

# データの相互運用性向上のためのガイド

## データ活用社会の実現へ向けた取組みの手順と実例

2021年3月25日 第二版

DOI : <https://doi.org/10.60430/digital.guidebook0010>

独立行政法人情報処理推進機構(法人番号 5010005007126)

# 目次

本ガイドについて .....	1
0.1.1 ガイドの目的 ～データ活用社会の実現へ向けて.....	1
0.1.2 本ガイドの構成 .....	1
0.1.3 ガイドの想定読者 .....	2
<b>第1章 データの相互運用性の必要性.....</b>	<b>3</b>
1.1 データの相互運用性とは .....	3
1.1.1 データの共通理解を実現 .....	3
1.1.2 相互運用を実現するデータセットの共通モデル.....	4
1.1.3 スマートシティとデータ連携基盤 .....	5
1.1.4 行政サービスを高度化する相互運用性の高いデータ.....	6
1.2 データの相互運用性の必要性 .....	8
1.2.1 データ活用社会の参加者として .....	8
1.2.2 2025年もうひとつの崖 データ資産の不活用 .....	8
1.2.3 政策における相互運用性の位置付け .....	9
<b>第2章 相互運用性向上に向けたアプローチ.....</b>	<b>12</b>
2.1 本ガイドにおける定義 .....	12
2.2 データの価値向上と効果的な連携のための整備.....	14
2.2.1 データ項目やデータ構造の整理 .....	14
2.2.2 データカタログ（データ管理台帳）の標準化 .....	17
2.2.3 ID・コードの標準化 .....	19
2.2.4 文字情報の標準化 .....	19
2.2.5 運用ルールと体制の整備 .....	20
2.3 データ連携の手法 .....	22
2.3.1 データ連携モデル .....	22
2.3.2 データ連携の実装 .....	24
2.3.3 共通フォーマットの活用 .....	25
2.4 技術標準／国際標準の活用 .....	28
2.4.1 本書における技術標準／国際標準の分類方法 .....	28
2.4.2 普及度を考慮した技術標準／国際標準の選択 .....	31
2.4.3 標準化活動との関係づけ .....	31
2.4.4 合意形成と運用 .....	35
2.5 データ連携の準備を行うための具体的な手順.....	37
2.5.1 語彙・ボキャブラリの検討 .....	37

2.5.2 表記・記法の決定 .....	37
2.5.3 データ構造の決定 .....	38
2.5.4 符号化文字集合（文字コード）とエンコーディング方式.....	39
<b>第3章 データ連携に向けた取組み例 .....</b>	<b>41</b>
3.1 「やさいバス」事業概要 .....	41
3.2 「やさいバス」の業務及びデータ連携.....	42
3.3 事業成功のポイント .....	46
3.4 データの相互運用性向上に関する今後.....	46
<b>最後に .....</b>	<b>48</b>
<b>付録 データ表現に関する国際標準化活動.....</b>	<b>49</b>

# 本ガイドについて

---

## 0.1.1 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<sup>1</sup>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。

一般にデータは何らかし目的を達成するために生成されている。そのためデータの生成または取得、移動、変換及び格納、維持及び共有、利用または適用、廃棄の一連のライフサイクル<sup>2</sup>は特定の業務目的を叶える情報システムで処理されてきた。しかしデバイスや流通技術の進展を経たこんにちのデータは、システムから自由である。市場取引や API を通じてシステムや組織を超え組み合わせることで処理することにより、生成時の目的に捕われない新たな発見や価値創造が期待されている。データが現代の石油と呼ばれるゆえんである。

来る官民データ活用<sup>3</sup>社会では、システムや組織を超えてデータという資産が活用可能な状態—すなわち、検索でき、所在がわかり、アクセスできる状態である必要がある。データの再利用が容易になれば、新たなデータ生成という社会コストは節約される。データ活用が可能な状態とは、データ資産を適切に守ることができる状態と同義でもある。そのためにデータに求められる新たな要件が、本ガイドで焦点を当てるデータの「相互運用性<sup>4</sup>」である。利害関係者(ステークホルダー)が合意と適切なルールのもとでデータ資産を共有し再利用するエコシステムを実感していただくため、データの相互運用に関する先行事例もとりあげる。

官民データ活用推進基本法<sup>5</sup>が掲げる産業・雇用の創出、及び我が国の国際的な競争力の強化のために、データの相互運用性に対する理解が進み、データの活用社会を推進することが、本ガイドの目的である。

## 0.1.2 本ガイドの構成

### 第1章 データの相互運用性の必要性

---

<sup>1</sup> JIS X 0001 における定義は「情報の表現であって、伝達、解釈又は処理に適するように形式化され、再度情報として解釈できるもの」とされているが、本ガイドではとくに記載がない限り、電子的に記録された数値や文字で表される記号を「データ」と称する。

<sup>2</sup> データマネジメント知識体系ガイド第二版の定義に基づく。欧州一般データ保護規則では「取得、記録、編集、構造化、保存、変更、復旧、参照、利用、移転による開示、周知または周知可能にする行為、整列または結合、制限、消去、破壊」を処理と定義している。

<sup>3</sup> 「データ活用」に明確な定義はないようだが、本ガイドでは、「1.1 データの相互運用性とは」等で説明するように、事業活動で付加価値を得る目的で、データの分析・加工などの処理を伴う営みと考えている。

<sup>4</sup> 本ガイドでは、既存データセットやシステムのデータ構造や設計を変更せず、データの意味構造や定義、典拠等を利用者が理解し合いコンピューターで処理、活用できることを指している。詳細は「1.1 データの相互運用性とは」を参照。

<sup>5</sup> 「1.2.3 政策における相互運用性の位置付け」を参照。

相互運用性と相互運用を仲介するデータ項目定義について概説する。官民データ活用推進基本法やスマートシティを支える官民データ連携基盤を含む、相互運用性向上に関する政策の動向に触れる。

## 第2章 相互運用性向上に向けたアプローチ

データを利活用するための準備として必要になるデータ整備の考え方や、相互運用する際のデータ連携の仕組み、データ連携を行う際に自社や業界の技術標準や国際標準をどのように取り込むか、などの手順について紹介する。

## 第3章 データ連携に向けた取組み例

業務をデジタル化し関係者間のデータ連携を実現するために、どのように取組み始めるか、推進していくための工夫、事業拡大等に伴う技術標準の適用などを実際の事例を用いて解説する。

### 0.1.3 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についても、あらためて自らの取組みの意義をご理解いただけるのではないかと考えている<sup>6</sup>。社会課題の解決に向けたシビックテック<sup>7</sup>活動を展開する市民も想定読者である。

オープンデータ、データ流通市場で取引されるデータ、社内システム上のデータなど種別を問わず、データに価値を見出すのは、需要者、利用者である。本ガイドを通じて、データを取り巻くあらゆるステークホルダーが、相互運用性についての理解を深めていただくことを期待している。

---

<sup>6</sup> オープンデータの相互運用性に関しては、「IMI 共通語彙に対応したデータ公開手順」を併せて参照されたい。<https://imi.go.jp/go/datalifecycle/>

<sup>7</sup> 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

# 第1章 データの相互運用性の必要性

本章では、データの相互運用性の必要性について説明する。適切かつ円滑なデータの流通・利活用を支える社会基盤の構築・運用には、官民の協力・協調が欠かせない。

## 1.1 データの相互運用性とは

### 1.1.1 データの共通理解を実現

私たちが営む社会・経済活動は、連続した複数の業務上の情報の流れで成り立っている。その営み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)<sup>8</sup>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"> <li>営業部門内の顧客リスト表</li> <li>紙の家計簿</li> </ul>	<ul style="list-style-type: none"> <li>社内顧客リストと連携した請求システム上のデータ</li> </ul>	<ul style="list-style-type: none"> <li>プロスポーツ共通の顧客管理システム連携ECのデータ</li> </ul>	<ul style="list-style-type: none"> <li>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</li> </ul>
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。

図 1-1 データ連携の範囲例と相互運用性

コンピューターがデータの意味に対する共通認識をもつための最も単純な手法は、データ要件定義やフォーマットの「統一」あるいは「互換<sup>9</sup>」である。組織の特定システム内ではデータ要件定義がそれに当たる。あるいは入力制限のかかった同一のワークシートを全ての参加者に利用させることができれば共通認識が可能である。ただし、このような「フォーマット統一」の手段は、こんにちのように多様な人びと、デバイス、

<sup>8</sup> 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に、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。

<sup>9</sup> 乾電池やネジ、ボルトやナットなどの寸法、仕様は、メーカーにかかわらず同じである。このように規格を共通化する考え方は、互換性(Compatibility)といわれる。

処理環境をまたぐデータ活用には適さないことが多い。既存システムのデータ要件定義を変更したり、既にある「貸付」「貸出」「ローン」などのデータセットのラベルを「融資」に統一したりする作業とコストを、全ての参加者へ強制することは現実的ではないからである。

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。「2.5 データ連携の準備を行うための具体的な手順」で解説するように、用語のもつ意味を付加することによって両者をつなぐことができる。意味が同じことをコンピューターが判読できる仕組みがあればデータセットのラベルが不統一であっても問題はなく、各組織が従来用いてきたデータの設計自体を変える必要はない。

### 1.1.2 相互運用を実現するデータセットの共通モデル

異なる用語を意味としてつなぐためにデータに付加する情報のセットを、本ガイドではデータセットの共通モデル(以下、共通モデル)と総称する。実装には様々な手法や事例があるが、ここでは基本的な考え方を示す。

ある財務会計システムに、次のような形式のデータ(項目名と値)があったとする。

売上高	1,000,000,000
-----	---------------

項目名は「売上高」で、値が「1,000,000,000」という数字である。だが、これだけを見ても、値の意味を理解できない。どの企業のなんの製品/サービスの売上高なのか不明である。「売上高」なので金額であろうと推察はできるが、定かではない。金額だったとしても1,000,000,000「円」なのか、1,000,000,000「千円」なのか桁数がわからないし、1,000,000,000「ドル」かもしれない。また、いつの売上高なのかもわからない。

このようなデータを正しく理解するためには、データ項目定義書などのドキュメントを確かめる、またはデータや財務会計システムの関係者に確認する必要がある。

だが、値に次のような情報が付加されていると、どうだろうか。

会社	株式会社〇〇〇〇
年度	2018年度
製品	XYZ
売上高	1,000,000,000
単位	円

金額項目
IFRS(国際会計基準)の売上高

この場合には株式会社〇〇〇〇における製品 XYZ の 2018 年度の売上高 1,000,000,000 円(IFRS に準拠)であることが読み取れる。

共通モデルは、このように複数のデータ項目を含む構造によって説明されている。このようなデータ項目を含む構造によってある用語が定義されていれば、「売上高」が「Sales」であったり、他の外国語表記にな

ったりしていても、人が理解できるだけでなく、コンピューターもその中に含まれる値を同一のデータとして処理することができる。

そのデータのもつ意味や型(数値型か、文字列型か、あるいは「あり、なし」など)、桁数、単位などが定義されていなければ、有益な情報として他者がデータを事業や研究などに活用することは難しい。データの共通理解を実現するためには、値とその定義情報をセットで相手に伝える必要がある。

### 1.1.3 スマートシティとデータ連携基盤

Society 5.0 の先行的な実現の場として進む取組みが、スマートシティである<sup>10</sup>。データと先進的技術の活用により、地域の機能やサービスを効率化・高度化し、社会課題の解決を図り、快適性や利便性を含めた新たな価値を創出する<sup>11</sup>。また、EBPM(Evidence Based Policy Making<sup>12</sup>)の観点で官民そして市民等がデータを共有し、データに基づいて生活の品質向上や経済発展を図る都市像である。

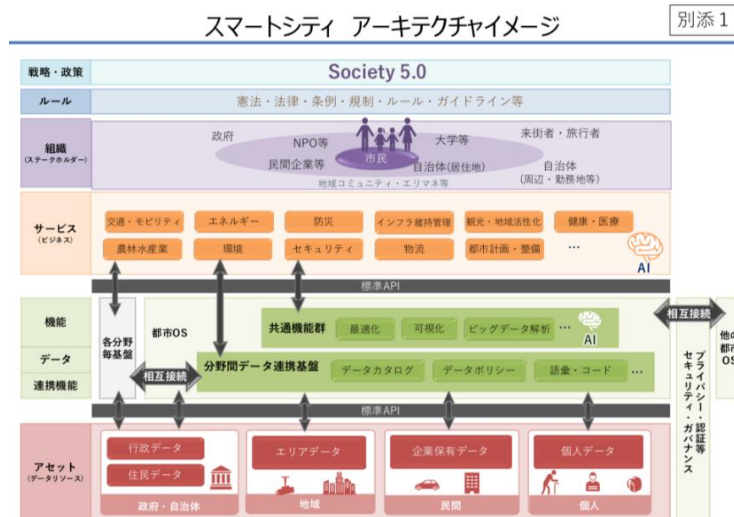


図 1-2 日本におけるスマートシティアーキテクチャのイメージ<sup>13</sup>

(出典)「戦略的イノベーション創造プログラム(SIP)第2期 ビッグデータ・AI を活用したサイバー空間基盤技術」におけるアーキテクチャ構築及び実証研究

スマートシティでの活用が想定されるのは、オープンデータや非識別加工匿名情報、スマートデバイスやIoT 機器などを通じて得られる交通、エネルギー、土地、建物、人流や医療・健康・教育といった行政サービス等に関わる官民データである。個別業務システムで生成されてきたこれらデータに相互運用性をもたせ高

<sup>10</sup> 例えば自治体及び企業・研究機関及び内閣府、総務省、経済産業省、国土交通省等を会員とする「スマートシティ官民連携プラットフォーム」が発足し、全国各地で取組みが進んでいる。<https://www.mlit.go.jp/scpf/>

<sup>11</sup> 物流や資源利用の最適化、防災、防犯、住民の生活品質向上や互助の促進、行政サービスの最適化と品質向上などが期待されている。

<sup>12</sup> 証拠に基づく政策立案。国や地方自治体が掲げる EBPM は、政策立案及び効果の測定に関連する情報や統計等のデータを活用し、政策の有効性を高め、国民の行政への信頼確保に資する。

<sup>13</sup> 2019 年時点のイメージ。内閣府実施、国立研究開発法人新エネルギー・産業技術総合開発機構が管理法人を務める「戦略的イノベーション創造プログラム(SIP)第2期/ビッグデータ・AI を活用したサイバー空間基盤技術におけるアーキテクチャ構築及び実証研究」にて、あらためて開発が予定されている。



度な処理を実現する都市 OS<sup>14</sup>の中核基盤として、データセット、語彙・コード、データカタログなどから構成される「分野間データ連携基盤」の構築が進んでいる。

都市 OS に含まれるデータ連携基盤を支えるデータの中核をなす部分が、政府が整備を進めてきた IMI (Infrastructure for Multilayer Interoperability: 情報共有基盤) である。IMI は、データに用いる文字や用語の同一性を特徴付ける概念を体系的に分析・整理し、コンピューター処理に適した定義情報としてデータに付加することで、既存システムに原則手を加えず運用を相互に維持しながら情報の共有や活用を円滑に行う<sup>15</sup>。

### 1.1.4 行政サービスを高度化する相互運用性の高いデータ

官民データ活用へ向けた相互運用性向上は、行政のサービスの品質向上と効率化を目指す世界的な要請である<sup>16</sup>。例えば米国では、連邦政府の出費を明細から開示する法律 (Digital accountability and Transparency Act: DATA Act) 及び DATA Act に基づくプロジェクト「USAspending.gov」が推進されている。

「USAspending.gov」は、連邦政府の支出という客観的なデータに基づく議会での討議などを可能にする EBPM の基盤である。企業開示情報の相互運用基盤である XBRL を用いて連邦政府横断で会計データとナレッジが開示・共有されており、データは各機関のデータベースから API などを通じて効率的に収集している。世界中の民間企業が利用する既存の共通モデル、情報共有基盤を政府が採用し、相互運用性の高いデータで行政サービスの品質向上と効率化を実現した好例である。

---

<sup>14</sup> 機能間連携、都市間連携、サービスシステムの共有・再利用というスマートシティの課題を解決するプラットフォームとして東京大学大学院情報学環・副学環長の越塚登教授が提唱。オープン API・標準 API、分野間データ連携基盤、デジタルツインがポイントとされている。「スマートシティ・アーキテクチャ都市間連携にむけて」[https://www.kantei.go.jp/jp/singi/tiiki/kokusentoc/supercity/supercityforum2019/190629\\_shiryuu\\_08\\_01.pdf](https://www.kantei.go.jp/jp/singi/tiiki/kokusentoc/supercity/supercityforum2019/190629_shiryuu_08_01.pdf)

<sup>15</sup> 共通語彙基盤と文字情報基盤からなり、行政データの相互運用性向上を図る。共通語彙基盤は、データに用いる様々な用語の表記、意味、構造の特徴を抽出し、体系的に整理した上で、分野を超えてデータの検索性向上やシステム連携強化を実現する。文字情報基盤は、行政で用いられる人名漢字等約 6 万文字の漢字を整備し、外字作成等のコストを解消した (2017 年 12 月に国際規格化)。

<sup>16</sup> 例えば米国では “Anticipate Future Uses: Create data thoughtfully, considering fitness for use by others; plan for reuse and build in interoperability from the start”, “Federal Data Strategy : Leveraging Data as a Strategic Asset <http://strategy.data.gov/principles/>” を公表している。

