

AI社会のためのデータ基盤やエコシステムの整備

2024-10-17

独立行政法人情報処理推進機構（IPA）
デジタル基盤センター長 平本 健二

誰でも簡単にビジネスを開始
デジタル空間の設計、データ供給・蓄積

最先端のビジネスに変革
組織や社会のデジタル改革の実現

誰でもアイデアを実現
革新的技術や人材の創出

データスペース・AI
(データ活用)

デジタル
トランスフォーメーション
(企業や組織のデジタル化)

イノベーション



AI

デジタル基盤

(データ供給、使える仕組み、標準化)

データ

ルール (制度)

ツール

方法論

事例

教材

ソフトウェア・エンジニアリング

(高速化する社会変革に必要なサービス実現方法)

セキュリティ

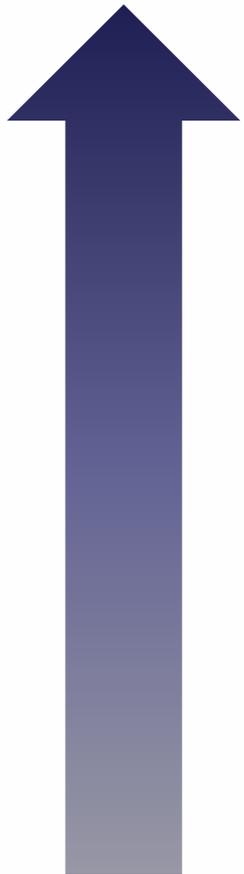
※セキュリティセンター

人材

※デジタル人材センター

AI時代におけるデータ基盤の必要性

データからAIまでは一つのストーリー



AIの本格的な導入

- 個社からデータスペースまで、データの価値をフルに活用

