

第 2 回 データ環境推進委員会 議事要旨

日時：2024 年 8 月 8 日(木) 16:00-18:00

場所：独立行政法人情報処理推進機構 15 階 委員会室 1 及びオンライン

出席者(敬称略)：

【委員長】

武田 英明 大学共同利用機関法人 情報・システム研究機構
国立情報学研究所 情報学プリンシプル研究系 教授

【委員(50 音順)】

瀬戸 寿一 駒澤大学 文学部地理学科 准教授
高木 祐介 株式会社自動処理 代表取締役
玉木 浩貴 独立行政法人国立印刷局 官報部
ベース・レジストリグループ グループリーダー
西村 正貴 独立行政法人統計センター
情報システム部システムソリューション課 課長
眞野 浩 一般社団法人データ社会推進協議会 専務理事

【傍聴】

阿部 隆弘 独立行政法人国立印刷局 官報部
ベース・レジストリグループ 主任専門官

【事務局】

独立行政法人情報処理推進機構

議題：

0. 議事次第・出席者確認
1. 国際データガバナンスガイドの骨子の確認
2. GIF 関連事業の確認
3. コード・コード一覧と公開方法の確認
国際データ語彙マッピングのスキームと実施方法の確認
4. 各種活動の状況について
5. クロージング

配布資料：

- 資料 0 議事次第
- 資料 1 委員名簿
- 資料 2 議論のポイント
- 資料 3 GIF 準拠ルール
- 資料 4 コアデータモデル改訂案(個人・法人)
- 資料 5 GIF-IMI の関係
- 資料 6 コード・コード一覧
- 資料 7 Core Data Model Mapping Project2
- 資料 8 ワンスオンリー入門ガイド
- 資料 9 データマチュリティ啓蒙リーフレット
- 資料 10 提供予定ツール等の概要
- 資料 11 データガバナンス・ガイドラインの整備について

議事概要：

■事務局連絡

- ・ 配布資料の確認
- ・ 本日の議事進行について、委員長に委ねる旨説明

■本日の議題ならびに議論のポイントの説明

- ・ 事務局より、本日の議題並びに各議題の議論ポイントを説明。

■議題 1. 国際データガバナンスガイドの骨子の確認

- ・ 事務局より、経済産業省におけるデジタルガバナンスコードの改訂状況、デジタル庁で行っている国際データガバナンス・ガイドラインの策定状況、および本日の論点である「データガバナンス・ガイドライン（案）」の目次案と各章における記載概要の説明を、資料 2 に沿って説明。
- ・ 続けて、目次案や各章における記載概要の案に対し、このような観点での内容も記載した方が良い、といった意見や、記載の上での留意点などあれば、意見をいただきたい旨を説明。
- ・ 上記説明後、下記の質問・コメント等が委員より寄せられた。
 - (委員)データ分析をする際に AI を使うことは当たり前となっており、データ分析の知見が十分でなくとも、AI を用いることにより、品質の良いデータがあれば、誰でもデータ分析が可能となっている。このため、データ単独でなく、データ活用の例など利用者目線で盛り込むと、データを使おうという気になるではないか。AI の普及は社会的な環境であると言えることから、2 章に追記が良いと考える。
 - (委員)一方で、ガイドラインであることを踏まえると、AI を利用した解析の話は、データ活用を行う上でのヒント的に、軽く 6 章でふれることが良いとも考える。

