

# セマンティックレイヤーをつなぐ (データ辞書)

氏名	利光 孝文
氏名ふりがな	としみつ たかふみ
所属・役職	IPA/DISC/DE部 データスペースグループ・研究員
略歴	物理学者、アプリケーションエンジニア、インフラ営業、コンサルタント、クライアント営業など (IBM, VMware, Accenture, NEC) を経て、研究員として日本電気株式会社から出向中。デジタル庁におけるアドレスベースレジストリの調査研究での活動を機に、現在、データモデリングや相互運用性などに関するデータ標準関連を取り扱う。

講演タイトル	セマンティックレイヤーをつなぐ (データ辞書)
講演概要 / 特に伝えたいポイント・メッセージ	言語処理に長けた人工知能の台頭により、海外 (主に欧州) はセマンティック (語彙) 基盤の重要性が再認識されている。特に生成AIとの関係性を意識する内容が多かった。また、日本においてメタデータ/レポジトリと呼称されるものに相当するデータ辞書を作成する取り組みを行った。これは、生成AIと国際標準規格 (ISO/IEC11179) を用いている。海外動向とデータ辞書作成の知見から今後の課題と考えうる打ち手を共有したい。

# セマンティック？ データ辞書？

semantic : 意味の、語義の、意味論の  
Layer : 層、階層

セマンティック？データ辞書？って、なんだ？？？なぜ必要？

## 講演タイトル

セマンティックレイヤーをつなぐ（データ辞書）

## 講演概要／特に伝えたいポイント・メッセージ （200字程度）

言語処理に長けた人工知能の台頭により、海外（主に欧州）はセマンティック（語彙）基盤の重要性が再認識されている。特に生成AIとの関係性を意識する内容が多かった。また、日本においてメタデータ/レポジトリと呼称されるものに相当するデータ辞書を作成する取り組みを行った。これは、生成AIと国際標準規格（ISO/IEC11179）を用いている。海外動向とデータ辞書作成の知見から今後の課題と考える打ち手を共有したい。



ChatGPT5.4にて作成

## バベルの塔

旧約聖書の「創世記」中に登場する巨大な塔



ChatGPT5.4にて作成

全ての地は、同じ言葉と同じ言語を用いていた。東の方から移動した人々は、シニアルの地の平原に至り、そこに住みついた。そして、「さあ、煉瓦を作ろう。火で焼こう」と言い合った。彼らは石の代わりに煉瓦を、漆喰の代わりにアスファルトを用いた。そして、言った、「さあ、我々の街と塔を作ろう。塔の先が天に届くほどの。あらゆる地に散って、消え去ることのないように、我々の為の名をあげよう」。主は、人の子らが作ろうとしていた街と塔を見ようとしてお下りになり、そして仰せられた、「なるほど、彼らは一つの民で、同じ言葉を話している。この業は彼らの行いの始まりだが、おそらくこのこともやり遂げられないこともあるまい。それなら、我々は下って、彼らの言葉を乱してやろう。彼らが互いに相手の言葉を理解できなくなるように」。主はそこから全ての地に人を散らされたので、彼らは街づくりを取りやめた。その為、この街はバベルと名付けられた。主がそこで、全地の言葉を乱し、そこから人を全地に散らされたからである。

—「創世記」11章1-9節

解釈：

バベルの塔の物語は、「人類が塔をつくり神に挑戦しようとしたので、神は塔を崩した」という解釈が一般に流布している。しかし『創世記』では、「塔を建てるのをやめた」と記述しており、「崩れた」とは書かれていない。

比喻として：

天にも届く神の領域まで手を伸ばす塔を建設しようとして、崩れてしまった（神に壊された）という故事にちなんで（ただし聖書内では先述したように建設を中断したという記述にとどまる）、空想的で実現不可能な計画の比喻としても用いられる。

## Gartner



Gartner Announces the Top Data & Analytics Predictions  
SYDNEY, Australia, June 17, 2025

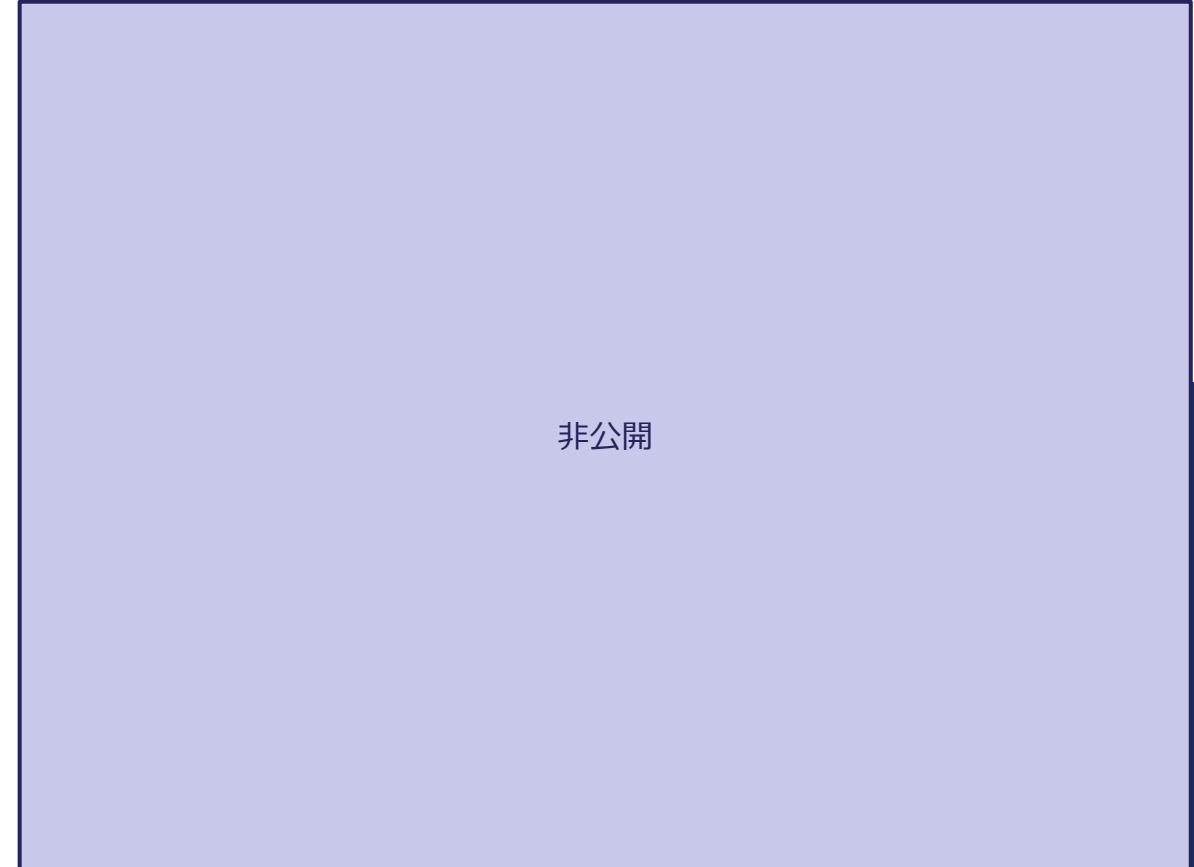
**2027年までに、AI-ready dataにおけるセマンティクスを優先順位付けする組織は、GenAIモデルの精度を最大80%向上させ、コストを最大60%削減する。**

GenAIにおける不適切なセマンティクスは、誤認の増加、不必要なトークンの増加、コスト増大招く。データ管理を見直し、能動的メタデータに焦点を当てる組織は、モデルの精度と効率性を高め、AIデータ対応力を向上させ、コンピューティングコストを削減できる。ガートナーによれば、これによりAIエージェントはより効果的に動作し、組織全体でより賢く迅速な意思決定を促進する。

