

データの相互運用性向上のためのガイド(概要版)

2022年8月8日 初版

独立行政法人情報処理推進機構(法人番号 5010005007126)




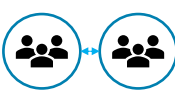
本ガイドについて

ガイドの目的 ～データ活用社会の実現へ向けて

私たちの暮らしは、データと切り離すことはできない。SNSの投稿、店頭やネット上の決済記録、デバイスやウェブサーバー等に蓄積される行動ログ等のパーソナルデータ、事業活動の証跡として業務システム等に記録される会計や製造データ、また建物や道路、車やドローン、センサー等を介して製品や設備等から収集される環境や位置情報、各種オープンデータなど、枚挙にいとまがない。

私たちが営む社会・経済活動の多くは情報システム等を通じてデータとして記録され、新たな価値創造へ直接的あるいは間接的に活用される。そうしたデータ駆動型のデジタル社会では、それぞれの組織が運用する情報システムを超えて、全ての参加者からそのデータに対する共通理解を得られなければならない。例えば「融資」と「貸付」は異なる用語だが、「災害被害者支援」という業務目的の文脈で公開される「復旧・復興支援制度情報サイト」においては、同等の意味をもつ概念として相互につながり合う必要がある。資金面での災害被害者支援において「融資」「貸付」と表現が異なっても、誰もがその意味を共通理解できる状態が、本ガイドでいう相互運用性(Interoperability)¹である。データ活用の現実解として本ガイドが主にとりあげる相互運用性向上の取組みは、前述の災害被害者支援のように同じ文脈で使われる「融資」と「貸付」を意味としてつなぐ手順や表現方法である。

とくに相互運用性が求められる

	連携なし	既知の2者間の連携	業界内の連携	分野横断連携
				
データの例	<ul style="list-style-type: none">営業部門内の顧客リスト表紙の家計簿	<ul style="list-style-type: none">社内顧客リストと連携した請求システム上のデータ	<ul style="list-style-type: none">プロスポーツ共通の顧客管理システム連携ECのデータ	<ul style="list-style-type: none">スマートシティの交通・エネルギー・気象・行動などのデータ
協調の枠組み	不要	当事者間の合意が必要	業界内のルールとして合意形成が必要	社会全体のルールとして合意形成が必要
解決手段の例	適用業務のなかでデータの意味づけを表現する。	相互に理解できるようにデータの意味づけを表現し共有する。	参加者全員が理解できるようにデータの意味づけを表現し共有する。	参加者全員が理解できるようにデータの意味づけを表現し共有する。

独立行政法人情報処理推進機構（IPA）は、データの活用社会を推進することを目的として、データの「相互運用性」に焦点を当て、「データの相互運用性向上のためのガイド」を発行することとした。本紙は、その概要版である。なお本編では、データ連携に向けた取組みなどを実際の事例を用いて解説している。

¹ JIS X 0001 は「それぞれの機能単位に固有な特性に関する知識を利用者がほとんど又は全く必要とせず、各機能単位が互いに通信し、プログラムを実行し又はデータを転送する能力」と定義している。また、ISO/IEC 19941:2017 の定義は「Ability of two or more systems to exchange information and mutually use the information that has been exchanged.」である。

ガイドの想定読者

本ガイドの主たる想定読者は産官学民いずれの立場においても、データ活用に携わるあらゆるステークホルダーである。特に CD0(Chief Data Officer)あるいは CD0(Chief Digital Officer)、CIO(Chief Information Officer)などの役職者の方に読んでいただくことを想定している。官民データ活用推進基本法に沿い相互運用性の高いオープンデータの作成・公開を求められる自治体と公共機関に所属されている方やシビックテック²活動を展開する市民も想定読者である。

データの相互運用性の必要性

政策における相互運用性の位置付け

データ利活用及びその手段としての相互運用性向上は、官民データ活用推進基本法及び Society 5.0 をはじめとする政策に基づいている。Society 5.0 の実現は、データ活用が前提である。特にフィジカル(現実)空間である製造、農業、医療等の現場で生成されるデータの活用は、サイバー(仮想)空間でのデータ活用に遅れる日本が、巻き返し可能な領域として期待されている。

