

建設分野における XML 記述仕様の考え方(案)

平成 21 年 3 月

国土交通省 国土技術政策総合研究所

はじめに

情報化の進展に伴い、建設分野においても様々な情報システムが構築され、利用されている。しかしながら、これまでは個別の情報システムごとで閉じられており、他の情報システムとの情報連携にはあまり配慮されてこなかった。このため、他のシステムではデータが活用できず、データの二重登録、二重管理等の無駄が発生している。

CALS/EC では、組織間、システム間の情報連携を可能とするデータの統合化技術の開発が目標として掲げられている。その1つの手段として、特定のシステムに依存しないデータ形式である XML を用いたデータ交換が考えられ、採用も増えている。

XML は、文書やデータの意味や構造を記述するためのマークアップ言語の1つであり、文書やデータの交換において有効なデータ形式と考えられる。しかし、XML は、利用する「タグ」がユーザ独自に指定できることから、同じ文書やデータでもタグが異なると、違う文書やデータと判断されてしまう。このため、タグの名称がシステム間のデータ交換にとって重要である。

本書は、システム間のデータ交換に XML を利用する際、データの交換がスムーズに実施できるように XML 記述仕様の考え方をまとめたものである。本書で示した標準的な考え方により XML の記述が行われると、他のシステムの XML 仕様が参考しやすくなり、タグ名称の統一が進んでいくと考えられる。

本書の策定にあたっては、国土交通省国土技術政策総合研究所が（財）日本建設情報総合センターに業務を委託し、社会基盤情報標準化委員会 情報連携基盤小委員会（柴崎亮介委員長） 建設 XML 検討 WG（田中成典座長）にご検討・ご審議をいただいた。多大なご協力をいただいた関係各位に深く謝意を表する次第である。

平成 21 年 3 月

国土交通省 国土技術政策総合研究所

目次

1 本書の概要・位置づけ.....	1
1.1 目的	1
1.2 用語の解説.....	2
1.3 本書の対象読者	5
1.4 本書の適用範囲	6
1.5 タグ作成に関する規定.....	6
1.6 引用・参照した規格等.....	7
1.6.1 関連する標準類.....	7
1.6.2 日本国内の関連事例.....	7
2 XML 記述に係る留意事項.....	8
2.1 スキーマ言語	8
2.2 文字コード.....	9
2.3 使用文字	10
2.4 大文字小文字の使い分け	12
2.5 単数形・複数形の使用.....	14
2.6 略語の使用.....	15
3 開発した XML 記述仕様の公開方法について.....	16
3.1 XML 記述仕様の公開.....	16

1 本書の概要・位置づけ

1.1 目的

現在、我が国の建設分野では、情報の共有・連携による生産性向上等の観点から、データベース仕様やデータ交換仕様におけるデータ記述方式について、OS等のプラットフォームに依存しないオープンな技術標準としてXMLを採用する動きが活発化しつつある。しかし、XMLは自由度の高い言語であるため、データベース仕様やデータ交換仕様ごとに異なった考え方に基づくタグ名称等が採用されることで、例えば同じ意味を持つデータであるのに利用することができない、もしくは違う意味のデータを誤って同一のデータとして取り扱ってしまうなど、逆に相互運用性を阻害してしまうことが懸念されている。

事実、既にXMLで開発されているデータベースシステムやデータ交換標準は、現状建設分野に何らかの統一的なXMLスキーマ言語に基づく仕様(以下、「XML記述仕様」という。)の開発指針がないため、いずれも独自に開発され、固有のタグ等が使用されている。

「建設分野におけるXML記述仕様の考え方(案)」は、建設分野で利用するXML記述仕様を開発する場合に最低限守っておくことが望ましい、タグ名称やこれらを組み立てて構築する構造定義に係る共通的な開発方法について解説したものである。

また、本書を参考にすることで、システム間でデータを交換するために必要な“データ辞書”の作成や閲覧が容易になるとともに、XML記述仕様の開発手順の理解を促進することも目指している。

