

データ標準化の取組について

平成17年5月24日

総務省自治行政局

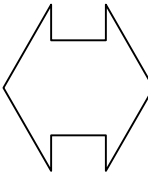
データ標準化の推進体制

電子自治体のシステム構築のあり方に関する検討会

学識経験者 : 須藤修(東京大学) 三好勝則(香川大学)
島田達巳(摂南大学) 大場充(広島市立大学)
南波幸雄(関東学院大学)
民間有識者 : 松尾明(公認会計士)
手島歩三(ビジネス情報システム・アーキテクト)
小町祐史(パナソニックコミュニケーションズ)
山口秀二(情報政策研究所)
地方公共団体 : 山谷吉宏(北海道) 溝江言彦(福岡県) 井堀幹夫(市川市)
吉田稔(西宮市)
総務省 : 牧慎太郎(自治行政局自治政策課)

データ標準化推進地方公共団体協議会

(メンバー)
全国知事会、全国市長会、全国町村会の代表団体
事務局: 総務省



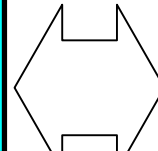
連携・調整

連携・調整

データ標準化WG

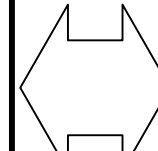
(メンバー)
学識経験者 : 松尾明(中央青山監査法人) <WG座長>
村田真(国際大学)
小町祐史(パナソニックコミュニケーションズ)
山口秀二(情報政策研究所)
平野洋一郎(インフォテリア)
地方公共団体 : 県 北海道、静岡県、高知県
市 市川市、横須賀市、西宮市
総務省 : 自治行政局、情報通信政策局
行政管理局、統計局

共同アウトソーシング評価委員会



連携・調整

国
地方
連携システム
のあり方
検討
WG



連携・調整

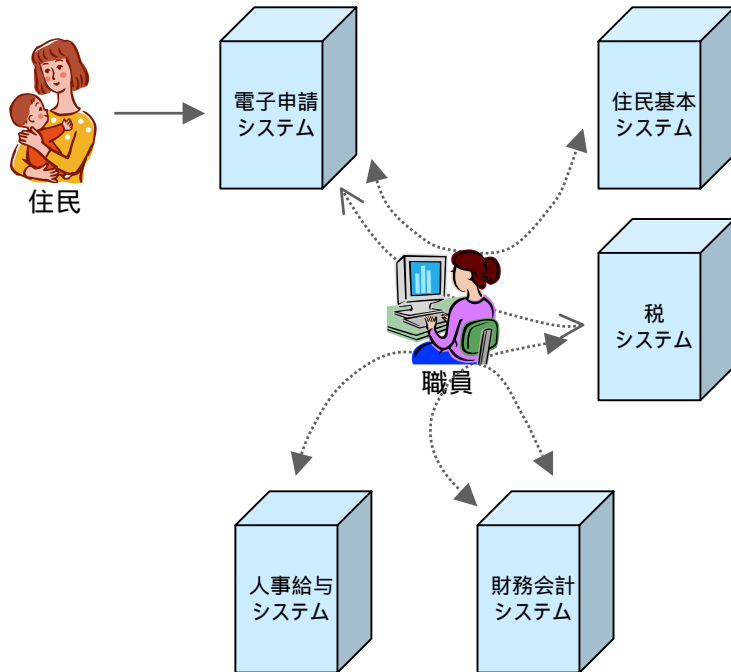
電子行政推進国・地方 公共団体協議会

霞が関WAN・
LGWAN連携部会

目指すべき電子自治体に向けた取組み(1)

業務の効率化

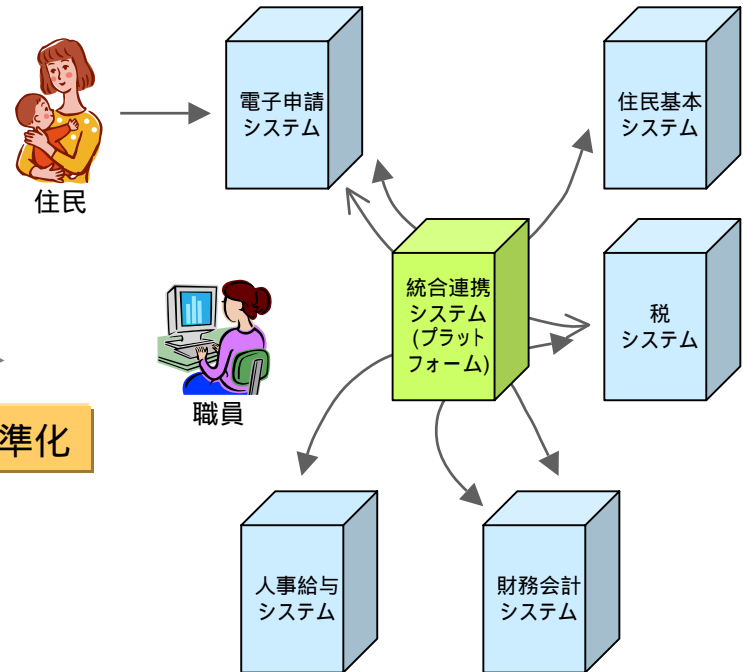
現状



各業務システムは個別に導入されているために、業務システム間の連携は職員が再入力したり、データの整形を行ったりして行う必要がある。
職員の業務手順を業務システムの機能に合わせる必要があり、効率的な業務手順とはなっていない。