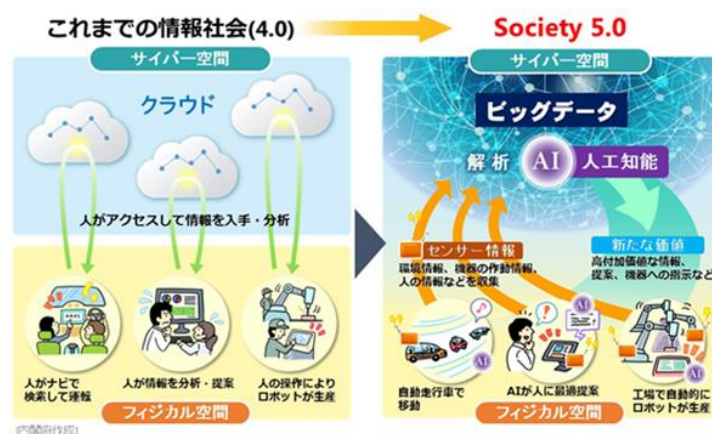


図 サイバー空間とフィジカル空間を高度に融合させる Society 5.0 のイメージ

(出典)内閣府 Society 5.0 https://www8.cao.go.jp/cstp/society5_0/

海外と比較すると、日本における取組みはまだ緒についたばかりだが、官民の協力と連携による国全体の DX 推進は喫緊の課題であり、政府もデジタル・ガバメントの実現を急いでいる。データの持続的な相互運用性向上は、官民データ活用社会を支えるあらゆるステークホルダーの参加と協力によって実現する。その恩恵は、個々の参加者そして社会全体へ還元されていく。

² 市民自身がテクノロジーを活用して、行政サービスの問題や社会課題を解決する取組み

相互運用性向上に向けたアプローチ

データ項目やデータ構造の整理

(1) データ項目の確認

ある企業における「仕入先」というデータセットに含まれるデータ項目には、次の図のような情報が含まれているとする。

仕入先
ID
仕入先情報 1
仕入先情報 2
仕入先住所 1
仕入先住所 2
使用通貨
支払条件
担当部署
担当者情報 1
担当者情報 2

図 仕入先に関するデータ項目の例

(2) データ項目の形式、付加情報の整備

上述した仕入先データを扱う際には、仕入先 1 件の情報を 1 レコードという単位で扱い、そのレコードの集合が仕入先のデータセットとなる。データ資産をデータベース等に入出力する際は、CSV、JSON、XML などの形式のファイルが用いられる。次の図は仕入先 1 件分(1 レコード)のデータをカンマ区切りで表した CSV 形式の例である。

```
01234,5010005007126,独立行政法人情報処理推進機構,東京都,文京区本駒込 2-28-8,JPY,A001,総務部,情報,朝陽
```

図 仕入先データの例

各データ項目に対して値の形式や意味を、次の図のように付加情報で定義する。人やコンピューターがデータを活用する際の情報源、つまりデータ項目定義書の基本的な整備方法である。

ID	仕入先を識別する 5 桁の整数(00000~99999)
仕入先情報 1	法人番号(13 桁の数字)
仕入先情報 2	社名(40 文字以内)
仕入先住所 1	都道府県
仕入先住所 2	市区町村名、番地等(40 文字以内)
使用通貨	通貨の略号(3 文字)
支払条件	当社購買部が定めた記号 (4 桁の英数字)
担当部署	担当者が所属する部署名(40 文字以内)
担当者情報 1	担当者の姓(20 文字以内)
担当者情報 2	担当者の名(20 文字以内)

図 基本的なデータ項目定義書

(3) 意味構造の整理

データ項目定義に定める型や桁数、単位に合致した「文法的に正しいデータ」であることは、データ資産に求められる最低限の要件である。「正しい」ためには、人とコンピューターがデータを活用しやすいように意味構造を定義する必要がある。意味構造を定義するためには、データ項目の分類と構造化が有効である。