加えて、今後は建設分野において共通的に使用することが望ましいタグ名称を収集し、逐次追加していく予定である。

1.2 用語の解説

(1) XML

XML(Extensible Markup Language)とは、W3C(World Wide Web Consortium、Web で使用される各種技術の標準化を推進する為に設立された非営利団体)が策定した、Web 上における国際標準の構造化文書の技術である。XML の最も重要な目的は、異なる情報システムの間で、特にインターネットを介して、構造化された文書や構造化されたデータの共有を容易にすることである。

(2) 名前空間

名前空間は、複数の文書内によく似た構造名が使われて名前が競合することを避けることができる。XML では要素名を自由に決められるので、同じ名前を持つ要素や属性でも、その意味や働きが異なる場合がある。要素名や属性名に重複があっても、名前空間を用いてそれぞれの属する空間を明示することによって、意味や内容を区別できる。

(3) W3C

W3C (World Wide Web Consortium) は、World Wide Web で使用される各種技術の標準化を推進する為に設立された非営利団体。マサチューセッツ工科大学(MIT)や欧州合同素粒子原子核研究機構(CERN) が中心となって 1994 年 10 月 1 日発足。2007 年現在、マサチューセッツ工科大学計算機科学研究所(MIT/LCS)、欧州情報処理数学研究コンソーシアム(ERCIM)、慶應義塾大学が中心となって活動している。

(4) XML Schema

XML Schema は、XML 関連技術の標準を策定している W3C (World Wide Web Consortium)が、XML1.0を拡張し標準化したスキーマ言語(XML Schema Part0,1,2 で定義)である。DTD より正確にデータの意味を定義できる。

(5) OASIS UBL

OASIS UBL(Universal Business Language)は、OASIS (Organization for the Advancement of Structured Information Standards、構造化情報標準促進協会)の UBL TC で標準化が行われている、企業間電子商取引のために定義された XML 形式の電子伝票の仕様である。

(6) UN/CEFACT

UN/CEFACT (United Nations Centre for Trade Facilitation and Electronic Business : 貿易簡易化と電子ビジネスのための国連センター) は、コンピュータ技

術による貿易の簡素化を目的とした機関であり、国連欧州経済委員会(UNECE)の中に設置されている。UN/CEFACT はこれまでに、企業間電子商取引の共通規格 ebXML(OASIS と共同)、電子データ交換標準の UN/EDIFACT などを策定し、国際的管理組織として世界規模で活動している。

(7) NDR

NDR (Naming and Design Rule : 命名設計規則) は、特定分野において、XML 文書を作成する際に準拠すべき規則を定めたものである。本書は、日本の建設分野において NDR に準じた役割を担うことを想定している。

(8) ISO/IEC11179

ISO/IEC11179 はデータの最小単位 (データ要素) の仕様を規定しており、情報に係る標準化作業は、全てこれに準拠して行う必要がある。メタデータを含めてデータ要素の構成の基本的な側面を規定する。これは、人と機械との間で共有するようなデータ要素の「表現形式及び意味」の定式化に適用し、機械水準でのビット及びバイトのようなデータの物理的な表現形式には適用しない。

(9) JIS X 4159

JIS X 4159 は、2000 年 6 月に 2nd Edition として W3C から公表された勧告 Extensible Markup Language (XML) 1.0 を翻訳し、技術的内容を変更することなく作成した日本工業規格である。

(10) ebXML

ebXML (Electronic Business using eXtensible Markup Language) は、UN/CEFACT と OASIS が共同で開発・保守を行なっている企業間電子商取引のための規格群で、情報項目モデリング手法から電文搬送手順までをカバーしており、その内コア構成要素技術仕様・レジストリ情報モデル及びサービス・電文搬送手順及び運用手順定義方式などが ISO 技術仕様 (ISO TS 15000 シリーズ) として公開されている。

(11) XML 日本語プロファイル

XML 日本語プロファイルは、日本規格協会情報技術標準化研究センター (INSTAC) に設立された高速 Web 環境における標準化に関する調査研究委員会でもまとめられた技術資料である。

XML のアプリケーションでは、Shift JIS などで入力された文字を ISO/IEC-10646 (Unicode) に変換する作業が行われるが、違うコードに変換されてしまうケースが