将来

データの標準化

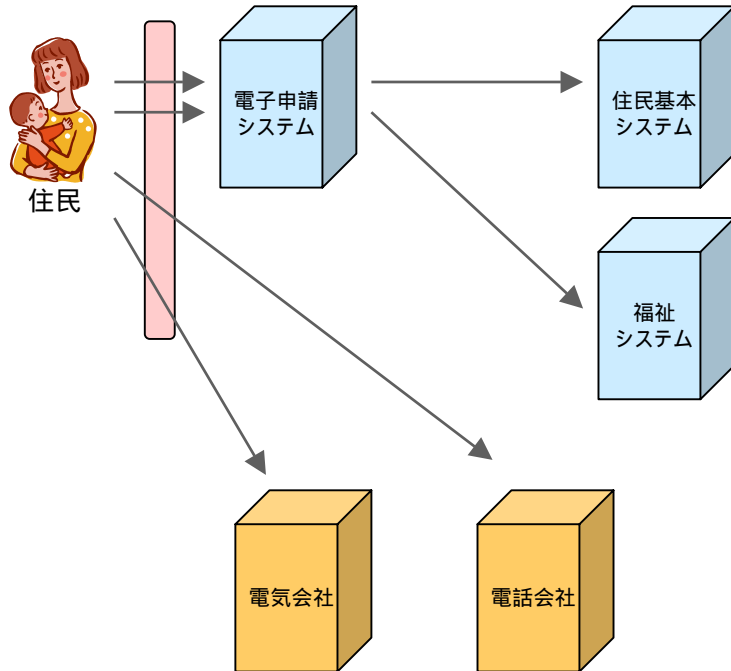


各業務システムは標準化されたデータを使って自動的に連携できるようになり、職員は本来の業務に時間を使うことができるようになる。
職員は業務システムにまたがった業務でも効率的な手順で行えるようになる。

目指すべき電子自治体に向けた取組み(2)

