

政府情報システムの移行のための データポータビリティ (ディスカッションペーパー)

2021 年 8 月

村田真¹、田丸健三郎²

要旨

古い政府情報システムから新システムに切り替えるには、古いシステムに蓄積された情報資産を新しいシステムに移行することがしばしば必要となる。政府情報システムの調達にあたって考慮すべきデータポータビリティを解説する。

本ディスカッションペーパーは、政府 CIO 補佐官等の有識者による検討内容を取りまとめたもので、論点整理、意見・市場動向の情報収集を通じて、オープンで活発な議論を喚起し、結果として議論の練度の向上を目的としています。そのため、ディスカッションペーパーの内容や意見は、掲載時期の検討内容であり、執筆者個人に属しており、内閣官房 情報通信技術 (IT) 総合戦略室、政府の公式見解を示すものではありません。

¹ 内閣官房政府 CIO 補佐官

² 内閣官房情報通信技術 (IT) 総合戦略室 プロジェクトマネージャー

目次

1	はじめに	1
1.1	背景と目的	1
1.2	適用対象	1
1.3	位置づけ	2
1.4	用語	2
2	データロックイン	2
3	データポータビリティとは	3
4	データポータビリティを実現するための要件	3
4.1	データの移行手段	3
1)	データの取だし手段があること	3
2)	現実的な時間で可能な移行手段であること	3
4.2	移行データの形式	4
1)	データの標準化	4
2)	文字の表現	4
3)	テキスト形式	5
4)	JSON、XML、CSVなどの採用	5
5)	スキーマ	5
6)	データフォーマット及びデータの権利	5
7)	文書化	6
4.3	完全性と安全性	6
	完全性	6
1)		6
2)	安全性	6
5	まとめ	6
付録 1:	政府・自治体情報システムのためのデータ標準化の取り組み	8
1)	行政データ連携	8
2)	中間標準レイアウト	8
3)	地域情報プラットフォーム標準仕様	8
4)	共通語彙基盤	8
5)	ベース・レジストリとしての住所・所在地マスターデータ	8
6)	海外の取り組み	8
	参照文献	10

1 はじめに

1.1 背景と目的

昨今、国内外の情勢や法律・制度の変化はますます速くなってきた。政府情報システムもこうした変化に柔軟に対応し、国民が必要とするサービスをすみやかに提供しなければならない。しかし、変化に柔軟に対応することは決して簡単ではない。政府情報システムが提供すべきサービスが何かは時に漠然としているし、いま必要なものが明日も必要とは限らない。

IT 技術の発展は目覚ましく、クラウド、モバイル、ビッグデータ、IoT(Internet of Things)、AI・機械学習などはこれまで不可能であったことを可能にしている。国民へのサービスのために必要なら今までのシステムを捨て、こうした IT 新技術を用いた情報システムに移行することも検討すべきである。そうした移行の例として、PaaS 又は SaaS などのクラウドサービスを利用して柔軟なスケラビリティコントロールや運用を可能にする方法、マイクロサービスやコンテナ技術などを利用してアプリケーションの展開と改良を容易にする方法、モバイル環境を活用して国民への利便性を高める方法などがある。移行には、オンプレミスシステムからクラウドベースへのシステムの移行だけではなく、クラウドベースの情報システムから別のクラウドベースの情報システムへの移行も含まれる。

しかし、クラウドサービスを含む新しい情報システムの実装はより簡単になったが、以前の情報システムから新しい情報システムへの移行は容易ではない。今日、移行を阻害する要因の一つとしてデータロックインがある。本文書では、データロックインとは何であるかを説明し、データロックインへの対策としてデータポータビリティの考え方を解説する。政府情報システムの調達にあたっては、要件としてデータポータビリティを含めることが求められる。

このディスカッションペーパーは、調達側にデータポータビリティの重要性を喚起することを目標とする。ベンダーがデータ移行をどのように行うかの詳細は扱わない。また、データポータビリティについて発注仕様書や契約書に何を盛り込むべきかについては今後の検討課題とする。

なお、データポータビリティは、パーソナルデータの主体である本人の意向を踏まえたデータ流通・活用を推進するために議論されることが多い [1]。しかし、本文書では政府情報システムのベンダーロックインへの対策としてのデータポータビリティを示す。

1.2 適用対象

政府情報システムであって、そこに情報資産が蓄積されるものすべてに適用できる。

1.3 位置づけ

本文書は、標準ガイドライン群の一つとして位置づけられる。

1.4 用語

本文書において使用する用語は、本文書に別段の定めがある場合を除くほか、標準ガイドライン群用語集の例による。

2 データロックイン

新しい政府情報システムへの移行を妨げるのがベンダーへのロックインである。これは、あるシステムを開発したベンダーが、その保守拡張だけではなく将来の作り直しにおいても圧倒的に有利になることを意味する。ロックインが起きれば、既存システムの開発ベンダーが今後の開発業務を独占する。ベンダー間の競争が起こらないのでコストは下がらないし、新たなユーザ要求に対処することも難しくなる。

