

Federal XML Naming and Design Rules Changes Proposed by the IC Metadata Working Group Staff

July 19, 2005

Position paper prepared by the
[Intelligence Community Metadata Working Group Staff](#)

ADNI/CIO/ICCIO/ISC – IC MWG

Table of Contents

Federal XML Naming and Design Rules Changes Proposed by the IC Metadata Working Group Staff.....	1
1 Summary	3
2 Background.....	3
2.1 Federal XML Naming and Design Rules.....	3
2.2 IC MWG	4
2.2.1 IC Metadata Standard for Publications (IC MSP)	5
2.2.2 IC Metadata Standard for Information Security Markings (IC ISM)	5
3 General Concerns from the IC MWG Staff	5
4 Major Incompatibilities Inhibiting IC Adoption.....	6
4.1 Define the Adoption Process	6
4.2 Allow Abbreviations and Acronyms	6
4.3 Need for DTDs.....	7
4.4 Clarify Uniqueness of Dictionary Entry Name (DEN).....	7
4.5 Need for xsd:choice	8
4.6 Need for xsd:any	8
4.7 Need for xsd:anyAttribute.....	9
4.8 Need for xsd:union.....	9
4.9 Need for Empty Elements	10
5 Other Significant Impediments to IC Adoption.....	11
5.1 Propose Guidelines Rather Than Enforce Rules.....	11
5.2 Allow for Possibility of Mixed Content.....	11
5.3 Versioning Scheme is Not Well Defined.....	11
5.4 Permit Non-Conformant Imports.....	11
5.5 Permit xsd:all	12
5.6 Permit xsd:appinfo	12
5.7 Elaborate on Code List Subsetting Mechanism	13
5.8 Run-time vs. Fully Documented Variants of Schemas.....	13
5.9 Oxford English vs. American Spelling	14
5.10 Modularity Section Needs Concrete Examples	15
5.11 Permit Prepositions, Conjunctions and Articles in Element Names.....	15
5.12 Voluntary Consensus Standards	16
5.13 Permit Specifications that are Not Quite Recommendations.....	16
6 How to Respond to This Position Paper	17

1 Summary

This position paper covers aspects of the draft *Federal XML Naming and Design Rules*,¹ dated June 9, 2005, which are incompatible with certain practices and needs of the Intelligence Community (IC), as promulgated by the IC Metadata Working Group (IC MWG).

Our community has been developing and using XML Schemas and DTDs for many years. We have enthusiastically promulgated the guidelines from the April 2002 *Draft Federal XML Developer's Guide* and would like to be able to leverage the emerging 2005 guidance as well. However, the June 2005 draft contains a number of rules that would severely conflict with work the IC MWG has already completed, impacting up to 35 organizations affected by IC policy on metadata and metadata markup. The rules also interfere at a technical level with some of the near term changes planned by the IC MWG.

Issues are divided into two categories:

1. True incompatibilities that are likely to inhibit IC adoption of this guidance; and
2. Aspects that, although possibly not “showstoppers” for the IC, are nonetheless perceived as significant impediments to our accepting this guidance.

We believe that the draft document as written will be unacceptable to the IC and could significantly hamper our efforts to create XML Schema pursuant to our mission objectives. We remain hopeful, however, that changes in tone and specific rules will result in guidance that will be truly beneficial to the federal government in general and to the Intelligence Community in particular.

2 Background

2.1 *Federal XML Naming and Design Rules*

The emerging *Federal XML Naming and Design Rules* have been under development since May 2005 with a draft released June 9, 2005. A new version is due in July with possible presentation to the Architecture and Infrastructure Committee (AIC) of the Federal CIO Council in August 2005. According to the June draft, the intention is to promote the document as government-wide policy.

1.2 SCOPE: This Federal XML Naming and Design Rules document is applicable to all Executive Branch Departments and Agencies (hereinafter referred to as Agency) that use XML, including all commercial and government off-the-shelf XML related product implementations. It is applicable to all contractors and vendors doing XML development work on behalf of Departments and Agencies.

¹ It is our understanding that the title of this document may be changed to, “US Federal XML Developer’s Guide.”

