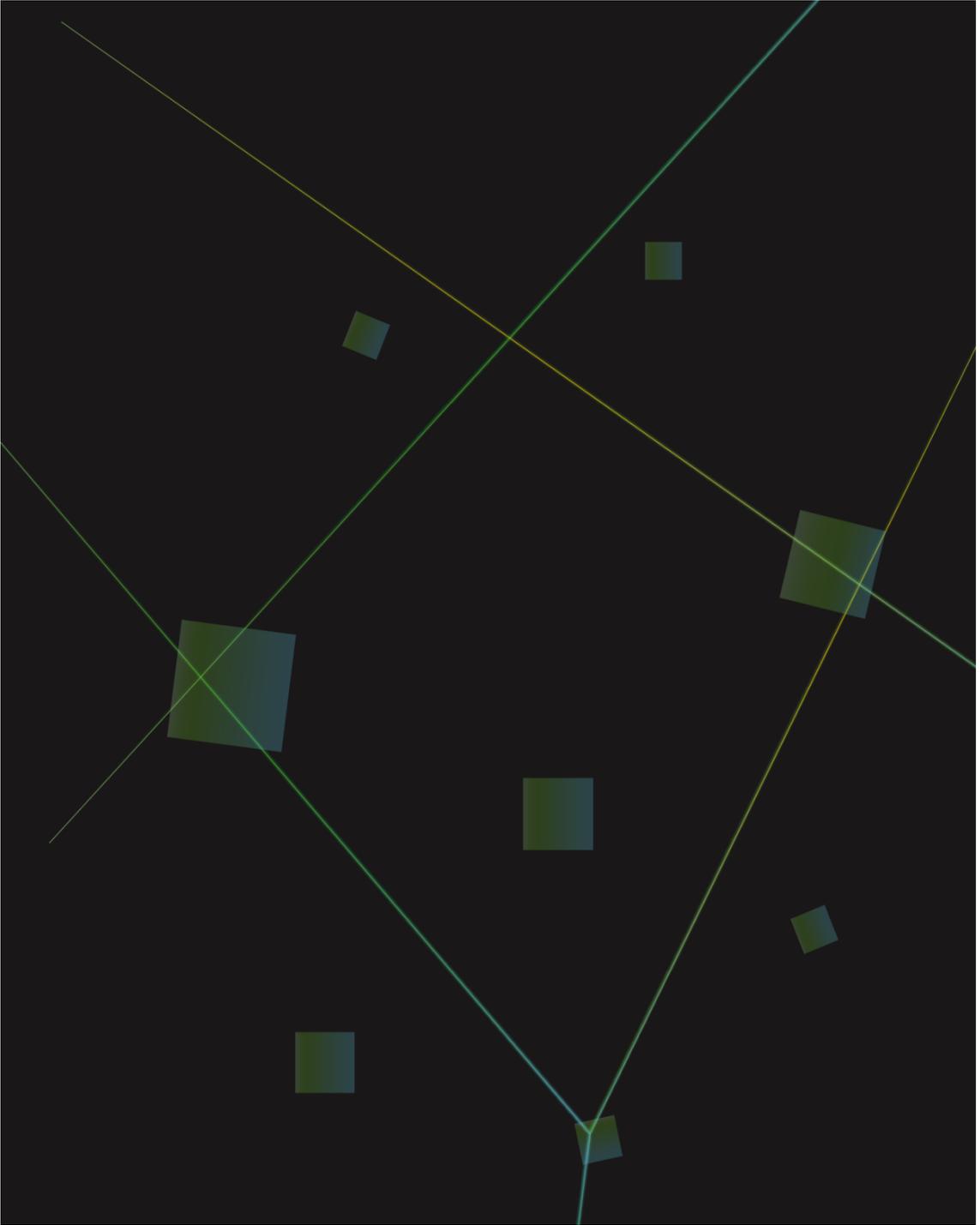




Project LINKS Technical Report  
国土交通分野のデータ整備・活用・オープンデータ化のための技術資料



# データ構築基盤 LINKS Veda 開発 技術検証レポート

Technical Report on the Development of the Data Infrastructure  
Platform "LINKS Veda"

series  
No. **01**

# 目次

---

---

1. 本業務の概要 .....	- 4 -
1-1. 調査の背景・目的 .....	- 4 -
1-2. 課題認識 .....	- 5 -
1-3. 課題解決のアプローチ .....	- 6 -
1-4. 創出価値 .....	- 7 -
1-5. 想定事業機会 .....	- 8 -
2. 実証実験の概要 .....	- 9 -
2-1. 実証仮説 .....	- 9 -
2-2. 実証フロー .....	- 10 -
2-3. 検証ポイント .....	- 11 -
2-4. 実施体制 .....	- 12 -
2-5. スケジュール .....	- 13 -
3. 開発スコープ .....	- 14 -
3-1. 概要 .....	- 14 -
4. 実証システム .....	- 16 -
4-1. アーキテクチャ .....	- 16 -
4-1-1. システムアーキテクチャ .....	- 16 -
4-1-2. データアーキテクチャ .....	- 17 -
4-1-3. システム構成 .....	- 18 -
4-1-4. ハードウェアアーキテクチャ .....	- 25 -
4-2. システム機能 .....	- 27 -
4-2-1. システム機能一覧 .....	- 27 -
4-2-2. 利用したソフトウェア・ライブラリ .....	- 39 -
4-2-3. 開発機能の詳細要件 .....	- 41 -
4-3. アルゴリズム .....	- 104 -
4-3-1. 利用したアルゴリズム .....	- 105 -
4-3-2. 開発したアルゴリズム .....	- 109 -
4-4. データインタフェース .....	- 112 -
4-4-1. ファイル入力インタフェース .....	- 112 -
4-4-2. ファイル出力インタフェース .....	- 114 -
4-4-3. 内部連携インタフェース .....	- 117 -
4-4-4. 外部連携インタフェース .....	- 168 -
4-5. 実証に用いたデータ .....	- 171 -
4-5-1. 活用したデータ一覧 .....	- 171 -
4-5-2. 生成・変換したデータ一覧 .....	- 179 -
4-6. ユーザーインタフェース .....	- 188 -

4-6-1. 画面一覧.....	188 -
4-6-2. 画面遷移図.....	192 -
4-6-3. 各画面仕様詳細.....	194 -
4-7. 実証システムの利用手順.....	222 -
4-7-1. 実証システムの利用フロー.....	222 -
4-7-2. 各画面操作方法.....	223 -
5. システムの非機能要件.....	224 -
5-1. 社会実装に向けた非機能要件.....	224 -
6. 品質.....	227 -
6-1. 機能要件の品質担保.....	227 -
6-2. 非機能要件の品質担保.....	229 -
7. 実証技術の機能要件の検証.....	230 -
7-1. 検証目的.....	230 -
7-2. 検証ポイントと KPI.....	231 -
7-3. 検証方法.....	232 -
7-3-1. 実証技術の検証：データ処理の正確性.....	232 -
7-3-2. 実証技術の検証：構造データ生成技術.....	234 -
7-3-3. 実証技術の検証：チャット生成技術.....	237 -
7-4. 検証結果.....	240 -
8. 実証技術の非機能要件の検証.....	246 -
8-1. 検証目的.....	246 -
8-2. 検証ポイントと KPI.....	247 -
8-3. 検証結果.....	248 -
9. 原課向け検証会のチャット機能有用性検証.....	249 -
9-1. 検証目的.....	249 -
9-2. 被験者.....	249 -
9-3. 検証方法.....	250 -
9-4. ヒアリング・アンケートの詳細.....	252 -
9-4-1. アジェンダの詳細.....	252 -
9-5. 検証結果.....	252 -
10. 原課向け検証会の Veda 本体機能有用性検証.....	256 -
10-1. 検証目的.....	256 -
10-2. 検証ポイントと KPI.....	257 -
10-3. 被験者.....	258 -
10-4. 検証方法.....	259 -
10-5. ヒアリング・アンケートの詳細.....	261 -
10-5-1. 検証項目と評価方法.....	261 -
10-6. 検証結果.....	262 -

11. 有識者とのディスカッション.....	- 265 -
11-1. ディスカッションの目的.....	- 265 -
11-2. ディスカッションから得られた結果.....	- 265 -
11-3. 成果と展望.....	- 268 -
12. 成果と課題.....	- 270 -
12-1. 本実証で得られた成果.....	- 270 -
12-1-1. AI を活用した自動データ生成における技術面での優位性.....	- 270 -
12-1-2. AI を活用した自動データ生成のビジネス面での優位性.....	- 272 -
12-1-3. AI を活用した自動データ生成の公共政策面での優位性.....	- 273 -
12-2. 実証実験で得られた課題と対応策.....	- 274 -
12-3. 今後の展望.....	- 276 -
EBPM アプリケーションにおけるパフォーマンス最適化.....	- 276 -
13. 用語集.....	- 278 -

## 1. 本業務の概要

### 1-1. 調査の背景・目的

国土交通省では、急速に進展する人口減少・少子高齢化に対応し、インフラ、防災、交通、まちづくり等の多様な分野における官民の生産性を向上させるため、政策やビジネスにおけるデータ活用や新サービス創出等を進める必要がある。幅広い施策・制度・手続を所管する国土交通省には膨大な行政情報が蓄積されているが、機械利用できる形で整備されておらず、二次利用を含め十分活用されずに宝の持ち腐れ状態となっている。これらを「データ」として整備し、官民が利用可能な基礎的な情報として提供するとともに、行政内での活用環境を整備することで、オープンデータを利用したビジネス創出や政策立案におけるデータ活用(EBPM)をし、社会全体の生産性向上の実現を図る取組(Project LINKS)をスタートする。本取組により、データが活用され、様々なサービスが生まれ出されていくオープン・イノベーションを創出し、更なる経済成長や生産性の向上、社会全体のデジタル化等の社会変革を実現していく。

## 1-2. 課題認識

国土交通省では、行政手続等から取得された膨大な行政情報が蓄積されているものの、これらはデータとして整備されていないため、全庁的な政策立案等への活用や、オープン・イノベーションに係る取組の本格的な展開には至っていない。本取組を通じてこれらの行政情報を「データ」として整備し、官民が利用可能な基礎的なデータとして提供するとともに、行政内での活用環境を整備することで、オープンデータを利用したビジネス創出や政策立案におけるデータ活用（EBPM）を促進し、社会全体の生産性向上の実現を図る。

表 1-1 本プロジェクトが取組むべき社会課題と解決手法

項目	内容
目的	Society5.0 の実現による更なる経済成長や生産性の向上、社会全体のデジタル化に向けた様々な施策の推進に向けて、国土交通省が保有する行政情報のデータ化、オープン化を推進すること。
現状分析	これまでの DX 施策はインフラ管理等の国土交通政策や所管業界に対するアプローチが中心であり、国土交通省に蓄積されている膨大な行政情報がデータとして整備がなされていなかったことによって、全庁的な政策立案等への活用、オープン・イノベーションに係る取組の本格的な展開には至っていない。
解決すべき社会課題	近年、デジタル化は急速に進展しており、国際社会や企業活動、そして一人ひとりのライフスタイルに至るまで、そのありようを変化させており、人口減少による地域の足の衰退や担い手不足、気候変動に伴う災害の激甚化・頻発化、脱炭素化等が大きな課題となっている。具体的な社会経済の課題としては、国民生活・経済活動の制約、デジタル競争力の低迷、労働力の減少・国内市場の縮小、災害被害の拡大・人命喪失、温暖化の進行が挙げられる。
想定する解決手法	<ul style="list-style-type: none"> <li>● 国土交通省が保有する行政情報の分野横断的なデータ化及びオープンデータ化。</li> <li>● 上記の課題から特に社会的ニーズが高く、即時性が期待できるテーマをピックアップした官民の多様な分野における活用事例（ユースケース）の開発。</li> </ul>

## 1-3. 課題解決のアプローチ

本業務では、国土交通分野の DX を推進し、分野横断的なデータ化の開発を進める新たな取組(Project LINKS)を展開していく。

本取組に必要なデータの作成、加工、ビジュアライズは「データ管理システムのプロトタイプ開発」にて可能な限り OSS を用いて実装することで、将来的なシステム自体の利用促進を見据えた開発を行う。国土交通省の分野横断的なデータ化の開発を並行して実施することで、更なる経済成長や生産性の向上、社会全体のデジタル化等の社会変革実現を図る。

### データ管理システムのプロトタイプ開発

- 国土交通省保有情報を機械的にデータ化、活用環境へシームレスに接続するためのデータ管理システムのプロトタイプ開発
- データ管理システムおよびデータ可視化システムは指定の機能要件／非機能要件を満たす
- 必要な機能要件、非機能要件、外部設計等を調査検討し提案
- ドキュメンテーション、開発及び技術検証
- プロトタイプを用いた国土交通省の関連部局に対するヒアリング等による有用性調査

## 1-4. 創出価値

本業務では、EBPM の推進およびオープン・イノベーションの創出に向けた方針について以下の通り想定する。特に、想定事業機会の観点では、積極的なオープンデータ化を通じて、民間企業や研究機関、シビックテック等でのアプリケーション開発や新たなビジネス創出の活性化が期待できる。

表 1-2 創出価値

創出価値	方針
1. 多様化する社会課題への柔軟な対応が可能な行政データの活用による EBPM の推進	<ul style="list-style-type: none"> <li>● 多様化する社会課題の解決に向けインフラ DX・まちづくり DX などの新たな取組と並行して、これまで蓄積された国土交通省保有行政データの新たな活用可能性と価値を見出すとともに、その活用環境を提供することで、効果的な EBPM を推進のための方法と環境を提供する。</li> </ul>
2. 行政データのデジタル化・標準化・一元化による行政手続の DX 推進	<ul style="list-style-type: none"> <li>● 行政情報の整備状況を含むデータ特性や、データ相互の関係性を考慮した上で、連携・統合が望まれるデータの棚卸しを行い、国土交通省及び地方公共団体の各種手続き行政業務の効率化・サービス向上、各種調査やデータ整備の負担軽減に資する行政データの整備・活用スキームを創出する。</li> </ul>
3. 適切なデータの秘匿化とオープンデータ化によるオープン・イノベーションの創出	<ul style="list-style-type: none"> <li>● オープンデータ化の推進により、多様な主体が活用することで、データそのものの信頼性や透明性の向上が期待できる。</li> <li>● 積極的なオープンデータ化は、民間企業や研究機関、シビックテック等でのアプリケーション開発や新たなビジネス創出の活発化が期待できる。</li> </ul>
4. 誰もがアクセスしやすく持続的なオープンデータ利活用環境の提供	<ul style="list-style-type: none"> <li>● Project LINKS で整備したオープンデータを広く活用してもらうためには、誰もが自由にデータにアクセスすることが必要である。持続的なオープンデータ環境を提供するため、利用ルールや運用ルールの設定、ダウンロード状況や活用状況等をモニタリング・評価し PDCA サイクルを回す。</li> </ul>

## 1-5. 想定事業機会

表 1-3 想定事業機会

項目	内容
利用者	<ul style="list-style-type: none"><li>● 国土交通省情報政策課</li><li>● 地方整備局や出先機関など、国交省内の他部局</li><li>● 行政文書の構造化・検索・可視化を必要とする地方自治体の行政部門</li></ul>
サービス仮説	<ul style="list-style-type: none"><li>● 行政文書の構造化・検索・分析を自動化し、職員の業務負担を軽減する</li><li>● 属人性を排除し、業務の標準化・現場性のある運用を可能とする</li><li>● EBPM（エビデンスに基づく政策形成）を支援するためのデータ活用基盤を提供する</li><li>● 再利用性の高い構造設計により、他部局・他自治体への展開も可能とする</li></ul>
提供価値	<ul style="list-style-type: none"><li>● 実証で得られた成果を基に、構造定義や処理手順をテンプレート化し、他部局・自治体への展開が可能</li><li>● 文書構造の違いにも柔軟に対応できる仕組みとすることで、導入時の設定コストを最小化</li><li>● 政策文書、計画書、審査資料など、定型文書の処理分野から段階的に展開し、標準的な行政支援ツールとして活用</li></ul>

## 2. 実証実験の概要

### 2-1. 実証仮説

国土交通省保有行政情報のデータ作成実証の観点の観点から、本業務におけるプロトタイプ開発を通じて以下を検証する。

- データ作成実証の観点

国土交通省は、構造化されていない大量の行政情報を保有している。これらの情報は、従来は人の目と手により確認・整理される必要があったため、情報を開示するまでに多くの労力とコストがかかる課題があった。本プロトタイプでは、LLM（大規模言語モデル）やNLP（自然言語処理）等のAI技術を活用してこれらを自動的に構造データ化しデータとして管理が可能な状態とすることで、大幅なコストの削減を実現する。さらには、関連するデータ同士の紐づけをした上で整備することで、データそのものの利用価値向上を目指す。また、データ作成における技術検証観点としては、オープンソースソフトウェアと有償ソフトウェアを用途に応じて使い分けることで、その品質とコストを検証し、継続的なシステム運用に向けた最適な技術要素を明らかにする。

## 2-2. 実証フロー

実証実験では、国土交通省が保有する情報を機械的にデータ化し、活用環境とシームレスに連携可能なデータ管理システムのプロトタイプを開発する。開発後は、省内の関連部署へのヒアリングおよび、有用性調査を行う。

### 1) データ収集

- 国土交通省が保有する行政情報を対象に、EBPM やオープンデータの観点から活用可能性の高いデータを選定・抽出する。対象データを体系的に整理したうえで、構造化や利活用に向けたデータを収集する。

### 2) プロトタイプ開発

- 収集・整理したデータを基に、活用環境とシームレスに接続可能なデータ管理システムのプロトタイプを開発する。あわせて、システム構成や運用方法を整理した技術ドキュメントを作成する。

### 3) 有用性検証

- 開発したプロトタイプを用いて、国土交通省内の関係部局を対象とした有用性検証を実施する。有用性・運用性・拡張性の観点からフィードバックを収集し、今後のシステム展開に向けた改善点を整理・検討する。

## 2-3. 検証ポイント

2-1. 実証仮説で述べた仮説を検証するにあたり、プロトタイプ開発における検証ポイントは以下の通りである。

表 2-1 国土交通省保有行政情報のデータ作成実証における検証ポイント

項目		内容
実証 1	仮説	<ul style="list-style-type: none"> <li>● 大規模言語処理モデル等の活用により、多種多様なフォーマットに対応し、柔軟性のあるシステムとすることができる。</li> </ul>
	検証ポイント	<ul style="list-style-type: none"> <li>● データ作成において、どのレベルまでノーコードによる自動化アプローチが有効か</li> <li>● どの程度複雑なデータセットに適用可能か</li> <li>● ユーザーが直面すると考えられる課題とそれに対する基準設定と運用フロー、ユーザービリティを高めるためのインタフェース等の解決策は何か</li> <li>● ユーザーが実運用に向けて判断の参考となるように、自動化できる範囲やユーザー側において用意する入力データの仕様やシステムへのスキーマ登録方法は何か</li> <li>● 手作業による修正が必要なケースは必ず発生しうること前提に、どのような作業項目において発生するか</li> </ul>
実証 2	仮説	<ul style="list-style-type: none"> <li>● 標準的なデータ仕様およびガイドラインの策定により、安定的にシステムが構造データを生成し、活用されるフォーマットを構築できる。</li> </ul>
	検証ポイント	<ul style="list-style-type: none"> <li>● 各フォーマットの特徴と適用シナリオはどのようなものか</li> <li>● 目的に応じた適切なデータ形式の選択基準はどのようなものか</li> <li>● 大規模言語処理モデル等の活用が不向きなデータ形式や事前に手動による作業が必要なケースはどのようなものか</li> </ul>
実証 3	仮説	<ul style="list-style-type: none"> <li>● オープンソースと有償 API の使い分けや併用が必要である。</li> </ul>
	検証ポイント	<ul style="list-style-type: none"> <li>● オープンソースでどこまで対応可能か、ショートリストごとに精度を算出</li> <li>● 有償 API を用いた場合の性能やコストの違いはどのようなものか</li> </ul>

