

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／
パーソナルデータアーキテクチャ構築

パーソナルデータ リファレンスアーキテクチャ概要書

**DFFT(Data Free Flow With Trust) 実現のための
アーキテクチャ設計と国際標準化推進の研究開発**

2020年3月18日初版

2020年3月24日改定

2020年3月末 改定

一般社団法人データ流通推進協議会（DTA）

第1章 はじめに

本概要書の位置付け

本書は「DFFT(Data Free Flow with Trust) 実現のためのアーキテクチャ設計と国際標準化推進の研究開発」の5つの成果物の内のパーソナルデータリファレンスアーキテクチャ書の概要書 (No1)である

N o	文書名	文書 形態	概要
1	パーソナルデータリファレンスアーキテクチャ概要書	PPT	下記2のリファレンスアーキテクチャ書（設計書）の概要書
2	パーソナルデータリファレンスアーキテクチャ書（設計書）	Word	パーソナルデータを扱う全ての事業者、ステークホルダが、ビジネスモデルや内部統制などのシステム設計を行うためのガイドとなる設計書
3	ユースケースシナリオテンプレート	PPT	パーソナルデータを取り扱う事業者が活用するテンプレートドキュメント
4	ユースケースシナリオ集	PPT	本プロジェクトの採択事業実施者が記載したユースケースシナリオのまとめ
5	ELSI検討報告書	Word 及び PPT	パーソナルデータの取扱いをELSI視点で必要事項をまとめたもの

リファレンスアーキテクチャ書の構成

導入編

- パーソナルデータを取り扱う事業者が理解し、または留意すべき事項を解説しており、**特定の事業や計画の実態の有無やその進捗に関わらず一読**することを想定している

活用編

- 各事業者が**自らの事業のアーキテクチャを設計・整理**し、パーソナルデータの取り扱いの適正性や潜在する課題を顕在化し、適切なパーソナルデータの利活用モデルを構築するためのリファレンスアーキテクチャとその記載方法について解説する

研究開発編

- 本書の作成にあたり、当協会では、国内での有識者会合および実証実験事業者との会合を開催するほか、国際標準化の調査と推進を実行した。これらの活動および、その調査内容について報告する
- 各事業者の事業実施における**国際展開を検討する際の参考**となることを想定している

付属資料

- リファレンスアーキテクチャ及び関連文書において用いられる用語の定義を取りまとめた「用語・定義書」
- データジャケットの国際標準化報告

パーソナルデータ分野に対する背景

課題

個人の課題：データ扱いを把握・制御できない不安
 企業の課題：企業・業界を超えたデータ流通・活用が進まない

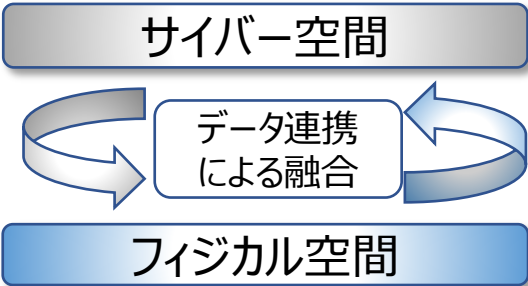
⇒便益の見える化
 ⇒魅力的なサービスの創出

背景

第5期科学技術基本計画(2016年)
Society5.0の提唱と実装

官民データ活用推進基本法

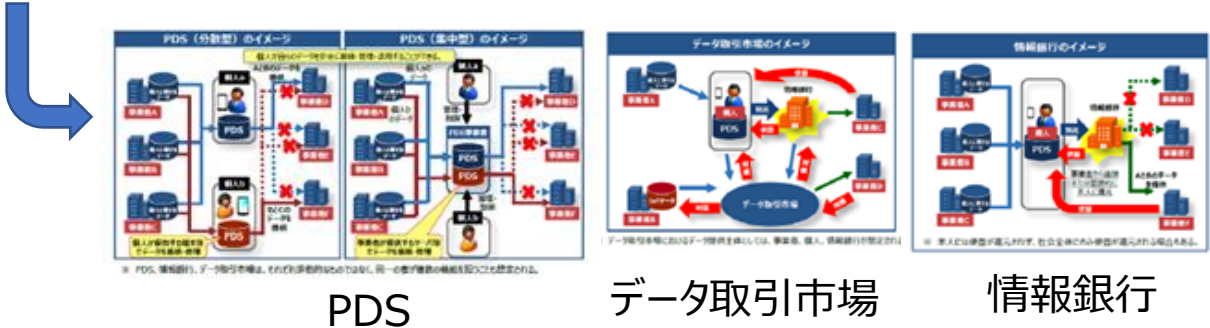
デジタル手続法



課題

個人の課題
 データ扱いを把握・制御できない不安
 ⇒便益を実感できない不満や不公平感

企業の課題
 企業・業界を超えたデータ流通・活用が進まない
 ⇒魅力的なサービスが創出がされない



研究開発の狙いと課題

狙い

データ活用の促進に資するための共通的なリファレンスアーキテクチャ設計手法を示す。以下の課題を克服する総合的アーキテクチャを構築するグランドデザインの整備

項目	課題
アーキテクチャ設計	アーキテクチャ構築による成果がどのように社会実装に寄与し、どう使われ、データ提供者にどのように価値を還元するかなどを明らかにするアプローチ（ユースケースシナリオの蓄積など）
情報銀行連携	実証研究テーマ実施者と連携し、実証研究における具体的な課題を顕在化し共有すること
分野間データ連携	分野にとらわれず、パーソナルデータであることに配慮したデータ連携ができるユースケースシナリオテンプレートの作成
国際標準化の推進	取組としてのグランドデザインの欠如、ユースケースシナリオの集積不足、成果物の形態定義不明瞭への対応、及び、国際標準化推進政策としてのデジュール・フォーラム包括的標準化推進体制の欠如、SDO連携の明確なミッションの欠如、リーダーシップ人材の不足への対応がある

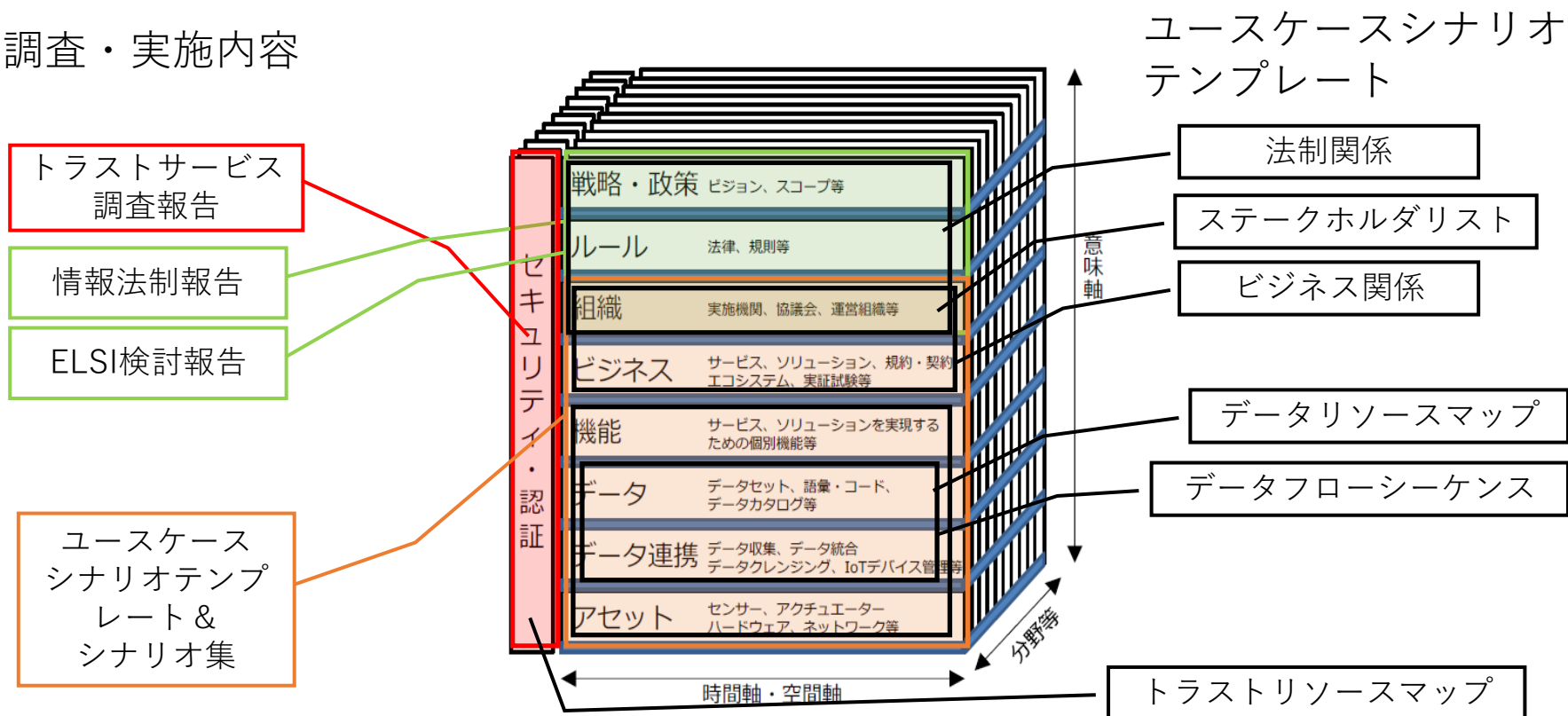
課題解決への対応

項目	課題解決への対応策
アーキテクチャ設計	アーキテクチャとは何かという基本的な概念の共有知の不足がある。アーキテクチャとは各事業者が自らの事業を設計・整理するための手引書であるとの定義を明確にし、設計手順を整理した（第8章）。また、実際の設計成果となる書面を定式化し、ユースケースシナリオテンプレートとしてまとめ、その使い方の実例を示した（第9章）
情報銀行連携	情報銀行とデータ取引市場の連携等について、調査によりパーソナルデータを扱う事業モデルを整理した（第3章）
分野間データ連携	設計書における定義設定（2.1）とパーソナルデータを扱う原則を示し（2.4）、パーソナルデータを扱うユースケースシナリオにおいては、ISO/IEC29100の分類に沿って整理することを推奨した（9.3）。また、テンプレートが有効に利用可能であることを示した
国際標準化の推進	”データ取引システム“標準化の必要性をIEEE-SA(Standard Association)に提案し、DTSI(Data Trading System Initiative)が設置された（10.1.2.1）。この活動を通して、リーダーシップ人材を育成し（10.1.2.1）、デジュール・フォーラム包括的標準化推進体制を構築し（10.1.2.2）、SDOとの連携の明確なミッションを定める一例としてW3Cとの関係構築を図った（10.1.2.3）

Society5.0 RAにおける設計書の位置づけ

本テーマの実施内容をSociety5.0リファレンスアーキテクチャと関係付けるなら以下のようになる。それぞれ複数の層に跨って対応する

調査・実施内容



(出所：NEDO 「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」公募要領におけるSociety 5.0リファレンスアーキテクチャ図を記載）

