

# MyDataとAI

2018-12-10 橋田浩一



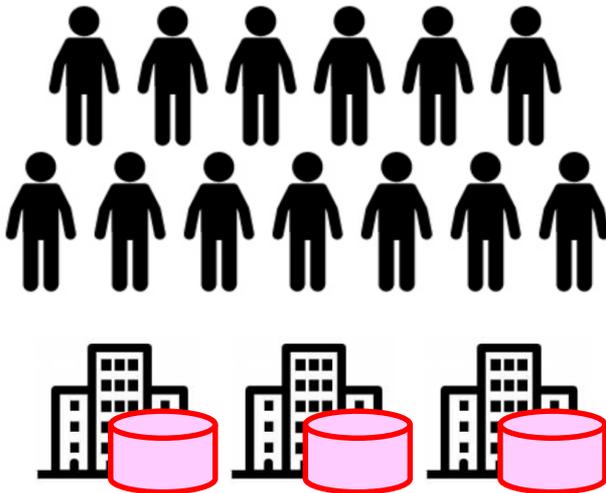
# 自己紹介：橋田 浩一

- 1981年 東京大学 理学部情報科学科 卒業
- 1986年 同 大学院理学系研究科 修了(理学博士)
- 1986～2001年 電子技術総合研究所
- 1988～1992年 (財)新世代コンピュータ技術開発機構  
に出向
- 2001～2013年 産業技術総合研究所
- 2013年～ 東京大学 大学院情報理工学系研究科ソー  
シャルICT研究センター 教授
- 2017年～ 理化学研究所革新知能統合研究センター社会  
における人工知能研究グループ分散型ビッグデータチー  
ムリーダーを兼任
- 専門は自然言語処理、人工知能、認知科学。現在の主な  
研究テーマはパーソナルデータの分散管理と意味的構造  
化およびそれに基づく人工知能。

# パーソナルデータの管理運用

## MyData

- 事業者にてデータを集約
  - ◆ 管理が集中するので危険
  - ◆ 本人同意だけでデータが使えないので不便
  - ◆ データが散在して価値が低い



- 本人にてデータを集約
  - ◆ 管理が分散するので安全
  - ◆ 本人同意だけでデータが使えるので便利
  - ◆ データが名寄せされて価値が高い



# データ管理を個人に分散 → 安全

- 集中管理されている1千万人以上の価値の高いデータが狙われる。
- そのデータをまとめて盗むコストは1億円未満。
- つまり1人分につき10円未満。
- 各個人から直接データを盗むコストは千円(1千万人のデータをまとめて盗む場合の100倍)をはるかに越える。
- それに、1千万人の各々からデータを盗もうとしたら、途中で発覚して逮捕されるのは必定。
- 各個人が本人の全データを管理していてその価値が集中管理に含まれる1人分の断片的なデータの100倍だとしても、集中管理された1千万人分のデータをまとめて盗む方がはるかに低コストかつ低リスク。

# MyData: 本人主導のパーソナルデータ活用

- 個人向けサービスの価値向上にMyDataが必要
  - ◆ パーソナルデータの使用は原則本人同意で
  - ◆ 無数の個別サービスにおけるデータの共有・活用の管理は現場に分散させるしかない
  - ◆ サービス受容者にとっての価値がサービスの価値だからその意思に沿ってデータを運用すべし
- 個人向けサービスが価値の大半(国内で600兆円/年)を生む
  - ◆ 生活者向けサービス(家計消費) 300兆円
  - ◆ 勤労者向けサービス 200兆円?
  - ◆ 家事や育児などの無償サービス 100~140兆円

# AIの運用と開発にMyDataが必須

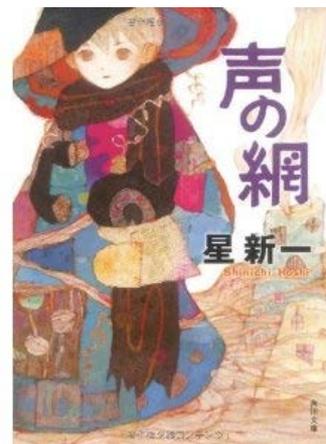
- AIが生み出す価値もほとんどが個人向けサービスによる
- AIの運用にMyDataが必須
  - ◆ AIが各個人に合った良質のサービスを提供するには、当該個人に関するリッチなデータが簡単に使える必要あり
- AIの開発にMyDataが必須
  - ◆ 多数の個人からデータを収集して分析するにも本人の意思でデータが簡単に使える必要あり

# 日本でも進むMyData

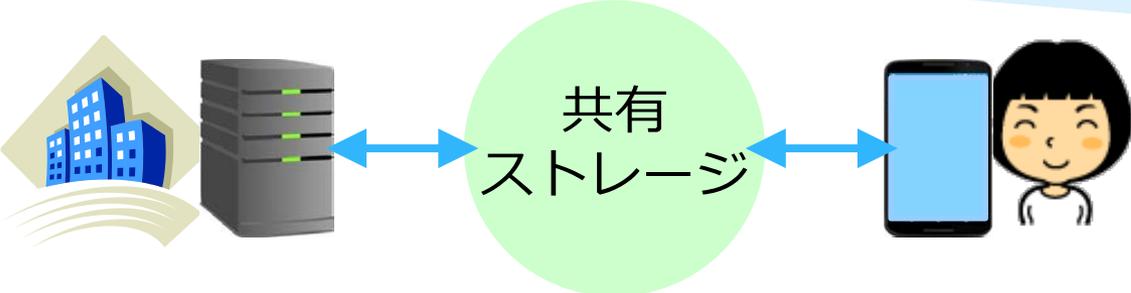
- 2005～2025年：医療制度改革等
  - ⇒ **ヘルスケアデータのポータビリティ**
  - ◆ 医療機関や介護施設の間でのデータ共有が必須に
  - ◆ 2020年から薬剤情報や特定健診等のデータをマイナポータルで本人に提供
- 2018年春：改正銀行法施行 ⇒ **購買データのポータビリティ**
  - ◆ API公開を銀行に事実上義務付け
  - ◆ キャッシュレス化+電子レシート
- 2019年：情報銀行・情報信託サービスの開始
  - ◆ 三菱UFJ信託銀行、三井住友銀行、電通、…
- 2021年：大学入試改革 ⇒ **教育データのポータビリティ**
  - ◆ eポートフォリオ(電子学習記録)を生徒本人が管理運用
  - ◆ eポートフォリオを生涯に拡張：スタディ・ログ
- 2020～2021年：個人情報保護法再改正
  - ⇒ **一般的データポータビリティ**