## 2-4. 実施体制

表 2-3 実施体制

役割	主体	詳細
全体管理	総合政策局 情報政策課	● プロジェクト全体ディレクション
	日建設計総合研究所	● プロジェクト全体マネジメント
実施事業者	マイクロベース	● プロトタイプ開発における開発・検証
	ユーカリヤ	● プロトタイプ開発における開発・検証
	ESRI ジャパン	● プロトタイプ開発における開発・検証
	三菱総合研究所	● プロトタイプ開発における開発・検証
実施協力	観光庁 観光戦略課	● データ提供・検証
	観光庁 観光戦略課 観光統計調査室	● データ提供・検証
	観光庁 観光地域振興部 観光地域振興課	● データ提供・検証
	物流・自動車局 貨物流通事業課	● データ提供・検証
	物流・自動車局 物流政策課	● データ提供・検証
	海事局 内航課	● データ提供・検証
	鉄道局 鉄道事業課	● データ提供・検証
	鉄道局 安全政策課	● データ提供・検証
	気象庁	● データ提供・検証
	大臣官房 総務課	● データ提供・検証
	大臣官房 地方室	● データ提供・検証
	道路局 企画課 道路経済調査室	● データ提供・検証
	水管理・国土保全局 河川計画課 河川情報企画室	● データ提供・検証
	航空局 安全部 無人航空機安全課	● データ提供・検証
	総合政策局 モビリティサービス推進課	● データ提供・検証

## 2-5. スケジュール

本業務の実行スケジュールは以下の通りである。

表 2-4 スケジュール

実施事項	2024 年									2025 年		
	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月
1. 要件定義書・基本設計書・内部設計書の作成		←→										
2. 【データ管理機能】(プロトタイプ) フロントエンド開発			←→									
3. 【データ管理機能】(プロトタイプ) バックエンド開発			←→									
4. 【データ生成機能】(プロトタイプ) バックエンド開発			←→									
5. 【データ配信機能】(プロトタイプ) バックエンド開発						←→						
6. 【データ可視化機能】(プロトタイプ) フロントエンド開発			←→									
7. データ管理システム(データ管理、データ生成、データ配信) 有用性検証								←→				
8. 有用性調査(データ活用関連の関係者ヒアリング)										←→		
9. 業務報告書作成等											←→	

## 3. 開発スコープ

### 3-1. 概要

本開発では、国土交通省が管理する行政情報をデータ化し、活用環境へシームレスに接続するためのデータ管理システム（LINKS Veda）のプロトタイプ開発を行う。本システムは大きく以下の6つの機能群で構成される。（以下、括弧内は仕様書番号を示す。）

- 1) 正規化データの管理機能（D003）
  - データの登録、編集、削除、検索、ダウンロード、公開 API の発行、バージョン管理等の機能を提供する。システム内で処理したデータは D004、D005、D006 で実装する省内利用の可視化アプリケーション、外部サードパーティシステムで利用可能なよう CSV 形式、JSON 形式、GeoJSON 形式、MVT タイル形式で作成し、データ取得可能な URL を発行する。
- 2) 非構造データの自動正規化機能（D001）
  - 国土交通省保有行政情報等の非構造データを LLM(大規模言語処理)、OCR(光学的文字認識)や NLP(自然言語処理)等の技術を活用して機械判読可能データ「正規化ローデータ (Processed Data)」をアウトプットする自動処理機能を提供する。
  - 正規化ローデータ (Processed Data) を後続の D002 で処理可能なようデータクレンジング、地図データにおいてはジオコーディング、座標系統一処理を提供する。
- 3) 正規化ローデータの集計・加工機能（D002）
  - 正規化ローデータ (Processed Data) を集計・加工し「最終構造化データ (Defined Data)」を生成する機能を提供する。
- 4) カスタムグラフ作成機能（D004）
  - D003 の機能と連携したグラフ作成機能を提供する。推移可視化や統計分析等の機能を備える。
- 5) マッピング機能（D005）
  - D003 の機能と連携したマップ作成機能を提供する。地図データの表示、重畳、スタイル編集等の機能を備える。
- 6) 帳票・レポート出力機能（D006）
  - D003、D004、D005 と連携した帳票及びレポート出力機能を提供する。

さらに、本プロトタイプの有用性検証を行うため、以下の工程を実施する。

- 1) データ作成実証
  - 開発したシステム及びデータを用いて、ユーザーによるデータ管理システムのユーザービリティ検証および作成されたデータの正確性の検証を行う。
- 2) 有用性検証

## LINKS24-03\_技術検証レポート\_データ管理システムのプロトタイプ開発

- 国土交通省職員へのヒアリングを通じて、システム及びデータの有用性を検証する。データの可視化環境や活用ソリューションの提案を含む。
- 3) 開発成果の公開
    - 開発したシステムを OSS 化し、チュートリアルや構築マニュアル等のドキュメントを作成する。また、サンプルデータの公開等の OSS 利用促進施策を実施する。
  - 4) 業務報告書の作成
    - 本業務で取得した知見を業務報告書としてまとめ、技術資料を作成する。技術資料は公表を前提とし、技術者向けに再現性を確保した形で作成する。

## 4. 実証システム

### 4-1. アーキテクチャ

#### 4-1-1. システムアーキテクチャ

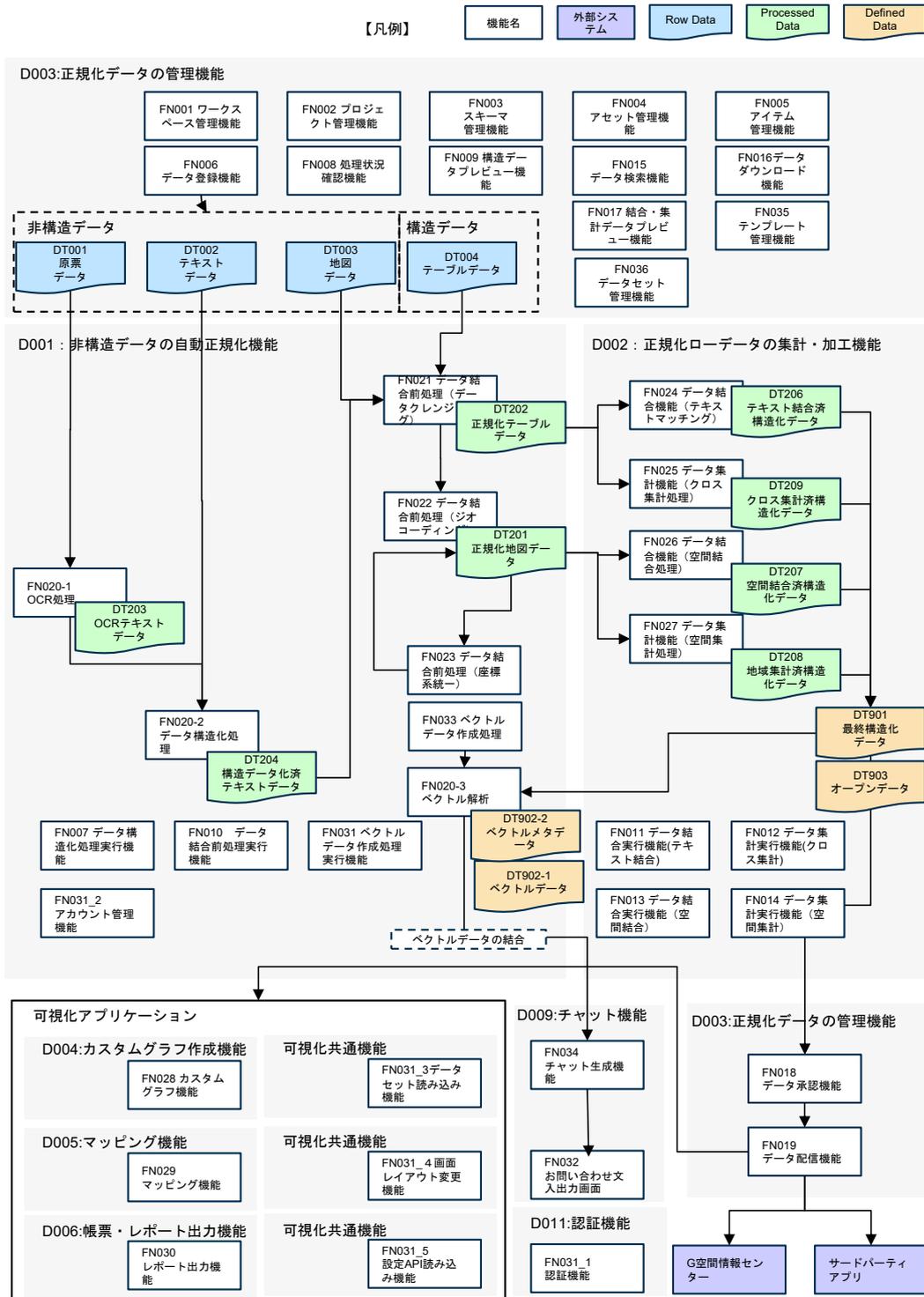


図 4-1 システムアーキテクチャ

4-1-2. データアーキテクチャ

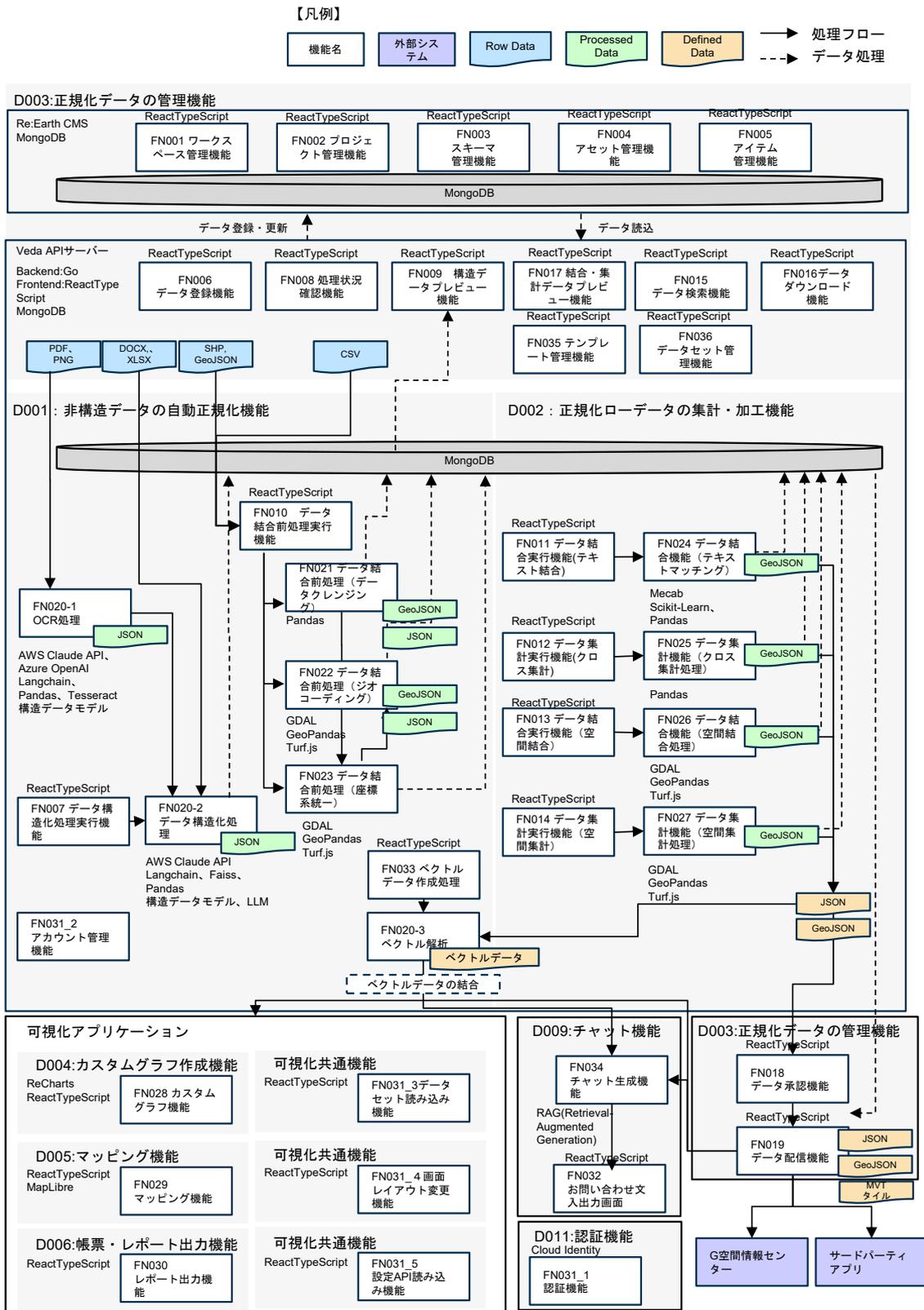


図 4-2 データアーキテクチャ

### 4-1-3. システム構成

本システムにおける Re:Earth CMS（以下、CMS）は、システム内に登録したデータ（アセット）、システム内で生成したデータ（コンテンツ）を一元的に管理する役割を持つ。

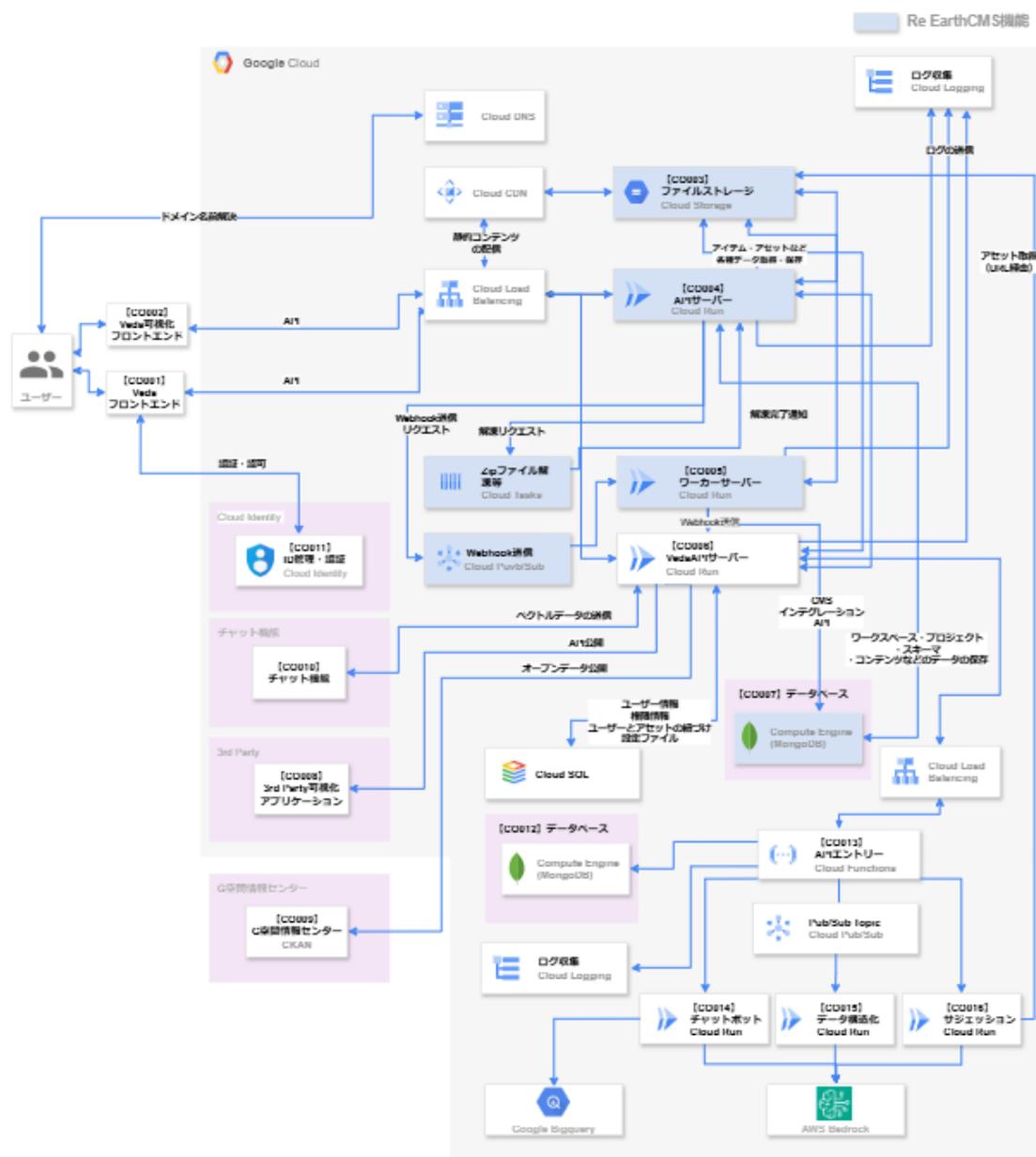


図 4-3 システム構成図

表 4-1 利用するコンポーネント一覧

ID	種別	コンポーネント名	用途
CO001	Veda フロントエンド	HTML・CSS・JavaScript	<ul style="list-style-type: none"> <li>Web ブラウザ上で動作する、(HTML・CSS・JavaScript による) フロントエンドアプリケーション。API サーバー・Cloud Identity・その他各種サーバーとの通信を行い、オペレーターの実行やプレビュー機能によるテーブルデータや地図データの表示をおこなうアプリケーション。</li> </ul>
CO002	Veda 可視化フロントエンド	HTML・CSS・JavaScript	<ul style="list-style-type: none"> <li>LINKS Veda 内で作成されたデータをもとに地図データ、グラフの可視化表現をおこなうアプリケーション。</li> </ul>
CO003	ファイルストレージ	Cloud Storage	<ul style="list-style-type: none"> <li>フロントエンドのアプリケーションのソースコードや画像、ユーザーによってアップロードされたアセットファイルを保存する、オブジェクトストレージサーバー。Google Cloud Storage (GCS) を使用している。容量は無制限、自動的にスケールリングし、バックアップも自動的に行われ、データは複数拠点に分散配置される。巨大なファイルを格納・配信することが可能。LINKS Veda では、静的ファイルを GCS に保存している。</li> </ul>
CO004	API サーバー	Cloud Run	<ul style="list-style-type: none"> <li>LINKS Veda は Re:Earth CMS と連携する。Re:Earth CMS の API サーバーであり、HTTP サーバーとして外部からのリクエストを受信している。MongoDB や Google Cloud Storage と連携して、アイテムの保存やプロジェクトの管理・公開などのさまざまなビジネスロジックを実行する。API サーバーは、Cloud</li> </ul>

			<p>Run 上で動作する。Cloud Run とは、Google Cloud で利用可能なサーバーレス (CaaS) プラットフォームであり、Docker コンテナをデプロイすることで、サーバーの保守管理の手間なしに、アプリケーションをクラウド上で動作させることができる。同時接続リクエスト数が規定数以上に達すると自動的にコンテナが増加し、より多くのトラフィックを自動的に分散処理することができる。Cloud Run は、デフォルト設定では、HTTP リクエストを受信して処理している間のみ CPU が動作し、リクエストを処理していない時は動作を停止するため、HTTP リクエストを実際に受信し処理するために動作した CPU 時間分のみが課金対象となる。</p>
CO005	ワーカーサーバー	Cloud Run	<ul style="list-style-type: none"> <li>● LINKS Veda における非同期バックグラウンド処理を行うサーバーであり、HTTP サーバーとして API サーバーからのリクエストを受信している。ZIP ファイル等の解凍処理や Webhook の送信などを Cloud PubSub と連携しながら行っている。ワーカーサーバーは、API サーバーと同じく、Cloud Run 上で動作する。</li> </ul>
CO006	VedaAPI サーバー	Cloud Run	<ul style="list-style-type: none"> <li>● Re:Earth CMS、Microbase バックエンドと連携し LINKS Veda で動作する API サーバーであり外部サービスとの連携を担う。詳細は以下の機能を持つ。 <ul style="list-style-type: none"> <li>◇ ユーザーの権限管理</li> <li>◇ Re:Earth CMS からのアセット、アイテムなど各種データ取得、保存。</li> </ul> </li> </ul>