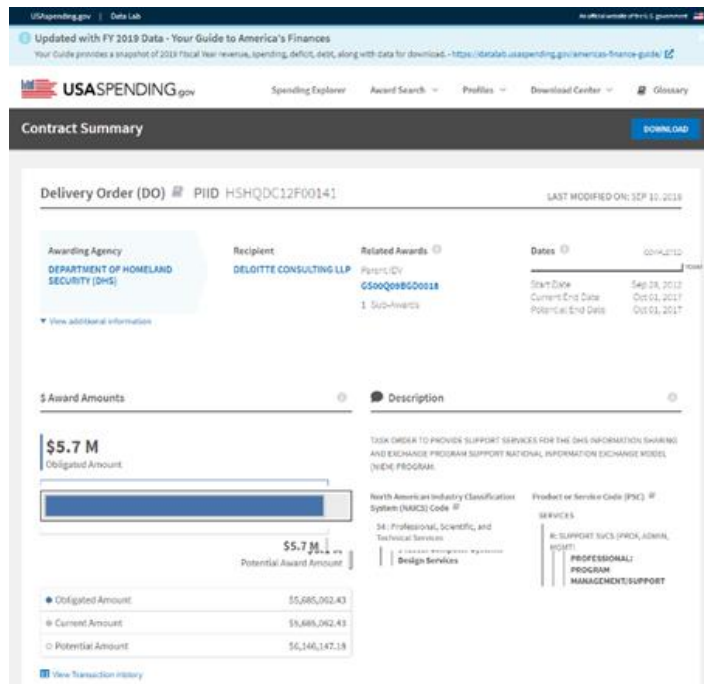


図 1-3 全政府機関の予実執行状況を公開する USAspending.gov  
 (出典) <https://www.usaspending.gov/>

## 1.2 データの相互運用性の必要性

なぜいま産官学そして市民等をあげデータの相互運用性に取り組むべきなのか概説する。

### 1.2.1 データ活用社会の参加者として

個別組織の情報システム内に閉じていたデータはいま、組織や時空の壁を超え流通し活用されようとしている。公共機関等によるオープンデータ公開の取り組みや、行政機関が保有する個人情報と特定の個人が識別できないよう加工した「非識別加工情報<sup>17</sup>」の活用も始まっている。また、データ取引市場や情報銀行<sup>18</sup>を通じてデータを売買する環境も整いつつある。今後は、組織内に必要以上のデータをもたず足りないデータは外部データで補い、それら内外のデータをつないで活用する社会となっていく。電子ファイル等で物理的に移転させず API 等でつなぐデータ活用も進むであろう。

つながる官民データ活用社会では、データ作成者・保有者・利用者といった参加者に、相互運用の理解と協力、相互運用性を高めるための参画が求められる。ひとつの相互運用性に欠けるデータセットは、伝搬されるたびに利用者による整形や意味の確認といった追加負担を産み、その積み重ねは膨大な社会コストになる。官民間問わず、誰もが相互運用性の高いデータ作成と運用の当事者なのである。

他方、EU の GDPR<sup>19</sup>等が求める個人情報ポータビリティ権保証への対応においても、相互運用性確保は不可欠である<sup>20</sup>。日本でも同様に情報銀行間のデータポータビリティ実現へ向けた仕様検討が始まっている。組織を超え、処理技術の変化を前提とするデータ活用社会では、あらゆるデータに相互運用性が求められる。

よりよいデータ活用社会の創造を担うのは、ほかでもないデータと関わる一人ひとりである。参加者の自発的かつ責任のある行動が欠かせない。

### 1.2.2 2025 年もうひとつの崖 データ資産の不活用

1.2.1 では官民データ活用社会参加者としての心構えに触れたが、データ相互運用性はデジタルトランスフォーメーション(DX)を迫られるビジネスのデータ活用力に直結する問題でもある。価値創造のための施策や根拠としてのデータ活用、そのためのリテラシー向上は喫緊の課題である。経済産業省「DX レポート<sup>21</sup>」は「既存システムが老朽化・複雑化・ブラックボックス化する中では、データを十分に活用しきれず、新しい

<sup>17</sup> 行政機関保有の個人情報を、利用希望者の提案に基づき個人が特定できないように加工した情報。[https://www.ppc.go.jp/personal\\_info/HishikibetsukakouInfo/](https://www.ppc.go.jp/personal_info/HishikibetsukakouInfo/)

<sup>18</sup> 個人ライフログや購買、Web 閲覧などのデータをデータ生成者自身の意志で信託され、本人同意の下で活用したい企業等の第三者へ提供する。ユーザーに報酬や信用スコアといった直接的な対価やデータ活用による社会的課題の解決といった間接的な対価を還元。

<sup>19</sup> General Data Protection Regulation。欧州一般データ保護規則。

<sup>20</sup> 例えば、Apple、Google、Facebook、Microsoft、Twitter は共同で各社サービス間でのポータビリティ実現へ向けた Data Transfer Project に取り組んでいる。<https://datatransferproject.dev/>

<sup>21</sup> 2018 年 9 月経済産業省が公表。[https://www.meti.go.jp/shingikai/mono\\_info\\_service/digital\\_transformation/pdf/20180907\\_03.pdf](https://www.meti.go.jp/shingikai/mono_info_service/digital_transformation/pdf/20180907_03.pdf)

デジタル技術を導入したとしても、データの利活用・連携が限定的であるため、その効果も限定的となってしまう」課題を指摘している。データという資産を活用できない状態こそが、「崖」の本質である<sup>22</sup>。

組織内でも異なる情報システム内に閉じられたデータをつないで活用できている企業はまだ決して多くない。データは営業支援システム、顧客管理システム、請求システム、会計システム等の業務システムに分散している。通常業務では処理ごとに必要なデータしか見ないし、つないで新たな価値を見出す発想を生みづらい。

「顧客A社」や「顧客A社B支店」との取引金額を、戦略や経営計画策定、マーケティング、商品開発、事業報告書作成といった異なる目的に応じてその都度データベースソフトや表計算ソフトで整形・集計する例も見られる。データサイエンティストの仕事は低品質データの除去など分析以前の処理が7-8割といわれる<sup>23</sup>ゆえんで、社外データを購入できたとしても、それらをつないだ分析、活用に多大なコストがかかる。

社内データ資産の相互運用性向上は決して簡単ではないが、データ活用経営において不可欠な取組みである。データ活用の用途によっては「0.1.1 ガイドの目的 ～データ活用社会の実現へ向けて」で挙げた「検索でき、所在がわかり、アクセスできる」という条件に加え、「データが正しい」ことも求められるが、その検証にあたっては相互運用性は有効である。

### 1.2.3 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。

官民データ活用推進基本法は、官民データの適正かつ効果的な活用の推進に関する施策を総合的かつ効果的に推進し、国民が安全に安心して暮らせる社会及び快適な生活環境の実現に寄与する目的で、2016年12月に制定された。データの「保護」一辺倒から、「適正な活用」への転換を目指すものである。相互運用性向上の取組みは、第15条(情報システムに係る規格の整備及び互換性の確保等)に該当する。

[第15条1項]

国及び地方公共団体は、官民データ活用に資するため、相互に連携して、自らの情報システムに係る規格の整備及び互換性の確保、業務の見直しその他の必要な措置を講ずるものとする。

[第15条2項]

国は、多様な分野における横断的な官民データ活用による新たなサービスの開発等に資するため、国、地方公共団体及び事業者の情報システムの相互の連携を確保するための基盤の整備その他の必要な措置を講ずるものとする。

図 1-4 官民データ活用推進基本法(抜粋)

他方、我が国では先端技術をあらゆる産業や社会生活に取り入れ、経済発展と社会的課題の解決を両立す

<sup>22</sup> DXレポートをふまえた「デジタル経営改革のための評価指標(「DX推進指標」)」でも、ビジョン、人材、ガバナンス・体制の評価軸としてデータ活用をとりあげ、ITシステムに求められる要素としてデータ活用を掲げている。<https://www.meti.go.jp/press/2019/07/20190731003/20190731003.html>

<sup>23</sup> 一例として The New York Times Aug.17,2014 “For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights” <https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html> など。

る新たな社会「Society 5.0」をビジョンとして掲げ、「1.1.3 スマートシティとデータ連携基盤」などの取組みが進んでいる。

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。

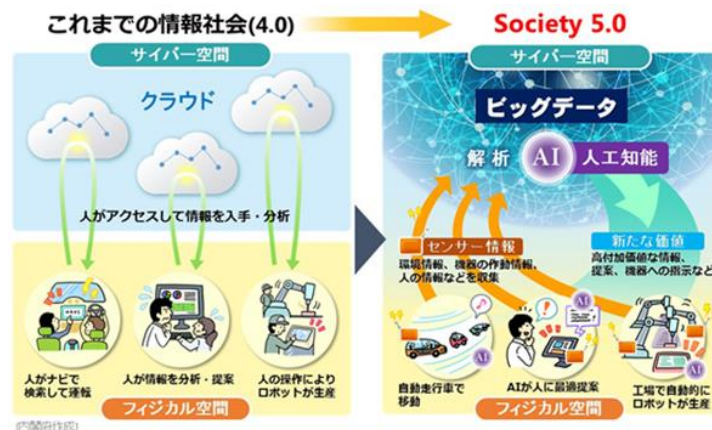


図 1-5 サイバー空間とフィジカル空間を高度に融合させる Society 5.0 のイメージ

(出典)内閣府 Society 5.0 [https://www8.cao.go.jp/cstp/society5\\_0/](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。2019年5月に公布されたデジタル手続法<sup>24</sup>では、図 1-6 情報通信技術を活用した行政の推進の基本原則の「デジタル化の基本原則」が掲げられている。

- |   |
|---|
| <ul style="list-style-type: none"> <li>①デジタルファースト：個々の手続・サービスが一貫してデジタルで完結する</li> <li>②ワンスオンリー：一度提出した情報は、二度提出することを不要とする</li> <li>③コネクテッド・ワンストップ：民間サービスを含め、複数の手続・サービスをワンストップで実現</li> </ul> |
|---|

図 1-6 情報通信技術を活用した行政の推進の基本原則

しかし、転居などのライフイベントに伴う個人の手続きや企業が会社を設立する際の手続きが全てオンラインで完結し、国民と行政がともに利益を享受する社会は、政府だけの取組みでは実現しない。官民データ活用推進基本法が第3条基本理念で「国、地方公共団体及び事業者の情報システムの相互の連携を確保するための基盤」を掲げているのは、公共性の高い事業者をはじめ、日々の活動でデータを取り扱うステークホルダーの参加が基盤の整備には不可欠だからである。同法はまた、第6条に事業者の責務を記し、民間の協力を求めている。

- |   |
|---|
| <p>(事業者の責務)</p> <p>第六条 事業者は、基本理念にのっとり、その事業活動に関し、自ら積極的に官民データ活用の推進に努めるとともに、国又は地方公共団体が実施する官民データ活用の推進に関する施策に協力するよう努めるものとする。</p> |
|---|

図 1-7 官民データ活用推進基本法第6条

<sup>24</sup> 正式名称は、「情報通信技術の活用による行政手続等に係る関係者の利便性の向上並びに行政運営の簡素化及び効率化を図るための行政手続等における情報通信の技術の利用に関する法律等の一部を改正する法律」。情報通信技術を活用した行政の推進の基本原則として、国、地方公共団体、民間事業者、国民その他の者があらゆる活動において情報通信技術の便益を享受できる社会の実現(社会全体のデジタル化)が掲げられている。

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。

## 第2章 相互運用性向上に向けたアプローチ

本章では、「第1章データの相互運用性の必要性」をふまえ、データ活用社会の参加者に求められるデータ資産の整備手法やデータ連携の仕組み、データ連携を行う際にどのように技術標準を取り込んでいくかなどの手順について紹介する。データ連携に重要なことは、データの意味することや表現方法が関係者間で合意されていることだが、相互運用性を向上させるには、その合意事項がより多くの人から参照可能な技術標準として共通認識できることにある。

### 2.1 本ガイドにおける定義

本ガイドにおいて文中で使用する用語をここで整理する。

図 2-1 は、ある企業において購買部に対し備品の購入を申請する際に提出する購買申請書のデータの例である。

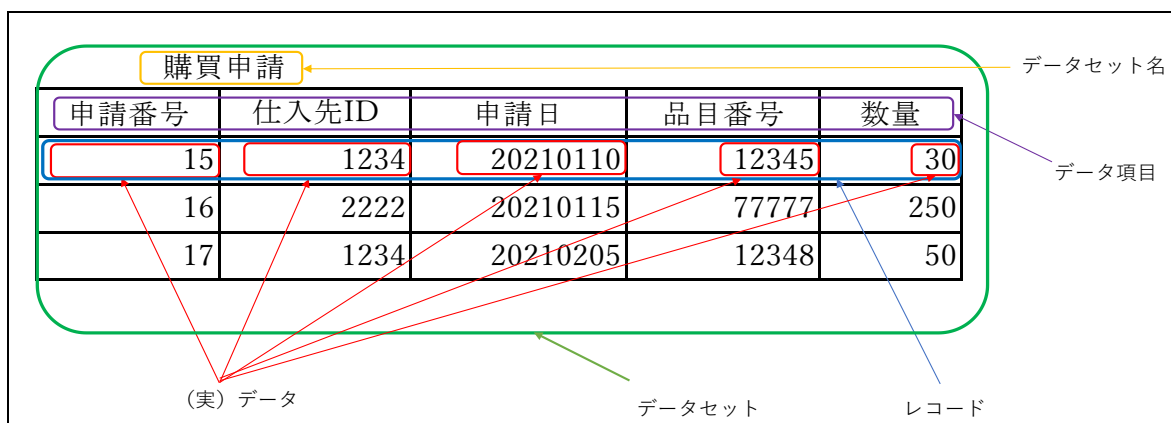


図 2-1 データセットの構成要素

本ガイドでは、このように特定の目的(購買申請)のために、ひとまとまりの情報として扱われる一覧表のセットを「データセット」と呼ぶことにする。図 2-1 の例でタイトル行に示す「申請番号」、「仕入先ID」、「申請日」、「品目番号」、「数量」は「データ項目」または「ラベル」と呼ぶ<sup>25</sup>。システム上のデータ要件定義等で利用する「データ項目名」と人間にとって使いやすい表記である「ラベル」は同じ場合と異なる場合がある<sup>26</sup>。各データ項目に入力された文字列や数値などの値は、「データ」と呼ぶ(実際に入力されているデータという意味合いで「実データ」や「インスタンス」と強調して呼ばれることもある)。

「レコード」はデータセット中のデータを構成する単位の一つで、1行分のデータを指す。

「キー」はデータセット中のレコードを一意に識別するために使うデータ項目であり、図 2-1 の例において特定の購買申請データを識別するためのキー項目は「申請番号」が該当する。

<sup>25</sup> データベースのテーブルにおいては「購買申請」をエンティティやクラス、データ項目名を属性などと呼ぶが、本ガイドでは特定の表記法に依存しないように努めている。

<sup>26</sup> 例えばシステム仕様で利用可能なデータ項目名が半角英数やコードのみに限定されているといったケースは少なくない。

私たちは物事を認識し、考え、コミュニケーションを図る際、物事を指し示すことばを用いる。そのことばとは、企業活動を例にとると、人、設備、資金、情報といったモノを対象としたり、注文、仕入、加工、販売などの出来事（コト）を対象としたりする。モノ・コトを指し示すためのデータは、範囲をどれだけの粗さ/細かさにするか、グループ化の粒度をどれくらいにするか、データ項目名として何を設定するか、などを関係者間で合意形成または共通の認識として了解しておく必要がある。了解されたデータを用いて、「当社に現在 20 件の注文があり、3 棟の工場をフル稼働して、3 日後までにうち 15 件の注文品を製造・発送する」といった活動を遂行していく。

この合意形成や共通の認識を得るための情報の一つとして「データ項目定義」がある。図 2-2 は、データセットの中のデータ項目だけを取り出して定義付けを行った「データ項目定義」の例示である。

ラベル		人向けの説明		型定義(人向け)	サンプル値	人向けの説明	意味構造	型定義(機械向け)
データ項目 (指定緊急避難場所一覧) (注1)								
項目 No.	項目名	区分	説明	形式	記入例	先進自治体 公開有無	共通語彙基盤	共通語彙基盤での値型
1	NO	○	地方公共団体内で指定緊急避難場所(注2)が一意に決まるよう、NOを設定し、記載。	文字列(半角数字)	3	有	施設>ID>識別値	xsd:string
2	名称	○	指定緊急避難場所の通称や建物等の名前を記載。	文字列	〇〇小学校	有	施設>名称>表記	xsd:string
3	名称_カナ	○	指定緊急避難場所の通称や建物等の名前をカナで記載。※記載方法について、「データ項目特記事項」シートの【共通ルール】を参照。	文字列(全角カナ)	〇〇ショウガッコウ	有	施設>名称>カナ表記	xsd:string
4	住所	○	指定緊急避難場所の住所を記載。※記載方法について、「データ項目特記事項」シートの【共通ルール】を参照。	文字列	北海道札幌市厚別区2-〇-〇	有	施設>住所>表記	xsd:string
5	方書	○	指定緊急避難場所の住所の方書を記載。	文字列	〇〇ビル1階	無	施設>住所>方書	xsd:string
6	緯度	○	指定緊急避難場所の緯度を記載。※記載方法について、「データ項目特記事項」シートの【共通ルール】を参照。	文字列(半角文字)	43.064310	有	施設>地理座標>緯度	xsd:string