データスペースの実現

- データをアセットや知識として使いこなせる社会・経済圏を構築

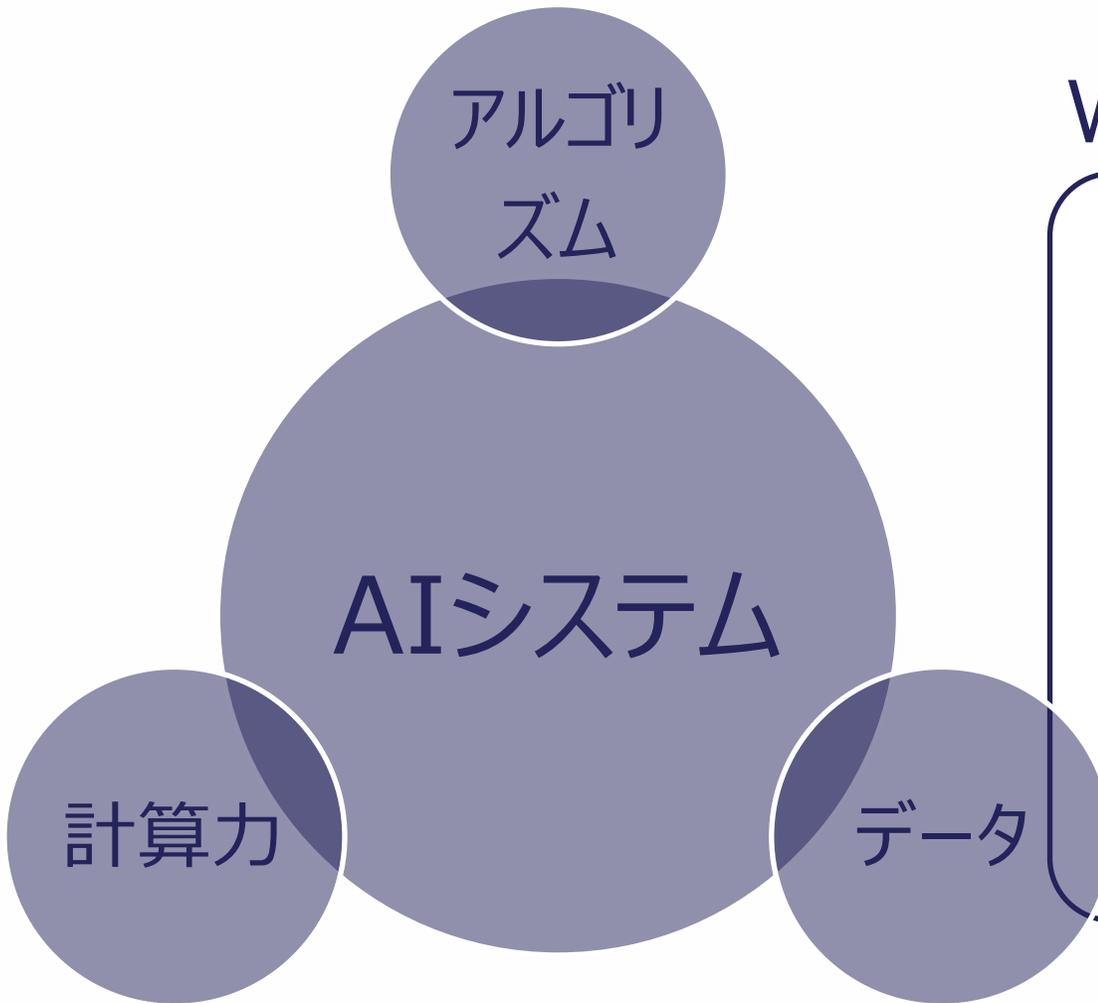
データ連携

- クロスインダストリ、クロスボーダーでのコネクタ型の連携へ

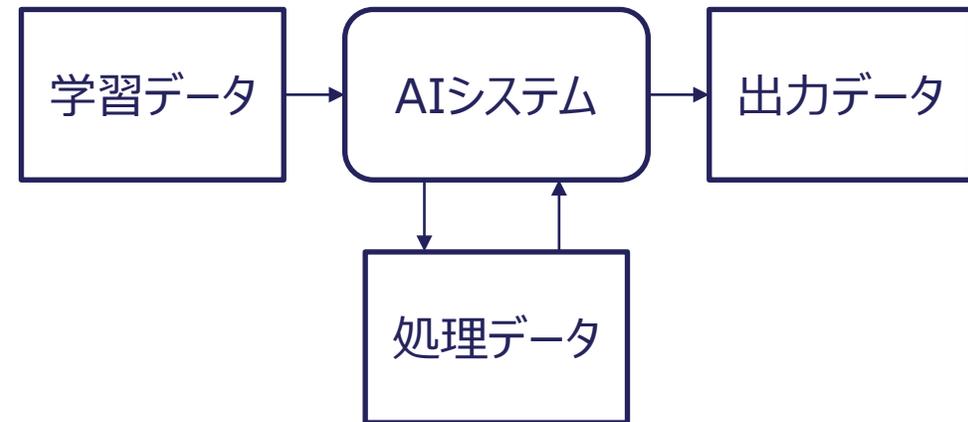
データマネジメント

- データ構造化から、品質の高いデータの安定的な供給へ

AIの発展を支える3要素でもあるデータ



Without data, there is no AI



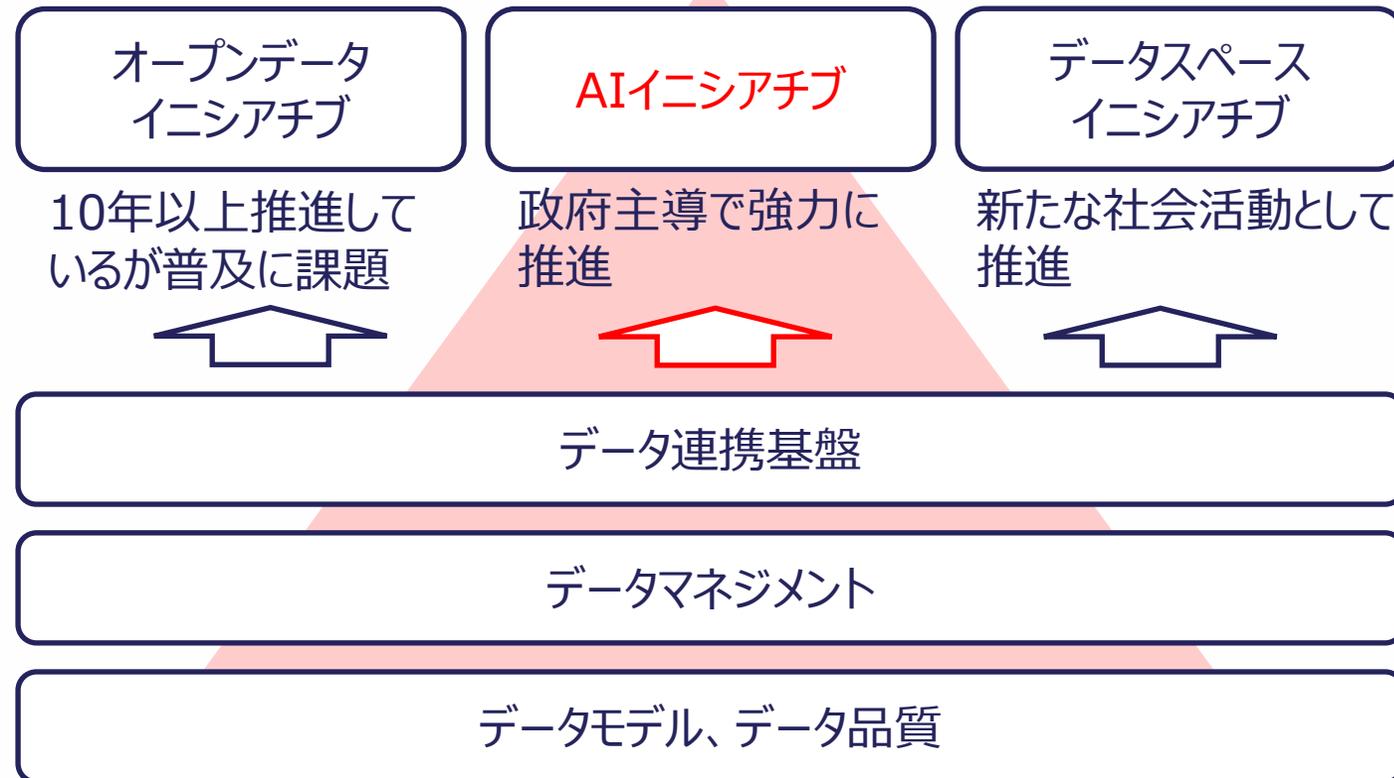
データを供給し、流通させる仕組みと、活用する能力が必要

- 社会全体のデータ
- 組織内のデータ

AIの推進と並行、先行してデータ整備を行う必要がある

- ◆ AIは大量のデータを必要とする

デマンドサイドからの強力なニーズ

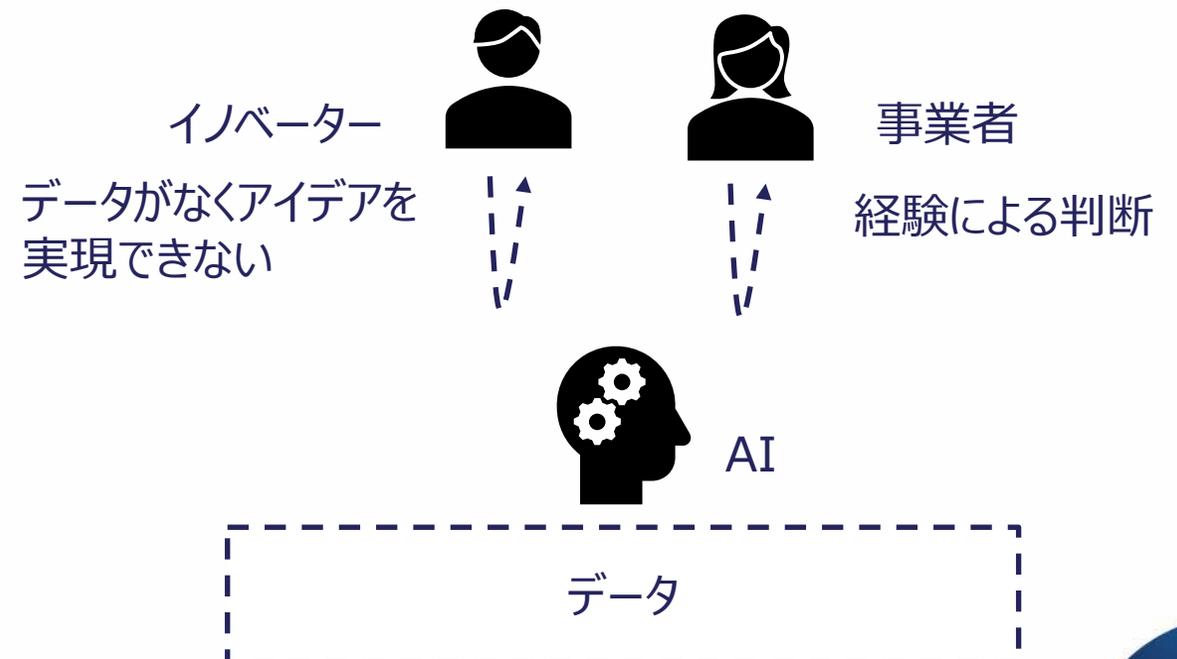
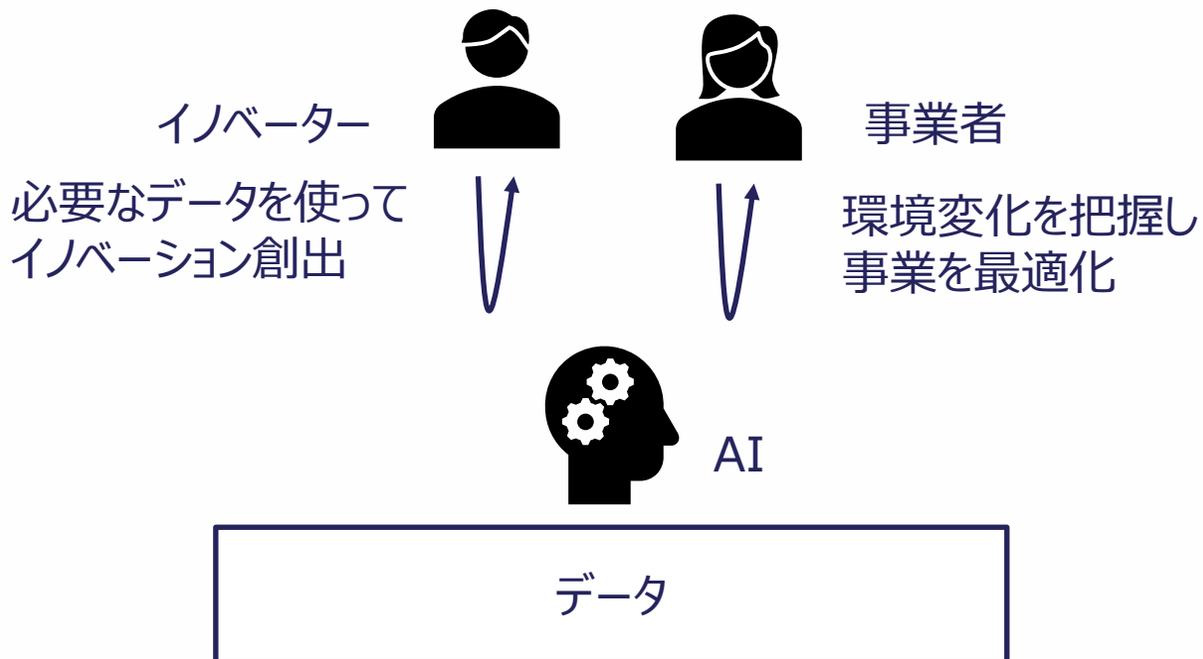