存在し、安全な情報交換はできないことになることがある。そこで、XML 日本語プロファイルでは、この問題を詳しく調査して、具体的な変換テーブルの相違と変換において発生する問題点を明らかにしている。

(12) DTD

DTD (Document Type Definition) は、SGML (Standard Generalized Markup Language) や XML において、文書構造 (文書型) 定義するためのスキーマ言語の一つである。文書内に記述することができる要素やその発生順序、発生回数、要素がもつ属性、属性の型などを記述することができる。

(13) 文字コード

コンピュータ上で文字を利用するために各文字に割り当てられるバイト表現もしくは、バイト表現と文字の対応関係。符号化文字集合 (coded character set、CCS) と文字符号化方式 (character encoding scheme、CES) という区別がなされることがある。

(14) メタデータ

データについてのデータ。あるデータそのものではなく、そのデータに関連する情報のこと。データの作成日時や作成者、データ形式、タイトル、注釈などが挙げられる。データを効率的に管理したり検索したりするために重要な情報である。文書や画像などを保存するファイル形式の多くは、ファイルの先頭にメタデータを格納し、続いてデータ本体を格納するようにできている。

1.3 本書の対象読者

【要点】

本書は、以下の読者を対象としている。

- XML 記述仕様を今後開発する、またはすでに開発している組織
(発注機関、業界団体等仕様開発組織)
- XML 記述仕様に基づいた共有データの作成・利用(システム開発)を支援する組織
(ITベンダ、コンサルタント等のシステム開発組織)
- 共有データを利用して業務を実施する組織
(発注機関、業界団体等のデータ利用組織)

【解説】

本書は、XML 記述仕様を開発する組織と、本書に則って作成された XML 記述仕様に基づいたデータを作成・利用する組織(システム開発者及びデータ利用者)を読者対象として作成している。

また、本書は XML 記述仕様作成の基本的な部分を示したものであるため、詳細の規定については「1.6 引用・参照した規格等」を参照すること。

1.4 本書の適用範囲

【要点】

本書は、建設分野において新規に作成される XML 記述仕様を対象とする。また、既存の XML 記述仕様を更新する場合については、本書を参考にする。

【解説】

本書は、建設分野における XML 記述仕様全般を対象としている。

今後作成される XML 記述仕様については、本書に示された作成の考え方に沿うことで、互換性の高い仕様を作成することが可能になる。

一方、既に作成された XML 記述仕様類については、緩やかな標準化を視野に入れて、システム更新期などのしかるべき時期に本書に沿った見直しを行うことなどで対応することとする。

1.5 タグ作成に関する規定

【要点】

基本的に自由だが、データ再利用性向上のためには先行事例を参照することが望ましい。

【解説】

タグの策定に関しては、基本的に自由とする。しかし、仕様間の互換性を高めるためには、過去に策定された XML 記述仕様を再利用し、先行事例を参照することが望ましい。

また、スキーマ言語に XML Schema を用いた場合は、名前空間の利用が可能であるため、複数のスキーマ仕様を混在して定義することができる。

1.6 引用・参照した規格等

本書で引用・参照した規格・文献は、次のとおりである。

1.6.1 関連する標準類

- [1] Extensible Markup Language (XML) 1.0 (Fifth Edition)
(W3C Recommendation 05 February 2008 : JIS X 4159)
- [2] XML Schema Part 0: Primer Second Edition
(W3C Recommendation 28 October 2004)
- [3] XML Schema Part 1: Structures (W3C Recommendation 2 May 2001)
- [4] XML Schema Part 2: Datatypes (W3C Recommendation 2 May 2001)
- [5] OASIS UBL(Universal Business Language) NDR CD, 1.0.1 15 November 2004
- [6] UN/CEFACT XML NDR Draft1.2, 8 September 2005
- [7] ISO/IEC11179 Metadata Registries
(Part 1 ~ 3) は JIS X 4181 (翻訳 JIS)
- [8] 日本工業規格 JIS X 4159:2005 拡張可能なマーク付け言語(XML)1.0