			<ul style="list-style-type: none"> <li>◇ データ構造化処理、結合処理、集計処理のリクエスト・レスポンスの受信と保存</li> <li>◇ チャット機能へのベクトルデータ送信。</li> <li>◇ G 空間情報センターへのデータ登録処理。</li> <li>◇ 3rd Party 可視化アプリケーション向けのデータ配信 API の発行、共有 URL データの保存</li> </ul> <ul style="list-style-type: none"> <li>● Veda API サーバーは、API サーバーと同じく Cloud Run 上で動作する。</li> </ul>
CO007	データベース	Mongo DB	<ul style="list-style-type: none"> <li>● MongoDB は、MongoDB 社が開発し、オープンソースとして提供される NoSQL データベース・ドキュメント指向型データベースである。</li> <li>● LINKS Veda ではデータベースとして Mongo DB を使用し、Re:Earth CMS、LINKS Veda（データ管理部分）でそれぞれデータベースを保持する。</li> <li>● Re:Earth CMS の DB には、ワークスペース、プロジェクト、アセット、コンテンツに関する情報が保存され、LINKS Veda（データ管理部分）ではユーザー、テンプレート、アクセス権限管理に関する情報が保存される。</li> <li>● Compute Engine 上で動作する</li> </ul>
CO008	3ed Party 可視化アプリケーション	-	<ul style="list-style-type: none"> <li>● ArcGIS、QGIS、ForePaas などの外部のプラットフォームを指す。LINKS Veda ではデータ利用のための公開 API を発行し、API からデータの読み込みを可能とする。</li> </ul>
CO009	G 空間情報センター	CKAN	<ul style="list-style-type: none"> <li>● 官民間問わずさまざまな主体により整備・提供される多様な地理空間情報を集約し、利用者がワンストップで</li> </ul>

			検索・ダウンロードし利用できる、産学官の地理空間情報を扱うプラットフォーム。LINKS Veda では、システム管理者がオープンデータ公開処理を行うタイミングで、自動的にG 空間情報センターへ登録を行う。
CO010	チャット機能	Cloud Run	<ul style="list-style-type: none"> <li>● LINKS Veda 内の機能の一つで、ベクトルデータを用いたチャットによる問い合わせのためのアプリケーションを指す。</li> </ul>
CO011	ID 管理・認証	Cloud Identity Platform	<ul style="list-style-type: none"> <li>● Google 社が提供する顧客 ID とアクセス管理 (CIAM) のシステムであり、ID 管理機能とアクセス管理機能をアプリケーションに組み込むことにより、ユーザーアカウントの保護と信頼性の高いスケーリングを可能にするものである。Cloud Identity Platform は、マルチテナント SaaS アプリケーション、モバイルアプリやウェブアプリ、ゲーム、API などに対応している。Cloud Identity を使用することで、開発者はアプリケーションに安全で使いやすく信頼性の高い認証認可機能を組み込むことができる。</li> </ul>
CO012	データベース	MongoDB	<ul style="list-style-type: none"> <li>● Re:Earth CMS とのやり取りで追従データを保管する。</li> <li>● Compute Engine 上で動作する</li> </ul>
CO013	API エントリー	Cloud Functions	<ul style="list-style-type: none"> <li>● Re:Earth CMS からのアセットデータを受け取る。</li> <li>● LLM とのやり取りが必要なアセット情報を VedaAPI サーバーから受け取る。</li> <li>● バックグラウンド処理が必要な場合、Pub/Sub 経由で CloudRun のジョブへ依頼する。</li> </ul>
CO014	チャットボット	Cloud Function	<ul style="list-style-type: none"> <li>● ユーザーからの問い合わせ内容を受</li> </ul>

			<p>け、ベクトルデータベース DB に関連データを取得し、LLM への回答生成を依頼する。</p> <ul style="list-style-type: none"> <li>● 処理後 Re:Earth CMS にデータを返却する。</li> </ul>
CO015	データ構造化	Cloud Run	<ul style="list-style-type: none"> <li>● Re:Earth CMS からのアセットデータを受け取る。</li> <li>● 取得したアセットを OCR 処理し、LLM を用いて構造データを生成する。</li> <li>● 処理後 Re:Earth CMS にデータを返却する。</li> </ul>
CO016	サジェッション	Cloud Run	<ul style="list-style-type: none"> <li>● Re:Earth CMS からのアセットデータを受け取る。</li> <li>● 取得したアセットを OCR 処理し、LLM を用いてファイルの構造化に必要なカラム名や説明文などを抽出する。</li> <li>● 処理後 Re:Earth CMS にデータを返却する。</li> </ul>
その他	-	Cloud DNS	<ul style="list-style-type: none"> <li>● Google Cloud の DNS。LINKS Veda で使用しているドメインに対応するレコードは全て Cloud DNS で管理されている。</li> </ul>
	-	Cloud CDN	<ul style="list-style-type: none"> <li>● Google Cloud の CDN (Content Deliver Network = ウェブコンテンツをインターネット経由で配信するために最適化されたネットワーク)。Google Cloud などと組み合わせて使用することで、リクエスト元から地理的に近いサーバーにコンテンツのキャッシュを自動的に配置し、コンテンツ配信を高速化・効率化させることができる。</li> </ul>
		Cloud Load Balancing	<ul style="list-style-type: none"> <li>● Google Cloud のマネージドなロードバランサ (負荷分散システム)。LINKS Veda で使用されるドメイン</li> </ul>

			の IP アドレスは全て Cloud Load Balancing に向いており、リクエストのホスト（ドメイン）に応じて API サーバーやストレージサーバーに自動的にルーティングされる。
		Cloud SQL	<ul style="list-style-type: none"> <li>● LINKS Veda におけるユーザー管理、ユーザーとアセットの紐づけ、EBPM Tools で利用する設定ファイル等を保存する。</li> </ul>
	ログ収集	Cloud Logging	<ul style="list-style-type: none"> <li>● Google Cloud のログ収集サービス。Cloud Run などから出力されるログを閲覧可能。システムのトラブルシューティング時に役立つ。</li> </ul>
	Webhook 送信	Cloud Pub/Sub	<ul style="list-style-type: none"> <li>● Google Cloud のメッセージングサービス。アプリケーション間の連携を実施するために利用される。LINKS Veda では、API サーバーとワーカーサーバー間の通信に使用している。</li> </ul>
	Pub/Sub Topic	Cloud Pub/Sub	<ul style="list-style-type: none"> <li>● Google Cloud のメッセージングサービス。API エントリからメイン処理へタスクを割り振る。</li> </ul>
	ベクトルデータベース	Google Big Query	<ul style="list-style-type: none"> <li>● ベクトル化したデータを格納するデータベースになる。</li> </ul>
		AWS Bedrock	<ul style="list-style-type: none"> <li>● LLM モデルへの問い合わせる際に使うサービスになる。</li> </ul>
		API Gateway	<ul style="list-style-type: none"> <li>● Google Cloud のフルマネージド API ゲートウェイ。Cloud Functions、Cloud Run、App Engine 向け API を管理する。システム上の API へアクセスする承認機能に対応し、パフォーマンス改善する技術などもサポートするサービスである。</li> </ul>

4-1-4. ハードウェアアーキテクチャ

利用したハードウェア一覧

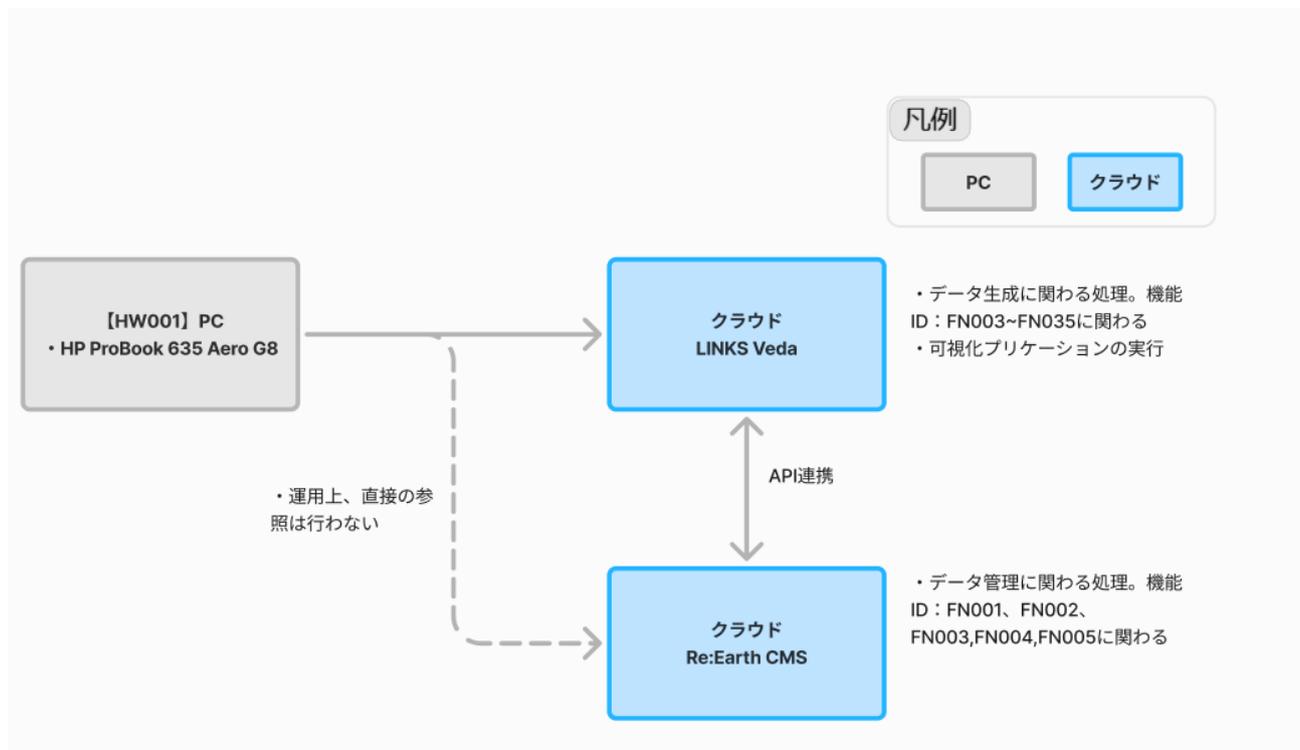


図 4-4 ハードウェアアーキテクチャ

表 4-2 利用するハードウェア一覧

ID	種別	品番	用途
HW001	実証用 PC	HP ProBook 635 Aero G8	<ul style="list-style-type: none"> <li>● LINKS Veda へのアクセス、データ生成に関わる処理の実行およびデータの公開処理</li> <li>● Re:Earth CMS へのアクセス</li> <li>● LINKS Veda 経由での各ユースケース EBPM Tools の実行</li> </ul>

利用したハードウェア詳細

1) HW001

- 選定理由
  - 実証を行う国土交通省職員の標準 PC のスペックであるため。
- 仕様・スペック
  - CPU：AMD Ryzen (TM) 3 5400U
  - メモリ：32GB
  - ブラウザ：Microsoft Edge、Google Chrome
  - SSD：512GB
  - ディスプレイ：13.3 インチ
  - 重さ：約 1000g
  - OS：Windows11 Enterprise
- 画面サイズの制約
  - EBPM Tools 検証時は、コンテンツを正常に表示させるため、以下の環境での実証を推奨する。最小サイズを下回る場合には、正しくコンテンツ表示ができず、オペレーションに影響が出る可能性がある。
    - ◇ 推奨動作環境
      - ブラウザ：1440 × 900、コンテンツ領域：1440 × 802
    - ◇ 最小動作環境
      - ブラウザ：1200 × 832、コンテンツ領域：1200 × 750

## 4-2. システム機能

## 4-2-1. システム機能一覧

表 4-3 機能一覧

大分類	分類名	ID	機能名	機能説明
D003	正規化データの管理機能	FN001	ワークスペース管理機能	<ul style="list-style-type: none"> <li>● 本機能は、CMS 内の機能である。複数ユーザーがデータ登録、管理するための作業場所としてワークスペースを作成・管理する。</li> </ul>
D003	正規化データの管理機能	FN002	プロジェクト管理機能	<ul style="list-style-type: none"> <li>● 本機能は、CMS 内の機能である。プロジェクトはワークスペースに複数作成可能で、本機能はデータ管理の目的に応じてプロジェクトを作成・管理する。</li> <li>● 1つのワークスペースに1つのプロジェクトは必須で作成となる。LINKS Veda では1つのプロジェクトを作成する。</li> </ul>
D003	正規化データの管理機能	FN003	コンテンツ管理機能	<ul style="list-style-type: none"> <li>● 本機能では、D001、D002 の処理実行によって作成されるコンテンツ情報を管理する機能を提供する。本機能はコンテンツメニューから呼び出す。コンテンツの検索、削除、名称変更、編集、ダウンロードが可能である。各コンテンツの編集・参照権限は、ログインユーザーの所属（原課レベル）によって制御する。</li> <li>● コンテンツの編集は、レコードのインサート、カラムの追加、削除、カラム名の編集、アイテムの追加、削除、フィールド内の値の編集を可能とする。</li> <li>● コンテンツの管理情報として以下を保持する。 <ul style="list-style-type: none"> <li>・ コンテンツは、オープンデータ、EBPM Tools、チャットでそれぞれ利用する。それらを管理する公開ステータスとして、「オープン</li> </ul> </li> </ul>

				<p>データ]、「可視化]、「チャット]を保持する。公開ステータスは、コンテンツ管理と【FN036】データセット管理機能の双方で保持する。</p> <ul style="list-style-type: none"> <li>・ 「オープンデータ」の公開ステータスは、データ作成、データ公開のステータスによって管理する。データ作成では、スイッチのオン/オフにより、オープンデータ向けの CSV、JSON、GeoJSON、MVT タイル形式データの作成を行う。データ公開では、スイッチのオン/オフを行うことでオープンデータの公開可否を管理する。オープンデータの公開ステータスは、システム管理者のみ更新を行うことができる。</li> <li>・ 「チャット」の利用ステータスは、ベクトルデータの作成スイッチのオン/オフによって管理する。</li> <li>● 公開ステータスの他、コンテンツ名、データサイズ、作成者、作成日、更新者、最終更新日などのメタデータとして保持する。メタデータの詳細は 4-2-3. 開発機能の詳細にて記載する。</li> <li>● コンテンツ情報は、スキーマ情報と合わせてユーザーがダウンロード可能とする。ダウンロードファイルの詳細は 4-2-3. 開発機能の詳細にて記載する。</li> </ul>
D003	正規化データの管理機能	FN004	アセット管理機能	<ul style="list-style-type: none"> <li>● 本機能では、アップロードされたファイルの単位 (Raw Data) であるアセットを管理する機能を提供する。本機能はアセットメニューから呼び出す。アセットの追加、検索、削除が可能である。</li> <li>● アセット管理のため、システム内でデ</li> </ul>

				<p>ータ管理するためのメタデータとして、ファイル名、ファイル形式 (PDF/CSV/JSON/GeoJSON/SHP/ZIP/DOCX/XLSX)、ファイルサイズ、作成者、作成日時の情報を保持する。</p> <ul style="list-style-type: none"> <li>● アセット画面では、アセットを選択することで、アセット一覧と同様の画面でファイルのプレビューを確認できる。GeoJSON の場合が地図ビューワーで地図のプレビューが可能である。</li> </ul>
D003	正規化データの管理機能	FN005	アイテム管理機能	<ul style="list-style-type: none"> <li>● 本機能は、CMS で管理するデータの最小単位で、コンテンツのレコードを管理する。</li> </ul>
D003	正規化データの管理機能	FN006	データ登録機能	<ul style="list-style-type: none"> <li>● 本機能は、アセットメニューから呼び出す。任意の静的ファイルをローカルから選択し、アセットとして新規アップロードしデータ登録をする機能である。登録データは編集、削除可能である。</li> <li>● アップロードファイルのデータ形式は、原票データ (PDF、PNG)、パッケージ (ZIP)、テキストデータ (DOCX、XLSX)、テーブルデータ (CSV、JSON)、地図データ (SHP、CSV、GeoJSON) での登録を許容する。</li> <li>● データ登録時に ZIP でファイルがアップロードされた場合には、登録時に解凍処理を自動で実行する。</li> </ul>
D001	非構造データの自動正規化機能	FN007	データ構造化処理 (UI)	<ul style="list-style-type: none"> <li>● 本機能では、非構造データ (原票データ、テキストデータ) に対し、構造化データ化 (コンテンツ作成) する処理を実行する画面を提供する。ユーザーは自身がアップロードしたアセット選択し、アセットビューワーで参照しながら、構造化するためのスキーマ構成を入力項目に則って指定する。なお、さらに詳細な条件文の記入やテンプレ</li> </ul>

				<p>トを利用しないプロンプト入力に慣れているユーザー向けの UI として自由文の記述項目を設ける。</p> <ul style="list-style-type: none"> <li>● 事前にテンプレートとして定義したプロンプトを利用する場合には、「テンプレート選択」ボタンから呼び出すことができる。実行時に利用したプロンプトを新規にテンプレートとして保存する場合には「テンプレートとして保存」ボタンを押下する。</li> <li>● スキーマ構成の入力サポートとして、サジェッションも可能である。「スキーマ提案の実行」ボタンを押下することで、【FN020_4】データ構造化前サジェッション機能を起動し、アセットから想定しうるスキーマ構成を LLM で提案することができる。</li> <li>● データ構造化は、1 ファイル 1 レコードで作成されるが、オプションとして 1 ファイル複数レコードを作成することもできる。</li> <li>● データ構造化に加えて、出典名の付与を行う。ユーザーは構造化対象のカラム名、自由入力のテキストを組み合わせることで出典名の作成を行う。本機能で付与した出典名は、【FN034】チャット生成機能のレスポンスで利用する。</li> </ul>
D003	正規化データの管理機能	FN008	処理状況確認機能	<ul style="list-style-type: none"> <li>● 本機能では、D001、D002 のオペレーターごとの処理ステータス（処理待/処理中/処理済/保存済/エラー）を実行ユーザーごとに一覧で確認できる機能を提供する。処理状況一覧メニューから遷移する。</li> <li>● 処理状況を押下することで、実行時のオペレーターの詳細画面に遷移することができる。</li> </ul>

D003	正規化データの管理機能	FN009	構造データ化プレビュー機能	<ul style="list-style-type: none"> <li>● D001 処理実行後の確認として、構造化されたデータプレビューをテーブル形式で行う機能を提供する。</li> <li>● 本機能はデータ構造化処理 (UI) に具備する。</li> </ul>
D001	非構造データの自動正規化機能	FN010	データ結合前処理実行機能 (UI)	<ul style="list-style-type: none"> <li>● 本機能では、D001 で構造化されたデータ、地図データ、テーブルデータに対し、【FN021】データ結合前処理 (データクレンジング) を実行する。ユーザーはデータクレンジングに必要な条件をオプションとして設定し、実行可能とする。</li> <li>● 本機能では、地図データでないコンテンツに対し、【FN022】データ結合前処理 (ジオコーディング) を実行する。ユーザーはジオコーディングを行う住所を保持するカラムを指定し、対象のカラムに対してジオコーディングを実行する。処理の結果、新規に緯度経度の情報とポイントのジオメトリが追加される。</li> <li>● カラムの秘匿化処理を行う。秘匿化処理には、ランク化、偏差値化、コードの新規採番、住所カラムの秘匿化の処理が含まれる。</li> <li>● 資料名の付与を行う。任意のテキストを入力することで資料名の作成を行う。本機能で付与した資料名は、【FN034】チャット生成機能のレスポンスで利用する。</li> <li>● 地図データの場合は、バックエンドで【FN003】データ結合前処理 (座標系統一) を実行し座標系を WGS84 に設定する。</li> </ul>
D002	正規化ローデータの集計・加工機能	FN011	データ結合実行機能 (テキストマッチング) (UI)	<ul style="list-style-type: none"> <li>● 本機能では、2つのコンテンツを対象に、結合処理を行う機能を提供する。ユーザーは、結合元のコンテンツ (以下、メインコンテンツ) と、情報を付</li> </ul>

				<p>加する結合先のコンテンツ（以下、結合先コンテンツ）を指定し、両者を結びつける基準となるカラム（以下、キーカラム）を選択して結合処理を実行する。なお、メインコンテンツのデータを基準として、結合先コンテンツのデータが付加される形で結合処理される。</p> <ul style="list-style-type: none"> <li>● 結合先コンテンツについては、結合後にメインコンテンツに保持させたいカラム（以下、結合カラム）を指定することができる。結合カラムを指定しない場合は、結合先コンテンツの全カラムが対象となる。なお、同時に結合できる結合先コンテンツは1つのみであり、複数のコンテンツと結合する場合は本機能を繰り返し使用する必要がある。</li> <li>● 結合条件に一致しなかったデータは後続の処理対象からは除外される。</li> </ul>
D002	正規化ローデータの集計・加工機能	FN012	データ集計実行機能（クロス集計）(UI)	<ul style="list-style-type: none"> <li>● 本機能では、選択したコンテンツに対し、集計単位と集計対象を指定し、カラムの合計・平均値、およびカウント値を算出する機能を提供する。</li> </ul>
D002	正規化ローデータの集計・加工機能	FN013	データ結合実行機能（空間結合）(UI)	<ul style="list-style-type: none"> <li>● 本機能では、地図データの空間結合を実施する。結合処理の基準となる地図データ(以下、メイン地図データ)と、情報を付加するために結合される地図データ(以下、結合地図データ)を指定し、結合方式を選択してデータを結合する。結合方式には交差結合、最近傍結合が存在し、ユーザーは任意の結合方式を選択し実行することができる。最近傍結合の場合には、検索上限範囲を指定する。</li> <li>● 【FN024】 データ結合実行機能（テキストマッチング）と同様、結合条件に一致しなかったデータは後続の処理対</li> </ul>