# PDS: Personal Data Store

- パーソナルデータを本人の意思で共有・活用する仕組み
- MyDataの必要条件
- 概念そのものは部分的には古い:
  - ◆ 星新一(1970) 声の網. (情報銀行)
- 主な機能
  - ◆ データのアクセス**権限定義**
  - ◆ データの**保管**
  - ◆ アクセス権限を持つ者によるデータの**共有**
  - ◆ アクセス権限を持たない者からのデータの**秘匿**



# パーソナルデータ管理運用法の分類



Personium、Cozy、OpenPDS、HAT、Meeco、SolidなどほぼすべてのPDS、顧客のデータを預る情報銀行

改ざん防止の仕組みが必要

アクセス権限を持たない者からのデータの秘匿

専用サーバなし

データのアクセス権限の定義と管理

大規模情報漏洩の防止

共有ストレージ	ID連携	ブロックチェーン	権限定義	保管	共有	個人間共有	秘匿	安全	安価
専用	なし			✓	✓	✓			
非専用	なし	なし	✓	✓	✓	✓	✓	✓	✓
なし	なし	なし	✓	✓	✓		✓	✓	✓
なし	あり			✓	✓				
		あり	✓						

PLR

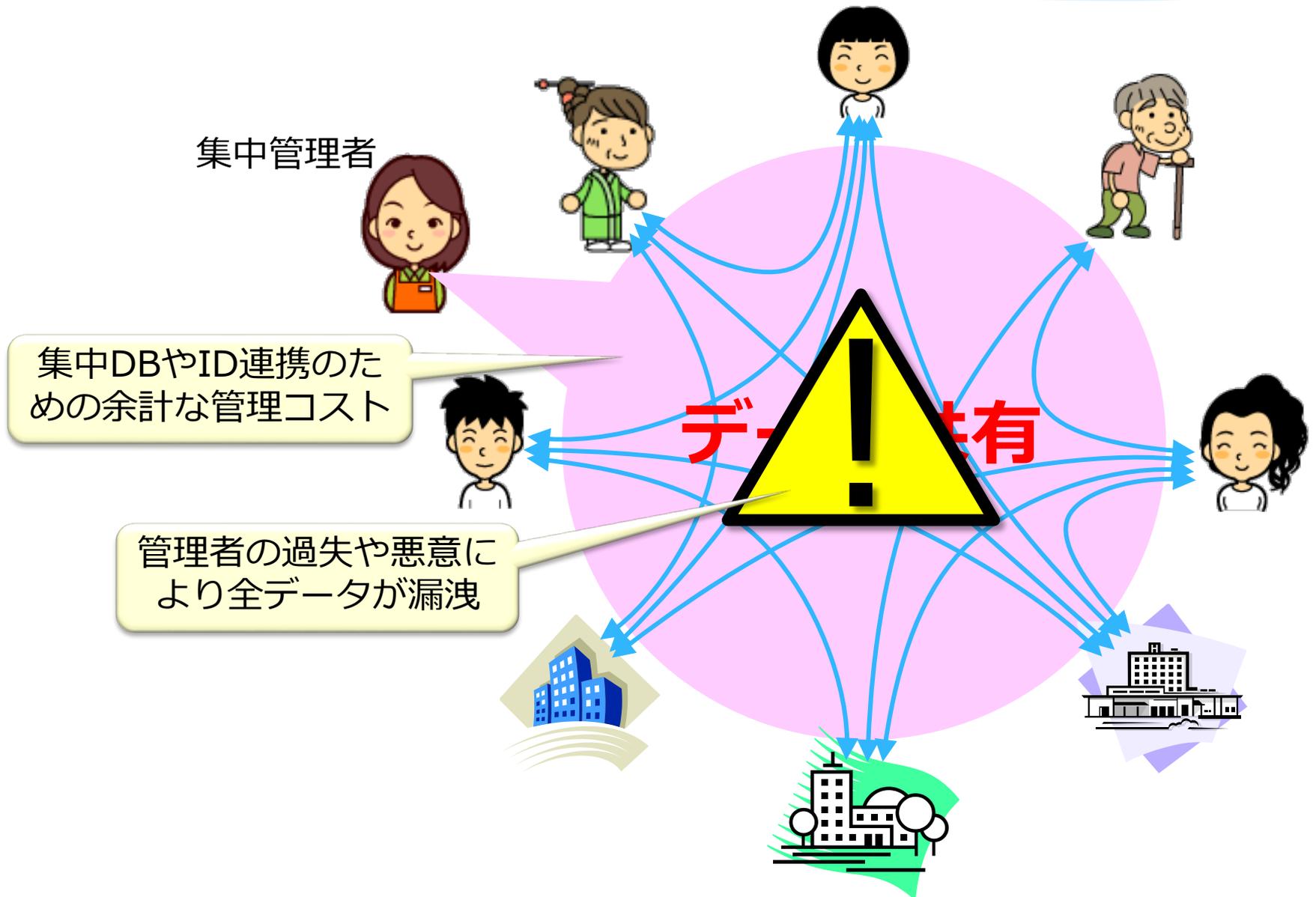
digi.me、CitizenMeなど

Facebookなど

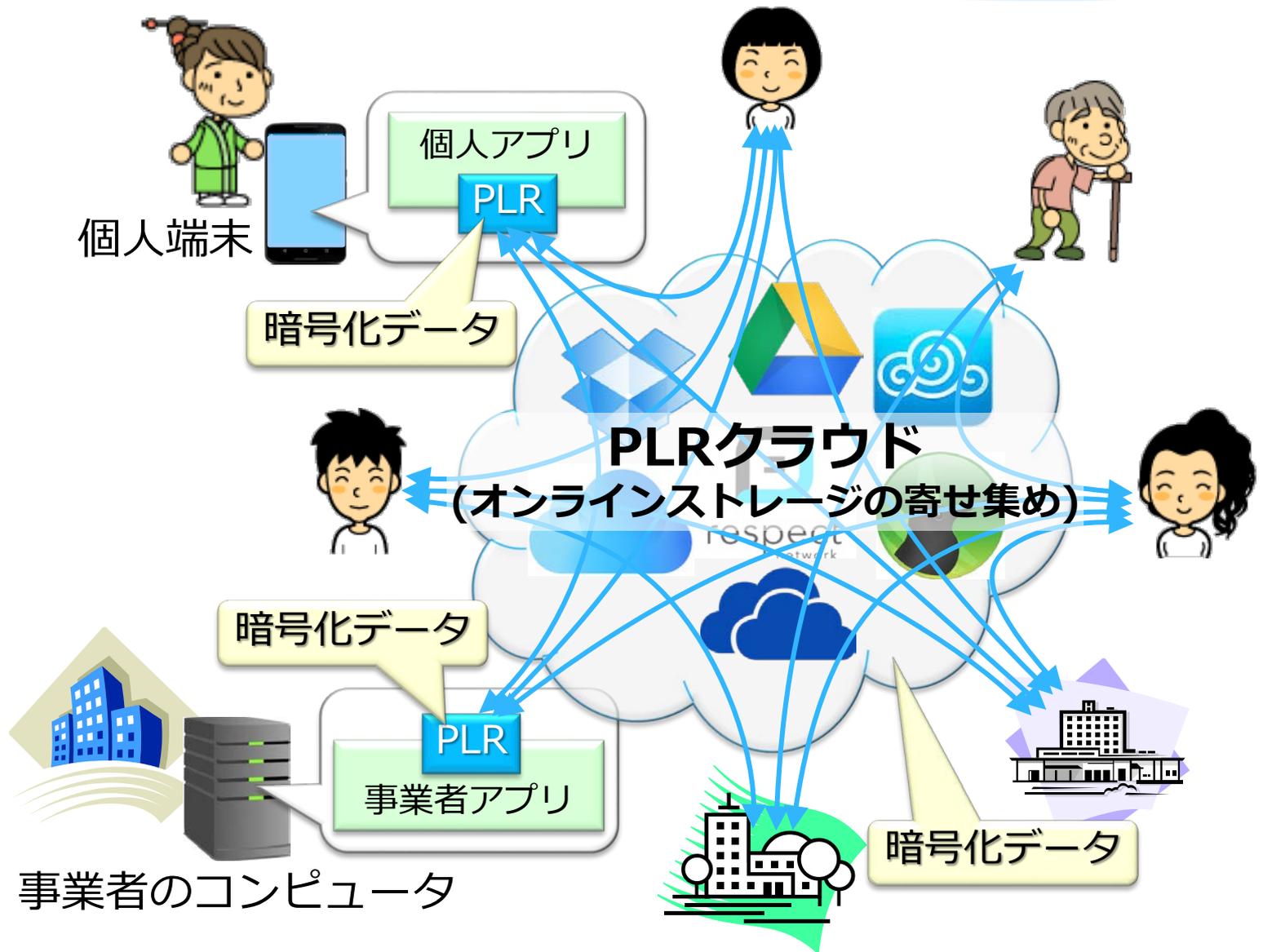
MedRec、MyDeeなど

専用共有ストレージ方式、ID連携方式と組合せ可