1.6.2 日本国内の関連事例

- [1] ECOM の報告書 : ebXML を用いた共有データ作成ガイドブック(平成 17 年 3 月)
- [2] XML 日本語プロファイル
- [3] 電子申請用 XML 様式的设计ガイドライン第 2 版

2 XML 記述に係る留意事項

2.1 スキーマ言語

【要点】

スキーマ言語は、XML Schema を採用することが望ましい。

【解説】

XML 記述仕様で構造を定義する言語である DTD と XML Schema について、それぞれの特徴と利用場面は以下のとおりである。

(1) DTD (Document Type Definition)

DTD は、XML の前身である SGML より基本的な構文を受け継いだ、文書構造を定義するためのスキーマ言語である。DTD では、要素やその発生順序、発生回数、要素が持つ属性、属性の型などを記述することができる。DTD 方式の特徴は以下のとおりである。

1) 特徴

- 利用され始めた時期が早く、ツールやノウハウなどの環境が整っている。
- 既存の利用できる DTD が多数存在する。
- データ交換には、機能や定義内容が不十分である。(型や名前空間の定義が不足)

2) 利用例

電子納品データ管理ファイルの記述形式、土木積算基準データ、電子入札情報システム 等

(2) XML Schema

XML Schema は、XML 関連技術の標準を策定している W3C が、XML1.0 を拡張し標準化したスキーマ言語 (XML Schema Part0,1,2 で定義) である。DTD より正確にデータの意味を定義できる。XML Schema の特徴は以下のとおりである。

1) 特徴

- DTD の持つ問題点に対応している。
- 名前空間を前提に設計されている。
- 仕様が複雑で解りにくい面がある。

2) 利用例

土木工事帳票様式共通タグ (案)、道路基盤データ交換属性セット (案)、道路中心線形データ交換標準 (案)、ボーリングデータ G-XML 等

以上より、XML Schema は DTD に比べて、より正確なデータの記述が可能となるため、データの再利用性が向上している。また、近年では、建設分野においても XML Schema が多く使われるようになってきている。

2.2 文字コード

【要点】

XML 記述仕様に用いる文字コードは、UTF-8 または UTF-16 とすることが望ましい。

【解説】

文字コードは、JIS X 4159 (XML1.0) 及び XML 日本語プロファイルの関連事例を参考として規定した。

JIS X 4159 (XML1.0) では、UTF-8 及び UTF-16 の処理をすべての XML プロセッサに義務づけている。XML 日本語プロファイルでも、UTF-8 または UTF-16 の利用を推奨している。

(関連事例での規約内容)

- JIS X 4159 (XML1.0)

2.2 文字

文字符号位置をビットパターンに符号化する機構は、実体ごとに違ってよい。すべての XML プロセッサは、Unicode 3.1 の UTF-8 符号化及び UTF-16 符号化を受け付けなければならない。二つのどちらが用いられているかを明示するため又は他の符号化を利用するための機構は、4.3.3 (実体における文字符号化) に記述する。

4.3.3 実体における文字符号化

XML 文書内の外部解析対象実体は、それぞれ別の文字符号化を用いてもよい。すべての XML プロセッサは、UTF-8 で符号化した実体、及び UTF-16 で符号化した実体を処理できなければならない。この規定での"UTF-8"及び"UTF-16"という用語は、他のラベルをもつ符号化には適用されない。その符号化がどれほど "UTF-8"又は"UTF-16"に類似していても、適用されない。

- XML 日本語プロファイル

- UTF-8 あるいは 16 (Unicode) の利用を推奨する。
- シフト JIS、日本語 EUC(圧縮形式)、ISO-2022-JP を推奨しない。

2.3 使用文字

【要点】

要素、属性、型名称に使用する文字は、半角英字（アルファベット[A-Z] [a-z]、ハイフン[-]、アンダースコア[_]、スペース[]）を用いることが望ましい。

【解説】

XML の記述は、利用者の目に触れる実データ（インスタンス部分）と、XML 文書の構造を記述するスキーマ部分（要素、属性、型名称）に分けることができる。このうち、実データは利用者にとって理解しやすい日本語で表記するのが普通であるが、スキーマ部分に関しては直接利用者の目に触れる部分ではないことから、使用文字については国際標準に準拠するなど、切り分けて考える必要がある。このため、要素、属性、型名称に使用する文字は、以下の関連事例を参考に、半角英字を用いることが望ましいと規定した。

