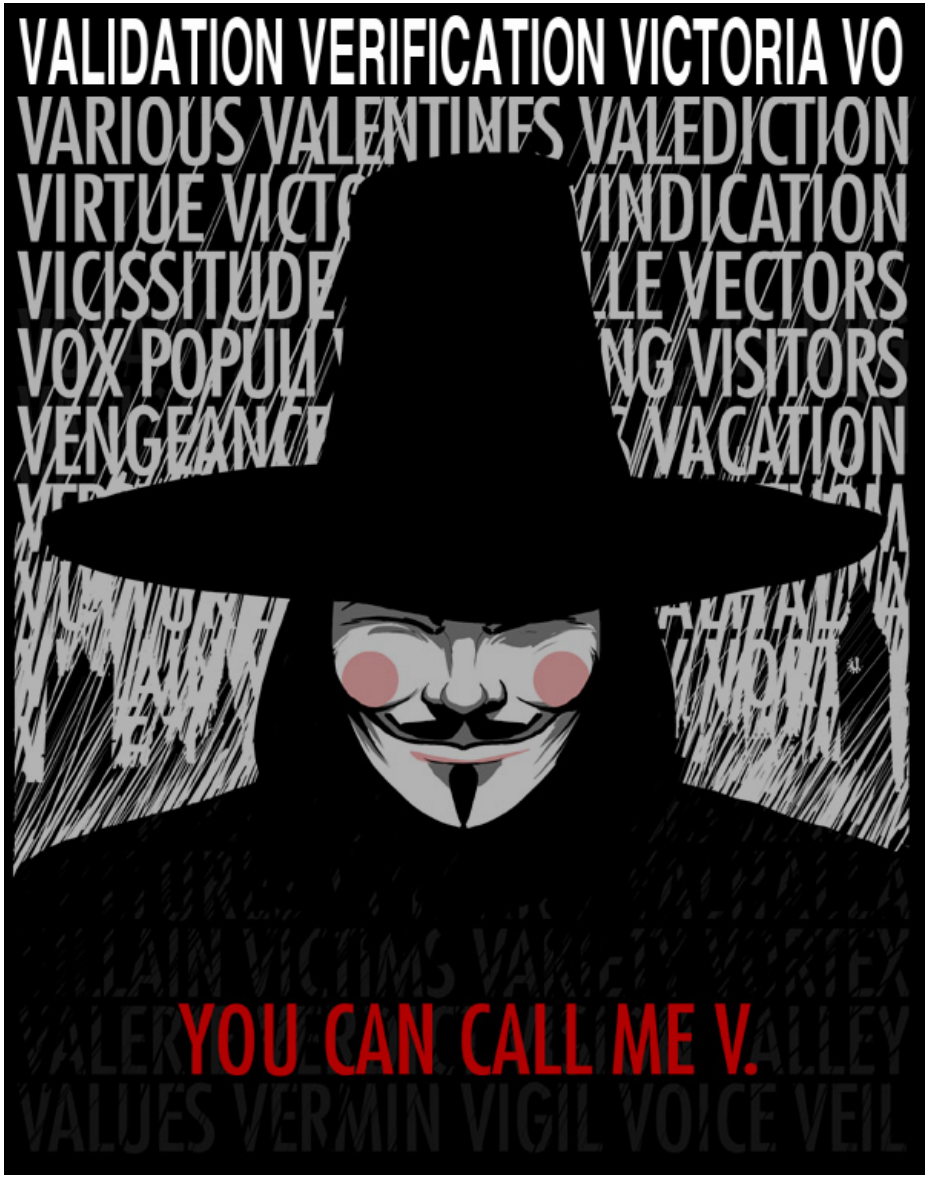
Validation and Verification: Applications Perspective

Mark Taylor (Bristol)

IVOA Interop,
Victoria
18 May 2010

$Id: vv.tex,v 1.9 2010/05/18 05:52:46 mbt Exp $

Validation & Verification
mean different things
to different people



I'll be talking about:

- (Non-) compliance of services to standards
- How to deal with it from an application's point of view

Non-compliant services

- Why are non-compliant services bad?
  - Functionality might not work
  - Functionality might work badly
    - wrong results to user?
    - defective data structures $\longrightarrow$ errors downstream?
    - . . .
  - Need to account for *possibility* of non-compliance in code
  - Users blame the tool

VOTable

- Many (most?) VOTables are broken
  - Badly-formed (incorrect XML)
    - `<PARAM name='Exp' datatype='char' arraysize='*' value='< 1min'/>`
      (unescaped '<' character)
  - Well-formed but invalid (does not validate against schema)
    - `<PARAM name='Note' value='Short exposure'/>`
      (missing `datatype` attribute)
  - Valid but incorrect (breaches rules in VOTable standard)
    - `<PARAM name='Exposure' datatype='int' value='small'/>`
      (non-numeric value in numeric field)
  - Correct but misleading (strict interpretation is not what's meant)
    - `<PARAM name='Note' datatype='char' value='Short exposure'/>`
      (missing `arraysize` attribute implies single character value, i.e. Note='S')

Policies

How to deal with non-compliant data?

- Actively look for non-compliance?
- In case of non-compliance:
  - Refuse to process altogether?
    - Might result in better compliance (more complaints to service owners)
    - But users don't like it
    - Also can make software less robust to standard version changes
  - Best effort to process if meaning is unambiguous?
  - Best effort to process guess what was meant?

Compare case of XML and HTML

TOPCAT Policy

TOPCAT's policy (not necessarily exemplary):

- In general:
  - Don't expend unnecessary effort to seek non-compliance
    - VOTable: look for relevant elements/atts; don't worry about irrelevant ones
    - . . .
  - Best efforts to process where can guess intended meaning
    - Cone: allow UCD1+ instead of UCD1
    - UCD: don't enforce case, allow any reasonable delimiters
    - VOTable: assume `"char"` for missing `datatype`
    - . . .
  - Warn through logging system for non-compliance *(but may go unseen)*
  - In case of failure, try to provide detail in error message (blaming service)
- In some cases (VOTable):
  - Option to enforce strict interpretation
    - `-Dvotable.strict` system property
  - Validator provided as part of software
    - STILTS `votlint` — written as necessary measure during development

Prevention is Better Than Cure

How to avoid non-compliant data?

- Validation:
  - Provision of good (comprehensive, easy to use) validation tools
  - Use of validation tools by service owners on their services
  - Use of validation tools by others, with feedback to service owners
    - Disgruntled users?
    - Disgruntled application authors?
    - IVOA?
    - Registry owners?
    - VO Vigilantes?
- Good standards:
  - Easy (and possible!) to implement
    - No inconsistencies
    - Clear, short, not too many options
    - Implementation early in standard development helps to ensure this!
  - Difficult to implement incorrectly
    - Avoid having to provide redundant information

Conclusions

Lessons:

- Having a comprehensive validation tool makes life much easier
  - ▷ can convince the user that the service (not the application) is at fault
  - ▷ can convince the service provider that the service is at fault
  - ▷ can aid the service provider in producing compliant services
- XML Schema Validation is of only limited help
  - ▷ just because it validates against the schema doesn't mean it's correct, or even usable
- Validators have to be used
- Specify standards which are easy to implement and hard to violate
- Non-compliant services are a fact of life ☹

Questions:

- What are the best policies for applications to adopt?
- How do we ensure that service providers make use of available validation tools?