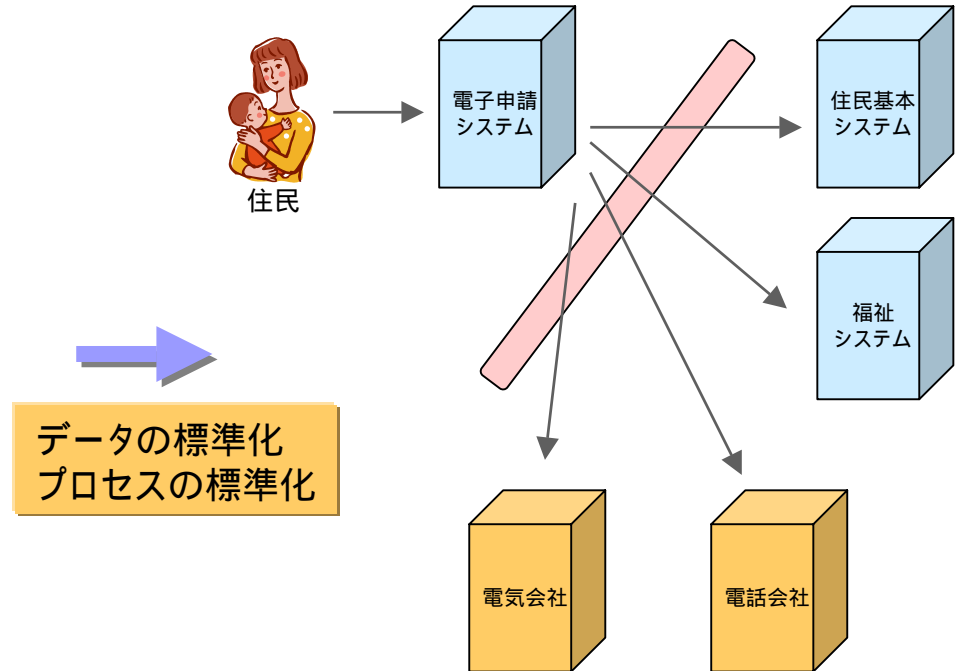
住民の利便性の向上

現状



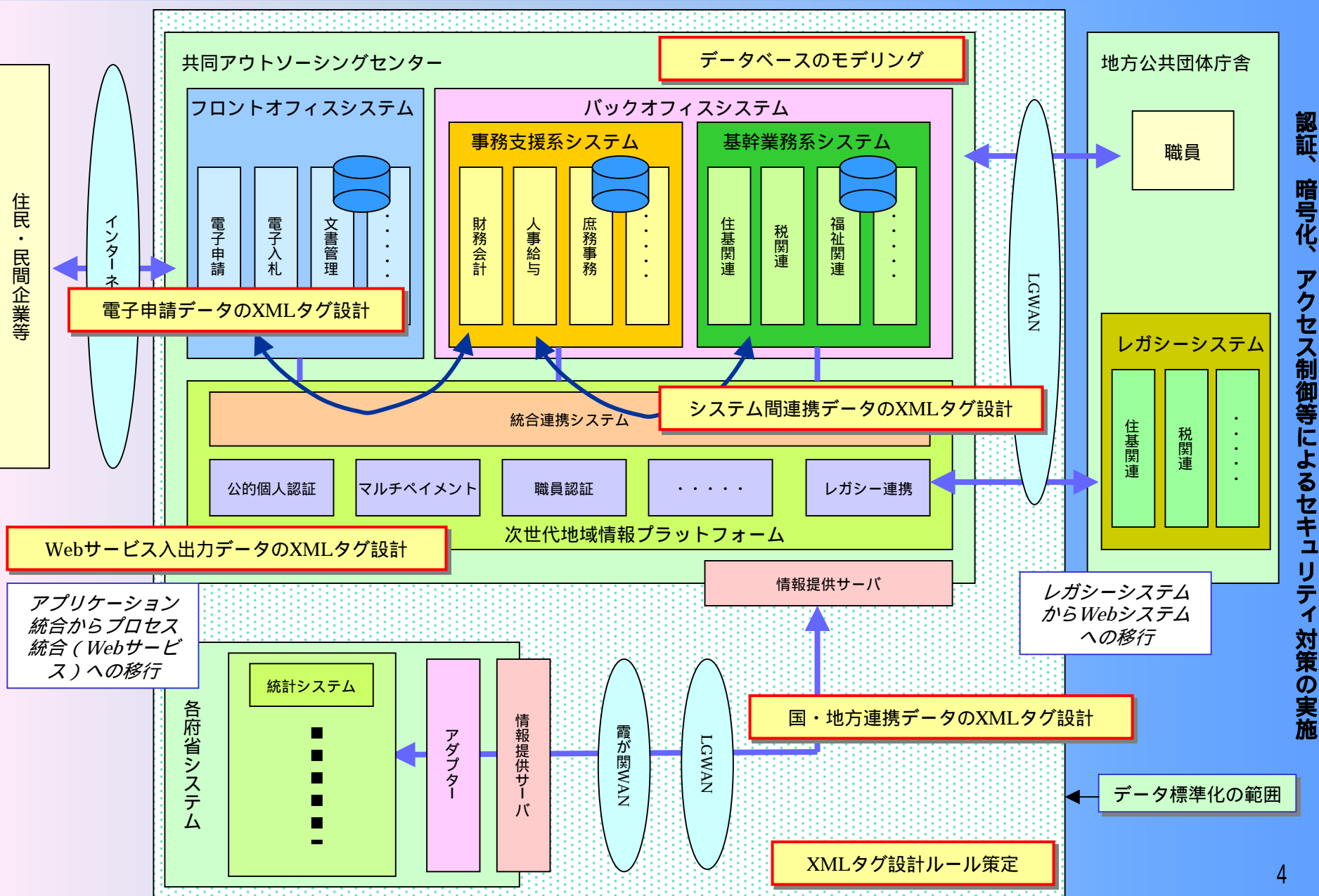
住民は転居や出産などの時に、必要となる手続き、申請についてそれぞれの窓口に出向いたり電子申請を行う必要がある。
業務システム間で連携が必要な場合には、職員が行う必要がある。

