

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 /
スマートシティ実証研究

スマートシティにおけるパーソナルデータ と産業データのデータ取引市場による 共有基盤の実証

2020年3月18日

エブリセンスジャパン株式会社

ソフトバンク株式会社

日鉄ソリューションズ株式会社

本実証研究の目的

スマートシティにおけるデータ提供者・利用者のビジネスモデルの実証によりデータ流通の課題抽出と解決方法を提示

データ取引市場が都市OSのIFに対応し「データ取引」という観点でデータ活用における出口戦略および官民・分野間のデータ連携を支援する

ネイティブデータ・オープンデータ・3rd Party データを活用し物流顧客へサービスを提供する

研究開発テーマ2

都市OS (FIWARE)

EverySense

民間生成データの自治体活用

NGSI

都市間のデータ取引

民間の利活用データを自治体へ展開

研究開発テーマ3

【5】利活用

1. 適用業務等の設定
2. 成果目標の設定
3. 利用～評価

全国情報産業連産連携協議会



サービス提供

購入企業へ
NS Solutions

【4】解析/予測

1. 目的変数設定
2. 説明変数設定
3. データ編集～予測

最終製品

調達企業へ
材料としてのデータ

価値化
データ販売

【3】データ取引

1. 受付～受入れ判断
2. データアップロード
3. 売買マッチング

納品/販売

製品としてのデータ

【2】データ加工

1. クレンジング
2. 統計処理
3. マッピング

データの製品化

原料としてのデータ
(生データ含む)

【1】データ収集

1. 収集可能データサーチ
2. 取得方法の検討
3. 取得・移転

原料調達

まちのデータ

研究開発テーマ1

複数のデータを自己調達し、一次処理して「中間製品」を製造し販売する

SoftBank

実験指導
日欧連携プロジェクトFESTIVALなどスマートシティ関連の実証研究の知見によるアドバイス

東京都における都市から収集される異分野データを用いたデータ活用ワークショップ

研究開発テーマ4

都市課題解決のためのアイデアを創出

スマートシティにおけるパーソナルデータと産業データのデータ取引市場による共有基盤の実証 コンソ各社の研究開発テーマと具体的実施概要

本研究開発テーマによる成果を得るため、スマートシティにおけるデータ提供者・利用者のビジネスモデルの実証によりデータ流通の課題抽出と解決方法を提示。コンソ各社の研究開発テーマと具体的実施概要を示す。

研究開発テーマ1：ソフトバンク株式会社

多様な産業データをデータ取引市場にて販売可能とする高付加価値化処理の研究開発

具体的な実施事項

複数のデータを自己調達し、一次処理して「中間製品」を製造し販売することにより、データ流通の促進に向けた問題・課題を洗い出す

研究開発テーマ2：エブリセンスジャパン株式会社

データ取引市場とスマートシティの国際標準インターフェイスによる相互接続手法の研究開発

具体的な実施事項

データ取引市場が都市OSのIFに対応し「データ取引」という観点でデータ活用における出口戦略および官民・分野間のデータ連携を支援する

研究開発テーマ3：日鉄ソリューションズ株式会社

データ取引市場を利用したデータの購入・分析加工・再販の研究開発

具体的な実施事項

顧客ネイティブデータ、オープンデータ、3rd Partyデータを活用し顧客（物流事業者）へサービス提供する。さらに都市課題解決に向けたデータセットを開発しデータ取引市場を経由して再販を試みる