次の図に先の仕入先情報の定義を記した。この定義の説明部分を読むことで「当社」独自の管理項目(赤字)、「仕入先」そのものに関わる項目(青字)、「仕入先」の「担当者」に関するサブ項目(緑字)に分類できる。

#	項目名	必須	説明	形式	記入例
1	ID	○	当社 の購買部門内で仕入先が一意になるように設定。	文字列 (半角数字 5 桁)	01234
2	仕入先情報 1	○	仕入先 の法人番号(13 桁の数字)。 仕入先が国税庁から指定された法人番号	文字列 (半角数字 13 桁)	5010005007126
3	仕入先情報 2	○	仕入先 の法人名。 gBizINFO ³ 掲載の法人名。	文字列 (40 文字以内)	独立行政法人情報 処理推進機構
4	仕入先住所 1	○	仕入先 の本店所在地の都道府県 都道府県を含んで入力(gBizINFO)	文字列 (4 文字以内)	東京都
5	仕入先住所 2	○	仕入先 の本店所在地の市区町村名、番地等 (gBizINFO)	文字列 (20 文字以内)	文京区本駒込 2- 28-8
6	使用通貨		仕入先 の使用する通貨コード 通貨コードは ISO 4217 にて確認	文字列 (半角英数字 3 桁)	JPY
7	支払条件	○	当社 の購買部門が管理する支払条件コード。 条件コードは当社購買部が定める記号	文字列 (半角英数字 4 桁)	A001
8	担当部署	○	仕入先 の 担当者 が所属する部署名	文字列 (20 文字以内)	総務部
9	担当者情報 1		仕入先 の 担当者 の姓	文字列 (20 文字以内)	情報
10	担当者情報 2		仕入先 の 担当者 の名	文字列 (20 文字以内)	朝陽

図 意味構造を加味したデータ項目定義書

³ 府省庁が保有する約 400 万団体の法人情報を一括検索しデータを取得できるデータベース。 <https://info.gbiz.go.jp>

前の図の分類をツリー状に展開すると、各データ項目の意味構造が理解しやすい。

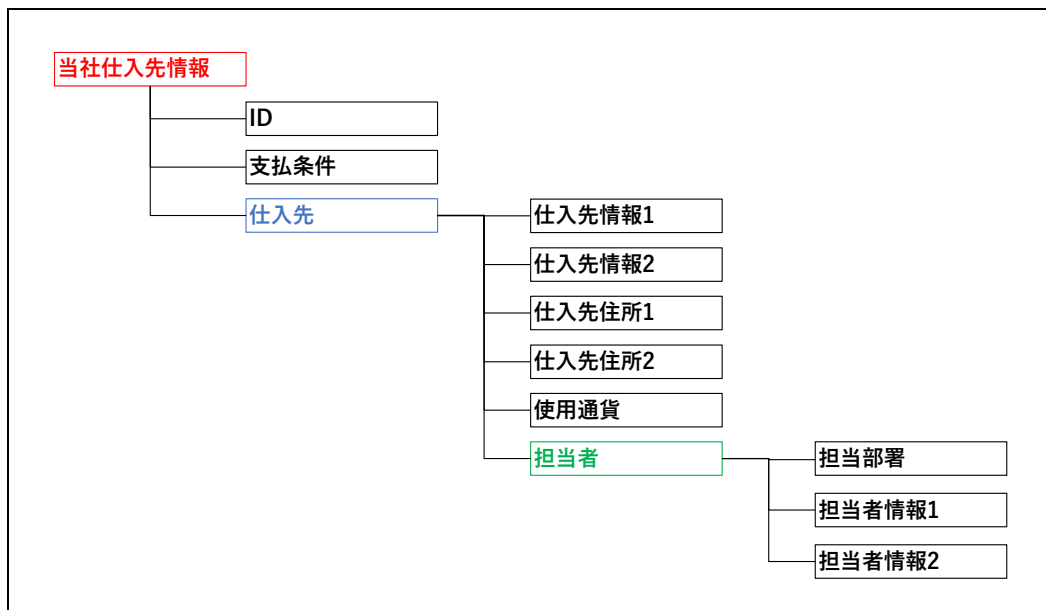


図 分類構造の展開

データ項目定義には、詳細な説明書きや入力形式の規定、あるいは入力例、そのデータ項目定義の典拠となる情報を記載することが多い。「仕入先情報1」は国税庁指定の法人番号を参照しており、「使用通貨」は通貨コードの国際規格である ISO 4217 を典拠とした分類である。