- (事務局)AI の活用によるデータ分析は、データ活用やデータ取引において重要な話題であることから、話の整合を意識しつつ、各章で繰り返し触れる方法もあり。
- (委員)扱うデータについても、従前の構造化データや数値データだけでなく、通常業務で用いるマニュアルや報告書といったテキスト文章にも価値がある時代となり、これらもデータ解析対象となり得る。このようなデータのストックも重要であること、AI の普及により、データサイエンティストがいなくともデータ分析ができるようになってきたこと、これらが企業の成長や業績拡大につながることも明示した方が良いと考える。
- (委員)AI の普及により、ツールやデータがよりリーズナブルに活用できるようになってきた状況の時に、今後どのような人材を育成していくべきと考えるか？
- (事務局)具体的な内容については、現在各所と議論中であるが、これまで人材育成はデータサイエンティストに偏っていたが、データマネジメントの人材も必要であることや、CDO の役割を定義すること、流動化している IT 人材の管理といった点の議論を踏まえ記載することを考えている。またリテラシーについても触れたいと考えている。IPA は、デジタルスキル標準や情報処理技術者試験がメインストリームとなるが、他にも金融のスペシャリスト、地理空間のスペシャリストなどといったデータのスペシャリストがいることを経営者向けに示すことも必要と考えている。
- (委員)データ分析をする際に、AI を使うことは当たり前となっており、データ分析の知見が十分でなくとも、AI を用いることにより、品質の良いデータがあれば、誰でもデータ分析が可能となっている。このため、データ単独でなく、データ活用の例を利用者目線で盛り込むと、データを使おうという気になるではないか。2 章に追記が良いと考える。
- (委員)データ管理は雑務ととらわれがちな面もあるが、これだけデータが重要視されるようになった時代であればこそ、データ管理が企業全体にとって、極めて重要な役割であること、日々のデータ管理がデータに価値を生み出すことを、4 章で示すとよいと考える。
- (委員)「データ主権」の観点から、データに関する基本的なキーワード、「データは無体物であること」、「データ排他的な所有をしかねない」といった法的な考え方を 3 章に示しておいた方が良いと考える。
- また、データ品質については、誰にとっても完全な品質を保証するというわけではなく、データを活用する側の要件やニーズに合致しているかどうかである。経営者に対しても、このような品質保証の考え方を、4 章に示しておく必要があると考える。
- 人材については、データの生成（収集）、分析、活用、抹消のデータライフサイクルを下敷きに、そのサイクルのすべてを把握し、適宜必要な意思決定等を行う役割が CDO、それぞれのライフサイクルステージで求められるスキル、能力を定義し、そこにエキスパートが置かれる、という形で示すのが良いと考える。
- (委員)DX やデータ活用のよりスムーズな推進するためには、スキルだけでなく、その前段にコンピテンシーやマインドが求められるように感じる。抽象的になるかも知れないが、この点にも触れられるとなおよいと考える。

- (委員)AI に任せておけば何でもできる、という風潮を昨今感じるが、AI は事実とは反する結果を返すこともある。AI により正確なアウトプットを求めるためには、統合化されたデータではなく、分散でもよいので構造化された品質の高いデータをインプットすることが重要であることや、分散化しているデータを有意に繋げるために、タミノロジー（オントロジー）の整備が重要になることにも触れるとよいと考える。
- (委員)AI は結果しか返さないため、結果に至るプロセスは人間が確認をし、判断をする必要がある。この部分のスキル、リテラシーがAI を扱う人間には求められることを人材に示しておくとういとお考える。
- (委員)データ活用に関して、ネガティブな面も示しておいた方がよいと考える。データを蓄積することによるウィルス感染や個人情報の漏洩、持ち出しだけでなく、判断に用いたデータの出所や品質を確認する必要があること等も、経営者は認識しておくべき事項である。
- ・ 提示した目次案に対し、今回の委員会で出たコメント・意見を踏まえ事務局で修正を加えたものを、再度メールにて今月末を目途に提示する。この修正案に対し、各委員は確認を行うことが、本委員会にて承認された。

■議題 2. GIF 関連事業の確認

【議題 2-1. GIF 準拠ルールの確認】

- ・ 事務局より、何を満たせば GIF 準拠と言えるのかが現状不明瞭であることの課題、GIF モデルの機能、および、課題に対する対応策として、GIF のデータモデル準拠ルール案として LV1、LV2、LV2s の 3 段階のレベルを設けることと、実際に準拠しているかどうかは、別途検討するツールを実行することで確認していく方針であることを、資料 3 に沿って説明。
- ・ 上記説明後、下記の質問・コメント等が委員より寄せられた。
 - (委員長)本日のゴールは、事務局から提示のあった GIF 準拠ルールの、仕様書に明示して調達等を行い、実際の納品物（成果物）に対し、ツールによる確認を行うことで、GIF に準拠したものであるかどうかを検証、判断することで問題はないかを確認する場となる。
 - (委員)GIF 準拠 LV1、LV2、LV2s について、レベル感の説明はざっくりし過ぎではないか？
 - (事務局)WG でもこの点は問題提起されており、各レベルで求める設計粒度等は、別途詳細な説明文書を作成する予定である。
 - (委員)ルールやツールができた段階で、例えば過去の政府調達の仕様書等を用いて、ボランティアを募っての検証（実際にツールが想定通りに機能するか、ツールの結果から正しく判断、検証ができるかなど）が必要であると考え。また、このルールやツールを用いて調達・検証するのは複数の行政機関となることから、官側に対してもルールやツールの使用方法についての正しい理解をさせるためのアクションが必要であると考え。
 - (委員長)検証は重要であると考え、ルールやツール整備の最終段階で IPA にて必要なイベント等を企画し、問題点の抽出や改善を図ってほしい。