研究開発テーマ4：全共同提案者 + 東京都

都市開発地域における総合的な流出入データ変動の活用研究

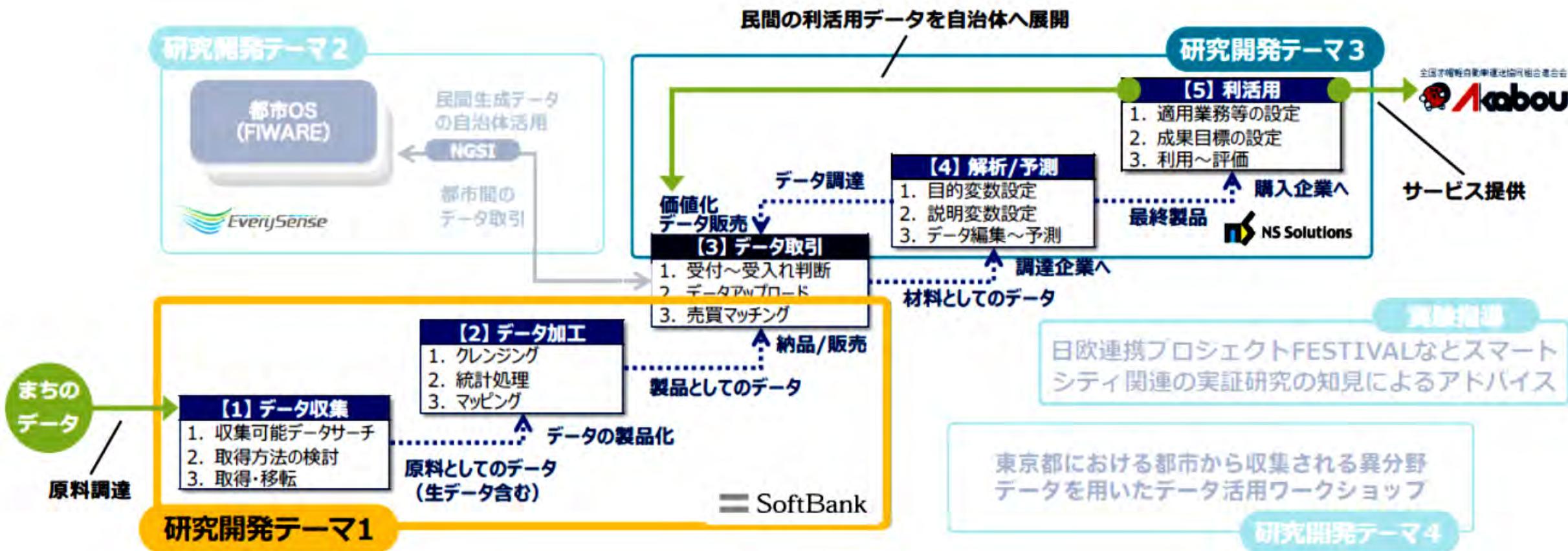
具体的な実施事項

都市課題解決のためのアイデアを創出ワークショップを実施
日欧連携プロジェクトFESTIVALなどスマートシティ関連の実証研究の知見によるアドバイス

ソフトバンク株式会社

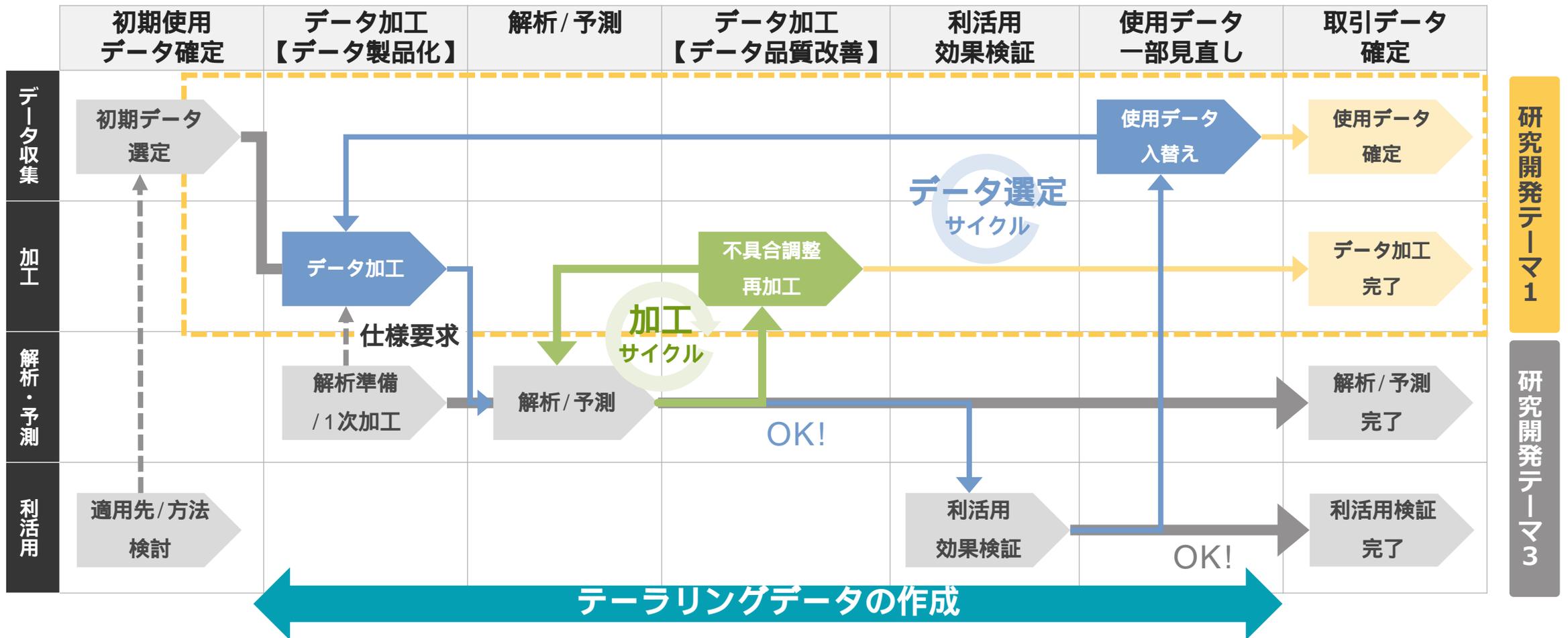
リアルタイム型データと蓄積型データを用いて、これらを再利用可能な形態に組み合わせ付加価値を与えデータ取引市場に提供する手法の研究開発