# データの集中管理



# PLR: Personal Life Repository



# PLRは安全

- 分散管理: データの管理者はデータ主体本人のみ
  - ◆ 本人同意がなければデータへのアクセスが技術的に不可能
  - ◆ **集中管理者の過失等による全データの漏洩があり得ない**
- DRM (デジタル権利管理)
  - ◆ 暗号化 + アプリの機能制限
  - ◆ **過失による情報漏洩があり得ない**
    - \* 個人: PLRアプリは平文データを保存・送信できない
    - \* 事業者: 不正なアプリをOSが排除すれば、不正なOSのインストールを防ぐ通常の管理でセキュリティを安価に担保できる
- 紙やCDでデータを受け渡すより安全
  - ◆ 端末が他人の手に渡ってもデータを取り出されない
  - ◆ データの開示先を制限できる

# マッチング

- 個人のニーズに適合するサービスの選定
- ニーズ = 取引条件 + サービスの評価 + ...
- CRM (customer relationship management)
  - ◆ 事業者がマッチング
- **VRM** (vendor relationship management)
  - ◆ **顧客がマッチング**

# マッチングアルゴリズム

## 下記の2種類の処理を統合

サービス参加者の取引条件を組合せて利得を全体最適化

● 協力ゲームの求解?

サービス評価と個人属性のうち既知のものを入力して未知のサービス評価を推定

● 制限ボルツマンマシン?