				象からは除外される。
D002	正規化ローデータの集計・加工機能	FN014	データ集計実行機能（空間集計）（UI）	<ul style="list-style-type: none"> <li>● 本機能では、地図データの集計を実施する。メイン地図データ、集計地図データ、集計単位、集計対象を指定し、カラムの合計・平均値、およびカウント値を算出する機能を提供する。</li> </ul>
D003	正規化データの管理機能	FN015	データ検索機能	<ul style="list-style-type: none"> <li>● 【FN003】コンテンツ管理機能、【FN004】アセット管理機能、【FN035】テンプレート管理機能のそれぞれにおいて、一覧から検索可能な機能を提供する。</li> </ul>
D003	正規化データの管理機能	FN016	データダウンロード機能	<ul style="list-style-type: none"> <li>● 本機能では、コンテンツ情報、スキーマ情報を【FN003】コンテンツ管理機能からダウンロードできる機能を提供する。</li> <li>● データダウンロードは、コンテンツ管理機能の管理パネルから実行できるデータ作成後に実行可能となる。</li> </ul>
D003	正規化データの管理機能	FN017	結合・集計データプレビュー機能（UI）	<ul style="list-style-type: none"> <li>● D002 処理実行後の確認として、結合/集計されたデータプレビューをテーブル形式、地図形式で行う機能を提供する。</li> </ul>
D003	正規化データの管理機能	FN018	データ承認機能	<ul style="list-style-type: none"> <li>● 最終構造化データに対し、システム管理者によるデータ承認を【FN003】コンテンツ管理機能から行う機能を提供する。【FN019】データ配信機能におけるオープンデータ公開処理と同義とする。</li> </ul>
D003	正規化データの管理機能	FN019	データ配信機能	<ul style="list-style-type: none"> <li>● データ配信には、①オープンデータ向けの配信（G 空間情報センターへの公開）、②EBPM Tools 向けの配信の2種類が存在する。</li> <li>● オープンデータ化向けの配信については、システム管理者によってのみ処理可能とする。②EBPM Tools 向けの配信については一般ユーザーにて可能とする。データの公開処理は【FN003】</li> </ul>

				コンテンツ管理機能の管理パネル、 【FN036】データセット管理機能のデータ公開スイッチから実施する。
D001	非構造データの自動正規化機能	FN020_1	OCR 処理	● 【DT001】原票データを対象に、画像データから文章および丸囲み等の図形オブジェクトを抽出し、テキストデータに変換する機能を提供する。
		FN020_2	データ構造化処理	● 【FN020_1】でテキスト化したデータおよび【DT002】テキストデータを対象に、【AL02】構造データ生成モデルによって、指定した出力スキーマに従い、構造データに変換する機能を提供する。
		FN020_3	ベクトル解析	● 【FN034】チャット生成機能にてユーザーが登録したデータの内容を問い合わせることができるよう、ベクトル化処理を実施する機能を提供する。
		FN020_4	データ構造化前サジェクション機能	● 【FN020_2】データ構造化処理に必要な出力スキーマの候補を提案し、ユーザーがデータ構造化処理の入力を支援する機能を設ける。
D001	非構造データの自動正規化機能	FN021	データ結合前処理（データクレンジング）	● 【FN024】データ結合処理（テキストマッチング）前にユーザーが指定したカラムを対象に表記ゆれの修正やデータクレンジング処理（スペースの削除や全角・半角の削除、テキストの置換処理など）を行う。 ● 表記ゆれの正規化には、番地表記、市区町村名に含まれる小書き文字表記、数字表記の統一、単位、日時表記の統一が含まれる。
D001	非構造データの自動正規化機能	FN022	データ結合前処理（ジオコーディング）	● 住所情報を保持するカラムに対してジオコーディング（座標付与）を実行し、新規に緯度経度のカラムを追加する機能を提供する。

D001	非構造データの自動正規化機能	FN023	データ結合前処理（座標系統一）	<ul style="list-style-type: none"> <li>● 登録データとなる【DT003】地図データにおいて複数の座標系が混在している場合に備え、WGS84 に座標系統一処理を実行する。</li> </ul>
D002	正規化ローデータの集計・加工機能	FN024	データ結合処理（テキストマッチング）	<ul style="list-style-type: none"> <li>● 任意のテーブルデータに対しテキストマッチングによるインデキシング処理を行う機能を提供する。</li> <li>● テキストマッチングにおいて、結合に用いるデータ間における類似度を算出し、指定した類似度の閾値以上の類似度となるデータを属性結合する。</li> <li>● テキストマッチングを利用した属性結合において、結合前後における総件数および結合率の表示を選択する。</li> </ul>
D002	正規化ローデータの集計・加工機能	FN025	データ集計処理（クロス集計）	<ul style="list-style-type: none"> <li>● ユーザーが指定した任意のカラムをキーに要約統計量（合計、カウント、平均）を算出する機能を提供する。</li> </ul>
D002	正規化ローデータの集計・加工機能	FN026	データ結合処理（空間結合）	<ul style="list-style-type: none"> <li>● 複数データを選択し、ポリゴンとポイントにおける交差結合をする機能を提供する。位置精度によるずれを防ぐため、最近傍結合も考慮する。</li> </ul>
D002	正規化ローデータの集計・加工機能	FN027	データ集計処理（空間集計）	<ul style="list-style-type: none"> <li>● ユーザーが指定した条件ののっとり（都道府県単位等）、ユーザーがアップロードした地域ポリゴンデータを結合し、地域単位で集計し、新規地図データとして保存する機能を提供する。</li> </ul>
D004	カスタムグラフ作成機能	FN028	グラフカスタマイズ機能	<ul style="list-style-type: none"> <li>● 本機能は、EBPM Tools における共通機能として提供する。【FN019】データ配信機能のアプリケーション利用で紐づけされたコンテンツに対し、事前に定義されたレイアウトパターンを選択しグラフ化する機能を提供する。</li> </ul>
D005	マッピング機能	FN029	マッピング機能	<ul style="list-style-type: none"> <li>● 本機能は、EBPM Tools における共通機能として提供する。【FN019】データ配信機能で紐づけされたコンテンツ</li> </ul>

				(地図データ) に対し、ダッシュボードに地図データをマッピングする機能を提供する。地図データの表示、レイヤの重畳、スタイル編集等の機能を提供する。
D006	レポート出力機能	FN030	レポート出力機能	<ul style="list-style-type: none"> <li>● 本機能は、EBPM Tools における共通機能として提供する。D004、D005 と連携したレポート出力機能を提供する。</li> </ul>
-	共通機能	FN031_1	認証機能	<ul style="list-style-type: none"> <li>● 任意のユーザーに対してアカウント発行、メールアドレスおよびパスワードの変更を行う機能を提供する。</li> <li>● 2段階認証による認証機能を提供する。</li> </ul>
		FN031_2	アカウント管理機能	<ul style="list-style-type: none"> <li>● ユーザーの権限管理を行う機能を提供する。</li> </ul>
		FN031_3	データセット読み込み機能	<ul style="list-style-type: none"> <li>● 本機能は、EBPM Tools における共通機能として提供する。ユーザー(=原課レベル)は、EBPM Tools 画面訪問時に、利用するデータセットをリストから選択して読み込み、EBPM Tools とデータの紐づけを行う。データセットにはユースケースで利用するコンテンツ情報が紐づけられている。</li> <li>● 選択するデータセットは【FN036】データセット管理機能にて作成したEBPM 用データセットのみ表示される。利用できるコンテンツは【FN004】コンテンツ管理の公開ステータスで「可視化」がオンであるものに限る。</li> </ul>
		FN031_4	設定 API 機能	<ul style="list-style-type: none"> <li>● 設定 API 機能は、ユースケース+ユーザーの組み合わせで設定情報を Veda API サーバー内に保存する機能である。初期設定後、画面再訪問時に自動で読み込みがなされるため、ユーザーによる都度の読み込みは不要である。</li> </ul>

				<ul style="list-style-type: none"> <li>● 取り込んだデータセット、グラフィックアウトの変更、色の変更、グラフ種類の変更などユーザーレベルで実施するカスタマイズはオペレーションの都度DBに保存される。</li> </ul>
		FN031_5	データ参照機能	<ul style="list-style-type: none"> <li>● Defined Data をブラウザ上で参照できる。</li> </ul>
D009	チャット機能	FN032	問い合わせ文入出力画面	<ul style="list-style-type: none"> <li>● ユーザーがお問い合わせ内容を入力し、入力内容に対して【FN034】での生成結果を出力する機能である。</li> </ul>
		FN033	ベクトルデータ作成機能 (ベクトル解析)	<ul style="list-style-type: none"> <li>● Defined データおよび原票データ・テキストデータの Raw Data に対して、【FN20-3】ベクトル解析を実行する機能である。</li> </ul>
		FN034	チャット生成機能	<ul style="list-style-type: none"> <li>● 【FN032】問い合わせ文入出力画面のユーザーからのお問い合わせ入力内容を元に関連性のあるデータを【FN033】ベクトルデータ作成機能(ベクトル解析)で生成したベクトルデータから検索・取得し、質問と取得したデータを埋め込んだプロンプトを作成し、LLM によって【FN032】問い合わせ文入出力画面で出力するユーザーへの返答文(自然言語)を生成する機能である。</li> </ul>
D003	正規化データの管理機能	FN035	テンプレート管理機能	<ul style="list-style-type: none"> <li>● 本機能では、データ生成に必要な処理を一連処理として実行するため、各オペレーターをシーケンシャルに設定したワークフローテンプレート(以降、ワークフロー)、個別のオペレーター起動に必要なパラメータ情報を設定する個別テンプレートを作成・実行する。</li> </ul>
D003	正規化データの管理機能	FN036	データセット管理機能	<ul style="list-style-type: none"> <li>● 本機能では、LINKS Veda 内で管理するコンテンツ情報を同一の属性情報を持ったデータ群として管理するため、ユースケースの単位でデータセットとして管理する機能を提供する。</li> </ul>

				<ul style="list-style-type: none"> <li>データセット管理画面では、G 空間情報センターに公開する際のメタデータ情報を入力する。</li> </ul>
		FN037	秘匿化処理機能	<ul style="list-style-type: none"> <li>データ公開に向けて、数値情報の階層化・偏差値化や新規コードの追加、住所の秘匿化を行い、元の情報を特定できないように秘匿化する機能を提供する。</li> </ul>
		FN038	英語カラム化機能	<ul style="list-style-type: none"> <li>データ公開に向けて、LLM を用いて日本語カラムを英語カラムへと変換する機能を提供する。</li> </ul>

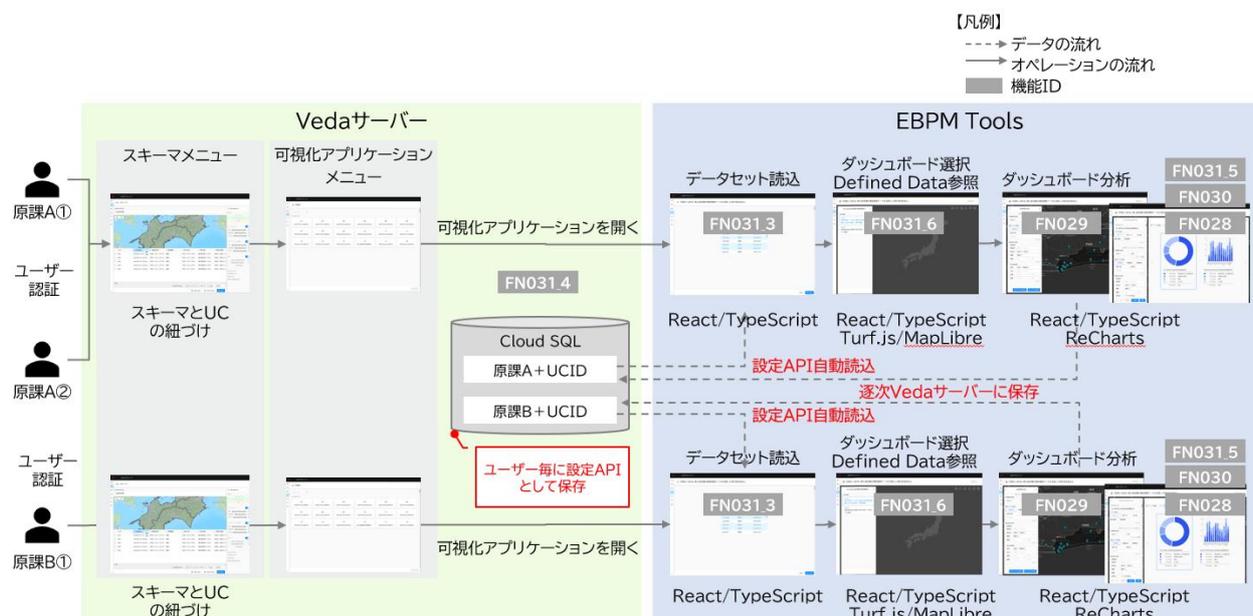


図 4-5 Veda と EBPM Tools 機能の関連

## 4-2-2. 利用したソフトウェア・ライブラリ

表 4-4 利用するソフトウェア・ライブラリ

ID	項目	バージョン	内容
SL001	Ubuntu	20.04 LTE	● Linux Distribution
SL002	miniconda	4. 9. 2	● Python ライブラリの管理
SL003-1	AWS Bedrock API	bedrock- 2023-05-31	● 大規模言語処理モデルを呼び出し、プロンプトに基づきデータ生成するための API。API には AWS 社の Bedrock API を利用、モデルには Claude3 を利用。 ●
SL003-2	Azure AI Vision API	3.1	● OCR の精度向上のため、マイクロソフト社が提供する Azure AI Vision API を利用。
SL004	Langchain	0.1.2	● オープンソース自然言語処理(NLP) LLM ライブラリ
SL005	Faiss	1.5.3	● Meta 社が提供する、高性能な類似性検索ライブラリ
SL006	Mecab	0.996	● 日本語のテキストの形態素解析を行う OSS ライブラリ
SL007	Scikit-Learn	1.4.2	● テキストマッチングを実施するための OSS ライブラリ
SL008	GDAL	3.8.5	● 地理空間情報データ操作のライブラリで、ラスタ・ベクター形式の変換等を実施
SL009	Pandas	2.2.2	● データフレームの操作を行うためのライブラリ
SL010	GeoPandas	0.14.3	● 地理空間情報データ操作のライブラリで、テーブルデータ処理等を実施
SL011	Tesseract	5.4.0	● オープンソースの OCR ライブラリ
SL012	React/TypeScript	-	● フロントエンドの Web アプリケーション開発言語
SL013	Turf.js	-	● オープンソースの地理空間解析ライブラリ
SL014	ReCharts	-	● オープンソースの統計グラフ可視化ライブラリ
SL015	Leaflet	-	● 2D 地図ライブラリ
SL016	Re:Earth CMS	-	● データ登録・管理・配信をノーコードで実行可能なオープンソースのヘッドレス CMS

## LINKS24-03\_技術検証レポート\_データ管理システムのプロトタイプ開発

SL019	MongoDB	-	● ドキュメント指向の NoSQL のデータベースで、データの柔軟性と拡張性が特徴である。
SL020	Go	-	● バックエンドの Web サーバー開発言語
SL021	GraphQL	-	● API のクエリ言語およびランタイムを指す。バックエンドとフロントエンド間の通信で利用する
SL022	Cloud Identity	-	● ユーザーの認証を行う外部サービス
SL023	Python	3.12	● オープンソースのプログラミング言語のひとつ
SL024	Remix	-	● Web アプリケーション開発フレームワーク
SL025	Hono	-	● Web アプリケーション開発フレームワーク
SL026	AWS Location API	4.0	● 住所情報から緯度経度(WGS84)を入手するジオコーダーAPI
SL027	MapLibre	-	● EBPM Tool で利用する地図ライブラリ。2.5次元での表現などが可能

### 4-2-3. 開発機能の詳細要件

#### 【FN001】 ワークスペース管理機能

- 機能概要
  - CMS におけるデータ管理用ワークスペース管理機能を指す。
- フローチャート
  - なし
- データ仕様
  - 入力
    - なし
  - 出力
    - なし
- 機能詳細
  - システム管理者のみが CMS 上にデータ管理用のワークスペースを作成する。本システムでは、ワークスペースを1つのみ作成する。
  - 利用するソフトウェア
    - Re:Earth CMS
  - 利用するアルゴリズム
    - なし

#### 【FN002】 プロジェクト管理機能

- 機能概要
  - CMS におけるデータ管理用プロジェクト管理機能を指す。
- フローチャート
  - なし
- データ仕様
  - 入力
    - なし
  - 出力
    - なし
- 機能詳細
  - システム管理者のみが CMS 上にデータ管理用のプロジェクトを作成する。本システムでは、プロジェクトを1つのみ作成する。
  - 利用するソフトウェア
    - Re:Earth CMS

- 利用するアルゴリズム
  - なし

【FN003】 コンテンツ管理機能

● 機能概要

- 本機能では、D001、D002 の処理実行によって作成されるコンテンツ情報を管理する機能を提供する。本機能はコンテンツメニューから呼び出す。コンテンツの検索、削除、名称変更、編集が可能である。各コンテンツの編集・参照権限は、ログインユーザーの所属（原課レベル）によって制御する。
- コンテンツを選択し、オペレーターボタンから各オペレーター画面（構造化処理画面、データ結合前処理画面、テキストマッチング画面、空間結合画面、クロス集計画面、空間集計画面）に遷移し、実行が可能である。
- コンテンツの編集では、インサート、カラムの追加、削除、カラム名の編集、コンテンツ内のデータ編集、コピーを可能とする。
- コンテンツの管理パネルにて、公開ステータスの管理等を実施する。

● フローチャート

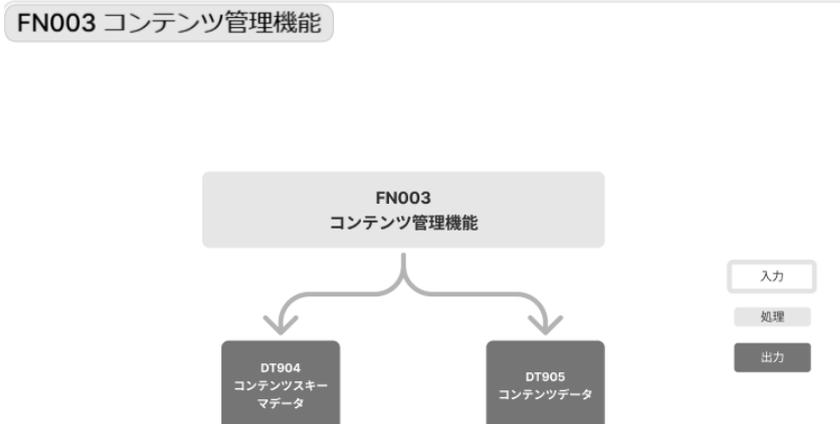


図 4-6 コンテンツ管理処理フロー

● データ仕様

- 入力
  - JSON 形式、GeoJSON 形式（D001、D002 で作成されたコンテンツ）
- 出力
  - 【DT904】 コンテンツスキーマデータ
  - 【DT905】 コンテンツデータ