別の公式発表<sup>\*1</sup>でも、data fabric は active metadata と semantic models を使って、consistent で trusted なデータへのリアルタイムアクセスを可能にすると説明している。「語彙・意味づけ・メタデータ」は、AI 活用の土台だとしている。

\*1 : Gartner Says Chief Supply Chain Officers Can Scale AI With Data Fabric Architecture (STAMFORD, Conn., October 14, 2025)

## IDC



組織の最重要のデータ投資分野として **52% が “data intelligence”** を挙げており、その中身に data quality, cataloging, lineage, metadata, master data が含まれている。さらに将来像の図の一案でも catalogs や semantics が主要構成要素に置かれている。

\*2 : [https://www.idc.com/wp-content/uploads/2025/09/DIR2025\\_TECHC\\_FutureData\\_SB.pdf](https://www.idc.com/wp-content/uploads/2025/09/DIR2025_TECHC_FutureData_SB.pdf)

# 相互運用性（Interoperability）の重要性

## OECD（経済協力開発機構）



OECD (2025), Enhancing Access to and Sharing of Data in the Age of Artificial Intelligence

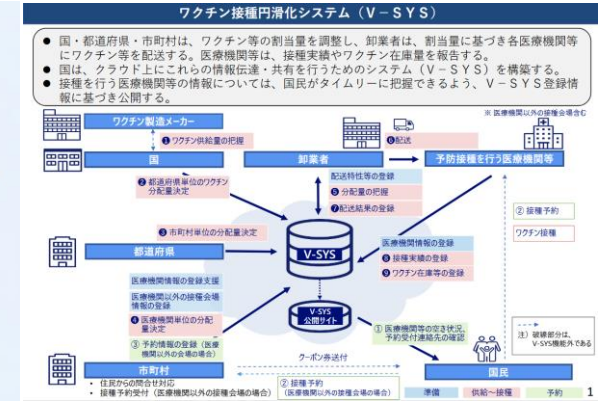
政府は、生成型AIを含むAIの台頭に伴い、データ（「FAIR」データ）の発見可能性、アクセス可能性、相互運用性、再利用性、そしてデータ品質の重要性が高まっていることから、積極的に推進している。この傾向は、公共部門、医療部門、そして機械可読で相互運用可能なメタデータの需要が高まっている科学研究の分野で特に顕著である。この文脈における政策措置には、公共部門、医療部門、研究データの検証、統合、および／または公開を行うための専用のデータプラットフォームとレポジトリの構築が含まれる。一部の国では、FAIRデータに焦点を当てた新しい機関を設立したり、国立統計局などの既存の機関の役割を拡大して信頼できるデータ仲介機関（TDI）として機能させたりしている。AIモデルを「オープンソース」として、さまざまなレベルのオープン性で公開するという傾向も、データとAIモデルの両方の発見可能性、アクセス可能性、相互運用性、再利用性の確保について疑問を投げかけている。

AIに必要なデータは FAIR（findable, accessible, interoperable, reusable : 発見可能性、アクセス可能性、相互運用性、再利用性）であるべきだ、とかなり明確に記載されている。

## 日本：COVID-19（V-SYS）



ChatGPT5.4にて作成



新型コロナウイルスワクチンの接種体制確保に係る自治体説明会（第1回）  
<https://www.mhlw.go.jp/content/10906000/000707430.pdf>

### 医療機関コード（10桁コード）

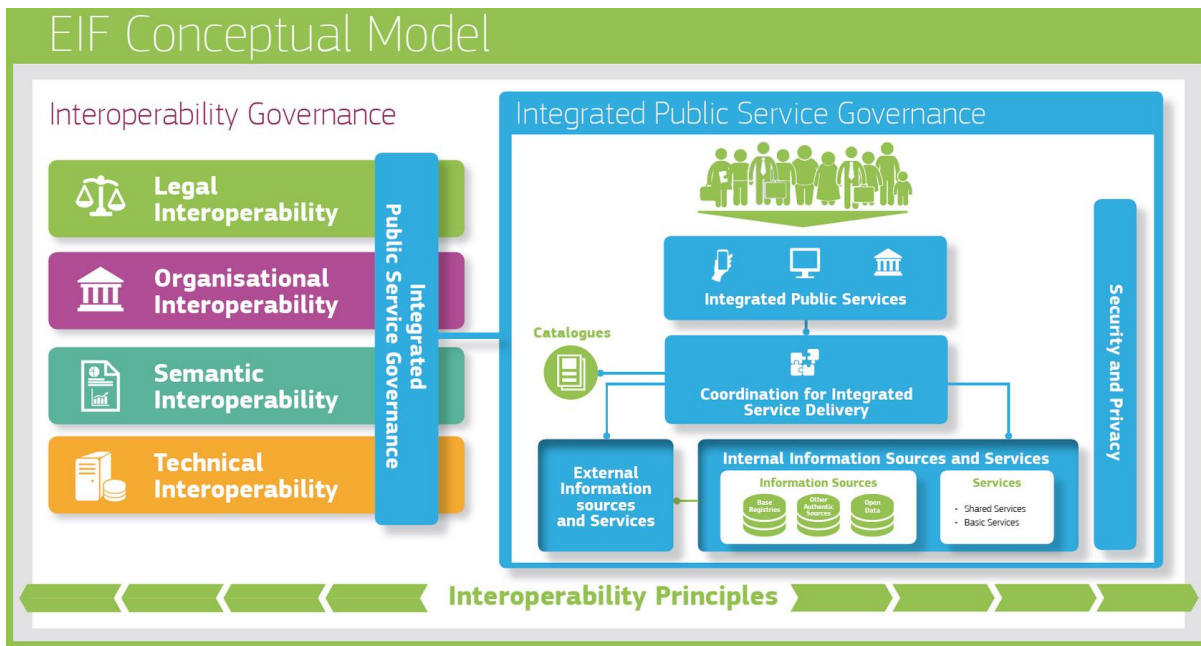
- 全国の医療機関を一意に識別するための最も標準的なコード
- 診療報酬請求や行政管理、医療機関データベースなどで利用され、病院・診療所・薬局・歯科医院などに適用されます。7桁のコードも存在し、地域ごとの管理目的で使用されることが多く、都道府県や自治体が医療機関情報を整理するために活用されている。

### V-SYS:ワクチン接種円滑化システム

約8万ほどあるコードは、各地方局、都道府県別で管理されていたため、全国統一的に拠点を把握することが難しく、また臨時施設も必要になっていたため、手動かつ人の手を介して、意思疎通・連携を行う必要があった。

コロナ禍のような緊急事態、被災などの状況において、適切な医薬品の配達が困難になるため、流通の拠点、物品などのリソース（ヒト、運送機能、経路把握も必要）など緊急時以外においても、事前にFAIR原則に沿ったデータ整備として、Semantic Layerの連携確保が必要である。

## EIF : European Interoperability Framework



The new EIF embodied in the Ministerial Declaration on e-Government as a guideline to the digitalization of public administrations  
Friday, 6 October, 2017

2017年に改訂された欧州相互運用フレームワーク（EIF）が、EU・EFTA32か国の電子政府担当相によるタリン宣言（タリン：エストニア首都）に盛り込まれ、**行政デジタル化の指針となった**ことを伝える内容である。EIFは47の提言を通じて、国境を越えた相互運用性と、開かれた・効率的で利用者本位の公共サービス実現を後押しすると位置づけられている。  
※EFTA:欧州自由貿易連合（4か国）

データの取扱いのみに限らず、法令、組織、セマンティック、技術の4つの階層での横断的なガバナンスを設けるフレームワークを掲示している。明示的に“Semantic”が含まれている。