導入編

導入編の構成

パーソナルデータを取り扱う事業者が理解し、または留意すべき事項を解説しており、特定の事業や計画の実態の有無やその進捗に関わらず一読することを想定している

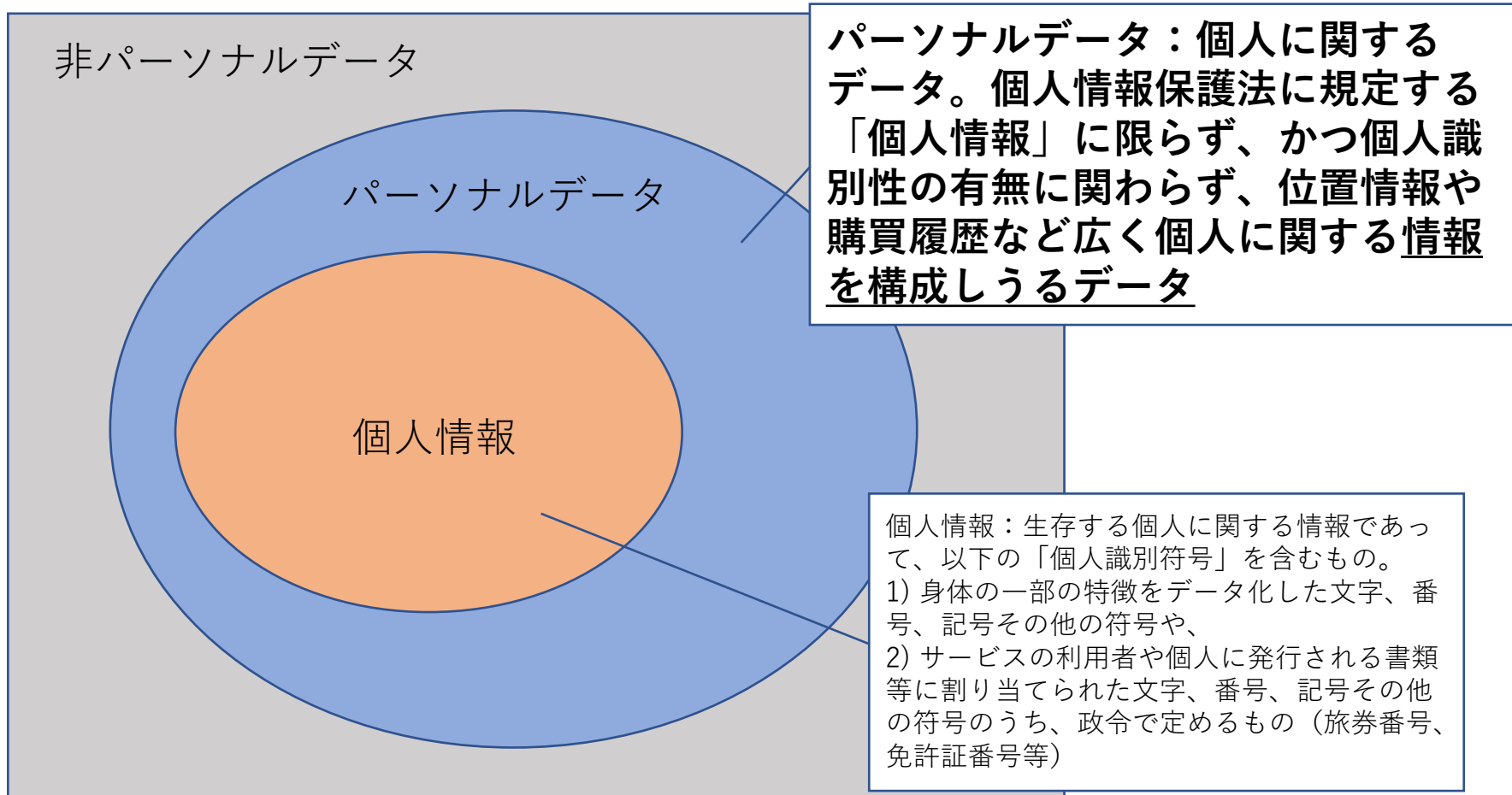
- 第2章 基本理念
- 第3章 パーソナルデータを扱う事業モデル
- 第4章 パーソナルデータを扱う上で必要なELSI
- 第5章 パーソナルデータと関連法制
- 第6章 トラストサービスの概要と現状



第 2 章 基本概念

パーソナルデータとは

- 改正個人情報保護法で定義される「個人情報」よりも範囲は広い



※改正個人情報保護法（平成29年施行）

データと情報

データ Data

- 「**データとは**、情報の表現を構成する要素であり、伝達、解釈または処理に適するように**形式化され**、複数のデータの組み合わせにより、**情報を構成しうるものである。**」

データセット Dataset

- データは、一般に名前と年齢などのように、複数の**データ**とデータの属性などを示す**メタデータから構成される**データセットとして取り扱われる

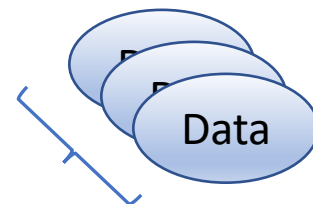
情報 Information

- 事実、事象、事物、過程、着想および概念により構成され、対象物に対して一定の文脈中で特定の意味をもつもので、**データセットを含むものもある**
- データセットと付帯情報(事実、事象、事物、過程、着想)および概念により構成され、対象物に対して一定の文脈中で特定の意味をもつもの
- (ISO/IEC 2382-1, JIS X 0001)

Data



Dataset



Information

いろはにほへと、
Xx??%*(

Dataset



データの特徴とオーナーシップ

データは無体物であり、民法上、**所有権や占有権、用益物権、担保物権**の対象とはならない

所有権や占有権の概念に基づいてデータに係る**権利の有無**を定めることはできない
(民法 206 条、同法 85 条参照)

知的財産権として保護される場合や、不正競争防止法上の営業秘密として**法的に保護**される場合は、**限定的**である

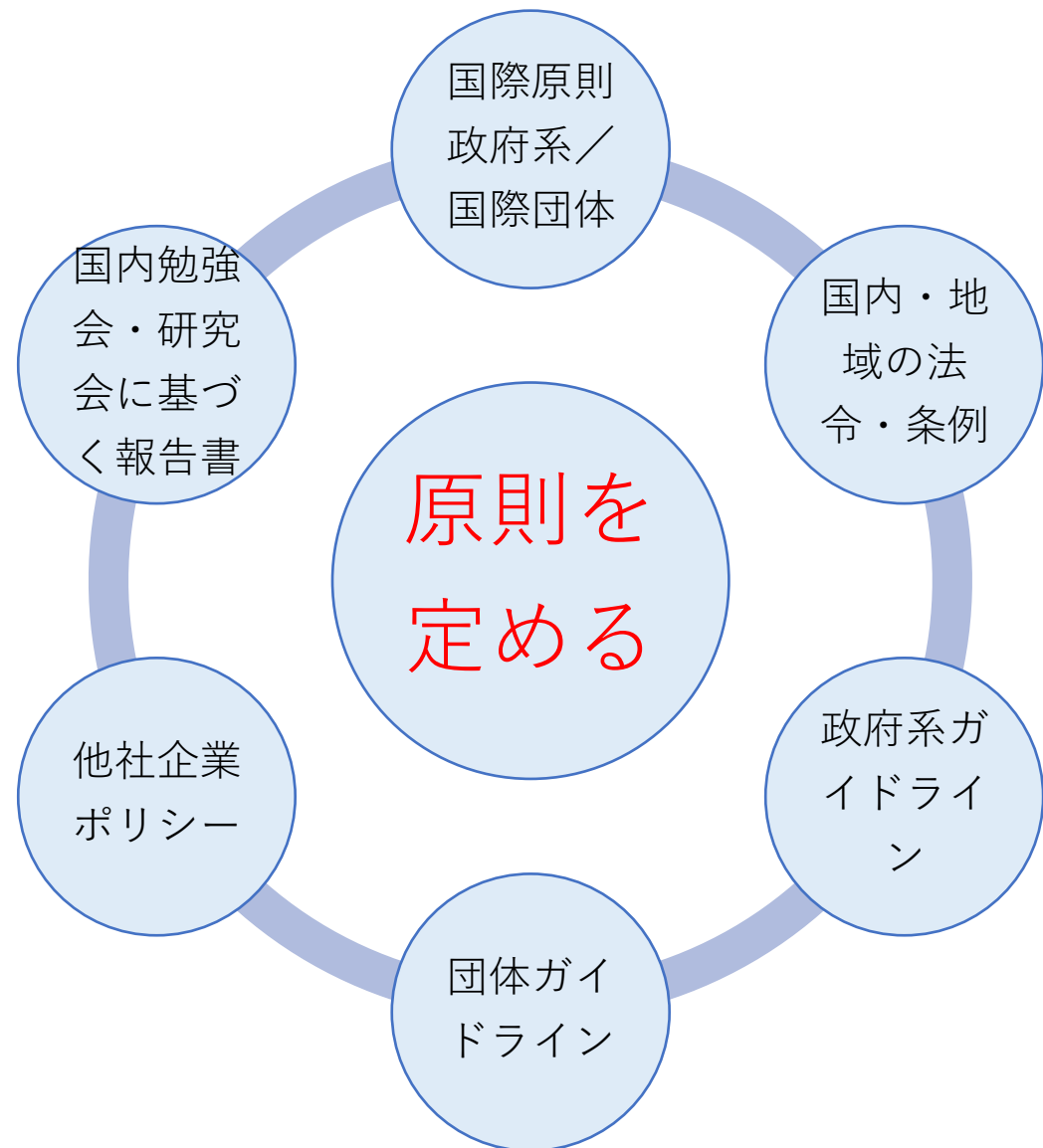
データの**保護**は原則として**利害関係者間の契約**を通じて図られる

「データ・オーナーシップ」という言葉の**法的な定義**はない

「データに対する**所有権を観念できる**」という意味ではない

パーソナルデータを扱う原則

- パーソナルデータの取り扱いについては、ISO/IEC 29100、OECDプライバシー8原則、米国FTC FIPPs(Fair Information Practice Principles), OECD8原則、など**様々な取り扱い原則**が存在する
- また、各国や団体、企業などが定めガイドラインなど、**事業に関する規範が多数**存在している
- これらにおいて、その分類などは異なるが、その**目指す**ところ、あるいは**前提**とするところに大きな**齟齬**はない
- そこで、パーソナルデータを取り扱う事業を行う者は、第五章および別冊の「パーソナルデータ分野に関するELSI検討会報告書」を一読し、**自らの原則を定める**ことが重要となる

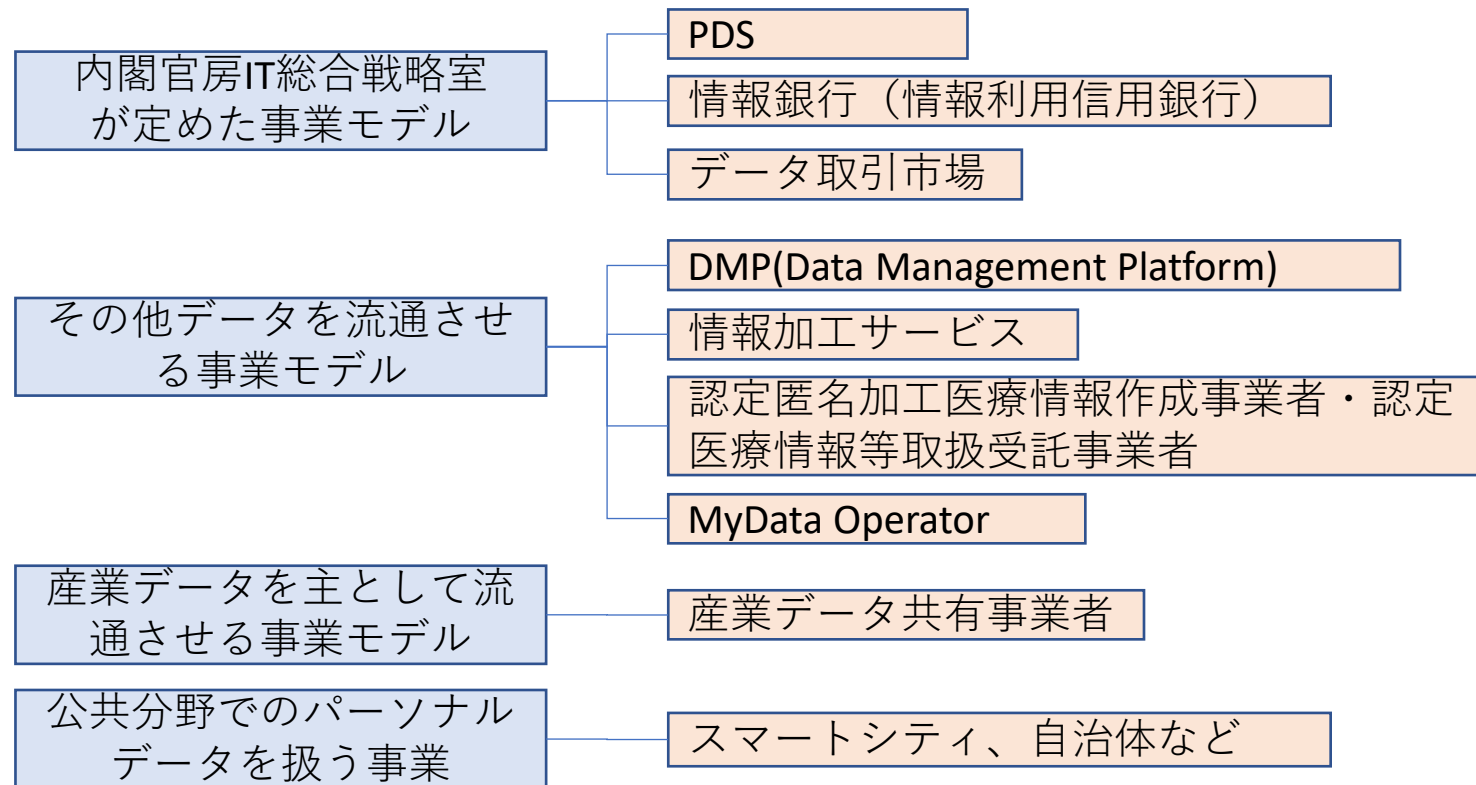




第3章 パーソナルデータを扱う事業モデル

事業モデルの確認

- 自らがパーソナルデータを扱う事業者なのかどうか、**パーソナルデータを取り扱う**としたらどのような**事業者**に**分類**されるかなどについて**明確な判断ができない**、あるいは**疑義がある**事業者は、この章の示す類型化を参考にして、**自らの事業モデルを確認**することができる



