

## 第二回 技術委員会 議事要旨

日 時：平成 25 年 1 月 9 日（水）13:30～15:30

場 所：三菱総合研究所 4 階 CR-C

出席者（敬称略）：

主 査：越塚 登（東京大学大学院情報学環 教授）

副 主 査：武田 英明（国立情報学研究所情報学プリンシプル研究系 教授）

委 員：平本 健二（経済産業省 CIO 補佐官）、深見 嘉明（慶應義塾大学大学院政策・メディア研究科 特任助教）、中尾 彰宏（東京大学大学院 情報学環 准教授）

オブザーバ：総務省 情報流通行政局、内閣官房情報通信技術（IT）担当室、経済産業省 商務情報政策局、国土交通省 国土地理院

そ の 他：新堂 克徳（横須賀テレコムリサーチパーク）

事 務 局：津國 剛（三菱総合研究所）

配布資料：

第 2 回技術委員会 議事次第

資料 2 -1 第 2 回技術委員会 座席表

資料 2 -2 第 2 回技術委員会 委員名簿

資料 2 -3 オープンデータの基盤であるボキャブラリに関する先進動向

資料 2 -4 技術委員会の論点

議 事：

1. 開会・資料確認

2. 中尾委員自己紹介

◇ 主査と同じく東京大学情報学環の総合分析情報学コースで、コンピュータサイエンスの応用学問について教鞭を執っている。専門はネットワークで、インターネットの研究を行っている。ネットワークのデータの方面から、オープンデータに少しでも貢献できればと考えている。

3. 技術委員会の論点について

◇ 本日の主な議題は 2 点ある。

まず、アウトプットになるところをどのような形で取りまとめていくかについて、事務局で整理したものについて説明していただき、その方向性で皆様にご了解いただきたいと考えている。

また、公共的な情報や政府の情報をオープンにしていく海外の取組みにおいて必ず出てくる、米国の NIEM やヨーロッパの ISA などの事例について、経済産業省で調査しておられるので、技術委員会で作っていく技術の規格にどう反映させていくか、相互運用性をとっていくかについて、ベースの情報共有を図るためにご紹介いただきたい。

コンソーシアム設立当初よりオープンデータに関する状況も変化しており、IT 戦略本部も動き始めている。そういったところへのコントリビューションもこのコンソーシアムの中で出していくべきではないかと考えている。そのため、前回よりも、本日お話しさせていただくことは多少範囲が拡大している。前回出された様々なご意見を反映した形で、今年度の議論をどのように集約していくかについての方向性を、委員の皆様と合意できればと思う。

当初は、総務省の情報流通連携基盤事業の標準のAPIとボキャブラリの規格を叩き台にブラッシュアップして、オープンデータをしようと思った人が使える本格的でレベルの高い技術になれば、ということに焦点を当てていた。しかし、他の取組みをいろいろ見ていくと、オープンデータにはいろいろな段階があり、そこにはいろいろなプレイヤーがいるので、ハイエンドだけではなく、そういったところへの対応も行った方が良いのではないかと思う。

全体としては一貫した体系を考え、その中でAPIの仕様と技術仕様を位置付け、必要であれば中間的で、本格的に技術的枠組みを使わない方々もオープンデータを使えるような規格を出したり、具体的で分かり易いガイドを作ってアウトプットしていく、というのが大きな方向性である。

資料2-4はかなり丁寧に書かれているため、2~3段階に分けて説明していきたいと思う。まずは、事務局より概要についての説明をお願いする。

- ・資料2-4 技術委員会の論点について、1~5 ページまで新堂氏より説明。

◇ ここまでのご意見はあるか。

データを出す時に、CSV ではないのか、という話がまず出てくる。CSV で出す場合は、どのように出したら良いのかということが、具体的な要望としてよく上がる。4 ページの論点4の「データ規格・API 規格のありかた」で、例えば、現在政府が出している白書の統計データはPDFのみである場合が多いが、オープンデータで出す際には具体的にどうすればよいのか、という疑問が出てくる。

段階を追って対処していく中で、理想的な段階まではいかなくても、最低限こうすれば良くなるという具体的なデータに即して説明するガイドを、この委員会の中で扱おうというのが、4と5の趣旨である。CSVのみでなく、他にもデータを出す際に要望が高い方式があれば、取り上げていきたいと思う。