Although the scope is quite broad, the number of people who have been actively following this government-wide effort is very small (roughly one dozen people have posted to the listserv's mailing list) and very few agencies are either officially or unofficially represented. (Those represented include GSA, IRS, EPA, DHS, and the IC). The adoption process is still being debated as of this writing. For example, it is particularly unclear if, when, and how all interested agencies will be given an opportunity to review and comment on the document before it is offered as government-wide policy.

This draft document is essentially a re-write of three earlier "NDR" (Naming and Design Rules) documents from [OASIS UBL, UN/CEFACT, and the Department of the Navy](#). In contrast to the April 2002 *Draft Federal XML Developer's Guide*, this new document proposes to mandate approximately 150 rules for XML Schemas that were originally developed in the context of EDI-like transactions (UBL and ebXML). While the earlier NDRs were acceptable to e-commerce communities of interest (CoI), this does not mean that the same (or very similar) rules are 100 per cent applicable to the all XML applications of all federal agencies. The heavy EDI and e-commerce influence is obvious in section 1.5.2, "Design For Extensibility." Although the Department of the Navy has adopted an NDR, it is our understanding that it is not universally applied throughout that department. It is encouraging that IRS has customized the NDR to meet its specific mission objectives, but it is unclear what would happen if the federal version becomes government-wide policy. It is our hope that agencies may view the federal version as guidance and choose to either tighten or relax some of its rules according to their own business needs. It is certainly difficult to imagine a "one-size-fits-all" approach for the entire federal government unless agencies are given sufficient latitude to tailor the guidelines.

As of July 12, 2005, even the title of the federal NDR document was still being debated, not to mention many of the rules. When even a small group of participants cannot reach consensus on a set of rules for XML Schema, it is hard to understand how limiting XML developers by prohibiting the use of some XML Schema constructs and making mandatory a certain rigid structure could possibly meet the needs of as diverse a user base as the US Federal Government, comprised as it is of numerous agencies with mission objectives that cover a gamut far broader than e-commerce.

2.2 IC MWG

To better understand the needs of the IC MWG which includes participants from all [15 of the IC agencies](#), this section provides brief background information, primarily about our XML schemas (XSDs and DTDs).

The IC MWG leads the metadata standards activities within the IC for the purposes of data interoperability and information discovery. The IC MWG Chair is Tim West (DIA). The IC Markup Languages Panel (ICML) is a working panel under the IC MWG convened for the purpose of review and oversight of the IC metadata standards described in the following subsections. (Two ICML efforts, IC Core and IC HTML, are not

discussed since they are not directly relevant to XML Schema development.) [Contact information](#) is available.

Note that the work of the IC MWG is directly relevant to compliance with [Executive Order 13356](#) (“Strengthening the Sharing of Terrorism Information To Protect Americans,” 27 August 2004) which mandates that a broad range of actions be undertaken culminating with providing the President a plan to create an interoperable “Trusted Terrorism Information Sharing Environment.”

2.2.1 IC Metadata Standard for Publications (IC MSP)

The [Intelligence Community Metadata Standard for Publications](#) (IC MSP) is a set of XML schemas for generic publication types, along with documentation in the form of a data element dictionary and an implementation guide. The schemas are implemented in two syntaxes; XML document type definitions (DTDs) and W3C XML Schemas. The standard has been developed as a cooperative effort in response to requests by numerous organizations within the IC to have an IC-wide mandated XML model to support interoperability of intelligence content across producers and consumers of information within the Community.

2.2.2 IC Metadata Standard for Information Security Markings (IC ISM)

The [Intelligence Community Metadata Standard for Information Security Markings](#) (IC ISM) consists of a set of XML attributes that may be used to associate security-related metadata with XML elements in documents, web-service transactions, or data streams. It is distributed as both an XML entity set and W3C XML Schema (WXS) so that the XML attributes defined in the standard can be incorporated into any XML document type definition (DTD) or schema. Made available along with the IC ISM entity set and WXS are controlled vocabularies of terms that are used as the sources for the values of the IC ISM attributes.

3 General Concerns from the IC MWG Staff

The IC MWG staff has thoroughly evaluated the June 9, 2005 draft of the Federal XML Naming and Design Rules. We do not believe this document presently meets the needs of the IC. The proposed rules are too restrictive for the many applications of XML currently in use or under development. This position paper details specific objections and changes required by the IC. Our expectation is that every requested change will be made, or, if not, each exception will be answered in a formal response.