- 機能詳細
  - コンテンツ管理機能では、システム内で作成された Processed Data、Defined Data の検索、削除、編集、コピー、ダウンロードが可能である。
  - コンテンツ一覧画面では、コンテンツのプレビューを確認することができ、画面下部のボタンからコンテンツの編集、コンテンツの削除、コンテンツダウンロード、オペレーターの起動、テンプレートの呼び出しが可能である。画面右部の管理パネルからは、コンテンツの管理情報を確認、更新することができる。コンテンツで保持しているメタデータは以下の通り。
    - 公開ステータス
      - オープンデータ、可視化、チャットのステータスを保持する。オープンデータ公開のオン/オフ、利用するユースケースの紐づけ、ベクトルデータの作成スイッチのオン/オフの操作によって更新される。
    - 「オープンデータ」ステータス
      - 本ステータスは、データ作成、データ公開のステータスによって管理する。
      - データ作成では、スイッチのオン/オフにより、オープンデータ向けの CSV、JSON、GeoJSON、タイル形式データの作成を行う。データ公開では、スイッチのオン/オフを行うことでオープンデータの公開可否を管理する。オープンデータの公開ステータスはシステム管理者のみ更新を行うことができる。
    - 「可視化」ステータス
      - 本ステータスは、可視化利用のユースケースを紐づけることによってオンになる。ユースケースは複数紐づけと削除が可能である。
    - 「チャット」ステータス
      - 「チャット」の利用ステータスは、ベクトルデータの作成スイッチのオン/オフによって管理する。ベクトルデータの作成スイッチを押下することで【FN033】ベクトルデータ作成機能（ベクトル解析）を起動し、【FN034】チャット生成機能で検索可能なベクトルデータを作成する。
  - 更新日
    - コンテンツの更新日を保持する。
  - 更新者
    - コンテンツの更新者を保持する。
  - 以下はコンテンツで設定するメタデータである。G 空間情報センターに公開される際には、リソースのメタデータとして公開する。

表 4-5 コンテンツ用メタデータ

ID	入力項目名称	デフォルト入力値	内容
1	管理 ID	ProjectLinks{ID となる数字}_{年度}	● G 空間情報センターで管理するデータセットを一意で判断するための ID。
2	タイトル（データセッ		● G 空間情報センターで管理

	ト名称)		するデータセットを一意で判断するための名称。
3	説明	<p>データの作成にあたっては、紙で保管された報告書をスキャンして PDF 化し、LINKS Veda を使用しました。</p> <p>LINKS Veda によるデータ作成の精度は紙資料の保存状態や記入状況に依存し、データの完全性や正確性を保証するものではありません。</p> <p><a href="https://www.mlit.go.jp/links/">https://www.mlit.go.jp/links/</a>本データは政府標準利用規約（第 2.0 版）に準拠しています。また、クリエイティブ・コモンズ・ライセンスの表示 4.0 国際と互換性があるとともに、利用者が Open Data Commons による ODC BY 又は ODbL での利用を希望する場合に、それを妨げるものではありません。</p>	<ul style="list-style-type: none"> <li>● G 空間情報センターに公開するデータに関する詳細説明。Project LINKS では、左記を固定とする。</li> </ul>
4	キーワード		<ul style="list-style-type: none"> <li>● G 空間情報センターでデータ検索を補助するためのキーワード情報。</li> </ul>
5	テーマ分類		<ul style="list-style-type: none"> <li>● G 空間情報センターでデータ検索を補助するためのタグで保持する情報。</li> </ul>
6	対象地域	日本全国	<ul style="list-style-type: none"> <li>● データ対象地域。Project LINKS では、左記を固定とする。</li> </ul>
7	対象期間		<ul style="list-style-type: none"> <li>● 公開するデータ対象期間を示す。</li> </ul>
8	提供者	国土交通省 総合政策局 情報政策課	<ul style="list-style-type: none"> <li>● データ提供者を示す。Project LINKS では、左記を固定とする。</li> </ul>
9	連絡先情報		<ul style="list-style-type: none"> <li>● データ提供者の連絡先情報を示す。</li> </ul>
10	作成者	国土交通省 総合政策局 情報政策課	<ul style="list-style-type: none"> <li>● データ作成者を示す。Project LINKS では、左記を固定とする。</li> </ul>

11	公開日		<ul style="list-style-type: none"> <li>● G 空間情報センターへのデータ公開日を示す。</li> </ul>
12	最終更新日		<ul style="list-style-type: none"> <li>● 作成されたデータセットの最終更新日を示す。</li> </ul>
13	更新頻度	不定期	<ul style="list-style-type: none"> <li>● 更新頻度を示す。Project LINKS では、左記を固定とする。</li> </ul>
14	言語	ja	<ul style="list-style-type: none"> <li>● データの言語を示す。Project LINKS では、左記を固定とする。</li> </ul>
15	公開範囲	制限付き公開	<ul style="list-style-type: none"> <li>● データの公開範囲を示す。Project LINKS では、左記を固定とする。</li> </ul>
16	公開条件		<ul style="list-style-type: none"> <li>● データの公開条件を示す。</li> </ul>
17	ライセンス	政府標準利用規約（第 2.0 版）	<ul style="list-style-type: none"> <li>● データのライセンスを示す。Project LINKS では、左記を固定とする。</li> </ul>
18	利用規約	<a href="https://www.mlit.go.jp/links/terms-of-use.html">https://www.mlit.go.jp/links/terms-of-use.html</a>	<ul style="list-style-type: none"> <li>● データの利用規約を示す。Project LINKS では、左記を固定とする。</li> </ul>
19	バージョン	ver1.0	<ul style="list-style-type: none"> <li>● データの利用規約を示す。データ更新の度にバージョンを上げて対応する。</li> <li>● データセットを変更した場合：ver 番号を更新する。（例：1.0→2.0）</li> <li>● データセット内のコンテンツを更新した場合：枝番を更新する。（例：1.0→1.1）</li> </ul>
20	タイプ	CSV、JSON、GeoJSON	<ul style="list-style-type: none"> <li>● G 空間情報センターに公開する際のリソースのファイル形式を示す。</li> </ul>
21	エンコーディング	UTF-8	<ul style="list-style-type: none"> <li>● G 空間情報センターに公開するファイルのエンコーディング。UTF-8 BOM 付のファイルで公開される。</li> </ul>
22	来歴情報		<ul style="list-style-type: none"> <li>● G 空間情報センターへの来</li> </ul>

			歴情報。空欄とする。
23	品質評価	正確性、完全性、一貫性	● LINKS Veda で作成したデータの品質評価を示す。 Project LINKS では、左記を固定とする。
24	データ品質（品質測定結果）	統計的な分析を行うのに十分な品質。ただし、データ型の統一などデータクレンジング処理を行っておりますが、紙資料の保存状態や記入状況によっては、正確な記載となっていない場合がございます。	● Project LINKS では、左記を固定とする。
25	制約	利用規約による	● Project LINKS では、左記を固定とする。
26	有償無償区分*	無償	● Project LINKS では、左記を固定とする。
27	災害時区分*	無償提供	● Project LINKS では、左記を固定とする。
28	価格情報	無償	● Project LINKS では、左記を固定とする。
29	使用許諾	Project LINKS の利用規約に従って、どなたでも、複製、公衆送信、翻訳・変形等の翻案等、自由に利用できます。 商用利用も可能です。 ( <a href="https://www.mlit.go.jp/links/terms-of-use.html">https://www.mlit.go.jp/links/terms-of-use.html</a> )	● Project LINKS では、左記を固定とする。
30	準拠する標準	<a href="https://www.w3.org/TR/vocab-dcat-3/">https://www.w3.org/TR/vocab-dcat-3/</a>	● Project LINKS では、左記を固定とする。
31	関連ドキュメント	<a href="https://www.mlit.go.jp/links/">https://www.mlit.go.jp/links/</a>	● Project LINKS では、左記を固定とする。

- コンテンツの編集では、コンテンツの詳細画面からコンテンツ名の変更、コンテンツ更新（インサート、カラム追加、カラム削除、コンテンツ内のセル単位の更新）を行う。更新結果は、上書き保存が可能である。
- コンテンツのインサート処理は、同一のスキーマ情報を保持するコンテンツに対してのみ実行可能である。コンテンツ定義が異なる場合、本処理は実行できない。
- セル単位の更新では、単一セルの選択、複数セルの選択、セル内の値の編集を可能とする。
- ダウンロード機能では、作成したコンテンツのスキーマ情報、およびコンテンツ情報の両方を

JSON 形式でダウンロード可能とする。

- 登録したコンテンツ情報は Re:Earth CMS へ連携する。
- 利用するソフトウェア
  - Re:Earth CMS
  - React/TypeScript
- 利用するアルゴリズム
  - なし

#### 【FN004】 アセット管理機能

- 機能概要

- 本機能では、アップロードされたファイルの単位（Raw Data）であるアセットを管理する機能を提供する。本機能はアセットメニューから呼び出す。アセットの追加、検索、削除が可能である。
- アセット管理のため、メタデータとして、ファイル名、ファイル形式（PDF/CSV/JSON/GeoJSON/SHP/ZIP/DOCX/XLSX）、ファイルサイズ、作成者、作成日時を保持する。
- アセットを選択することで、アセット一覧と同様の画面でアセット内ファイルのプレビューをできる。
- テンプレートボタンを押下することで、適用するテンプレートを選択した状態で後続処理となる【FN007】データ構造化処理（UI）、【FN010】データ結合前処理実行機能（UI）を起動できる。
- オペレーターボタンを押下することで、後続処理となる【FN007】データ構造化処理（UI）、【FN010】データ結合前処理実行機能（UI）を起動できる。

- フローチャート



図 4-7 アセット管理機能処理フロー

- データ仕様

- 入力

- 【DT001】 原票データ (PDF、PNG 形式)、原票データの ZIP 形式
- 【DT002】 テキストデータ (DOCX、XLSX 形式)、テキストデータの ZIP 形式
- 【DT003】 地図データ (SHP 形式、GeoJSON 形式)
- 【DT004】 テーブルデータ (CSV 形式)

- 出力

- なし

- 機能詳細

- アセット管理機能では、【FN006】データ登録機能でアップロードされたファイルをアセットとして管理、CMS 上に登録する機能を指す。ZIP 形式でファイルがアップロードされた場合は、ZIP 形式で1つのアセットの扱いとなる。
- アセット一覧より登録済みアセットを選択すると、同一画面からアセット詳細をプレビューできる。画面からアセットに含まれるファイルの1つ1つのファイル名、ファイルサイズ、メタデータ (オープンデータ)、作成者、作成日時、ファイルの中身のプレビューを確認することができる。
- アセット一覧画面より、後続処理となるオペレーターの起動を実行することができる。【FN007】データ構造化処理 (UI) は【DT001】原票データ (PDF、PNG 形式)、【DT002】テキストデータ (DOCX、XLSX 形式) のアセットのみが対象であり、【FN010】データ結合前処理実行機能 (UI) では、【DT003】地図データ、【DT004】テーブルデータが対象となる。
- 利用するソフトウェア
  - Re:Earth CMS
  - React/TypeScript

- 利用するアルゴリズム
  - なし

#### 【FN005】 アイテム管理機能

- 機能概要
  - 本機能は、CMS で管理するデータの最小単位で、コンテンツのレコードを管理する機能を提供する。Veda 内では、コンテンツ管理画面から参照可能である。
- フローチャート
  - なし
- データ仕様
  - 入力
    - JSON 形式、GeoJSON 形式 (D001、D002 で作成されたコンテンツ)
  - 出力
    - なし
- 機能詳細
  - **【FN003】** コンテンツ管理機能においてユーザーが実行したコンテンツ更新の結果をレコード単位で CMS に反映する。
  - 利用するソフトウェア
    - Re:Earth CMS
  - 利用するアルゴリズム
    - なし

#### 【FN006】 データ登録機能

- 機能概要
  - 本機能は、アセットメニューから呼び出す。任意の静的ファイルをローカルから選択し、アセットとして新規アップロードする機能を提供する。
  - アップロードファイルは、アップロードファイルのデータ形式は、原票データ (PDF、PNG)、パッケージ (ZIP)、テキストデータ (DOCX、XLSX)、テーブルデータ (CSV)、地図データ (SHP、CSV、GeoJSON) を許容する。
  - データ登録時に ZIP でファイルがアップロードされた場合には、登録時に解凍処理を自動で実行する。
  - 登録したアセット情報は Re:Earth CMS へ連携する。

● フローチャート



図 4-8 データ登録機能処理フロー

● データ仕様

➢ 入力

- 【DT001】 原票データ (PDF、PNG 形式)
- 【DT002】 テキストデータ (DOCX、XLSX 形式)
- 【DT003】 地図データ (SHP、GeoJSON)
- 【DT004】 テーブルデータ (CSV 形式)

➢ 出力

- なし

● 機能詳細

- ユーザーはアセットメニューのアップロードボタンを押下し、アセットをアップロードする。アップロード後に削除する場合には、同画面のアセット削除ボタンを押下し削除する。
- ZIP でファイルを圧縮する際には、同様の種類のファイルのみ許容する。
- 【DT003】 地図データ (SHP) は、.SHP：図形の情報を格納する主なファイル、.SHX：図形のインデックス情報を格納するファイル、.DBF：図形の属性情報を格納するテーブルで構成される。アップロードする際には、これらは ZIP で1 アセットにまとめた上でアップロードをする。
- 利用するライブラリ
  - 【SL009】 Pandas
- 利用するアルゴリズム
  - なし

【FN007】データ構造化処理実行機能(UI)

● 機能概要

- 本機能では、非構造化データ（原票データ、テキストデータ）に対し、構造化データ化（コンテンツ作成）する処理を実行する画面を提供する。ユーザーは自身がアップロードしたアセット選択し、プレビュー画面で参照しながら、抽出するためのプロンプト情報を入力項目に則って指定する。なお、さらに詳細な条件文の記入やテンプレートを利用しないプロンプト入力に慣れているユーザー向けのUIとして自由文の記述項目を設ける。
- 事前にテンプレートとして定義したプロンプトを利用する場合には、テンプレート選択ボタンから呼び出すことができ、新規に入力した項目をテンプレートとして保存する場合には保存テンプレートとして保存ボタンを押下する。
- データ構造化に加えて、出典名の付与を行う。ユーザーは構造化対象のカラム名、自由入力のテキストを組み合わせることで出典名の作成を行う。本機能で付与した出典名は、【FN034】チャット生成機能のレスポンスで利用する。

● フローチャート

- なし

● データ仕様

- 入力
  - 【FN020\_2】データ構造化処理に準ずる
- 出力
  - 【FN020\_2】データ構造化処理に準ずる

● 機能詳細

- 【FN007】データ構造化処理実行機能(UI)へは、アセット登録直後の実行、もしくはオペレーター単体処理での実行が想定されることから、アセットメニュー、もしくはホーム画面のオペレーターメニューから遷移する。
- データ構造化対象となるアセットは、【DT001】原票データ（PDF、PNG形式）、【DT002】テキストデータ（DOCX、XLSX形式）である。アセット選択ボタンよりアセットを選択し、データ構造化処理を実行する。
- アセット選択後、アセットビューワーからファイルの中身をプレビューで参照することができる。ユーザーは設定タブより構造化に必要な以下のパラメータを入力し、構造化処理を実行する。本パラメータをもとにプロンプトがシステム内で作成され、コンテンツデータが生成される。構造化するカラムを追加する場合には、カラムを追加ボタンよりカラムを追加する。（\*は必須項目を表す。）
  - パラメータ
    - カラム名\*：作成コンテンツのカラム名。生成されるコンテンツのスキーマ情報となる。
    - タイプ\*：作成コンテンツカラムのデータ型（String、Number、Boolean）
    - ポジション：抽出元のアセット内で、抽出する値が存在する章立て等の情報
    - 単位：円、千円、百万円など数値データを抽出する値の単位

- キーワード：抽出元のアセット内で、抽出する値を特定する際に参考となる情報。近くに存在するテキスト情報
- 自由文：ユーザーが任意に入力できるプロンプトエリア
- 処理実行時のパラメータ情報は、事前にテンプレートとして登録しているものからテンプレート選択ボタンを押下し、呼び出すことができる。処理実行時に新規にテンプレート保存したい場合にはテンプレートとして保存ボタンを押下し、保存することができる。
- データ構造化処理と同タイミングで、出典名の作成を必須で行う。出典名の作成に必要なパラメータは以下の通り。
  - パラメータ
    - テキスト/カラムリスト：出典名生成に必要な情報をテキストによる自由入力、もしくはカラム名による選択をする。
    - 入力項目：テキストを選択した場合、自由入力のテキスト、カラムを選択した場合はカラム名を入力する。カラム名は、構造化実行時に生成するカラム名とする。
- 実行ボタンより【FN020\_2】データ構造化処理を実行する。実行後に処理結果タブに自動で遷移し、実行ログを確認できる。実行ログは、【FN008】処理状況確認機能からも確認可能である。
- アウトプットとしては、1 ファイル 1 レコードとしてコンテンツが作成される。作成されたコンテンツをアウトプットフィールドから確認し(【FN009】構造データプレビュー機能(プレビュー))、問題がなければ保存を押下し保存する。内容に問題がある場合には、パラメータを再設定、処理を再実行し、期待する結果が得られるまで本処理を繰り返す。
- データ構造化処理結果として正しく抽出されたかを確認するための指標として複数の LLM の判定結果によって付与される一致度によってアウトプットのカラムを色づけする。一致度は 0.0~1.0 の範囲で抽出結果の正しさを示し、1.0 が最も正しい値を示す。ユーザーはカラムの色から抽出結果の正確性を判断する。(一致度の設定ロジックは、【FN020\_2】データ構造化処理にて詳細を記載する)
  - White : 1.0
  - DustyRed02 : 0.9~0.7
  - DustyRed04 : 0.6~0.4
  - DustyRed06 : 0.3~0.0
- 登録したコンテンツ情報は Re:Earth CMS へ連携する。
- 利用するソフトウェア
  - React/TypeScript
- 利用するアルゴリズム
  - なし

【FN008】 処理状況確認機能

- 機能概要
  - 本機能では、D001、D002 のオペレーターごとの処理ステータス（処理待/処理中/処理済/保存済/エラー）を実行ユーザーごとに一覧で確認できる機能を提供する。
  - 処理状況を押下することで、実行したオペレーターのログ詳細に遷移することができる。
- フローチャート
  - なし
- データ仕様
  - 入力
    - なし
  - 出力
    - なし
- 機能詳細
  - 処理状況一覧では、ユーザー自身が発行した処理結果のみ確認/編集できる。ファイル名、ファイル形式、処理内容、ステータス、実行者、実行日を一覧で確認できる。処理状況を押下することで、同一画面内で処理結果を確認できる。
  - 処理状況を選択し、開くボタンを押下することで、処理実行したオペレーターの詳細を開くことができる。
  - 処理ステータスの詳細は以下の通り
    - 処理待：【FN035】テンプレート管理の一部であるワークフロー処理内で、複数のオペレーターの処理を連続して実行する場合、前段の処理の完了を待つ状態を指す。
    - 処理中：各オペレーターで実行ボタンが押下され、処理が完了していない状態を指す。
    - 処理済：各オペレーターで実行ボタンが押下され、処理が完了した状態を指す。
    - 保存済：各オペレーターで処理が完了し、作成されたコンテンツに名前をつけて保存した状態を指す。
    - エラー：各オペレーターで実行ボタンが押下され、処理の結果何らかの不具合が発生している状態を指す。
  - 利用するソフトウェア
    - React/TypeScript
  - 利用するアルゴリズム
    - なし

【FN009】 構造データプレビュー機能（プレビュー）

- 機能概要
  - D001 処理実行後の妥当性検証として、構造化されたデータプレビューをテーブル形式で行う機能である。
- フローチャート
  - なし
- データ仕様
  - 入力
    - なし
  - 出力
    - なし
- 機能詳細
  - データプレビュー時点のデータは、すでに CMS に保存された状態のものであり、ユーザーは画面から任意の名称を入力して保存する。
  - ユーザーが任意の名称を入力して保存したタイミングで、【FN008】 処理状況確認機能におけるステータスを「保存済」に更新する。
  - プレビュー画面でのコンテンツの直接編集を行うことは許容しない。
  - 利用するソフトウェア
    - Re:Earth CMS
    - React/TypeScript
  - 利用するアルゴリズム
    - なし