将来



住民は転居や出産などの時に、1回の届出を行うだけで、電子申請システムが必要な業務システムや民間企業への必要な手続き、申請を自動的に行ってくれるようになる。

目指すべき電子自治体の全体像



認証、暗号化、アクセス制御等によるセキュリティ対策の実施

XML仕様の作業項目と時期

標準化作業項目	概要	時期	備考	
XMLタグ設計ルール	要素・属性の命名規則	XML文書やタグ名等の名前の付け方に関する規則を規定する。	2005年3月	協議会了承済み
	名前空間の命名規則	名前空間名(URI)として、どのようなものを用いるかを規定する。	2005年3月	協議会了承済み
	文字コード・外字に関する規則	文字コードといわゆる“外字”の取り扱いについて規定する。	2005年3月	協議会了承済み
	スキーマ言語	どのスキーマ言語を選択するか、どの範囲の機能(サブセット)を使用するかについて規定する。主なスキーマ言語としてDTD, W3C XML Schema, ISO/IEC (JIS) RELAX NGの三つがある	2005年3月	協議会了承済み
	XMLプロファイル	XMLや名前空間のどの範囲の機能を使うかを規定する。XML文書において日本語を正しく取り扱うために従うべき規則を規定する。	2005年3月	協議会了承済み
	スキーマ言語プロファイル	XML文書の構造を定義するスキーマ言語の利用方法について従うべき規則を策定する。	2005年4月以降	2005年度から、まずはRELAX NGについて検討を開始する
XMLタグとスキーマ	XMLタグとスキーマの共通化	具体的な業務について、XMLタグ名およびXML文書の文書構造(スキーマ)を標準化する。標準化に先立って、業務プロセスとやり取りする情報(各種帳票、コンピュータ画面など)を調査・分析する必要がある。スキーマの定義は、DTD, W3C XML Schema, ISO/IEC (JIS) RELAX NGの3種類の形式で作成する	2004年から順次	共同アウト事業、レガシー連携事業、国・地方連携事業の成果を取り込み、一部の業務について試行中
	コード表	地方公共団体コードなど、既に政府・自治体で利用しているコードを洗い出し、XMLから利用できる形に整理する。	2004年から順次策定	できるところから順次作成中
	データ型ライブラリ	文字列によって表現されるデータとして、どのようなもの(単位、値のとりうる範囲等)があるかを規定する(例: 単位を百万円とする金額)。	2004年から順次策定	できるところから順次作成中
	共通部品スキーマ	多くのスキーマから使われる標準的な部品を策定する。人名や日付などを規定する小さな部品から、申請書などを規定する大きな部品までが考えられる。標準化に先立って、業務プロセスとやり取りする情報(各種帳票、コンピュータ画面など)を調査・分析する必要がある。	2004年から順次検討	共同アウト事業、レガシー連携事業、国・地方連携事業の成果を取り込み、一部の業務について試行中
XMLタグ・スキーマ等の標準XML文書の維持管理の枠組み	標準化したXML文書の維持・管理の枠組み(体制、仕組み、サービス等)を策定する	2005年4月以降	データ標準の維持管理の枠組みを策定後、ステークホルダ間の調整を行い、実際の枠組みを構築する	

データ標準化のスケジュール

	平成16年度	平成17年度	平成18年度	平成19年度以降
XMLタグ設計ルール策定	<ul style="list-style-type: none"> XMLタグ付けルールの策定 標準化するデータの範囲の策定 XMLタグの構造化、整理方法の策定 	<ul style="list-style-type: none"> 標準XMLの維持・管理の枠組みの策定 		
XMLタグ設計	<ul style="list-style-type: none"> 国・地方連携データ（一部分）のXMLタグ システム間連携データ（一部分）のXMLタグ 電子申請情報（一部分）のXMLタグ バックオフィスシステム（一部分）のデータベースのモデリング 	<ul style="list-style-type: none"> 国・地方連携データのXMLタグ システム間連携データ（順次）のXMLタグ 電子申請情報（順次）のXMLタグ バックオフィスシステム（順次）のデータベースのモデリング システム内データのXMLタグ（一部分） Webサービス入出力データ（一部分）のXMLタグ 	<ul style="list-style-type: none"> システム間連携データのXMLタグ 電子申請情報のXMLタグ バックオフィスシステムのデータベースのモデリング システム内データのXMLタグ Webサービス入出力データ（順次）のXMLタグ 	<ul style="list-style-type: none"> 官民連携によるWebサービス入出力データのXMLタグ
地方公共団体にとってのデータ標準化の効果	業務の効率化に寄与			
		<ul style="list-style-type: none"> 電子申請や庶務事務などと他システムとの連携の自動化が始まる。 システムの構築、連携などで業者に標準仕様を示すことでコスト削減につながる。 	住民の利便性の向上に寄与	
			<ul style="list-style-type: none"> 共同アウトソーシングセンターで稼動するシステムについてはデータの連携が自動化されている。 ほぼリアルタイムで住民からの申請、届出が完了ようになる。 	<ul style="list-style-type: none"> 住民へのワンストップサービスの実現を目指す。

データ標準化WGの進捗状況について

平成17年5月24日

総務省自治行政局

データ標準化WGの進捗状況

<第1回 データ標準化WG>

日時:平成16年11月30日 18:00～

場所:総務省第4特別会議室

概要:

- ・「データ標準化WG」を設置し、電子自治体推進パイロット事業等のこれまでの取組みの成果を十分に活用し、データ標準化に関する検討及び作業を実施していくことを確認
- ・データ標準化WGの成果をオーソライズする場である「データ標準化推進地方公共団体協議会」を設置
- ・岡山県、北海道、国・地方連携事業等における取組みを紹介