## SEMIC2025



Towards the next generation of the European Interoperability Framework (EIF) SEMIC2025、パネルディスカッション

2017年版EIFは、相互運用性という共通言語を広め、各国の制度整備を後押ししてきた。しかし、生成AI、データスペース、デジタルウォレットなど**新たな環境の中で、EIFは原則集ではなく、現場で使え成果を測れる実践的な道具へ進化する必要**がある。課題は技術だけでなく、自治体で優先されにくい現実や、言語・文化・スキル不足にもあるため、国・自治体・民間の協働や共通基盤が鍵となる。デンマークやエストニアの事例も、その方向性を示した。なお、今後は、相互運用性への投資理由を政治的に説明し、データ保護やAI透明性も織り込んだうえで、**作って公開して終わりにしない継続運用とワーキンググループによる伴走**が求められる。

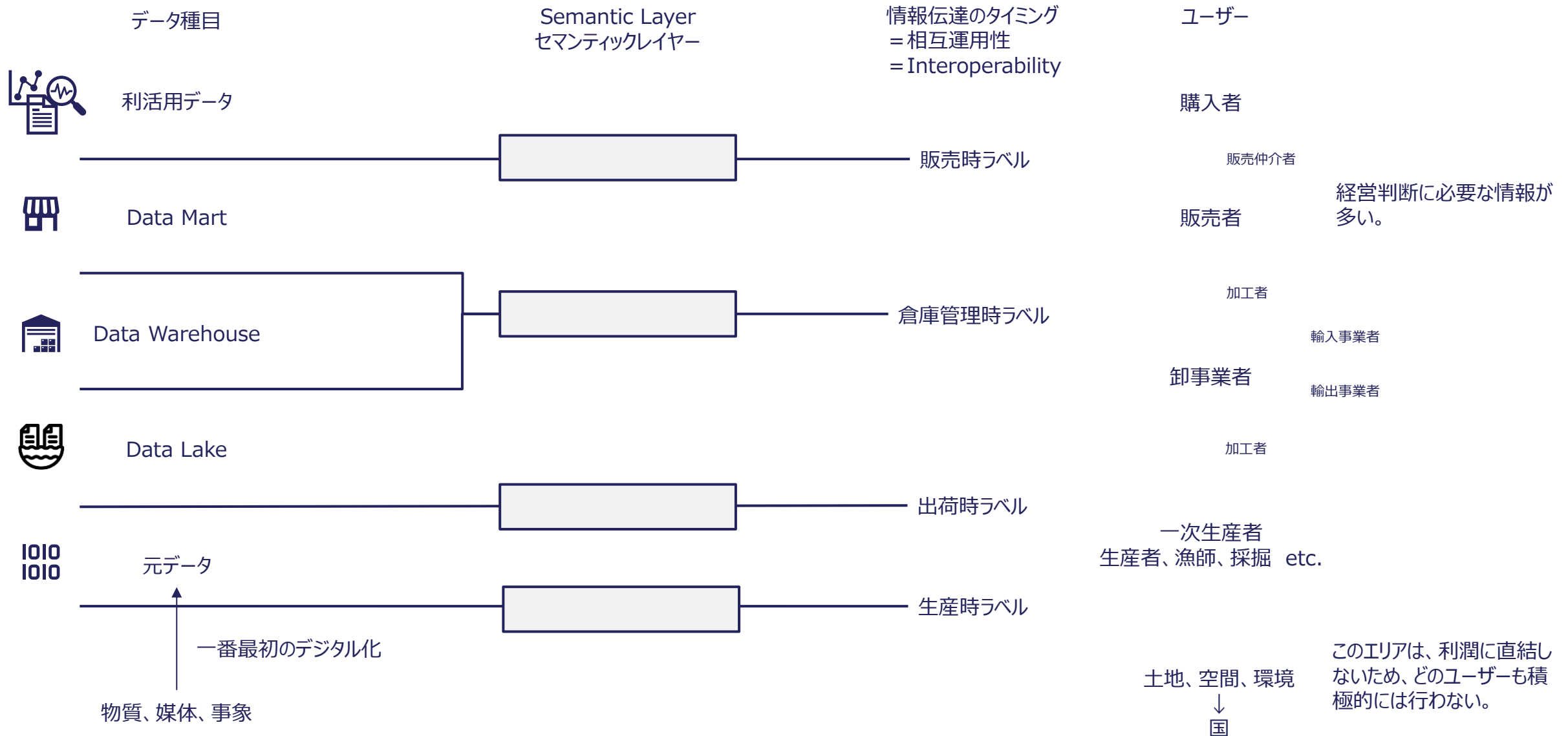
毎年イベントが継続的に実施されており、フレームワークの浸透やその取り組みなどを、継続的に共有し、毎年、改善され続けている。