### サービス受容条件

目的: 私の健康管理  
開示データ: 病歴、服薬、  
バイタル、運動  
支払: ≤300円/月  
相手の信用度: ≥8

個人

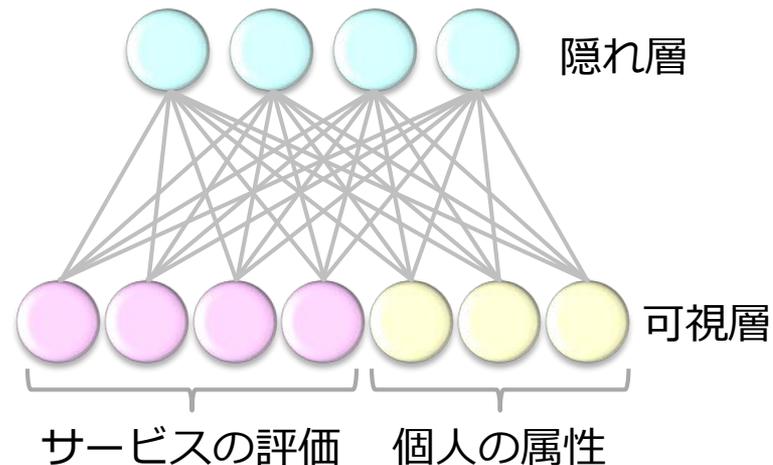


最適結合

### サービス提供条件

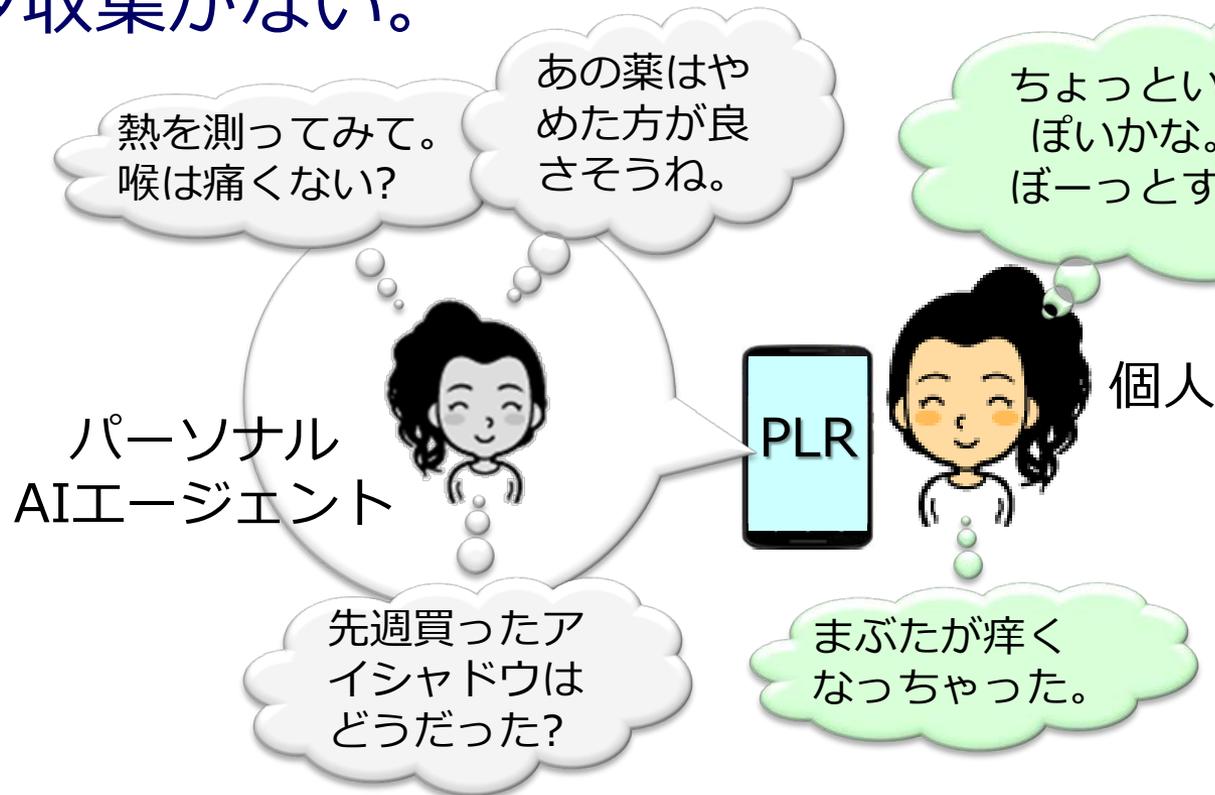
目的: 顧客の健康管理  
取得データ: 病歴、バ  
イタル、運動  
価格: 200円/月  
信用度: 11

事業者



# パーソナルAIエージェント

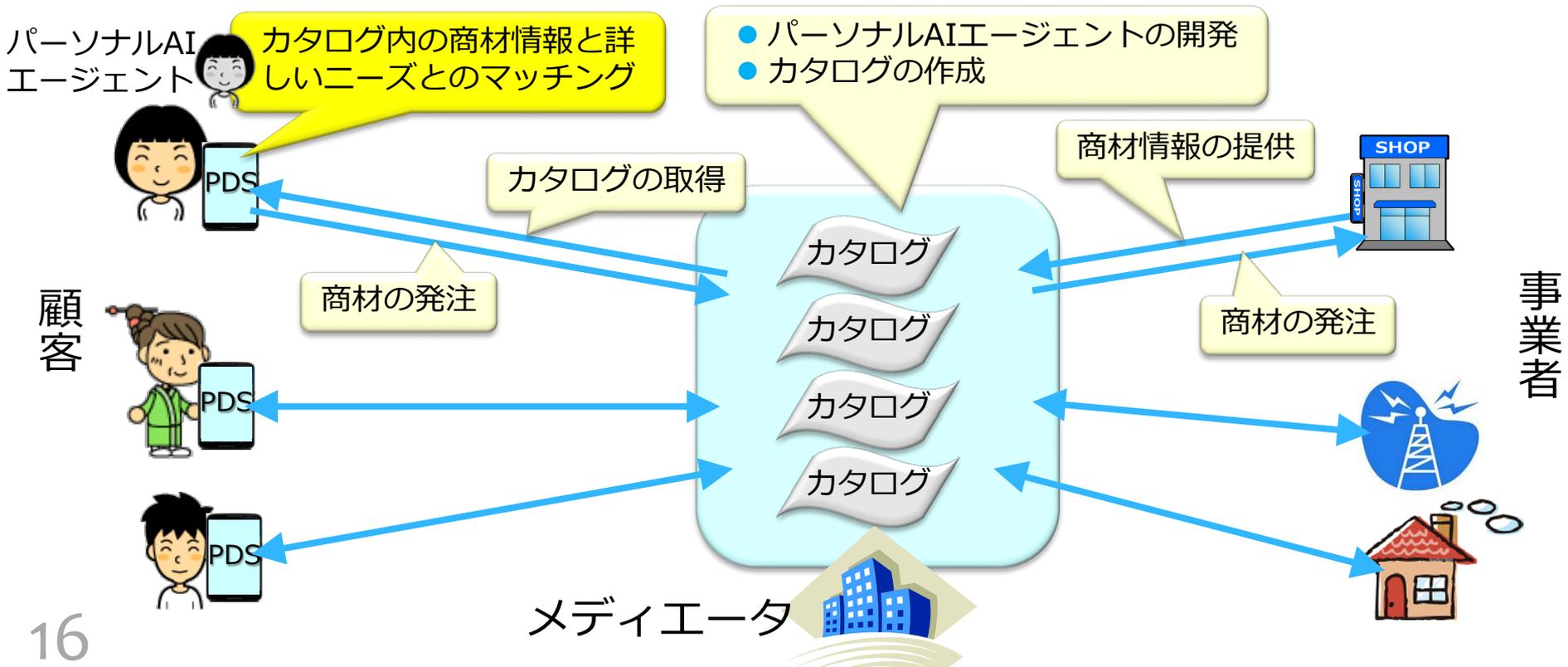
- PLRに付随し個人に属するパーソナルAIエージェントが本人から情報(商材の評価や本人の属性)を得てPLRに蓄積してマッチング(本人に適した商材等の選定)に活用。
- Amazon EchoやGoogle Homeと違って事業者によるデータ収集がない。