これまでロックインを引き起こすのは OS であった。すなわち、ある OS 上で動作するシステムとして作成されたものを異なる OS 上で動作するように作り直すことは困難であった。しかし、今日のネットワーク環境においては、システムに関わるすべてのコンピュータが同じ OS で動く必要はない。ユーザはクラウドの OS を意識しないし、マルチクラウドならクラウドごとに OS が別であっても何ら問題ない。コンテナ技術を用いてシステムを別の OS 上に配備することもできる。つまり、OS はもはやロックインを引き起こさなくなった。同様に、プログラミング言語もロックインをもはや引き起こさない。OS やプログラミング言語などの技術は常に進化しており、新しい政府情報システムを開発するときには最適なものを選べばよい。

今日問題になるのはデータロックインである。データロックインとは、既存のシステムに蓄積されたデータが新しいシステムに移行できない状態を意味する。データロックインは、オンプレミスでも存在したが、クラウドでも存在する。

データの移行が難しい理由を説明する。今日の政府情報システムのデータは複雑になることが多い。そうした複雑なデータは、同じ情報を表現する方法が毎回少しずつ違っていたり、仕様が更新されず実情とずれていたりする。さらに、データの在処や仕様が分からない場合もある。そのような状態に陥ってしまえば、データ移行は困難になる。また、データフォーマットそのものが契約により非公開とされ、開示に別途コストを要する場合もある。

3 データポータビリティとは

本文書でいうデータポータビリティは、政府情報システムに蓄積されたデータが、新しい政府情報システムに完全かつ実用的な形で移行する手段及び必要な権利が提供されていることである。データポータビリティはデータロックインが起こらないことを保証する。システム更改が想定される政府情報システムの調達にあたっては、データポータビリティを非機能要件として含めることが求められる。前もって準備すれば、データ移行コストを大きく削減できる可能性がある。

データポータビリティは、政府情報システムをその開発ベンダーが拡張することを妨げるものでもなく、現在のベンダーが新しい政府情報システムを開発することを妨げるものでもない。したがって、データポータビリティとして提供される移行手段は使われないこともある。使われない可能性がある手段を用意することは無駄に思われるかもしれないがデータロックインを避けるためには必要なことである。

データポータビリティは、オンプレミスとクラウドのどちらにおいても必須の要件である。データが取り出せることは政府・民間によるデータ活用にも役立つ。

4 データポータビリティを実現するための要件

4.1 データの移行手段

1) データの取出し手段があること

情報システムから移行用データを取り出す手段が提供されていること。ファイルとしての取出し、データ取出しのためのプロトコル、データベースの提供するエクスポート機能などが考えられる。

2) 現実的な時間で可能な移行手段であること

大量のデータを取り出す場合であっても、現実的な時間で行えることを確認する必要がある。

4.2 移行データの形式

1) データの標準化

多くの政府情報システムで使われるデータなら、その形式を標準化することが考えられる。簡単に言えば、同じ情報を表すデータの形式を一つに決めることが標準化である。移行元のシステムと移行先のシステムが標準のデータ形式をサポートすれば、原理的にデータ移行の問題は起きないはずである。

国内における政府情報システムまたは自治体情報システムのためのデータ標準化の取組みの一覧を付録に示す。これらの標準化されたデータ形式を採用することが望ましい。しかし、政府情報システムで扱うデータのすべてが標準化可能なわけではなく、独自のデータ形式を使わざるを得ないことは当然考えられる。

2) 文字の表現

移行用データにおける文字の表現については、文字環境導入実践ガイドブック [2]に従う。さらに、IVD/IVS [3]を併せて利用すれば外字を回避することがさらに容易になる。ベース・レジストリに文字情報基盤が含まれているので、その所管部門である文字情報技術促進協議会³から IVS/IVD については詳細な情報が提供されている。

ISO/IEC 10646 の私用領域⁴に属する文字が移行用データに存在してはならない。これは、移行先システムにおいて私用領域の文字をこれまでと同様に表示・印刷することが困難だからである。ただし、業界で認められた中立的な団体（一般社団法人文字情報技術促進協議会など）の取決めに従って私用領域を用いる場合はこの限りではない。