データ基盤があるかないかが国の競争力を左右

- ◆ AIは重要なツールであるが、それだけでは機能しない。

データ基盤のある国

データ基盤のない国



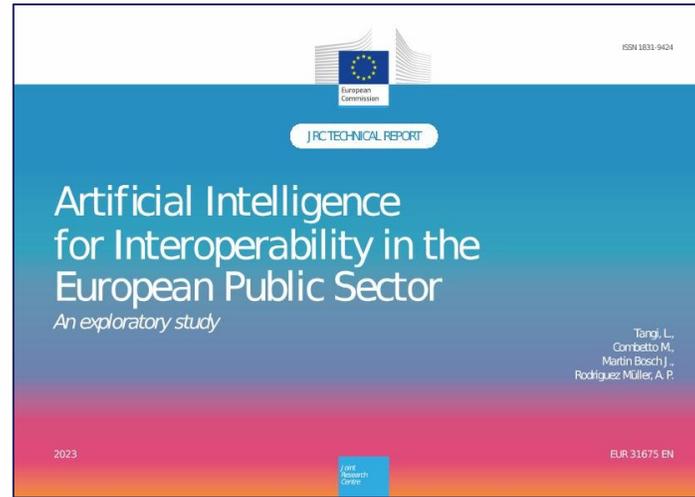
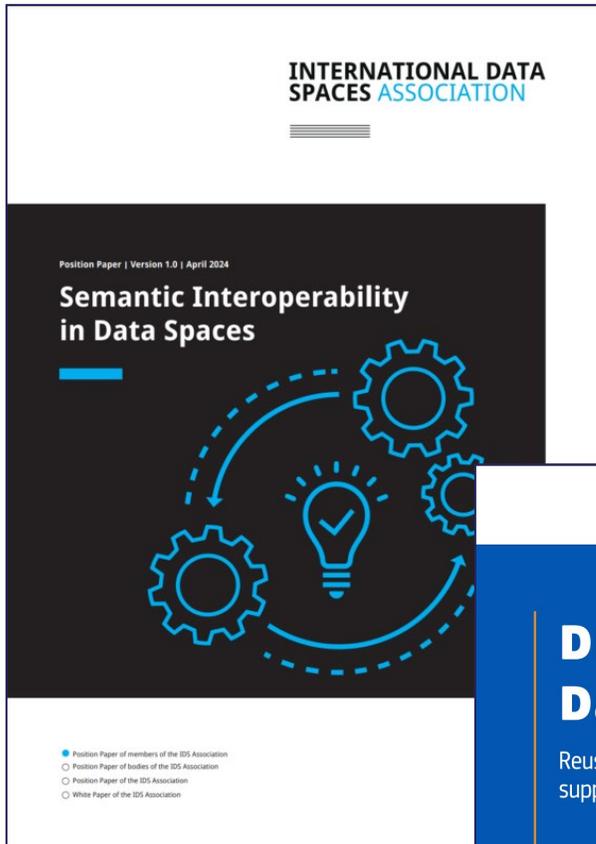
Global Digital Compact (2024-09)

- 国連の未来のための協定の付属文書である「グローバル・デジタル・コンパクト」は、**デジタル協力とAIガバナンスに関する初めての包括的な世界的枠組み**である。
- 以下のコミットメントが含まれる。
 - すべての人々、学校、病院をインターネットに接続する。
 - デジタル協力を、人権と国際法に根差したものににする。
 - 各国政府、テクノロジー企業、ソーシャルメディアの行動を通じて、すべての人々、特に子どもたちにとって、オンライン空間を安全なものにする。
 - 国際科学パネル並びにAIに関するグローバル政策対話を含めたロードマップにより、AIを統治する。
 - **オープンソースのデータ、モデル、標準に関する合意**を基に、データをよりオープンでアクセス可能なものににする。
 - このコンパクトはまた、**データガバナンスに関する初の国際的なコミットメント**である。データガバナンスを国連の議題に載せるとともに、各国に対し、2030年までに具体的な行動をとるよう求めている。

欧州の体系的アプローチ

欧州では様々なストーリーが繋がってきている

Data (Semantics)



AI



Data Spaces



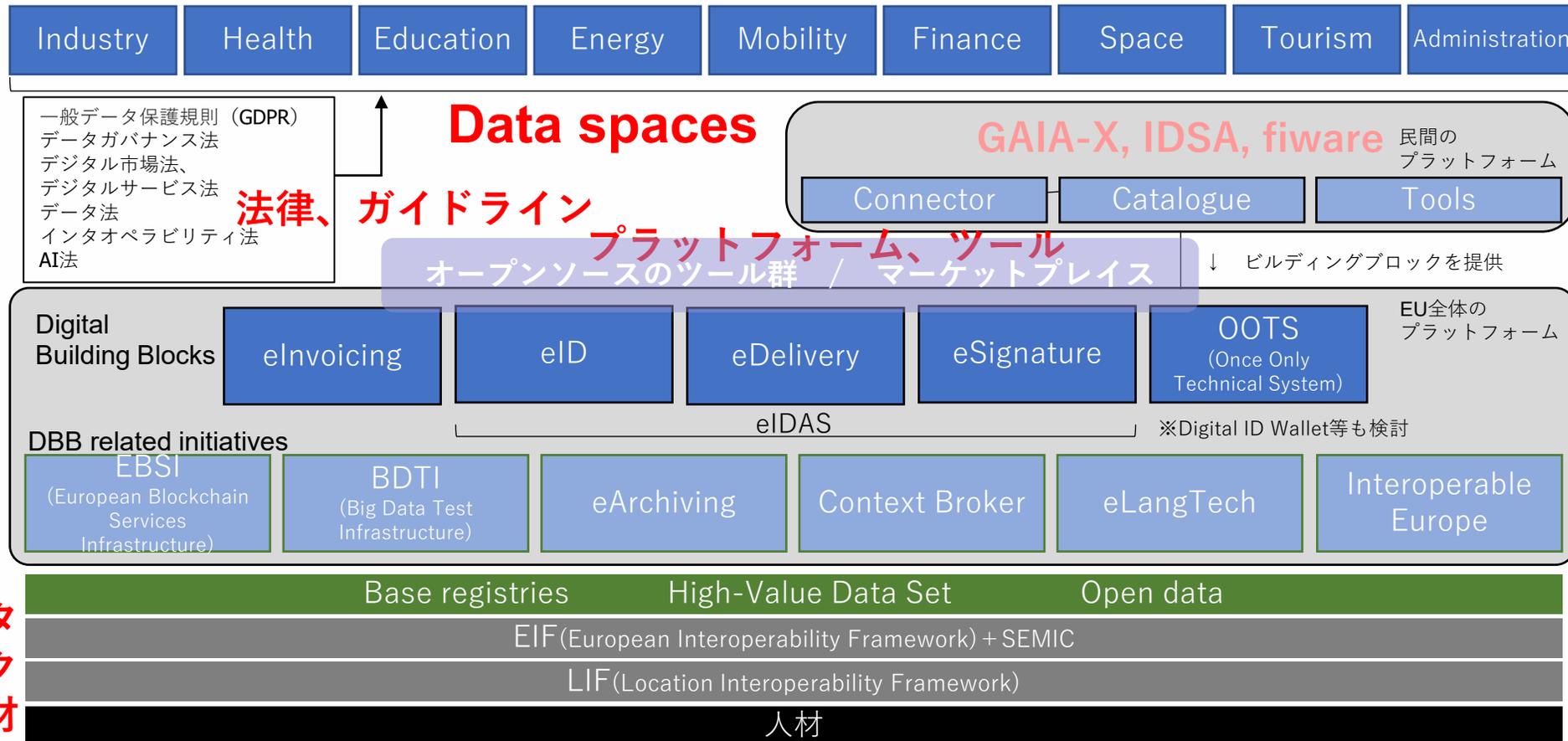
2030年に向けた、欧州の着実な布石

- ◆ 戦略 European Data Strategy、European AI Strategy
- ◆ ルール Interoperable Europe Act、・・・、AI act
- ◆ 組織 Data Maturity、
The Digital Competence Framework (DigComp)
- ◆ ビジネス Once Only Technical System (OOTS) 、 Use cases
- ◆ 利活用環境 Digital Building Blocks[eID、DIW等](DBB)、
Trusted Data Transaction (TDT) 、
Digital Market Place(DMP) 、AI-on-Demand(AIoD)
- ◆ 連携基盤 SIMPL、Connector
- ◆ データ Open Data、High Value Dataset(HVDs)、Base Registries
Quality、Graph Technology、Ontology