パーソナルデータを扱う事業の範囲

- 前スライドの事業は、もっぱらパーソナルデータの取り扱いを、その事業の主たる取り扱い対象とするもの
- 全てのビジネスモデルが、これらのモデルのいずれかに単純に類型化されるものではない
- むしろ、パーソナルデータは多くの事業に利用されるため、多くのビジネスモデルは前スライドの示す事業モデルの一部または全部を包含或いは組み合わせにより実現される

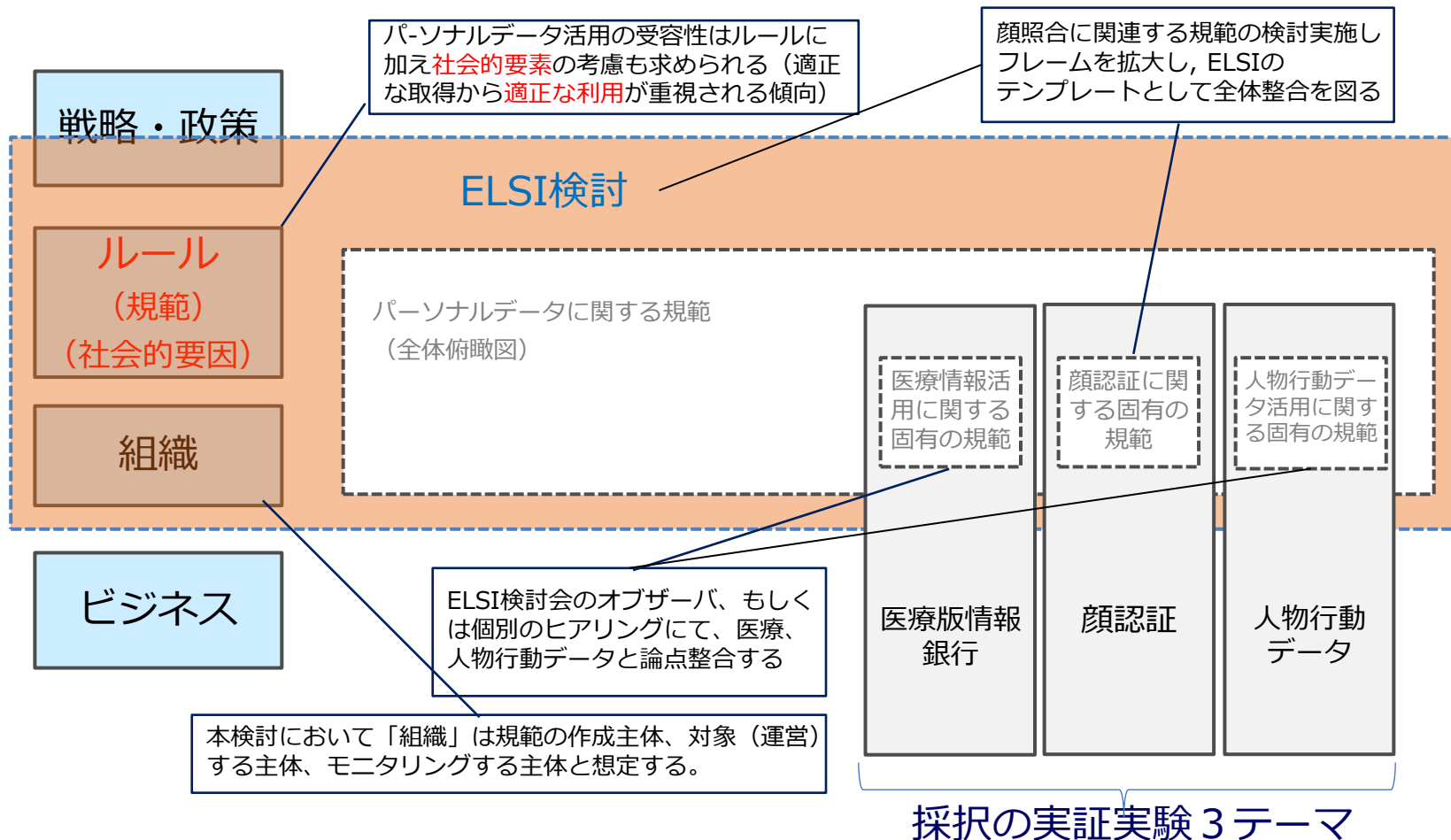
第4章

パーソナルデータを扱う上で必要なELSI

-Ethics, Legal, and Social Issues-

ELSI検討会の位置付け

- 医療版情報銀行、顔認証、人物行動データという異なるタイプのパーソナルデータを扱うテーマと論点を共有し全体整合を図った



パーソナルデータのアーキテクチャ（ルール層）

ルール層においては、**法制度に留まらず**、広義の社会規範やPEST分析における**社会的要因をスコープに入れるべき**

P:Politics

（政治的要因）

規制など市場のルールを変化させるもの

- ・ 法律、法改正（規制・緩和）・ 税制、現在・増税
- ・ 政治、政権交代 ・ 裁判制度 ・ 政治団体、デモ

E:Economy

（経済的要因）

景気や経済成長など、価値連鎖に影響を与えるもの

- ・ 景気動向 ・ 経済成長率 ・ 物価 ・ 為替 ・ 株価
- ・ 金利 ・ 原油 ・ 消費動向

S:Society

（社会的要因）

人口動態の変化など需要構造に影響を与えるもの

- ・ 人口動態 ・ 密度 ・ 構成 ・ 流行 ・ 世論 ・ 世帯
- ・ 宗教 ・ 教育 ・ 言語 ・ 高齢人口 ・ 少子化

T:technology

（技術的要因）

I T など、競争ステージに影響を与えるもの

- ・ インフラ ・ I T 活用 ・ イノベーション ・ 特許
- ・ 新技術、技術開発

6つの基本要件

- 基本要件① グローバルな目線の必要性
- 基本要件② 責任あるビジネス、バリューチェーンの推進
- 基本要件③ 消費者（個人）を主役に据えた事業全体のデザイン
- 基本要件④ 消費者目線を踏まえた通知および同意
- 基本要件⑤ フェアネス
- 基本要件⑥ 透明性と説明責任

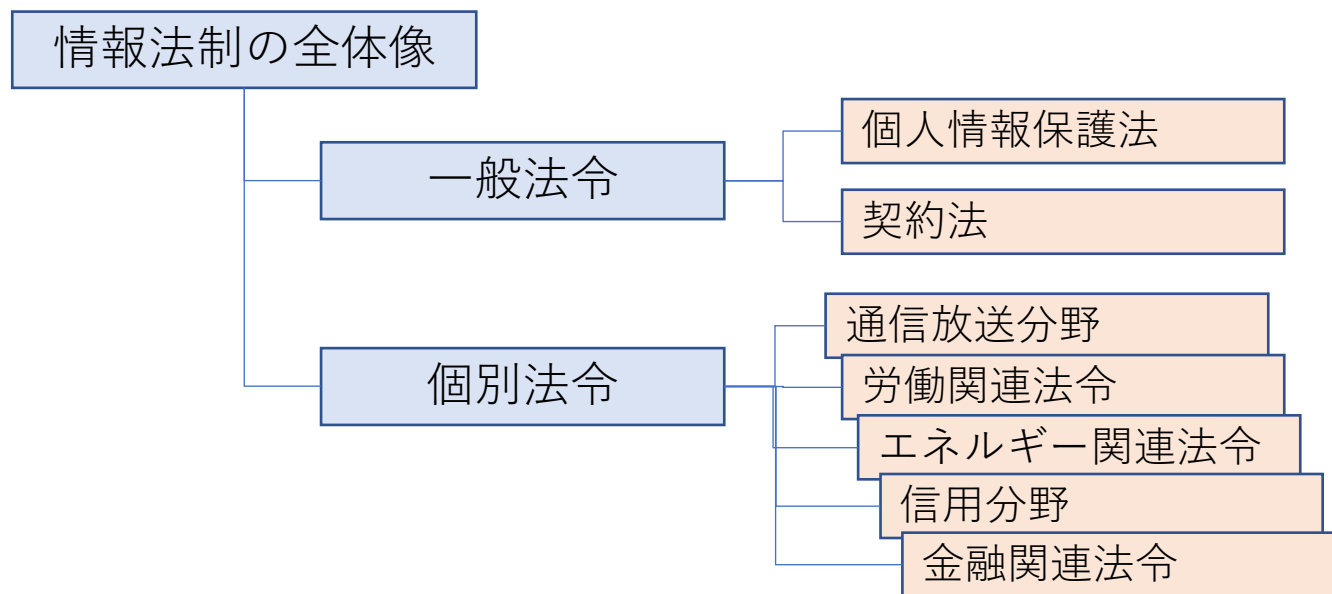
行動に関する提言

- 認定制度の創設
- 認定制度自体のブランド力強化
- 市民社会との対話と共働



第5章 パーソナルデータと関連法制

- パーソナルデータの取り扱いでは、個人情報保護法その他、各種業法に対する遵法性や各国や地域の法制に照らした事業内容との整合性が重要
- 情報法制の全体像は以下のように一般法令と個別法令に大別される
- ユースケースシナリオ集の法制関連表では法律（個人情報保護法等）と契約法の観点でまとめることが推奨される





第6章 トラストサービスの概要と現状

トラストサービス

- パーソナルデータを取り扱う事業では、パーソナルデータを提供する個人と事業者に限らず、複数の事業者が連携して事業を行うことが想定される。このような複数のステークホルダーが連携して一つのシステムを構成する場合には、**各機関や個人との間での認証や認可**といった信頼関係の構築が重要となる
 - 6.1 認証の種類
 - 6.2 データに対する認証と技術
 - 6.3 認証と認可
- このような信頼関係を確立するために用いられる電子署名や認証などのトラストサービスは、国内外を問わず広く検討され、その導入や法令による導入なども進められている。そこで、本書の作成において参照とするために、外部の学術研究者により実施した「トラストサービス調査報告」から「トラストサービスの各国の取り組み」を解説する
 - 6.4.1 欧州の取り組み（eIDAS規制）
 - 6.4.2 米国の取り組み（トラストフレームワーク）
 - 6.4.3 日本の取り組み

活用編

活用編の構成

各事業者が自らの事業のアーキテクチャを設計・整理し、パーソナルデータの取り扱いの**適正性や潜在する課題を顕在化**し、適切なパーソナルデータの**利活用モデルを構築**するためのリファレンスアーキテクチャとその記載方法について解説する

- 第7章 用語・定義
- 第8章 リファレンスアーキテクチャ本体（設計部）
- 第9章 ユースケーステンプレートの使い方



第7章 用語・定義

用語・定義集

定義項目

分類

- 当該用語が主として利用される範囲を以下に類型化したもの。当該用語の利用範囲を規定するものではなく、定義集の整理の目的として付したもので、利用者が任意の用語を検索することを容易にするための分類である