In general, our chief concerns are (a) whether this document is intended to become strict policy or useful guidance and (b) why the scope is so broad. With respect to the former, it is our considered opinion that the goal should be best-practices guidance. In our opinion, the current draft version of document suffers from a lack of examples and justifications for its “rules”. Earlier XML guidance documents, such as from EPA in 2003 (available [here](#) and [here](#)) and from the Justice XML Structure Task Force, were more acceptable as guidance since they provided detailed explanations and justifications for each rule, providing developers a good sense of the issues and rationale behind each decision.

With respect to scope, we are especially concerned whether strict rules are truly necessary on a government-wide basis, including applications internal to agencies. If they are, then they must be formally vetted with each federal agency to reach consensus. Such rules should only be as restrictive as absolutely necessary to encourage interoperability and data sharing. In particular, technical implementation policy that infers with critical mission objectives is unacceptable to the Intelligence Community. It is therefore imperative that any federal XML guidance or policy address the actual needs of those who are to follow it.

4 Major Incompatibilities Inhibiting IC Adoption

The following rules are perceived by the IC MWG staff to be, in all likelihood, unacceptable to the IC members. We recommend changing the rule as indicated to make the guidance acceptable to the IC.

Alphanumerics in brackets such as “[IND1]” are rule identifiers based on the June 9th draft. “IC CR” stands for *Intelligence Community Change Request*.

4.1 Define the Adoption Process

It is imperative that the road to government-wide adoption is clearly paved. At this time, neither the exact steps necessary to reach consensus nor the timeline for the adoption process is well defined.

4.2 Allow Abbreviations and Acronyms

Rule [DEN13]: “Abbreviations and acronyms that are part of the *Dictionary Entry Name* shall be expanded or explained in the definition.”

Rule [GNR4]: “XML element, attribute, and simple and complex type names MUST NOT use acronyms, abbreviations, or other word truncations, except those in the list of exceptions published in Appendix XX.”

Rule [GNR5]: “Acronyms and abbreviations MUST only be added to the federal approved acronym and abbreviation list after careful consideration for maximum understanding and reuse.”

While we would definitely agree that abbreviations and acronyms should be fully documented (either in a Data Element Dictionary and/or in an `xsd:documentation` element within the schema itself), we strongly disagree that element, attribute, and complexType names MUST NOT use abbreviations or acronyms. Every government agency and every CoI has its own list of shortened terms that are widely known by its members.

IC CR for [GNR4]: We request requiring [DEN13] to guarantee that the shortened form is expanded for the benefit of anyone who is new to that CoI or agency. *We can therefore eliminate* [GNR4] which results in long XML names that are redundant information to the vast majority. Interoperability across different CoIs and agencies can still be achieved because the definitions will contain the expansion. Furthermore, the definition includes

more about the semantics than can be gleaned merely by expanding the abbreviation or acronym. An example from IC-ISM-v2.xsd illustrates these points:

```
<xsd:attribute name="SARIdentifier">
  <xsd:annotation>
    <xsd:documentation xml:lang="en">
      Authorized Special Access Required program digraph(s) or
      trigraph(s) preceded by "SAR-". Either (a) a single digraph or
      trigraph or (b) a space-delimited list of digraphs or trigraphs.
      Example: "SAR-ABC SAR-DEF ..."
    </xsd:documentation>
  </xsd:annotation>
<xsd:simpleType>
  <xsd:restriction base="xsd:NMTOKENS"/>
</xsd:simpleType>
</xsd:attribute>
```

To simply rename the attribute `SpecialAccessRequiredIdentifier` would tell the user nothing about digraphs, trigraphs, or lists of same. Interoperability without the complete definition is not likely.

In fact, in the Intelligence Community, there are many acronyms and abbreviations for which the expansion of the shortened form is classified even though the shortened form is unclassified. To expand these terms would be a security violation.

Therefore, we strongly oppose [GNR4] (especially since the 2002 draft guidance endorsed the use of commonly-understood abbreviations and acronyms).