▼データ収集・加工（データの製品化）からデータ取引市場に販売/納品するまでがテーマ1のタスクである。

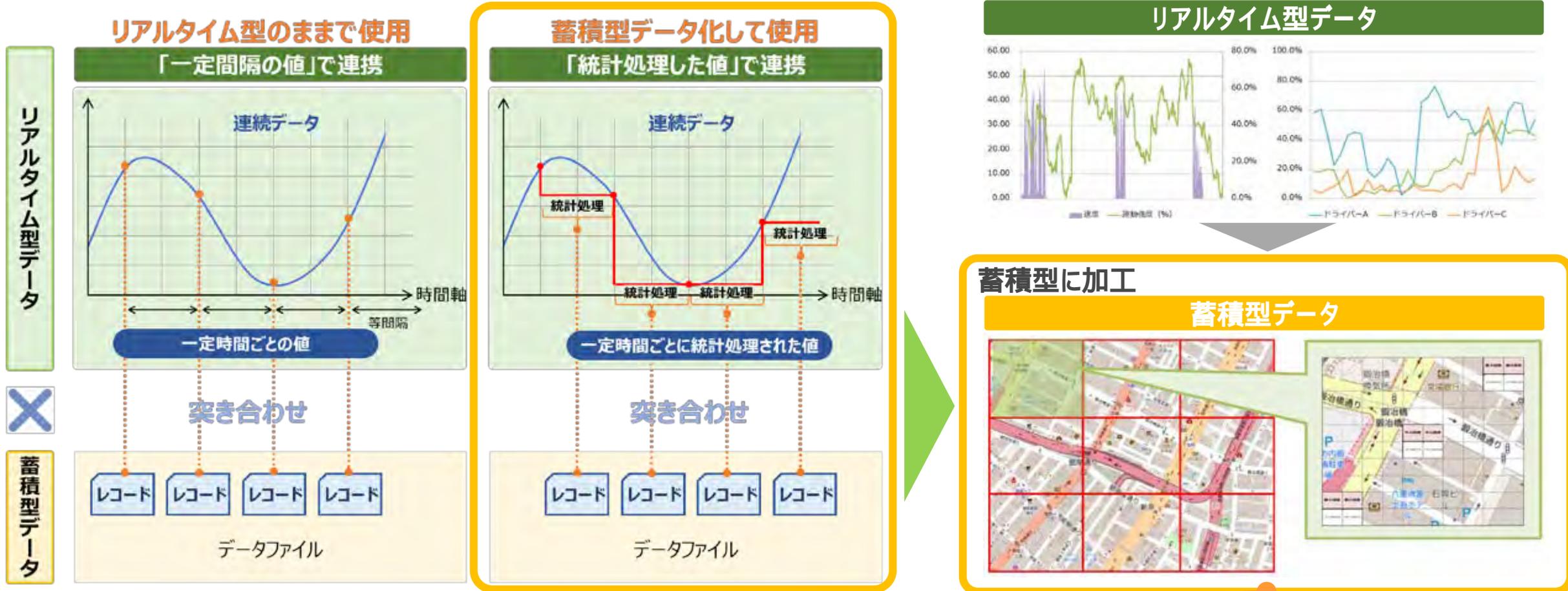


研究開発テーマ1の実施計画・推進方法

本実証研究では、データ利用者にとって利便性の高い製品としてのデータ生成を目指すために、研究開発テーマ3より1stPartyデータを譲り受けて、以下の図の通り研究開発テーマ3と研究開発テーマ1が連携し最適なデータ加工を確認しながらテラリングデータを作成した。



研究開発テーマ 1 の実施内容 ・ データ加工の概念図

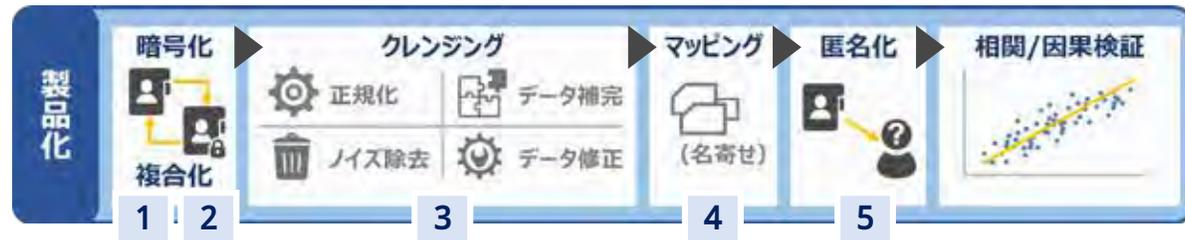


メッシュID	中心緯度	中心経度	時間帯	換算人口 (平均)
533934794203	35.623123	139.735306	10時台	90.8人

平均作業時間	運動強度レベル	イライラ件数	事故リスク件数	振動件数
23.0時間	5.0	4件	6件	1件

研究開発テーマ1の実施内容 ・ データ加工手法プロセス整理

データ加工のプロセス及び方法についての調査・整理を行い、本研究開発において不可欠なもの及び最適と思われる方法を選択の上加工処理を行った。



個人特定可能な情報はID削除の上特定不能とした状態のデータを利用したため、匿名化は不要。データの受け渡しは相対で直接行ったため、暗号化/複合化も実施していない。

4 データマッピング

同じファイル内のレコードの名寄せ処理

ファイル名	実証参加登録車種	時間	緯度	経度	住所	...
車両位置情報ファイル						
車両位置データレコード01	B-007	11時30分	35.683766	139.753609	...	
車両位置データレコード28	A-002	11時30分	36.287652	137.394532	...	
車両位置データレコード42	B-007	12時00分	35.683766	139.753609	...	

