



[Editorial Draft] Extending and Versioning Languages: Terminology With comments from Noah

Draft TAG Finding 18 May 2007

This version:

<http://www.w3.org/2001/tag/doc/versioning-20070518.html> ([xml](#))

Latest version:

<http://www.w3.org/2001/tag/doc/versioning>

Previous versions:

Unapproved Editors Drafts: <http://www.w3.org/2001/tag/doc/versioning-20070326.html>, <http://www.w3.org/2001/tag/doc/versioning-20061212.html>, <http://www.w3.org/2001/tag/doc/versioning-20060726.html>, <http://www.w3.org/2001/tag/doc/versioning-20060717.html>, <http://www.w3.org/2001/tag/doc/versioning-20060710.html>, <http://www.w3.org/2001/tag/doc/versioning-20031116.html><http://www.w3.org/2001/tag/doc/versioning-20031003.html>

Editor:

David Orchard, BEA Systems, Inc. <mailto:David.Orchard@BEA.com>

Copyright © 2003 W3C® ([MIT](#), [INRIA](#), [Keio](#)), All Rights Reserved. W3C [liability](#), [trademark](#), [document use](#), and [software licensing](#) rules apply.

Formatted: French (France)

Field Code Changed

Formatted: French (France)

Formatted: French (France)

Abstract

This document provides terminology for discussing language versioning. Separate documents contains versioning strategies and XML language specific discussion.

Status of this Document

This version includes comments from Noah on the first few pages. I'm circulating this now because it may be several weeks before I get to transcribe more of my comments.

Comments with more asterisks (***) indicate particularly important points or global issues. Fewer asterisks (*) or none indicate correspondingly less important points.

This document has been developed for discussion by the [W3C Technical Architecture Group](#). It does not yet represent the consensus opinion of the TAG.

Publication of this finding does not imply endorsement by the W3C Membership. This is a draft document and may be updated, replaced or obsoleted by other documents at any time.

[Additional TAG findings](#), both approved and in draft state, may also be available. The TAG expects to incorporate this and other findings into a Web Architecture Document that will be published according to the process of the [W3C Recommendation Track](#).

Please send comments on this finding to the publicly archived TAG mailing list www-tag@w3.org ([archive](#)).

Table of Contents

- 1 [Introduction](#)
 - 1.1 [Terminology](#)
 - 1.1.1 [Compatibility](#)
 - 1.1.1.1 [Composition](#)
 - 1.1.2 [Partial Understanding](#)
 - 1.1.3 [Divergent Understanding and Compatibility](#)
 - 1.1.4 [Open or Closed systems](#)
 - 1.1.5 [Compatibility of languages vs compatibility of applications](#)
- 2 [Conclusion](#)
- 3 [References](#)
- 4 [Acknowledgements](#)

Appendix

- A [Change Log](#) (Non-Normative)

1 Introduction

The evolution of languages by adding, deleting, or changing syntax or semantics is called versioning. Making versioning work in practice is one of the most difficult

Formatted: French (France)

Formatted: French (France)

Field Code Changed

Formatted: French (France)

Comment: General comments

●**Not all languages are at the full document level. We should be clear that the finding applies to any language consisting of sets of texts, whether whole document, subtree in XML, just the text content of some tag (e.g. the format of a floating point number), or plain text files.

●Do you want to reference RFC 2119?

●* I think you need to define "TEXT" before this. There is a formal definition, but it comes after this first use

●**The relationship between information sets and semantics seems to be unclear, yet this one or both of them is crucial to the story you're telling about versioning.

problems in computing. Arguably, the Web rose dramatically in popularity because HTML and HTTP provide effective support for extensibility and versioning. Both systems provide explicit extensibility points and rules for understanding extensions that enable their decentralized extension and versioning.

Deleted: evolution and versioning were built into

This finding describes terminology of languages and their versioning.

1.1 Terminology

Suggested terminology for describing languages, producers, consumers, information, constraints, syntax, evolvability etc. follows. Let us consider an example. Two or more systems need to exchange information about peoples' names. Names may not be the perfect choice of example because of internationalization reasons, but it resonates strongly with a very large audience. The Name Language is created to be exchanged. [Definition: A **producer** is an agent that creates text.] Continuing our example, Fred is a producer of Name Language text. [Definition: An **Act of Production** is the creation of text.]. A producer produces text for the intent of conveying information. When Fred does the actual creation of the text, that is an act of production. [Definition: A **consumer** is an agent that consumes text.] We will use Barney and Wilma as consumers of text. [Definition: An **Act of Consumption** is the processing of text of a language.] Wilma and Barney consume the text separately from each other, each of these being a consumption event. A consumer is impacted by the instance that it consumes. That is, it interprets that instance and bases future processing, in part, on the information that it believes was present in that instance. Text can be consumed many times, by many consumers, and have many different impacts.