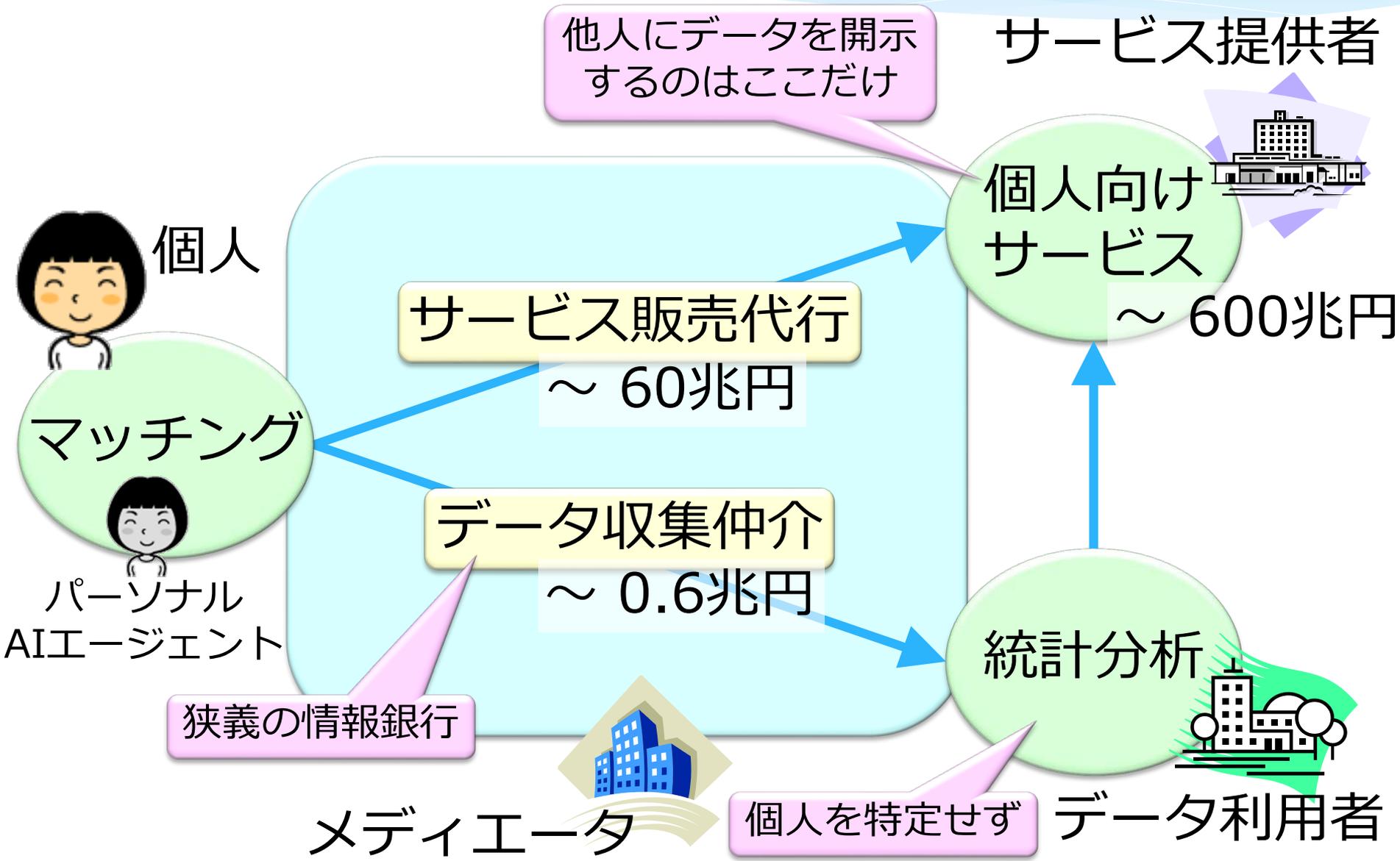
# 分散VRM

各個人のアプリが商材のカタログを個人端末にダウンロードして詳しいニーズと商材の情報をマッチングした結果に従ってメディエータ経由で商材を購入

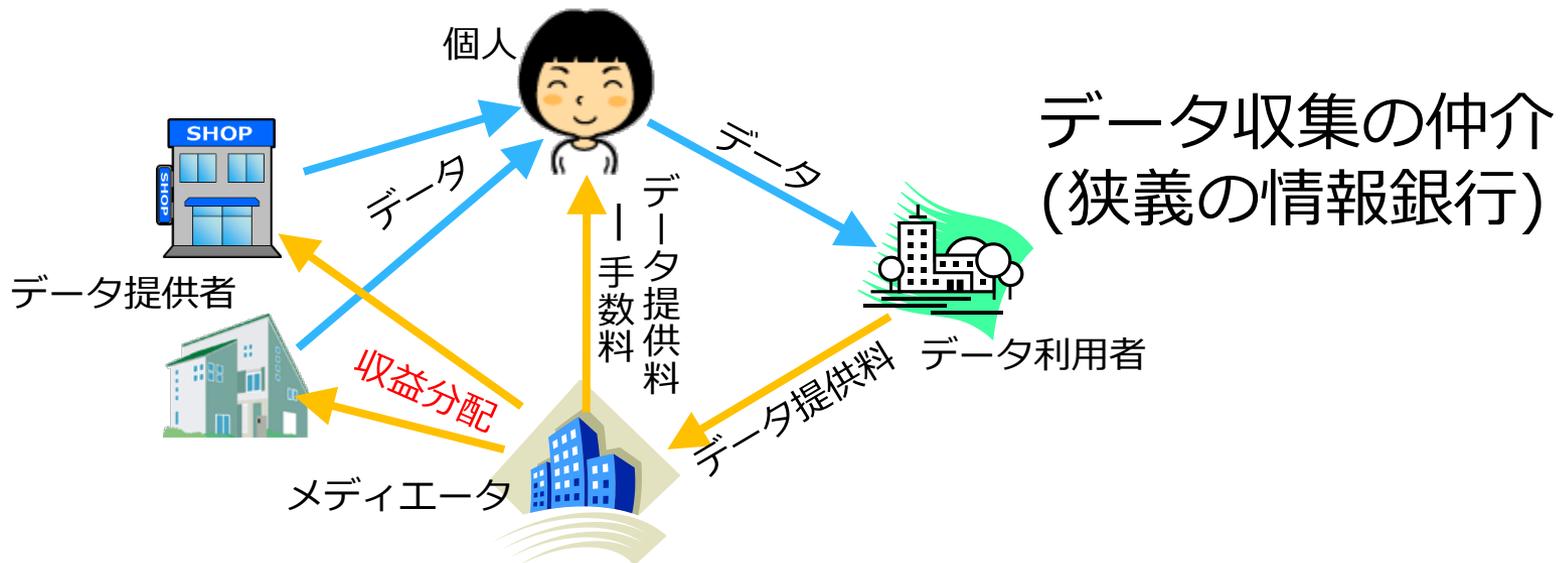