語彙定義

図 2-2 データ項目定義の例

本ガイドにおける「語彙」は、人間可読の部分と機械可読の部分で構成されており、両者が一体となった一つの用語である<sup>27</sup>。データの意味を表す用語やそれが取る値の型(単位や小数点以下何位まで表示するのか、端数は四捨五入するのかなど)また、用語間の関連や具体的な表記などをまとめて本ガイドでは「データ構造」と呼ぶ。

<sup>27</sup> データ項目定義の厳密さや粒度は用途や運用コストにより異なる。人向けの説明、機械向けの定義いずれも重要だが、強制するための説明ではない。



## 2.2 データの価値向上と効果的な連携のための整備

本節ではデータ連携を進めるにあたって、データのもつ価値や相互運用性を向上させるために必要なデータの整備や、データ運用環境の整備について説明する。

### 2.2.1 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、図 2-3 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 2-3 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。図 2-4 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 2-4 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、図 2-5 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 2-5 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。しかしながら、データ項目に「担当者情報」とだけ記していても、「仕入先の営業部門担当者」なのか、それとも「商品を発送する物流部門担当者」なのか、仕入先情報の作成関係者以外の人は正確に判断できない。「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。これは、そのデータが表す意味をどのような社員でも誤解なく認識できるように、詳細な説明文を書いておくことにも通じる。

反対に、各種定義書などが整備されていない、または組織内で一元的に管理されていない（組織外のベンダーに丸投げしている等）場合、相互運用性は低下する。画面や帳票の名称やデータ項目名の意味、入力される数値や文字などの書式が統一されていない場合も同様である。社員間での認識違いにより不正確な情報が入力された結果、業務処理が停滞したり顧客に迷惑を及ぼすようなトラブルが起きやすい。仮に、組織内で管理されていたとしても、データ構造やマスターデータの定義情報などが標準化されず独自の流儀や伝統的な慣習を踏襲している場合は、異なる組織間でのデータ共有・移転が困難になるため「2.4 技術標準／国際標準の活用」で示すような技術標準を参照する配慮が必要である。

図 2-6 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	○	<b>当社</b> 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	○	<b>仕入先</b> 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	○	<b>仕入先</b> の法人名。 gBizINFO <sup>28</sup> 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	○	<b>仕入先</b> 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	○	<b>仕入先</b> 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		<b>仕入先</b> 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	○	<b>当社</b> 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	○	<b>仕入先</b> の <b>担当者</b> が所属する部署名	文字列 (20 文字以内)	総務部

<sup>28</sup> 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>

9	担当者情報 1		仕入先の <b>担当者</b> の姓	文字列 (20 文字以内)	情報
10	担当者情報 2		仕入先の <b>担当者</b> の名	文字列 (20 文字以内)	朝陽

図 2-6 意味構造を加味したデータ項目定義書

図 2-6 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。データ構造の検討や、「2.5 データ連携の準備を行うための具体的な手順」で既存の語彙等を使って機械可読可能なデータの説明を記述する際は、「仕入先」という法人の情報や、「担当者」という連絡先をひとつの塊<sup>29</sup>として検討できる。各項目にどのような値が記述されるのか、実際のデータを使って検証することもポイントのひとつである。

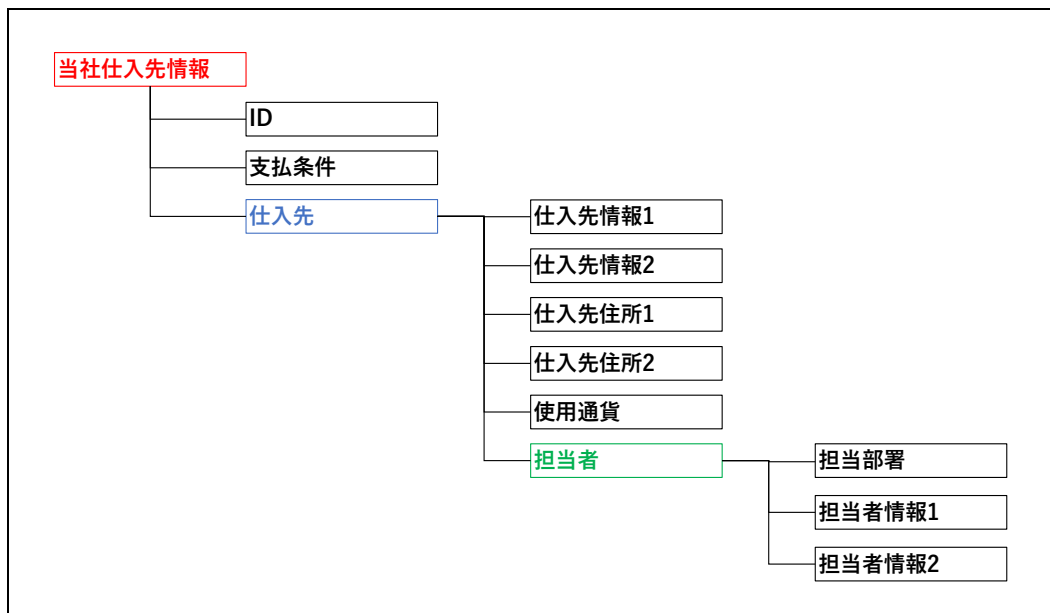


図 2-7 分類構造の展開

データ項目定義には、図 2-6 のように、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。図 2-6 の「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。伝える手段は図 2-6 のようなデータセットと別の定義書の作成、「2.4.3 標準化活動との関係づけ」で述べる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<sup>30</sup>。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。

<sup>29</sup> いわゆるエンティティの考え方に近い。

<sup>30</sup> 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。

## 2.2.2 データカタログ（データ管理台帳）の標準化

データカタログ<sup>31</sup>とは、企業や組織が保有するデータの管理台帳であり、利用者に対して開示されるデータの概要説明である。データに関する情報の例を以下に列挙する。

- データセットの識別子（URI などの ID）
- データセットの名称
- 含まれるデータ項目の内容説明
- 出所や来歴
- 発行日や更新スケジュール
- 管理者の連絡先

データカタログはデータ項目とそれぞれに対応する実データが記入されたデータセットである。図書館に収蔵された書籍でいうと、書籍を識別する「通し番号」「タイトル」「目次」「著者」「発行元」「発行年月日」などを記した書誌情報データベースに相当する。データカタログによってデータ資産を可視化すると、多くの利用者が検索によって必要なデータを発見、二次利用することが容易になる。オープンデータの利用促進やデータ流通市場での円滑な取引において欠かせないものである。

データカタログがない場合、組織内にあるデータの継続的な棚卸は容易ではなく、データが資産として可視化されなくなる。発見されないデータの価値は眠ったままになる。組織内に「前任者から引き継いだため」等、出所や経緯、利用目的、ステークホルダーが曖昧なままのデータが存在することは少なくない。

これらは組織外に開示・共有・移転できるデータかどうか、二次使用の際に著作権などへ抵触しないか、などが適切に判断できない。また、データカタログが存在する場合でも、継続的な棚卸がなされていなかったり、その書式が標準化されていなければ、連携先の組織が必要とする情報を提供できず、データの検索・比較・移転・二次利用などが困難な状況になる。

データカタログの設計に先立って、掲載する情報の種類や書式、カタログ項目名の付け方などを検討しなければならない。これらの対応にはコストと時間を要するが、その結果として、異なるデータカタログ間の照合が容易になり、開示したデータの二次利用が促され、新たな事業機会の拡大などのメリットが得られる可能性が高い。

データカタログを検討する際に、図 2-8 に示すような既存のデータカタログ定義や ISO/IEC 11179（メタデータレジストリ標準）、W3C が発行する Data Catalog Vocabulary (DCAT)などを参照する。作成にあたって相手先とどの標準を参照したかの情報を共有したり、また一般に公開することで相互運用性が高まる。

---

<sup>31</sup> 一般社団法人データ流通推進協議会では、データカタログとはデータの所在や内容等の概要情報を項目別に記入する書式の総称と定義している。

データカタログ作成ガイドラインV1.1（中間とりまとめ） 項目定義書

※領域（データタイプ）に、リンクと記載している項目は、カタログの全体構造をリンクとして表現しているものであり、入力対象項目ではありません。

※サンプル値に「※」が入っているものは、値ではなく、注意書きです。

項目 No	大分類	見出し	出現回数	属性 (データタイプ)	説明	サンプル値A (バイタル値直系) フローID	サンプル値B (バイタル値直系) ストック量	サンプル値C (バイタル値直系) ショッピング
①データカタログ本体部								
101	①	カタログ	1..1		<説明> データカタログ本体部を定義します。データカタログ本体部は、カタログ、カタログレコード、データセット、配信から構成します。			
102	①	カタログID	1..1	文字列型 (xsd:string)	<説明> このカタログをユニークに識別するための管理IDです。 <入カールール> カタログの作成者が、ユニークになるような規則(例えばURIなど)を決め、発行することを推奨します。	http://example.com/dataservice/catalog/A-BC-00613479/	http://example.com/dataservice/catalog/A-BC-00613480/	http://example.com/dataservice/catalog/A-BC-00613480/
103	①	カタログのタイトル	1..n	文字列型 (xsd:string)	<説明> このカタログの名称です。 <入カールール> 登録しようとしているデータをひと言で言い表すタイトルを記載してください。	健康データマーケットプレイス	健康データマーケットプレイス	健康データマーケットプレイス
104	①	カタログの説明	1..n	文字列型 (xsd:string)	<説明> このカタログの内容の説明です。 <入カールール> このカタログに、収録されているデータを踏まえて、データの特徴を第三者に理解してもらうための説明を記載してください。	インターネットで取得可能な健康測定データのカタログ	インターネットで取得可能な健康測定データのカタログ	インターネットで取得可能な健康測定データのカタログ
105	①	カタログの発行日	0..1	日付型 (xsd:date)	<説明> このカタログを発行した日付です。 <入カールール> このカタログを発行した日付を記載してください。データセットの発行日や配信の開始日は、別に記入欄があります。 入力形式:YYYY/MM/DD	2018/04/01	2018/04/01	2018/04/01
106	①	カタログの最終更新日	0..1	日付型 (xsd:date)	<説明> このカタログを最後に更新・修正した日付です。 <入カールール> このカタログを最後に更新・修正した日付を記載してください。データセットや配信のほうだけを変更した場合は、そちらの記入欄を扱い、この項目は修正しないでください。	2018/05/01	2018/05/01	2018/05/01

図 2-8 データカタログ定義の例

(出典) データ流通推進協議会「データカタログ作成ガイドライン V1.1(中間とりまとめ)項目定義書」

[https://data-trading.org/wp-content/uploads/2020/11/DataCatalogGuidelineV11\\_definition.pdf](https://data-trading.org/wp-content/uploads/2020/11/DataCatalogGuidelineV11_definition.pdf)

DCAT は、データカタログ間の相互運用性を促進することを目的に W3C で作成され、ウェブで公開されている。データカタログの作成にあたって DCAT を用いたデータを公開すると、そのデータを必要とする利用者が検索しやすくなる。図 2-9 は DCAT を使用して記述したデータカタログの一部である。記されたデータセットは架空のものだが、タイトル、キーワード、作者、発行年月日、変更年月日、問い合わせ先などの項目とそれに対応するデータが記入されている。

```

:dataset-001
a dcat:Dataset ;
dct:title "Imaginary dataset"@en ;
dcat:keyword "accountability"@en, "transparency"@en, "payments"@en ;
dct:creator :finance-employee-001 ;
dct:issued "2011-12-05"^^xsd:date ;
dct:modified "2011-12-15"^^xsd:date ;
dcat:contactPoint <http://example.org/transparency-office/contact> ;
dct:temporal <http://reference.data.gov.uk/id/quarter/2006-Q1> ;
dcat:temporalResolution "P1D"^^xsd:duration ;
dct:spatial <http://sws.geonames.org/6695072/> ;
dcat:spatialResolutionInMeters "30.0"^^xsd:decimal ;
dct:publisher :finance-ministry ;
dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-W> ;
dcat:distribution :dataset-001-csv ;
.
    
```

図 2-9 DCAT に基づくカタログとデータセットの例

(出典) Data Catalog Vocabulary (DCAT) - Version 2 W3C Recommendation 04 February 2020

<https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/>

## 2.2.3 ID・コードの標準化

ID(Identifier：識別子)とは、その適用範囲の中で、対象を特定し他のものと曖昧さなく区別できるようにするための情報(符号)である。図 2-6 の例であれば、項目名”ID”（仕入れ先を識別する 5 桁の整数)や”仕入れ先情報 1”（法人番号)がその例である。データ連携で用いられる ID には、メーカーなどにおける製品や部品に割り振られる番号のように内部で設計されたものと、国、市町村の分類に用いられるリストのように参照すべき公知の標準があるものが存在する。参照すべき標準がある場合は積極的に参照して設計することで、データの相互運用性が格段に向上する。

独自に設計した ID についても、複合キーであればどのような構成であるのか、また ID の桁数、用いる文字種などの定義情報を開示することが相互運用性の向上につながる。社内外の複数のデータセットを組み合わせる際には、ID により対象物同士を効率的に関連付けることができる。

コードは、分類に用いるもので、図 2-6 の場合、項目名”使用通貨”（ISO 4217 に準拠した通貨コード）、”支払条件”（独自基準）がそれらに該当する。コードには国コードリスト（ISO 3166-1）など標準が多数ある。独自にコードを発番する前にぜひ参照をしていただきたい。仮に独自作成のコードを用いる場合にも、データの更新や二次利用を考えるデータ作成者以外の担当者がそのコードを見て「これは何を意味するコードなのか」が適切にわかるように情報を開示できる状態にしておくことが大切である。

また、標準か独自基準かを問わず ID・コードともに組織的に一元管理しておくことが望ましい。政府 CIO ポータル「コード（分類体系）導入実践ガイドブック<sup>32</sup>」などを参照のうえ、ID・コードの設計及び導入、運用ルールを策定する。

上記のような ID（対象に付与される識別子）や区分（種別）コードの発番・管理が組織内で一元的になされていない場合、各部署で属人化、暗黙知化されたルールでバラバラに ID やコードが割り振られ、組織間でデータの照合ができないなどの問題が生じる。たとえば、国コードリスト（ISO 3166-1）や、日本の法人番号など標準的な参照データを利用する ID やコードが採用・参照されておらず、独自の基準や運用ルールで付番され、なおかつ、その基準やルールが判らないために、データセットに含まれる同一のデータが適切に名寄せできなくなる。

## 2.2.4 文字情報の標準化

文字情報の整備にあたっては、漢字、代替文字、フリガナ、ローマ字等の文字情報の現状や導入方法に関するガイドとして政府により策定された「文字環境導入実践ガイドブック<sup>33</sup>」をご観いただきたい。

文字情報においては、相互運用性を考慮して JIS X 0213 の利用を原則としている<sup>34</sup>。これは一般的なパソ

---

<sup>32</sup> [https://cio.go.jp/sites/default/files/uploads/documents/1018\\_code\\_guidebook.pdf](https://cio.go.jp/sites/default/files/uploads/documents/1018_code_guidebook.pdf)

<sup>33</sup> [https://cio.go.jp/sites/default/files/uploads/documents/1014\\_moji\\_guidebook.pdf](https://cio.go.jp/sites/default/files/uploads/documents/1014_moji_guidebook.pdf)

<sup>34</sup> デジタルガバメント実行計画では、JIS X 0213 を原則とし、UCS (Universal multiple-octet coded Character Set) に従った表現（符号化及び記述法）で情報システムの整備を行う、とある。UCS は ISO/IEC 10646 で規定された文字コードの国際標準である。

コンやスマートフォン等において標準的に搭載されている文字の集合である。戸籍情報に記載される、画像データとして管理された外字（がいじ）など JIS X 0213 で不足するようなケースにおいて文字情報基盤（ISO/IEC 10646）を利用することが望ましい。文字コードやエンコーディング方式の技術標準については、「2.4 技術標準／国際標準の活用」も合わせて参照願いたい。

組織内の業務担当部署のシステムごとに用いられる文字コードやエンコーディング方式が統一されていないか、把握されていない場合、データを共有・配信できない事象が起きやすい。外字が一元的に管理されていないと、文字化けなどの表示エラーを起し、コンピューターによってデータを処理できない。もし組織内で使用する文字コードやエンコーディング方式が統一され、外字が管理されていても、組織外の標準とは異なる独自仕様を参照していると、他の組織でそれらの文字情報を他の標準的な規格に容易に変換することができないために相互運用性が低下する状態に陥りやすい。

## 2.2.5 運用ルールと体制の整備

### (1) 組織的なマネジメント

法改正やルール変更、事業戦略やビジネスモデルが変化すれば、それを表すデータ構造やデータ項目も変わる。常に変化し続けることを前提として、活用ニーズを汲みつつ、組織内でデータを一元管理することが相互運用性の向上につながる。そこで組織内でのオペレーション及び、データ構造やカタログを含むデータ資産の定義・設計開発、統合管理、ならびにデータマネジメントを遂行する人材育成、教育研修を統括する部署を整備することが望ましい。