データスペースの基本構造

- ◆ 変化に応じ、迅速に実装可能な仕組みに基づくデータ連携環境
 - 組み立て構造：インタフェースを標準化し、コネクタなども複数から選択可能

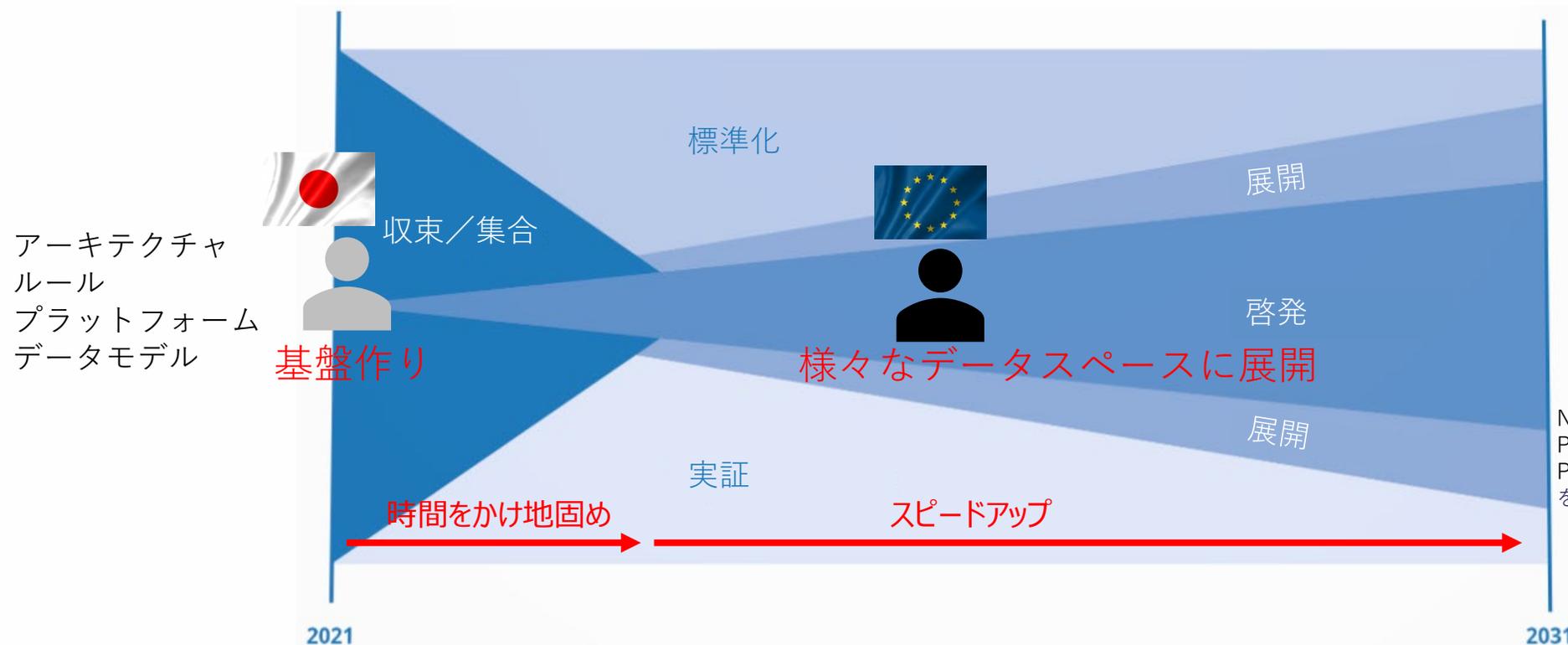
欧州の例



データ
フレームワーク
人材

基盤や体制整備が進んだ欧州に引き離されつつある

- ◆ 欧州は、制度や技術などの個々の取組を集約し体制も強化することで、データ戦略を一気に展開。
- ◆ 日本は個々の取組をしており、体制も小さく、地道に推進。



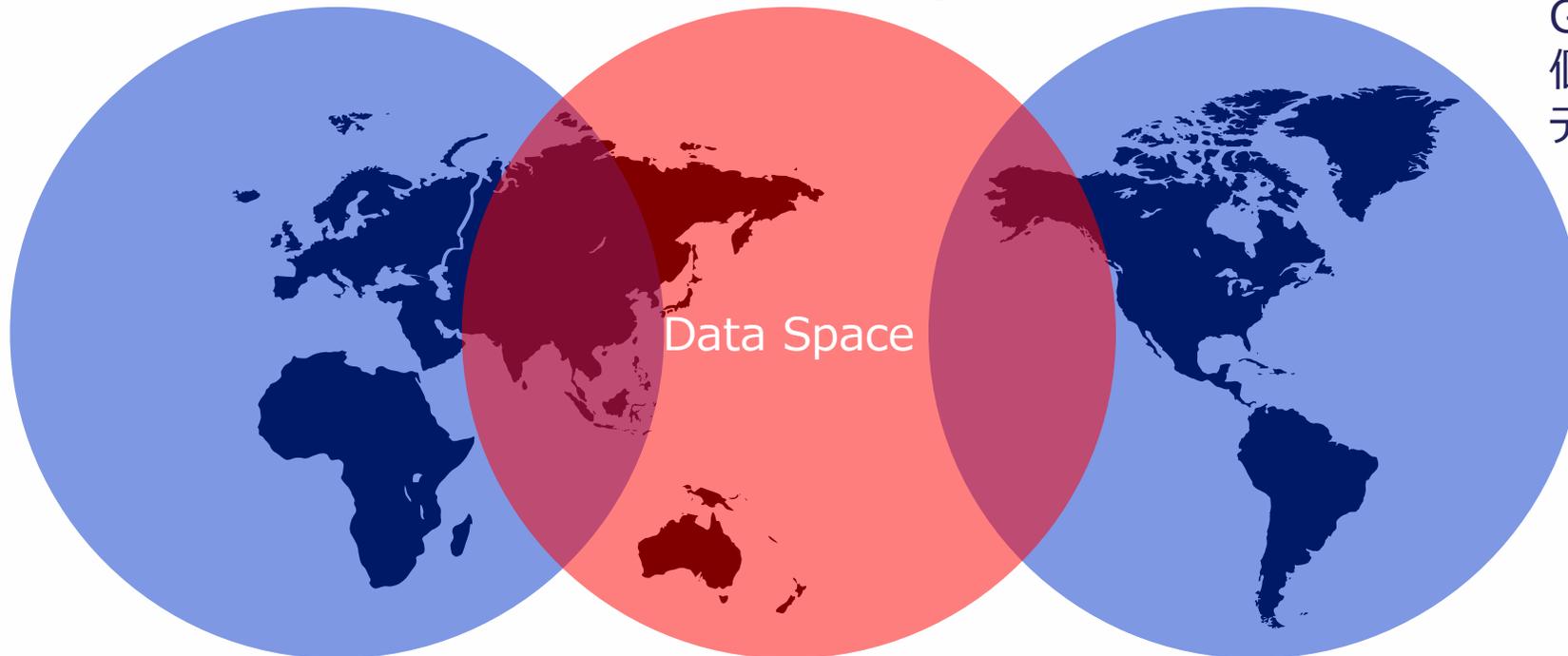
Nagel L., Lycklama D. (2021): Design Principles for Data Spaces. Position Paper. Version 1.0. Berlin
を元に加筆作成

国境や分野を越えたデータスペースの誕生

- ◆ 企業や個人の活動は、気が付かないうちに、日常的に国境を越えています。
 - 1企業による独占的サービスではない、**国境や分野を超えた新しい社会、経済活動経済圏**が生まれてきている。