用語

- 当該用語の日本語表記

英語表記

- 当該用語の英語表記

本書での定義

- 当該用語の本書における定義

リファレンス

- 当該用語について、既出の定義などがある場合の外部参照情報

リファレンス先での定義

- 当該用語について、外部参照先での定義

アイコン

- 本書およびユースケースなどで当該用語を表す場合のアイコン

分類の内訳

全般

- 当該用語がアーキテクチャ全般にて利用されているまたは、適切な分類が未定のもの

データ種類

- データまたは、データセットの構造、及びその内容に関する用語

情報種類

- データまたは、データセットにより構成される情報に関する用語

データ処理

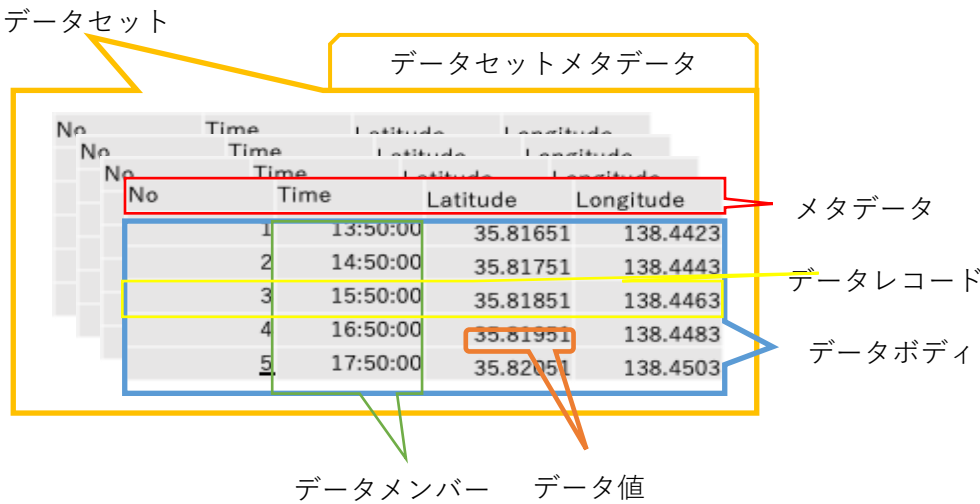
- データまたは、データセットの処理に関する用語

契約・トラスト

- 契約・トラストに関する用語

用語・定義集（一部抜粋）

分類	用語	英語表記	本書での定義	リファレンス	リファレンス先での定義	アイコン
全般	データ	Data	データとは、情報の表現であって、伝達、解釈また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	ISO/IEC 2382-1,JIS X0001 情報処理用語-基本用語	A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.情報の表現であって、伝達、解釈また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	
全般	データボディ	Data Body	1以上のデータの集合でメタデータを含まない	No definition in ISO		
全般	メタデータ	Meta Data	データのうち、データの属性などを示すデータ	ISO/IEC 11179-3:2013, 3.2.74	data that defines and describes other data	
全般	データ値	Data Value	個々のデータの持つ値	ISO/IEC 25000:2005	content of data item	
全般	データメンバー	Data Member	同一のメタデータに紐づくデータの集合	No definition in ISO and ITU		
全般	データレコード	Data Record	共通の識別子により関連づけられたデータメンバーの集合	ISO 18739:2016(en), 3.1.13	one or more data items treated as a unit within a data set	
全般	データセット	Data Set	データボディ、メタデータの集合で、データセット自体にもメタデータが含まれる	ISO 8000-2:2018, 3.2.4	logically meaningful group of data	



本書でのデータセットのアイコン定義

-  個人データを含まない場合
-  個人データを含む場合
-  仮名化された情報の場合



第 8 章

リファレンスアーキテクチャ本体（設計部）

設計部の位置付け

自らの事業を設計・整理するための以下のような手引書である

設計部の 定義

- ・パーソナルデータを扱う全ての事業者、ステークホルダが、ビジネスモデルや内部統制などのシステム設計を行うためのガイド

利用目的

- ・リファレンスアーキテクチャを設計・整理することで、各事業者がパーソナルデータの取扱いの適正性や潜在する課題を顕在化し、適切なパーソナルデータの利活用モデルを普及させる
- ・パーソナルデータを取り扱う事業の共通要件を明確にすることで、分野・事業間の一定の協業を推進する

対象者

- ・パーソナルデータを取り扱う事業者(事業の計画時、検討時も含む)、パーソナルデータを取り扱う可能性を有する事業者(事業の計画時、検討時も含む)

制約事項

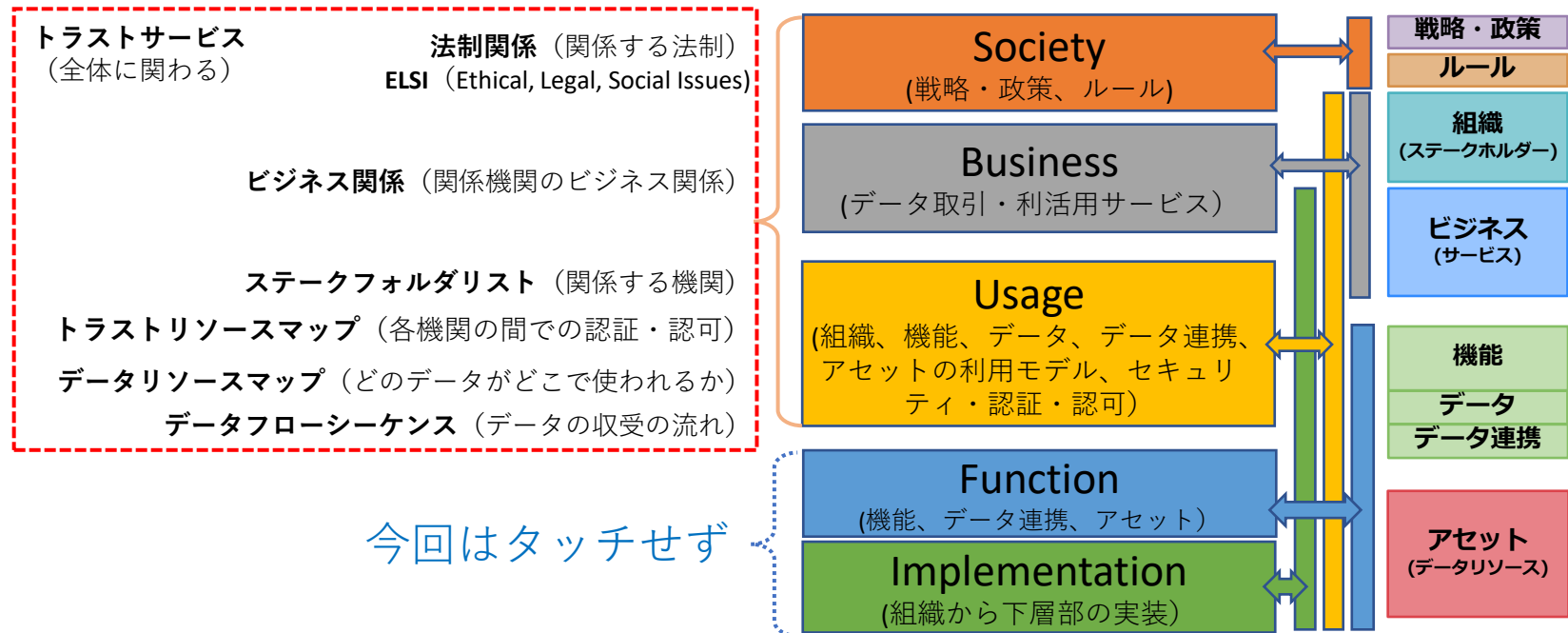
- ・設計部を含む本書は、パーソナルデータを取り扱う事業に対して、免責を付すものではない。本書は、パーソナルデータを取り扱う事業に対して、実装を制限するものではない

使い方

- ・本書記載のユースケースシナリオテンプレートに照らして、自身のビジネスのユースケースシナリオを構築する
- ・ユースケースシナリオテンプレート記載のチェックポイントを参照しながら、自身のビジネスを解析し課題を明確にしていく。解決すべき課題がどこに存在するかを明確化するチェックリストとして使用する

Usage視点中心のアーキテクチャ

- 国際標準ISO/IEC/IEEE 42010は複雑なSoSシステムのアーキテクチャ構築指針を与える。これをベースに開発されたIIRA*を参考に5つ視点(Implementation, Function, Usage, Business, Society)を定義できる
- 以下はSociety5.0RAとの比較である。本事業ではUsage視点を中心にBusiness, Societyに跨ってとりまとめた



5つの視点とSociety5.0 RAとの対応関係

*IIRA: Industrial Internet Reference Architecture. IICが開発する参照アーキテクチャ

アーキテクチャの設計手順

事業者は以下の手順に従って、そのアーキテクチャを明確にできる

(1) 対象事業の適用性確認

- 対象とする事業の概要を掴む。パーソナルデータの関与があるかを確認する
- データジャケットによる新事業領域の創出を検討してみる

(2) プライバシー原則の制定

- 自身の事業のプライバシー原則を定める

(3) ユースケースシナリオテンプレートに従った記載

- 「ステークホルダリスト」、「ビジネス関係図」、「データリソースマップ」、「トラストリソースマップ」、「データフローシーケンス」、「法制関係表」の各アーキテクチャ図面を記載する

(4) プライバシー原則や関連法性に照らして、作成した各図面の整合性を評価

- 第9章及び別冊「ユースケースシナリオ集」を参考に実施する

(5) 確認の結果

- 評価結果を確認し、修正すべき事項があれば修正し、(3) からを繰り返す
- 最終段階では、ELSI視点での問題がないかを確認する

第9章

ユースケースシナリオテンプレートの使い方

ユースケースシナリオテンプレート

テンプレートとして以下の6つを用意している



1.ステークホルダリスト

どのような機関が関係するのかを洗い出す



2.ビジネス関係

ステークホルダリスト間のビジネス関係（契約や同意の取得、物販や役務等）



3.データリソースマップ

どのエンティティがどんなデータを持つことになるのか、どこに転送されるのかなど



4.トラストリソースマップ

ステークホルダ間の認証関係の有無、認証方法を整理する



5.データフローシーケンス

誰がどの情報をどういう順序で流通させるのかを時系列で記載する



6.法制関係図

遵守事項とその拠り所を明確にする。契約や法制の記載

1. ステークホルダリスト

ドライブレコーダビジネスの事例

目的：関与する個人、事業者の一覧表を作成することで、パーソナルデータの取り扱われる範囲を明確し、プライバシー原則などを遵守すべきプレイヤーに抜けがないかを確認する

名称	概要	ISO/IEC 29100での分類(次ページ参照)
ドライバー	ドライブレコーダーで録画した映像を提供する	PII principal
ドライブレコーダー販売事業者	ドライブレコーダーを販売する	非該当
データ蓄積事業者	ドライバーから提供された映像を蓄積管理する。映像加工（非個人情報化）をデータ加工事業者へ委託する。映像を購入したい事業者へ販売する。個人情報保護法上の個人情報取扱事業者に該当	Data controller
データ加工事業者	データ蓄積事業者から映像加工（非個人情報化）を受託する 個人情報保護法上の委託先に該当	Data processor
データ購入事業者	データ蓄積事業者から映像を購入する	非該当（PII扱わない）
通行人	ドライブレコーダーが録画した映像に映り込んでいる人	PII Principal

チェックポイント

- ✓ ドライブレコーダー販売事業者は、パーソナルデータの取り扱いについて、特段の役目を持たないのか？
- ✓ 通行人は、パーソナルデータの視点では、システムを構成する一構成者としてリストされる

ISO/IEC 29100におけるアクターとその役割定義

PII Principals

- PIIが関係する自然人

PII controller

- PII処理の行われる理由（目的）及び方法（意味）を決定する

PII processor

- PIIコントローラに代わってPII処理を実行し、またPIIコントローラの指示に従って動作し、規定のプライバシー要件を順守し、対応するプライバシーコントロールを実装する