「データ相互運用性」の取組みとは、こうした意味構造や定義、典拠等を利用者やコンピューターへ正しく伝える努力のことである。伝える手段はデータセットと別の定義書の作成、その他、データ構造定義を利用した電子ファイルへの同梱、ウェブサイト上のデータ項目定義へのリンク、決まった相手とのデータ連携ならば社内システム上のデータセット同様ドキュメントの事前共有等、用途に応じて様々な手段がある⁴。いずれにせよ、データがもつ背景や正確性の判断材料をステークホルダーとコンピューターが共有できる工夫が、データを利活用するためには欠かせない取組みである。

⁴ 例えば、人間の閲覧を想定した統計表であれば、表中の縦横のラベルから属性、単位など主要な値の意味を読み取れるが、詳細な分析に必要なそもそもの調査対象や調査手法は別途「調査概要」「調査手法」などのテキストで補足するという方法が、広く実践されている。

データ連携の手法

データ連携モデル

相互運用性を高めるデータ連携モデルとして、ここでは3つを取り上げる。ポイント・トゥ・ポイント(Point to Point)モデル、ハブ&スポークモデル、協調モデルである。それぞれ異なる特徴を有するため、設計や実装においては目的と照らして選択、または複数を適切に組み合わせる。

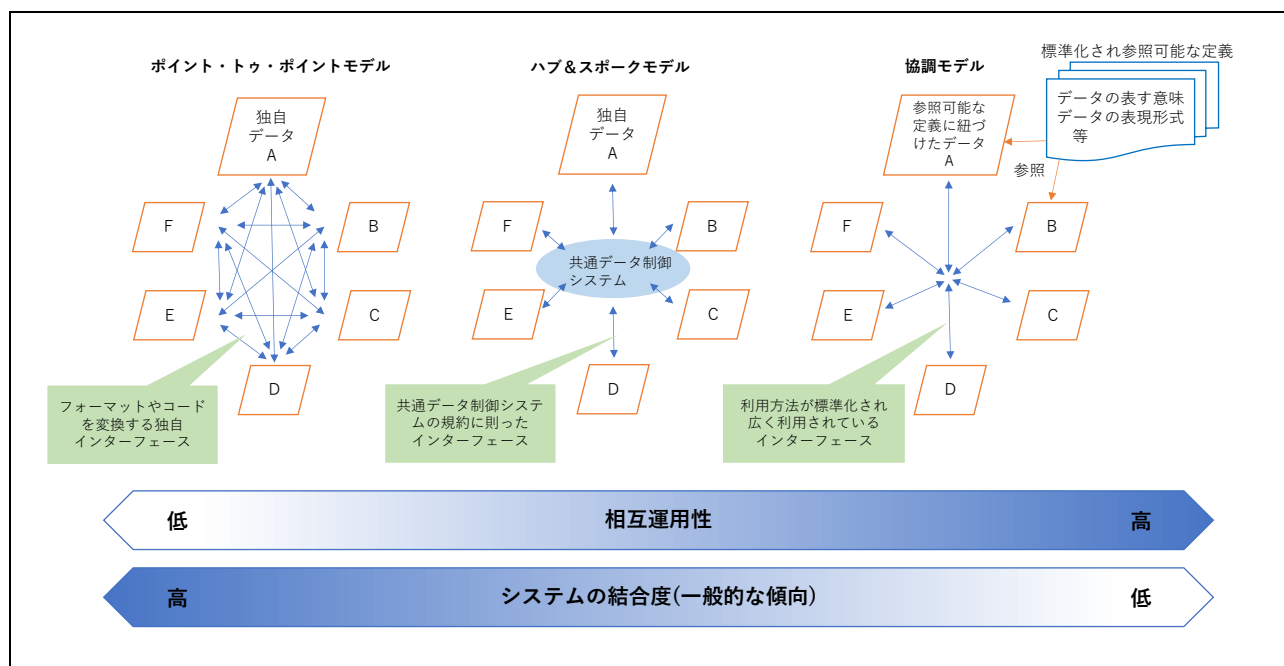


図 データ相互運用性の考え方

(1) ポイント・トゥ・ポイント (Point to Point) モデル

「Point to Point」は、異なる形式のデータを相手先と1対1でやり取りする手法を示す。相互のデータを変換するためには、個別にインターフェースを開発・運用する。各システムの詳細仕様の把握が必要になることが多いため、一般にシステムの結合度は高い。