デジタル競争の加速

GAIA-Xなどの
法人向けの
データスペース



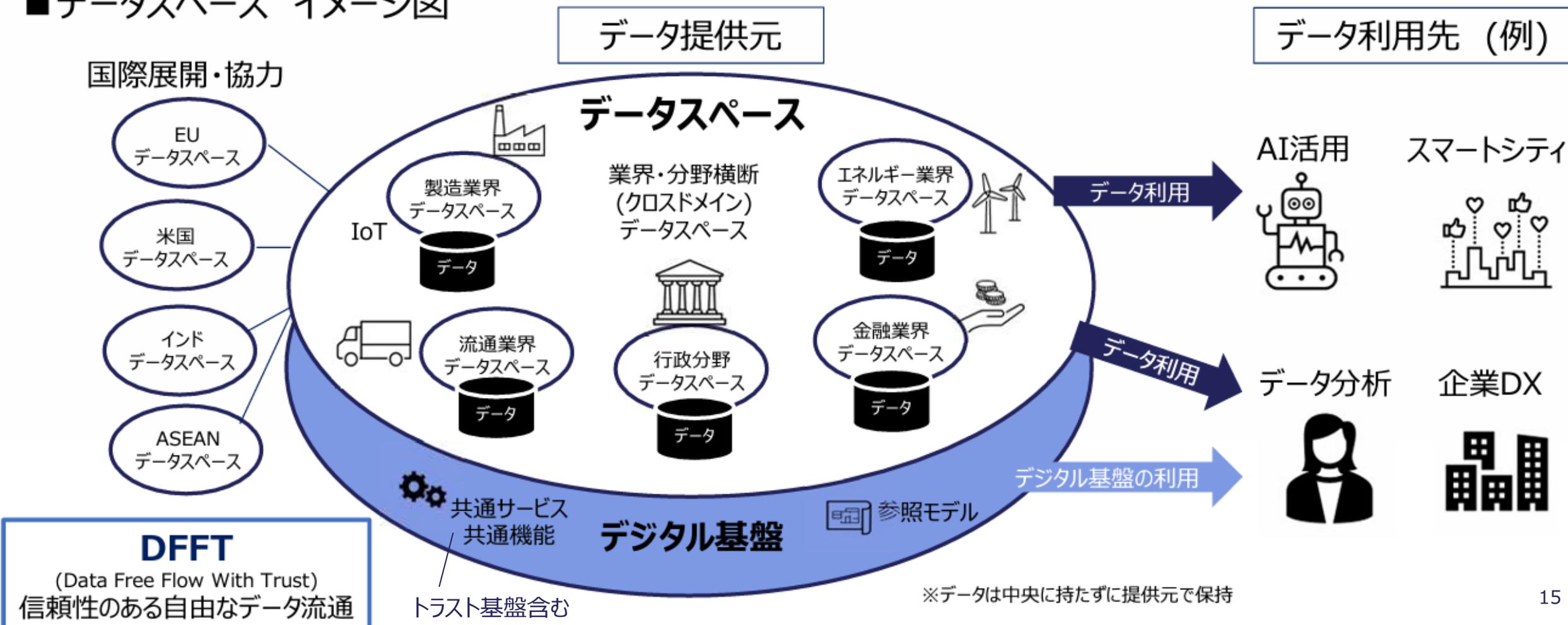
GAFAMを中心とした
個人向けの
データスペース

取引規約等、**データスペースのルール**に支配される

データスペースのイメージ

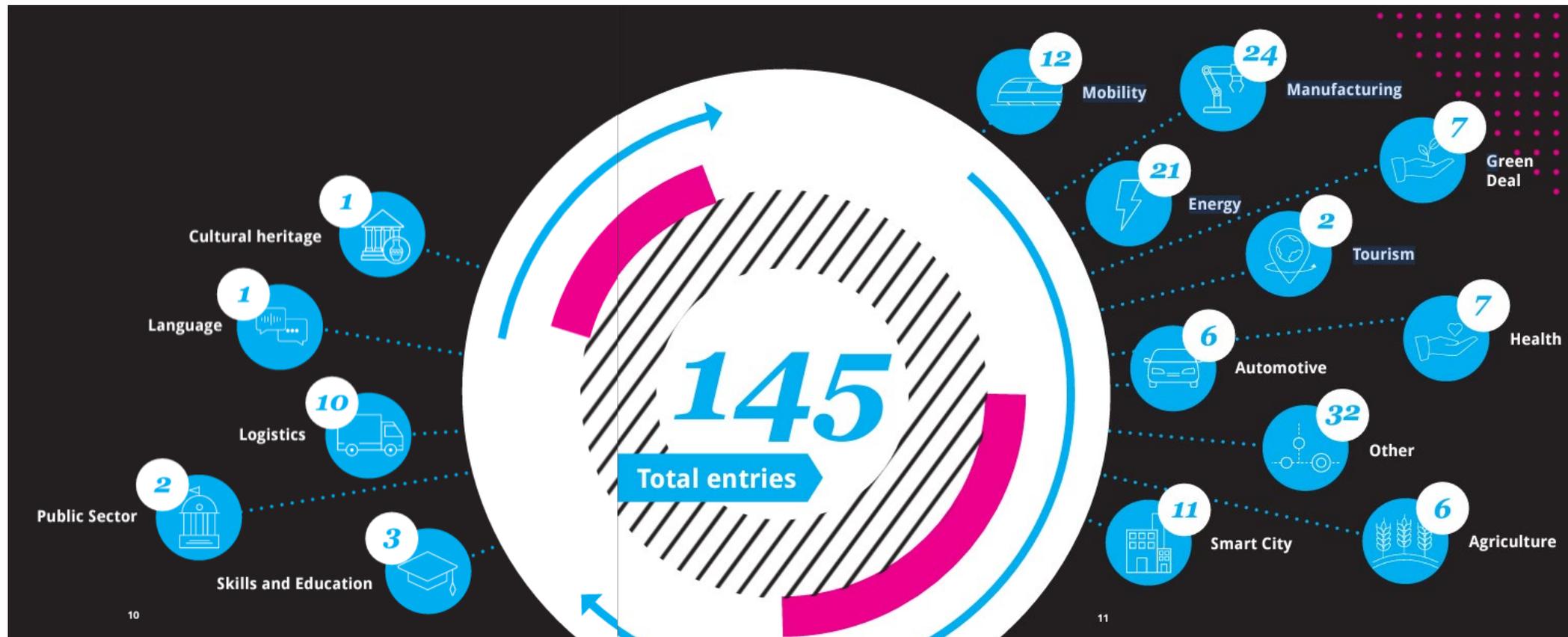
- ◆ 「多種多様」で「信頼性のある」大量のデータが安心して利用できる

■ データスペース イメージ図



データスペースはたくさんある

2024年3月で145の事例登録（9月で185の事例登録（89データスペース））



資料
IDSA Data Space Rader

GAIA-XやCATENA-Xが注目されるが、GAIA-Xはイニシアチブであり、CATENA-Xは1つの主要プロジェクト

データスペースの接続の仕組みもたくさんある

- ◆ 基本仕様はあるが、プロジェクトごとに多くのコネクタがある。
 - コネクタとは、データ交換の接続ツール。電気のコンセントのようなもの

Advaneo Open-Source EDC Connector
Advaneo-X Connector
Advaneo Trusted Connector
AI.SOV Connector
Boot-X Connector
Data Space Integration by SAP
ECI Gatewise IDS Connector powered by TNO
Eclipse Dataspace Components (EDC) – Framework
EdgeDS Connector
EGI DataHub Connector
EONA-X EDC Connector
FIWARE Data Space Connector
GATE Dataspace Connector
GDSO Connector - Tyre Information Service
HEALTH-X dataLOFT EDC
IIOC IoT Connector
Kharon IDS Connector powered by the Dataspace Connector
Mitsubishi Electric Dataspace Connector
MPAD-C by Mondragon

OneNet Connector
Prometheus-X Dataspace Connector
Silicon Economy EDC
sovity CaaS (Connector-as-a-Service)
sovity Open-Source EDC Connector
Tech2B SCSN Connector
Tekniker IDS Connector
Telekom DIH Connector
TNO Security Gateway (TSG)
Tritom Enterprise Connector
TRUE Connector
Trusted Connector
Trusted Supplier Connector
VTT DSIL Connector
Dataspace Connector
GAIAbOX
Teralab Connector
WeTech Connector



- ◆ システム間のデータ連携は、N対Nで、都度調整する方式が主流
- ◆ 多様なデータ連携が増えてくる中で、調整が業務の妨げになっている
- ◆ 様々なシステムをつなぐ支援システム、サービスもあるが抜本的に解決していない