<第1回 データ標準化推進地方公共団体協議会>

日時:平成17年3月22日 10:00～

場所:合同庁舎2号館1階 共用会議室3

概要:

- ・データ標準化推進地方公共団体協議会の立ち上げを確認
- ・データ標準化の今後の取組み及びスケジュールを確認
- ・データ標準化WGで策定した、XML設計ルール(要素・属性の命名規則、名前空間URI命名規則、文字コード・外字に関する規則、スキーマ言語プロファイル、XMLプロファイル)を協議会にてオーソライズ

<第2回 データ標準化WG>

日時:平成17年4月25日 13:30～

場所:総務省共用第2会議室

概要:

- ・第1回データ標準化推進地方公共団体協議会会合にて、平成16年度におけるデータ標準化WGでの取組みがオーソライズされたことを確認
- ・情報通信政策局、自治行政局、行政管理局、統計局におけるデータ標準化の取組みについて報告。今後も協調しながらデータ標準化を推進していくことを確認

要素・属性などの命名規約

1. はじめに

この文書は、電子自治体における XML 文書において、要素、属性、スキーマ等に現れるシンボル（型名、パターン名など）にどのような名前を用いるべきかを規定する。

2. 国外との整合性を必要としない場合

ひらがな・カタカナ・漢字を積極的に用いる。今後の電子自治体を推進するには、業務に関する名称をすべて英語にすることは現実的ではない。現在、ひらがな・カタカナ・漢字を要素名・属性名・シンボル名として許さない XML ソフトウェアはほとんど存在しない。

以下に、使用できる文字を要約する。正確な定義は、XML 日本語プロファイルを参照されたい（4. 参考文献）。

- 原則として日本語とする。
- 全角英数字は使用できない。
- 半角カナは使用できない。
- 中黒（・）括弧は使用できない。

備考：この場合は、ファイル名やディレクトリ名にもひらがな・カタカナ・漢字を積極的に用いることが考えられる。

3. 国外との整合性を必要とする場合

英数字とする。Universal Business Language (UBL) Naming and Design Rules (OASIS standard) など、いくつかの国際的な命名規約がある。国際整合のために、これらの命名規約に従うことが必要な場合がある。

4. 参考文献

W3C ノート XML Japanese Profile, 14 April 2000
<http://www.w3.org/TR/2000/NOTE-japanese-xml-20000414/>

標準情報（TR） TR X 0015:2002 XML 日本語プロファイル
http://www.y-adagio.com/public/standards/tr_xml_jpf2/toc.html

Robin Cover, XML Naming and Design Rules Specifications Published by OASIS, UN/CEFACT, and Navy CIO, <http://xml.coverpages.org/ni2005-01-31-a.html>, 2005

Universal Business Language (UBL) Naming and Design Rules, OASIS Standard,
<http://www.oasis-open.org/committees/download.php/9943/cd-UBL-NDR-1.0Rev1b.pdf>,
2004

名前空間 URI 命名規約

1. はじめに

この文書は、電子自治体における XML 文書において、どのような名前空間名を用いるべきかを規定する。

2. 使用する文字

英数字のみを使用し、平仮名・片仮名・漢字は使用しない。

理由

XML 名前空間を規定する仕様書 (W3C 勧告 XML 名前空間 1.0) では、英数字のみしか使用できないように解釈できる文面になっている。実際、英数字しか受け付けない実装が存在する。

3. URN と URL の選択

URN (Uniform Resource Name) を利用する。

理由

総合行政ネットワーク (LGWAN : Local Government Wide Area Network) および霞が関 WAN では、インターネットとは DNS (1) が異なる。したがって、URL (Uniform Resource Locator) は、本運用環境とテスト環境とで動作が異なる危険性がある。

4. URN の構文

URN スキーム名としては、go.jp を利用する。本来は、IANA (2) に登録申請を行うべきである。

その次の名前として xmlns を指定し、XML 名前空間であることを表す。

最後に、年度と月を 200411 のように指定する。xmlns との間には英数字または一部の記号 (" (", ")", "+", ",", "-", ".", ":", "=", "@", ";", "\$", "_", "!", "*", "") が指定できる。

5. 例

URN の記述例を以下に示す。

```
urn:go.jp:xmlns:Chiiki-Kasseika-Jigyousai-Saishujutougaku:Youshiki2-2_Soukatsuhyou_Ic  
hiranhyou_Shichouson-Tokubetsu-Kubun:200411
```

6. 注意事項

URN を用いるという決定によって、名前空間 URI の指す場所に、RDDL(Resource Directory Description Language) 文書を置くことができなくなる。しかし、現実的な問題はないと判断する。RDDL は、一部では採用されているが、どの仕様制定団体からも公式には承認されていない。