## 対象セマンティックレイヤー&データ辞書

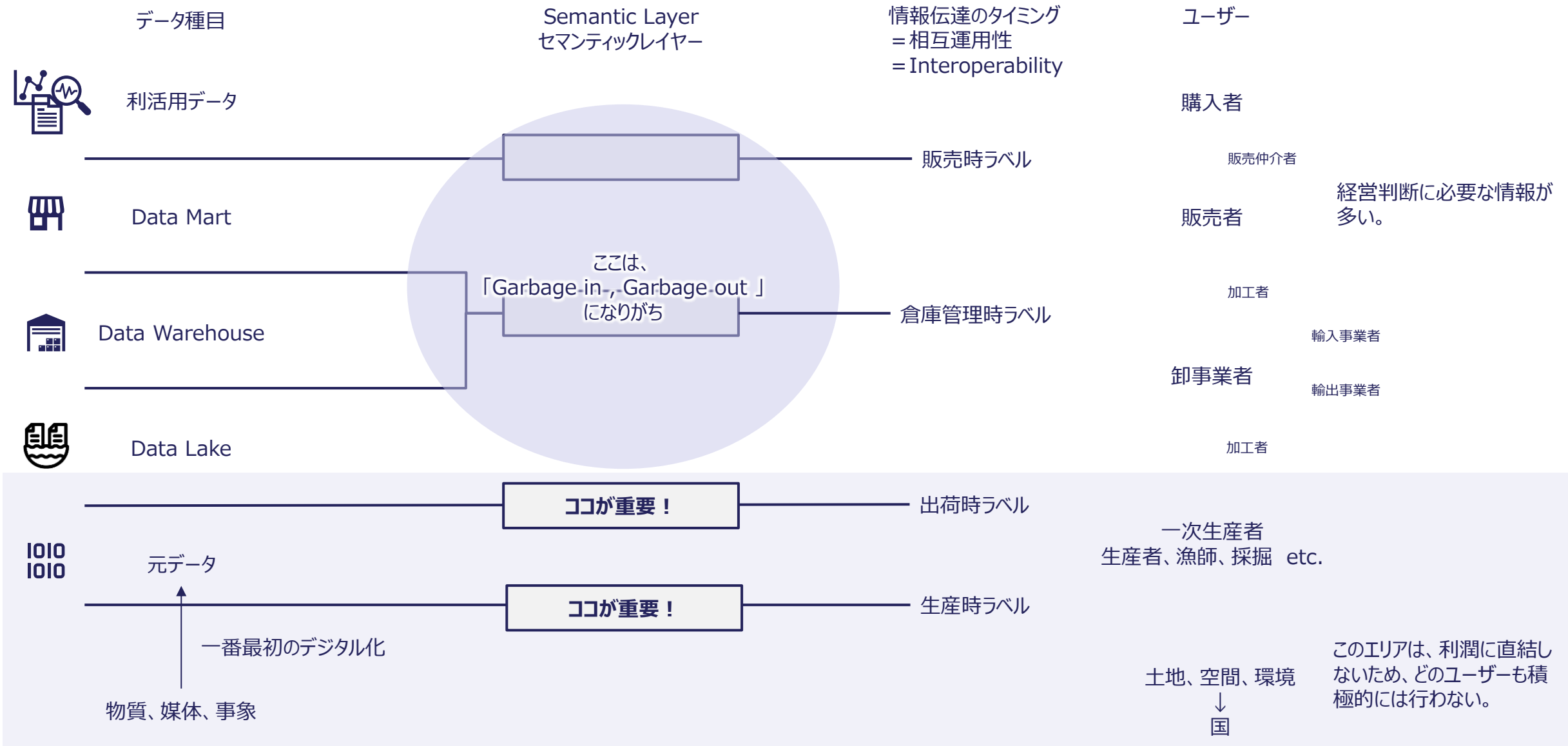
取り組みの対象、データ辞書の定義ってなに？

講演タイトル	セマンティックレイヤーをつなぐ（データ辞書）
講演概要／特に伝えたいポイント・メッセージ （200字程度）	言語処理に長けた人工知能の台頭により、海外（主に欧州）はセマンティック（語彙）基盤の重要性が再認識されている。特に生成AIとの関係性を意識する内容が多かった。また、日本においてメタデータ/レポジトリと呼称されるものに相当するデータ辞書を作成する取り組みを行った。これは、生成AIと国際標準規格（ISO/IEC11179）を用いている。海外動向とデータ辞書作成の知見から今後の課題と考える打ち手を共有したい。

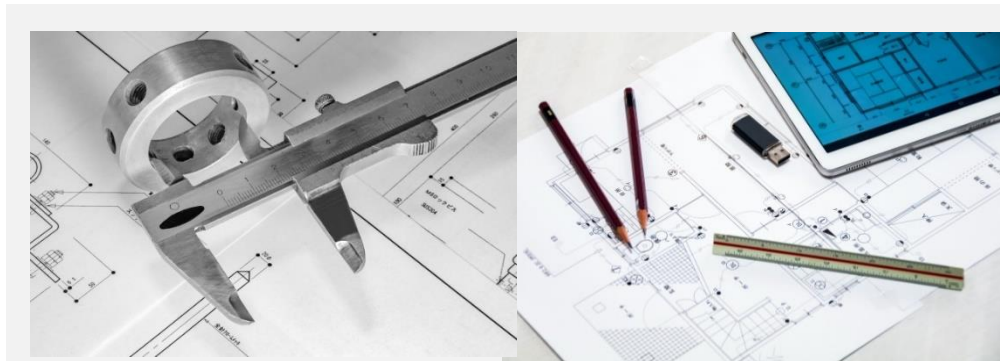
# どこのセマンティックレイヤーか？



# どこのセマンティックレイヤーか？ - 注目するレイヤー



- GIFは、政府相互運用性フレームワーク（Government Interoperability Framework）の略称であり、政府情報システムが保有するデータの相互運用性を確保し、効率的な情報共有と情報連携を実現するためのデータモデルをひな形としてまとめたもの。
- 「デジタル社会推進標準ガイドライン」中で、実践ガイドブックに位置付けられている **ドキュメント** です。



建物や製品や作るには設計の記法、取引ルールは、100年以上前から確立しています。そして、モデルや部品等が用意されてきました。このため、大量生産、継続的改善やイノベーションができてきました。



デジタル社会のための設計の記法は、発展途上です。GIFは、デジタル社会のデータモデル等を提供するものです。サービスの広域展開や持続、発展をさせるためには欠かせない基盤です

# データ項目説明文の当時の状況 (1/2)

No.	プレフィクス	項目名	必須 (オン1) 推奨:高	必須 (オン2) 推奨:中	必須 (オン3) 推奨:中	最大 回数	項目名 (英語)	説明	形式	記入例	名称・形式等の参考元
1	pd	ID群	必	必		1	idGroup	個人番号、教員番号など	pd:ID情報型	ID情報型のデータモデルを格納 例)	IME
2	pd	氏	必	必	必	1	familyName	氏	文字列	山田	schema.org
3	pd	名	必	必	必	1	givenName	名 (ミドルネームを有している場合は、区切り文字として半角空白(*2)を使用した上でミドルネームを記載してもよい。例: ミドルネーム+区切り文字+名)	文字列	花子	schema.org
4	pd	氏(カナ)	必	必	必	1	familyNameKana	氏のカナ表記	文字列 (全角カナ)	ヤマダ	IME
5	pd	名(カナ)	必	必	必	1	givenNameKana	名のカナ表記	文字列 (全角カナ)	ハナコ	IME
6	pd	氏(英字)	必			1	familyNameEn	氏のローマ字表記	文字列 (半角英数字)	Yamada	
7	pd	名(英字)	必			1	givenNameEn	名のローマ字表記	文字列 (半角英数字)	Hanako	
8	pd	氏名				1	fullName	氏名 (氏、名のセット)、区切り文字は半角空白(*2)とする	文字列	山田 花子	IME
9	pd	氏名(カナ)				1	fullNameKana	氏名 (氏、名のセット) のカナ表記、区切り文字は半角空白(*2)とする	文字列 (全角カナ+区切り文字; 半角空白)	ヤマダ ハナコ	IME
10	pd	氏名(英字)				1	fullNameEn	氏名 (氏、名のセット) のローマ字表記	文字列 (半角文字)	Yamada Hanako	
11	pd	国籍				1	nationality	国籍名	pcd:国コード一覧 (コード、名称) 例)	コード一覧のコードと名称の構造を格納	
12	pd	性別			必	1	sex	性別名	pcd:性別コード一覧 (コード、名称) 例)	コード一覧のコードと名称の構造を格納	
13	pd	生年月日		必		1	dateOfBirth	生年月日	日付 (YYYY-MM-DD)	2000-01-01	
14	pd	死亡年月日		1		1	dateOfDeath	死亡している場合、死亡年月日	日付 (YYYY-MM-DD)	2000-01-01	
15	pd	備考				1	remark	その他特筆事項	文字列		
16	pd	住所 (*1)		必		1	address (*1)	居住住所情報	pd:住所 コアデータモデル: 住所のデータモデルを格納		IME
17	pd	連絡先情報				1	contactPointInformation	連絡先住所の情報	pd:連絡先 コアデータモデル: 連絡先のデータモデルを格納		schema.org

項目名	説明
ID群	個人番号、教員番号など
氏	氏
名	名 (ミドルネームを有している場合は、区切り文字として半角空白(*2)を使用した上でミドルネームを記載してもよい。例: ミドルネーム+区切り文字+名)
氏(カナ)	氏のカナ表記
名(カナ)	名のカナ表記
氏(英字)	氏のローマ字表記
名(英字)	名のローマ字表記
氏名	氏名 (氏、名のセット)、区切り文字は半角空白(*2)とする
氏名(カナ)	氏名 (氏、名のセット) のカナ表記、区切り文字は半角空白(*2)とする
氏名(英字)	氏名 (氏、名のセット) のローマ字表記
国籍	国籍名
性別	性別名
生年月日	生年月日
死亡年月日	死亡している場合、死亡年月日
備考	その他特筆事項
住所 (*1)	居住住所情報
連絡先情報	連絡先住所の情報