Deleted: The

Deleted: name

Comment: I still don't think it's the best example, but I think we've agreed to disagree on that. So, I'll assume it stays.

Comment: *No. The language is not exchanged. Maybe: "Documents conformant with the name language are intended for exchange between computer applications."

Comment: Define TEXT first.

Comment: What do we mean by "act of creation"? Encoding as a string of bits? Deciding what the content is going to be?

Comment: Sounds too much like it relates to tuberculosis.

Comment: Antecedent of "these" is ambiguous. Furthermore, the sentence parses to suggest that something is an event, but the nouns and noun phrases earlier in the sentence are Wilma, Barney and "each other", none of which could be an event.

Comment: ***No, no, no. We keep disagreeing on this, and others (Stuart?) have disagreed as well. I strongly believe we'll do better to say: The language is a set of texts and the mapping of texts to information. Once you know the texts, the constraints are redundant. Same for information resulting from the mappings, and constraints that those results happen to obey.

I would buy saying: "one way to specify the set of texts that comprise a language is intensionally, using a constraint language such as regular expressions, W3C XML schemas, etc. In other cases, a language may be defined extensionally, by just listing the texts that are legal (this is often practical with very small languages, such as "red" "yellow" "green", which might be legal texts for a language capturing the state of a traffic light).

Comment: Definition of TEXT is needed earlier.

Deleted: Given that there are constraints on a language, a

Comment: Why have we switched from text to string of chars?

Deleted: may

Deleted: the

Deleted: Documents are texts of a language.

[Definition: A **Language** consists of a set of text, any syntactic constraints on the text, a set of information, any semantic constraints on the information, and the mapping between texts and information.][Definition: **Text** is a specific, discrete sequence of characters]. Any particular text may or may not have membership in a language. Indeed, a particular string of characters may be a member of many languages, and there typically will be many different strings of characters that are members of a given language. The texts of a language are the units of exchange between a producer and consumer. [Definition: When a text is the outermost unit of exchange, we call it a **document**] (documents, in turn may employ use smaller languages internally: so, for example, a document language might use a number language to represent integer values as strings of digits).

The Name Language consists of text set that have 3 terms and specifies syntactic constraints: that a name consists of a given and a family. [Definition: A language has a set of **constraints** that apply to the set of strings in the language.] These constraints can be defined in machine processable syntactic constraint languages such as XML Schema, microformats, human readable

textual descriptions such as HTML descriptions, or are embodied in software. Languages may or may not be defined by a schema in any particular schema language. The constraints on a language determine the strings that qualify for membership in the language. Vocabulary terms contribute to the set of strings, but they are not the only source of characters to the set of strings in a given language. The language strings may include characters outside of terms, such as punctuation. One reason for additional characters is to distinguish or separate terms, such as whitespace and markup.

Example 1: Name examples.

```
<name>
  <given>Dave</given>
  <family>Orchard</family>
</name>

name="Dave Orchard"

<span class="fn">Dave Orchard</span>

urn:namespace:given:Dave:family:Orchard
```

The set of information in a language almost always has semantics. In the Name Language, given and family have the semantics of given and family names of people. The language also has the binding from the items in the information set to the text set. Any potential act of interpretation, that is any consumption or production, conveys information from text according to the language's binding. The language is designed for acts of interpretation, that being the purpose of languages. In our example, this mapping is obvious and trivial, but many languages it is not. Two languages may have the exact same strings but different meanings for them. In general, the intended meaning of a vocabulary term is scoped by the language in which the term is found. However, there is some expectation that terms drawn from a given vocabulary will have a consistent meaning across all languages in which they are used. Confusion often arises when terms have inconsistent meaning across language. The Name terms might be used in other languages, but it is generally expected that they will still be "the same" in some meaningful sense.

*****TRANSCRIBED COMMENTS END HERE FOR NOW*****

These terms and their relationships are shown below

Comment: ***Whoa! You're saying "information has semantics"? Either that's wrong, or it needs more careful explanation. I think I know what you're trying to say, but first of all it's unclear, and secondly I'm not sure we want to get into semantics. If we can just formulate an explanation of which documents convey the same information (the last name value is Mendelsohn) I think that's enough to tell the versioning story. Going into semantics (in Western societies, one part of the name is traditionally taken from the father, so saying that the last name field is Mendelsohn suggests that I actually have such a family name) we should most avoid, I think. Semantics is important, but I think that if we show how to convey information reliably, others can build semantics and reasoning on top of that.

Comment: ***What's an Information Set? It seems to be a key abstraction, but you're using it without introducing it.

Comment: First of all, this seems a bit smug in tone. Secondly, it's not entirely correct. Languages are intended, I think, as a means by which information can be set down or encoded, as well as to support interpretation. I'd kill the whole sentence.