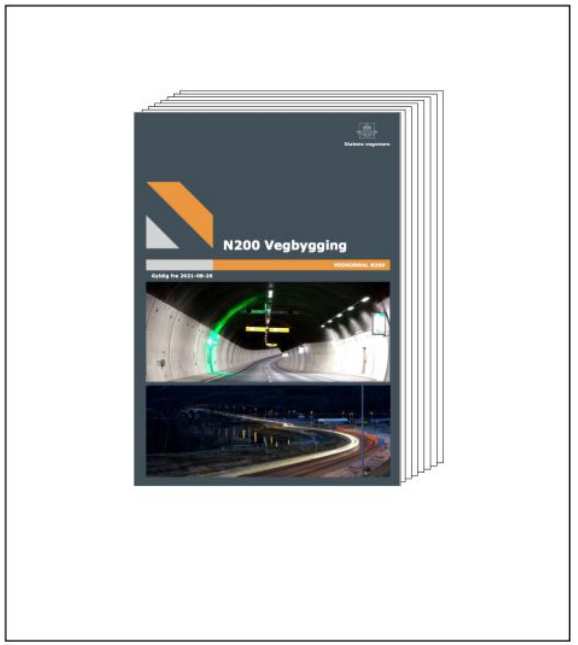
Réalta's clients' publications

Client publications need to support different renderings of same JATS semantics

Client adoptions of international standard:



Independent client publications:



Réalta's clients' publications (cont.)

Réalta or client specifies the fetching of XML from different sources using ISO-STS

client-adoption.xml

```
<standard xml:lang="en">
    <front>
        <nat-meta originator="NEK">
            <std-ref type="dated">NEK EN IEC 60034-33:2022</std-ref>
        </nat-meta>
        <std-meta std-meta-type="european">
            <custom-meta-group originator="realta">
                <custom-meta>
                    <meta-name>realta-fetch cenelec xml</meta-name>
                    <meta-value>68956</meta-value>
                </custom-meta>
            </custom-meta-group>
        </std-meta>
        <std-meta std-meta-type="international">
            <custom-meta-group originator="realta">
                <custom-meta>
                    <meta-name>realta-fetch iec xml</meta-name>
                    <meta-value>23163</meta-value>
                </custom-meta>
            </custom-meta-group>
        </std-meta>
        <sec id="sec_nat-foreword_en" sec-type="foreword" originator="NEK">
        <sec id="sec_nat-foreword_no" sec-type="foreword" originator="NEK" xml:lang="nb">
    </front>
    <body/>
</standard>
```

Réalta's clients' publications (cont.)

Réalta or client specifies the fetching of XML from different sources using NISO-STS

client-adoption-niso.xml

```
<adoption xml:lang="en">
    <adoption-front>
        <std-meta>
            <std-ref type="dated">NEK EN IEC 60034-33:2022</std-ref>
        </std-meta>
        <sec id="sec_nat-foreword_en" sec-type="foreword">
        <sec id="sec_nat-foreword_no" sec-type="foreword" xml:lang="nb">
    </adoption-front>
    <adoption>
      <adoption-front>
        <std-meta std-meta-type="european">
            <custom-meta-group originator="realta">
                <custom-meta>
                    <meta-name>realta-fetch cenelec xml</meta-name>
                    <meta-value>68956</meta-value>
                </custom-meta>
            </custom-meta-group>
        </std-meta>
      </adoption-front>
      <standard>
        <front>
          <std-meta std-meta-type="international">
              <custom-meta-group originator="realta">
                  <custom-meta>
                      <meta-name>realta-fetch iec xml</meta-name>
                      <meta-value>23163</meta-value>
                  </custom-meta>
              </custom-meta-group>
          </std-meta>
        </front>
        <body/>
      </standard>
    </adoption>
</adoption>
```

Réalta's clients' publications (cont.)

The fetched files are merged into a monolithic NISO-STS file with adoption layers

client-niso.xml

```
<adoption>
  <adoption-front>
    <std-meta>
      <std-ref type="dated">NEK EN IEC 60034-33:2022</std-ref>
    </std-meta>
    <sec id="sec_nat-foreword_en" sec-type="foreword">
    <sec id="sec_nat-foreword_no" sec-type="foreword" xml:lang="nb">
  </adoption-front>
  <adoption xml:lang="en">
    <adoption-front>
      <std-meta std-meta-type="european">
        <std-ref type="dated">EN IEC 60034-33:2022</std-ref>
      </std-meta>
    </adoption-front>
    <back>...</back>
    <standard>
      <front>
        <std-meta std-meta-type="international">
          <std-ref type="dated">IEC 60034-33:2022</std-ref>
        </std-meta>
      </front>
      <body>...</body>
      <back>...</back>
    </standard>
  </adoption>
</adoption>
```

Fidelity and flexibility for clients and their products leveraging XSLT and XSL-FO

G. Ken Holman
Réalta Online Publishing Solutions Limited
https://RealtaOnline.com

JATS-Con 2023 - June 13-14, 2023