この組織的なマネジメントを行う部署が主体となって、経営視点を交えて以下のような取組みを推進する。

- 各業務部門を統括する責任者に対し、扱っているデータの所在把握やデータに用いている用語の説明文などをまとめた辞書の組織横断的な整備についての協力を依頼
- 所在把握や辞書整備における目標設定やスケジュール管理を行う業務部門参加型の共通の「場」や推進体制の設置
- 整備したデータの所在や辞書などを管理するルールの策定とルールに則った維持管理の継続
- データ作成や入力業務を担う現場の担当者を含め、組織全体に対するデータ教育研修の企画と実施
- 対外的に開示できる定義情報や採用する標準を積極的に公開
- データ連携を望む他の組織との協業検討の窓口
- データ利活用を検討するコミュニティへの参画や円滑な活動に対する支援
- データ品質を維持するためのオペレーション自動化の検討と実装

### (2) オペレーションの自動化

データの入力・転記において人手が介在するケースは多い。たとえば、FAX で受信した発注情報を見ながらコンピューターに担当者がデータを入力する、他の情報システムから印刷出力された資料を参照しながらデータを再入力するという作業である。これは負担が大きいだけでなく、個人の習熟度や認識の違い、うっか

りなどによるミスを招きやすい。その結果、データ資産全体の品質や二次利用における信頼性の低下を引き起こす原因になる。

組織的なマネジメントの一つとして、オペレーションの自動化の検討と実装を挙げたが、入力や転記などのオペレーションをできるだけ自動化することで、データの信頼性低下を防ぐことが望ましい。

業務システムとして、データの入力や転記が自動化されずに人の手が介在しているプロセスでは、省略表記・誤入力などのヒューマンエラーに起因したデータ品質の低下により、以下のような事象が発生することが多く、データの二次利用を妨げる。

- ・ 姓をローマ字転記する際にヘボン式と日本式が混在するなど、システムによる文法上のチェック機能では問題ないが、内容が不正確または不一致となることで、業務の遂行やサービスの提供で支障を来す。
- ・ 値がないデータ項目は「ゼロ (0)」なのか「空白」なのかが曖昧で、データ分析などを行う場合に結果の信頼性が揺らぐ。

オペレーションの自動化の検討はペーパーレス化だけでなく、業務フロー全体の見直しが大切である。



## 2.3 データ連携の手法

ここでは、システム間でデータを連携するための手法と実装における留意点を説明する。

### 2.3.1 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。

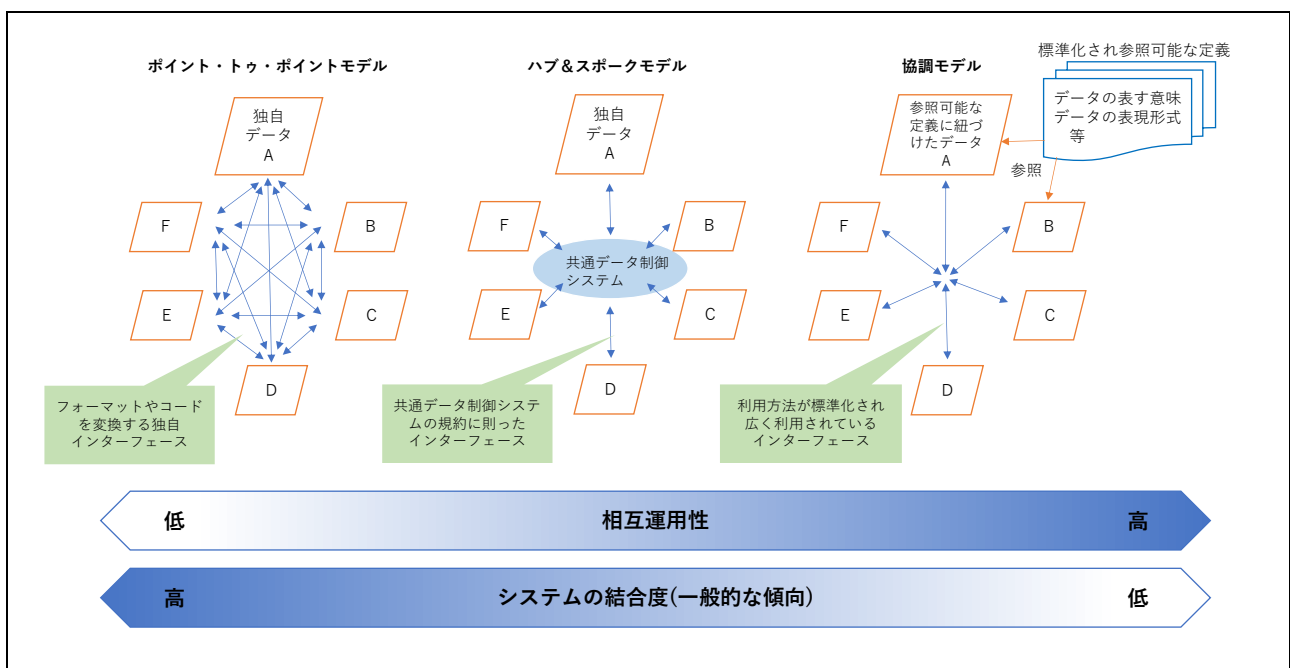


図 2-10 データ相互運用性の考え方

#### (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。小規模の社内システムなど相手先が少ない場合や固定化されている場合はよいが、相手先のシステムが多くなるとそれぞれの仕様に合わせたインターフェースの数が急増し、開発や維持・保守に係るコストの増加、高頻度なデータの取り込みによる処理負荷の増大など、業務に及ぼす影響が課題になる。

#### (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

のみとやり取りする。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。この考え方はデータを連携する組織間で共通データ制御システムの規約に則ったインターフェースを介してデータを連携する仕組みにおいて用いられている。仲介システムの仕様策定や開発運用における関係者の調整や合意形成及び共通データ制御システム開発はたいへんな取り組みであるが、作ることができればその効果は極めて大きい。サプライチェーンを構成する複数の同一グループ企業間や業界団体におけるデータの相互運用などで利用される。ただし、共通データ制御システム（外部企業の提供するサービスを含む）の技術仕様や運営管理、課金ルール、利用規約などが将来的な拡張や相互運用の制約因子になる場合がある。

### (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。データを受け渡す際に必ず、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あるのか）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<sup>35</sup>や中間標準レイアウト<sup>36</sup>はその例である。

仮に、データ項目名は異なるが同じものを指し示している場合、そのことがそれぞれの定義情報から読み取ることが出来れば、関係者がともに参照できる辞書データに登録することで以後、コンピューターも同一データ項目として処理することができる（たとえば、「融資」「貸付」「ローン」など）。

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。

定義情報の記述方法については、仕様開示された標準をできる限り参照・利用する。データに付記する定義情報が、コンピューターで自動処理できるような標準に沿って記述されていると、相手先がそれを手がかりに機械的にデータを自動変換するなど、データの連携や二次利用がやりやすく、コストを抑えられる。

さらに、標準化はされていないが関係者間で高頻度利用されるデータ項目などを、既存の標準の内容に追加・統合することで以後、参照できるデータが拡充されるメリットなどが得られる。標準の改定提案や、フィードバックを通じた標準化活動への貢献を通じて、より広く社会的に認知・利用される好循環が期待される。グループ企業を超えたサプライチェーンなど、多くの組織が参加し、協調領域として戦略的に連携して取り組む相互運用の考え方である。

---

<sup>35</sup> 政府 CIO ポータル オープンデータ。 <https://cio.go.jp/policy-opendata>

<sup>36</sup> 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](https://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/lg-cloud/02kiban07_03000024.html)

## 2.3.2 データ連携の実装

データ連携機能をシステムに実装する場合、どのような方式でその機能を実現するのかはシステム設計者にとっての課題となる。データ連携モデルに関わらず、データ連携機能を実装する方式は多数存在する。主要な方式として以下の3パターンに大別される。

### (1) リソース共有

複数のシステムが同一のデータベースやデータを格納したファイル自体を参照してデータを共有する方式。

### (2) ファイル連携

複数のシステム間で取り決めたフォーマットに従って格納したデータファイルをシステム間で受け渡す方式。

### (3) アプリケーション連携

システムがデータ連携用インターフェースを公開し、別システムがそのインターフェースを利用してデータを受け渡す方式。

	システム開発コスト	リアルタイム性	主な用途	相互運用性	実装方式など
リソース共有	低	高 (高められる)	共有リソースに対して全システムが物理的にアクセスするため、組織内連携にほぼ限定される	各システムの詳細仕様の把握が必要	連携システム数が増えるとデータ更新・参照のタイミング制御が複雑化しコストは増大する
ファイル連携	中	低	大量データの授受	システム間でファイルフォーマットと連携方法を共有する	各システムは対外的なインターフェースとしてファイル転送機能(FTP等)を実装するだけなので、組織内/組織外を問わず連携可能
アプリケーション連携	高	高 (高められる)	組織内外を問わず提供されるWebサービス(SaaS)やAPI	各システムはデータ授受のインターフェース仕様を公開する	インターフェースにHTTPを用いる前提のWebAPIは組織内/組織外を問わず連携可能。インターフェースの実装方式にはRPC, CORBA, WebAPI(SOAP, REST)など各種あるが、最近の傾向としてはシステム結合度がより疎となり、開発も容易なRESTが用いられることが多い

図 2-11 各方式の主な特徴や用途

組織間のデータ連携を考えた場合、ファイル連携、またはアプリケーション連携が採用されることが多いと考えられる。一方、データ連携の観点からは、連携方式に依らず、受け渡すデータそのものを平素から適切に整備することの重要性は変わらない。

### 2.3.3 共通フォーマットの活用

異なるシステム間においてデータを連携するアプローチとして、情報の表現方法を共通化していく考え方がある。データ資産そのものを書き換えるのではなく表現方法を揃える手法は、ビジネス及びデータ流通のグローバル化が進展する中での現実解として位置付けられる。

#### (1) 共通化の考え方

図 2-12 の例では、住所データを管理するシステム A とシステム B のデータ構造が大きく異なっており、コンピューター間でそのままデータを受け渡し、すぐ処理することができない。人間が読むと同じ住所であることはおおむね理解できるが、コンピューターで処理するには、どの項目がどの項目に一致するのか、あらかじめ逐一指示する必要がある。データを構成する文字列を分割、置換、追記するなどの事前処理も必要になる。

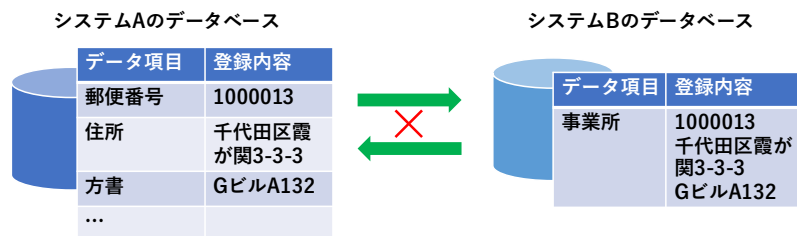


図 2-12 データ構造が異なるためにそのまますぐにコンピューター処理できないレコードの例

これに対して次のように、双方のシステム開発・運用側で合意した共通フォーマットを作成し、それに準じたデータと定義情報を付加したファイルをそれぞれ準備して、相互運用を行う例が次の図である。

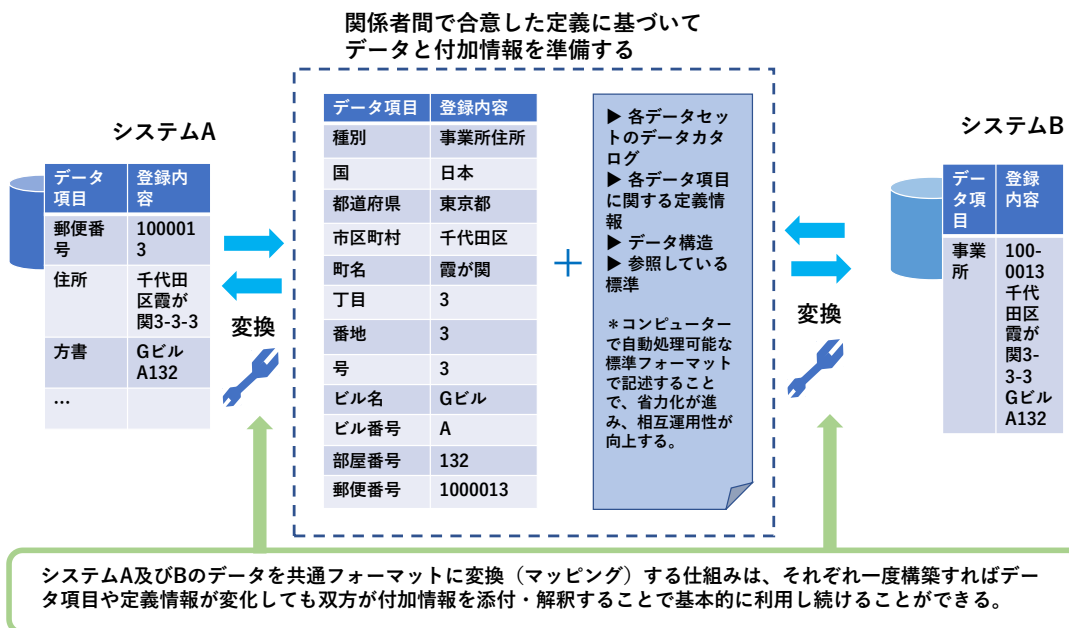


図 2-13 共通フォーマットを利用した相互運用の考え方

システム A 及び B のデータを共通フォーマットに変換（マッピングと呼ぶ）する仕組みは、それぞれ一度構築すれば、データ項目や定義情報が変化しても双方が付加情報を添付・解釈することで基本的に利用し続けることが可能で、システム A、B のデータ構造を維持した状態でデータの相互運用を実現する。また、関係者が増えても合意された標準に準拠してもらえれば、マッピングの仕組みは流用することができる。

## (2) 標準化された共通フォーマット

標準として合意されたモデルや共通フォーマットを使うと、社会全体でデータの相互運用性が高まりやすい。政府によるデータ連携基盤、共通語彙基盤や「推奨データセット」をはじめとする具体的なモデル整備は、データの相互運用という協調領域における参加者の負担を抑制し、活用されやすいデータの公開、流通を促進するために進められている。

日本政府が進めるモデルの重要な構成要素としては、図 2-14 にある「推奨データセット」と「データモデル記述」が挙げられる。日付、電話番号などの記述形式を示したデータ記述の原則<sup>37</sup>に基づき、共通語彙基盤の基本項目を「避難場所」など現場で利用するデータセットの目的に合わせてモデル化したデータ項目定義書と入力テンプレートが「推奨データセット」であり、推奨データセット等に基づき作成されたデータセットを、社会全体でデジタルに活用しやすいよう定義するための枠組みが「データモデル記述」である。

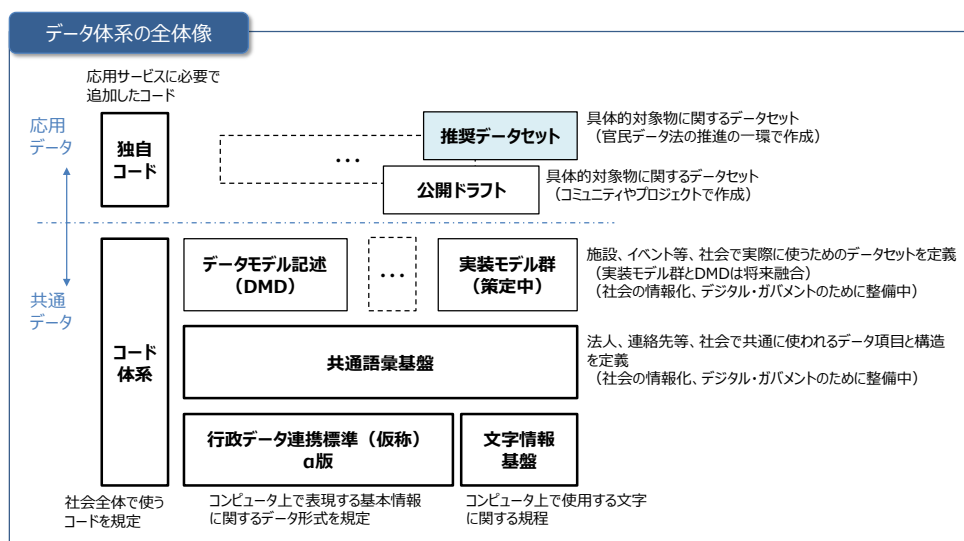


図 2-14 政府 CIO ポータルで示されるデータ体系の全体像

(出典)推奨データセットについて [https://cio.go.jp/sites/default/files/uploads/documents/opendata\\_suisyou\\_dataset.pptx](https://cio.go.jp/sites/default/files/uploads/documents/opendata_suisyou_dataset.pptx)

## (3) 推奨データセットにみる共通フォーマットの活用例

推奨データセットは、あらかじめ共通語彙基盤に則ったデータモデルや典拠が定義しており、データ入力の Excel シートと併せて提供されている。オープンデータ推進を求められる自治体はもちろん、施設情報や避難場所等関連するデータセットのモデルを検討する民間企業にも参考になる。共通フォーマットの存在