前回に比べると、IT戦略本部の部分が、念頭に置くべきプレイヤーとして増えた。技術的な部分を行っている委員会なので、アウトプットを出していければと思う。

続いて、オープンデータのデータ規格とAPI規格についてだが、一般論として、オープンデータが対象とするデータにはどういったものがあり、扱えるデータの規格にはどういったものがあるのか、それら全体を分類学のように俯瞰して見た上で、必要に応じて最適なものを段階的に上手く使っていこうという意図がある。事務局からご説明頂きたい。

- ・資料2-4 技術委員会の論点について、6~56 ページまで新堂氏より説明。

#### 【資料に関するコメント及び質疑】

◇ 技術委員会で扱うオープンデータで対象とするデータや、それを表す既存のデータ規格の概要について説明していただいた。API規格に関しても、時代によって流行り廃りはあるが、いろいろなスタイルのものがあり、特徴によって分類することができる。データをオープンにしていく際に、データの特性に合わせてこれらの中から適切なものを使っていくのが基本的な考え方で、新しいデータフォーマットやプロトコルを一から考える必要はない、というのがこの分類から得られる知見である。

ただし、最適なものを選んだとしても、データを運用する側の人や、ボキャブラリ、ID、日本の場合は漢字の扱い、それらを公開する際のデータカタログやメタデータの作り方等がある程度決めていく必要があり、これらが技術委員会での議論の焦点になっていくのではないかと思う。

API 規格については、比較的決めることは多くはないが、具体的なレベルでの検討項目について 54 ページにまとめられている。さらに、現在動いている他のシステムの動向や一貫性、相互運用性も考えて行く必要がある。  
今までのところで、ご意見等があればお願いしたい。

- ◇ 6 ページの本委員会が扱うオープンデータの全体像のところ、前回は本委員会の対象データが、統計、地理、センサーデータ、物流、防災関係等と上がっていたが、今回防災が落ちて行政情報データがかなり増えているのは、環境変化によるものか。
- ◇ 対象としたものが変わったわけではなく、単に例を増やしたかっただけである。
- ◇ この書き方だと全部行うように見える。むしろ、この中で をつけた部分を集中的に行う、という書き方が正しいのではないか。後ほど紹介させていただく、経済産業省の取組みと重複しているところはかなりあるので、対象データをすっきりさせておく必要があるのではないかと思う。
- ◇ 扱うという意味も色々あり、ここでダイレクトに扱うものと外にあるものをそのまま使わせていただくものも含めている。
- ◇ 少ないリソースで行うので、この部分はどこから持ってくる、このデータは誰が行う、ということを書いたほうが良いのではないか。
- ◇ これをベースに、いろいろな情報を載せて対応等を整理して、将来マップになれば良いと考えている。
- ◇ 例えば、3 ページの部分に経済産業省も一緒に行っていくという雰囲気が出ると良い。行政情報については、経済産業省の取組みと一緒にいう形にしてしまった方が世間受けもするし、無駄な労力を省くことができる。  
次に、10 ページについてだが、「データ」と「データのコンテンツ」という分け方は、微妙ではないか。例えば、ツイッターのつぶやきのような、データのテキストデータはどちらに入るのか。
- ◇ オープンデータにおいては、機械が処理しやすい符号化したものをデータと呼ぶものと思うが、行政のオープンデータの流れの中では文書を出していくことも大きな枠組みとなっている。確かに文書はコンテンツだが、そういうものも扱うということを伝えたい。
- ◇ 上段はジェネリック、下段はデータを含むコンテンツになるのだろうか。ここで全てを扱いたいわけではないので、網をかけるのであれば、「データを含むデジタル情報」くらいにすれば問題ないのではないか。
- ◇ 表現の問題とスコープの問題があると思う。スコープの問題としては、符号というよりもグラフ化できる数字に集約されてしまい、データという言葉が独り歩きする可能性がある。それを、出来るだけ広げていきたいという意図があるのだと思う。それに関しては賛成だが、表現をどうするかは考えなければならない。その上で、どこから手を広げて、優先順位をどう付けていくのかは別の問題としてある。大前提は、数値や符号化されたものではなく、公共に資する資源に関しては、すべからく国民が利用出来る形でデ