## コアデータモデル: 個人

説明文は、どちらかというと補足や備考に近い位置づけの記載となっており、記入例などを示していた。

# データ項目説明文の当時の状況 (2/2)

No.	プレフィックス	項目名	必須 ターン1 注:高	必須 ターン2 注:中	必須 ターン3 注:中	最大 回数	項目名 (英語)	説明	形式	記入例	名称・形式等の参考元
1	pd	法人番号	必	必		1	legalEntityNumber	法人番号 (13桁) を記載。法人番号が指定されている場合は必須。	文字列 (半角数字)	8000012010038	
2	pd	ID群	必		必	n	idGroup	法人に付与した一意のIDを記載。インボイス登録番号 (登録番号) が指定されている場合は必須。	d:ID情報型	ID情報型のデータモデルを格納例) 00000000000000000000	DM1
3	pd	商号または名称	必	必	必	1	tradeName	商号または名称を記載。英字は全角英字とする。	文字列	株式会社 G I F	
4	pd	商号または名称(カナ)	必	必	必	1	tradeNameKana	名称をカナで記載。	文字列 (全角カナ)	カシキガイシャアイエフ	
5	pd	商号または名称(英字)	必	必	必	1	tradeNameEn	名称を英語で記載。	文字列 (半角英字)	GIF Co., Ltd.	
6	pd	組織種別	必	必	必	1	categoryOfOrganization	組織種別 (株式会社、有限会社、合名会社、合資会社、合同会社等) を記載。	d:法人種別コード一覧 (コード)	301	DM1
7	pd	通称				n	alternateName	通称や略称アルファベットを記載。	文字列	ジーアイエフ	schema.org
8	pd	事業活動状況				1	activityStatus	活動状況 (倒産、破産、休眠、休業、廃業等) を記載。	文字列	営業中	DM1
9	pd	説明				1	description	説明を記載	文字列	株式会社GIFは情報連携サービスの提供を中心としたDX事業を推進する企業です。	DM1
10	pd	WebサイトURL				1	webSiteUrl	法人に関する情報源を示すWebサイトURLを記載。	文字列	https://www.xxx.com	schema.org
11	pd	社員数				1	numberOfCompanyEmployee	社員数 (役員数を含む) を記載。	数値 (半角数字)	1550	
12	pd	代表者 (*1)				1	representative (*1)	代表者の情報を記載。	d:個人	山田太郎 (コアデータモデル: 個人) のデータモデルを格納	
13	pd	代表者役職名				1	positionOfRepresentative	代表者役職名を記載。	文字列	代表取締役	
14	pd	設立年月日				1	dateOfEstablishment	設立日を記載。	日付 (YYYY-MM-DD)	1990-04-01	DM1
15	pd	事業種目				n	categoryOfBusiness	事業種目 (日本標準産業分類) を記載。	文字列	3921 情報処理サービス業	DM1
16	pd	登記住所 (*1)				1	registeredAddress (*1)	登記住所を記載。	d:住所	コアデータモデル: 住所のデータモデルを格納	DM1
17	pd	事業所情報				n	businessPlaceInformation	事業所情報を記載。	d:事業所	コアデータモデル: 事業所のデータモデルを格納	DM1
18	pd	連絡先情報	必	必	必	1	contactPointInformation	連絡先情報を記載	d:連絡先	コアデータモデル: 連絡先のデータモデルを格納	schema.org

項目名	説明
法人番号	法人番号 (13桁) を記載。法人番号が指定されている場合は必須。
ID群	法人に付与した一意のIDを記載。インボイス登録番号 (登録番号) が指定されている場合は必須。
商号または名称	商号または名称を記載。英字は全角英字とする。
商号または名称(カナ)	名称をカナで記載。
商号または名称(英字)	名称を英語で記載。
組織種別	組織種別 (株式会社、有限会社、合名会社、合資会社、合同会社等) を記載。
通称	通称や略称アルファベットを記載。
事業活動状況	活動状況 (倒産、破産、休眠、休業、廃業等) を記載。
説明	説明を記載
WebサイトURL	法人に関する情報源を示すWebサイトURLを記載。
社員数	社員数 (役員数を含む) を記載。
代表者 (*1)	代表者の情報を記載。
代表者役職名	代表者役職名を記載。
設立年月日	設立日を記載。
事業種目	事業種目 (日本標準産業分類) を記載。
登記住所 (*1)	登記住所を記載。
事業所情報	事業所情報を記載。
連絡先情報	連絡先情報を記載。

- コアデータモデル : 個人

説明文は、どちらかという補足や備考に近い位置づけの記載となっており、記入例などを示していた。

# データ辞書とは、

データ辞書（Data Dictionary System）とは、  
「組織内外で使用/再利用されるデータについてのメタデータ（データを説明するデータ）を一元管理するための構造化されたレポジトリ」と定義します。

※レポジトリ：ソフトウェア開発プロジェクトにおけるソースコード、ドキュメント、設定ファイルなど、様々な関連ファイルを一元管理する場所。本来は、バージョン管理システムと連携し、変更履歴の追跡や複数人での共同作業を効率的に行うための仕組みとして利用されることが好ましい。