ISO/IEC11179 では、スキーマ言語部分には半角英字を用いることとしている。OASIS UBL NDR、UN/CEFACT XML NDR 及び ebXML を用いた共有データ作成ガイドブックがあるが、これらでも英字を用いることになっている。

使用が望ましい文字の JIS X 0201 文字コード（16 進数）は、20（スペース）、2D（ハイフン）、41～5A（A～Z）、5F（アンダースコア）、61～7A（a～z）である。

使用が望ましい文字には、二重引用符["]、引用符[']、コンマ[,]、ピリオド[.]、スラッシュ[/]、コロンの[:]、セミコロン[;]、小なり[<]、大なり[>]などは含まない。

例えば、ローマ数字（Ⅰ、Ⅱ、Ⅲ、Ⅳ、Ⅴ、Ⅵ、Ⅶ、Ⅷ、Ⅸ、Ⅹ、Ⅺ、Ⅻ など）、ギリシャ文字（α、β、γ、δ、ε、ζ、η、θ、ι、κ、λ、μ、ν、ξ、ο、π、ρ、σ、τ、υ、φ、χ、ψ、ω など）を表現したい場合は、one、five、ten、alpha、beta、gamma など英語表記を用いる。

また、順序を表現したい場合は、数字ではなく one、two などの英語表記を用いることが望ましい。

（関連事例での規約内容）

- ISO/IEC11179

・アルファベット（A-Z, a-z）及びハイフン（-）、アンダースコア（_）、スペース（ ）のみで構成する。

- OASIS UBL NDR

オックスフォード英語必須（ISO/IEC11179 準拠）

・要素、属性、型名称はオックスフォード辞書スペルの英語でなければならない。

- UN/CEFACT XML NDR

オックスフォード英語必須 (ISO/IEC11179 準拠)

- 要素、属性、型名称はオックスフォード辞書スペルの英語でなければならない。

- ebXML を用いた共有データ作成ガイドブック

規則 10 言語固有の規則がない限り，名前には英大文字 (A-Z) 及び英小文字 (a-z) 以外は使ってはならない。

2.4 大文字小文字の使い分け

【要点】

大文字と小文字を使い分ける場合については、ISO/IEC11179 に準拠することが望ましい。

【解説】

同一の文字であっても、大文字で表記した場合と小文字で表記した場合、ツールやシステムでは別のタグとして認識されてしまうため、システム間の連携が困難になる。このため、大文字小文字の使い分けは、関連する国際標準（ISO/IEC 11179）に準拠することが望ましいと規定した。

（ISO/IEC 11179 の規約内容）

要素名（エレメント名） 半角英字、先頭の英語は大文字とすること 属性名（アトリビュート名） 半角英字、先頭の英字は小文字とすること
--

また、関連事例として、OASIS UBL NDR、UN/CEFACT XML NDR 及び ebXML を用いた共有データ作成ガイドブックがあるが、これらでも ISO/IEC 11179 に準拠した規約内容になっている。

（関連事例での規約内容）

- OASIS UBL NDR

要素及び型名は先頭大文字、属性名は先頭小文字でなければならない。 （ISO/IEC11179 準拠）

- UN/CEFACT XML NDR

要素及び型名は先頭大文字、属性名は先頭小文字でなければならない。 （ISO/IEC11179 準拠）

- ebXML を用いた共有データ作成ガイドブック

規則 7 アトリビュート (xsd:attribute)の命名には、先頭小文字キャメルケース(LCC: Lower Camel Case) を使わなければならない。 例：“codeListName Text” 規則 8 エレメント(xsd: element)と型の命名は先頭大文字キャメルケース(UCC: Upper Camel Case) を使わなければならない。 例：“ByerPartyOrganization”
--