ータ化し、かつオープンにしていく、という打ち出しができた上で、こういった手順を取っていくのかという議論やアピールをしていければ良いと思う。

- ◇ スコープを決める際に、頭の中では先回りして優先度の高いものが入るような規準を考えたいと思っていたが、総務省の情報流通連携基盤事業や IT 戦略本部の動きを見ると、優先度の高いものが思っていたよりもバラエティに富んでいた。CSV のように比較的わかりやすいフォーマットになっている統計データがある反面、白書のような文書のものもあり、両極端であるがどちらもニーズの高いものである。これらは、両方スコープに含めて、優先的に取り上げて良いのではないかと考えている。
- ◇ 今までの政府が行ってきた情報公開という流れの中で、本来終わっているはずの内容であるが、現状、日本における情報公開は違う意味で使われているように思う。用語の定義と活動が混乱している。今言われているオープンデータには、単純に情報を公開することまで広く含まれてしまっている。この点については、本来なら実務者会議の仕事かもしれないが、クリアに分けるのか、敢えて分けなくていいのかは考えどころである。  
かつて、文書の公開に関してあまり適切でない方法が取られてきたが、今回のスコープにデータとしてのドキュメントのような位置づけで文書を入れて、データとして扱う、と言ってしまっても良いかもしれない。
- ◇ オープンデータと今までの情報公開との違いは、オープンデータは単に情報を出せば良いのではなく、出したものが機械可読で、プログラムが書けるような形で出せることが重要だと散々言ってきたが、周囲からそれは言い過ぎという反応もあり、自分自身で妥協した部分がある。
- ◇ 過去のこともあるが、それは今、オープンデータに混ぜることによって挽回できる。ライセンスの問題をクリアにすることが前提条件である。結果的に、白書をどう再利用して良いかわからなかったこともクリアになり、同時に解決出来る。そういう意味では、データとしての文書やドキュメントとして入れておくのは悪くないと思う。
- ◇ 資料を作っている際に、ドキュメントファイルが出てくると、こういったものも含めておかないと矛盾してしまうと思った。
- ◇ 9 ページの静的・集約データとリアルタイムデータについて、データの集め方とアクセス方法等、実装の違いによって類型をしているのはよく分かるが、10 ページはどのような目的でこのような分け方をしたのか疑問に思った。全体として、白書のようなデータも扱う、としておけば、無理に分けなくても良いのではないかと思う。
- ◇ 分け方の問題というよりも、姿勢を促すようなものになっていくのだと思う。極論をすると、条例に基づく請求を行い、法令や判例として認められた場合に、役所に赴いて鉛筆で写すことも情報公開である。それに対して、積極的に使ってもらえる形でデータを出そうというのは、技術的な話というより姿勢の話である。その姿勢に応える適切な技術に関して、我々が真意を示すということだと思う。
- ◇ 3 ページの国際標準化について、2 点気になる。  
他国で、API 規格等を ITU T に持って行った事例はあるのだろうか。大きな枠組では、

RDFのような一般的なものについては W3C があるが、オープンデータに限った形で先行しているところがあるとしたら、そこでの統合が気になる。もしない場合には、急先鋒になって対応したほうが良いのではないか。

また、標準化とは何を標準化するのか。新たなデータフォーマットや API 規格を作って標準化するというのはわかりやすいが、今あるものを適材適所で使うのであれば、使い方を提案するのか、システムを提案するのか、何を標準化するのか出口のところ気になる。