⇒ 11時30分～12時00分の間、この車両の位置情報に変化ないので駐車中である。
⇒ 異なるファイルにあるレコードの名寄せ処理

ファイル名	実証参加登録車種	車両種別	燃料全量名	...
実証参加登録車両位置情報ファイル				
登録車両データレコード01	A-003	高所作業車	〇〇産業	...
登録車両データレコード33	B-007	貨物トラック	△△運輸	...

⇒ 11時30分～12時00分の間駐車していた実証登録車両B-007は、貨物トラックである。

5 匿名化

個人情報保護、企業機密情報保護のためにケースごとに最適な手法により行う

マスキング

- ▶ ダミー値匿名化: 山田 太郎 → 固定値に置換 → ダミー シロー
- ▶ 連結匿名化: 01234567 → 連結表で置換 → 38245864
- ▶ 連携不可能匿名化: 01234567 → ハッシュ値で置換 → 195bd578qdg470k kne368k890egged

非特定化

- ▶ 年齢の場合: 57歳 → 区分化 → 50～59歳
- ▶ 住所の場合: 東京都港区新橋1-2-3 → 切り捨て → 東京都港区
- ▶ 病名の場合: 心房中隔欠損症 → 一般化 → 心臓病

・氏名 ⇒ ヤマダ タロウ
・住所 ⇒ 東京都千代田区1丁目1-1

1 暗号化・・・データ受領時

データ受領時にデータを適切に保管する

暗号化

・氏名 ⇒ U2FsdGVkX18HGpxTLq8Z33noKq...
・住所 ⇒ FsdGVlrdV8MpdhWEnNSHkd7/U2oqRU7i5...

2 複合化・・・加工処理前

受領したデータを加工可能に戻す

複合化

・氏名 ⇒ ヤマダ タロウ
・住所 ⇒ 東京都千代田区1丁目1-1

3 データクレンジング

同一データ、記載方法複数パターンでの共通化対策

正規化
例) 住所処理
・千代田区千代田1-1-1
・東京都千代田区千代田1丁目1番1号
↓
・東京都千代田区1丁目1-1

データ補完
例) 住所を緯度経度
・東京都千代田区1丁目1-1
↓
・緯度: 35.683766
・経度: 139.753609

情報欠損 (不足) データへの補完処理

エラー/破損などのゴミデータ、不要記号などの削除

ノイズ除去
例) カラム処理 (氏名)
・(性) ヤマダ (名) タロウ
・山田 (ヤマダ) 太郎 (タロウ)
↓
・ヤマダ タロウ

正式名称への書き換えなど誤記修正

データ修正
例) 西暦和暦変換
・令和元年7月17日
↓
・2019年7月17日

研究開発テーマ1の実施内容 ・ 利用者の分析要求に着眼した加工

データ利用者（研究開発テーマ3）が様々な目的に応じて分析・解析・予測・視覚化などを行うことを想定し、利便性の高い製品としてのデータへの加工を目指し、以下の処理を行いテラーリング加工をした。

1. データ活用の利便性向上のための正規化処理

▶連続値のレベル分けとフラグ化

作業時の「負荷」や「きつさ」を表す『運動強度』に対し、5段階のレベル(フラグ)情報を付与

運動強度の値（範囲）	運動強度フラグ
60.0%以上	5.0
60.0%未満～40.0%以上	4.0
40.0%未満～30.0%以上	3.0
30.0%未満～20.0%以上	2.0
20.0%未満	1.0

▶素材データ間で異なる表記の統一

異なる表記ルールで記載された「年月日」情報を統一した記載に変換

処理前	処理後
2019-12-14T08:34:50	2019/12/14
2019年12月14日	

2. 情報価値を高めるための補完処理

▶座標情報（緯度経度）に関連情報を付与

位置情報に対し、エリア特有の属性情報を付与し、データの分析価値を向上

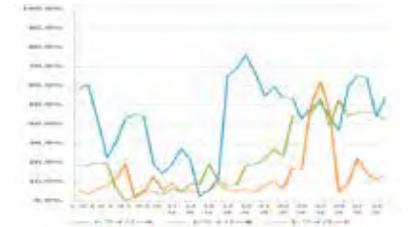


3. 傾向比較を容易にするための演算処理

▶分析指標の生成（様々な演算処理を用いた分析指標の生成）

個人差のある「心拍数」のデータから、分析に活用可能な指標（運動強度）を生成

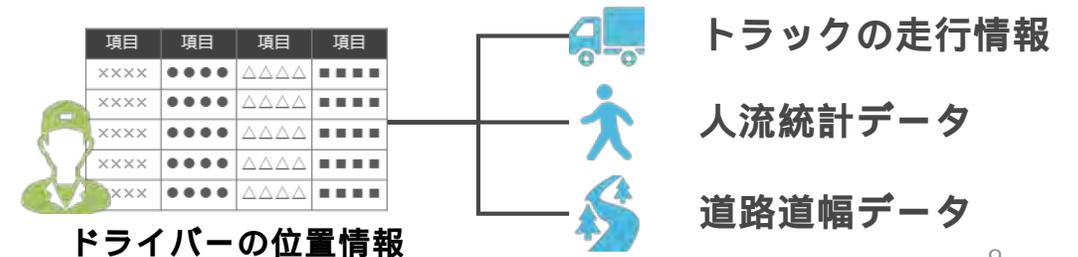
ユーザ	年齢	心拍数	分析指標（運動強度）
ユーザA	60代	96	54%
ユーザB	40代	78	18%
ユーザC	30代	75	3%