データ辞書であるレポジトリは、人と機械をつなぐ情報群で構成されるものであり、以下のような体系において、構成されると定義します。

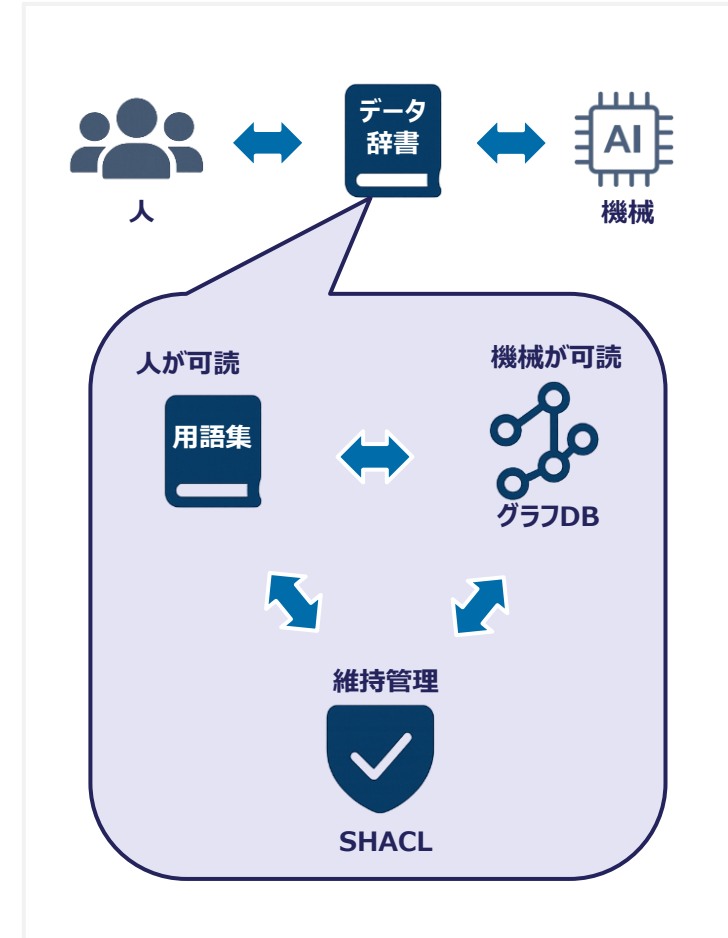
「組織内外で再利用するデータ項目について、

**用語（人が可読）**、

オントロジー記述（機械が可読）、

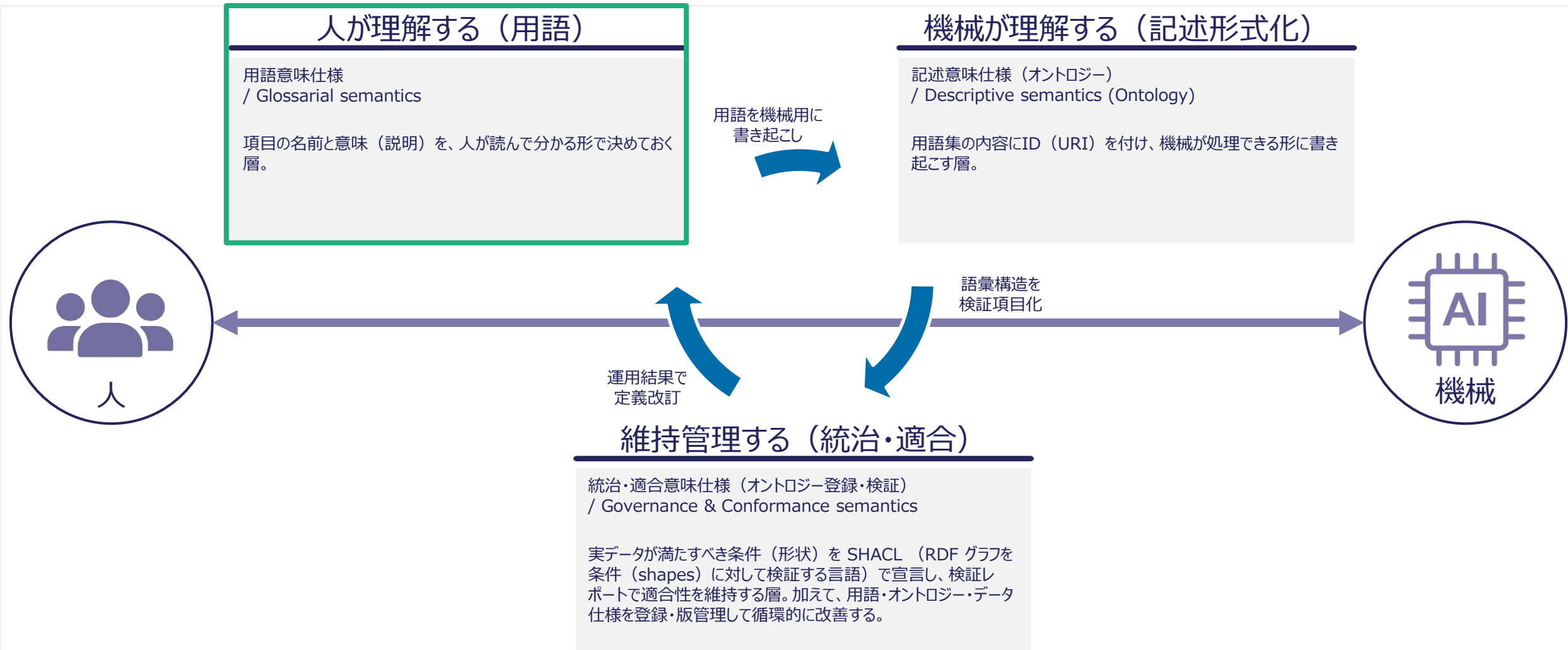
適合検証/登録管理（維持管理）

を統合的に扱う体系」



図表：データ辞書のイメージ

「組織内外で再利用するデータ項目について、用語（人可読）、オントロジー記述（機械可読）、適合検証/登録管理を統合的に扱う体系」で構成することにより、**人と機械の情報連携を支援**します。



図表：データ辞書の構成体系

## データ辞書の作成

どうやって作成したの？

講演タイトル	セマンティックレイヤーをつなぐ（データ辞書）
講演概要／特に伝えたいポイント・メッセージ （200字程度）	言語処理に長けた人工知能の台頭により、海外（主に欧州）はセマンティック（語彙）基盤の重要性が再認識されている。特に生成AIとの関係性を意識する内容が多かった。また、日本においてメタデータ/レポジトリと呼称されるものに相当するデータ辞書を作成する取り組みを行った。これは、生成AIと国際標準規格（ISO/IEC11179）を用いている。海外動向とデータ辞書作成の知見から今後の課題と考える打ち手を共有したい。

メタデータレジストリで指定されるデータ定義の定式化に関する要件と推奨事項を指定しています。これらを定義する目的は、データの意味を指定、記述、説明を明確にし、データ要素の標準化または再利用を促進し、データの共有と情報システムの統合を促進することです。

必須項目：5つ

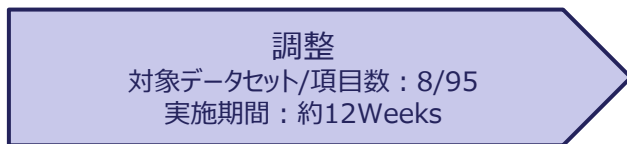
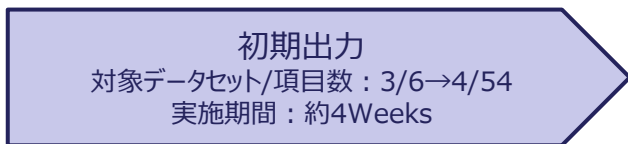
必須 shall	要件
a	単数形で記述される説明 be stated in the singular
b	概念が何であるかを述べる、それが何でないかだけでない説明 state what the concept is, not only what it is not
c	説明的なフレーズまたは文として記述される（ほとんどの言語で）説明 be stated as a descriptive phrase or sentence(s) (in most languages)
d	一般的に理解されている略語のみを含む説明 contain only commonly understood abbreviations
e	他のデータまたは基礎なる概念の定義を埋め込まずに表現する説明 be expressed without embedding definitions of other data or underlying concepts

- 前述のとおり、データの意味を“文”として記述することがおろそかにされがちであるため、「人が理解する」との観点から、再注目されるものとして、選択しています。
- ISO/IECの製造、金融に関連する規格の定義において、一部、参照されています。
- 日本における小委員会（SC32/WG2）は、2023年に解散しており、現在、日本からの検討発信はない様子です。  
※独自にワーキンググループを設置し、当時の小委員会主査に協力要請を行いました。

推奨項目：8つ

推奨 should	要件
a	概念の本質的な意味を述べる説明 state the essential meaning of the concept
b	正確かつ明確な説明 be precise and unambiguous
c	簡潔な説明 be concise
d	独立できる説明 be able to stand alone
e	理論的根拠、機能的用法、ドメイン情報、または手続き型情報を埋め込まずに表現する説明 be expressed without embedding rationale, functional usage, domain information, or procedural information
f	循環論法の回避説明 avoid circular reasoning
g	関連する定義に同じ用語と一貫した論理構造を使用する説明 use the same terminology and consistent logical structure for related definitions
h	定義されているメタデータアイテムのタイプに適している説明 be appropriate for the type of metadata item being defined

一定の成果物を作成できるものの、生成AIでは、ランダム性を排除できないため、精度・品質を上げるためには、生成AIからプログラム出力を行い、再利用のできるプログラムベースのモジュールが必須。



## INPUT

- ・ISO/IEC11179-4
- ・データメイン名、項目名

## OUTPUT

指定した語彙の定義文、適合結果・理由

## 指示プロンプト

適合評価をさせるプロンプトを作成し、定義するうえでのルールとレイアウトをプロンプトに追加していった（会議で結果を見てブラッシュアップを繰り返す）

## 成果：

- ・ それらしい説明文の生成を行うことができる。

## 課題・対応

- ・ 項目のドメイン（一つ上のレイヤの情報）が必須であった。
- ・ 全体的な表記ゆれが大きく、また、余計な文言などもみられる。
- ・ すべて適合にする傾向がみられる。  
→上記を2周目で修正。対象拡張。