- ◆ 政府のデータ戦略の一環でデータ環境の整備を推進
- ◆ 欧州のマーケットからの要請により産業界も対応

- ◆ データ連携基盤の整備は進んできたが、持続的な展開に向けてはまだ課題がある

データスペースの基本機能である**トラスト基盤**が整備できていない

- データにアクセスするための認証用IDの仕組みの整備が不十分

識別用IDの整備に方針がない／整備の必要性の認識がない

- 自動車、自転車、製品番号等、一意に判断できる情報統合のための仕組み
- 海外ではオブジェクトID等の付与も開始している

データスペースでのデータ連携が目的であるが、使いやすい形で**データが整備されていない**

- 社会基盤である行政データ：行政は国内最大で信頼できるのデータオーナーである
- 企業のデータ：データが、独自に設計されていて、再利用しにくい場合が多い

人材と経験が不足しており、データスペースに十分対応できていない

- データサイエンティストなどに人材が偏在しており、データの全体設計できる人が少ない

国内はデータ供給が不足

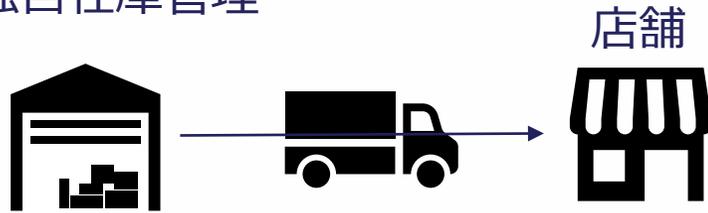
- ◆ 検索のためのメタデータ標準のDCATの導入が進んでいない。また、データ量が少なくPDFが多いため、実用性向上が必要。

フォーマット(2024-03-02)	米国	フォーマット(2024-03-02)	日本	フォーマット(2024-03-02)	欧州
XML	143,908	PDF	9,679	WMS	392,735
HTML	101,279	HTML	7,068	CSV	304,961
ZIP	94,631	XLS	4,650	WFS	278,678
PDF	37,259	XLSX	1,733	JSON	128,419
CSV	24,769	CSV	983	HTML	89,056
EsriREST	17,448	ZIP	510	ZIP	87,937
JSON	15,899	JPEG	431	EXCEL XLSX	76,005
TEXT	15,577	XML	153	PDF	66,250
TIFF	14,423	GIF	88	EXCEL XLS	56,493
SID	12,795	KMZ	71	Esri Shape	40,415
RDF	11,913	PNG	47	XML	39,968
WMS	9,997	DOC	26	Atom Feed	34,349
ArcGIS GeoServices REST API	9,625	DOCX	24	TIFF	29,950
KML	9,188	EXE	16	KML	24,854
XYZ	7,924	epub	13	GeoJSON	21,488
EXCEL	7,700	PPTX	12	TSV	16,076
GeoJSON	6,109	TXT	9	Application/JSON	12,760

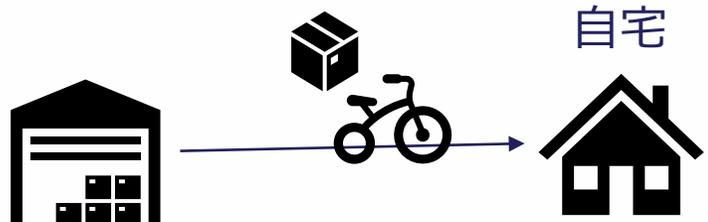
JSONでの提供は0件

データハンドリングが変わってきている

独自形式データと
独自在庫管理



標準化されたデータと
メタデータ



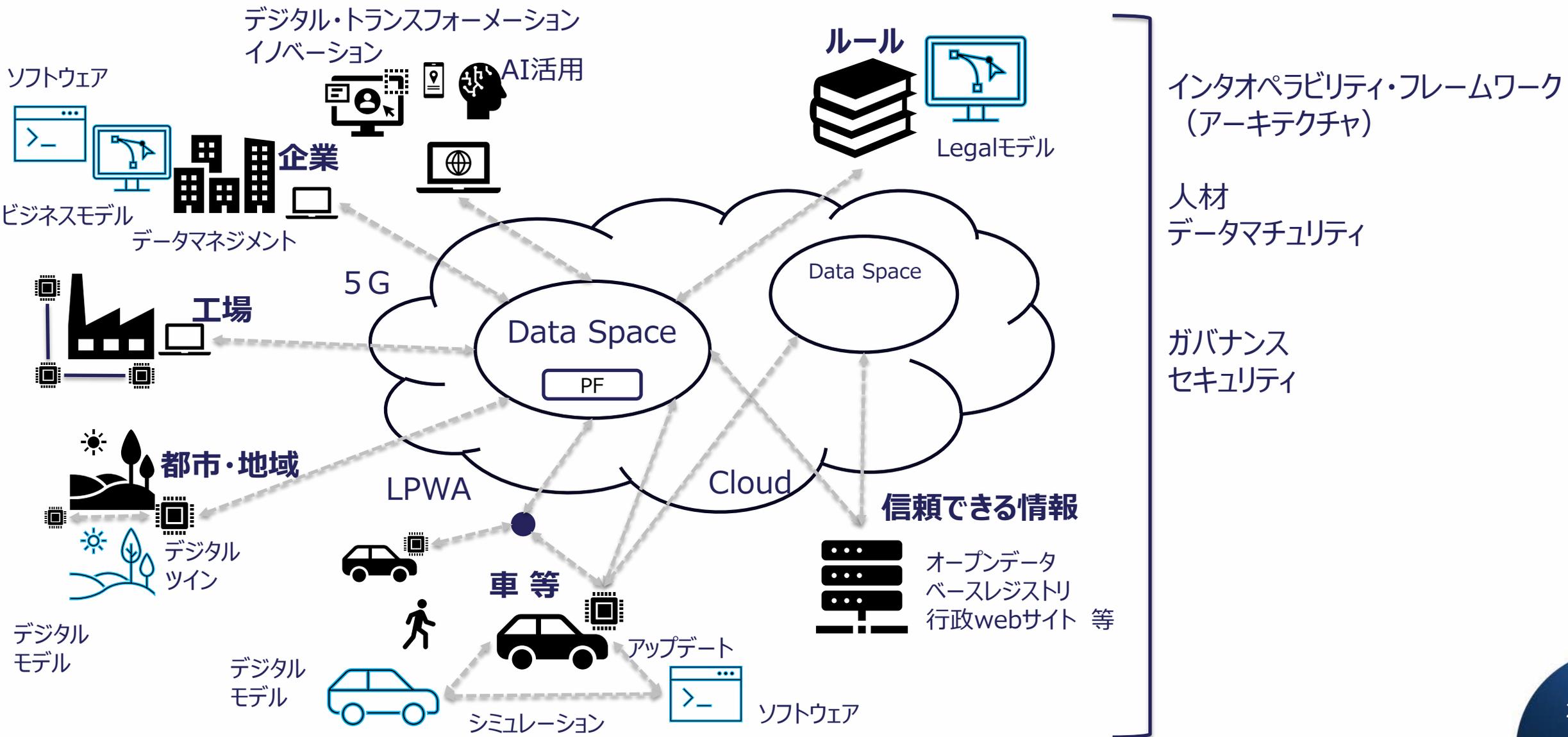
- ◆ 国内は、RDBとデータレイク、ファイル転送、高速ネットを軸にした設計思想
 - **大量消費時代の思想**
 - データ保護が起点
 - ・ 持っていかれるときは全部持っていかれる
- ◆ 世界は、従来の思想に加えグラフ、アクセス、低速ネットを組み合わせた設計思想
 - **大量消費と個別消費を組み合わせたベストチョイスの思想**
 - データ共有が起点、
 - ・ しかも全体のセキュリティを強化

- ◆ 米国は、現時点ではデータスペースとは関連付けていないが、デジタルエンジニアリングの動きが注目される。AIの活用も視野に入る。
 - あらゆるものを、モデル化、データ化し、からシミュレーション、導入、フィードバックを一貫の流れで検討し、ゴールとしてのデジタルツインを目指す
- ◆ 欧州は、データスペースとCEI（Cloud Edge IoT）やAIとの融合を推進。
 - リアルタイムデータのデジタルツインへの取り込みを目指す
 - 制度のデジタル対応もデジタル社会のゲームチェンジャーとして強力に推進