4. 異種データの関係付けのためのマッピング

▶異種レコード間の関連付け

様々なセンサーで収集した計測データをマッピング（関連付け）



研究開発テーマ 1 の実施内容

・ 利便性に着眼した生成データ

加工した製品としてのデータの受け渡しにおいては、前頁の処理結果を踏まえたデータ仕様書を作成し分析等テスト用のデータサンプルを生成した。

生成データ仕様書

流通データA (ファイル名) distribution_data_A.csv

項目名	単位	形式	説明
meshID		10桁整数	メッシュID
H_flag		0/1	加工済フラグ
hour		0-23	時間
WorkID (cum)		整数	作業ID (累積)
workingtim e (cum)		実数	作業時間 (累積)
workingtim e (ave)		実数	作業時間 (平均)
workingtim e (min)		実数	作業時間 (最小)
workingtim e (max)		実数	作業時間 (最大)
E_flag (mesh_ave)		0/1	エラーフラグ (メッシュ平均)
E_flag (mesh_min)		0/1	エラーフラグ (メッシュ最小)
E (mesh_max)		実数	エラー (メッシュ最大)

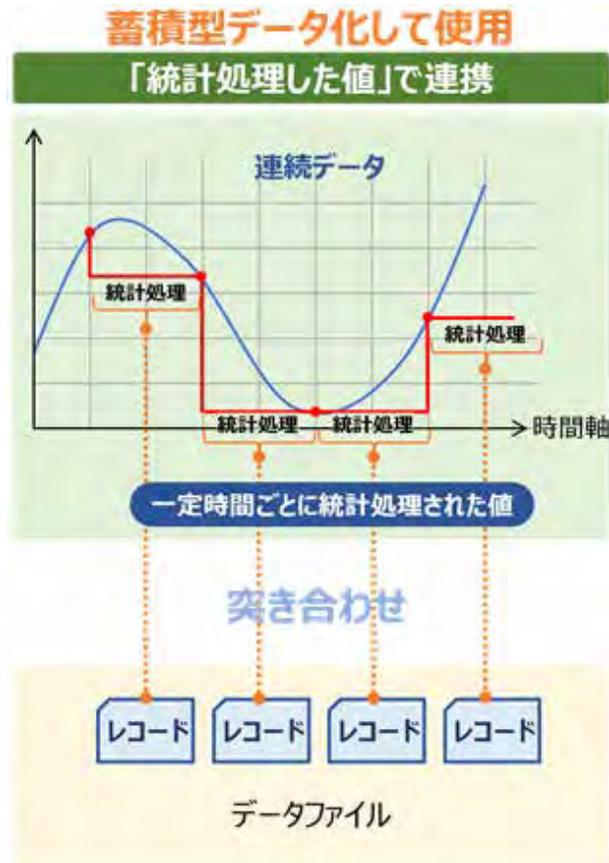
生成データサンプル

流通データA

meshID	H_flag	hour	WorkID (cum)	workingtim e (cum)	workingtim e (ave)	workingtim e (min)	workingtim e (max)	E_flag (mesh_ave)	E_flag (mesh_min)	E (mesh_max)
533936903143	0	08	1	0.0	0.0	0.0	0.0	3.0	3.0	3.0
533936903324	0	08	5	7.0	7.0	1.0	20.0	4.0	3.0	3.0
533946001011	0	08	7	13.0	6.5	3.0	17.0	5.0	5.0	5.0
533935584212	0	09	8	11.0	11.0	1.0	10.0	3.0	4.0	3.0
533935584221	0	09	5	0.0	0.0	0.0	0.0	2.0	4.0	3.0
533936922321	0	09	1	0.0	0.0	0.0	0.0	1.0	5.0	3.0
533945093104	0	10	1	0.0	0.0	0.0	0.0	5.0	3.0	3.0
533945093202	0	10	5							
533945093211	0	10	6	13.0	13.0	1.0	15.0	4.0	1.0	3.0
533936802334	0	11	12	14.0	7.0	1.0	16.0	4.0	3.0	3.0
533936922223	0	12	1							
533936922321	0	12	12	0.0	0.0	0.0	0.0	3.0	5.0	3.0
533936922412	0	12	6							
533936924004	0	12	6							
533936924013	0	12	1	9.0	9.0	1.0	20.0	2.0	3.0	3.0
533936924102	0	12	2							
533936924111	0	12	2							
533936924113	0	12	1							
533936924113	0	12	1							
533936924113	0	14	1	1.0	1.0	1.0	20.0	3.0	2.0	3.0
533936903433	0	14	1	0.0	0.0	0.0	0.0	3.0	1.0	3.0
533934794203	1	09	8	23.0	11.5	1.0	20.0	5.0	5.0	5.0
5339334823003	1	10	2	30.0	15.0	1.0	20.0	4.0	4.0	4.0
533943023113	1	11	2	44.0	22.0	1.0	20.0	4.0	4.0	4.0
533936922321	1	19	1	0.0	0.0	0.0	0.0	5.0	5.0	5.0
533945093211	1	19	1	0.0	0.0	0.0	0.0	3.0	3.0	3.0
533945093211	1	20	1	18.0	18.0	3.0	20.0	1.0	1.0	3.0
533935584212	1	21	1	0.0	0.0	0.0	0.0	2.0	1.0	3.0
533935584213	1	21	1	8.0	8.0	1.0	20.0	1.0	1.0	3.0