Rule [GNR5] is simply impractical without a clearly defined abbreviation and acronym registry, not to mention an acceptance process and an arbitration process when two shortened forms collide. The IC does not feel that this can happen in a timely manner and is largely unnecessary since namespaces associated with different schemas should suffice to disambiguate collisions. In other words, let each CoI, LOB, or agency define its own shortened forms, include expansions in element definitions, and publish and distribute lists in a manner consistent with their other needs.

4.3 ***Need for DTDs***

Rule [IND1]: “All instance documents MUST validate to a corresponding XSD schema.”

Discussion: The IC has members who, for various reasons, cannot use the current XSD tools and therefore MUST rely on DTDs which have a longer history of tool support.

IC CR of [IND1]: “All instance documents MUST validate against either an XML Schema or a DTD, with XML Schema being preferred for new development efforts.”

4.4 ***Clarify Uniqueness of Dictionary Entry Name (DEN)***

Rule [DEN7]: “The *Dictionary Entry Name* shall be unique.”

What is the scope of this statement? Unique across the entire federal government? Per agency? Per project? Per namespace? Per schema? We hope it only means per schema or namespace, otherwise it is virtually impossible to control or even ascertain uniqueness until there is some reliable, federated registry accessible to all government developers and presumably contractors. Perhaps this rule needs to be explicitly related to the Global Element rule?

However, if the scope of uniqueness is intended to be government-wide, which agency receives the honor of registering the DEN "Person"? DOJ? IRS? DHS? HHS? The IC? Is it simply "first come, first served"? Presumably each of these agencies (and others) will all want to use the "Person" DEN and each is likely to have different definitions both at the semantic level as well as at the child and attribute level?

Where is the process defined in the NDR for registering these "unique" DENs? Who is the authority? Which agencies are on the Control Board? How long does the process take? All of these questions must be answered before the IC can accept this rule.

4.5 Need for *xsd:choice*

Rule [GXS9]: "The `xsd:choice` element SHOULD NOT be used where customization and extensibility are a concern."

While SHOULD NOT is less restrictive than MUST NOT, the draft does not present a very strong case why `xsd:choice` is even less preferred than `xsd:sequence`. Our IC MSP schema collectively have 34 instances of `xsd:choice`. In fact, the OASIS Table Model, a familiar and widely used standard, has 3 instances of `xsd:choice` as well.

IC CR for [GXS9]: Consider eliminating this rule completely. It is not relevant for government-wide guidance or policy.

4.6 Need for *xsd:any*

Rule [ELD8]: "The `xsd:any` element MUST NOT be used."

There are situations in which any well-formed content is acceptable. In such situations, the inner content need not be checked for validity or, if necessary, validity of the inner content can be ascertained by other application-specific means. For example, the XSD language itself allows `xsd:any` children of the `xsd:documentation` element, so schema authors can add text or HTML or even some other custom XML markup within.

Furthermore, `xsd:any` is needed by the IC to permit our users to add their own extensions to our schema. This is different from merely extending our custom datatypes. For example:

```
<xsd:complexType name="ExtensionElementsType">
```



```
<xsd:sequence>
  <xsd:any namespace="##other" processContents="lax"/>
</xsd:sequence>
</xsd:complexType>
```

Our purpose for using such wildcards is to allow domain-specific data—such as geospatial, signals or other data—to be included in instance documents that conform to a common core data model. Parsers will validate the domain-specific data if an applicable schema is accessible. To preclude indeterminate models we require extension elements to be from a different namespace and we enclose the extensions in a special wrapper element, as shown above. The DoD’s Defense Discovery Metadata Specification (DDMS) uses the same wildcard-based approach.

IC CR for [ELD8]: “The `xsd:any` element SHOULD NOT be used, unless the content of the parent element can be any well-formed XML for which validity is not a primary concern, or in cases where providing an extension mechanism is necessary.”

4.7 Need for `xsd:anyAttribute`

Rule [ATD5]: “The `xsd:anyAttribute` MUST NOT be used for data centric schema. The `xsd:anyAttribute` MAY be used for document-centric schema if consistency in [sic] not an issue.”

Similar to the IC need for `xsd:any`, `xsd:anyAttribute` is needed by the IC to permit users to extend our schema to add any special attributes they require. There are 46 instances of `xsd:anyAttribute` used in IC MSP, all of which are as follows:

```
<xsd:anyAttribute namespace="##other" processContents="lax"/>
```

IC CR for [ATD5]: “The `xsd:anyAttribute` SHOULD NOT be used for data centric schema. The `xsd:anyAttribute` MAY be used for document-centric schema especially when providing an extension mechanism is necessary.”