規則 13 頭字語および省略形がアトリビュート宣言の先頭に現れるときは小文字で書かなければならない。その他の頭字語および省略形は大文字でなければならぬ。

規則 14 すべてのエレメント宣言および型定義中で、頭字語は大文字で書かなければならぬ。

キャメルケースとは、複合語をひと綴りで表現するために、要素語の最初を大文字で書き表すことをいう。

2.5 単数形・複数形の使用

【要点】

単数形・複数形を使用する場合には、ISO/IEC11179 に準拠することが望ましい。

【解説】

同一の単語であっても、ツールやシステムでは単数形と複数形を別のタグとして認識してしまうため、システム間の連携が困難になる。このため、単数形・複数形の使用は、関連する国際標準（ISO/IEC 11179）に準拠することが望ましいと規定した。

（ISO/IEC 11179 の規約内容）

本来複数形のもの以外は、単数形でなければならない。

また、関連事例として、OASIS UBL NDR、UN/CEFACT XML NDR 及び ebXML を用いた共有データ作成ガイドブックがあるが、これらでも ISO/IEC 11179 に準拠した規約内容になっている。

（関連事例での規約内容）

- OASIS UBL NDR

本来複数形のもの以外は、単数形でなければならない。（ISO/IEC11179 準拠）

- UN/CEFACT XML NDR

本来複数形のもの以外は、単数形でなければならない。（ISO/IEC11179 準拠）

- ebXML を用いた共有データ作成ガイドブック

規則 9 元来複数形のもの以外，エレメント及びアトリビュートの名前はすべて単数形で表す。

2.6 略語の使用

【要点】

略語を使用する場合には、ISO/IEC11179 に準拠することが望ましい。

【解説】

何の規定もなく略語を開発者が自由に作成すると、それぞれの観点で作成された略語が乱立し、同内容のタグ名称であっても別のタグであると認識されてしまうため、システム間の連携が困難になる。このため、略語の使用は、関連する国際標準 (ISO/IEC 11179) に準拠することが望ましいと規定した。

(ISO/IEC 11179 の規約内容)

略語は、共通に理解できるものだけを使用すること。

頭字語(Acronym)や頭文字(Initialism)を含む、略語(Abbreviation)の意味の理解は、通常は特定の環境に限定される。異なる環境では、同じ略語が、誤解を招く可能性がある。それゆえ、あいまいさをさけるため、定義中には完全な言葉で (略語ではなく) 記述されるべきである。ただし、"i.e.", "e.g."といった一般的に理解できるような略語や、複雑な単語で構成された完全形よりも略語のほうがわかりやすく、しかも世の中に受け入れられているもの、たとえば"radar" (radio detecting and ranging の短縮形)などは、この規則の例外である。すべての略語は、最初に出現したときに展開されなければならない。

また、関連事例として、OASIS UBL NDR、UN/CEFACT XML NDR があるが、これらでも ISO/IEC 11179 に準拠した規約内容になっている。

(関連事例での規約内容)

- OASIS UBL NDR

“ DUNS ” “ ID ” “ URI ” 以外の略語を使用してはならない。(ISO/IEC11179 準拠)

- UN/CEFACT XML NDR

“ DUNS ” “ ID ” “ URI ” 以外の略語を使用してはならない。(ISO/IEC11179 準拠)

3 開発した XML 記述仕様の公開方法について

3.1 XML 記述仕様の公開

【要点】

本書を参考に XML 記述仕様を開発した場合については、可能な限り公表することが望ましい。

【解説】

開発した XML 記述仕様はインターネット上に公開することを想定している。XML 記述仕様を公開することで、XML 記述仕様の開発者が先行事例を参照しやすくなり、タグ名称の引用や、データ辞書の作成が容易になると考えられる。

国内の建設分野の XML 記述仕様を統一的に登録するサイトが構築された場合、そこに登録するが、公開サイトが未整備なために、XML 記述仕様を開発した場合、当面は国土交通省国土技術政策総合研究所高度情報化研究センター情報基盤研究室に電子メール（jyouhou@nilim.go.jp）で連絡するものとする。