7. 参考文献

標準情報 (TR) TR X 0015:2002 XML 日本語プロファイル

http://www.y-adagio.com/public/standards/tr_xml_jpf2/toc.html

IETF RFC 2141, URN Syntax, R. Moats, <http://www.ietf.org/rfc/rfc2141.txt>, 1997

IETF RFC 3986, Uniform Resource Identifier (URI): Generic Syntax T. Berners-Lee, R. Fielding, and L. Masinter, <http://www.ietf.org/rfc/rfc3986.txt>, 2005

IETF RFC 3987, Internationalized Resource Identifiers (IRIs), M. Duerst and M. Suignard, <http://www.ietf.org/rfc/rfc3987.txt>, 2005

(1) DNS (Domain Name System):

インターネット上のホスト名と IP アドレスを対応付ける仕組み。

(2) IANA (Internet Assigned Number Authority):

IP アドレスやドメイン名等のインターネット上で利用される資源の標準化や割り当てを行う組織。

文字コード・外字に関する規則

1. はじめに

この文書は、電子自治体における XML 文書において、文字コード・外字をどのように利用すべきかを規定する。

2. XML 文書の符号化

JIS X 0213:2004 の範囲の文字だけを、UCS (UTF-8 または UTF-16) によって符号化する。

3. 一般文章中の漢字の扱い

JIS X 0213:2004 に含まれる範囲の漢字だけを使用する。

4. 人名及び地名用漢字の扱い

情報交換用の符号化文字としては、JIS X 0213:2004 の範囲の漢字だけを使用することを原則とする。ただし、窓口における納税者からの要望などにより、JIS X0213:2004 で包摂される字体や字形の区別が必要な場合は、XML の要素を用いて情報を付加する。この目的のため、EGIX (JIS X 4166 として制定予定) が制定されつつある。

5. 参考文献

日本工業規格 JIS X 0213:2004, 7 ビット及び 8 ビットの 2 バイト情報交換用符号化格調漢字集合, 日本規格協会

日本工業規格 (予定) JIS X4166, XML 文書へのグリフ識別子の埋め込み, 日本規格協会

スキーマ言語プロフィール

1. はじめに

この文書は、電子自治体における XML 文書において、どのスキーマ言語をどの範囲で用いるべきかを規定する。

2. 一般事項

1. すべての制約条件をスキーマで表現しようと試みてはならない。ある制約条件をスキーマで表現してもよいのは、そのために必要なコストを上回るだけのメリットがあると確信できる場合に限る。
2. XML 文書の内容を変えてしまう機能(デフォルト値など)は使用しない(XML プロファイルの原則 1, 2 による)
3. 大規模なコード表(たとえば市町村コード)については、すべての値を列挙することはしない。スキーマ以外の手段によってチェックする。
4. ユーザ定義型には、原則として名前を付けることによって、意図を明確化する。例えば、小数点以下一桁までの decimal には名前を付ける。ただし、名前を付けても意図を明確にできない場合(いくつかの値を列挙するだけの場合など)は、名前を付けない。

備考： 単位を百万円とし小数点以下一桁までを表す有理数 = `xsd:decimal { fractionDigits = "1" }`

3. スキーマ言語の選択について

DTD (Document Type Definition) , W3C XML Schema, RELAX NG の三つの有力なスキーマ言語が存在する。それぞれの特徴を次の表に示す。

	長所	短所
DTD (ISO/IEC, W3C)	<ul style="list-style-type: none"> • 利用者が最も多い • 実装が多い • 採用事例が極めて多い 	<ul style="list-style-type: none"> • 機能が貧弱(名前空間とデータ型がない)
W3C XML Schema (W3C)	<ul style="list-style-type: none"> • 実装が多い • 機能が豊富 • 採用事例が多い 	<ul style="list-style-type: none"> • 利用者が多いとはいえない • 機能が複雑で使いにくい • 異なる実装が別の動作をする場合がある
RELAX NG (ISO/IEC, JIS)	<ul style="list-style-type: none"> • 機能が強力かつ単純で使いやすい 	<ul style="list-style-type: none"> • 利用者がさらに少ない

以下の三つのいずれかを選択するものとする。

1. RELAX NG で作成し、DTD と W3C XML Schema に自動変換する。
2. W3C XML Schema で作成する。
3. DTD で作成し、RELAX NG と W3C XML Schema に自動変換する。

4. RELAX NG プロファイル

1. XML 構文ではなく、短縮構文を用いる。
2. 属性の有無や値によって子要素が変わるパターンは、使用しない。また、いくつかの属性のうちの一つを選択するパターンは使用しない。

備考： この制限は、他のスキーマ言語への変換のためである。

3. ファイル名拡張子として、rnc を使用する（短縮構文）。
4. interleave (&) は使用しない。
5. データ型として、W3C XML Schema Part 2 (Second edition) を利用する。