ガイドラインを作って実務者会議に出すのは非常に分かりやすい出口だが、標準化には既に何が存在していて、この委員会で何を新たに標準化するのがわかりにくい。

- ◇ どういう組織・場所で、どういう作業が行われているのかについては、今年度調査していくことになっている。  
知っている範囲で言うと、静的・集約的データについてはあまり詳しくないが、リアルタイムデータの標準化に関して、ITU-T に提出されたことはないと思う。W3C は全て把握しているわけではないので分からない。  
静的・集約的データに関して、コンテンツのボキャブラリの標準化については、ITU-T でも随分行っているようだが、マルチメディアに特化しており、今回の対象とは大分異なっている。リアルタイムデータ、センサーから出てきたデータを、クラウド等に集約する際にどうするのか、民間の通信事業者が取り扱う際に、センサーデータのプロトコルをどう集約するかに関しては、今、最もホットなトピックとなっている。
- ◇ スマートグリッドでも行っているのか。
- ◇ ITU、ISO、IEC 等いろいろな分野で、センサーデータを広域ネットでどう集めてくるかに関しては取り上げられており、すぐにはまとまらないくらい様々なものが出ている。
- ◇ W3C で標準化されたものの一部は、ISO のファストトラックとして認められているが、W3C レコメンデーションになった後に、ISO 標準になるには 4 ~ 5 年かかっている。HTML4.01 が ISO から JIS に落ちてくるまでも年単位でかかっている。  
W3C 自体がデファクトの団体であり、産業コンソーシアムであると言い切っており、強制力はない。WEB 標準というのは基本的にデファクトで、皆が使うから使う。実質的には、パテントフリー仕様の集約機関としての位置づけになっている。RDF のような基礎的なもの、マークやダブリンコアもコミュニティで利用されており、W3C もコミュニティだが、相互のリエゾンをどう組んでいくかについては、W3C の中でも課題となっている。デファクトだがある程度公的な形で活動している W3C、IETF、ISOC 等が、ISO、ITU、IEEE のような形のところと、どうリエゾンを組んでいくのかについては、未だに議論の段階である。
- ◇ 標準化の出口の可能性で言うと、データの形式は既存のものを使うとして、標準化の中心はボキャブラリだと思う。それをどこで標準化するかは別の話だが。  
例えば、総務省の事業で行っている食品トレーサビリティで使うようなボキャブラリに国際標準があるかといえば、ない。データを広げていくと、ボキャブラリがある分野はかなり限られている。そういう意味では、挙げられている 4 つの課題は、標準化の可能性があるところである。  
政府のデータをどう扱うのか、人をどう表すのか。今後、仮に政府が国民総背番号を付けたとして、そこで扱う人間のデータとモデルが一致しているかどうか曖昧である。

それが一致しなくなると、総背番号が付いている前提での人の在り方が変わってくる。ボキャブラリ系はまだまだ標準化する余地がある。

プロトコルに関して、REST はモデルレベルのものであり、具体的なプロトコルは個別に決めなければならない。標準化した方が良いかどうかは別にしても、規格としては有り得る。

また、オープンデータ化を行う際の手順やガイドライン等、高次元のものも国際標準化していくというのが ISO の流れであるので、こういった手順を社会的なプロセスも含めて行っていくというのも 1 つのパターンである。

メタデータ間の調整の仕方や相互運用に関しては、まだまだ標準化しなければならない。センサネットワーク関係の標準化についても、機械間で標準化しなければ通信できないので、かなり具体的に決めなければならないが、その部分についての標準化は全く終わっていない。