<sup>37</sup> 政府 CIO ポータルでは、手続や情報提供において分野を問わず使用される基本的なデータの形式について、データ連携を円滑に行えるよう、基本的なデータの記述形式を示したモデルを「行政基本情報データ連携モデル」として公開している。2021年3月現在、日付時刻、住所、郵便番号、地理情報、電話番号、POIコード、POIコード一覧がある。

により相互運用性の高いデータ公開が進んでいる。

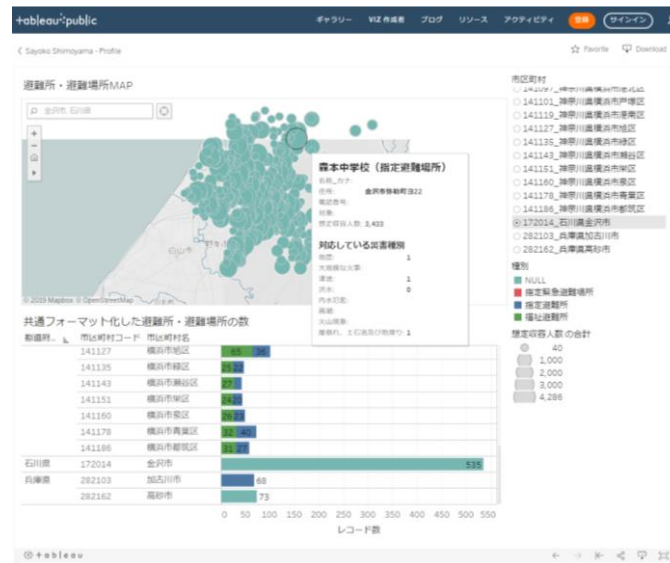


図 2-15 推奨データセットをもとに相互運用性の高いデータ公開が進みアプリに活用された例  
 (出典) 避難所 MAP [https://public.tableau.com/profile/sayoko.shimoyama.linkdata#!/vizhome/PJ\\_1/PJ](https://public.tableau.com/profile/sayoko.shimoyama.linkdata#!/vizhome/PJ_1/PJ)

## 2.4 技術標準／国際標準の活用

---

ここでは、データ表現に関わる技術仕様として使われている技術標準／国際標準にどのようなものが存在するかについて整理し、データ連携における現在の標準利用状況を把握する。そして、データの相互運用性を向上させる上で標準の利用方法をまとめる。

### 2.4.1 本書における技術標準／国際標準の分類方法

本書では、技術標準／国際標準を整理するにあたって、以下の2つの観点で標準を分類することとする。

- 規定内容による標準の分類
- 制定団体のタイプによる標準の分類

これらは、それぞれの標準の特性を知り、相互運用性向上とどのような関係があるかを理解するのに役立つ情報である。以下にそれぞれの観点での分類を説明する。

#### (1) 規定内容による標準の分類

技術標準／国際標準を体系的に整理するにあたっては、標準の規定する内容による分類項目を決め、それぞれの分類項目に対してデータ連携に関連する標準を選択するものとする。具体的には以下のような分類項目を用いる。

##### A) 「データ」の表現手段

汎用的なデータ表現のための標準の分類項目。以下の2つのサブ項目をもつ。

###### ■ 表記・記法

データ表現のための形式（例：XML、JSON）

###### ■ 符号化文字集合（文字コード）

文字形状に対してコード番号を割り振った文字の集合（例：ISO/IEC 10646、JIS X 0213）

##### B) 「データ」の伝送・格納に関する物理構造表現

文字として表現されたデータをコンピューター上に格納したり、インターネットで伝送する物理的表現（ビット列）に関する分類項目。サブ項目は以下の1つである。

###### ■ エンコーディング方式

個々の文字データをビット列として表す仕組み（例：UTF-8、Shift\_JIS）

## C) データの内容の定義

実際に交換されるデータの内容を定義するために使われる標準の分類項目。以下の 3 つのサブ項目をもつ。

### ■ データモデル

データとして表現される用語及びそれらの関係性を表す概念的なモデル。具体的な表記 (XML、JSON など) とは独立した抽象的な規定 (例: UML)

### ■ データ構造

データの意味を表す用語やそれが取る値の型、また、用語間の関連や構造について、具体的な表記を対象とした定義。(例: XML Schema)

### ■ 語彙・ボキャブラリ

実際のデータ連携、データ公開の際に記述されるデータの意味を表す用語 (「人」「氏名」など) の集合。(例: IMI コア語彙、Dublin Core など)

## (2) 制定団体のタイプによる標準の分類

標準の制定団体には、それが標準を公開するまでの「制定プロセス」と、その団体によって制定される標準の「適用範囲」の 2 つの特徴から成るタイプがある。

### A) 標準の制定プロセス

制定団体は、それぞれ標準の制定プロセスが定まっており、その特徴から、その団体が公開する標準は「デジュール標準」「フォーラム標準」「デファクト標準」に分けられる。

#### ■ デジュール標準

公的な団体の中での審議を経て制定されるコンセンサスベースの技術仕様。

#### ■ フォーラム標準

民間の企業や個人から成る会員制の組織 (フォーラムやコンソーシアム) において会員による審議を経て制定されるコンセンサスベースの技術仕様。

#### ■ デファクト標準

特定ベンダーの製品で使われている仕様が市場の支持を得て事実上の標準とみなされている技術仕様。

### B) 標準の適用範囲

標準の適用範囲として、「国際」「国内」「地域」の 3 つがある。



## ■ 国際

標準の適用範囲が、特定の国や地域の枠を越えた世界的なものであること。

## ■ 国内

標準の適用範囲が、1つの国に限られたものであること。

## ■ 地域

標準の適用範囲が、近隣の複数の国のまとまり（例えば欧州）内であること。

「A)標準の制定プロセス」と「B)標準の適用範囲」の特徴で標準の制定団体をタイプ分けすると、9のタイプに分かれることになるが、一般によく使われる表現として「国際標準」「国内標準」「フォーラム標準」「デファクト標準」がある。

「国際標準」「国内標準」という場合には、公的な団体で制定された「デジュール標準」の中で、それぞれ適用範囲（及び制定団体）が「国際」的、「国内」的なものを指すのが通例である。

「国際標準」としては、ISO（国際標準化機構：International Organization for Standardization）や IEC（国際電気標準会議：International Electrotechnical Commission）などが制定する技術標準がある。

「国内標準」としては、日本の JISC（日本産業標準調査会：Japanese Industrial Standards Committee）が制定する JIS 規格（日本産業規格：Japanese Industrial Standards）、米国の ANSI（米国国家規格協会：American National Standards Institute）が制定する ANSI 規格などがある。

「フォーラム標準」は、特に適用範囲に限らず使われる表現であるが、その中でも国際的に使われる技術標準を制定する団体としては、XML や RDF を制定した W3C（ワールド・ワイド・ウェブ・コンソーシアム：World Wide Web Consortium）などがある。なお、インターネットのプロトコルなどの標準化を行う IETF（Internet Engineering Task Force）は、個人参加で、コンセンサスの取り方が投票によるものではなく、ラフコンセンサス（例えば、参加者が挙手やハミングなどで同意を表現し、コンセンサスが得られたかどうかは議長が判断する）という独特の方式を採用している。また、IETF で公開される仕様は、RFC（Request for Comments）と呼ばれるが、最終的な Internet Standard として STD 番号が割り振られるものは少なく、その他の状態の RFC も事実上の標準（デファクト標準）として使われている場合が多い。

「デファクト標準」も適用範囲に限らず使われる表現である。国際的なデファクト標準としては、Adobe 社の PDF 仕様などがある（これは後に ISO 32000-1 として国際標準化された）。

上記の PDF の例のように、デファクト標準やフォーラム標準については、政府調達などでより広く採用されることを目指し、ISO などの国際標準化団体に提出されてデジュール標準となる場合もある。

なお、ここで説明した「A)標準の制定プロセス」と「B)標準の適用範囲」の特徴をもつ制定団体のタイプについては、制定団体によって区別がつくので、ひとつひとつの標準に「デジュール標準」「フォーラム標準」「デファクト標準」あるいは「国際標準」「国内標準」「地域標準」の種別は明示しない。

## 2.4.2 普及度を考慮した技術標準／国際標準の選択

技術標準／国際標準は、改定作業により複数のバージョンが存在する 경우가ほとんどである。また、同種の機能をもった競合する標準が複数存在する場合もある。このような状況の中で、標準を実際を選択して使う際には、データの相互運用性を高める標準を選択するために、標準の「普及度」を考慮に入れる必要がある。

必ずしも、制定時期が新しいものが実際のデータ連携で最も活用されているとは限らないので注意が必要である。標準の利用実態をふまえて、最も普及している標準を選択することが望まれる。従って、このようにして選択された標準を参照するにあたっては、バージョンを特定する情報も必要となる。

## 2.4.3 標準化活動との関係づけ

データの相互運用性を向上させるには、受け取ったデータがどの技術標準／国際標準に準拠した語彙・ボキャブラリで記述されているかを知ることが必要である。それによって、データの意味を正しく解釈し、活用することが可能となる。

この項では、データが準拠する技術標準／国際標準をどのように参照し、データ連携の相手に伝えるか、という方法について枠組みを示す。

### (1) データ解釈に必要な情報項目

送られてきたデータを解釈するにあたって必要な情報には以下のものがある。

- 符号化文字集合（文字コード） 例：ISO/IEC 10646:2017、JIS X 0213:2012
- エンコーディング方式 例：UTF-8、Shift\_JIS
- データ表記 例：XML1.0 第4版、JSON
- データ構造 例：XML Schema、JSON-LD
- 語彙・ボキャブラリ 例：IMI 共通語彙基盤 コア語彙 バージョン 2.4.2

文字コードとエンコーディング方式は、そもそもデータの文字を読むために必要な情報である。文字を読み取れるようになったところで、データ表記が分かれば、意味と値を表す部分に分解（構文解析）することができる。データ構造については、データ表記に対応したデータ構造定義言語を使うことにより、想定される構造定義にデータが適合したものであるかどうかを検証することができる。そして、意味を表す部分が実際何を表しているのかを語彙・ボキャブラリによって理解することができる。

### (2) 技術標準／国際標準の参照方法

標準を参照する際には、該当する標準を正確に特定する必要がある。前述「2.4.2 普及度を考慮した技術標準／国際標準の選択」のように、技術標準／国際標準は、類似したものも多く、同一標準においても度重なる改定によって様々なバージョンが存在するため、それらのどの標準、バージョンを使用しているかを正確

に伝えなければ、データの解釈を誤り、相互運用性が損なわれるからである。

このため、標準を参照する記述には、次の4つの項目が何らかの方法で分かるようにする。

- 制定団体
- 標準名称
- バージョン
- 取得方法（インターネット上で公開されている場合は URL）

なお、バージョン情報については、呼び方（「版」、「バージョン」）やその表記法が様々であり（番号の場合や年号の場合がある）、初版にはバージョン情報がない、あるいは、標準名称の一部に含まれている場合もある。したがって、バージョンの判定には標準名称及びバージョン情報の両方を見て判断する必要がある。

また、標準仕様の取得方法として一般に使われている URL は、W3C などでは最新バージョンを指す URL が使われることもある（このため、個別のバージョンの URL は別途用意されている）。従って、URL が、参照したい実際のバージョンの標準仕様を指す URL であることを確認して記述する必要がある。

### (3) 準拠標準情報の記述方法

「(1)データ解釈に必要な情報項目」で挙げた情報項目は、これも標準化した形で記述することにより、データ連携の利便性が高まる。

そこから参照されるのが技術標準／国際標準の場合、それらの標準を特定するための情報として「標準名称」「URL」の2つを記述し、「(2)技術標準／国際標準の参照方法」で挙げた情報項目（「制定団体」「標準名称」「バージョン」「取得方法」）が分かるようにする。「取得方法」は「URL」に対応し、「制定団体」と「バージョン」は「標準名称」「URL」のいずれかで知ることができる。参照情報項目の記述の標準化には、例えば表記に XML を使うのであれば、情報項目の構造を XML Schema で定義することになる。

XML を使ってデータカタログ (DCAT) を利用した表記例を以下に示す。データカタログにないものはここで新たに定義している（新たに定義した項目は日本語にし、接頭辞に ex を付けている）。ここで、データ解釈に必要な情報項目及び標準の参照情報をまとめて「準拠標準情報」と呼ぶこととし、以下の例では、それをタグで表現している。

```
<ex:準拠標準情報>
  <ex:文字コード>
    <ex:通称>ISO/IEC 10646 第 5 版</ex:通称>*
    < dct:title>ISO/IEC 10646:2017 Information technology — Universal Coded Character Set (UCS)</ dct:title>
    < dcat:accessURL>https://standards.iso.org/ittf/PubliclyAvailableStandards/c069119_ISO_IEC_10646_2017.zip</ dcat:accessURL>
  </ex:文字コード>
  <ex:エンコーディング方式>
    <ex:通称>UTF-8</ex:通称>
    < dct:title>ISO/IEC 10646:2017 Information technology — Universal Coded Character Set (UCS)</ dct:title >
```

```

    <dcat:accessURL>https://standards.iso.org/ittf/PubliclyAvailableStandards/c069119_ISO_IEC_10646_
2017.zip</dcat:accessURL>
  </ex:エンコーディング方式>
  <ex:データ表記>
    <ex:通称>XML1.0 第4版</ex:通称>
    <dct:title>Extensible Markup Language (XML) 1.0 (Fourth Edition) W3C Recommendation 16 August
2006</dct:title>
    <dcat:accessURL>https://www.w3.org/TR/2006/REC-xml-20060816/</dcat:accessURL>
  </ex:データ表記>
  .
  .
  .
</ex:準拠標準情報>

```

※「dct:title」で記述される標準の正式名称は一般に馴染みがなく、また、対象となる標準が何らかの仕様内の一部の規定である場合などは、「dcat:accessURL」によって参照される仕様書の情報だけではその標準を識別できないことがある。このため、対象となる標準を識別できる一般的な呼称を「ex:通称」で記述する。これは説明的な情報なので、バージョン情報も含めた記述とすることが望ましい。

図 2-16 データカタログ (DCAT) を利用した表記例

#### (4) データと準拠標準情報との関連付け

データを受け取る際、「(3) 準拠標準情報の記述方法」で標準化された情報を得る必要があるが、これにはいくつかの方式が考えられる。

- データと一緒に準拠標準情報を記述する
- データに準拠標準情報へのリンク情報を記述する
- データと準拠標準情報を記述したファイルをパッケージ化する

以下に、それぞれの方式について説明する。

##### A) データと一緒に準拠標準情報を記述する

これは、準拠標準情報をデータの特記箇所（メタデータ）に記述するというものである。

この方式のメリットは、データを流通させるときに、1つのファイルでデータ本体とその準拠標準情報を伝えることができる点にある。

次の例は、語彙を使ったデータが XML 表記されている場合を想定し、XML 形式で、そのデータの前に「準拠標準情報」を含めたものである。

```

<Container>← データのコンテナ
  <Header> ← ヘッダー部
    <ex:準拠標準情報>
      <ex:文字コード><ex:通称>JIS X 0213:2004</ex:通称> . . . </ex:文字コード>
      <ex:エンコーディング方式><ex:通称>UTF-8</ex:通称> . . . </ex:エンコーディング方式>
      <ex:データ表記><ex:通称>XML1.0 第4版</ex:通称> . . . </ex:データ表記>
      .
      .
    </ex:準拠標準情報>
  </Header>

```

```

<Body>← データ本体部
  <ei:EventInfo> ← イベント情報を例示するためのサンプル語彙であり、接頭辞として ei を付けた
    :
    :
  </ei:EventInfo>
</Body>
</Container>

```

図 2-17 準拠標準情報を含めたデータの例

この場合、「Container」「Header」「Body」などの記述方式はあらかじめ決めておく必要がある。

### B) データに準拠標準情報へのリンク情報を記述する

これは、準拠標準情報を表現した記述は、別ファイルとして特定の場所に格納しておき、データの特記箇所ですそれを参照するというものである。

この方式のメリットは、同一の準拠標準情報を複数のデータで共有する場合に流通させるデータの量を軽減できる点にある。

以下にこの方式の XML での記述を例示する。

【準拠標準情報ファイル】 --<https://example.org/standardsInfo/A.xml> --

```

<ex:準拠標準情報>
  <ex:文字コード><ex:通称>JIS X 0213:2004</ex:通称> . . . </ex:文字コード>
  <ex:エンコーディング方式><ex:通称>UTF-8</ex:通称> . . . </ex:エンコーディング方式>
  <ex:データ表記><ex:通称>XML1.0 第 4 版</ex:通称> . . . . . </ex:データ表記>
  :
  :
</ex:準拠標準情報>

```

【語彙を使ったデータ本体を含むファイル】

```

<Container>
  <Header ref=" https://example.org/standardsInfo/A.xml" /> ← ヘッダー部
  <ei:EventInfo> ← 語彙を使ったデータ本体
    :
    :
  </ei:EventInfo>
</Container>

