Validation for Authoring VO Standards

Mark Taylor (University of Bristol)

Operations IG

IVOA Interop

Online

26 April 2022

$Id: docvalid.tex,v 1.10 2022/04/26 20:24:57 mbt Exp $

Validation during document authoring:

- Why?
- What?
- How?

Why Validate While Authoring?

It's a good idea to perform automatic checking of document contents while authoring it, or at least before it's published, in order to:

- Check the standard can be implemented
- Check you're not misleading standard users
- Save time later (avoid Errata)

Not mandatory, just (my idea of) good practice

- Tackles some operational issues by prevention rather than cure

Why: Errata

Some examples of Errata that early validation would have avoided:

- DataLink-1.0 Erratum 1: *illegal VOTable RESOURCE/type attribute value*
- DataLink-1.0 Erratum 4: *bad ref/ID linkage in VOTable example*
- VOSpace 2.1 Erratum 1: *invalid XML amongst other things*
- SSA-1.1 Erratum 2: *illegal UCDs*
- SODA 1.0 Erratum 1: *illegal UCD*
- ObsCore 1.1 Erratum 1: *illegal UCDs*
- SLAP 1.0 Erratum 1: *illegal UCDs*

What to Validate

Some things you can validate during document authoring:

- Normative text
  - ▷ Schemas
  - ▷ Prescribed UCDs
  - ▷ Prescribed (VO)Units

- Examples
  - ▷ XML schema instance documents
  - ▷ VOTables
  - ▷ Whatever it is you are standardising

Validation Mechanics

Separate validatable text from main document file:

- Content (e.g. XML documents) is easier to validate in free-standing files than embedded in LaTeX
- You can store validatable text in separate files and import:
  - ▷ `\lstinputlisting[language=XML]{example.xml}` (LaTeX *listings* package)

If possible, integrate into document build process

- ▷ IvoaTex provides Makefile `test` target
- ▷ Consider making document build dependent on tests?
  - ○ In some cases can be a bit of a pain (system-dependent tools, network dependency, slow runtime)
- ▷ At least: perform validations regularly, and immediately before publication

VO Validators

VOTable

- STILTS `votlint` validates VOTables
- Well-formed XML, XSD conformance, namespaces, UCDs and VOUnits, ID/refs, cell datatypes, ...

UCDs and VOUnits *especially where mandated by DMs*

- Ucidy Java library (Grégory Mantelet) validates UCDs
- Unity Java/C library (Norman Gray) validates VOUnits
- Use libraries directly, or via STILTS harness functions (see e.g. UCD/VOUnit usage report Nov 2021)

Others listed at IvoaValidatorsSummary wiki page

- There *should* be one for each applicable standard in PR/REC
- Some are more convenient for document build integration than others (standalone vs. web interface)

XML Validation

What to validate:

- If defining an XSD schema:
  - ▷ check the schema itself is correct XSD
  - ▷ check example XML documents conform to the schema, especially examples in text, maybe other test docs too

- If examples in text use an existing schema:
  - ▷ check XML examples conform to external schema(s)

- At least check that XML examples are well-formed!

How to validate:

- Fairly easy to check well-formed XML
  - ▷ Any XML parser should report errors

- Harder to validate against schemas
  - ▷ Tools exist (xmlstarlet, xmllint, others??)
  - ▷ Can be hard to configure (e.g. to use local copies of schemas under development)
  - ▷ STILTS 3.4-4 (04/2022) has command `xsdvalidate`
    - ○ XML Schema 1.0 only
    - ○ Somewhat experimental — please complain to me if it doesn't behave

Or maybe ... schematron?

Resources

- IvoaTex Note

  - in production: `https://github.com/ivoa-std/ivoatexDoc` (see especially Section 3.11)

- List of validators by standard

  - IvoaValidatorsSummary wiki page

- STILTS manual

  - Sections on `votlint`, `xsdvalidate`, `ucdStatus/vounitStatus`, `datalinklint`

Summary

- Standard has associated validator?
  - ▷ if not, should it?
  - ▷ run validation as part of build if possible/applicable

- Document defines XML Schema?
  - ▷ validate schema XSD itself
  - ▷ validate at least one schema instance against XSD
  - ▷ validate any example instances in the text against XSD

- Document prescribes UCDs?
  - ▷ check UCDs are legal UCD1+

- Document prescribes units?
  - ▷ check units are legal VOUnit (unless some reason why not)

- Document has example (complete/part) VOTables?
  - ▷ validate VOTables

- Document has other example XML snippets?
  - ▷ check XML is well-formed
  - ▷ validate if possible against existing XSDs

Discussion

- Opinions?

- Other tips, tools, suggestions?

- What is missing to make this easier for document authors?

CSP Discussion Input

Comments arising from CSP *"Publishing Data in the VO"* panel discussion (15:00 UTC today)

Alberto Micol: (presentation)

    ▷ Not always crystal clear standards (especially when dependending on cascade of standards), get ready to contact standard editors for clarifications

    ▷ At times standards evolve in unexpected ways

Tom Donaldson: (paraphrased in session notes)

    ▷ Standards should be very simple without fancy addons to address specific science cases

Anne Raugh: (zoom chat)

    ▷ As a reader of standards without years of IVOA background, I have noted that some standards are quite readable and self-contained in their descriptions, while others heavily reference other standards, to the point that they are incomprehensible without having memorized the contents of the references.

Renaud Savalle/**Paul Harrison**: (informal discussion)

    ▷ IVOA standards beauty(/ugliness) contest?