【FN010】 データ結合前処理実行機能（UI）

- 機能概要
  - 本機能では、D001 で構造化されたデータ、地図データ、テーブルデータに対し、【FN021】 データ結合前処理（データクレンジング）を実行する。ユーザーはデータクレンジングに必要な条件をオプションとして設定し、実行可能とする。
  - 本機能では、地図データでないコンテンツに対し、【FN022】 データ結合前処理（ジオコーディング）を実行する。ユーザーはジオコーディングを行う住所を保持するカラムを指定し、対象のカラムに対してジオコーディングを実行する。処理の結果、新規に緯度経度の情報とポイントのジオメトリが追加される。
  - カラムの秘匿化処理を行う。秘匿化処理には、ランク化、偏差値化、コードの新規採番、住所カラムの秘匿化の処理が含まれる。
  - 資料名の付与を行う。任意のテキストを入力することで資料名の作成を行う。本機能で付与した資料名は、【FN034】 チャット生成機能のレスポンスで利用する。

- 地図データの場合は、バックエンドで【FN003】データ結合前処理（座標系統一）処理を実行し座標系を WGS84 に設定する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - 【FN021】データ結合前処理（データクレンジング）、【FN022】データ結合前処理（ジオコーディング）、【FN023】データ結合前処理（座標系統一）に準ずる。
  - 出力
    - 【FN021】データ結合前処理（データクレンジング）、【FN022】データ結合前処理（ジオコーディング）、【FN023】データ結合前処理（座標系統一）に準ずる。
- 機能詳細
  - 本機能では、【DT003】地図データ、【DT004】テーブルデータ、【DT204】構造データ化済テキストデータに対し、データクレンジング、ジオコーディング、地図データに対し座標系統一の処理を実行する。D001 を介さないデータは構造化未実行の状態であるため、アセットを選択し実行する。アセットとコンテンツの選択の違いは以下の通り。
    - アセット選択：【DT003】地図データ、【DT004】テーブルデータ
    - コンテンツ選択：【DT204】構造データ化済テキストデータ
  - 上記に則りアセット、もしくはコンテンツを選択後、本機能の処理オプションとして、資料名作成、データクレンジング、ジオコーディング、秘匿化処理を選択する。各処理オプションの処理内容は以下の通り。
    - 資料名作成：ユーザー任意のテキストにて資料名を入力する。コンテンツ作成時、資料名カラムは必須とする。本機能で付与した資料名は、【FN034】チャット生成機能のレスポンスで利用する。
    - データクレンジング：コンテンツ内の値更新として、以下のクレンジングオプションを準備する。以下のオプション以外のクレンジング処理が必要な場合には、プロンプト入力内に自由文で入力しデータクレンジングを実行する。
      - 削除：カラムを指定し、削除条件に則って削除する。
        - 削除条件：一致する文字列、カンマ、全角スペース、半角スペース、タブ
      - 置換：カラムを指定し、特定の文字列と一致する文字列を置換する。
      - 表記ゆれの正規化：住所カラムを指定し住居表示又は地番住所の表記を正規化する。
    - ジオコーディング：【DT202】正規化テーブルデータ、【DT204】構造データ化済テキストデータに対し、指定した住所カラムを元に【FN022】データ結合前処理（ジオコーディング）を起動しジオコーディングを行う。ジオコーディングを行い、緯度経度のカラムとジオメトリ（ポイント）を新規に追加する。カラム名は「元のカラム名\_lat、「元のカラム名\_lon」とする。
    - 秘匿化処理：オープンデータ用、EBPM Tools 利用用にランク化カラム、偏差値化カラム、新規 ID カラム、秘匿化住所カラムの付与を行う。

- ランク化カラム付与：ランク化対象のカラムを指定し、新規にランク化されたカラムを追加する。
  - 偏差値化カラム付与：偏差値化対象のカラムを指定し、新規に偏差値化されたカラムを追加する。
  - 新規 ID カラム付与：個人や企業等を特定できる ID のカラムを指定し、新規にシステム内で採番した ID カラムを追加する。
  - 秘匿化住所カラム付与：住所を都道府県市町村のレベルに秘匿化したカラムを追加する。
- 本処理内では、【DT201】正規化地図データに対して【FN023】データ結合前処理（座標系統一）をシステム内で自動実行し座標系を WGS84 に統一する。
  - 【DT003】地図データ、【DT004】テーブルデータの正規化処理をシステム内で自動実行し、【DT201】正規化地図データ、【DT202】正規化テーブルデータを作成する。
  - 登録したコンテンツ情報は Re:Earth CMS へ連携する。
  - 利用するソフトウェア
    - React/TypeScript
  - 利用するアルゴリズム
    - なし

#### 【FN011】データ結合実行機能（テキストマッチング）（UI）

- 機能概要
  - 本機能では、結合元のコンテンツ（メインコンテンツ）、結合先コンテンツと結合するカラムをキーカラムと指定し結合する機能を提供する。
  - 結合コンテンツでは、結合後に保持するカラムを結合カラムとして指定することができる。結合カラムの指定のない場合は、全カラム結合となる。なお、結合コンテンツは 1 つのみ指定することができ、複数のコンテンツを結合する場合は本処理を繰り返し実行する。
  - 結合条件に一致しなかったデータは後続の処理対象からは除外される。
- フローチャート
  - なし
- データ仕様
  - 入力
    - 【FN024】データ結合処理（テキストマッチング）に準ずる。
  - 出力
    - 【FN024】データ結合処理（テキストマッチング）に準ずる。
- 機能詳細
  - 本機能では、【DT201】正規化地図データ、【DT202】正規化テーブルデータ、【DT206】データ結合済構造化データ、【DT207】空間結合済データ、【DT208】地域集計済構造化データ、【DT209】クロス集計済構造化データを対象にコンテンツ同士の結合処理を行う。ユーザーはコンテンツ選択から結合元とするメインコンテンツと結合先とするコンテンツ選択し、それぞれ結合するカラ

ムをキーカラムとして指定し【FN024】データ結合処理（テキストマッチング）を実行する。なお、結合コンテンツは1つのみ指定することができるため、複数を選択しようとした場合は警告が出る。

- 処理の結果作成された結合コンテンツは、新規コンテンツとして登録され、Re:Earth CMSへ連携する。
- 利用するソフトウェア
  - React/TypeScript
- 利用するアルゴリズム
  - なし

#### 【FN012】データ集計実行機能（クロス集計）（UI）

- 機能概要
  - 本機能では、選択したコンテンツに対し、集計キーごとに指定カラムの合計・平均値、およびカウント値を算出する機能を提供する。ユーザーが任意のカラムを指定し、指定したカラムをキーに集計を行う。
- フローチャート
  - なし
- データ仕様
  - 入力
    - 【FN025】データ集計処理（クロス集計）に準ずる。
  - 出力
    - 【FN025】データ集計処理（クロス集計）に準ずる。
- 機能詳細
  - 本機能では、【DT201】正規化地図データ、【DT202】正規化テーブルデータ、【DT206】データ結合済構造化データ、【DT207】空間結合済データ、【DT208】地域集計済構造化データ、【DT209】クロス集計済構造化データを対象に単一コンテンツのクロス集計処理を行う。
  - ユーザーはコンテンツ選択から集計するコンテンツを選択し、集計単位と集計対象と指定し、【FN025】データ集計処理（クロス集計）を実行する。
  - 集計単位、集計対象はコンテンツのカラムから複数指定可能であり、集計対象の要約統計量（合計、平均、カウント）を集計できる。
    - 合計/平均：数値のカラムを対象とする。
    - カウント：テキストのカラムを対象とする。
  - 処理の結果作成された集計コンテンツは、新規コンテンツとして登録され、Re:Earth CMSへ連携する。
  - 利用するソフトウェア
    - React/TypeScript
  - 利用するアルゴリズム

- なし

#### 【FN013】データ結合実行機能（空間結合）（UI）

- 機能概要

- 本機能では、地図データの結合を実施する。メイン地図データ、結合地図データ、結合方式を選択しデータを結合する。結合方式には交差結合、最近傍結合が存在し、ユーザーは任意の結合方式を選択し実行することができる。最近傍結合の場合には、検索上限範囲を指定する。
- 【FN024】データ結合実行機能（テキストマッチング）と同様、結合条件に一致しなかったデータは後続の処理対象からは除外される。なお、結合コンテンツは1つのみ指定することができ、複数のコンテンツを結合する場合は本処理を繰り返し実行する。

- フローチャート

- なし

- データ仕様

- 入力
  - 【FN026】データ結合機能（空間結合処理）に準ずる。
- 出力
  - 【FN026】データ結合機能（空間結合処理）に準ずる。

- 機能詳細

- 本機能では、【DT201】正規化地図データ、【DT207】空間結合済データ、【DT208】地域集計済構造化データに対し、【FN026】データ結合処理（空間結合）を行う。
- ユーザーはコンテンツ選択から結合元とするコンテンツをメインコンテンツと結合するコンテンツ、結合方式を選択し【FN026】データ結合機能（空間結合処理）を実行する。なお、結合コンテンツは1つのみ指定することができるため、複数を選択しようとした場合は警告が出る。
- 結合方式には交差結合、最近傍結合を選択でき、最近傍結合の場合には検索上限範囲もあわせて指定する。
- 【FN024】データ結合処理（テキストマッチング）と同様、結合条件に一致しなかったデータは後続の処理対象からは除外される。
- 処理の結果作成された結合コンテンツは、新規コンテンツとして登録され、Re:Earth CMSへ連携する。
- 利用するソフトウェア
  - React/TypeScript
- 利用するアルゴリズム
  - なし

#### 【FN014】データ集計実行機能（空間集計）（UI）

- 機能概要
  - 本機能では、地図データの集計を実施する。メイン地図データ、集計地図データ、集計単位、集計対象を指定し、カラムの合計・平均値、およびカウント値を算出する機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 【FN027】データ集計機能（空間集計処理）に準ずる。
  - 出力
    - 【FN027】データ集計機能（空間集計処理）に準ずる。
- 機能詳細
  - 本機能は、【DT201】正規化地図データ、【DT207】空間結合済データ、【DT208】地域集計済構造化データを対象に空間集計を行う。
  - ユーザーはコンテンツ選択から集計する地図データ（ポリゴンデータとポイントデータ）、集計単位と集計対象と指定し、【FN027】データ集計機能（空間集計処理）を実行する。
  - 集計単位は、メイン地図データのカラムから選択し、集計対象は集計地図データのカラムから選択する。集計単位、集計対象はコンテンツのカラムから複数指定可能であり、集計対象の要約統計量（合計、平均、カウント）を集計できる。
    - 合計/平均：数値のカラムを対象とする。
    - カウント：テキストのカラムを対象とする。
  - 処理の結果作成された集計コンテンツは、新規コンテンツとして登録され、Re:Earth CMS へ連携する。
  - 利用するソフトウェア
    - React/TypeScript
  - 利用するアルゴリズム
    - なし

#### 【FN015】データ検索機能

- 機能概要
  - 【FN003】コンテンツ管理機能、【FN004】アセット管理機能、【FN035】テンプレート管理機能のそれぞれにおいて、一覧から検索可能な機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 入力

- なし
- 出力
  - なし
- 機能詳細
  - 本機能は、【FN003】コンテンツ管理機能、【FN004】アセット管理機能、【FN035】テンプレート管理機能に登録された情報を名称およびシステム内メタデータにより検索機能を提供する。  
【FN035】テンプレート管理機能は、テンプレート名称での検索のみとする。
  - アセット、コンテンツでは、それぞれ以下の情報をシステム内メタデータとして保持する。
    - アセット：ファイル名、ファイル形式、ファイルサイズ、更新者、更新日時
    - コンテンツ：ファイル名、公開ステータス（オープンデータ/可視化/チャット）、ファイルサイズ、作成者、作成日、更新者、更新日時
  - 利用するソフトウェア
    - React/TypeScript
  - 利用するアルゴリズム
    - なし

#### 【FN016】データダウンロード機能

- 機能概要
  - 本機能では、スキーマ定義、コンテンツ情報（正規化ローデータ、最終構造化データ）を【FN003】コンテンツ管理機能からダウンロードできる機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - なし
  - 出力
    - JSON形式、GeoJSON形式
- 機能詳細
  - 本機能では、正規化ローデータ（【DT206】テキスト結合済構造化データ、【DT207】クロス集計済構造化データ、【DT208】空間結合済データ、【DT209】地域集計済構造化データ、最終構造化データ（【DT901】最終構造化データ）をコンテンツメニューから選択し、ダウンロードボタンを押下することで、スキーマ情報、コンテンツ情報をJSON形式、GeoJSON形式ファイルでダウンロードする。
  - 利用するソフトウェア
    - React/TypeScript
  - 利用するアルゴリズム
    - なし

【FN017】 結合・集計データプレビュー機能 (UI)

- 機能概要
  - D002 処理実行後の妥当性検証として、結合・集計されたデータプレビューをテーブル形式、地図データ形式で行う機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - なし
  - 出力
    - なし
- 機能詳細
  - データプレビュー時点のデータは、すでに CMS に保存された状態のものであり、ユーザーは画面から任意の名称を入力して保存する。
  - ユーザーが任意の名称を入力して保存したタイミングで、【FN008】 処理状況確認機能におけるステータスを「保存済」に更新する。
  - プレビュー画面でのコンテンツの直接編集は行うことは許容しない。
  - プレビュー画面で確認可能な地図データは、2次元地図のみ対応する。
  - 利用するソフトウェア
    - React/TypeScript
  - 利用するアルゴリズム
    - なし

【FN018】 データ承認機能

- 機能概要
  - 最終構造化データに対し、システム管理者によるデータ承認を行う機能を提供する。本機能は【FN019】 データ配信機能における G 空間情報センターへの公開処理と同義とする。
- フローチャート
  - なし
- データ仕様
  - 入力
    - なし
  - 出力
    - なし

- 機能詳細
  - データ公開をする場合、システム管理者はコンテンツ一覧から「オープンデータ」ステータスがオンになっていないコンテンツを検索し、「オープンデータ」公開ステータスをオンにする。本ステータスは G 空間情報センターのリソース公開を制御する。なお、本ステータスをオンにしても即座に G 空間情報センターに公開されるものではない。
  - 利用するソフトウェア
    - React/TypeScript
  - 利用するアルゴリズム
    - なし

## 【FN019】データ配信機能

- 機能概要
  - データ配信には、①オープンデータ向けの配信（G 空間情報センターへの公開）、②EBPM Tools 向けの配信の 2 種類が存在する。
  - ①オープンデータ向けの配信については、システム管理者によってのみ処理可能とする。②EBPM Tools 向けの配信については一般ユーザーにて可能とする。これらの公開処理は【FN003】コンテンツ管理機能の管理パネル、【FN036】データセット管理機能の公開ステータス制御から実施するものとし、G 空間情報センターへは【FN036】データセット管理機能の公開ステータスがオンになったタイミングで公開されるものとする。
  - データ配信制御に関する処理権限は以下の通り。すでに EBPM Tools やオープンデータ向け G 空間情報センターに配信済みのコンテンツ、データセットについては、一度ステータスをオフにしない限り削除できないものとする。

表 4-6 コンテンツ管理/データセット管理の処理権限

処理権限	システム管理者	一般ユーザー
コンテンツ：EBPM Tools 公開	✓	✓
コンテンツ：オープンデータ公開	オン	-
データセット：EBPM Tools 公開/非公開	✓	✓
データセット：オープンデータ公開/非公開	オン	-

表 4-7 コンテンツ/データセットの配信制御

配信状況の制御	公開ステータス	G 空間情報センター	
		リソース	データセット
コンテンツ	オフ	非公開	非公開
データセット	オフ		
コンテンツ	オン	非公開	非公開

データセット	オフ		
コンテンツ	オフ	非公開	公開
データセット	オン		
コンテンツ	オン	公開	公開
データセット	オン		

- 配信処理はデータ作成処理時に公開 URL を発行する機能を提供する。外部よりこれらを利用する場合には、本公開 URL を利用してデータ利用を行う。
- フローチャート
  - なし
- データ仕様
  - 入力
    - なし
  - 出力
    - テーブルデータは CSV 形式、JSON 形式、JSONL 形式、地図データの場合は GeoJSON 形式、GeoJSONL 形式、MVT 形式での配信とする。
    - オープンデータ公開、EBPM Tools 向けの座標系およびファイル形式は以下の通り。
      - 座標系：WGS84
        - フォーマット：GeoJSON RFC7946
      - 座標系：JGD2011
        - フォーマット：GeoJSON 2008
- 機能詳細
  - 公開 URL はコンテンツ単位で発行するものとする。
  - 公開データは上記の出力形式に則り作成する。
  - 利用するソフトウェア
    - React/TypeScript
  - 利用するアルゴリズム
    - なし

## 【FN020\_1】 OCR 処理

- 機能概要

- 【DT001】原票データを【FN020\_2】データ構造化処理を適用できるように、OCR処理によりテキストデータ化する機能。読み取り対象として、画像内にある活字・手書き文字の両方を対象とするとともに、チェックボックス等の文字列以外も読み取り、テキストデータ化を行う。

- フローチャート

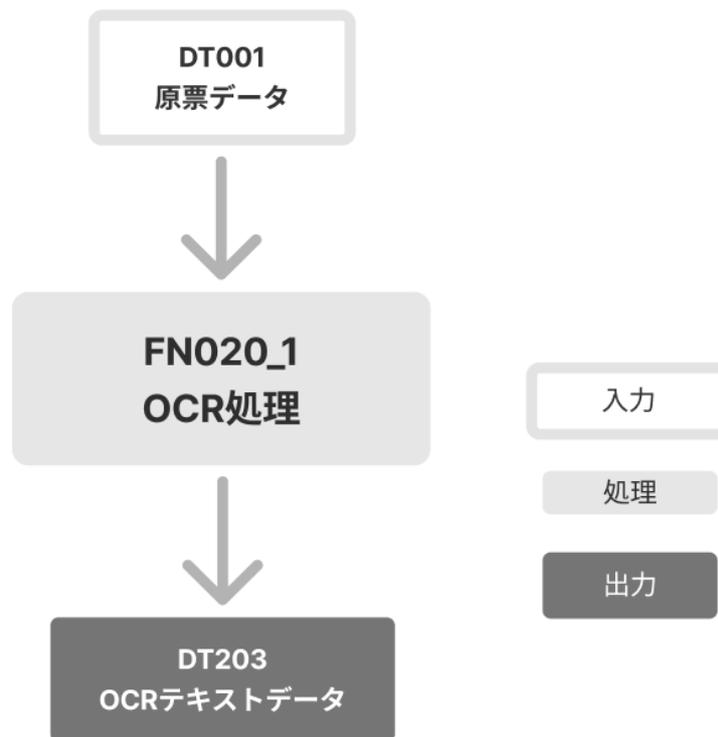


図 4-9 OCR 処理フロー

- データ仕様

- 入力

- ◇ 【DT001】原票データ

- 形式

- PDF 形式、PNG 形式

- データ詳細

- アセットとして登録された PDF 形式や PNG 形式のデータ。
- OCR 処理を実行しテキスト化した後、【FN020\_2】データ構造化処理を行う。
- 最終的にデータの可視化を行う際に必要となる、名称、座標、日時などの情報を含む。

➤ 出力

◇ 【D203】 OCR 処理済テキストデータ

● 形式

➤ JSON 形式

● データ詳細

➤ 【DT001】 原票データに対し、OCR 処理を行いテキストデータ化したデータ。

➤ 原票データ内で画像データ化されていた文字情報を含む。

● 機能詳細

➤ OCR 処理

◇ 画像データである【DT001】 原票データに対し、【SL011】 Tesseract もしくは【SL003】 AWS Bedrock API (OCR 精度向上のため、Azure OCR API もユーザーによる選択で利用可) を用いて OCR 処理を行いテキストデータ化する。

◇ 画像の前処理 (リサイズ、回転、トリミング、ノイズの除去等) により、文字認識に適した状態に加工後、画像内の文字のセグメンテーションを実施。

◇ 抽出結果から抽出した文字形状やサイズなどの特徴を取得し、機械学習により文字認識モデルから文字を識別。

◇ 画像内にある活字・手書き文字の両方を対象とするとともに、チェックボックス等の文字列以外も読み取り、テキストデータ化を行う。

➤ 利用するライブラリ

◇ 【SL003】 AWS Bedrock API / Azure OCR API

◇ 【SL004】 Langchain

◇ 【SL009】 Pandas

◇ 【SL011】 Tesseract

【FN020\_2】データ構造化処理

● 機能概要

- 【DT002】テキストデータを、構造データとして既存プログラミング言語や既存ソフトで読み取ることができるように、【DT004】テーブルデータ形式に変換を行う機能。
- テーブルデータの生成にあたり、ユーザーは出力コンテンツを定義・入力することで、出力コンテンツにしたがってテーブルデータが生成される。