```

図 2-18 データから準拠標準情報を参照する例

### C) データと準拠標準情報を記述したファイルをパッケージ化する

これは、データと準拠標準情報のファイルをまとめて ZIP 化などしてパッケージとして流通させるというものである。

この方式は、他にもデータの利用時に必要であったり有用だと思われる情報（データ解説書など）がある場合に、データと準拠標準情報を別個に扱い、その他の情報を合わせた一連のデータ関連情報群をパッケージとして配信することによって、データ利活用に必要な情報群をまとめて伝えたり入手できるというメリットがある。

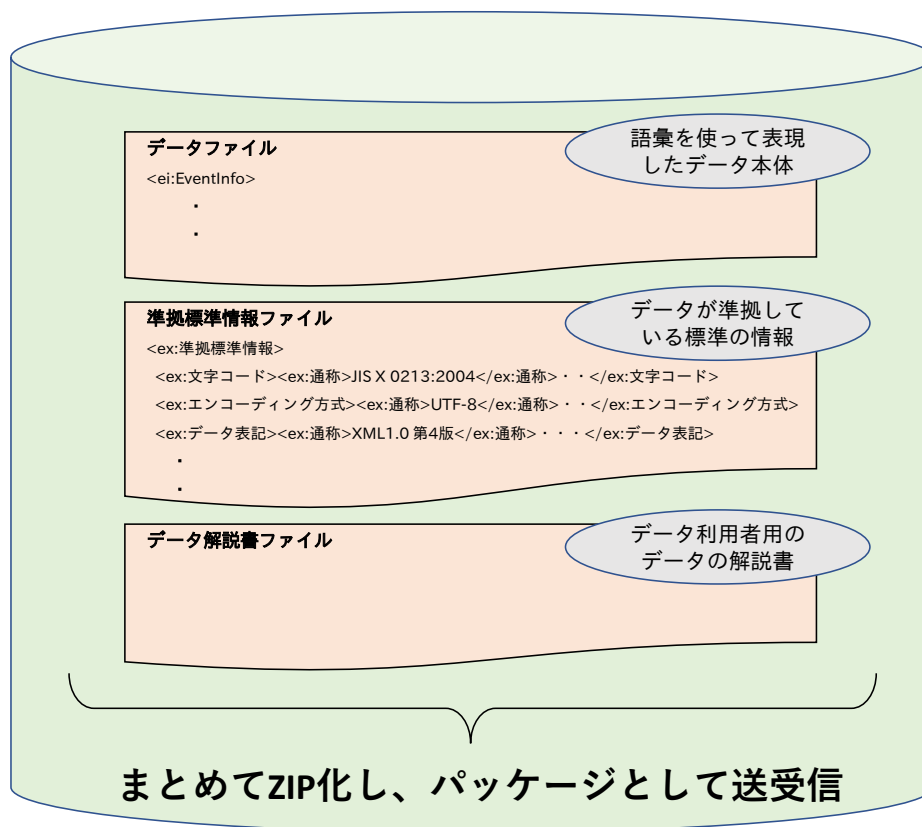


図 2-19 パッケージ化のイメージ

## 2.4.4 合意形成と運用

「2.4.3 標準化活動との関係づけ」では、受け取ったデータがどのような技術標準／国際標準に準拠しているかを記述する方法や、それを伝達する仕組みについて提案を行ったが、そのような方法論が存在するだけでは、データの相互運用性を向上させることはできない。データの相互運用は、送り手と受け手の間での共通の理解によって初めて成立するものだからである。

例えば、データと共に伝える情報として何があるかについては、まず、「(1)データ解釈に必要な情報項目」「(2)技術標準／国際標準の参照方法」を相互に理解していなければならない。そして、それがどのような形式で記述され、「準拠標準情報」として提供されるかについては、「(3)準拠標準情報の記述方法」を理解しておかなければ、必要な情報を解釈することができない。さらに、実際のデータ送受信において、データを含む、関連する情報がどのような形でやり取りされるかについては、あらかじめ「(4)データと準拠標準情報との関連付け」を参考にして決めておく必要がある。

つまり、「2.4.3 標準化活動との関係づけ」で提案した仕組みの共通理解が相互運用の基盤となる。そして、2者間での相互運用を3者間、4者間、・・・と広げてゆけば、相互運用によるデータ利活用の範囲が大きくなる。そのようにして、様々なデータが広範囲に相互運用されることによって、最終的に、提供される情報の恩恵を受ける社会全体でデータの利活用が大いに促進されることになる。

このため、それらステークホルダーが集まりデータ利活用のための合意形成の場をもつことが望ましい。

そのような合意形成の場において、共通理解の対象となるデータに関連した技術、あるいは他の運用面での規約、仕様、ガイドラインを策定することにより、参加しているステークホルダー間で、データを正しく解釈するための共通理解を得ることができるようになる。

このようにして、データ利活用に関する共通理解を土台に、データの相互運用性が向上する。

## 2.5 データ連携の準備を行うための具体的な手順

ここでは、データ連携を行うにあたって、ここまでに提案したデータの相互運用性を向上させるための枠組みを活用し、具体的な仕様をどのように決めてゆけばよいかについて説明する。この考え方は、そもそも情報をデータ化する際に、内部処理を含めてデータ処理全般において必要なものであるが、本ガイドでは、主にデータ連携を対象として記載している。

### 2.5.1 語彙・ボキャブラリの検討

#### (1) 連携対象データの分析及びデータの意味の整理

連携したいデータを分析し、その意味を伝えるための用語の整理を行う。

例：「人」「氏名」「年齢」「住所」など

#### (2) 既存標準の適用可能性の検討

整理された用語に対し、既存の技術標準／国際標準の語彙・ボキャブラリの中に適用可能なものがあるかどうか検討し、適用可能であればそれを採用する。それらは、データの意味が標準化されているので、データの共通理解が容易であり、相互運用性の向上につながる。

標準化された語彙・ボキャブラリ仕様が多数の語彙・ボキャブラリを規定しており、実際の連携対象データに対し、採用した語彙・ボキャブラリの集合が大き過ぎる場合がある。そのような場合、採用した標準の語彙・ボキャブラリの一部だけを使用することも考えられる。

例：「共通語彙基盤コア語彙 2.4.2」など

#### (3) 新たな用語の策定

採用した既存の語彙・ボキャブラリだけでは扱う用語全てを表現できない場合、表現できない用語については、データ連携に関わるステークホルダーによって検討を行い、新たな語彙・ボキャブラリ用語として策定する。

例：「共通語彙基盤コア語彙 2.4.2」にはない「自動車」「税理士」など

### 2.5.2 表記・記法の決定

次に、扱う用語の内容を実際に記述する表記・記法を決める。これは、XML、JSON など、付録に挙げたように広く認知され、その表記・記法を解釈し処理を行うプログラムが充実した標準を採用することが望ましい。

例：XML など



## 2.5.3 データ構造の決定

データの表記・記法の標準には、用語（XML の場合はタグで表される要素）やそれらの間の関係を構造（用語同士の階層関係及び用語の出現回数など）として決め、さらに用語の特性や属性（数値の桁数など）を定義するための規定（XML の場合は XML Schema）などがある。これらを用いて、連携したいデータの語彙・ボキャブラリの具体的な定義を行う。これによって、データの作成時や解釈時に、データの書き方が意図したものであるかどうかをチェックすることができるようになる。

以下に、「人」という用語のデータ構造が、その一部として「氏名」「住所」「年齢」で構成されるという階層関係を持ち、それぞれの用語の出現回数も指定されていることを「共通語彙基盤コア語彙 2.4.2」の XML Schema を使って例示する。また、特定の場面で、「共通語彙基盤コア語彙 2.4.2」には含まれていない「自動車」「税理士」といった用語を使いたい場合に、それぞれ独自に用語を策定し、XML Schema を使ってデータ構造を定義する例も示す。なお、ここでは、何らかの規定を用いて用語のデータ構造を定義する必要があるということを示すことが目的であり、読者は、XML Schema の構文を厳密に理解する必要はない。

【「共通語彙基盤コア語彙 2.4.2」の XML Schema より】

```
<xsd:complexType name="人型">
  .. 中略 ..
  <xsd:complexContent>
    <xsd:extension base="ic:実体型">
      <xsd:sequence>
        <xsd:group ref="ic:人型-固有要素グループ"/>
      </xsd:sequence>
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>
<xsd:group name="人型-固有要素グループ">
  <xsd:sequence>
    .. 中略 ..
    <xsd:element ref="ic:氏名" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element ref="ic:住所" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element ref="ic:年齢" minOccurs="0" maxOccurs="1"/>
    ↑
    用語を表す XML の要素「ic:年齢」が 0 回以上 1 回以下出現することを指定
  </xsd:sequence>
</xsd:group>

<xsd:element name="年齢" type="ic:数量型" nillable="true">
  .. 中略 ..
</xsd:element>

<xsd:complexType name="数量型">
  .. 中略 ..
  <xsd:complexContent>
    <xsd:extension base="ic:概念型">
      <xsd:sequence>
        <xsd:group ref="ic:数量型-固有要素グループ"/>
      </xsd:sequence>
    </xsd:extension>
  </xsd:complexContent>
```

```

</xsd:complexType>
<xsd:group name="数量型-固有要素グループ">
  <xsd:sequence>
    <xsd:element ref="ic:数値" minOccurs="0" maxOccurs="1"/>
    . . . 中略 . . .
  </xsd:sequence>
</xsd:group>

<xsd:element name="数値" type="ic:数値-単純型">
  . . . 中略 . . .
</xsd:element>

<xsd:simpleType name="数値-単純型">
  <xsd:restriction base="xsd:decimal"/> ← データが 10 進数であることを指定
</xsd:simpleType>

【「共通語彙基盤コア語彙 2.4.2」に存在しない新たな用語の定義】
<xsd:element name="自動車">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element ref="ic:名称" minOccurs="0" maxOccurs="1"/>
      <xsd:element ref="ic:製造日" minOccurs="0" maxOccurs="1"/>
      .
      .
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>

<xsd:element name="税理士">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element ref="ic:氏名" minOccurs="0" maxOccurs="1"/>
      <xsd:element ref="ic:組織" minOccurs="0" maxOccurs="1"/>
      .
      .
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>

```

図 2-20 XML Schema による定義例

## 2.5.4 符号化文字集合（文字コード）とエンコーディング方式

### (1) 符号化文字集合（文字コード）

実際のデータを定められた表記・記法に従って記述する際に使う符号化文字集合（文字コード）を選択する。日本語データを記述するには漢字・ひらがななどを含む文字コードを使わなければならない。

例：JIS X 0213:2012

## (2) エンコーディング方式

符号化文字集合（文字コード）は、使うことのできる文字の字形の集合であり、それをコンピュータで扱えるビット列にするために、特定のエンコーディング方式を選択する。これによって、コンピュータはデータとして記述された個々の文字を認識することができるようになる。

例：UTF-8

## 第3章 データ連携に向けた取組み例

本章では、業務のデジタル化を推進していくために、いかにしてステークホルダーの意識を変えていったか、データ連携を進めていく上で発生するデータの認識齟齬を排除する工夫、技術標準の利用などを、やさいバス株式会社<sup>38</sup>様の事例を用いて解説する。

### 3.1 「やさいバス」事業概要

2017年4月に静岡県でスタートした「やさいバス」は、地域内を循環する保冷車を用いて、青果物など生鮮食品を共同配送する仕組みである。「やさいバス」を通じて品物をやり取りする生産者は主に農家で、購買者はレストランなどの飲食店のほか、スーパーなどの小売店が中心である。

約40km圏内の配送エリアに10カ所ほどの野菜の集配所（停留所）を設置し、保冷車（やさいバス）が週5日ほど定期巡回する。保冷車の到着予定時刻をめぐり停留所に生産者は商品を置き、購買者は都合の良い時に注文した品を受け取りに行く。この「やさいバス」の仕組みを立案し、サービス運営を担うのが、やさいバス株式会社（以下、やさいバス社）である。

共同配送の仕組みにより、生産者は鮮度の高い青果物を購買者に届けられる。購入者は信頼する生産者から安定した品質と量の商品を手に入れられる。

一般に生鮮食品の流通では、東京など大都市圏の市場に一度卸売りし、仲卸等を経由して飲食店やスーパーに商品が入る。一方、「やさいバス」は生産者と購買者の間の距離と時間を共同配送によって短縮している。一回の配送料はケース単位で決まる。ロットのまとまらない小口の商品もケースに詰めれば配送しやすく、輸送にかかるコストも下げられる。足が早い青果物の輸送時間も短縮できるので鮮度を維持できて商品の廃棄率が減少するほか、長距離物流に伴って生じていたCO<sub>2</sub>の排出量も削減できる。さらに物流コストの削減は、生産者が得られる価値分配比率（手取り収入）の向上につながっている。静岡県のある地域では2台の保冷車が月6500ケースを運ぶと物流面で採算が取れる試算だという。

「やさいバス」は農家と飲食店や小売店などの法人とを結ぶBtoBに特化したビジネスモデルである。生産者による商品情報の発信及び、購買者による注文の受付はそれぞれが事前に登録したECサイトを介して行われている。このECサイトでは購買者から生産者への「野菜がおいしかったよ」といった評価を投稿できるほか、お気に入りマークが増えることで生産者の評判アップにもつながる。また、購買者側も生産者との接点が生まれることで飲食店やスーパーへの集客や収益向上が可能になっている。

同社は海外でも同様の事業を展開しており、ケニアやインドなどで取組みを進めている。途上国では、生産者の価値分配比率を高めることは持続可能な一次産業を支え、国全体の食糧自給率の向上に結びつく。バスは食材以外に、畑での収穫や物流倉庫でのピッキング作業といった関連業務の従事者を運ぶことができ、

<sup>38</sup> <https://www.vegibus.co.jp/>

地域経済の活性化にも貢献している。さらに今後は青果物以外にも食肉、鮮魚など対象とする食材の種類を拡げようと試みている。

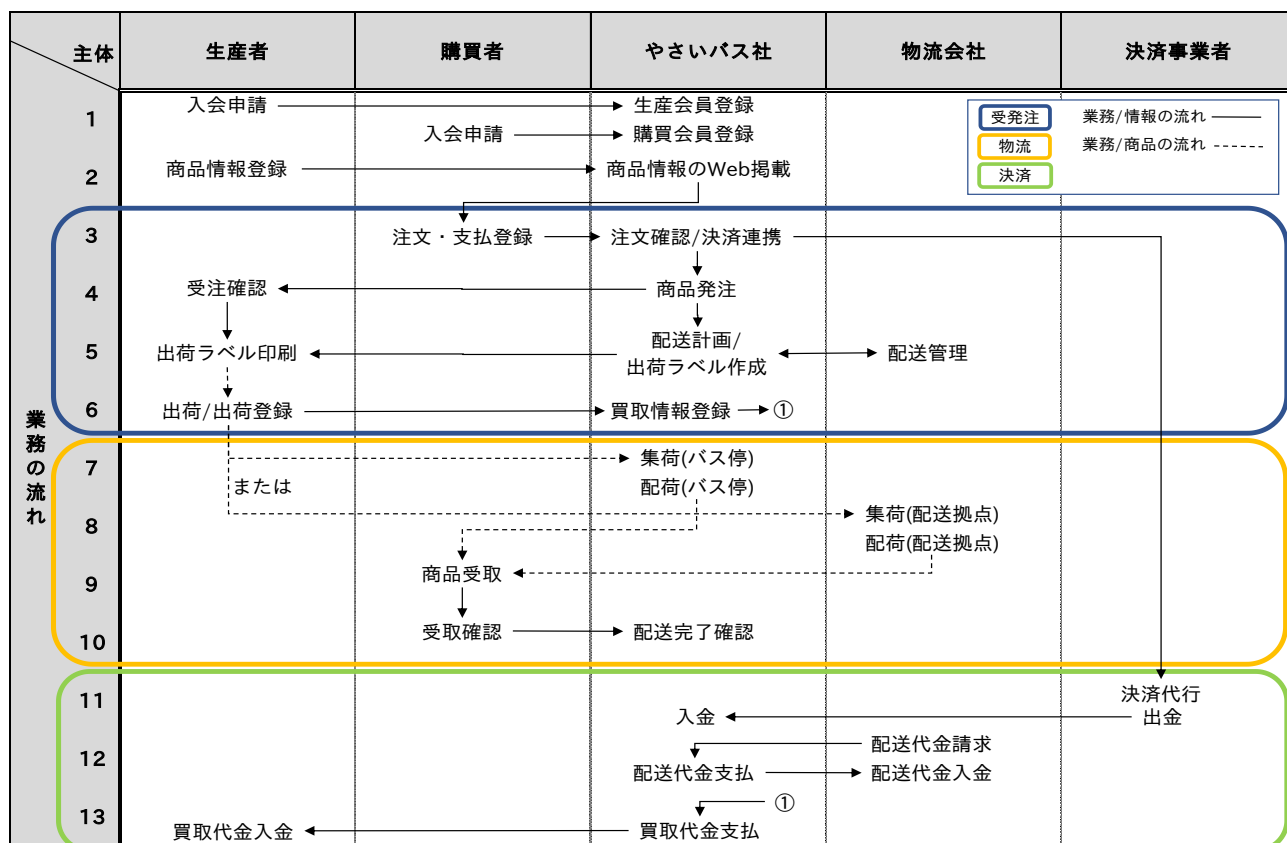
## 3.2 「やさいバス」の業務及びデータ連携

共同配送網とECサイトを一体化した「やさいバス」の仕組みは、地域コミュニティのインフラとして位置付けられる。現在、静岡県、長野県、神奈川県、茨城県など各地で地域開発が展開されている。