4.8 Need for `xsd:union`

Rule [GXS11]: “The `xsd:union` technique MUST NOT be used except for Code and Identifier Lists. The `xsd:union` technique MAY be used for Code and Identifier Lists.”

In government forms, there are cases in which either the value is from a controlled vocabulary (or enumeration), or it could be a generic response, such as “N/A” or “other”. The `xsd:union` construct is useful for defining such specialized types.

To rule out `xsd:union` would be unacceptable to the IC; `xsd:union` is being used in several schemas by several IC-wide programs. If the types that are joined by the union are strongly typed, the result is also. In version 3.0 of IC MSP schema, the IC MWG found the W3C built-in date/time types lacking, so we added three `xsd:unions` to

extend three XSD date types to accommodate date/times with hours and minutes but not seconds, a la ISO8601:

```
<!-- **** v3.0: Added a data type pattern for date/time specifica-
      tions (YYYY-MM-DDThh:mm±hh or YYYY-MM-DDThh:mm±hh:mm)
      without seconds. -->
<xsd:simpleType name="dateHourMinType">
  <xsd:restriction base="xsd:string">
    <xsd:pattern
      value="\d{4}-\d{2}-\d{2}T\d{2}:\d{2}(Z|[\-\+]\d{2}:\d{2})"/>
    </xsd:restriction>
  </xsd:simpleType>

<!-- **** v3.0: Added a union data type for date specifications. -->
<xsd:simpleType name="ISO8601DateType">
  <xsd:union memberTypes="xsd:date xsd:gYearMonth xsd:gYear"/>
</xsd:simpleType>

<!-- **** v3.0: Added a union data type for date/time specifica-
      tions. -->
<xsd:simpleType name="ISO8601DateTimeType">
  <xsd:union memberTypes="xsd:dateTime dateHourMinType xsd:date
xsd:gYear xsd:gYearMonth"/>
</xsd:simpleType>
```

The DDMS uses `xsd:union` as well to allow strings like “none” or “not applicable” to be used in conjunction with built-in types like `xsd:dateTime` or with enumerations.

IC CR for [GXS11]: "If you use `xsd:union`, you SHOULD ensure that the new type is globally defined and is comprised of types that are as tightly constrained as possible. For example, a union of `xsd:date` and `xsd:string` is too loose. A union of `xsd:date` and an `xsd:string` restricted to a specific pattern would be acceptable."

4.9 *Need for Empty Elements*

Rule [ELD7]: “Empty elements MUST not be declared.”

The IC sees no reason to categorically rule out empty elements (which are devoid of content but may contain attributes). It is possible to declare a given element to be empty yet required. It is also possible to declare that the empty element contain either optional or required attributes. The only kind of empty element that leads to ambiguity is one which is both optional and does not have attributes. In publishing practice, an empty and attributeless element, such as `<toc>`, is often used as a placeholder to trigger stylesheets or composition software to generate a table of contents and insert it in place of the element.

IC CR for [ELD7]: “In general, optional empty elements without attributes SHOULD NOT be used. However, empty elements that are either required or that contain one or more attributes are acceptable.”

5 Other Significant Impediments to IC Adoption

The objections listed in this section are also important to the IC, although they are secondary to those in the preceding section. However, a formal response will be appreciated for each objection that is not addressed by changes to the *Federal XML Naming and Design Rules*.

5.1 **Propose Guidelines Rather Than Enforce Rules**

The IC MWG does not believe firm XML Schema rules can be established government-wide without significant participation from all federal agencies. Until that time, we request that the NDR be renamed the “US Federal XML Developer’s Guide” to reflect that it is guidance not requiring mandatory compliance.

5.2 **Allow for Possibility of Mixed Content**

Rule [MDC2]: “Mixed content **MUST NOT** be used in data centric schema except where contained in an `xsd:documentation` element.”

IC MWG acknowledges that in general data-centric schema should not use mixed content. However, when serializing XML from a database, it is possible that text content will contain XML markup.