● フローチャート

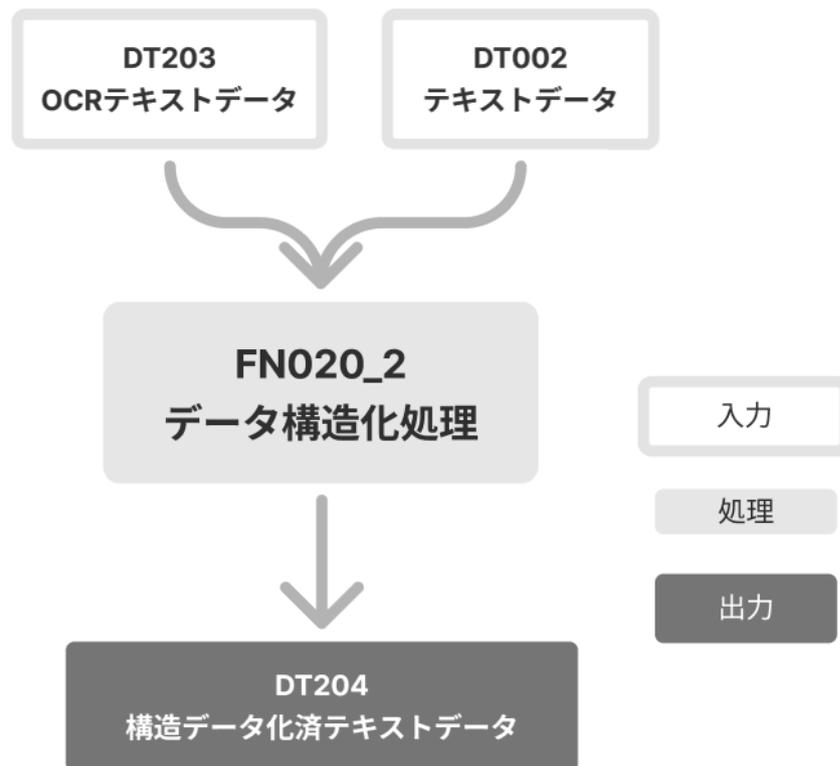


図 4-10 データ構造化処理フロー

➤ 入力

◇ 【DT002】テキストデータ

- 形式
  - DOCX 形式、XLSX 形式
- データ詳細
  - アセットとして登録された DOCX 形式、XLSX 形式のデータ。
  - 【AL01】LLM や【AL02】を用いて構造データ化を行う。

- XLSX 形式ですでに構造データ化されているデータもテキストデータとして入力される。
- ◇ 【DT203】 OCR 処理済テキストデータ
  - 形式
    - JSON 形式
  - データ詳細
    - 【DT001】 原票データに対し、【FN020\_1】 OCR 処理を行いテキストデータ化したデータ。
    - 【DT001】 原票データ内で画像データ化されていた文字情報を含む。
- 出力
  - ◇ 【DT204】 構造データ化済テキストデータ
    - 形式
      - JSON 形式
    - データ詳細
      - 【DT203】 OCR 処理済テキストデータおよび 【DT002】 テキストデータを、構造データ化処理をしたデータ。
- 機能詳細
  - データ構造化処理
    - ◇ 構造データ化されていないテキストデータとして、【DT203】 OCR 処理済テキストデータ・【DT002】 テキストデータを【AL01】 LLM ならびに【AL02】 構造データ生成モデルを用いて構造データ化させる機能。
    - ◇ ユーザーが生成したい構造データ構成（カラム構成、データ型等の詳細条件）を指定すると、LLM に投入するプロンプトを自動生成し、プロンプトに従い、構造データを自動生成する。生成にあたり、【AL02】 構造データ生成モデルに基づき、生成結果が入力データとの一致度合いを示す類似度を出力する。
    - ◇ プロンプトの入力補助として、【FN035】 テンプレート管理機能で事前に登録したコンテンツテンプレートを利用し、所定のカラム構成、データ型、単位の調整などを行うことができる。
    - ◇ さらに詳細な指定（住所のハイフンの有無、氏名の中の半角スペースの有無など）が必要な場合には、自由文を入力することでプロンプトを制御・拡張する。
    - ◇ 出力時にはハルシネーションの検出に向けて、【AL02】 構造データ生成モデルにより信頼度 confidence を出力する。
    - ◇ チャット機能で出典情報を閲覧できるようにする為、構造化処理の段階でユーザーが命名したい情報を入力すると、構造化処理によって生成されたカラム名と結合し、出典名をカラムとして登録できる。
  - 利用するライブラリ
    - ◇ 【SL003】 AWS Bedrock API
    - ◇ 【SL004】 Langchain
    - ◇ 【SL005】 Faiss

- ◇ 【SL009】 Pandas
- 利用するアルゴリズム
  - ◇ 【AL01】 LLM
  - ◇ 【AL02】 構造データ生成モデル

【FN020\_3】ベクトル解析

- 機能概要
  - 【FN034】チャット生成機能において、ユーザーが問い合わせ内容へ【AL04】チャット生成により自動返答ができるように、【DT901】最終構造化データをベクトル化処理する機能。
- フローチャート

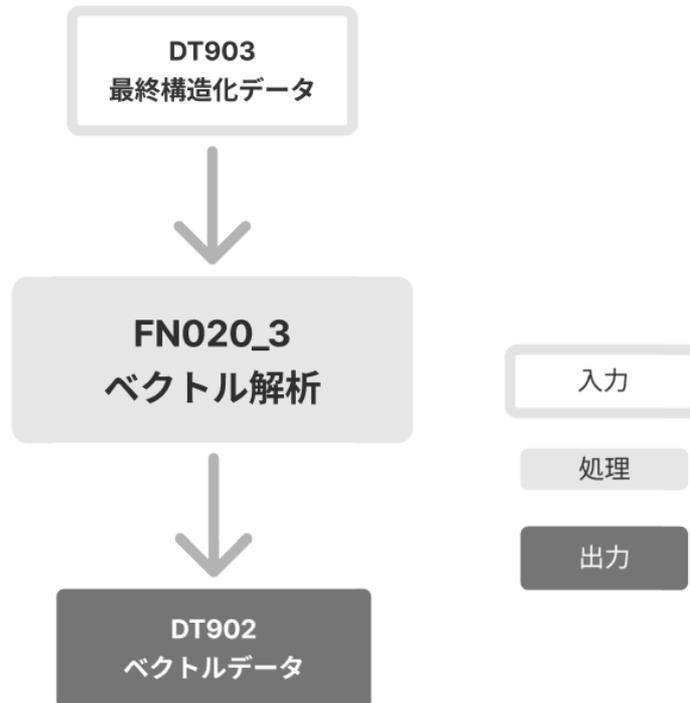


図 4-11 ベクトル解析処理フロー

- データ仕様
  - 入力
    - ◇ 【DT901】最終構造化データ
      - 内容
        - 【DT206】テキスト結合済構造化データ、【DT207】クロス集計済構造化データ、【DT208】空間結合済データ、【DT209】地域集計済構造化データのうち、最終的に出力されるデータ。
      - 形式
        - JSON 形式
        - GeoJSON 形式
      - データ詳細

- D002 正規化ローデータの集計・加工機能の処理を完了し、【FN018】にてデータ承認を行うことができる状態のデータ。
- 出力
  - ◇ 【DT902】 ベクトルデータ
    - 内容
      - 【DT901】 最終構造化データをベクトル解析によって、高次元空間上における数値ベクトルに変換したデータ。意味的に近い単語は互いに近い位置に配置され、遠い単語は遠くに配置される。
      - 【AL01】 LLM において、本データによって入力テキストの意味と文脈を理解し、適切な応答を生成可能とするために作成される。
    - 形式
      - PKL 形式
    - データ詳細
      - 【DT901】 最終構造化データを、ベクトル解析したデータ。
- 機能詳細
  - ベクトル解析
    - ◇ 【AL03】 ベクトル解析により、【AL01】 LLM が入力テキストの意味と文脈を理解できるように、テキストデータとなる【DT901】 最終構造化データを高次元空間上における数値ベクトルに変換する。
    - ◇ 生成されたベクトルデータは Python オブジェクトとして PKL 形式にて出力され、【FN032】 チャット機能の処理において利用される。
  - 利用するライブラリ
    - ◇ 【SL003】 AWS Bedrock API
    - ◇ 【SL004】 Langchain
    - ◇ 【SL005】 Faiss
  - 利用するアルゴリズム
    - ◇ 【AL01】 LLM(Large Language Model)

#### 【FN020\_4】 データ構造化前サジェッション機能

- 機能概要
  - 【FN020\_2】 データ構造化処理に必要な出力スキーマの候補を提案し、ユーザーがデータ構造化処理の入力を支援する機能。

- フローチャート

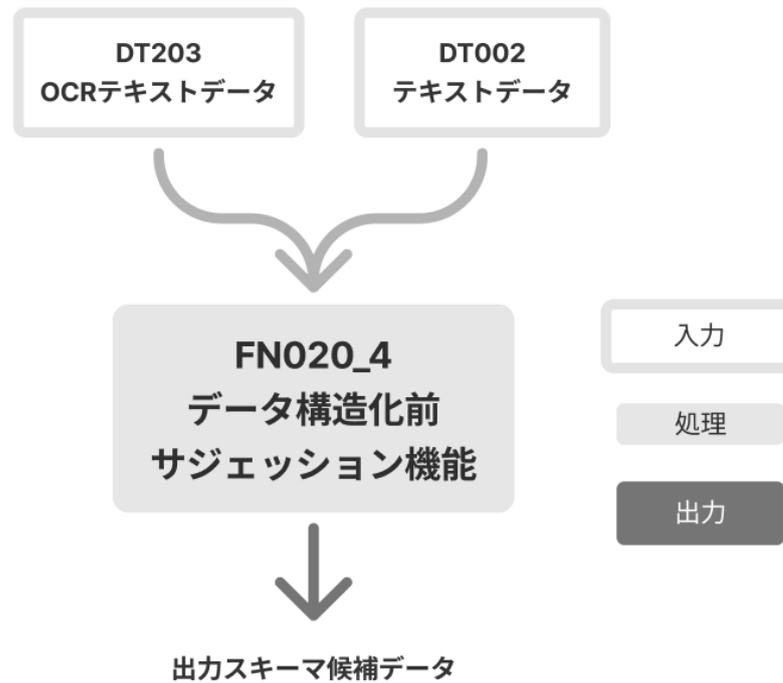


図 4-12 データ構造化前サジェッション処理フロー

- データ仕様

- 入力

- ◇ 【DT002】 テキストデータ

- 形式

- DOCX 形式、XLSX 形式

- データ詳細

- アセットとして登録された DOCX 形式、XLSX 形式のデータ。

- 【AL01】 LLM や 【AL02】 を用いて構造データ化を行う。

- XLSX 形式ですでに構造データ化されているデータもテキストデータとして入力される。

- ◇ 【DT203】 OCR 処理済テキストデータ

- 形式

- JSON 形式

- データ詳細

- 【DT001】 原票データに対し、【FN020\_1】 OCR 処理を行いテキストデータ化したデータ。
- 【DT001】 原票データ内で画像データ化されていた文字情報を含む。
- 出力
  - ◇ 出力スキーマ候補データ
    - 形式
      - JSON 形式
    - データ詳細
      - 【DT203】 OCR 処理済テキストデータおよび【DT002】 テキストデータを対象に、ヘッダにあたる情報を自動的に読み込み、生成された出力スキーマ候補となるデータ。
- 機能詳細
  - スキーマサジェッション機能
    - ◇ 【DT203】 OCR 処理済テキストデータおよび【DT002】 テキストデータを対象に、ヘッダにあたる情報を自動的に読み込み、生成された出力スキーマ候補となるデータを生成する。生成されたデータは【FN007】 データ構造化処理実行機能(UI)に送られ、出力スキーマ候補としてユーザーの入力支援に用いられる。
  - 利用するライブラリ
    - ◇ 【SL003】 AWS Bedrock API
    - ◇ 【SL004】 Langchain
    - ◇ 【SL005】 Faiss
    - ◇ 【SL009】 Pandas
  - 利用するアルゴリズム
    - ◇ 【AL01】 LLM
    - ◇ 【AL02】 構造データ生成モデル

【FN021】データ結合前処理（データクレンジング）

● 機能概要

- データ結合の処理、および、データ集計が可能なデータセットを作成するため、【DT003】、【DT004】、【DT204】を対象にデータの正規化、およびデータフォーマットの統一を行う機能。
- ユーザーが命名したい資料名を入力する事でカラムとして追加できる機能。本資料名は、チャット機能でも利用する。

● フローチャート

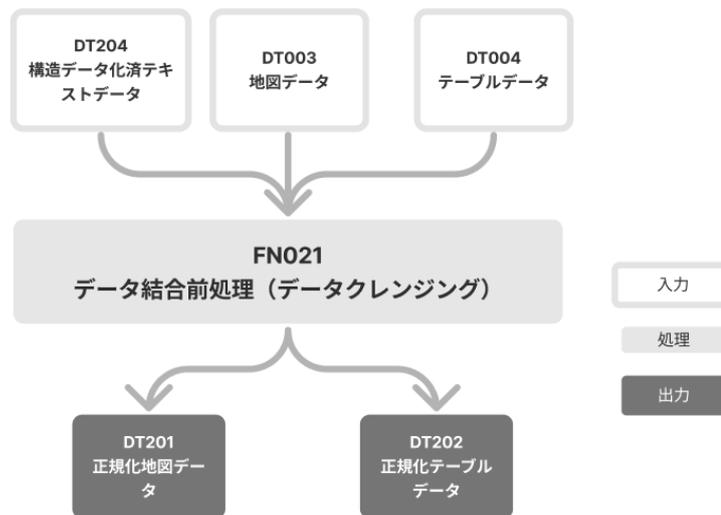


図 4-13 データ結合前処理（データクレンジング）フロー

● データ仕様

➤ 入力

- ◇ 【DT003】地図データ
  - 形式
    - SHP 形式、GeoJSON 形式
  - データ詳細
    - ジオメトリを含み、カラムとローが整理されテーブル化されたデータ。
- ◇ 【DT004】テーブルデータ
  - 形式
    - CSV 形式
  - データ詳細
    - データ処理がしやすいように、カラムとローが整理されテーブル化されたデータ。
- ◇ 【DT204】構造データ化済テキストデータ
  - 形式

- JSON 形式
- データ詳細
  - 【DT203】 OCR 処理済テキストデータおよび【DT002】 テキストデータを、構造データ化処理をしたデータ。
- 出力
  - ◇ 【DT202】 正規化テーブルデータ
    - 形式
      - JSON 形式
    - データ詳細
      - 【DT004】 テーブルデータをもとに、正規化を行ったデータ。
  - ◇ 【DT201】 正規化地図データ
    - 形式
      - GeoJSON 形式
    - データ詳細
      - 【DT003】 地図データをもとに、正規化を行ったジオメトリを含むデータ。
- 機能詳細
  - データクレンジング
    - 結合時にテキストマッチングを行う際に、【FN020\_1】 OCR 処理、および、【FN020\_2】 データ構造化処理の過程において発生した不要文字列や、元々のデータが保持していた表記揺れを、適切にデータ結合を行うために表記を統一させる。
    - データクレンジングでは、表記揺れの正規化、削除・置換、異常値の除去を実施する。
    - 表記揺れの正規化では小書き文字表記、数字表記を統一、置換処理を行い特定の用語を変換・除去させる。置換処理などの具体の指示は【FN010】 データ結合前処理実行機能 (UI) で記載の通りデータクレンジングテンプレートをを用いて行う。クレンジングイメージの具体例は以下の通り。
      - ○番地○を○-○に変換
        - ◇ 例：「○○町 1 丁目 3 番地 2 号」を「○○町 1-3-2」
      - 全角表記を半角表記に変換
        - ◇ 例：「○○町 1 - 3 - 2」を「○○町 1-3-2」
      - ガ・ケ・ツ・ノをひらがな表記へ変換
        - ◇ 例：「旭ヶ丘」を「旭が丘」
      - 置換処理キーワード入力による置換処理・除去
        - ◇ 例：「【(かっこ)】」を除去
      - 日付において、和暦・西暦が含まれる場合に西暦に統一
        - ◇ 例：「令和 2 年」を「2020 年」
      - 日付において、年月日表記を Date 型として読込可能な yyyy/mm/dd 表記に統一。
        - ◇ 例：「2023 年 8 月 11 日」を「2023/08/11」
    - 削除・置換ではユーザーがアプリ上で指定して文字列を全データ、もしくは、特定のカラ

ムに絞り、削除・置換を実施する。

- 異常値の除去では数値型でのカラムを対象に、単位の違いや OCR の抽出ミス等によって著しく値が大きい、または、小さくなってしまっているレコードを除外対象として検出する。検出には、対象カラムのデータ分布をコルモゴロフ・スミルノフ検定によって正規分布を対象とした場合の p 値が有意水準（例：5%）以下である場合に外れ値として検出する。
- 資料名作成
  - ◇ チャット機能で資料情報を閲覧できるようにする為、全ての資料に対してクレンジング処理の段階でユーザーが命名したい資料名情報を入力する事でカラムとして登録される。
- 利用するライブラリ
  - ◇ 【SL009】 Pandas
- 利用するアルゴリズム
  - ◇ なし

#### 【FN022】 データ結合前処理（ジオコーディング）

- 機能概要
  - 【DT202】 正規化テーブルデータ、および、座標情報が不完全な【DT201】 正規化地図データを地図に読み込み可能とし、【FN026】 【FN027】 の処理が可能となるように、住所情報をもとに緯度経度の座標情報を付与する機能。
- フローチャート

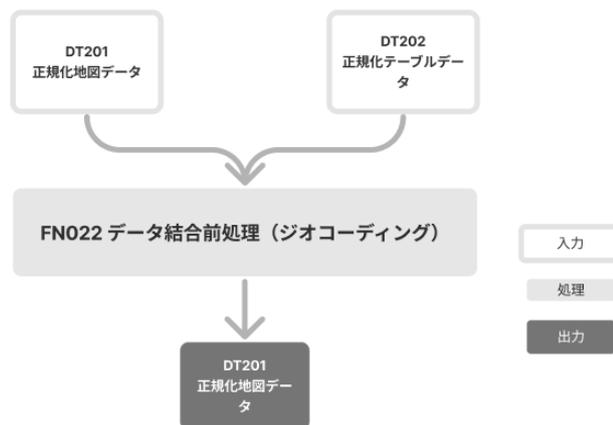


図 4-14 データ結合前処理（ジオコーディング）フロー

- データ仕様
  - 入力
    - ◇ 【DT202】 正規化テーブルデータ
      - 形式
        - JSON 形式
      - データ詳細
        - 【DT004】 テーブルデータをもとに、正規化を行ったデータ。
  - 出力
    - ◇ 【DT201】 正規化地図データ
      - 正規化地図データ形式
        - GeoJSON 形式
      - データ詳細
        - 【DT003】 地図データをもとに、ジオコーディングにより座標情報を付与したデータ。
- 機能詳細
  - ジオコーディング
    - ◇ 処理内容
      - ユーザーは【SL026】 Amazon Location API を利用して、各データに住所情報から座標情報を取得する。
    - ◇ 利用するライブラリ
      - 【SL008】 GDAL
      - 【SL010】 GeoPandas
      - 【SL013】 Turf.js
      - 【SL026】 Amazon Location API
    - ◇ 利用するアルゴリズム
      - なし

【FN023】データ結合前処理（座標系統一）

- 機能概要
  - 【DT201】正規化地図データを対象に、位置情報の座標系を統一する機能。
  - 【FN021】データクレンジング実施時に同時に実行される。
- フローチャート

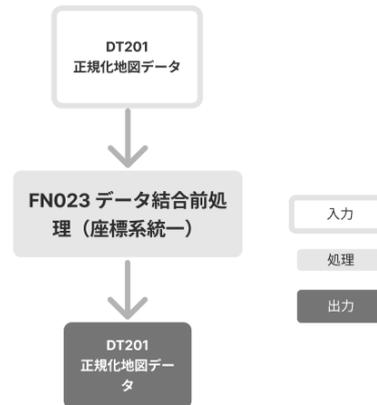


図 4-15 データ結合前処理（座標系統一）フロー

- データ仕様
  - 入力
    - ◇ 【DT201】正規化地図データ
      - 形式
        - GeoJSON 形式
      - データ詳細
        - 【DT003】地図データをもとに、正規化を行ったジオメトリを含むデータ。
  - 出力
    - ◇ 【DT201】正規化地図データ
      - 形式
        - GeoJSON 形式
      - データ詳細
        - 【DT003】地図データをもとに、正規化を行ったジオメトリを含むデータ。
- 利用するライブラリ
  - ◇ 【SL008】GDAL
  - ◇ 【SL010】GeoPandas
  - ◇ 【SL013】Turf.js