研究開発テーマ1の実証結果・実施したデータ加工の俯瞰図

膨大な連続性のあるリアルタイム型データを蓄積型データに変換し、研究開発テーマ3の利便性を高めるための様々な加工プロセスを経ることにより、「実施内容・市場化に向けたデータ加工」の一例としてのテラーリングデータを作成した。



データ加工

リアルタイム型データ

下記のような連続性を持つリアルタイム型データを蓄積型データへ変換

車両の加速度	車両のGPS速度
ドライバーのGPS位置	ドライバーの心拍数
ドライバーの歩数	荷室の衝撃データ

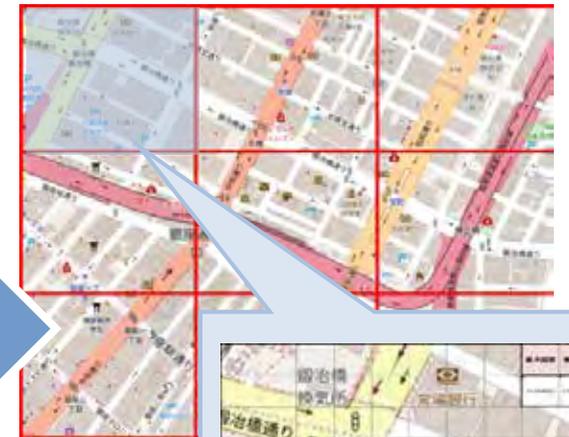


蓄積型データ

上記、蓄積型データとの紐づけを行うためのマッピング処理を実施

地図データ	人流統計 (50mメッシュ)
-------	----------------

利便性の高いデータの生成



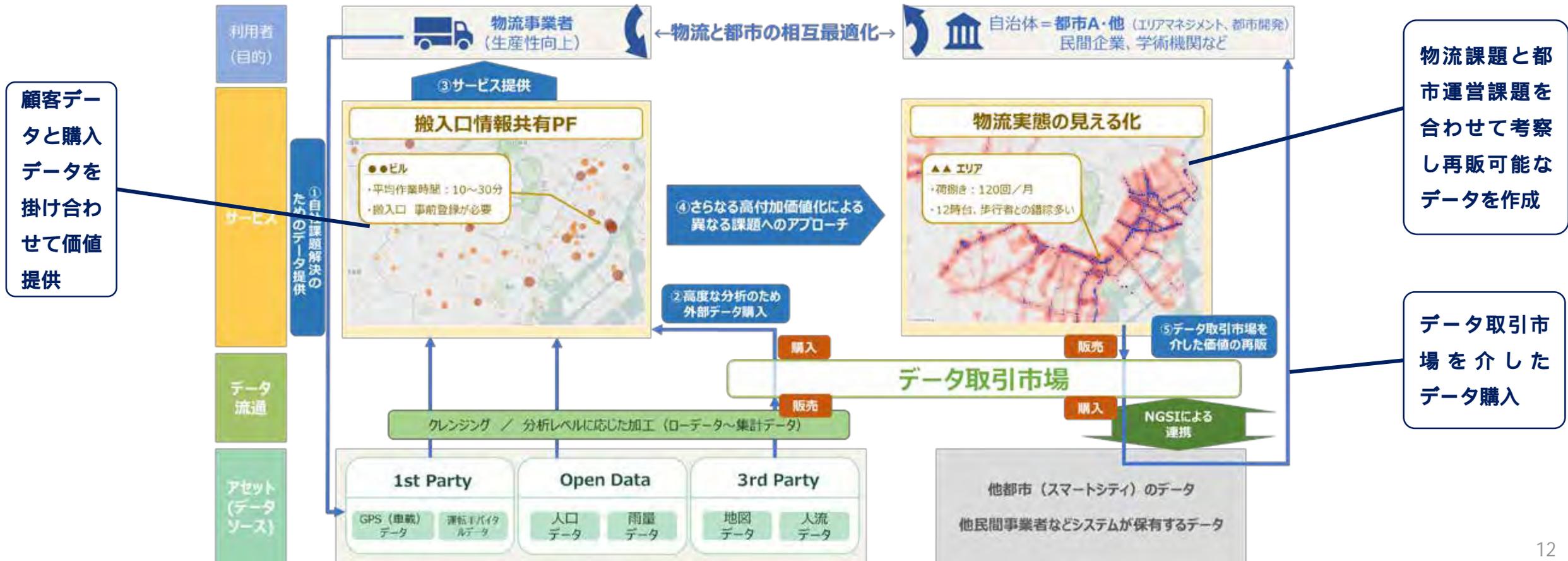
研究開発テーマ 1 の考察の結果

異分野データの掛け合わせによる活用（分析/予測/視覚化等から社会課題解決）を推進させるため、匿名化や異常値除去等の最低限の処理の他、以下のデータ加工処理の実施が有効と考える。