## INPUT

- ISO/IEC11179（英字原文）
- ※項目対象を4→8データセットへ。

## OUTPUT

定義文、備考、適合結果・理由

## 指示プロンプト

- ・ いい人フィルター外し
- ・ 役割明示
- ・ 固有設定変更（温度、Top-p）
- ・ 公式ガイド（？？？）

## 成果：

- ・ 説明文は問題はない。表現統一は人の手での指示は必須。
- ・ **備考欄の記載がカオスになる。**

## 課題・対応：

- ・ 定義は違和感がないものになったが、**備考に曖昧な記載や不要な情報まで事細かに記述されるようになった。**

## INPUT

試作データ辞書

## OUTPUT

備考文書

## 指示プロンプト

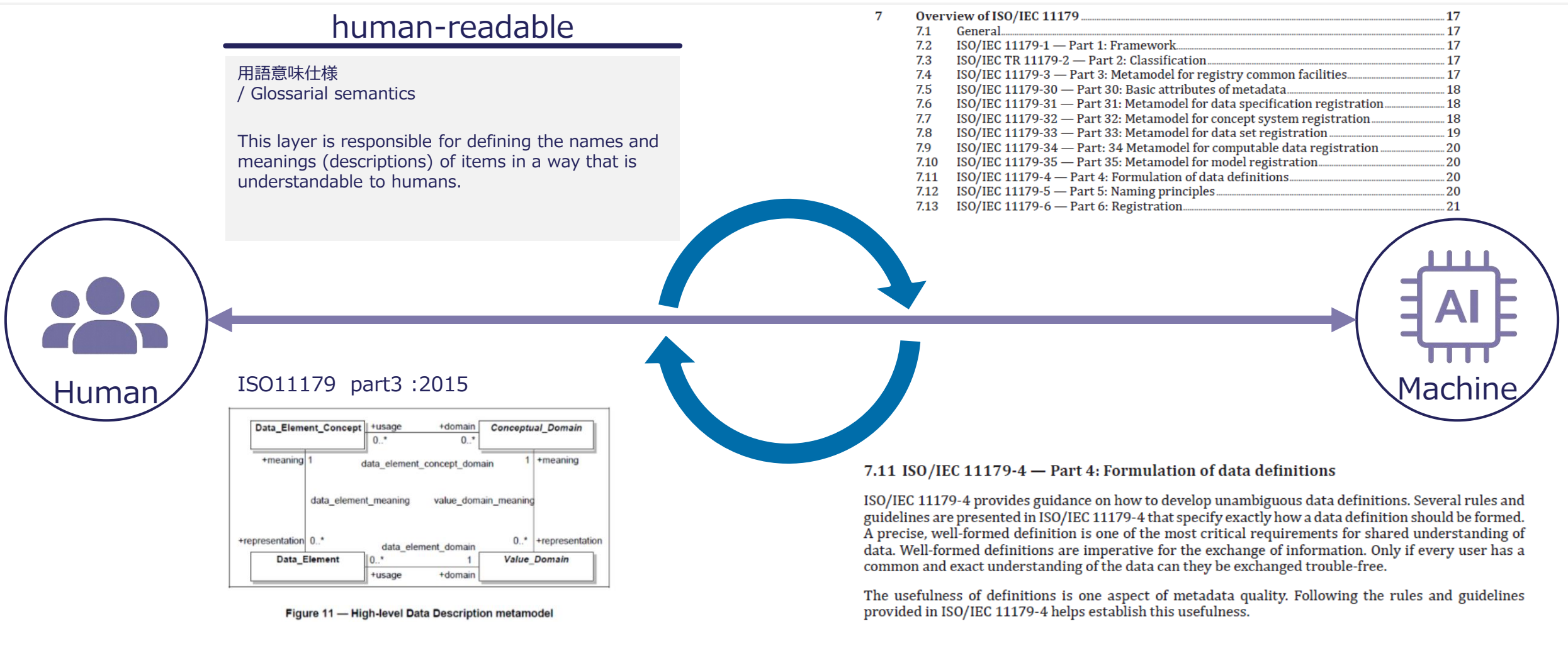
- ・ 表現統一には、「中学生でもわかる」を追加
- ・ 統一用プロンプトを自作はしたものの、最終的な品質が上がらず、別生成AIから作成

## 得られた知見：

- ・ 一定の品質ではできたが、備考欄などは人による手修正で全修正。
- ・ 備考記載事項のルール化が必要。
- ・ そもそもデータ項目がISO/IEC11179にそぐわない項目もある。
- ・ 適合表をどのようにすれば生成AIと親和性の高い取り込みができるのか工夫が必要。
- ・ 「人と機械」の作業分担を明確に位置付ける必要がある。
- ・ モジュール化（ビルディングブロック化）が好ましい。

### 初期出力

ISO11179



# Data dictionary structure:

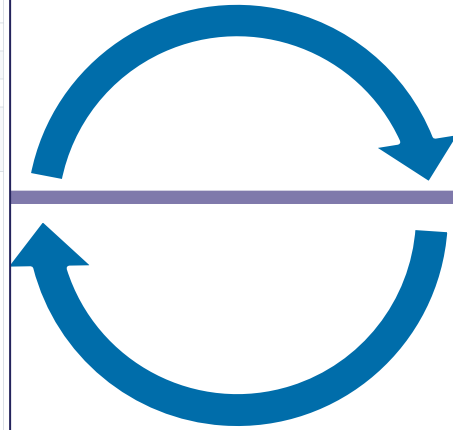
## The data dictionary supports the use of generative AI. @GIF

### 調整

ISO11179 PART4



概念データ要素 (Data Element Concept) : 国籍	
項目	内容
名称	国籍
名称 (英語)	nationality
定義	個人がどの国の国民であるかという法的所属の区分。
概念定義域 (Conceptual Domain)	国家に対する国民としての法的所属 (citizenship) の意味の集合。
備考	<ul style="list-style-type: none"><li>・【範囲】 適用範囲: 国際一般概念。日本の制度に依存する場合は国籍法等の要件に従う。</li><li>・【値域】 本DECは意味のみを規定する。許容値・表記・コードの採用可否 (例: ISO 3166による国名コード、複数国籍や無国籍の扱い) は値域 (Value Domain) で定義する。</li><li>・【値域】 「国籍」は法的地位 (citizenship) を指す概念であり、「出身国」や「居住国」とは異なる。混同を避けるため、実務での区分意図は値域 (Value Domain) 側で明示する。</li><li>・【値域】 国や制度により国籍の取得・喪失や表示方針が異なるため、適用する法制度に依存する事項は値域 (Value Domain) または実装規約で扱う。</li><li>・【参照: 法令】 e-Gov法令検索 (国籍法)。</li><li>・【参照: 規格】 ISO (ISO 3166)。</li><li>・【運用】 英語名は添付ファイルの「個人」シート「項目名 (英語)」への厳密一致要件に従う。現時点は暫定設定であり、提供された語彙に合わせて更新する。</li></ul>



2) ISO/IEC 11179-4 (定義説明文規格) の適合根拠			
No.	必須	適合評価	適合根拠
1	必須-a: 単数形	N/A (日本語) / 英語で担保	日本語は単数の形態的区別がないため判定対象外。英語名 (nationality) の単数形で担保。
2	必須-b: 本質の説明	適合	否定列挙ではなく、概念の中核 (法的所属の区分) を肯定形で特定。
3	必須-c: 説明的フリーズ/文	適合	同義語列挙を避け、簡潔な一文の説明的文で記述。
4	必須-d: 略語の扱い	適合	略語は一般に理解されるもののみ使用し、規格名は正式名称で参照 (ISO=International Organization for Standardization)。
5	必須-e: 他用語の定義を埋め込まない	適合	実務や法制度の詳細は備考/値域 (Value Domain) へ分離し、本文は意味のみを規定。