(2) ハブ&スポークモデル

相互運用性の向上を図るために、共通データ制御システムを仲介させる。各システムは、各種データに用いられる多様な文字情報やフォーマットの変換などを処理するハブシステムが公開するインターフェースとのみやり取りする。システム間でのファイル連携方法のみ共有したり、インターフェース仕様を公開したりするだけよく、システム結合度は低い。

(3) 協調モデル

相互運用の実現にあたって、関係者間であらかじめ合意された標準的な規約や開示された定義情報を参照してデータを作成・変換し、公開・授受することで社会的コストの抑制を図る手法である。データを受け渡しする際に必ず、どのような標準を参照・利用したか（独自の場合はどのような決まり事でデータを記入しているのか）という定義情報を付記した上で、データを公開し、互いにデータを受け渡す。推奨データセット⁵や中間標準レイアウト⁶はその例である。

協調モデルでは、ハブ&スポークモデルのようにデータの受け渡しを仲介する制御システムは基本的に存在しない。既存ファイルを標準フォーマットに変換する仕組み（コンバーターなど）は各社がそれぞれ都合のよい方法で用意すればよい。システムの結合度は低い。

最後に

データ活用やそのための相互運用性向上は一過性の取組みではなく、維持管理を前提とした継続的な営みである。社会環境や経営環境、またデータの処理技術に変化が生じてもデータの相互運用性が確保され続けることが重要である。

データ資産の運用管理は、あくまでビジネス、業務を起点に企画、推進することが重要である。個別最適を避けるため、業務横断的なデータ資産管理部門を設置することが望ましい。実際に経営層と連携して部署横断的に活動する権限をもつ統括的なCD0(Chief Data Officer)の設置も一部で進んでいる。また、組織全体そして社会の人材育成も必須である。長期的に見たデータ活用人材育成への戦略的な投資が不可欠である。

データ活用社会の推進は、データの相互運用性を高めることで社会的損失の発生を回避しながら各組織と社会全体の創造力、生産性を向上していく協調的な取組みである。そのため、組織内に閉じない広範囲なステークホルダーを巻き込んだ一段次元の高い視点が求められる。公共機関や地方自治体はもちろんのこと、民間事業者もデータの相互運用性を高めるための社会貢献が自らの持続的な成長をもたらすことを認識し、データ活用社会の実現に向けた積極的な活動を期待したい。

⁵ 政府 CIO ポータル オープンデータ。 <https://cio.go.jp/policy-opendata>

⁶ 総務省 中間標準レイアウト仕様。 https://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/lg-cloud/02kiban07_03000024.html

この文書について

■ 表題

- ・ データの相互運用性向上のためのガイド(概要版)

■ 公開履歴

- ・ 初版 2022年8月8日

■ 監修 (50音順,所属は公開時のもの)

- ・ 齊藤 浩 独立行政法人情報処理推進機構
- 萩原 正規 独立行政法人情報処理推進機構
- 堀越 秀朗 独立行政法人情報処理推進機構
- 森貞 夏樹 独立行政法人情報処理推進機構
- 我妻 浩子 独立行政法人情報処理推進機構

■ 編集・発行

独立行政法人情報処理推進機構(IPA) (法人番号 5010005007126)

©Information-technology Promotion Agency, Japan (IPA)

東京都文京区本駒込 2-28-8 文京グリーンコートセンターオフィス

<https://www.ipa.go.jp/>

この文書のご利用にあたって

本書の内容を適用した結果生じたこと、また適用できなかった結果について、IPAは一切の責任を負いませんのでご了承ください。