IC CR for [MDC2]: “Mixed content **SHOULD NOT** be used in data centric schema except where contained in an `xsd:documentation` element or where serialization may contain markup.”

5.3 **Versioning Scheme is Not Well Defined**

All [VER*] rules need further elaboration with detailed examples before the IC can decide whether the rules are acceptable. We believe it is extremely difficult to determine versioning rules that work well across agencies.

Rule [VER8]: “A minor version document schema **MUST** import its immediately preceding version document schema.”

The IC believes that [VER8] assumes very strict configuration of incremental versions that may not accommodate some mission critical needs. Please relax this rule.

IC CR for [VER8]: “A minor version document schema **SHOULD** import its immediately preceding version document schema.”

5.4 **Permit Non-Conformant Imports**

Rule [SSM4]: “All imported schema modules **MUST** be fully conformant with the Federal XML naming and design rules. [sic]”

This is simply not practical. By definition, importing another schema means reusing it “as is” from another namespace. The IC may import something from DoD, or vice versa. Or the IC may import a schema that is already considered a voluntary consensus standard (VCS). The IC as well as any other organization may have no control over the degree of conformance of imports in various situations, yet we may need to leverage them anyway. For example, an imported controlled vocabulary (code list) from an outside authoritative source is not likely to conform to federal rules. We completely defeat the reuse concept if we have to re-write a schema from another namespace just so it conforms to the NDR.

IC CR for [SSM4]: “Whenever possible, imported schema modules SHOULD be fully conformant with the rules set forth in this document, except when imports are from sources over which the organization has no control.”

5.5 Permit *xsd:all*

Rule [GXS8]: “The `xsd:all` element MUST NOT be used in data centric schema.”

While the IC MSP schemas do not presently use `xsd:all`, the IC MWG staff does not believe it is wise to rule out its potential use. The IC acknowledges that `xsd:all` has certain limitations that make it undesirable but we would prefer to simply cite the limitations and let developers decide whether the construct is useful regardless.

IC CR for [GXS8] The `xsd:all` element SHOULD be used sparingly if at all in data-centric schema. The `xsd:all` element MAY BE used in document-centric schema.

JUSTIFICATION (please add to NDR document): Since `xsd:all` must appear as the first part of a content model, since it cannot be used in an extension, since it permits its children to appear in any order, and since it does not permit cardinality other than 0 or 1 for its children (no repeated children), this compositor is of limited utility in data-centric applications.

5.6 Permit *xsd:appinfo*

Rule [GXS12]: “Federal or Agency schema MUST NOT use `xsd:appinfo`.”

The IC notes that IRS has modified this rule as follows:

[GXS8] “IRS schemas SHOULD NOT use `xsd:appinfo`. If used, `xsd:appinfo` MUST only be used to convey non-normative information.”

IC MSP uses `xsd:appinfo` in 7 places. In each case, it merely identifies the version number of a schema module. The IC MWG could represent the version by other methods, but we see no harm in indicating such normative information via `xsd:appinfo` since it is not crucial to the use of the schema.

IC CR for [GXS12]: “Agencies SHOULD NOT use `xsd:appinfo` for normative information, except if such information is not crucial to the interoperability of the schema.”

5.7 **Elaborate on Code List Subsetting Mechanism**

Rule [CIL7]: “Users of the Federal or Agency Library MAY identify any subset they wish from an identified code or identifier list for their own trading community conformance requirements.”

Code lists, also known as *controlled vocabularies* or *authority files*, are very important to the IC MWG, particularly to our [Enablement Strategy](#). It is the goal of the IC MWG [Authority File Management System](#) to facilitate the central management of these controlled lists of terms for code lists representing values from FIPS, IANA, ISO, and so on.

How exactly will this federal NDR subsetting mechanism work, in particular if we want to subset an externally maintained list? For example, in the [US Postal Service's state abbreviations](#), what if we wish to disallow the last three (or six?) codes which are not actually part of the 50 states? Or, for any given intelligence analyst, what if we wish to provide a restricted list of ISO country codes?

IC CR for [CIL7]: The IC requests a detailed example of code lists including an example that demonstrates how to subset a list. We would also like an example of identifier lists and an explanation of how they differ from code lists.

5.8 **Run-time vs. Fully Documented Variants of Schemas**

Rule [GXS2]: “Federal and Agency schema should provide two normative schemas for each transaction. One schema shall be fully annotated. One schema shall be a run-time schema devoid of documentation.”