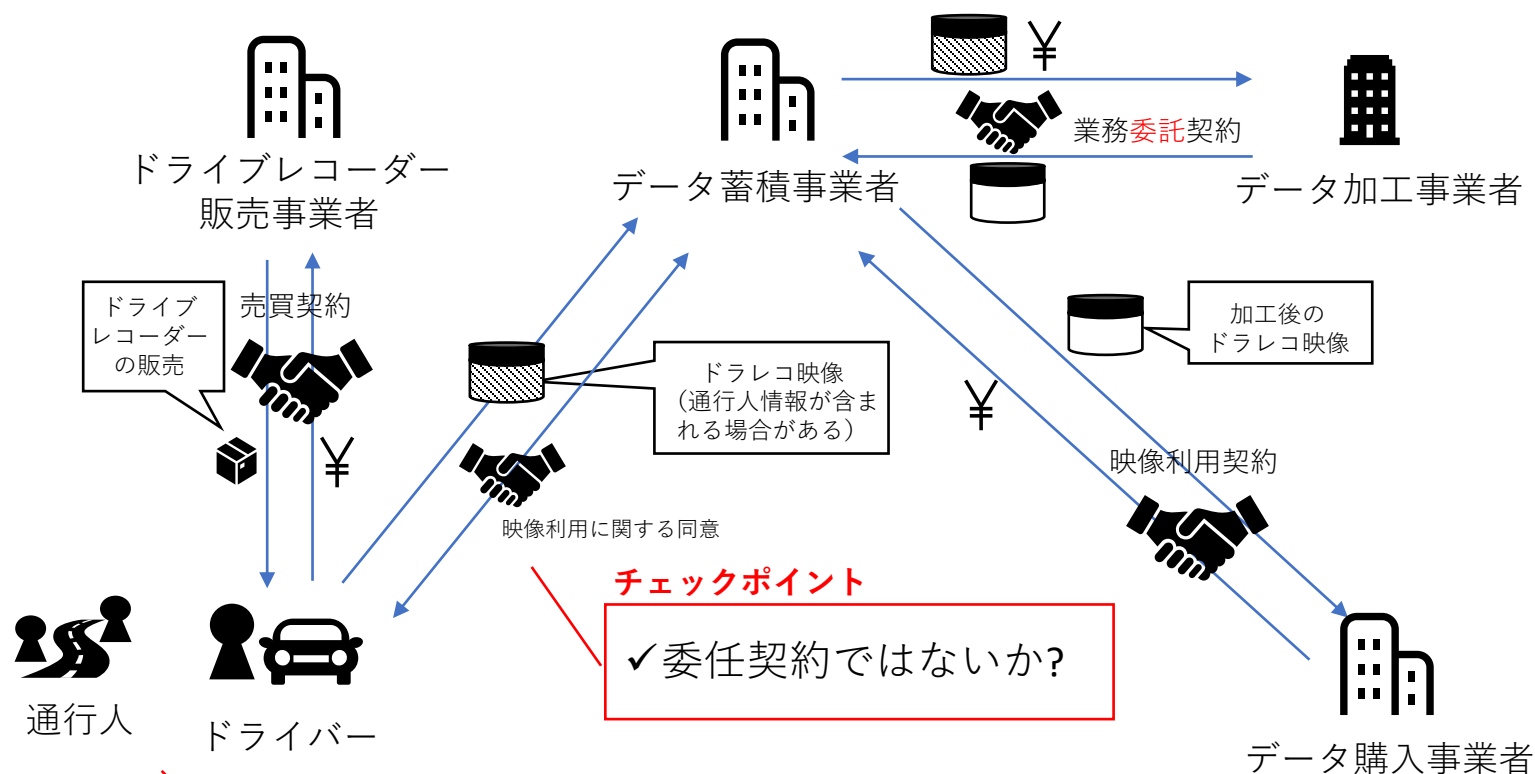
3rdParty

- PIIをPIIコントローラやPIIプロセッサから受け取ることができるが、処理はしない

2. ビジネス関係

ドライブレコーダビジネスの事例

目的：関与する個人、事業者間のビジネス関係（契約など）を明確化する



チェックポイント

✓委任契約ではないか?

チェックポイント

- ✓通行人とは、何の約定も、結ばれていない
- ✓プライバシー原則に照らしてドラレコ映像を利用している旨の通知が通行人他に対しては必要ではないか?

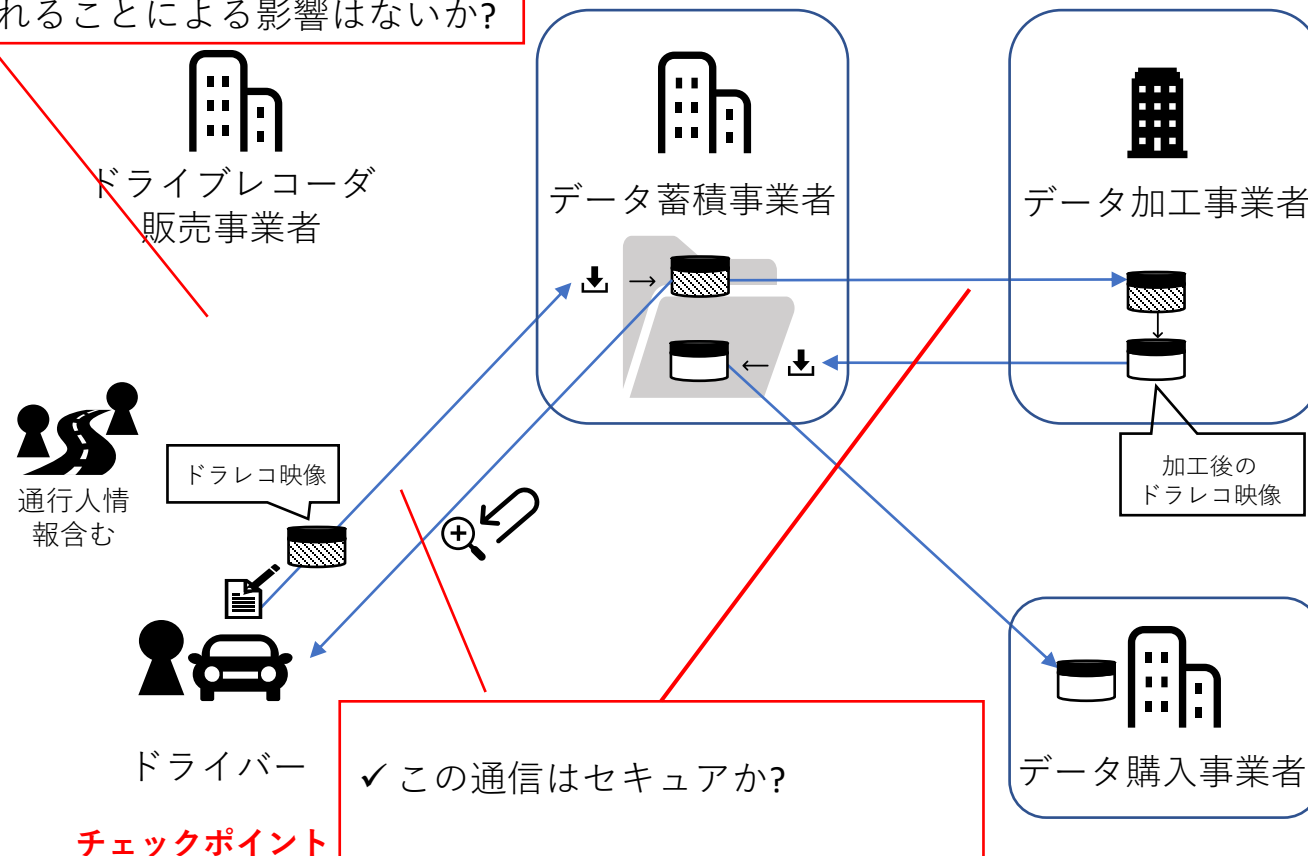
3. データリソースマップ

ドライブレコーダビジネスの事例

目的：パーソナルデータを含むデータセットがどこに存在するのかを明確にする。
事業遂行する上で、セキュリティを確保すべき箇所や、インシデント発生時の影響
範囲、事業譲渡や事業終了などに伴う処理範囲を明確に把握する。

チェックポイント

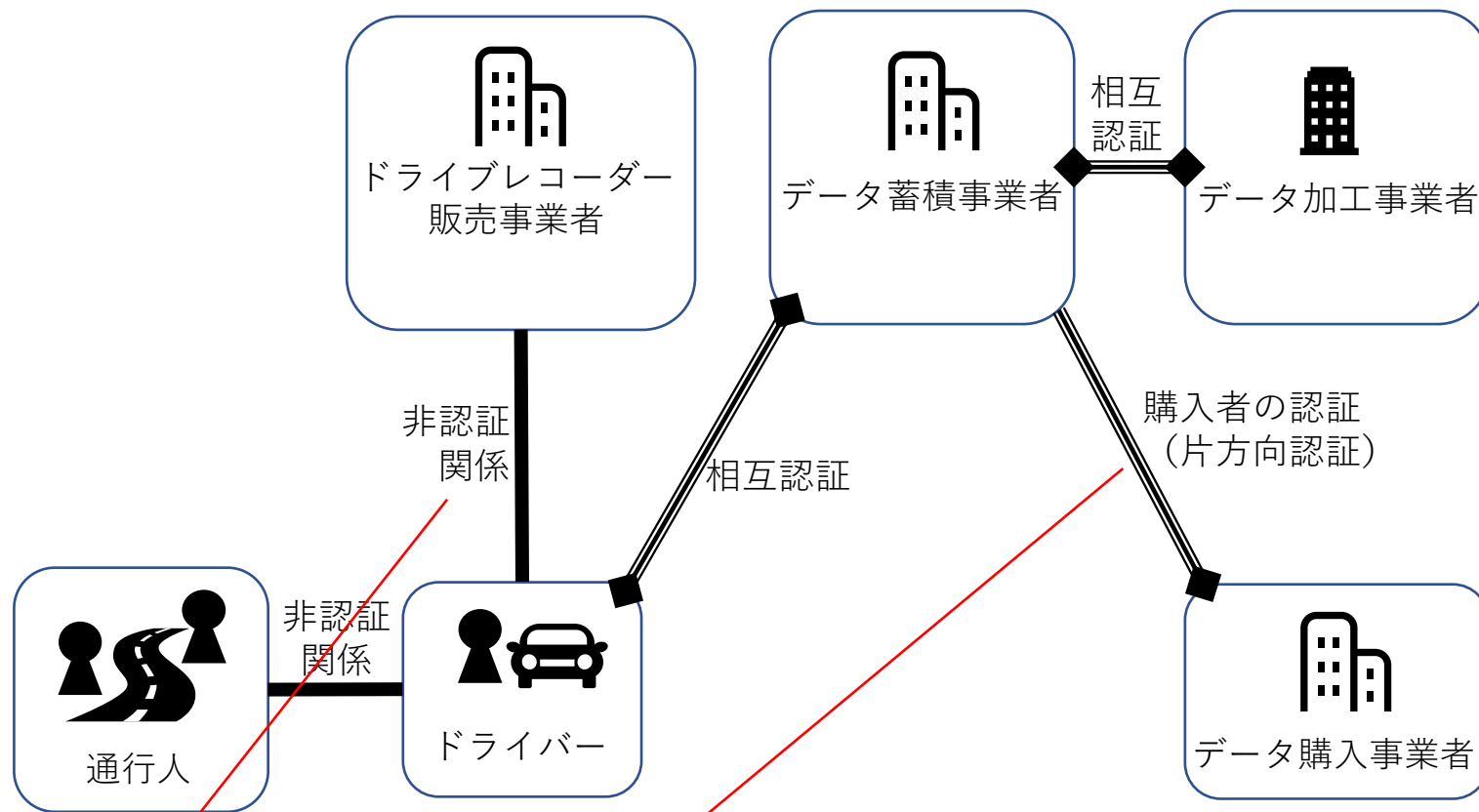
✓ ドライブレコーダ販売事業者はステークホルダーには入っているが、まったくパーソナルデータに関与しないのか？(例えば、機器番号がと個人情報情報が紐づけられることによる影響はないか？)