やさいバス社は共同配送を支える受発注、物流、決済のデータを連携するシステムを構築・運用している。システムの特徴は、物流会社や決済会社など様々な事業主体とのデータ連携がやりやすいシンプルな仕組みである。各地域の特性に合わせて柔軟に導入しやすい特徴を有している。

### (1) 「やさいバス」の業務概要

「やさいバス」の業務は、生産者・購買者・やさいバス社・物流会社・決済事業者の間で、情報や商品が受け渡されることで遂行される。図 3-1 に「やさいバス」の業務の概要を示す。



#### [業務の流れについて]

##### 1. 生産者及び購買者の入会申請

生産者及び購買者が出荷や受け取りで利用するやさいバスのバス停や物流会社の配送拠点、決済手段などを登録する。

##### 2. 生産者の商品情報登録

生産者が登録を行うことで、やさいバス社のECサイトに商品情報が掲載される。

出荷可能な期間、一日あたりの出荷可能数、あらかじめ決められたコンテナに箱詰め可能な点数などを登録する。

<p>3. 購買者の注文・支払登録  購買者は登録された商品を開覧し商品注文すると同時にクレジットによる支払い登録を行う。  支払い情報はやさいバス社から決済事業者へ連携し処理する。</p> <p>4. やさいバス社の商品発注  やさいバス社が生産者に対し商品発注し、生産者が受注する。</p> <p>5. 生産者の出荷ラベル印刷  やさいバス社は出荷情報を集約し、バス運行情報や物流会社システムと連携し配送を計画するとともに、出荷ラベルを作成。  生産者は「やさいバス」が作成した出荷ラベルを印刷する。</p> <p>6. 生産者の出荷と出荷登録  生産者はコンテナに商品を入れ、出荷ラベルを貼付する。  生産者は出荷ラベルを貼付したコンテナをやさいバスのバス停もしくは物流会社の配送拠点に持ち込むことで出荷する。  生産者は出荷完了後に出荷登録を行い、やさいバス社は出荷された商品全量を買取りを登録する。</p> <p>7. やさいバスによる配送  やさいバス社はやさいバス(保冷車)を巡回させ、バス停に持ち込まれたコンテナを集荷し、指定されたバス停で荷卸しする。</p> <p>8. 物流会社による配送  物流会社は配送拠点に持ち込まれたコンテナを指定された配送拠点まで配送する。</p> <p>9. 購買者の商品受け取り  購買者はやさいバスのバス停または物流会社の配送拠点で、注文した商品の受け取りを行う。</p> <p>10. 購買者の受取確認  購買者は商品を受け取った後に受取確認を行い、やさいバス社は配送完了を確認する。</p> <p>11. 決済事業者による決済代行  決済事業者は、やさいバス社から連携された購買者からの支払い情報に基づき決済を代行する。  決済事業者は決済した商品代金を月締めでやさいバス社へ出金し、やさいバス社はそれを入金する。</p> <p>12. 物流会社の決済  物流会社はやさいバス社へ配送代金の月締め請求を行う。  やさいバス社は請求に基づき配送代金を物流会社に支払い、物流会社はそれを入金する。</p> <p>13. やさいバス社の買取代金支払い  やさいバス社は先に登録した買取情報に基づき、月締めで生産者に買取代金を支払い、生産者はそれを入金する。</p>
---

図 3-1 「やさいバス」の業務概要

## (2) 「やさいバス」における主要なデータ連携について

「やさいバス」事業を支える情報システムは、ビジネスプロセスに基づいて、受発注、物流、決済の主に3つで構成される。どのようなデータをシステム間で入出力するかを検討は、円滑な業務処理を遂行する観点から進めている。データを入力する主体のITスキルやデータ連携の整備状況も考慮されている。

たとえば物流業者や決済事業者は、既に自社のデータ連携基盤が整備されている関係者である。物流会社の配送システムとやさいバス社のシステム連携では、リアルタイム性に優れ開発コストが抑えられる標準的インターフェイスであるAPI方式を用いて接続しており、物流業者の配送伝票を「やさいバス」の出荷ラベルと合わせて出力することができる。決済事業者との与信・収納処理についてもAPI連携によって処理が自動化されている。

一方、生産者が商品を登録する画面や、保冷車（やさいバス）の運転を委託する地元運送業者などが利用する端末などの操作画面は、ユーザーが面倒と感じないように必要最小限のシンプルなデータ項目に絞り込んでいる。入力できる実データの選択肢をあらかじめ用意する項目もある（他方で商品名や商品説明など生産者の創意工夫を尊重した自由記述式のデータ項目も用意している）。

物流における停留所の地名や住所は、地図情報のデファクト標準である GoogleMap のロケーション情報をインポートしている。ケースに添付する出荷ラベルには、集荷をする停留所と配荷をする停留所の名称、購買者名、及び管理情報としてケースの識別コードを含んだ QR コードが記されている。QR コードは信頼性の高いデータエラー訂正機能などを備えており、配送時に多少の汚れ等が生じてもデータ連携が可能となる。これにより、保冷車のドライバーがラベルの QR コードを読み取ることでやさいバス社側で物流のステータス管理を可能にしている。また GPS と連動して保冷車の位置情報もタイムリーに把握することができる。

### (3) 生産者と「やさいバス」におけるデータ認識の共通化

データを異なる組織間で扱う際に特に気を付けるべきポイントとして、データが表しているもの(意味)が何であるのかを齟齬無く認識する必要がある。「やさいバス」では認識齟齬が生じうる可能性のあるものについて、あらかじめ地域共通の言葉を用いたり、「やさいバス」で定めた選択肢しか利用できないようにすることでこれを排除している。以下に、商品登録画面を通じて入力される商品データの例を記す。

主要データ項目	定義
商品 ID	「やさいバス」独自に付与する商品識別番号（利用者は意識しない）
商品名	自由記述
商品画像	商品の画像データ
商品説明	自由記述
タグ	自由記述
品目	農協の規定に基づく大分類から選択
栽培方法	「水耕」または「土耕」から選択
1点当たりの量	単位は、kg、g、個、袋、本、パック、束、ケース、セット、等から選択
地域内での販売価格	1点当たりの販売価格（円）
全国での販売価格	1点当たりの販売価格（円）
サイズ 120 の箱 1 つに入る点数	点数
サイズ 100 の箱 1 つに入る点数	点数
購入締め切り日時	出荷可能日付からの購入締め切り日時(2 日前 17:00,1 日前 17:00 選択)

図 3-2 「やさいバス」商品データの例

「やさいバス」は地域内での流通を基本としており、商品登録画面でも商品名をあえてリスト化（マスターデータ化）せず、商品の名称については地域で最も一般的なものを自由記述で入力してもらうとともに、出来るだけ商品の画像を登録してもらうようにしている。生産者や購買者が抱えるデジタル化の障壁を低くするために、商品 ID を意識させないように工夫されている。利用者は商品名と画像の組み合わせで商品特定し、システムは商品 ID で商品特定する。

利用者間やシステム間で認識の相違が発生する可能性がある項目については、既存の標準を用いたり選択肢を固定することでデータのもつ意味の共有を図っている。例えば品目については地域の農協が定めている規定を用いて選択肢として提示している。商品の量については生産者と購買者の間で認識が異なると円滑な

取引に支障を来すため、「やさいバス」では齟齬が生じないように使える単位をあらかじめ商品登録画面にて固定化している。また、日時を指定するような項目であれば、24時間制の表記を選択肢として提示している。併せて、「やさいバス」では生産者や購買者に対してデータ入力に関する詳細なマニュアルを提供しており、このマニュアルがデータを共通認識するための標準として機能している。

①商品登録する。ホーム画面より、出品ボタンを押してください。

商品画像をアップしてください。写真があると購入されやすくなります。※後日、対応可

品目  
野菜や、魚など品目を選択してください。

栽培方法を選択してください。特にない方は、その他を選択してください。

1点あたりの量  
単位はプルダウンより選択してください。

全国に向けて出品される方は、チェックが入っている状態をお願いします。

全国にチェックを入れた方は、全国での販売価格も入力してください。※相場に合わせて価格を変えても大丈夫です。

税込の価格が自動的に入力されます。

箱の占有率を自動計算するための参考値にするため、記載をお願い致します。地域内では120サイズ、全国の場合は80、100、120、140、160サイズで出荷することができます。ご都合に良いサイズに入力してください。

他サイズも設定する

図 3-3 「やさいバス」商品登録マニュアル(抜粋)

(出典) <https://vegibus.s3-ap-northeast-1.amazonaws.com/html/upload/新規登録出荷者.pdf>



## 3.3 事業成功のポイント

---

### (1) ステークホルダーとの合意形成

「やさいバス」の地域への導入では、共同配送の仕組みづくりが重要になる。そこで欠かせないのが、住民主体の「地域チームづくり」である。やさいバス社では、地元の生産者、購買者、物流業者、自治体、また地域の活性化などに前向きな金融機関や有力企業などの協力を募り、生産者と購買者の信頼を高めていくような地域に根差した仕組み作りを行っている。仕組みだけ先走って作っても地域住民の支持が得られなければ、生産者も商品データなどを入力・公開してくれず、購買者からの注文も入らないことが理由である。また、一部の人だけ利用しても高い相乗効果は得られない。そこで住民にまずはメリットを体感してもらい、全体最適で地域に実装していく、という順序を大切にしている。そして、地元のメディアや有力企業、行政の首長を巻き込んだ情報発信にも力を入れている。

共同配送の仕組みは購買者が支払う配送料と、生産者が支払う手数料をベースに全国で共通化・規格化されている。シンプルなルールにすることで農業経験者でなくても「やさいバス」の仕組みづくりに携われるようにするねらいがある。

### (2) 地域への定着

同社ではシステムに搭載する機能を高度化していない。「やさいバス」利用の敷居を低くすることを優先して、必要最小限の機能から設計・実装している。生産者や購買者が利用する EC サイトへのデータ入力などについても先述のように簡素化しているが、IT が苦手な人に対してはやさいバス社のスタッフなどが慣れるまで支援することもある。生産者は一般に、FAX、または電話口で購買者からの注文を手書きで書き留めるなど、事務処理のほとんどを手作業で行っていることが多い。注文の訂正依頼などが入ると事務作業に追われ、本来の農業に専念できなくなることも少なくない。「やさいバス」の仕組みで受発注などの業務がデジタル化されることで、仕事にゆとりが生まれるケースが増えている。

「やさいバスがなくなると困ってしまう」と地域の生産者や購買者が思うように地域に根付くにはだいたい3年程度かかるという。鮮度の高い野菜を、物流コストを抑えた適切な価格・量で取引できる購買者側のメリット、そして煩雑な事務作業を減らすことで仕事に余裕が生まれ、手取り収入が増える生産者側のメリットが周知された地域では、「やさいバス」を通じた取引でないと困る、と条件を出す生産者、購買者も増えており、地域への定着が進んでいる。

## 3.4 データの相互運用性向上に関する今後

---

### (1) 技術標準の適用

「やさいバス」の取扱品目も青果物だけでなく、鮮魚や畜産品、また生鮮品以外の商材を運びたいというニーズも生まれている。商品を識別するための方法として、日本で最も普及している標準規格である JAN コ

ード（事業者番号9桁+商品アイテム番号3桁+チェックディジット1桁）の書式に則り、事業者や商品アイテムを整備することを検討している。これにより国内外<sup>39</sup>に向けたデータ相互運用性の向上を図るとともに、データ連携にかかるコストが削減可能となる。

## (2) 商品の共通認識

商品の識別については、今後「やさいバス」の取引・流通が地域を越えて進展する場合、論点になりえる。現在、商品名は自由記述、品目については地域の農協が定めている規定を用いて選択肢として提示しているが、地域性による名称表記の「揺れ」が大きいと同一商品の横断的な検索などが難しくなる可能性もある。仮に、商品画像をAIで解析して分類するにしても、誤りや情報の信頼性を考慮する必要がある。やさいバス社では、その際にどのような商品名、品目にするか、新たなデータ項目が必要か、名称の定義やデータの関連付けを含めたデータ標準化の議論について、取引関係者を交えて進めたい考えである。現時点では、タグを手がかりとした地域横断的な商品検索を実装している。

## (3) 連携組織の多様化

今後、事業拡大に伴う新たな取引先事業者とのデータ連携も想定されている。これには、データの標準化や関係者間でのデータカタログの共有をはじめ、多様な事業者とのデータ相互運用性を図る際の基礎資料の整備が重要となる。そのため、受発注、物流、決済を連携する業務プロセスや主要データ及びデータ間の関連図などを整理している。

---

<sup>39</sup> JANコード(13桁)は、国際的にはGTIN-13と呼ばれ、海外で利用されているEANコード、UPCコードなどと互換性がある。

# 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。

どのような組織にもデータ資産は存在する。また、データ資産にはデータの生成過程において、組織内で生成されるデータと、組織外から定期または不定期に組織内に取り込まれるデータが存在する。多くのデータセットは業務システムで定義されたデータ構造その他仕様に基づく設計・実装に即して処理されており、データ項目定義やメタデータなどが一元的に管理されていない。すなわち、資産として活用したり保護したり出来ない状態のデータである。

データセットを資産として活用できる状態(検索でき、所在がわかり、アクセスできる)にするために、組織内のデータ資産を把握することから始める。組織内にどのようなデータセットが存在するのか、システムや部署ごとに、第2章でとりあげたようにデータカタログ等を活用して整備する。その際にはデータ生成の起点であるビジネスモデルや業務プロセスの定義を確かめ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 CDO(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。

第2章ではデータを表現する方法として語彙を利用することにも触れた。語彙は単なるラベルではなく①データの本質的な意味を表す概念、②概念間の関係性、③付加情報で構成されるナレッジを伝える重要な役割を有している。ことばは常に変化し続けるものであり、語彙もまた社会的なルールや文化など人を取り巻く環境の変化に伴って変化し続ける。したがって、語彙を体系化するデジタル的な技術論に加え、語彙の運用管理やステークホルダー間の調整などを行う合意形成の場が重要である。この合意形成の場の代表例として、NIEM<sup>40</sup>は10年、XBRL<sup>41</sup>は20年、UN/CEFACT<sup>42</sup>は30年以上、データの相互運用性を確保するためにステークホルダーが協力しながら、環境変化に合わせて内容を更新し続けている。

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。

<sup>40</sup> 本ガイド別冊の「米国「NIEM」に見るコミュニティ主導の情報交換」を参照

<sup>41</sup> 本ガイド別冊の「企業情報開示の相互運用基盤 XBRL」を参照

<sup>42</sup> United Nations Centre for Trade Facilitation and Electronic Business。貿易簡易化と電子ビジネスのための国連センター。国際連合傘下で、商取引や貿易の促進を目的として電子データ交換(EDI)などの標準を整備している

# 付録 データ表現に関する国際標準化活動

この付録では、データ連携における現在の標準利用状況を把握するため、「2.4 技術標準／国際標準の活用」の「2.4.1 本書における技術標準／国際標準の分類方法」で記した標準の分類や、「2.4.2 普及度を考慮した技術標準／国際標準の選択」で記した標準の選択基準に基づいて、データ表現に関連する技術標準／国際標準をリストアップする。

## A) 「データ」の表現手段

### ■ 表記・記法

#### ● XML (Extensible Markup Language)

コンピューターが処理するデータをテキスト形式で記述するための言語。データのもつ意味をタグと呼ばれる記法で表し、階層的なデータ構造を表現することができる。タグによって値の意味を表現することができる。

XML のデータ構造を定義するには関連標準の XML Schema を用いる。

元々、文書データを表現するために開発されたが、インターネット上での電子商取引やデータ公開、パソコン上で操作するワープロなどのソフトウェアの内部データなどにも使われるようになった。

(参照情報)

制定団体：W3C

名称：Extensible Markup Language (XML) 1.0 (Fourth Edition)

W3C Recommendation 16 August 2006

URL：<https://www.w3.org/TR/2006/REC-xml-20060816/>

#### ● JSON (JavaScript Object Notation)

コンピューターが処理するデータをテキスト形式で記述するための記法。データのもつ意味を名前と値のペアで表現し、階層的なデータ構造を表現することができる。名前によって値の意味を表現することができる。

JSON のデータ構造を定義するには JSON Schema を用いる。

JSON は、JavaScript Object Notation という名称から分かるようにプログラミング言語 JavaScript をベースにして開発されたが、主要なプログラミング言語でサポートされているので汎用性が高い。

コンピューター間でのメッセージ／データ交換に用いられる。文書処理を除いて XML と同じ用途であるが、XML に比べてデータ量が少なく済むという特徴がある。

Ecma International の JSON 仕様「ECMA - 404」と IETF(Internet Engineering Task Force)の JSON

仕様「STD90/RFC8259」との間で統一が図られた上で、Ecma International から ISO/IEC JTC1 に提出され、「ISO/IEC 21778:2017 Information technology — The JSON data interchange syntax」として国際標準となっている。

(参照情報)

制定団体：Ecma International、IETF、ISO/IEC JTC1

名称：ISO/IEC 21778:2017 Information technology - The JSON data interchange syntax

URL：<https://www.rfc-editor.org/info/rfc8259>

<http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>

<https://www.iso.org/standard/71616.html>

- JSON-LD (JSON for Linked Data)