米国、欧州の流れから、データを軸にリアルタイム性を持った社会全体のデジタルツインが進むと考えられる。

その土台がデータスペースであり、データ流通量も今後増大すると予測されていることから、AI活用も含め大きなビジネスチャンスになると予測される。

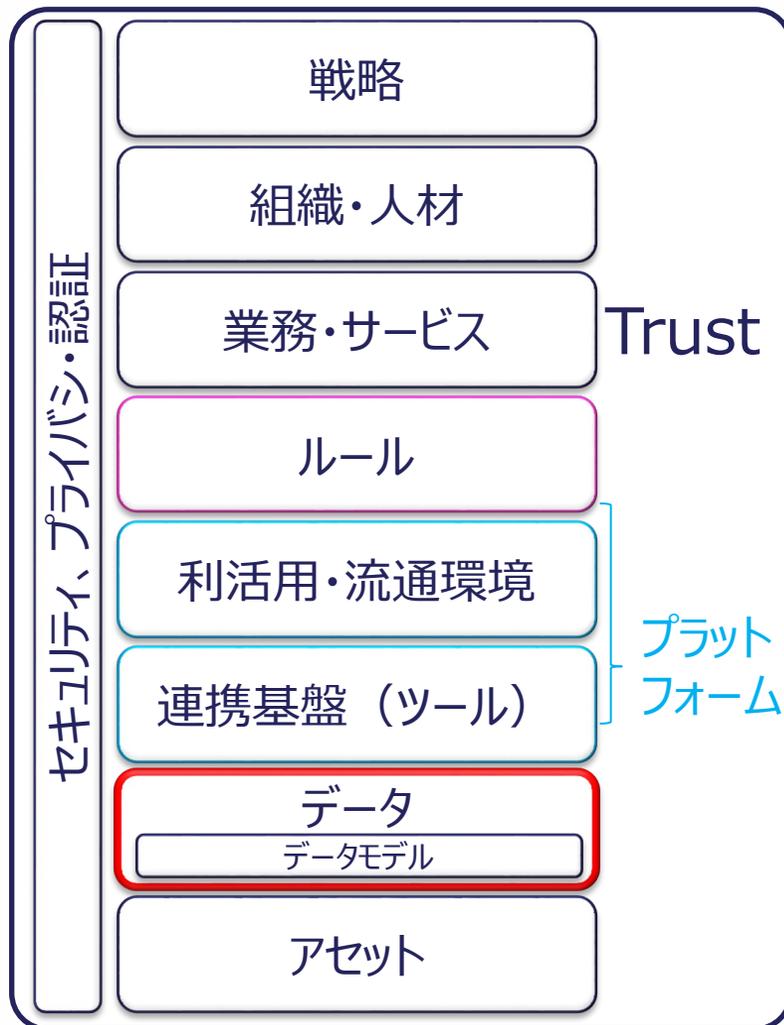
イメージ



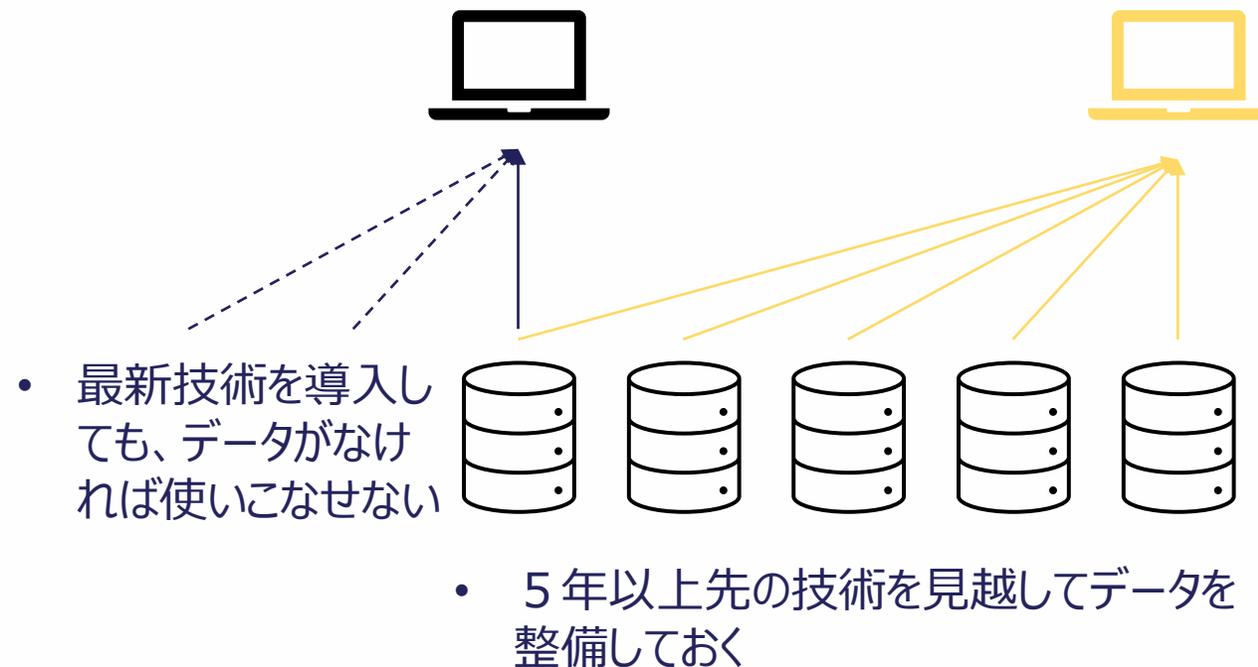
アーキテクチャで考え、長期的投資を行う必要がある

◆ アーキテクチャで考える

バランスの取れた
整備が必要



◆ 5年以上先を見た投資を行う



各レイヤーでの取り組み

- ◆ ルールはなるべく標準のものを使う。調整コストと時間を節約する。
 - 契約
 - 利用規約
 - アクセス規約
- ◆ グローバルな動向にも注意が必要。
 - GDPR、データ法等
- ◆ 世界では、ルールのモデル化が進展
 - 合理性の追求
 - 透明化・可視化、それに伴う各種対応の迅速化
 - 事前シミュレーションの実現

- ◆ 汎用的なものを利用。データは転送するという思想からの脱却も必要。アクセスするほうが正確性や最新性を確保しやすい。
 - 商用プラットフォームサービスを利用
 - OSSも含むビルディングブロック（コンポーネント等）を活用する
- ◆ いずれにしてもインタフェースが重要。
 - API
 - データモデル
- ◆ トラストサービスを早急に整備していくことが求められている。

- ◆ 汎用的なデータモデルを徹底して活用し、データの再利用性を高める。
 - 一部を活用したり拡張して利用する参照データモデルを活用
 - 政府相互運用性フレームワーク・データモデル
 - Schema.org
 - Smart Data Models
- ◆ 参照データモデルを使ってデータ設計することで、ビジネスの迅速性、アジリティ、データを使ったビジネス創出が簡単にできるようになる。
 - また、データ品質の向上が期待できる

例 日付

2024年10月17日

20241017

2024-10-17

- ◆ 行政機関は、その国にとって最大のデータオーナーでありデータ供給者であり、各国は基盤強化を図っている。

	先進国	日本
ベースレジストリ ※行政が正確性や最新性を保証する社会の基本データ	欧州はほぼすべての国で基本 5 情報の整備が終わり、連携しながら運用、さらに拡張中。	法人と住所の2つのベースレジストリを試行中
ハイバリューデータセット ※行政が業務で保有する社会的価値あるデータ	欧州は、2024年6月からオープンデータにすることを法制化	方針なし
オープンデータ ※行政が保有する一般データ。原則オープン化されることになっている	データカタログが充実。訓練コースも提供。	カタログサービスを設置。
プラットフォーム、データモデル	Digital Building Blocksの整備。インタオペラブル・ヨーロッパ法の推進。	DATA-EXやウラノスが稼働。