Discussion of this rule on the NDR listserv indicates that the two versions of each schema differ only in that the run-time versions are devoid of `xsd:annotations` elements, which contain `xsd:documentation` children (and possibly `xsd:appinfo` children). At a minimum, the wording of the rule needs to be clarified accordingly.

However, the IC MWG is unclear what the compelling use case makes this rule necessary? Even if such a use case can be presented, in order to minimize the extra configuration management work this causes for each agency, the IC feels strongly that an XSLT stylesheet MUST be included with the guidance document that will strip `xsd:annotation` elements and their children from a schema. (IC MWG staff has already posted such a stylesheet to the NDR listserv.)

IC CR for [GXS2]: The IC MWG recommends eliminating this rule. However, if a good use case can be presented as to why it is necessary, please change the wording to: “Agencies SHOULD provide two normative variants of each XML Schema. One variant

shall be fully annotated. The other variant SHOULD be a run-time schema with all `xsd:annotation` elements and any `xsd:documentation` and `xsd:appinfo` children elements removed.”

5.9 Oxford English vs. American Spelling

Section 4.1.1 of the NDR states: “All official dictionary entries will be in English. Due to the growing exchange of data between federal agencies and coalition partners, an authoritative source that will ensure absolute clarity and understanding of the names and definitions is required. The Oxford English Dictionary is that authoritative source. Specifically, Oxford English Dictionary American spellings will be used as the primary source.”

Rule [DEN1]: “The dictionary content, with the exception of *Business Terms*, shall be in the *English Language* following the primary *Oxford English Dictionary* American spellings to assure unambiguous spelling.”

Rule [GNR1]: “XML element, attribute and type names MUST be in the English language, using the primary American spellings provided in the Oxford English Dictionary for writers and editors.”

While the IC MWG staff understands and agrees with the need to keep spelling variations of XML Names (elements, attributes, and datatypes) to a minimum, the rules and guidance provided in the NDR draft are very impractical. No exact reference to a specific edition of the dictionary is given, nor is a URL provided for easy access. There is no book listed on Amazon with the exact title "Oxford English Dictionary American Spellings". Amazon searches for the "Oxford English Dictionary" (OED) results in a number of variants, with list prices shown:

- [Oxford English Dictionary, Second Ed.](#) (1989; 20 volumes, 22,000 pages, \$1,500; Addition Series volumes appear every few years for another \$65)
- [Compact OED](#) (1991; \$395)
- [Shorter OED](#) (2002; \$150)
- [Concise OED](#) (2004; \$30; 1700 pages)
- [Pocket OED](#) (2002; \$18)
- [New Oxford American Dictionary](#) (2nd book and PDA/CD edition; 2005; \$60; 2000 pages)

There are also related books published recently, such as *The Oxford American Desk Dictionary* (1998), *The Oxford American Dictionary of Current English* (1999), *The Oxford American Dictionary and Language Guide* (1999), and *The Oxford Dictionary of American Usage and Style* (2000).

Since there are so many OED versions, we must pick one specific one. It would be hard to justify the expensive editions to most program managers. It also seems unwise to select any edition that is more than a few years old, since new terms may be introduced, especially in the rapidly changing IT world.

IC CR for [DEN1] and [GNR1]: Given price and recentness considerations, the IC MWG staff recommends specifying either the *Concise Oxford English Dictionary* (2004 edition) or the *New Oxford American Dictionary* (2005 edition). Since the intent is to use American rather than British spelling, the IC MWG staff also recommends including either a non-normative list of common American vs. British spellings (e.g., labor, center, organization, etc.) or links to same, such as:

- <http://esl.about.com/library/weekly/aa110698.htm>
- <http://www.askoxford.com/betterwriting/us/?view=uk>
- <http://scientific.thomson.com/support/patents/dwpioref/reftools/usukdict/>
- <http://www2.gsu.edu/~wwwesl/egw/jones/spelling.htm>
- <http://www.scit.wlv.ac.uk/~jphb/american.html>

5.10 **Modularity Section Needs Concrete Examples**

Rules [SSM16 to SSM22]

The whole modularity discussion in Section 3.6.* may seem like the right idea on the intuitive level, but it is hard to see how it will work without a concrete, well-thought-out example. For example, using the old UBL/CCTS `Address` complexType, how would this be defined as a Federal object, how would it be imported, how could an Agency modify it at the Agency level and then how could someone else modify it at a lower level? Will the Federal level `Address` allow for military bases that aren't US states? Or is that up to DoD to do by extending the Federal `Address`?