流通製品としての価値を高める加工処理	加工による効果と本研究開発で行った処理
リアルタイム型データを蓄積型データへ変換	<p>▶膨大な連続データの分析等には多大な負荷が掛かる、予め蓄積型データに変換しておくことで利用者の前処理の作業を軽減することができる。</p> <p>本研究開発では、車両の加速度・位置、ドライバーの位置・心拍数・歩数、荷室の衝撃度を蓄積型データへ返還した。</p>
分析作業の効率化のための段階分けとフラグ付与	<p>▶強弱や大小などの程度を測定したデータについて、最小値と最大値間の段階分けを行い、各段階を表すフラグを付与することで分析等の作業の効率化に寄与する。</p> <p>本研究開発では、ドライバーの運動強度を5段階に分けフラグ化した。</p>
情報価値を高めるための補完処理	<p>▶異種データ間の共通キー項目に、使用目的分野において参照価値のある情報を付与することで、分析等の想定を超えたアウトプット創出の可能性に寄与する。</p> <p>本研究開発では、共通キーである座標（緯度経度）に道路幅員情報を付与した。</p>
傾向比較を容易にするための演算処理	<p>▶物体や個体に定常時等の数値格差がある場合、単純比較が困難。演算処理により分析指標を生成することにより、物体・個体の差を意識することなく比較可能となる。</p> <p>本研究開発では、個人差のある「心拍数」のデータから、ドライバーの作業負荷などの分析に活用可能な指標（運動強度）を生成した。</p>

データ取引市場を利用したデータの購入・分析加工・再販の研究開発

研究開発テーマ3では、物流事業者（事業組合）である顧客の1stPartyデータとしての車載データ、ドライバのバイタルデータにデータ取引市場を介して購入する3rdPartyデータを用いて価値化しサービス提供する。また、さらに異なる都市課題の解決のために分析・解析を加えてできたデータ、自治体などにデータ取引市場を介して再販する。



研究開発テーマ3の総括

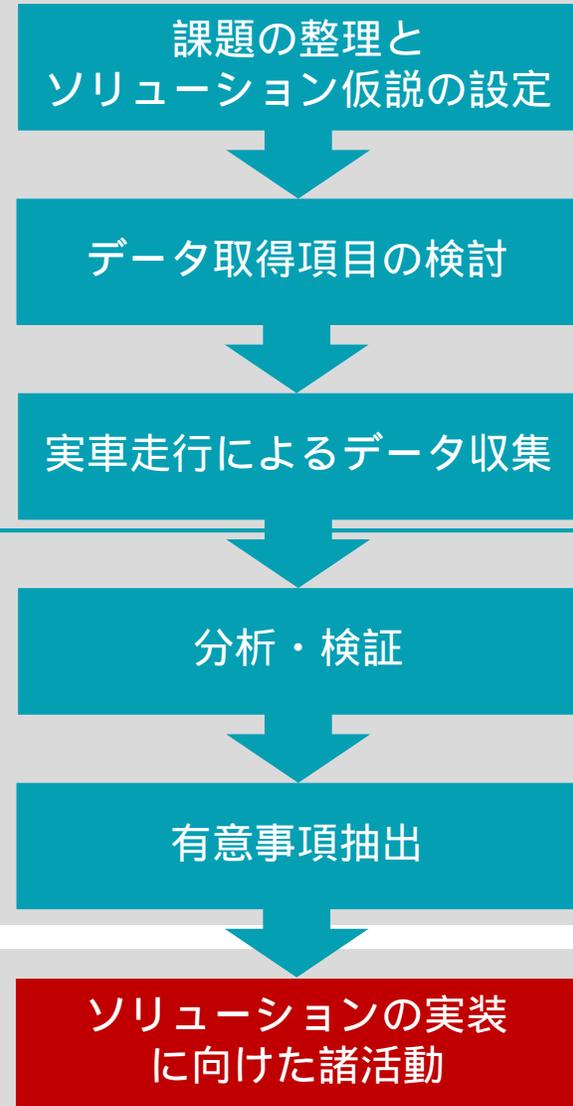
顧客ネイティブデータ、オープンデータ、3rd Partyデータを活用し顧客（物流事業者）へサービス提供する。さらに、都市課題解決に向けたデータセットを開発しデータ取引市場を経由して再販を試みる

研究開発テーマ3 今年度実施範囲

- 物流課題について解決につながる試案をソリューション仮説として設定した
- その有効性を検証するために取得するデータと分析方法について整理した
- 実走車両よりデータを取得し、オープンデータ、及び、データ取引市場を介して取得した外部の産業データを合わせて分析・検証を行いソリューション仮説を証明する有意事項を抽出した。その際に、研究開発テーマ1に協力してテラーメイドデータを作成し、分析・検証の精度向上・迅速化を果たした
- 都市課題に訴求し、自治体などに向けた再販を行うべくデータ取引市場を介して販売する準備を行った