No.	推奨	適合評価	適合根拠
1	推奨-a: 本質・主特徴を含む	適合	対象 (個人) と主特徴 (どの国の国民であるかという法的所属) を明示。
2	推奨-b: 明確で一義的	適合	「国籍」を「出身国」や「居住国」と混同しないよう範囲を規定。
3	推奨-c: 簡潔	適合	冗長な但し書きを避け、一文で過不足なく表現。
4	推奨-d: 独立可読	適合	外部参照がなくても定義のみで理解可能。
5	推奨-e: 理由・機能・手続・ドメイン情報の非混入	適合	rationale/functional/procedural/domain情報は定義に含めず備考へ分離。
6	推奨-f: 循環論法の回避	条件付き適合	上位概念 (法的所属の区分) で規定し自己語反復を避けているが、関連定義群の横断レビューで最終確定。
7	推奨-g: 用語と文型の一貫性	条件付き適合	本系列の他項目との一貫性は横断レビューで確定 (「〜という法的所属の区分」の定型に整合)。
8	推奨-h: メタデータタイプへの適合	適合	DECは意味のみを扱い、許容値・表記は値域 (Value Domain)、表現実装はデータ要素に委譲。

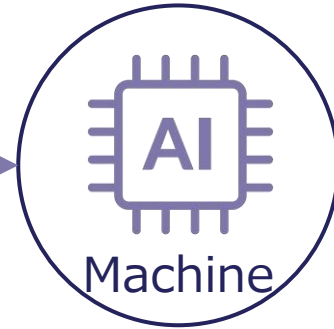


Diagram: Structure of the data dictionary

# Data dictionary structure: The data dictionary supports the use of generative AI. @GIF

## 精度上げ

人力です、、、

The image displays a grid of 40 data dictionary pages, arranged in 4 rows and 10 columns. Each page contains technical details, likely related to data fields and their attributes. In the center of the grid, there is a prominent blue circular icon. The icon features three stylized human figures in purple, with the word "Human" written in black text below them. Two large, thick blue arrows form a circle around the icon, pointing clockwise, which suggests a continuous cycle or process involving human input and output.

## データ辞書の目指すところ

全体的にどうしたい？ 今後は？

講演タイトル	セマンティックレイヤーをつなぐ（データ辞書）
講演概要／特に伝えたいポイント・メッセージ（200字程度）	言語処理に長けた人工知能の台頭により、海外（主に欧州）はセマンティック（語彙）基盤の重要性が再認識されている。特に生成AIとの関係性を意識する内容が多かった。また、日本においてメタデータ/レポジトリと称されるものに相当するデータ辞書を作成する取り組みを行った。これは、生成AIと国際標準規格（ISO/IEC11179）を用いている。海外動向とデータ辞書作成の知見から今後の課題と考える打ち手を共有したい。

# Data dictionary structure:

## The data dictionary supports the use of generative AI. @GIF

An integrated framework linking terminology (human-readable), ontology (machine-readable), and unified validation supports seamless information sharing between people and machines.

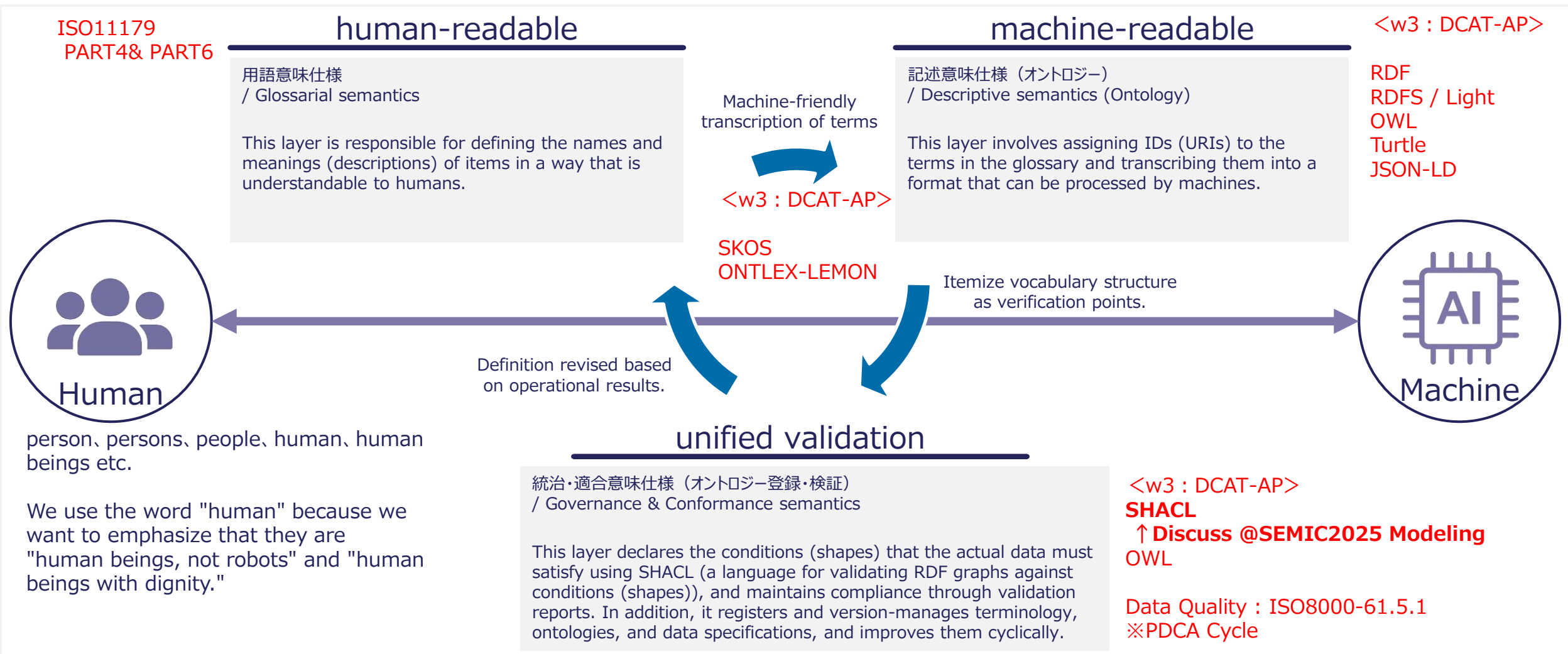


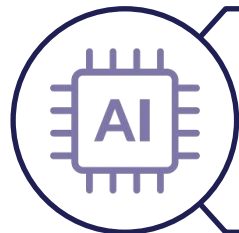
Diagram: Structure of the data dictionary



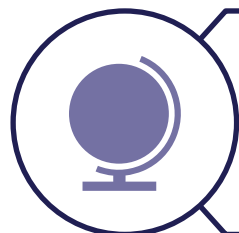
規格・規約・ルール・チェックリストなど、第三者が確認し、その適合を確認する作業を生成AIやプログラムを用いることで、その規格などの商業的価値を創出できる。



規格などの作り自体を機械判読可能な形式にする必要がある。日本語からpythonコードなどのスクリプト言語での生成を行い、適切なチェックが入っているのは可読性とそのテストを人間が行う必要がある。この際、複数の生成AI同士のチェックとしてもよい。



セマンティックレイヤーでは、RDF(OWL etc.), SHACL, などリンクドデータを用いたデータの可視化を並行させ、積極的な利用を進めるのが好ましい。



海外動向に合わせ、SEMIC/NIEM/schema.orgの3つに絞って、連携を毎年確認し、語彙のずれを課題として、提起する。



“推進”機構としての打ち手は、以下を考えているが、いずれも、政策予算に依存する。



## 政策予算次第...

※講演機会、個別企業・組織実践の支援  
情報発信の機会をいただければ、取り組む。

データ品質の観点から  
複数年計画が必要だが、  
現状、未定。

データスペースアカデミーなどの教材への協力  
ArchiMate や UML などの各種モデリング技法を周知できる情報  
発信が必要。

※今回の生成AI利用などの知見の公開。“再利用”を意識。

## Linked DBの構築と公開

Neo4j Auraにて、簡易的にも公開できるデータ構成モデルを作成し、発信が必要。（商用DBの試行も必要）

※簡易的でも動くものを公開。

IPA