4. トラストリソースマップ

ドライブレコーダビジネスの事例

目的：関係者間の認証・認可の関係を明らかにする



チェックポイント

チェックポイント

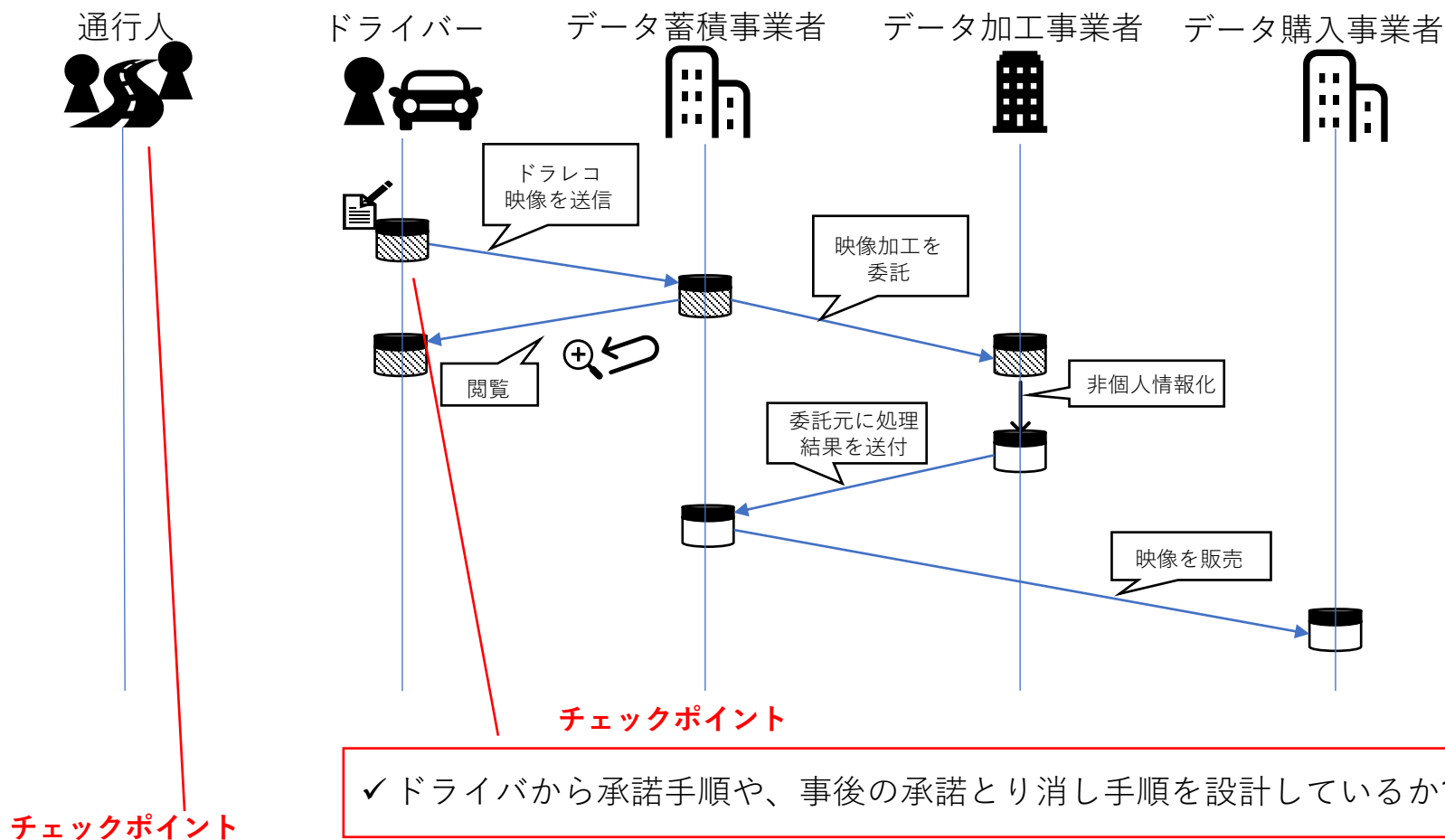
✓片方向認証としているが、相互認証は必要なのか？

✓ドライバーと、ドライブレコーダー販売事業者や通行人との間には認証関係が存在するのか？

5. データフローシーケンス

ドライブレコーダビジネスの事例

目的：関係者間でのデータセットの移動、処理フローを、許可などの手順を明確化する



チェックポイント

✓ 通行人は全くデータフローには入ってこないのか?

6. 法制関係図

ドライブレコーダビジネスの事例

目的：各ステークホルダの事業ならびに相互の契約などにおいて、遵守すべき法制の範囲を明確にする

	ドライバー	ドライブレコーダー販売事業者	データ蓄積事業者	データ加工事業者	データ購入事業者	通行人
ドライバー	NA	販売契約	映像利用に関する同意 個人情報保護法			
ドライブレコーダー販売事業者	販売契約	NA				
データ蓄積事業者	映像利用に関する同意 個人情報保護法		NA	業務委託契約 個人情報保護法	映像利用契約	個人情報保護法
データ加工事業者			業務委託契約 個人情報保護法	NA		
データ購入事業者			映像利用契約		NA	
通行人			(個人情報保護法)			NA

チェックポイント

- ✓ ドライバーとデータ蓄積事業者との間には映像利用に関する同意書が存在。そこには、二次利用（データの加工・販売）の旨の記載が必須であることがわかる。
- ✓ 通行人とデータ蓄積事業者の間には、個人情報保護法に関わる可能性があることがわかる。

研究開発編

研究開発編の構成

本書の作成にあたり、当協議会では、国内での有識者会合および実証実験事業者との会合を開催するほか、国際標準化の調査と推進を実行した。これらの活動および、その調査内容について報告する。各事業者がより深い検討や国際展開を検討する際の参考となることを想定している

- 第 1 0 章 パーソナルデータに関わる標準化
- 第 1 1 章 今後の進め方



第10章

パーソナルデータに関わる標準化

国際標準化等の推進活動

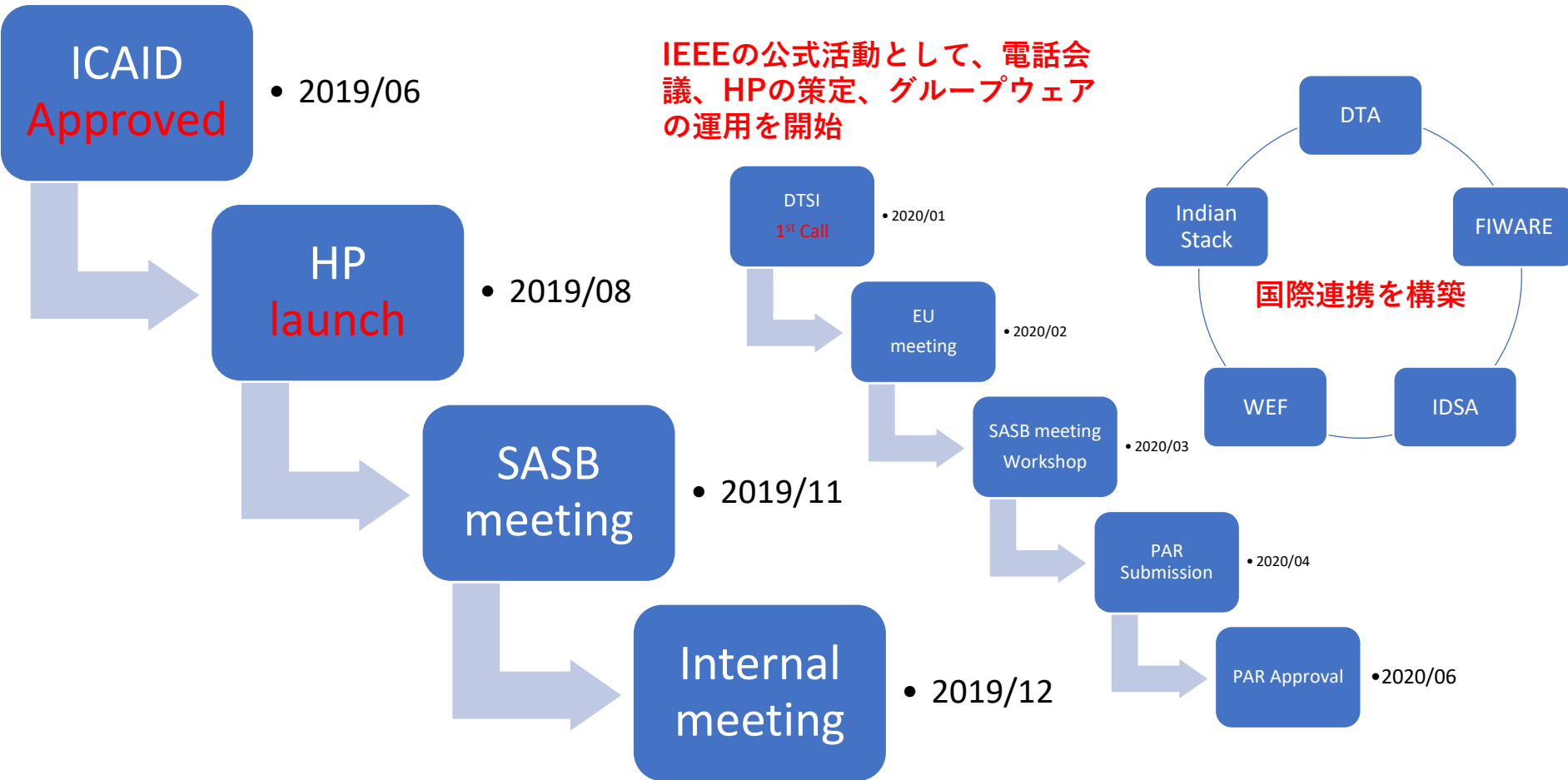
- 目的

- IEEE-SA DTSI（Data Trading System Initiative）活動におけるリーダーシップの発揮
- ISO/TC設立の動きに対する協力と連携
- W3C（World Wide Web Consortium）との関係構築の提案

- 実施事項

- IEEE-SA DTSI活動の推進
 - データ流通に関する標準化を行うPAR(Project Authorized Request)の起草を行うためのICAIDドキュメントが承認された。NesComへのプロジェクト開始提案を目指す
- ISO/TC設立の動きに対する協力と連携
 - DTA内に国際標準化推進委員会を発足させ体制を整えた
- W3Cとの関係構築の提案
 - DTAとしてW3Cに加入し、会員としてW3Cにおける標準仕様策定に参与できる体制を整えた

IEEE DTSI(Data Trading System Initiative)の設置 実施内容と今後の予定



新規WGの設置のため、ICComの承認によるDTSI設置を提案し承認を得た

PARの策定と承認がゴール

総務省の「データ流通に関する国際標準化の推進と関連動向調査」事業と連携しながらこの活動を持続的に推進している

データジャケットの国際標準化

- 目的

- データ取引においてデータジャケット（DJ）は、人間中心のダイナミックな創造活動の場を生み出すためのデータセットのダイジェストである。データジャケットを中心とした標準化インプット案の文書を作成して、将来の標準化への寄与を目指す

- 実施事項

参加者のコミュニケーションが重要なデータ市場において、データジャケット（DJ）を中心とした標準化文書を提示した。データの中身を明かさずにデータのダイジェストを共有可能とする記法、そして、DJを介したコミュニケーション過程とそのレギュレーション（信頼性・創造性を高めるための制約）、その他約束事、DJの利用事例、関連して有効となる技術について記載した。具体的には、以下を標準化対象とする

- DJが含む要素やその論理的位置づけ：コミュニケーションの前提知識となる
- 信頼性と創造性を高めるコミュニケーションIMDJ（Innovators Marketplace on Data Jackets）のプロセスと参加者同士の約束事
- DJの使用事例
- IMDJにおけるコミュニケーションと思考の支援技術



第11章 今後の進め方

今後の進め方

アーキテクチャの継続的な維持・発展

- 普及促進に向けた啓蒙
- アーキテクチャの持続的改版
- 国際標準化への寄与と推進
- 採用に向けたコンFORMANCEテストや認定の検討

メンテナンス

- 内閣府およびDTAの用意するホームページにて公開
- 改版や新しいユースケースの登録などのプロセスを明確化など



第12章 【付属資料】

付属資料

下記 2 種の成果物が、内閣府HP及びDTAのHPより公開予定である

リファレンスアーキテクチャ用語・定義書

- 本書公開時点で 83 用語、分類別で、全般 16、情報 15、データ 7、処理 14、契約・トラスト 12、事業モデル 12、その他 7

データジャケットの国際標準化報告

- 成果報告書とその概要書を公開
- 本ページ以降に概要を記載

データジャケットの国際標準化

Data Jackets as a Global Technology for Trading Data with Trust

Yukio Ohsawa (大澤幸生), Teruaki Hayashi (早矢仕晃章), Gensei Ishimura (石村源生), The University of Tokyo

A Data Jacket (DJ) is a piece of digest information of a dataset, that does not open the content of the data but includes the title, the abstract, and variables, that may represent the subjective expectation of data owner or potential data users about the utility of the data.