JSON の表記を使ってリンクトデータ (Linked Data : Web 上の情報をつなげる仕組み) を表現する記法。

リンクトデータは RDF のフレームワークに基づくものであり、Web 上で公開されるオープンデータにリンクトデータの考え方を適用したものは LOD (Linked Open Data) と呼ばれるが、その表現形式として JSON-LD を使うことができる。

JSON の名前と値のペアでは、名前がデータの意味を表すが、その意味は、JSON-LD では IRI (URL の拡張) で表す。ただし、IRI の文字列が長くなり、直感的に意味を捉えることができないので、データを読み易くするために、コンテキスト (@context) という部分で、厳密な意味定義である IRI と簡略化された用語 (「name」など) の対応を取ることができる。これによって、データ構造も表現できる。

(参照情報)

制定団体：W3C

名称：JSON-LD 1.0 A JSON-based Serialization for Linked Data<sup>※</sup>

W3C Recommendation 16 January 2014

URL：<https://www.w3.org/TR/2014/REC-json-ld-20140116/>

※JSON-LD は、最新バージョンの JSON-LD 1.1 が 2020 年 7 月 16 日に W3C Recommendation となっているが、普及するには時間がかかると思われるため、JSON-LD1.0 を挙げた。

- CSV (Comma-Separated Values)

テキスト形式で、カンマで値を区切ってデータを表現する記法。値の意味は表現できない。

Microsoft 社の表計算ソフト Excel は、データを CSV 形式で保存し、読み込みもできる。

特に規格がない状態で普及した後、2005 年に IETF が「RFC4180」を公開し、一応 CSV の構文を明文化したが、様々な方言があるので注意が必要である。

(参照情報)

制定団体： IETF

名称： Common Format and MIME Type for Comma-Separated Values (CSV) Files

URL： <https://www.ietf.org/rfc/rfc4180.txt>

## ■ 符号化文字集合（文字コード）

### ● ISO/IEC 10646 Universal Coded Character Set (UCS)

全世界の文字を対象とし、個々の文字形状に対してコード番号を割り振った文字の集合。

国際標準の内容（追補（Amendment）を含む）は、ユニコードコンソーシアム（Unicode Consortium）の策定するユニコード（Unicode Standard）とほぼ同期を取っている。

対応する日本の国内標準は、「JIS X 0221:2014 国際符号化文字集合（UCS）」であるが、対応する国際標準のバージョンは「ISO/IEC 10646:2012[第3版]」であり、国際標準の最新バージョン「ISO/IEC 10646:2017[第5版]」との差異がある。

（参照情報）

制定団体： ISO/IEC JTC1

名称： ISO/IEC 10646:2017 Information technology - Universal Coded Character Set (UCS)[第5版]

URL： <https://www.iso.org/standard/69119.html>

### ● JIS X 0208

日本語の文字集合。第1水準漢字と第2水準漢字を含む6,879個の文字が規定されている。

（参照情報）

制定団体： JISC

名称： JIS X 0208:1997 7ビット及び8ビットの2バイト情報交換用符号化漢字集合

JIS X 0208:1997/AMENDMENT 1:2012 7ビット及び8ビットの2バイト情報交換用符号化漢字集合（追補1）

URL： [https://webdesk.jisa.or.jp/books/W11M0090/index/?bunsyo\\_id=JIS+X+0208%3A1997](https://webdesk.jisa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+X+0208%3A1997)

[https://webdesk.jisa.or.jp/books/W11M0090/index/?bunsyo\\_id=JIS+X+0208%3A1997%2FAMENDMENT+1%3A2012](https://webdesk.jisa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+X+0208%3A1997%2FAMENDMENT+1%3A2012)（追補1）

### ● JIS X 0213

JIS X 0208:1997を拡張した日本語の文字集合。JIS X 0208:1997に対して4354個の文字が追加され、11233個の文字が規定されている。第1水準漢字と第2水準漢字に加えて第3水準漢字と第4水準漢字が含まれる。

2000年に制定された後、2004年と2012年に追補の形で改正されている。2004年の改正版は「JIS2004」

と呼ばれている。

なお、マイナンバーでは、漢字の相互運用性を高めるために、申請の際の氏名及び住所の漢字を JIS の第 1 水準漢字と第 2 水準漢字及び補助漢字に収め、一般的な端末で表現できるようにしている。

(参照情報)

制定団体：JISC

名称：JIS X 0213:2000 7ビット及び8ビットの2バイト情報交換用符号化漢字集合

JIS X 0213:2000/AMENDMENT 1:2004 7ビット及び8ビットの2バイト情報交換用符号化拡張漢字集合 (追補 1)

JIS X 0213:2000/AMENDMENT 2:2012 7ビット及び8ビットの2バイト情報交換用符号化拡張漢字集合 (追補 2)

URL：[https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo\\_id=JIS+X+0213%3A2000](https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+X+0213%3A2000)

[https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo\\_id=JIS+X+0213%3A2000%2FAMENDMENT+1%3A2004](https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+X+0213%3A2000%2FAMENDMENT+1%3A2004) (追補 1)

[https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo\\_id=JIS+X+0213%3A2000%2FAMENDMENT+2%3A2012](https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+X+0213%3A2000%2FAMENDMENT+2%3A2012) (追補 2)

## B) 「データ」の伝送・格納に関する物理構造表現

### ■ エンコーディング方式

#### ● Shift\_JIS

日本語文字用のエンコーディング方式であり、マイクロソフト、アップルなどが採用したデファクト標準だったが、JIS X 0208:1997 の附属書 1 「シフト符号化表現」に規定された。

JIS X 0201 (半角英数字と半角カナ) 及び JIS X 0208 (第 1 水準と第 2 水準の漢字と非漢字) の文字を扱うことができる。

(参照情報)

制定団体：JISC

名称：シフト符号化表現

URL：[https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo\\_id=JIS+X+0208%3A1997](https://webdesk.jsa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+X+0208%3A1997) (要購入)

#### ● EUC(Extended Unix Code)

オペレーティングシステム UNIX で使われるエンコーディング方式。EUC には各国定義部分があり、そこを使った日本語 EUC、韓国語 EUC、簡体字中国語 EUC などがある。

日本語 EUC (EUC-JP) は、JIS X 0208 の文字を表現している。

JIS X 0213:2004 の文字に対応した EUC は、それとは別のものであり、EUC-JIS-2004 と呼ばれてい

る。EUC-JIS-2004 は、JIS X 0213 の附属書 3 で規定されている。

- ISO-2022-JP

インターネット上で使われる日本語文字用のエンコーディング方式であり、一般に「JIS コード」と呼ばれている。ISO/IEC 2022 のエスケープシーケンスを使って文字集合を切り替える。

1993 年に RFC1468 で規定された後、JIS X 0208:1997 の附属書 2「RFC 1468 符号化表現」に規定された。

(参照情報)

制定団体： IETF、JISC

名称： Japanese Character Encoding for Internet Messages

URL：<https://www.ietf.org/rfc/rfc1468.txt>

[https://webdesk.jisa.or.jp/books/W11M0090/index/?bunsho\\_id=JIS+X+0208%3A1997](https://webdesk.jisa.or.jp/books/W11M0090/index/?bunsho_id=JIS+X+0208%3A1997) (要購入)

- UTF-8、UTF-16

ISO/IEC 10646 で規定された文字のコード番号を、1 バイト～3 バイトのビット列に対応させるエンコーディング方式。

ISO/IEC 10646 の仕様内及び、UTF-8 については IETF の RFC3629「UTF-8, a transformation format of ISO 10646」、UTF-16 については IETF の RFC2781「UTF-16, an encoding of ISO 10646」に規定がある。

(参照情報)

制定団体： ISO/IEC JTC1

名称： ISO/IEC 10646:2017 Information technology - Universal Coded Character Set (UCS)[第 5 版]

URL：<https://www.iso.org/standard/69119.html>

## C) データの内容の定義

### ■ データモデル

- RDF (Resource Description Framework)

Web 上のリソース同士の関係を記述するフレームワーク。

RDF はリソース及びリソース間の関係を記述するモデルのみの規定であり、そこで使われる creator (作成者), date (日付) などのリソース記述の語彙は RDF Schema を用いて定義する。

(参照情報)



制定団体：W3C

名称：Resource Description Framework (RDF):Concepts and Abstract Syntax W3C  
Recommendation 10 February 2004

URL：<https://www.w3.org/TR/rdf-concepts/>

- OWL (Web Ontology Language)

Web 上のリソースの分類、関係、推論のためのルールをオントロジーの考え方にに基づき定義する言語。記述モデルとして RDF を使う。

オープンデータ用の語彙である IMI 共通語彙基盤 コア語彙で使われている。

(参照情報)

制定団体：W3C

名称：OWL Web Ontology Language Semantics and Abstract Syntax W3C Recommendation  
10 February 2004

URL：<https://www.w3.org/TR/owl-semantics/>

- UML (Unified Modeling Language)

オブジェクト指向のソフトウェア開発において、データ構造や処理の流れなどを図示する際に用いる統一記法。

UML の様々な記法のうち、「クラス図」はデータ構造を定義するために用いられ、それを基に XML、RDF、OWL などのデータ構造定義を生成することができる。

(参照情報)

制定団体：OMG (Object Management Group)、ISO/IEC JTC1

名称：ISO/IEC 19505-1, Information technology - OMG Unified Modeling Language (OMG UML) Version 2.4.1 - Part 1: Infrastructure

ISO/IEC 19505-2 , Information technology - OMG Unified Modeling Language (OMG UML) Version 2.4.1 - Part 2: Superstructure

URL：(OMG での公開資料)

<https://www.omg.org/spec/UML/ISO/19505-1/PDF> (Part1)

<https://www.omg.org/spec/UML/ISO/19505-2/PDF> (Part2)

(ISO での公開資料)

<https://www.iso.org/standard/32624.html> (Part1)

<https://www.iso.org/standard/52854.html> (Part2)

## ■ データ構造

- XML Schema

XML で記述される要素や属性及びその構造の定義を行うスキーマ定義言語。仕様の Part1 で構造定義、Part2 でデータ型が規定されている。

機能的に XML 仕様で規定されている DTD (Document Type Definition) を拡張しており、現在は XML の構造定義には XML Schema が広く使われている。

(参照情報)

制定団体：W3C

名称：XML Schema Part 1: Structures Second Edition W3C Recommendation 28 October 2004

XML Schema Part 2: Datatypes Second Edition W3C Recommendation 28 October 2004

URL：<https://www.w3.org/TR/xmlschema-1/> (Part1)

<https://www.w3.org/TR/xmlschema-2/> (Part2)

- RDF Schema

RDF で表現される creator (作成者), date (日付) などのリソース記述の語彙を定義するためのスキーマ定義言語。

(参照情報)

制定団体：W3C

名称：RDF Vocabulary Description Language 1.0: RDF Schema W3C Recommendation 10 February 2004

URL：<https://www.w3.org/TR/2004/REC-rdf-schema-20040210/>

## ■ 語彙・ボキャブラリ

- Dublin Core

文書や Web リソースのタイトル、作者、主題、説明、作成日付などのメタデータを記述するための語彙。Dublin Core Metadata Initiative という団体が保守を担当している。

中核となる 15 の語彙要素が「Dublin Core」と呼ばれるもので、RFC5013、ANSI/NISO Z39.85-2012 の策定を経て ISO 15836:2009 として国際標準となった。他の拡張語彙は「DCMI Metadata Terms (省略形は Dublin Core terms)」と呼ばれるが、これも国際標準化され、ISO 15836-2:2019 として出版された。「Dublin Core」の方の最新版は、Part1 として分離され、ISO 15836-1:2017 として出版されている)。

国立国会図書館は、「Dublin Core」及び「DCMI Metadata Terms」を基にして独自に拡張したメタデータを使用している。

(参照情報)

制定団体：Dublin Core Metadata Initiative、ISO

名称：ISO 15836-1:2017 Information and documentation - The Dublin Core metadata

element set - Part 1: Core elements

ISO 15836-2:2019 Information and documentation - The Dublin Core metadata element set - Part 2: DCMI Properties and classes

URL : <https://www.iso.org/standard/71339.html> (Part 1)

<https://www.iso.org/standard/71341.html> (Part 2)

- **ebXML Core Component Library / Core Components Technical Specification**

企業間電子商取引で使われる語彙（コア構成要素（Core Component））のライブラリ。UN/CEFACT の Web サイトで公開されている（年に 2 回更新される）。

語彙は、汎用的な CC（Core Component：コア構成要素）と、特定の応用分野のための BIE（Business Information Entity：ビジネス情報エンティティ）という 2 つに分類されている。

それらの語彙は、「Core Components Technical Specification」（ISO 15000-5 として国際標準化されている）という技術仕様で規定されたコア構成要素の開発方法論に基づいて開発され蓄積されている。

（参照情報）

制定団体：UN/CEFACT

名称：[語彙ライブラリ] ebXML Core Components Library

[技術仕様] Core Components Technical Specification - Part 8 of the ebXML Framework, 15 November 2003 Version 2.01 (ISO 15000-5:2014)

URL：[語彙ライブラリ]「Core Components Library」

[https://www.unece.org/cefact/codesfortrade/unccl/ccl\\_index.html](https://www.unece.org/cefact/codesfortrade/unccl/ccl_index.html)

[技術仕様]「Core Components Technical Specification」

（UN/CEFACT での公開資料）

[https://www.unece.org/fileadmin/DAM/cefact/codesfortrade/CCTS/CCTS\\_V2-01\\_Final.pdf](https://www.unece.org/fileadmin/DAM/cefact/codesfortrade/CCTS/CCTS_V2-01_Final.pdf)

（ISO での公開資料）

<https://www.iso.org/standard/61433.html>

- **IMI 共通語彙基盤 コア語彙**

氏名、住所、組織等、あらゆる社会活動で使用される中核的な用語の集合。

多くのシステム間で情報交換のための基礎となる語彙で、データ連携、オープンデータの二次利用等の効率化に役立つ。

国内において、社会活動に関連したオープンデータ表現の標準として活用されている。

（参照情報）

制定団体：情報共有基盤推進委員会

名称：IMI 共通語彙基盤 コア語彙 バージョン 2.4.2

URL：<https://imi.go.jp/core/core242/>

- Schema.org

Web 上にあるコンテンツなどにメタデータを与え、検索エンジンなどがコンテンツとして記述された情報の意味を認識しやすくするためのタグ付けに使う用語。

HTML 文書に対し、例えば、書籍について記述した HTML で、各タグのテキストについて、「ここは著者を表している」「ここは編集者を表している」「ここは出版者を表している」「ここは初版出版の日付を表している」「ここは編集された日付を表している」など、具体的な意味を細かくタグ付けによって示すことができる。Schema.org では、RDF Schema をベースとしたデータモデルを使っている<sup>43</sup>。

Schema.org は、オープンなコミュニティであり、仕様策定については、2015 年 4 月からは、W3C の Schema.org Community Group<sup>44</sup>が主な議論の場となっているが、具体的な議論のやり取りは、メーリングリスト public-schemaorg@w3.org を使って行われている。また、GitHub 上で、検討課題 (Issue) が管理されている。Schema.org の活動は、Google、Microsoft、Yahoo、Yandex からの出資を受けている。

- public-schemaorg@w3.org のメールのアーカイブ：

<https://lists.w3.org/Archives/Public/public-schemaorg/>

- GitHub：

<https://github.com/schemaorg/schemaorg/issues>

(参照情報)

制定団体：Schema.org

名称：Schema.org

URL：<https://schema.org/>

---

<sup>43</sup> <https://schema.org/docs/datamodel.html>

<sup>44</sup> <https://www.w3.org/community/schemaorg/>

# 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  
データ活用社会の実現へ向けた取組みの手順と実例

## ■ 公開履歴

- ・ ワーキングドラフト 2019年12月20日
- ・ 初版 2020年3月27日
- ・ 第二版 2021年3月25日

## ■ 初版及びワーキングドラフト監修・協力 (50音順,所属は公開時のもの)

- ・ (監修) IMI 検討部会
  - 主査 大槻文彦 株式会社富士通研究所
  - 委員 篠井大祐 一般社団法人XBRL Japan
  - 伊藤昭仁 一般社団法人インダストリアル・バリューチェーン・イニシアティブ
  - 柏崎吉一 合同会社エクリュ
  - 北原圭 株式会社日立製作所
  - 佐伯芳教 SBテクノロジー株式会社
  - 畠山暖央 農林水産省
  - 守屋貴治 SBテクノロジー株式会社
  - 山田隆之 一般社団法人データ流通推進協議会
  - 山本暖 株式会社日立製作所
  - 鷲尾傑 富士通株式会社
  - オブザーバー 経済産業省
  - 事務局 独立行政法人情報処理推進機構
- ・ (事例協力) 岩田孝一 日本電気株式会社
- ・ (編集協力) 柏崎吉一 合同会社エクリュ

## ■ 第二版監修・協力 (50音順,所属は公開時のもの)

- ・ (監修協力) 大槻文彦 株式会社富士通研究所
- ・ (事例協力) 岩田孝一 日本電気株式会社
- 加藤百合子 やさいバス株式会社
- 畠山暖央 農林水産省
- ・ (編集協力) 奥井康弘 株式会社ティージェイ総合研究所
- 柏崎吉一 合同会社エクリュ

- ・ (監修) 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)  
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

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。