- ・ 事務局より提示された方針に沿って、GIF 準拠ルールや準拠を確認するツールの検討・整備を事務局で引き続き進めていくことが、本委員会にて承認された。

【議題 2-2. GIF コアデータモデル(個人・法人)の確認】

- ・ 事務局より、現在 GIF で公開しているコアデータモデル（社会の様々な場面で用いられている、個人や法人、住所や連絡先などの基礎的な情報）について、利用者などからいくつかの指摘を受けていることや、欧米のデータモデルも公開時から更新がされていることから、インタオペラビリティを確保する観点からも、改訂が必要であることを説明。
- ・ WG にて検討したコアデータモデル（個人・法人）の案について、資料 4 に沿って説明。ポイントは下記の通り。
 - データ項目について、データ設計に必ず組み込むべき「基礎項目」と、用途や業界ごとの事情に応じ任意に追加することが可能な「拡張項目」の 2 つに分離。
 - 基礎項目について、設計時に参考としてもらうべく、データの利用用途や目的等に応じ、データ設計への組み込みを必須とする項目のパターンを設定。
- ・ 上記説明後、下記の質問・コメント等が委員より寄せられた。
 - (委員)基礎項目についてはミニマムが良い。数多くの項目を必須としたり、提示したりすることで、空白が生じたり、この項目はないので GIF に準拠できない、といったことの発生を抑制できる。
 - (委員)「形式」という項目があるが、ここは、機械判読性の点からは、型・TLV という観念（たとえば、整数か小数か、など）からの観点と、全角や半角といった、表記方法の観点の、2 つの観点から書き分けた方が良いと考える。
 - (委員)この資料は、プログラミング向けの資料でなく、データモデル検討のための資料の位置づけと考えていることから、あまり詳細な資料としたりたくないこともあり、どこまでの粒度で書くか、悩みどころである。
 - (委員)一方で、この資料にどこまで役割を持たせるか、という議論でもある。例えば、一旦は項目に入る、というレベルでの情報を提供とし、別途バリデーションツールを提供し、各項目に入ったデータが有効でない場合は、エラーでその旨知らせるといったようなこともできるのではないか。JSON Schema で定義できるところまでは書きたい。
 - (委員)詳細な情報までは書かない方針であれば、リコメンデーションとして、形式や文字長ということを記載しておく、例えば、今後業界ごとに個々の拡張項目の検討を進める際に、有益な情報提供になると考える。
 - (委員)免許取得時や更新時、不動産取引時等の行政手続時に登場する、本籍地がないが、あった方がよいのではないか。
 - (事務局)現在の行政手続時には、本籍地が登場するシーンもあるが、本籍地そのものが手続きの本質的な部分で必須であるとは考えにくい面もあり、行政手続の BPR 的な点も踏まえ、

今回の案では掲載していない。

→(委員)今後、本籍地の扱いが必要となった場合には、拡張項目として議論すればよいと考える。

→(委員)「住所」とは、現在住んでいる場所を指すのか、住民票登録をしている場所を指すのか？

→(事務局)説明に「居住住所」とあるとおり、現在住んでいる場所を入れてもらう認識である。

→(委員)現在、複数拠点に居住している方も少なくなく、このような方に対し、複数の住所の記入ができるようにすべきではないか。このため、「主な住所」、「他の住所」等の項目名が良いと考える。複数住所の記入を考慮すべき場合は、拡張項目において、適切な項目名とともに、複数の住所に対応できる項目を今後検討すればよいと考える。

→(委員)項目名に、例えば身長は cm で入力するといった、記入上のルールを今回設けた点は、データモデル設計時に、どのような単位で入力することを想定したものであるが、データモデル設計者が把握できる点で、有効であると考ええる。

- ・ 今回事務局にて提示した、個人・法人のデータモデルの案について、IPA よりデジタル庁に対し、GIF のコアデータモデル改訂版として提案することについて、本委員会にて承認された。

【議題 2-3. GIF と IMI の今後の整備方針の確認】

- ・ 事務局より、GIF と IMI の両者は、現状明確な関係性を持っておらず、両者の定義のずれが発生していることから、その解決のために、WG にて検討し整理を行った事項を、資料 5 を基に説明。ポイントは下記の通り。
 - IMI のコア語彙から、GIF のコアデータモデルが参照する中間語彙（コアデータモデル語彙）を作成
 - 上記仕組みにより、GIF と IMI の関係性を明確化させるとともに、データモデル利用者とデータモデル設計者の双方、それぞれが意識しなくとも、「GIF のデータは、実は IMI でつながっている」ことを実現
- ・ 加えて、既存の IMI サイトが提供する技術や語彙などを、GIF を支える基盤として再整理し、IMI3 としてリニューアルし、公開することを考えており、コンテンツごとに IMI3 ではどのような公開方針となるか、案を資料 5 に沿って説明。特に、コア語彙の英語化により、GIF の基軸となる英語用語名との整合が高まることを説明。
- ・ 議題 2-3 については、本日は全体方針の確認であり、この方針に沿って進めていく上で、気にすべきこと、考慮すべきこと等あれば、後日、事務局に委員がメール等でコメントを寄せることを確認。