これらについて、政府が音頭を取ってどこを狙っていくのか。取りやすい、取りにくい、やるべき、やるべきでないということもあるので、いろいろと考えた方が良いと思う。

- ◇ API について、ここではデータのアクセスの仕方が列挙してあるだけだが、実際にはこの上に、どういった検索を行うのかといったことを載せていかなければならない。
- ◇ 標準化といえば標準化だが、古典的な標準化とは状況が異なる。オープンデータや LOD は、古典的な標準化を行わなくても、きちんとデータ交換できるようにしようというのが本来の趣旨である。データ交換するためには、レイヤの異なるメタなところで、何らかの取り決めは必要となる。  
現在、標準化団体もそういった新しい部分に踏み込んで来ているので、広い意味での標準化、国際連携になるのではないかと。有り方については、この委員会の中で、皆様のご意見を伺いながら進めて行きたい。
- ◇ 先ほどの HTML4.01 の ISO 化についてだが、W3C でレコメンデーションが標準化されたのが 1999 年 12 月で、HTML4.01 の ISO 版のファーストファイルである ISO/IEC15445 が出たのが翌年 2000 年 3 月である。ただし、各国に持ち帰ってフィードバックする期間があるため、ファイナルバージョンができたのは 2003 年で、3 年かかっている。持ち帰って議論する場合は、実装されたものを使ってみてのフィードバックとなる。HTML4.01 を作り始めたのは 1990 年の半ばなので、最終的に使われるようになるまで、10 年ほどかかっている。議論が前倒しに行われて、最終的に ISO 標準できちんと反映してもらうためのコメントを出すためにも、デジュールになる前の段階で如何に実装を先に進められるかが肝になってくる。標準化と実装の順番は逆だと思ってい。
- ◇ ITU T と書いてあると、デジュールに持って行く古いタイプの標準化を思い浮かべがちだが、もしかするとデファクトなのかもしれない。  
例えば、M2M やスマートグリッドなど、センサー系の機械同士で通信をしているところは、こういった話が必要なので、そちらの話も注視しながら進めて行かなければならない。
- ◇ M2M は、出口を考えると、単にローレベルでデータ交換のプロトコルを決めるだけではなく、上位で交換されるデータの意味やセマンティックスが自動交換されるようにしなければならない。最上位では、ここで扱っているようなボキャブラリを決める必要があるが、センサネットワークはそこまで到達していない。まだ、ネットワークウェアや

トランスポート等の議論を行っている。全体的なフレームワークのアーキテクチャの話になると、そういったものも必要だと話にのぼるぐらいである。そこへ、どんどんインプットするというのもあるかと思う。

- ◇ 提案としてはどこでも行っているが、結果的にデジュールになってから、もしくはデファクトも含めて標準の間で、どれが一番シェアを取るか、その先に進んでいくのがポイントである。標準を取ることを目的にするだけではなく、標準をきちんと実効性のあるものにしていくところまで見据えた形で作っていかなければ、作っても結局使われなかったり、世界の趨勢から外れてしまう。
- ◇ 政府が大きなプレーヤーということもあり、オープンデータのシステムは公共調達になるので、WTO 絡みが必ず出てくる。国内で行っているデータがあるのであれば、そこに引っ掛かるところは優先度が高いと思う。
- ◇ 43 ページの論点3の文字コードについてだが、論点として落としても良いのではないかと思う。「Unicode で表現できない部分がある」と書いてあるが、戸籍の文字まで全てトラックに載っているので、あと数年で戸籍データまで全て Unicode に載る。戸籍に載っていない外字は、法務省的には認めないと言っているので、この点についてはほぼ解決している。Unicode と同じところに CHN と3文字入っているものがある。文章そのものに JA、CN と入れてしまえば解決できるので、あまり深く掘らなくて良い気がする。
- ◇ ここで扱うつもりはなく、単に指摘の意味で載せたので、論点に書かない方が良かったかもしれない。

#### 4. ボキャブラリに関する先進動向について

- ・資料 2-3 オープンデータの基盤であるボキャブラリに関する先進動向について、平本委員より説明。

##### 【資料に関するコメント及び質疑】

- ◇ 毎年11月前後に TPAC という年次総会があり、その中でオープンガバメントに関連する議論のトラックも設けられている。その中でオープンデータ関係のトピックは、2つに大別される。一つが、ユースケースを放り込んでまとめていくというもので、比較的ボキャブラリ、データセットに寄った形で進んでいる。もう一つは、ピュアテクニカルなもので、XML-RDF ベースだけではなく、JSON を如何に取り扱うかとか、Five Star の枠組の中で如何に拡張して集めていくかとか、LOD アプリケーションを如何に良く作っていくか、といったことを行っている。それぞれキャラクターも異なり、いろいろと並行して走っているのが現状である。  
ヨーロッパからユースケースの事例、提案をどんどん放り込んでいるという話があったが、基本的にはユースケースベースでの話になる。なぜなら、コアテクノロジーは、シンプルに三つ組や RDF/XML にリンクするだけである。如何にユースケースをベースにして、適切なボキャブラリを標準化していくか、アプリケーションを作る上でのノウハウを如何に標準化して普及促進していくかに、変わってきている。標準化を本当に狙うのなら、ゼロから一つ一つ検討するのではなく、今あるものを使いながら足りないものをプラスして、如何に早く実効性のあるユースケースを作って、そこからの知見を投げ込むかという話になる。