備考： 平成 16 年度の国・地方連携プロジェクトでは、同一名前空間に属する同一のタグ名（正確には局所名）を、文脈によって異なる用途に使用してはならない。たとえば、title 要素を person 要素と section 要素とで共有してはならないという制限を設けている。

5. 参考文献

日本工業規格（予定） JIS X4177-2, 文書スキーマ定義言語（DSDL） - 第 2 部：正規文法に基づく妥当性検証 - RELAX NG, 日本規格協会

日本工業規格 JIS X 4159:2002 拡張可能なマーク付け言語（XML）
http://www.y-adagio.com/public/standards/jis_xml/toc.html

XML Schema Part 0: Primer Second Edition, W3C Recommendation, <http://www.w3.org/TR/xmlschema-0/>, 28 October 2004

XML Schema Part 1: Structures Second Edition, W3C Recommendation, <http://www.w3.org/TR/xmlschema-1/>, 28 October 2004

XML Schema Part 2: Datatypes Second Edition, W3C Recommendation, <http://www.w3.org/TR/xmlschema-2/>, 28 October 2004

XML プロファイル

1. はじめに

このプロファイルは、電子自治体における XML 文書が従うべき一般的な規則である。

W3C XML 1.0 勧告および W3C XML 名前空間勧告には、相互運用性が不十分であったり、総合行政ネットワーク（以下 LGWAN：Local Government Wide Area Network）および霞が関 WAN に適さない機能を含んでいるという問題がある。本プロファイルは、これらの問題を回避することを目的とする。XML 文書は、W3C XML 1.0 勧告および W3C XML 名前空間勧告に準拠するだけでなく、このプロファイルで追加される制約を満たさなければならない。

2. 原則

本プロファイルには、以下の原則に基づいて制約を設ける。

原則 1: 相互運用性が保証されていない機能は使用しない

XML 仕様に準拠するパーサ（ 1 ）であれば、結果が同一であることが保証されている範囲の機能だけを利用する。それ以外の機能を用いると、パーサによって結果が変わることがある。特に、電子署名・暗号化がまったく機能しなくなる恐れがある。

原則 2: ネットワーク依存性とローカル環境依存性の排除

LGWAN（または霞が関 WAN）上にある場合とそうでない場合、LGWAN（または霞が関 WAN）上にある一部のマシンが落ちている場合とそうでない場合とで、XML パーサの動作が変わる機能は使用しない。

原則 3: XML 仕様で誤って許しているものは禁止する

XML や名前空間の仕様には、本来許すべきでないものを誤って許してしまったところがある。そう広く認識されている部分については禁止する。

原則 4: DTD を利用可能にする

DTD（Document Type Definition）は、W3C XML Schema や RELAX NG 等のスキーマ言語に比べると十分な機能を備えているとは言えない。しかし、機能が単純で XML 規格の一部として古くから利用されているので、ユーザ数は最も多く、使いこなすためのノウハウも蓄積されている。XML 名前空間の用法によっては、DTD を用いることが不可能になるので、このような用法は避けることが望ましい。

この制限によって名前空間の本質的な機能が損なわれることはない。しかし、すでに実装されているプロトコルによっては、この制限を満たすことはできないことがある（すなわち DTD を利用できない）。

3. 規定

1. 外部解析対象実体（ 2 ）を使用してはならない。（原則 1,2 による）

2. 文書型宣言を文書に含めてはならない。(原則 2 による)

備考: ローカルな環境において、文書型宣言を追加して検証することは自由である。

3. 外部 DTD サブセットにおける内部解析対象実体(3)を宣言してはならない。(原則 1 による)
4. 属性のデフォルト値を使用してはならない。(原則 1 による)
5. DTD のデータ型 (属性の正規化) を使用してはならない。(原則 1 による)
6. W3C XML Schema のデータ型によって属性を正規化してはならない。(原則 1 による)
7. ルート要素の後にコメントや処理命令を書いてはならない(原則 3 による)
8. 名前空間 URI (Uniform Resource Identifier) として、相対 URI を使用してはならない(原則 3 による)
9. xsi:schemaLocation 属性を使用して W3C XML Schema スキーマを参照してはならない(原則 2 による)
10. Unicode との変換のとき、Unicode のコード位置と JIS X 0208 の区点位置との対応関係が、実装によって異なる文字 (JIS X 0208 の波ダッシュ、双柱、減算記号、セント記号、ポンド記号、否定、ダッシュ) がある。これらのコード位置を直接使用してはならない(原則 1 による)。詳細は、付録 1 と XML 日本語プロファイル(4 . 参考文献) を参照。
11. C1 制御文字 (U+0080 から U+009F) を使ってはならない。これは、文字参照による参照も禁止する。(原則 3 による)
12. 名前空間を持たない要素を使用してはならない。
13. 名前空間は、ルート要素で必ず宣言する。ルート以外の要素で宣言してはならない (別の言い方をすると名前空間を再定義してはならない)。(原則 4 による)
14. 同一文書内で同一名前空間 URI に対応する複数の名前空間プレフィックスを利用してはならない。(原則 4 による)