Definition a data jacket (DJ_i) for a dataset d , suffixed by i is defined by information corresponding to $DJ_i(d) := \{F_i(d), P_i(d), V_i(d), U_i(d)\}$ where

$F_i(d)$: the set of functions defined on the variables in $V_i(d)$

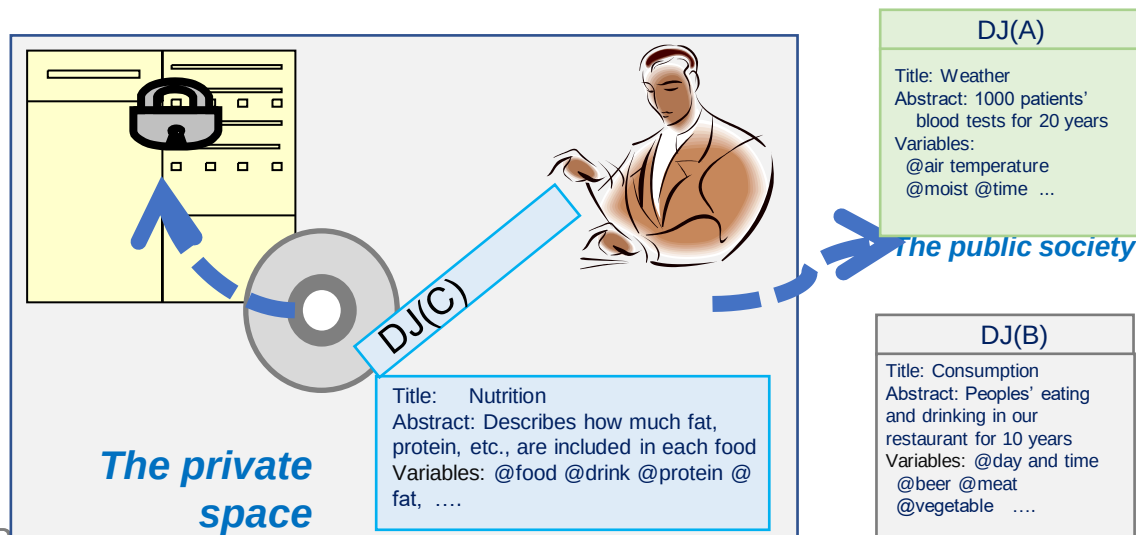
$P_i(d)$: the set of predicates that express relations among variables in $V_i(d)$

$V_i(d)$: the set of variables in $DJ_i(d)$

$U_i(d)$: the set of use cases of (d) and their values based on the information in $DJ_i(d)$.

In $U_i(d)$, hypothetical scenarios are expressed in logical formulae using the elements of $F_i(d)$, $P_i(d)$, and $V_i(d)$ or corresponding explanation by human(s) in natural language.

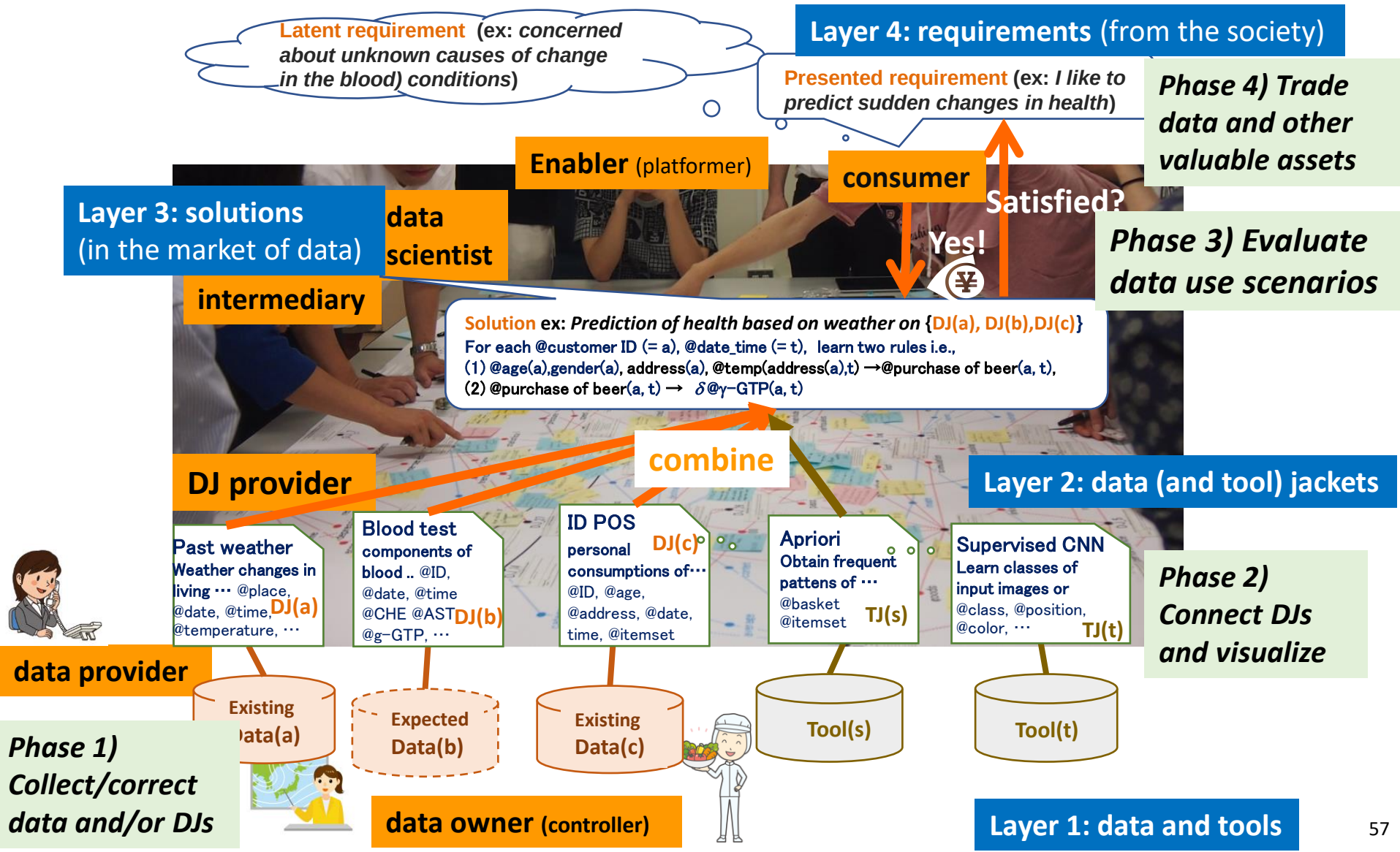
Ohsawa Y., Hayashi T., Kido H., Restructuring Incomplete Models in Innovators Marketplace on Data Jackets.
Magnani L., Bertolotti T. (eds) *Handbook of Model-Based Science*, 1015-1031 Springer (2017)



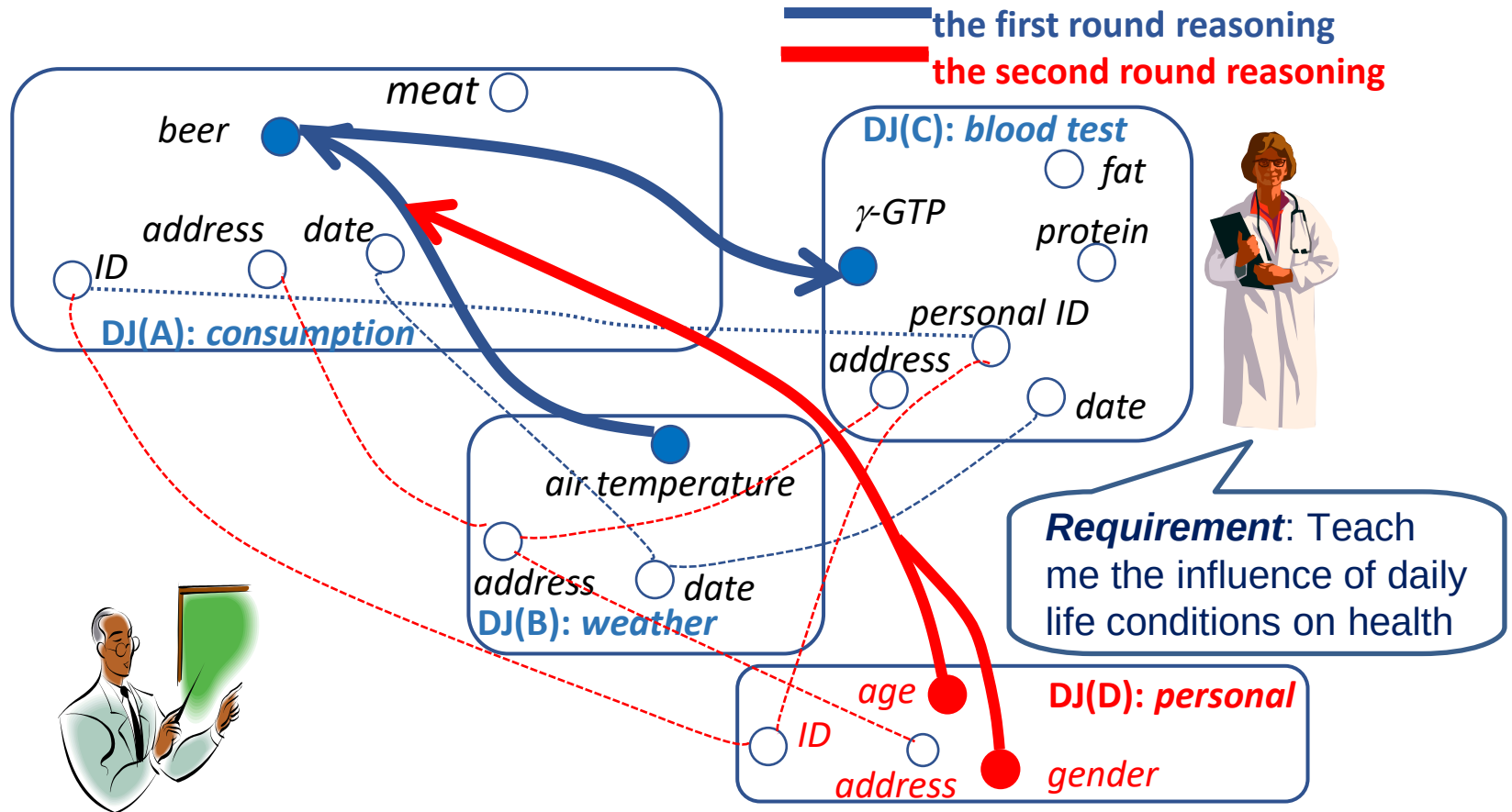
⑫	Power demand and weather	
Summary of data	Electricity demand, and temperature and rain data in [in Tokyo] the Tohoku Electric Power's managing regions from 2010 till 2011	
Attributes in the data	Region Year Month, date, and time Load of power	Monthly average temperature Lowest regional temperature Highest regional temperature Rain quantity of the day
Source and the methods for collection	The data of an electric power company and the data of the Meteorological Agency were suitably combined by Excel.	
Tool for analysis	Regression analysis	Decision tree
Expectation for analysis	Power load implies people's activities	The significant factors among others about weather, affecting the demand for power
Unexpected results	Weather is not so informative as expected, at least in summer and winter	The tree comes to be too complex. Data should be approximated e.g., by

The **Four Phases** crossing the **Four Layers** of IMDJ for the **Seven Stakeholders** in the market of data

Extending Ohsawa, et al, Innovators Marketplace on Data Jackets for Externalizing the Value of Data via Stakeholders' Requirement Communication AAAI 2014 Spring Symposium on "Big data becomes personal: Knowledge into Meaning, March 2014, Stanford (2014)