■議題 3. コード・コード一覧と公開方法の確認

- ・ 事務局より、現在、GIF の情報として公開されている、コード・コード一覧（Excel 表形式での公表）について、現状、GIF での公開後、一覧のメンテナンスが行われておらず、情報が陳腐化しており、公開情報の定期的なメンテナンスが必要性と、閲覧者にとって分かりやすい公開が必要であることと、IPA にて整備した今後 GIF サイトで公表予定のコード・コード一覧の整備状況について、サンプルを交えての説明を行った。
- ・ あわせて、今後の公開方法として、引き続き GIF サイトの GitHub 上でのコード一覧の公開を行うとともに、IMI サイト上にて、コード・コード一覧の検索を可能とできるよう、公表を行う予定であることを説明。
- ・ 上記説明後、下記の質問・コメント等が委員より寄せられた。
 - (委員)メンテナンスはどのように行うのか？
 - (事務局)リンク切れを起こしていること自体は、機械的に確認ができる一方で、どのリンクが正しいかといった判断は、人が行う必要がある。このため、メンテナンスの最適な頻度などは、今後検討の必要あり。
 - (委員)現状、GIF サイトでの情報が最新化されていないのはなぜか？
 - (事務局)現状、デジタル庁において、最新情報を定期的に自身で収集する、もしくは広く社会で利用されているコードを強制的に提出させる、といった仕組みが構築されていないためである。
 - (委員)そうすると、利用者はどのコードを使えばよいか、あるいはどのコードを使うことがより望ましいか、分からないのではないか。
 - (事務局)その通りである。
 - (委員)IPA で公開予定のコード・コード一覧で、可能であれば、コードが取りうる値を知ることができたらなおよい。
 - (事務局)それは次のステップであると考えている。
 - (委員)現状だと、コードの統一はかなり難しいのではないか。
 - (委員)コードの統一は、現状これだけのコードが発行され、利用されている中では現実的に難しく、まずは社会で使われているコードを収集・集約するとともに、今後はコードを新規に作成しないことと、使いたいコードとの紐づけ等は、例えば GIF サイトや IMI サイトに公表されているので、開発者等は必ず確認するように周知するのが次のステップ、その中で、使われていないコードは次第に淘汰していく、このような複数段の構えで行かないと難しいのではないかと考える。
 - (事務局)コードの統一が難しい一方で、現状、多くのコードが発行されていることそのものが明らかになることは、問題提起という点では意味があるとも考える。
 - (委員)他に、ユニーク性が担保されていることや、アップデートの方法（頻度）、再利用は可能か、といったコードのポリシーも知りたいところである。
 - (委員)統計センターで、例えば、現在のコード一覧の中から、これは比較的安心・安定して使用できるコードである、統計に使いやすいコードである、といった情報を付加し、「推

奨」という形で公表をしてもらえると、コードの統一やデータ品質の向上にもつながるのではないかと。コードごとに集計を取り、参照されたランキングを発表することも有効ではないかと。

- ・ 今回事務局にて説明した内容に沿って、IPA にて公開に向けた準備を引き続き進めていくことを、本委員会にて承認された。

■議題 4. 国際データ語彙マッピングのスキームと実施方法の確認

- ・ 事務局より、IPA が今回行う “Core Data Model Mapping project2” について、初回のマッピングから 10 年経過しており、現状を踏まえた再調査が必要であること、必須項目はシンプルな欧州のモデルに合わせる方針であること、また、日欧米のデータモデルを比較した際に、空白の項目や一致しない項目等がある場合、情報を正確に把握することができず、相互に投資をしようとする際の信用調査等にも影響が出ることから、単にマッピングを行うだけでなく、インタオペラビリティを確保したデータモデルの作成のためのガイドラインの整備も必要であることなど、資料 7 の内容に沿って説明。
- ・ 上記活動のアウトプットについては、日本とアメリカ、日本と欧州、それぞれのデータマッピング結果と、ガイドラインのドラフト版（ベータ版）の作成を 2025 年 3 月、そのマッピング結果に基づく個人、施設、土地の語彙マッピングと、ガイドラインの正式版の作成を 2026 年 3 月で進める予定。
- ・ 上記データマッピングプロジェクトについて、時間の都合から、本日は事務局からの説明のみとし、意見その他コメント等あれば、後日、事務局にメールにて連絡する。

■議題 5. 各種活動の状況について

- ・ 事務局より、IPA が行っている、データ環境の推進に向けた各種情報公開の状況、関係団体との連携状況、GIF の適用促進やデータの相互運用性を高めるための各種ツールの提供予定について、資料 8、9、10 の内容を交え説明。
- ・ これまでのデータ環境の推進に向けた IPA の取り組み、ならびに今後の取り組み予定について、本委員会にて承認された。

■クロージング

- ・ 参加者全員で、次回のデータ環境推進委員会は、2024 年 11 月～12 月頃の開催予定とすることを確認した。