備考: この制限のもとでは名前空間 URI と名前空間プレフィックスが一対一に対応する。しかし、アプリケーションプログラムが名前空間プレフィックスを直接扱ってはならない。名前空間 URI を扱うようにプログラムしなければならない。また、要素名を文字列としてそのまま比較してはならない。名前空間 URI と局所名の対として比較しなければならない。
15. 文字コードは UTF-8 または UTF-16 とする。UTF-8 の Unicode signature はある場合もない場合も認める。UTF-16 は BOM (Byte Order Mark) を必須とするが、big endian と little endian の両方を許容する。これらは、XML のデフォルトである。また、Unicode への変換揺れを避けることもできる。
16. XML 仕様には、バージョン 1.0 とバージョン 1.1 があるが、1.1 は現時点では普及していないので、1.0 を利用する。同様に、XML 名前空間にも二つのバージョンがあるが 1.0 を利用する。

- 17.XML 宣言は指定を必須とする。宣言は、<?xml version="1.0"?>または<?xml version="1.0" encoding="utf-8"?> である。スタンドアロン宣言 (4) は含めない。XML 宣言の中に改行を入れてはならない。また、XML 宣言の後ろには必ず改行を入れなければならない。
- 18.名前空間 URI は、URN (Uniform Resource Name : urn://で始まるもの) とする。URN 命名についての規則は別に定める。

4. 参考文献

W3C 勧告 Extensible Markup Language (XML) 1.0 (Third Edition)
<http://www.w3.org/TR/REC-xml/>

日本工業規格 JIS X 4159:2002 拡張可能なマーク付け言語 (XML)
http://www.y-adagio.com/public/standards/jis_xml/toc.html

W3C 勧告 Namespaces in XML <http://www.w3.org/TR/REC-xml-names/>

標準情報 (TR) TR X 0023:1999 XML 名前空間
http://www.y-adagio.com/public/standards/tr_xml_ns/toc.htm

W3C ノート XML Japanese Profile, 14 April 2000
<http://www.w3.org/TR/2000/NOTE-japanese-xml-20000414/>

標準情報 (TR) TR X 0015:2002 XML 日本語プロファイル
http://www.y-adagio.com/public/standards/tr_xml_jpf2/toc.html

(1) パーサ :

ソースコードや文書を読み込んでアプリケーションソフトウェアから利用しやすい形に変換するソフトウェア。XML パーサは、XML 文書を解釈し変換する。XML の基本的なタグ付け規則に従っているかを検証するだけでなく、あらかじめ定義された文書構造との整合性を検証する機能を備えているものもある。

(2) 外部解析対象実体 :

参照する実体が XML データの一部と見なすことができるテキストデータであり、かつ外部ファイルであるものを指す。

(3) 内部解析対象実体 :

参照する実体が XML データの一部と見なすことができるテキストデータであり、かつエンティティ宣言の中に直接データが書かれるものを指す。

(4) スタンドアロン宣言 :

外部の文書を参照しなくても正しい文書が得られる XML 文書をスタンドアロン文書という。スタンドアロン宣言は、当該文書がスタンドアロン文書であることを通知するものであり、XML 宣言の中で "standalone = yes" のように記述する。

付録 1: Unicode のコード位置と JIS X 0208 の区点位置との対応関係が、実装によって異なる文字

JIS X 0208 の区点位置

- 0x815C(EM DASH/ダッシュ(全角))
- 0x815F(REVERSE SOLIDUS/逆斜線)
- 0x8160(WAVE DASH/波ダッシュ)
- 0x8161(DOUBLEVERTICAL LINE/双柱)
- 0x817C(MINUS SIGN/負符号, 減算記号)
- 0x8191(CENT SIGN/セント記号)
- 0x8192(POUND SIGN/ポンド記号)
- 0x81CA(NOT SIGN/否定)

Unicode のコード値

- U+00A2(CENT SIGN)
- U+00A3(POUND SIGN)
- U+00A5(YEN SIGN)
- U+00AC(NOT SIGN)
- U+2014(EM DASH)
- U+2015(HORIZONTAL BAR)
- U+2016(DOUBLEVERTICAL LINE)
- U+203E(OVERLINE)
- U+2212(MINUS SIGN)
- U+2225(PARALLEL TO)
- U+301C(WAVE DASH)
- U+FF0D(FULLWIDTH HYPHEN-MINUS)
- U+FF3C(FULLWIDTH REVERSE SOLIDUS)
- U+FF5E(FULLWIDTH TILDE)
- U+FFE0(FULLWIDTH CENT SIGN)
- U+FFE1(FULLWIDTH POUND SIGN)