政府情報システムが内部で保持する文字と、移行用データに現れる文字とはできるだけそろえることが必要である。政府情報システム内部では文字を非標準的な形式で保有しておいて移行時に標準的なテキストデータに変換するという方法もあるが、この方法ではデータポータビリティを保証したとは言えないので注意が必要である。実際に、テキストデータの自動変換が難しいことは文字情報基盤関連事業でも実証されている。

³ 一般社団法人文字情報技術促進協議会 <https://moji.or.jp>

⁴ 私用領域とは、UCS 基本多言語面の U+E000 ~ U+F8FF 並びに第 15 面及び第 16 面 (U+F0000 ~ U+10FFFF) を指す。

3) テキスト形式

標準化された符号化方式⁵によるテキスト形式であることが望ましく、バイナリ形式は好ましくない。これは、移行用データを読み込めないというトラブルがあったとき、テキスト形式のほうが対処しやすいからである。

4) JSON、XML、CSVなどの採用

JSON、XML、CSV⁶などの標準化されている形式であること。特定のベンダーや特定のソフトウェアでしか扱えない独自形式であってはならない。

5) スキーマ

移行用データが満たすべき制約を表現したものをスキーマという。JSON 用のスキーマならどのようなキー名を用いるかなどを表現し、XML 用のスキーマならどのようなタグ名・属性名を用いるかなどを表現する。スキーマを記述するための言語として JSON Schema、W3C XML Schema、RELAX NG、CSV Schema などがある。

スキーマがあれば、移行用データがそれに従っているかどうかをプログラムによって検証できる。スキーマは、移行先クラウドの開発者が移行用データを理解するためにも役立つ。

実際には、スキーマを作成せずに文書化だけに頼ることも多い。これはスキーマを運用するコストが高く、利点を帳消しにしてしまう場合に行われる。しかし、文書化だけに頼ると移行用データのスキーマによる検証ができなくなるという問題点があるので、移行用データ検証用のプログラムを用意するなどの対策が必要になる。

6) データフォーマット及びデータの権利

データフォーマット及びデータの権利がベンダーに属している場合は、移行用データを移行先のシステムに取り込むことができない。移行用データを妨げなく利用する権利を政府が保持していることが必要である。特に、あらかじめ開発されたパッケージ製品を用いる場合には、データフォーマット及びデータの権利について発注時点で明確にしておく必要がある。

⁵ UTF-8 及び UTF-16 など。

⁶ CSV は、普及した後に IETF RFC 4180 によって後付けで標準化されたので、実装と IETF RFC 4180 の間に差異があることが多い。CSV を利用する場合は、典型的な差異について明確化しておくこと（例えば文字符号化方式、引用符の扱い、ヘッダの有無など）。

7) 文書化

移行用データの所在や形式について、文書で詳しく記述することが必要である。この文書は、情報システムの業務要件を理解しているユーザなら理解できるように書くこと。この文書は、新規開発のときに一度書き上げればよいというものではなく、システムの運用中に継続して改定する必要がある。

4.3 完全性と安全性

1) 完全性

情報システムに蓄積された情報資産を移行用データがもれなく表現していることを完全性と呼ぶ。データポータビリティには、移行用データの完全性を客観的に確認するための方法が提示されていること、その方法に従って実際に確認がとれていることが必要である。

2) 安全性

情報システムに蓄積された情報資産が移行完了後に、確実に復元不可能な形で消去できることが必要である。物理ディスクをユーザが管理できるオンプレミスシステムの場合は、物理ディスクの消去や破壊も可能である [4]。クラウドベースの場合は別の方法をとる必要がある [5]。

5 まとめ

これから調達する政府情報システムも、いずれは時代に合わなくなり更改の対象となる。システム更新の際にシステムに蓄積されたデータが別のシステムに移行できなければ（すなわちデータがロックインされれば）システムの刷新は困難になる。

本文書では、データロックインへの対策としてデータポータビリティを解説した。データポータビリティとは、システムに蓄積されたすべてのデータが新しいシステムに完全かつ実用的な形で移行する手段が提供されていることである。データポータビリティを実現するための要件として、データの移行手段、移行データの形式、データの完全性と安全性（移行済みデータが復元不可能な形で消去されていること）を整理した。

情報資産が蓄積される政府情報システムを調達するときは、調達仕様書で要件としてデータポータビリティに言及されたい。移行すべきデータを列挙できればさらによい。最初からシステム移行を意識し、データポータビリティを準備しておくことによってデータロックインを避けることができる。

付録1: 政府・自治体情報システムのためのデータ標準化の取り組み

1) 行政データ連携

内閣官房情報通信技術（IT）総合戦略室はデジタル・ガバメント推進標準ガイドラインをまとめている。そのうち、行政基本情報データ連携モデル [6] と行政サービス・データ連携モデル（β版） [7] がデータ標準化である。