- ストレス、急減速と事故多発地点などに一定の関連性が確認できた
- 荷捌き場所・時間を視覚化し、新しい視点からドライバーが抱えている問題を浮き彫りにできた
- 経済的な路面データ取得について遡及できる成果を得た
- 物流と人流の関係など都市課題の解決につなげられる示唆を得た

- 今年度の結果と各協力機関の意見・ニーズなどに基づきより大規模な実証を行う。
AIによるデータ分析を行うことで精度を高めて、提言するソリューションの質を向上させる



今後の取り組み

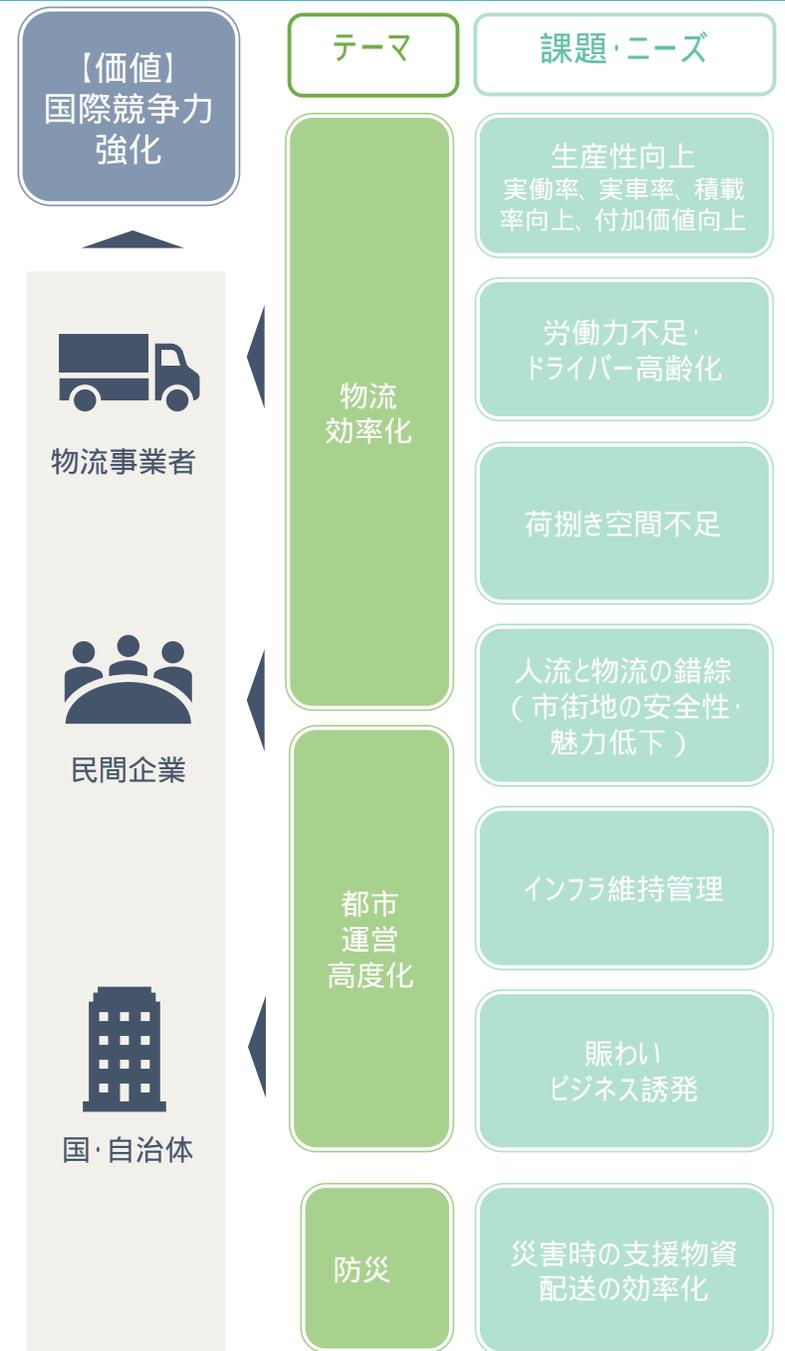
課題の整理とソリューション仮説の設定 (1/2) | 課題の整理

本研究開発テーマでは一義的に**物流事業の効率化**をテーマに課題解決に臨む。そのため、協力機関（赤帽首都圏軽自動車運送協同組合）へのインタビュー、官の公開資料に基づき課題を整理した。

また、協力機関から取得する実走データは道路インフラ維持管理、人流・物流の錯綜における安全性に関わることなど街の課題について言及可能であると考え、都市課題の解決に訴求することを目指した。最終的には実走データから「**物流と都市課題の相互解決**」が行えるソリューション化を目標に検討をすすめる。

さらに、災害時の救援活動、支援物資運搬に関わる事項について示唆の提示を試みた。

これらの検討結果から今回はソリューション仮説（1）～（3）を導いた。



課題の整理とソリューション仮説の設定 (2/2)

物流事業者向けに提供するサービス仮説 (ソリューション仮説) として

- ・ 物流事業者向けの情報共有
- ・ 物流運行計画 + 搬入支援

また、都市課題解決に向けて基礎情報収集と物資輸送支援を検討に加えた



物流効率化のために具備するシステムが、災害時にも役立つという整理