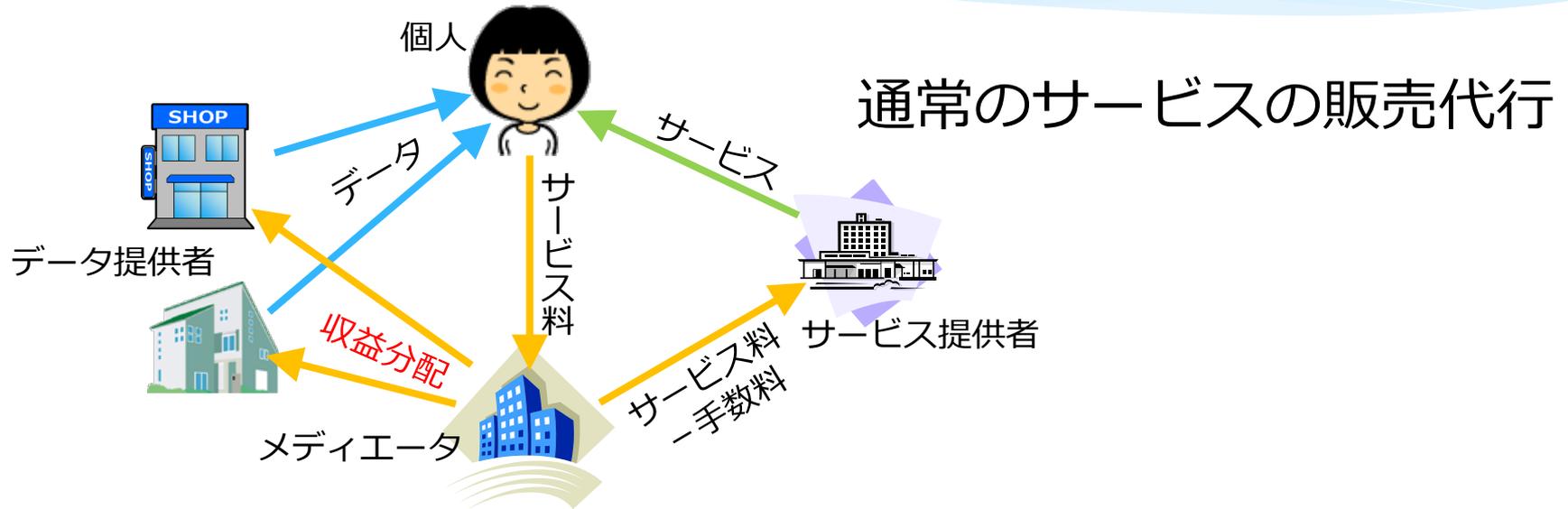
- 詳しいパーソナルデータを本人だけが使うので高精度かつ個人にとっても事業者にとっても安全
- メディエータにパーソナルデータを開示せずにメディエータのサーバでマッチングすることも可能