2) 中間標準レイアウト

中間標準レイアウト [8] は、総務省の「自治体クラウドの円滑なデータ移行等に関する研究会」（平成 23 年度）がまとめ、その後は地方公共団体情報システム機構（J-LIS）が維持管理している。

3) 地域情報プラットフォーム標準仕様

地域情報プラットフォーム標準仕様 [9] は、総務省事業として策定され、（一財）全国地域情報化推進協会（APPLIC）が公開し運用している。

4) 共通語彙基盤

情報処理推進機構（IPA）の共通語彙基盤 [10] は、データで用いる様々な用語の表記、意味、構造を統一し、分野を超えてデータの検索向上やシステム連携強化を実現するための標準である。

5) ベース・レジストリとしての住所・所在地マスターデータ

住所・所在地の情報をベース・レジストリとして整備・更新するために考慮・検討すべき事項を整理したディスカッションペーパー [11] が政府 CIO ポータルに掲載されている。

6) 海外の取り組み

海外においても、政府・自治体情報システムのためのデータ標準化の試みは数多く存在する。ここでは住所のデータ標準化に絞って取り上げる。

アメリカ合衆国郵便公社は、住所を文字列として表現するための標準 [12]

を定めている。文字列として与えられた住所を正規化するためのソフトウェアも開発されている。

住所を XML で表現する仕様も数多く存在する。米国連邦地理データ委員会が 2011 年に制定した United States Thoroughfare、Landmark、and Postal Address Data Standard[13]はその一つである。住所を XML で表現するさまざまな仕様の一覧[14]もある。

参照文献

- [1] JETRO、“EU 一般データ保護規則（GDPR）について” [オンライン].
<https://www.jetro.go.jp/world/europe/eu/gdpr/>から入手可能。閲覧日：2021 年 6 月 10 日
- [2] 政府 CIO ポータル、“文字環境導入実践ガイドブック” [オンライン].
<https://cio.go.jp/guides> から入手可能。閲覧日：2021 年 6 月 10 日
- [3] 文字情報技術促進協議会、“IVD/IVS とは” [オンライン].
<https://moji.or.jp/mojikiban/aboutivs/>から入手可能。閲覧日：2021 年 6 月 10 日
- [4] 一般社団法人コンピュータソフトウェア協会、“データ消去証明ガイドブック” [オンライン].
https://www.csaj.jp/documents/NEWS/committee/datadelete/datadelete_guidebook_20170614.pdfから入手可能。閲覧日：2021 年 6 月 10 日
- [5] Amazon Web Service ブログ、“クラウドにおける安全なデータの廃棄、” [オンライン].
https://aws.amazon.com/jp/blogs/news/data_disposal/から入手可能。閲覧日：2021 年 6 月 10 日
- [6] 政府 CIO ポータル、“行政基本情報データ連携モデル” [オンライン].
<https://cio.go.jp/guides> から入手可能。閲覧日：2021 年 6 月 10 日
- [7] 政府 CIO ポータル、“行政サービス・データ連携モデル” [オンライン].
<https://cio.go.jp/guides> から入手可能。閲覧日：2021 年 6 月 10 日
- [8] 総務省情報流通行政局、“中間標準レイアウト仕様” [オンライン].
https://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/lg-cloud/02kiban07_03000024.htmlから入手可能。閲覧日：2021 年 6 月 10 日
- [9] 総務省情報流通行政局、“地域情報プラットフォーム標準仕様について” [オンライン]. https://www.soumu.go.jp/main_content/000579848.pdf 閲覧日：2021 年 6 月 10 日
- [10] IPA、“IMI とは” [オンライン]. <https://imi.go.jp/imi/> から入手可能。閲覧日：2021 年 6 月 10 日
- [11] 政府 CIO ポータル、“ベース・レジストリとしての住所・所在地マスターデータ整備について” [オンライン]. https://cio.go.jp/dp2021_03 から入手可能。閲覧日：2021 年 6 月 10 日
- [12] アメリカ合衆国郵便公社、“Postal Addressing Standards” [オンライン].

<https://pe.usps.com/text/pub28/welcome.htm> 閲覧日：2021年6月10日

- [13] 米国連邦地理データ委員会、“United States Thoroughfare、Landmark、and Postal Address Data Standard”、February 2011 [オンライン].
<https://www.fgdc.gov/standards/projects/address-data> 閲覧日：2021年6月10日
- [14] The Cover Pages、“Markup Languages for Names and Addresses” [オンライン].
<http://xml.coverpages.org/namesAndAddresses.html> 閲覧日：2021年6月10日