The abductive communication in IMDJ creating use cases of datasets



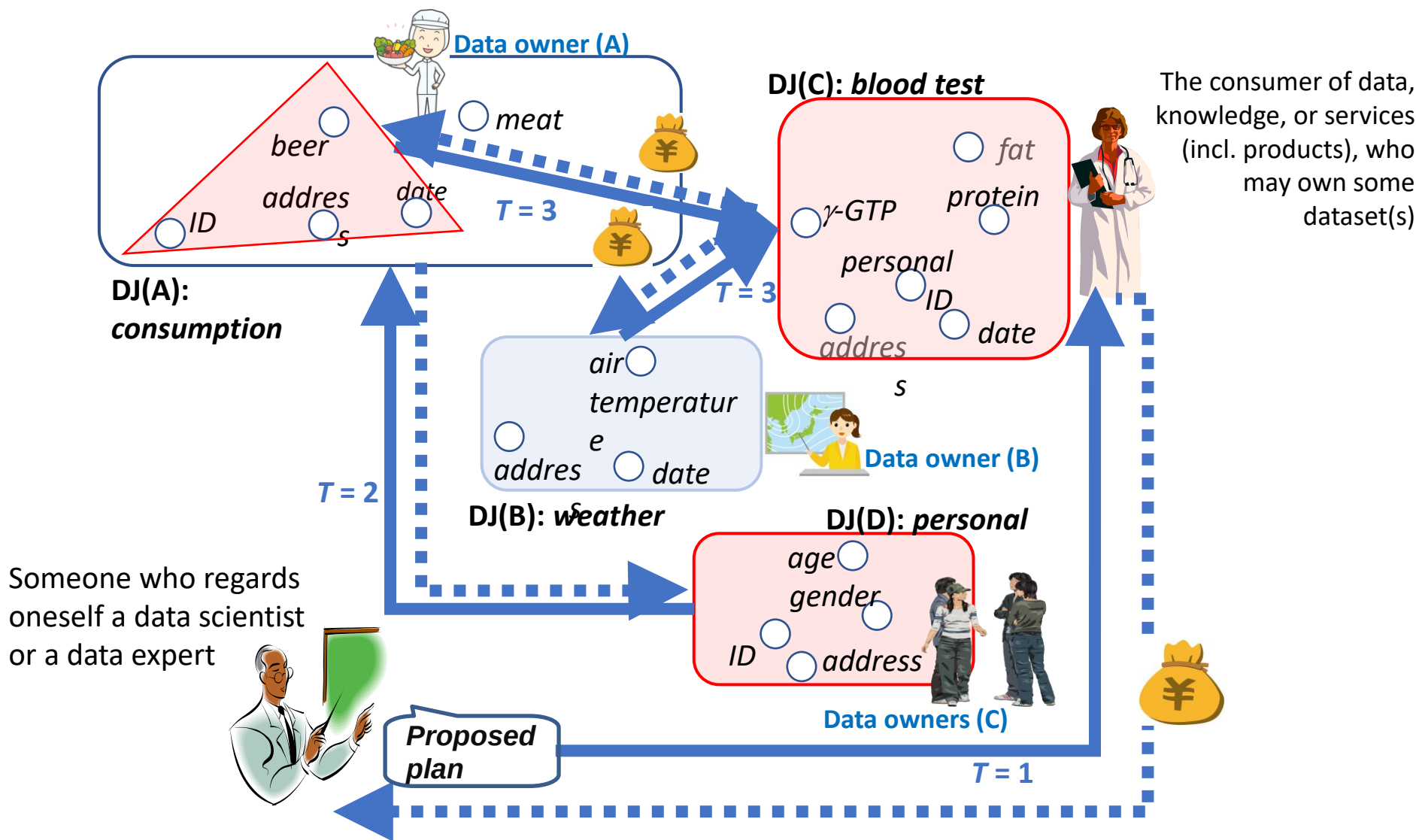
Proposed plan: Combine data behind DJ(A) though DJ(C), to learn

(1) `tempr_up(adr=Tokyo, date) -> beer_drink(adr=Tokyo, gndr:man, age:30, date)` (DJ: A, B, **D**)

(2) `beer_drink(Tokyo, man, 30, date) -> up_γ-GTP (Tokyo, man, 30, date)` (DJ: A, C, **D**)

Confidence increases (e.g. from 1% to 9%) by adding the red variables in the personal dataset **D**.

The stakeholders' value exchange after the abductive communication



Someone who regards oneself a data scientist or a data expert

The consumer of data, knowledge, or services (incl. products), who may own some dataset(s)

The definition of Trust in the context of IMDJ

The trust in/of IMDJ is established *iff* the following rights and reliabilities are ensured.

1. Right of DJ's providers in IMDJ (conditions promised to providers)

- DJ's providers are ensured that they can control how their DJ should be used.
 - The information of DJ is absolutely ***kept among defined stakeholders***. The definition is noticed to the providers in advance. If no definition is made, it means that DJ are open to the public. Participants in IMDJ are restricted to the defined stakeholders.
 - The provider can request ***withdrawal*** of their DJ and/or data sets and ***restriction*** of their circulation/use whenever they want after IMDJ. Their request is reported immediately to the related stakeholders, and is executed appropriately.

2. Right of participants in IMDJ (conditions promised to participants)

- Participants in IMDJ, ***and only they***, are ensured to be able to use any ideas of solutions for given problems proposed in the IMDJ, under the CROP principle (see the next page).
- They are ensured to express opinions freely in IMDJ except for illegal or unethical ones.

3. Reliability of DJ treated in IMDJ (conditions promised to DJ's users)

- DJ's users are ensured that DJ which they use are ***sufficiently*** reliable.
 - Data sets represented by DJ are authentic (not falsified, fabricated, nor stolen).
 - Any unethical intension such as exposure of confidential information is prohibited.
 - The reliability doesn't necessarily mean "accuracy" - DJ can be described subjectively.

4. Reliability of IMDJ in the society (conditions promised to the society)

- The society is ensured that IMDJ follows laws, social norm, and ethics (referring to "ELSI").

The rules for the participants in IMDJ

The CROP principle in IMDJ

Based on all the definition for trust in and of IMDJ (see the last page), all the ideas created in IMDJ are right-protected under the boundary defined by the Controlled (in or after the session) Reach Of the Presentation.

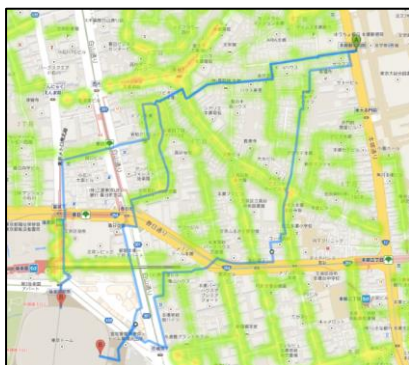
- (1) Each participant, and each participant only, has the right to use any idea for any solutions, and to use related DJs for any problems proposed by the others in the IMDJ freely.
- (2) The above right is assured only if it is exercised fairly and ethically, and it is allowed by all of the providers of the referenced DJ.
- (3) Any proposer of any idea in the IMDJ can request the restriction of the rights of the others at any time later, and this should be realized as far as possible.
- (4) However, the proposers have to accept that the use and propagation of their idea might not be able to be interrupted by their ex-post facto requests.

The cases of IMDJ in businesses and sciences

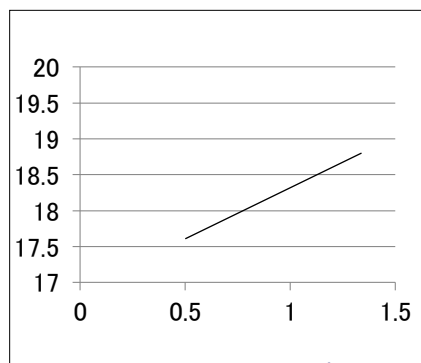
Six cases: bold letters show datasets corresponding to the combined DJs

- A. Visualized safe walking paths by streetlight locations + Google maps
- B. Discovered positive correlation of stock price vs the number of payed holidays
- C. Football coaching system from **full-view** video images of a soccer stadium
- D. Change explanation in markets from **POS** or stock chart data
- E. Change explanation in earthquakes diverting from D using DJ store (next page)
- F. City planning in Tokyo and Yokohama using datasets from various industries

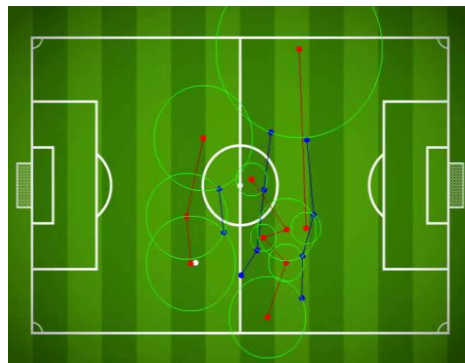
A



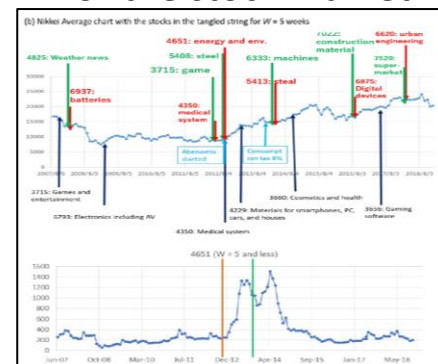
B



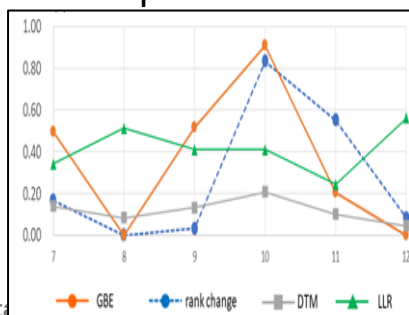
C



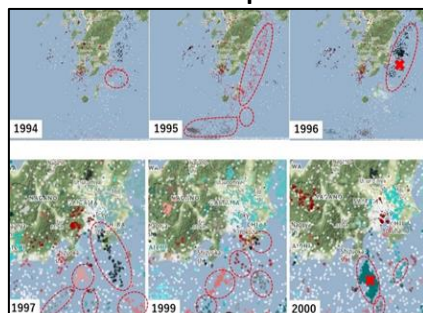
D for the stock market



D for supermarkets



E CE in earthquakes

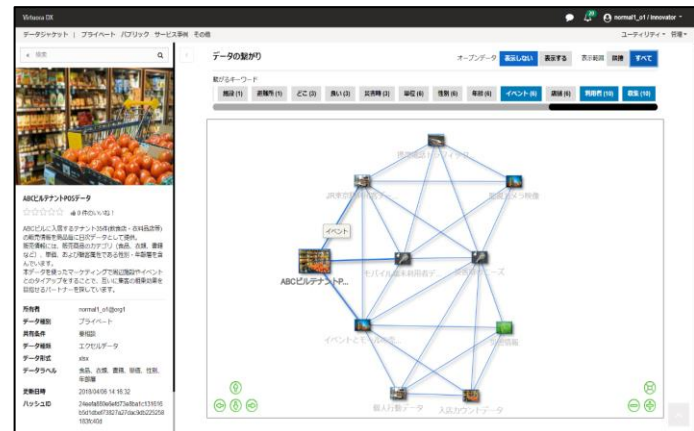
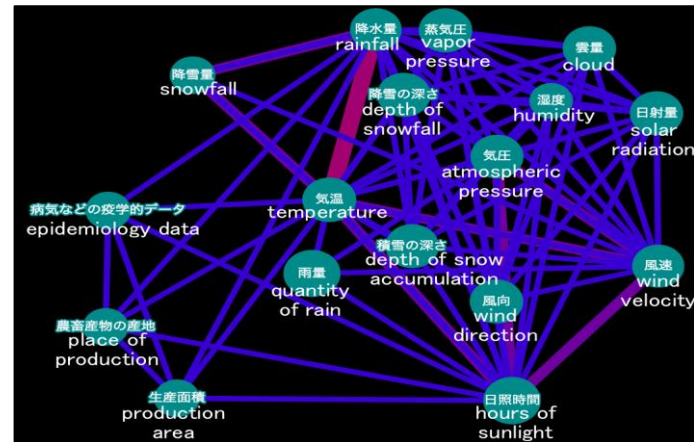


F IMDJ in Marunouchi Data Consortium



ing
nce

- C. **Web-based IMDJ:** for supporting the communication on DJs reducing the load for crossdisciplinarity data collaborations.
- D. **Virtuora DX:** A digital transformation system introducing DJs and visually connecting them, and DJ Store functions.



EOF