# パーソナルデータの活用と価値



# データポータビリティと収益分配



# パーソナルデータエコシステム

データ活用を**分散PDS**で追跡

価値を貢献  
度に応じて  
分配

事業者はパーソナル  
データを保管するリス  
クとコストを免れる

**分散PDS**で安全・  
安価にデータ共有

- 個人アプリでのマッチング
- 通常の個人向けサービス

**一次利用** » 二次利用

社会全体で  
価値が増大

パーソナル  
データを本  
人に集約

個人ごとにデータを名寄せ

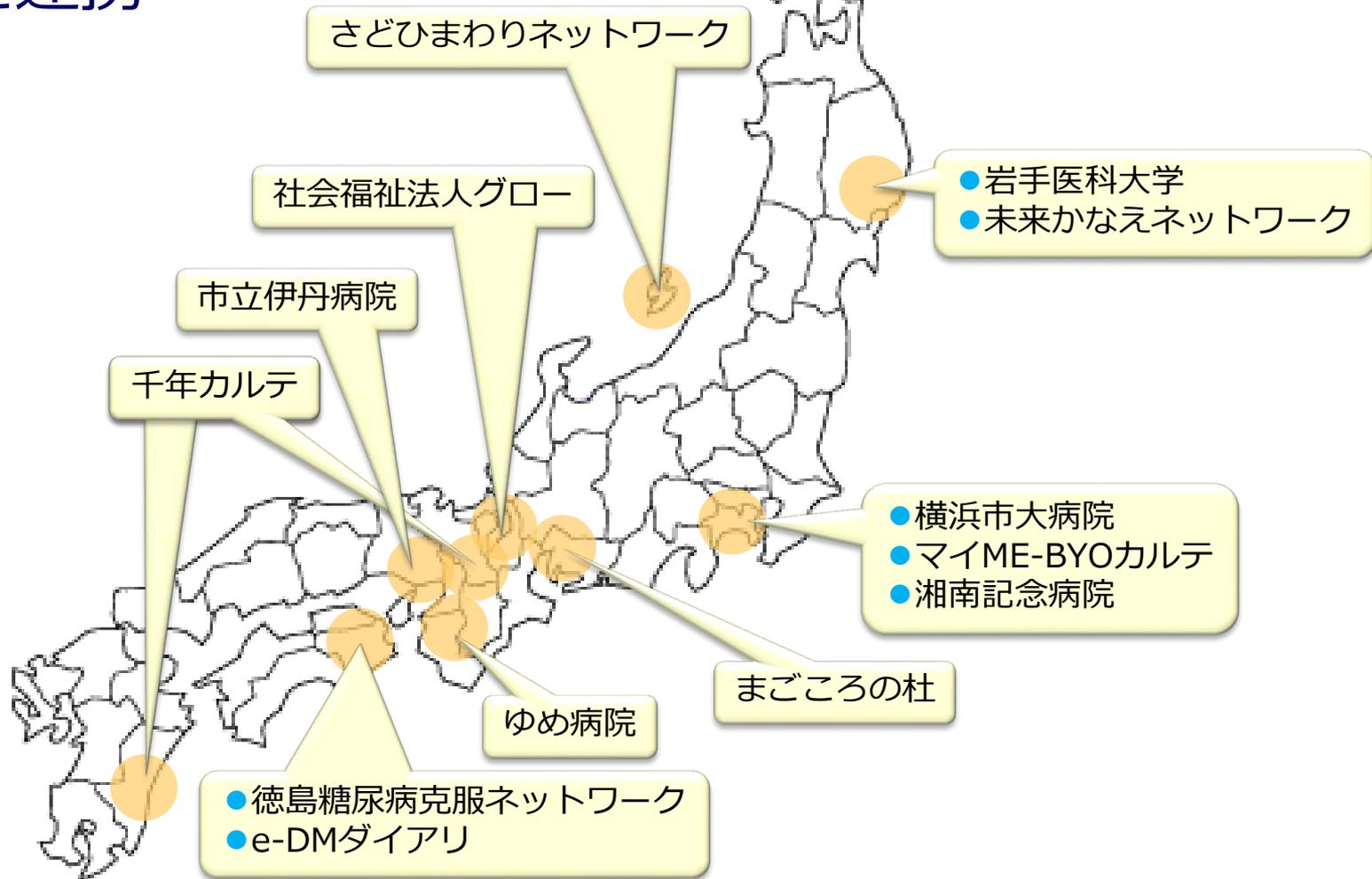
# まとめ

- MyData
  - ◆ 価値の大半を生む個人向けサービスの質を高める
  - ◆ AIの運用と開発に必須
- メディエータ(分散情報銀行)
  - ◆ パーソナルAIエージェント: マッチング + 購買
  - ◆ 販売代行: 国内市場規模 > 60兆円/年
  - ◆ 収益分配 → パーソナルデータエコシステム
- マッチング
  - ◆ ゲームの求解 + 協調フィルタリング?
  - ◆ **AIの要素技術のうち経済的価値が最大**
  - ◆ **中立性・公益性・公正競争の確保**
    - \* オープンソース + オープン学習データ
    - \* パレート最適性?
      - 適正価値での売買など

# 付録: PLRのユースケース

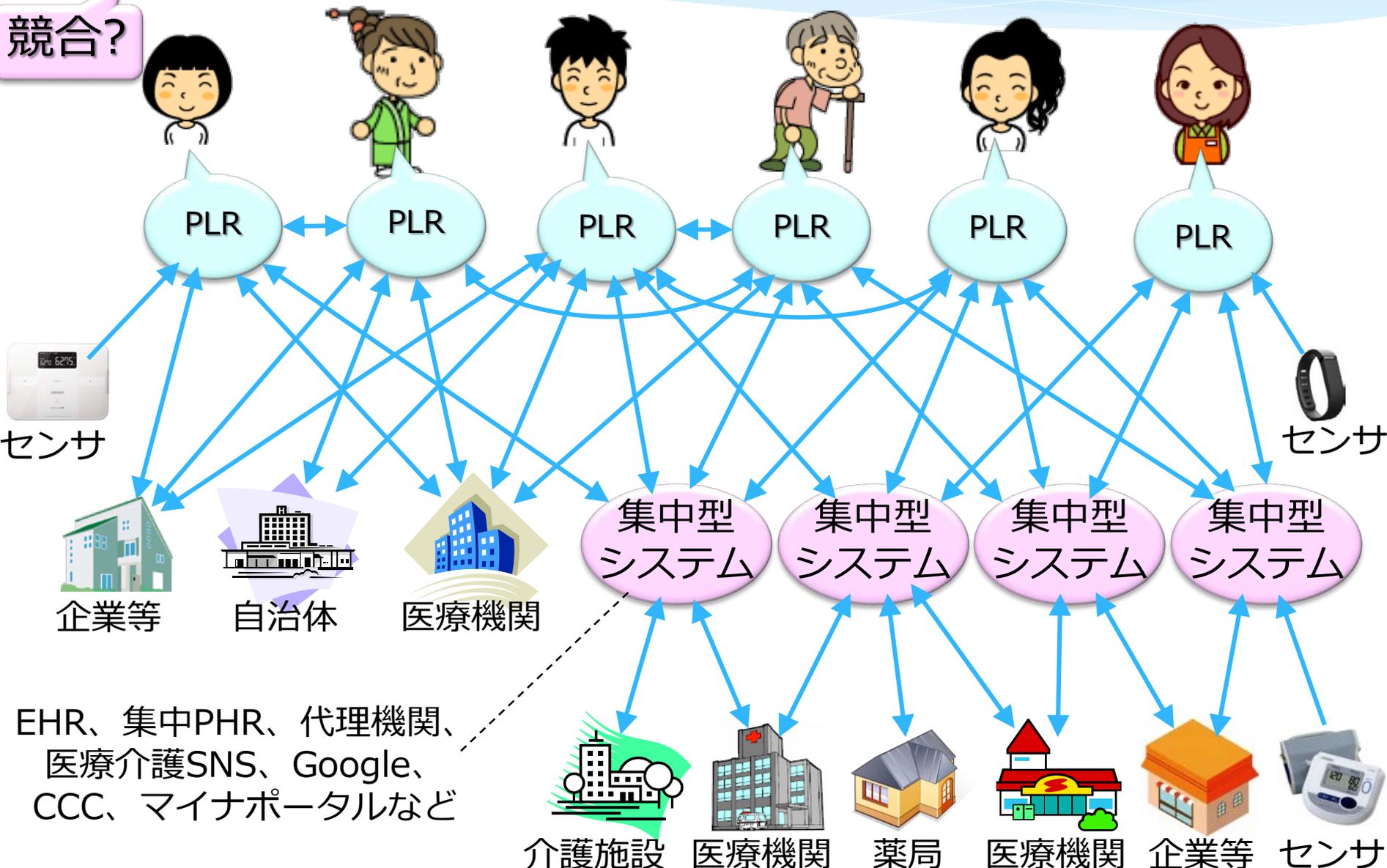
# 医療等での活用(AMEDプロジェクト等)

各実証フィールドにおいて既存の地域医療連携システムや病院とPLRとを連携



# 事業者の間の連携を個人がPLRで仲介

競合?