- ◆ グローバルに競争するには社会基盤として、データ供給の充実が必要。

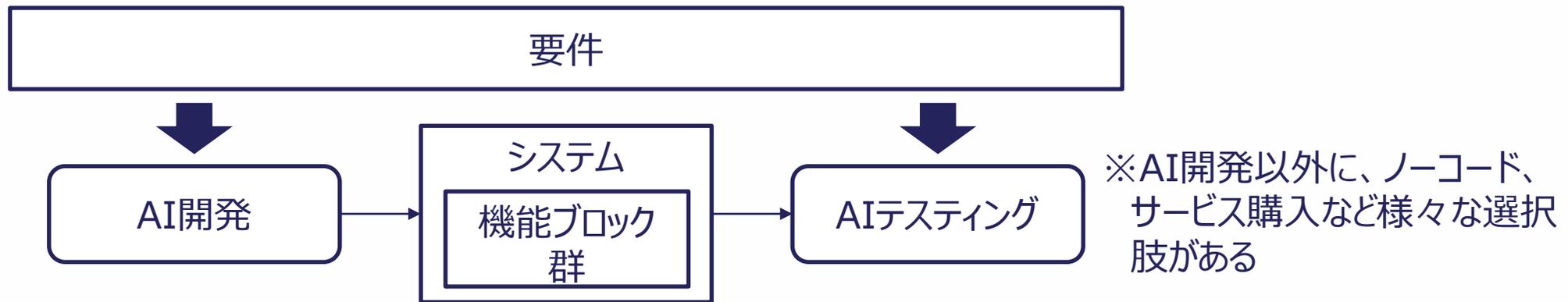
- ◆ データマネジメントやガバナンスの重要性が注目されている。
 - 組織にどのようなデータがあるのか？
 - データは正確か、最新か？
 - そのデータの権利は明確か？
 - 昔のデータに引きずられていないか？
- ◆ データのライフサイクルで見直していく必要がある。
- ◆ AIを活用するためには、その基盤となる品質の高いデータが重要である。
 - ISO25012、ISO25024、ISO8000、ISO5259

データマチュリティに立ち返ってみよう

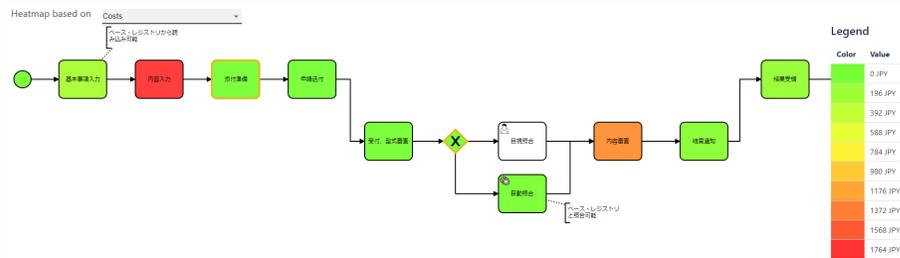
AI社会のためのデータのエコシステム構築に向けて

要件定義ベース、組み立てベースになる

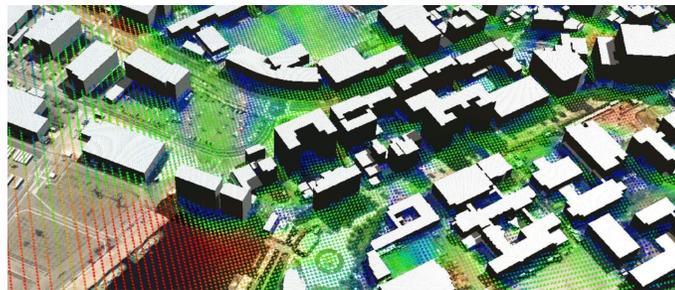
- ◆ プロンプトエンジニアリングとか「はやり言葉」を言う以前に、要件定義の基礎を固める必要がある。
 - 要件が定義されるから、プロンプトも明確に書け、AI開発やAIテストができる。
- ◆ もちろん、要件定義自体をAIでサポートすることも可能。
- ◆ 要件が定義できたからと、なんでも作ればよいというものではない。
- ◆ 再利用できる部品を活用することが必要。



- ◆ ガバナンスの基本は可視化である。経営ダッシュボードだけではなく、具体化が必要。
- ◆ 可視化されているから、共通の理解ができ、課題が明確化されるとともに、効率的に議論ができる。さらに、解決策を作り、検証することができる。
- ◆ AIに指示を的確に出すにも要件定義の考え方が必要になる。



業務プロセスモデルの可視化

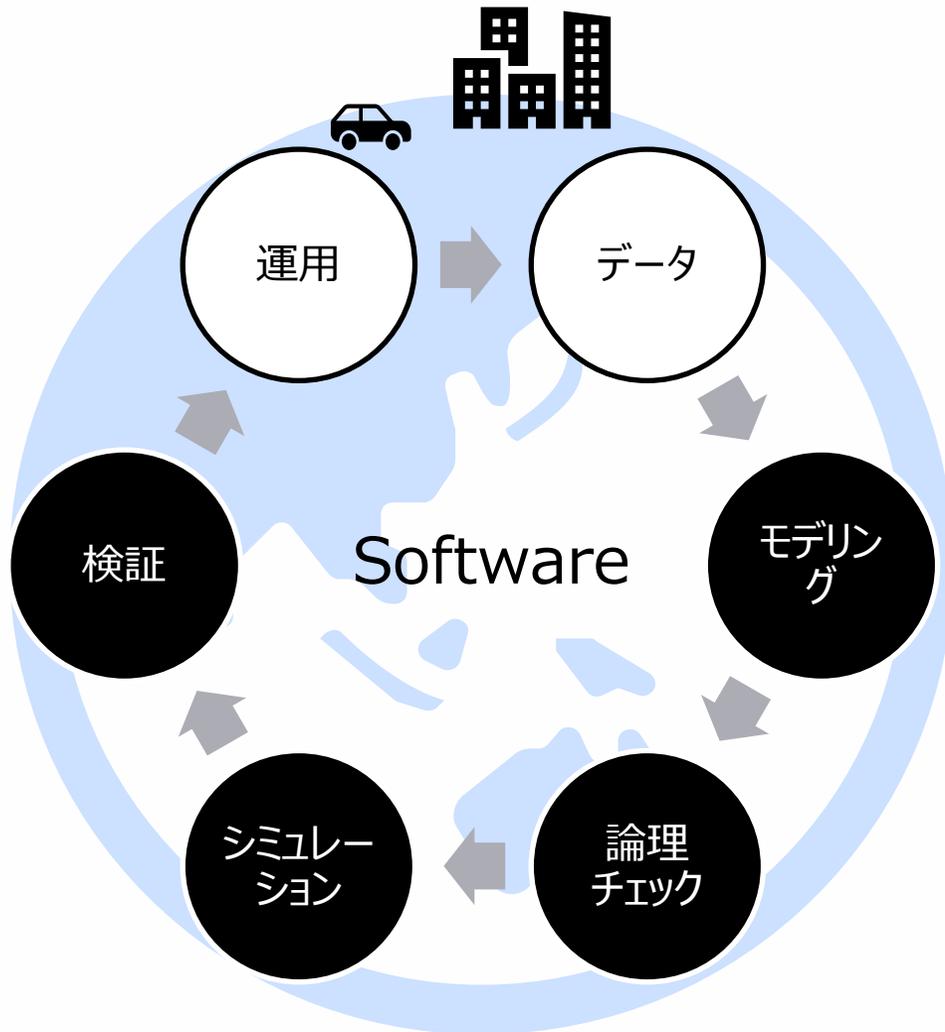


都市モデルの可視化 (Plateau)

モデリングで活動を可視化

- 可視化することで事前検証が容易になる
- トラブル発生時に原因究明が容易になる
- 事前検証や予見が可能になる

グローバル起点のアクションが必要



- ◆ AIやデータ活用、ソフトウェアの開発ではグローバルに競争が行われており、企画やプロト検証などでもグローバルを意識しなければならない。
- ◆ 日本でPoCしてからというスピードではついていけない。
 - グローバルなニーズや知見の取り込み

The logo for IPA (Information Policy Agency) is displayed in a bold, orange, sans-serif font.

IPAでは世界レベルで活躍したい
データ専門家を大募集中です。