- ◇ UCore はなくなってしまうが、UCore はアプリケーション作りまでセットで考えていたので、そのノウハウがここに入ってくるのではないかと思う。  
W3C と言えば、今月 21 日に国際会議のテレカンがあるが、今回は日本の発表になっている。各国が順番にオープンデータの各国の取組みについて説明するのだが、日本はアメリカのコンサルタント会社が代行して発表することになっている。
- ◇ NIEM や ISA も、政府のデータをオープンにすることが最初の主眼だったというよりは、電子政府を実現するための省庁間のデータ交換フォーマットとして始まっており、オープンデータの際に、交換したものをそのまま使って公開すれば良いのではないか、という流れだと理解して良いか。  
経済産業省の取組みは、道具としての NIEM なのか、目的としての NIEM なのか、どちらなのか。
- ◇ 目的として作っても仕方がないので、道具として使えるものを作ろうということである。
- ◇ NIEM の枠組みを使ってオープン METI として行っていくということか。
- ◇ データ項目が揃っていけば、オープンデータにも繋がるし、IEPD のような形にもなる。
- ◇ いきなりオープンデータから行うのではなく、電子政府で政府間のデータ交換が電子的に行われるという前提があって、外に出すというのは正論だと思う。やはりアメリカはすごいと思った。

## 5. 技術委員会における外部仕様書案について

- ・資料 2-4 技術委員会の論点について、55～74 ページまで新堂氏より説明。

### 【資料に関するコメント及び質疑】

- ◇ 前回も外部仕様書について説明したが、今回は少し広がっている。56 ページにあるように、コンソーシアム全体のアウトプットとしてはいろいろなタイプの技術ガイド等も出して、電子行政オープンデータ実務者会議へインプットしていきたいと思う。  
72～73 ページの CSV 規格については、まだまだアイデアの段階だが、RDF を最上位規格に置くとスマートに行くといったようなガイドラインを作っても良いのではないかと考えている。RDF ベースの外部仕様書のみでなく、表形式のものをそのままシンプルに扱う形のガイドラインを、最終的にわかりやすい形でアウトプットしていければと思う。  
ここまでのところで、ご意見があればお願いしたい。
- ◇ 56 ページの技術ガイドのところ、「公共文書のオープン化推奨ガイド」となっているが、ライセンス等ガバナンス委員会が関わってくるものが多いと思う。これは技術ガイドとして書くのか、それとも全体のコンソーシアムとして書くのか。表形式と書いてあればわかりやすいが、これは難しい気がする。
- ◇ 社会的プロセスも含めたガイドになっていたほうが良いかもしれないが、技術に関しては我々も IT 戦略本部も共通だが、プロセスに関してはこのコンソーシアムと相手の状況は違うので、こちらでインプットしたものがそのまま当てはまるのかどうかわからない。  
ライセンス等は扱う主体にかなり依存するし、ガバナンス委員会で出てきたことをこの

場で取りまとめてインプットするのは逸脱する気がするので、「公共文書のオープン化推奨」というタイトルを変えて、技術に絞った方が良くかもしれない。

- ◇ そうした方が良くと思う。
- ◇ 何を選択するかについては書けないが、ライセンスにはこういったものがあり、それらの中から選択して使う、一般的にはこのように使われる等、データを公開するための技術的な手順としては書くことが出来ると思う。データを公開する際に、担当者が何を入力すればできるのかについては首尾一貫したほうが良いので、そういう意味では入れた方が良くと思う。
- ◇ コンソーシアムは技術委員会とデータガバナンス委員会、実務者会議もデータ WG と普及 WG がある。現在、ガバナンス委員会でも同様に、担当者のためになるようなものがどういうものかについて議論が行われているが、そこは何らかの形でインプットされて行くことになると思う。
- ◇ 最終的にはきちんとマージされると思うが、それをどこがやるかは要調整かと思う。
- ◇ 越境していると言われるくらい、どんどん出していったほうが良くと思う。流動的な部分なので、お互い遠慮し合っているとカバーできなくなってしまう。
- ◇ 他にご意見等があれば、事務局までメールをいただければと思う。

## 6. 次回スケジュールについて

- ◇ 第3回、第4回の開催日程については、改めてメールにて調整をお願いするので、ご協力いただきたい。

以上