5.11 **Permit Prepositions, Conjunctions and Articles in Element Names**

Rule [DEN12]: ‘The *Dictionary Entry Name* shall only contain verbs, nouns and adjectives (i.e. no words like *and*, *of*, *the*, etc.).’

While this rule seems correct in principle, it breaks down when one considers specifics, especially if there is a need to expand acronyms and abbreviations.

- `CostOfGoodsSold`
- `PointOfContact`
- `DepartmentOfTheNavy`
- `ProductAndServiceCode` (PSC code in federal procurement)
- `CommercialAndGovernmentEntityID` (CAGE code in federal procurement)

IC CR for [DEN12]: ‘The *Dictionary Entry Name* SHOULD contain nouns, verbs and adjectives. Prepositions, conjunctions and articles SHOULD be used sparingly (i.e., words like *of*, *and*, *the*, etc.).’

5.12 ***Voluntary Consensus Standards***

Rule [STR1]: “To ensure conformance with both statutory and policy requirements contained in Public Law 104-113 and Office of Management and Budget Circular A-119, all Federal XML implementations must adhere to the following hierarchy of standards in creating and using XML:

- De jure Voluntary Consensus Standards
- Cross-sector Voluntary Consensus Standards
- Sector specific Voluntary Consensus Standards
- Federal Enterprise Wide Standards”

IC CR for [STR1]: Consider changing the hierarchy of standards to the following and provide examples of each term:

- De facto Voluntary Consensus Standards (e.g., SAX)
- Horizontal Voluntary Consensus Standards (e.g., ebXML)
- Vertical Voluntary Consensus Standards (e.g., XSLT)
- Government-wide standards (e.g., TBD)
- CoI-specific standards (e.g., IC MSP, TWPDES)

5.13 ***Permit Specifications that are Not Quite Recommendations***

Rule [STA2]: “All schema and messages **MUST** be based on the W3C suite of technical specifications holding recommendation status.”

The current wording of this rule seems to prohibit the use of many other VCS standards (i.e., those from OASIS). For example, if taken literally, it would disallow using UBL or CCTS (even though they are based on W3C Recommendations). This exclusivity does not agree with the text appearing before and after the rule.

The IC MWG staff fully appreciates the importance of waiting for W3C/OASIS/IETF specifications to mature before they are widely adopted. We also understand the efforts necessary if one incorporates working draft technology into an application. On the other hand, there are many cases in which a VCS takes so long in reaching that final stage with relatively minor changes in the last few iterations that waiting is not always the best course of action. If an agency needs the technology for a specific mission critical application and determines that the specification, although not a fully mature “Recommendation,” is stable enough such that a subsequent retrofit will not be as costly as doing without that technology. XQuery is a case in point, as evidenced by the numbers of vendors who already have implemented this in their product lines (e.g., IBM, Oracle, Microsoft, DataDirect, etc.).

While interoperability is extremely important, there are times when implementing a not-yet-finalized specification does not significantly impact interoperability and yet achieves the pressing application-specific goals. For example, since XQuery returns results as XML, other interoperating applications won't be impacted by how that XML result was obtained even if they are in agency systems that don't support XQuery.

IC CR for [STA2]: “All schema and messages SHOULD be based on technical specifications holding recommendation status from any of the Voluntary Consensus Standards bodies identified (i.e., W3C, OASIS, IETF, etc.). However, agencies MAY use technical specifications that have not yet reached recommendation status if the technology is needed for mission critical applications or if the specification is judged to be reasonably mature and nearing recommendation status.”

6 How to Respond to This Position Paper

Tim West (DIA, Chair of the Intelligence Community Metadata Working Group) has requested a formal response from the Federal XML NDR Project Team. Please send this response to the IC MWG Secretariat:

Karen Stevens
(703) 874-8264
karen.h.stevens@saic.com
stevnsk@cia.ic.gov