EHR、集中PHR、代理機関、  
医療介護SNS、Google、  
CCC、マイナポータルなど

# フレイル予防

PLRカード



個人



フレイルチェック

- 地域データ解析
- 健康づくり・介護  
予防政策の立案

役所

市区町村

個人アプリ

PLR

PLR

地域健康管理  
システム等

中継アプリ

PLR

PLRクラウド

PLR

介護アプリ

介護施設

中継事業者



PLR

分析アプリ

PLR

事業者アプリ

PLR

電カル等

分析事業者



顧客別サービス、  
商品開発、事業  
戦略立案、他

企業等

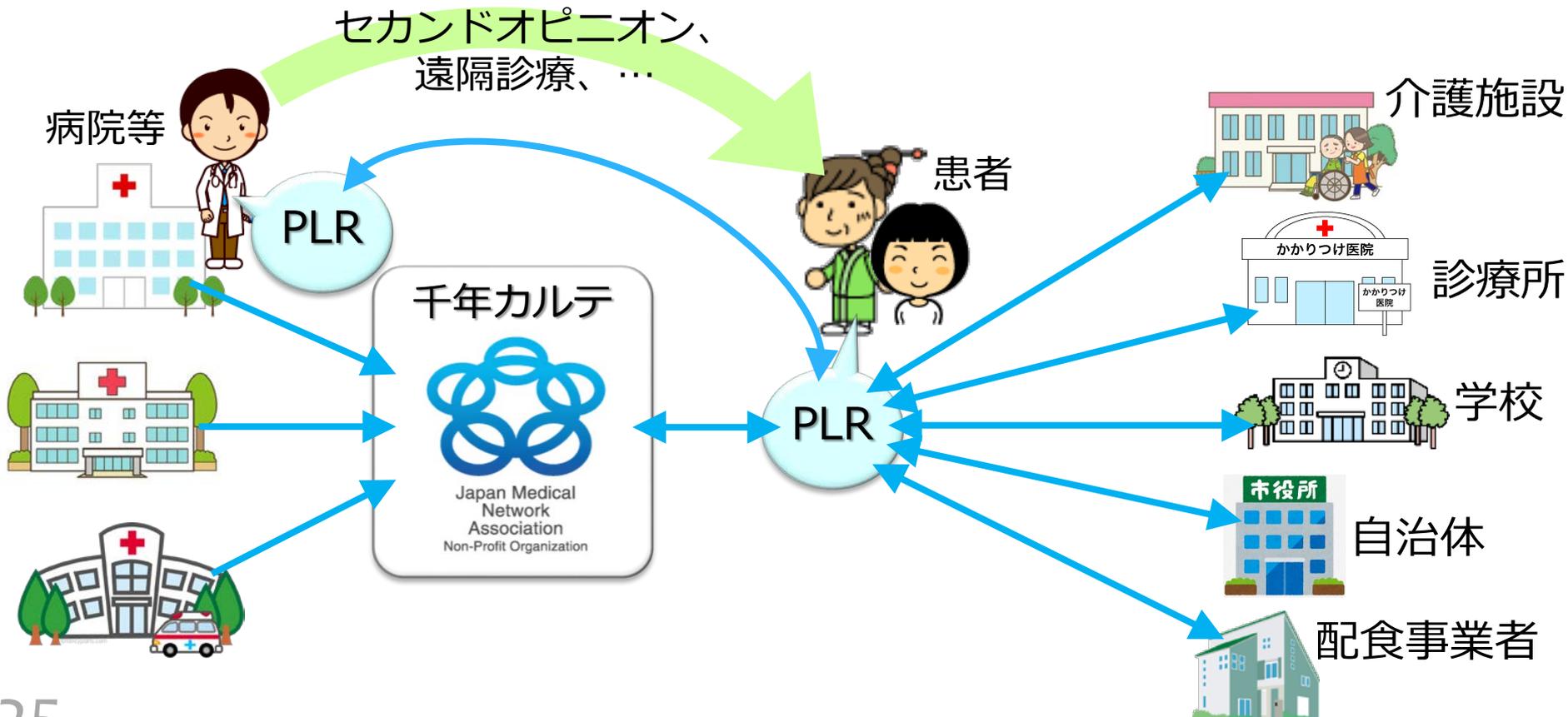


医療機関



# 千年カルテとの連携

- 次世代医療基盤法に基づく医療情報匿名加工・提供機関が病院等から医療データを収集
- そのうちデータポータビリティに対応する千年カルテと連携することにより、遠隔診療や多職種間連携を実現

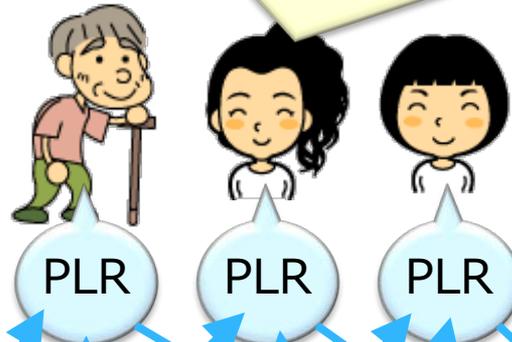


# がんの全ゲノム解析

- 保険診療でのゲノム解析は手遅れになりがちで部分的
- なるべく早期の全ゲノム解析が望ましい

- 確定診断後すぐに自由診療で全ゲノム解析
- 個別化医療の可能性

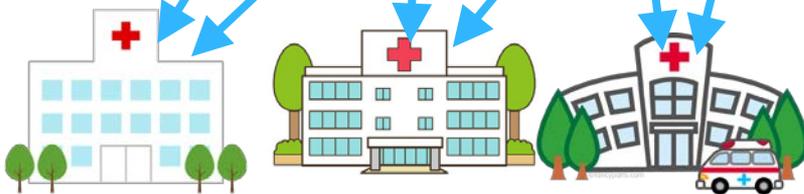
がん患者



がん保険の特約?

- 全ゲノム解析無料
- 製薬会社への売上を還元

病院



製薬会社

支援者  
(SMO/CRO)



# 生徒本人主導のeポートフォリオ

- 2020年度からの新制度の大学入試で運用されるeポートフォリオ(電子学習記録)をPLRで実装
  - ◆ データポータビリティとセキュリティを確保
- 埼玉県教育局の主導で2019年2～3月に実証実験、2019年度から実運用
  - ◆ 埼玉の県立高校の校務システムは全国の高校5,300校のうち2,000校が利用
- 教材、塾、進学先、就職先等とのマッチング