- 機能詳細
  - 座標系統一
    - ◇ 処理内容
      - 地理座標系か投影座標系など、各データの座標情報の座標系を統一させる。
      - システム側で各データセットに使用している座標系を確認し、空間解析において一般的に用いられる座標系である WGS84 (EPSG:4326) に変換する。この座標系の統一によって、ユーザーがアップロードする地図データの座標系にかかわらず、【FN026】空間結合処理および【FN027】空間集計処理が可能となる。
    - ◇ 利用するライブラリ
      - 【SL008】 GDAL
      - 【SL010】 GeoPandas
      - 【SL013】 Turf.js
    - ◇ 利用するアルゴリズム
      - なし

【FN024】データ結合機能（テキストマッチング）

● 機能概要

- 複数の異なるデータセットを結合するため、ユーザーが結合元データと結合対象データを入力し、結合条件（結合に用いるカラムや部分一致の類似度の閾値の条件）を設定することで、文字列カラムの一致するレコードを対象に属性結合する機能。

● フローチャート

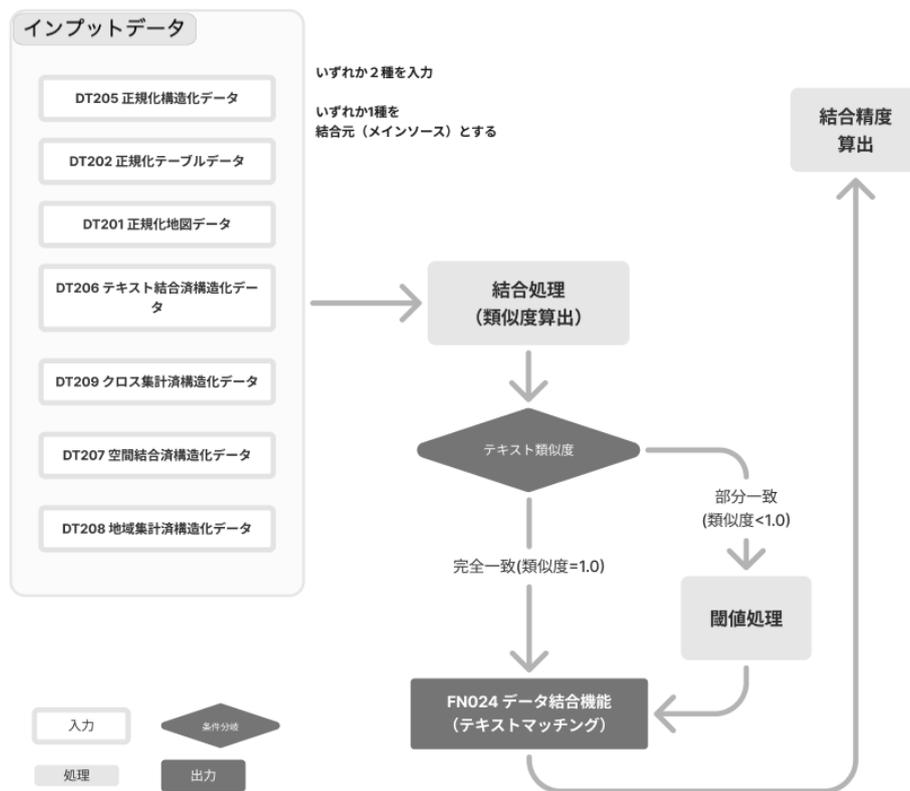


図 4-16 データ結合機能（テキストマッチング）フロー

● データ仕様

➤ 入力

- ◇ 【DT201】正規化地図データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 【DT003】地図データをもとに、正規化を行ったジオメトリを含むデータ。
- ◇ 【DT202】正規化テーブルデータ
  - 形式

- JSON 形式
- データ詳細
  - 【DT204】 構造データ化済テキストデータをもとに、正規化を行ったデータ。
- ◇ 【DT206】 データ結合済構造化データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 【DT201】 正規化地図データ、【DT202】 正規化テーブルデータ、【DT205】 クロス集計済データに対し、テキストマッチングを行ったデータ。
- ◇ 【DT207】 空間結合済データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 【DT201】 正規化地図データをもとに、空間結合によってポリゴンデータが結合されたデータを含む構造化データ。
- ◇ 【DT208】 地域集計済構造化データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 【DT201】 正規化地図データをもとに、都道府県や市区町村、メッシュ単位など地域ごとにデータが集計されたデータ。
- ◇ 【DT209】 クロス集計済構造化データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 入力されたデータに対し、クロス集計処理がなされた構造化データ。
- 出力
  - ◇ 【DT206】 データ結合済構造化データ
    - 形式
      - GeoJSON 形式
    - データ詳細
      - 【DT201】 正規化地図データ、【DT202】 正規化テーブルデータ、【DT205】 クロス集計済データに対し、テキストマッチングを行ったデータ。
- 機能詳細
  - テキストマッチング
    - ◇ 処理内容

- 2種類の共通のキーカラムを持つデータセットを対象に、メインコンテンツ（結合元）と結合コンテンツ（結合対象）を指定し、結合条件となる住所などの結合に用いるカラムを指定することで、文字列の一致による結合を行う。
  - 出力結果には結合したデータとともに、文字列の一致を示す類似度を出力する。文字列類似度の計算手法の一つである N-gram 法により、各データの住所と結合先の住所との類似度を算出し、その一致度が設定した閾値以上のデータを抽出する。なお類似度は文字列が完全に一致する場合には 1.0、部分的に一致する場合には 1.0 未満の小数值となり、値が大きいほど高い一致度となる。
  - 文字列の一致には、完全一致と部分一致の2種類の方式がある。前者は文字列が完全に一致している場合の結合方式である。後者は文字列の一致度を示す類似度を結合対象レコードごとに算出し、類似度がもっとも高いレコードを候補として抽出する。ユーザーは閾値を指定し、閾値以上であった場合に結合を行う。なお、類似度が閾値を下回るデータについては結合されず、結合後のデータにおける該当箇所は欠損値となる。
  - 結合条件となる住所などの結合に用いるカラムは、複数指定することができる。
  - 結合方法は、left join で行い、メインコンテンツに対して、上記の結合方式で一致する結合コンテンツを追加する。
- ◇ 利用するライブラリ
- 【SL006】 Mecab
  - 【SL007】 Scikit-Learn
  - 【SL009】 Pandas
- ◇ 利用するアルゴリズム
- なし

【FN025】データ集計機能（クロス集計処理）

● 機能概要

- ユーザーが指定したデータの要約統計量（合計、カウント、平均）を確認できるように、指定した任意のカラムを集計単位とし集計する機能。

● フローチャート

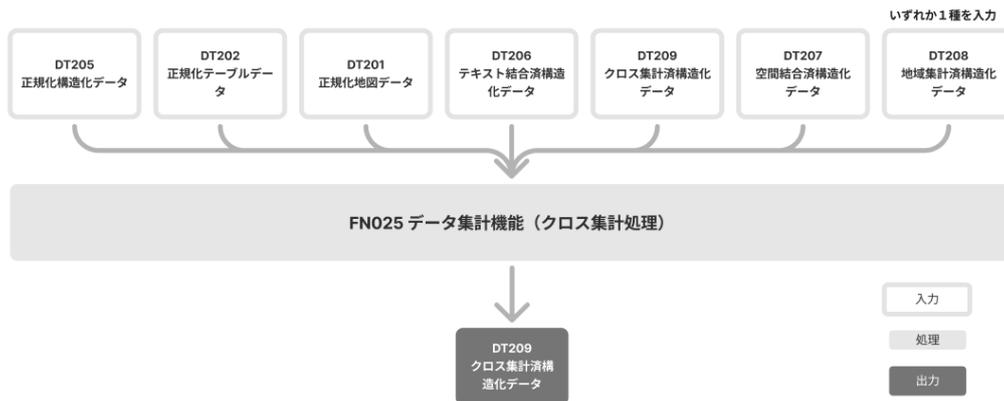


図 4-17 データ集計機能（クロス集計処理）フロー

● データ仕様

➤ 入力

◇ 【DT201】正規化地図データ

- 形式
  - GeoJSON 形式
- データ詳細
  - 【DT003】地図データをもとに、正規化を行ったジオメトリを含むデータ。

◇ 【DT202】正規化テーブルデータ

- 形式
  - JSON 形式
- データ詳細
  - 【DT204】構造データ化済テキストデータをもとに、正規化を行ったデータ。

◇ 【DT206】データ結合済構造化データ

- 形式
  - GeoJSON 形式
- データ詳細
  - 【DT201】正規化地図データ、【DT202】正規化テーブルデータ、【DT205】クロス集計済データに対し、テキストマッチングを行ったデータ。

◇ 【DT207】空間結合済データ

- 形式

- GeoJSON 形式
- データ詳細
  - 【DT201】正規化地図データをもとに、空間結合によってポリゴンデータが結合されたデータを含む構造化データ。
- ◇ 【DT208】地域集計済構造化データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 地図データで保持するジオメトリの単位ごとにデータが集計されたデータ。
- ◇ 【DT209】クロス集計済構造化データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 入力されたデータに対し、クロス集計処理がなされた構造化データ。
- 出力
  - ◇ 【DT209】クロス集計済構造化データ
    - 形式
      - GeoJSON 形式
    - データ詳細
      - 入力されたデータに対し、クロス集計処理がなされた構造化データ。
- 機能詳細
  - クロス集計処理
    - ◇ 処理内容
      - 1種類の入力データを対象とし、入力データ内にあるカラムを対象に、ユーザーは集計対象とするカラムおよび集計時にキーとするカラムを指定し、クロス集計する機能。
      - カテゴリごとのサンプル数（出現回数・頻度、UI上ではカウントと表現）や合計・平均・カウントを算出することができる。集計対象が数値の場合には、合計・平均・カウントを適用可能とし、文字列の場合にはカウントのみとする。
    - ◇ 利用するライブラリ
      - 【SL009】Pandas
    - ◇ 利用するアルゴリズム
      - なし

【FN026】データ結合機能（空間結合処理）

● 機能概要

- 複数の異なるデータセットを結合するため、ユーザーが結合元データと結合対象データを入力し、ポリゴンとポイントにおける交差結合をする機能を提供する。位置精度によるずれを防ぐため、最近傍結合も考慮する。

● フローチャート

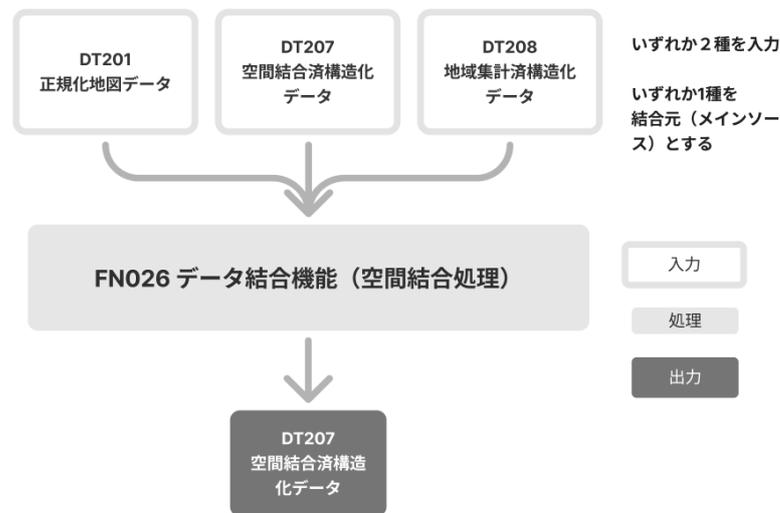


図 4-18 空間結合機能フロー

● データ仕様

➤ 入力

- ◇ 【DT201】正規化地図データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 【DT003】地図データをもとに、正規化を行ったジオメトリを含むデータ。
- ◇ 【DT207】空間結合済データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 【DT201】正規化地図データをもとに、空間結合によってポリゴンデータが結合されたデータを含む構造化データ。
- ◇ 【DT208】地域集計済構造化データ
  - 形式

- GeoJSON 形式
- データ詳細
  - 地図データで保持するジオメトリの単位ごとに集計されたデータ。
- 出力
  - ◇ 【DT207】空間結合済データ
    - 形式
      - GeoJSON 形式
    - データ詳細
      - 【DT201】正規化地図データをもとに、空間結合によってポリゴンデータが結合されたデータを含む構造化データ。
- 機能詳細
  - 空間結合
    - ◇ 処理内容
      - 2種類の共通の座標系のジオメトリ情報を持つデータセットを対象に、空間上で交差することを条件として、結合させる機能。座標の僅かなずれによってデータが結合されない場合は、最近傍結合を行い結合させる。
      - なお、結合条件には交差や最近傍の他にも包含や重なり等があるが、交差結合と最近傍結合においてカバーが可能であり、ユーザーに専門知識を持たずとも操作しやすくなることを考慮し、交差結合と最近傍結合を結合条件として想定する。
      - 結合時において、ユーザーはメインコンテンツ（結合元）と結合コンテンツ（結合対象）を指定し、結合条件に合致した結合コンテンツのレコードを、メインコンテンツのレコードの横方向に1対1の結合が行われるように実施する。
      - 結合させるジオメトリデータの組み合わせは、ポイントデータとポリゴンデータ同士が想定される。ポリゴンデータは行政区界データやメッシュデータ等の【DT003】地図データが想定される。ポイントデータはPOIや施設の位置を示す【DT003】地図データや、【FN022】データ結合前処理（ジオコーディング）によって【DT201】正規化地図データとなったデータが該当する。
  - 利用するライブラリ
    - ◇ 【SL008】GDAL
    - ◇ 【SL010】GeoPandas
    - ◇ 【SL013】Turf.js
  - 利用するアルゴリズム
    - ◇ なし

【FN027】データ集計機能（空間集計処理）

● 機能概要

- ユーザーが指定した地域データによって、要約統計量（合計、カウント、平均）を再集計できるように、ユーザーが指定した条件にのっとり（都道府県単位等）、ユーザーがアップロードした地域ポリゴンデータを結合させ、地域単位で集計し、新規地図データとして保存する機能。

● フローチャート

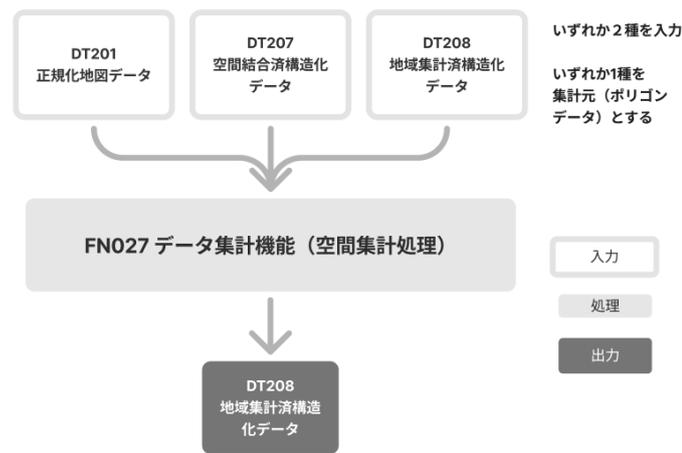


図 4-19 データ集計機能（空間集計処理）フロー

● データ仕様

➤ 入力

- ◇ 【DT201】正規化地図データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 【DT003】地図データをもとに、正規化を行ったジオメトリを含むデータ。
- ◇ 【DT207】空間結合済データ
  - 形式
    - GeoJSON 形式
  - データ詳細
    - 【DT204】構造データ化済テキストデータをもとに、正規化を行ったデータ。
- ◇ 【DT208】地域集計済構造化データ
  - 形式
    - GeoJSON 形式
  - データ詳細

- 地図データで保持するジオメトリの単位ごとに集計されたデータ。
- 出力
  - ◇ 【DT208】 地域集計済構造化データ
    - 形式
      - GeoJSON 形式
    - データ詳細
      - 地図データで保持するジオメトリの単位ごとに集計されたデータ。
- 機能詳細
  - 空間集計処理
    - ◇ 処理内容
      - 2種類の共通の座標系のジオメトリ情報を持つデータセットを対象に、空間上で空間複数交差することを条件として、都道府県や市区町村、メッシュ単位などユーザーが指定した地域単位ごとにデータを集計する機能。
      - なお、結合条件には交差や最近傍の他にも包含や重なり等があるが、交差結合と最近傍結合においてカバーが可能であり、ユーザーに専門知識を持たずとも操作しやすくなることを考慮し、交差結合と最近傍結合を結合条件として想定する。
      - 結合時において、ユーザーはメインコンテンツ（結合元）と結合コンテンツ（結合対象）を指定する。これらは1対多の比率となることを想定する。結合条件に合致した場合に、メインコンテンツと交差する複数の結合コンテンツのレコードを集計した情報、結合元のレコードの横方向に追加が行われるように実施する。
      - 地域単位ごとのサンプル数（出現回数・頻度、UI上ではカウントと表現）や合計・平均・カウントを算出することができる。結合させるジオメトリデータの組み合わせは、メインコンテンツがポリゴンデータまたはラインデータ、結合コンテンツがポイントデータとなる場合において集計対象が数値の場合には、合計・平均・カウントを適用可能とし、文字列の場合にはカウントのみとする。上記以外となる場合には、カウントのみとする。
    - ◇ 利用するライブラリ
      - 【SL008】 GDAL
      - 【SL010】 Ge、oPandas
      - 【SL013】 Turf.js
    - ◇ 利用するアルゴリズム
      - なし

#### 【FN028】 グラフカスタマイズ機能

- 機能概要
  - 本機能は、EBPM Tools における共通機能として具備する。【FN019】 データ配信機能のアプリケーション利用で「公開」となったデータに対し、事前に定義されたレイアウトパターンを選択しグラフ化する機能を提供する。

- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ パラメータ
  - 出力
    - ◇ なし
- 機能詳細
  - ダッシュボード上でパラメータを指定して集計した内容を元に、グラフを作成する。
  - グラフはあらかじめ定義されたグラフパターン（円グラフ、棒グラフ、折れ線グラフ、レーダーチャートなど）に則り作成する。
    - ◇ 利用するライブラリ
      - ReCharts
    - ◇ 利用するアルゴリズム
      - なし

#### 【FN029】 マッピング機能

- 機能概要
  - 本機能は、EBPM Tools における共通機能として具備する。【FN019】データ配信機能のアプリケーション利用で「公開」となった地図データをダッシュボード上にマッピングする機能を提供する。地図データの表示、レイヤの重畳、スタイル編集等の機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ パラメータ
  - 出力データ仕様
    - ◇ 【DT906】属性情報出力データ
      - 形式
        - CSV 形式
      - データ詳細
        - 【FN029】マッピング機能で集計した属性データ
- 機能詳細
  - ダッシュボード上でパラメータを指定して集計した内容を元に、マップエリアにマップを表示、選択したジオメトリをハイライトする。
  - 選択したジオメトリの属性情報、パラメータに則り集計された属性情報をサイドパネルに表示し、CSV形式で出力する。

- マップエリア上で複数のジオメトリを重畳する。
- パラメータ
  - ◇ 利用するライブラリ
    - MapLibre
  - ◇ 利用するアルゴリズム
    - なし

#### 【FN030】レポート出力機能

- 機能概要

本機能は、EBPM Tools における共通機能として提供する。【D004】、【D005】と連携したレポート出力機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ なし
  - 出力
    - ◇ 【DT906】出力レポートデータ
    - ◇ 【DT907】属性情報出力データ
      - 形式
        - PDF 形式、CSV 形式
      - データ詳細
        - 【FN028】グラフカスタマイズ機能で作成したグラフ
        - 【FN029】マッピング機能で可視化したマップデータ、ジオメトリの属性情報
- 機能詳細
  - 作成したマップ及びグラフを PDF 形式、CSV 形式で出力する。
    - ◇ 利用するライブラリ
      - ReCharts
      - MapLibre
    - ◇ 利用するアルゴリズム
      - なし

### 【FN031\_1】 認証機能

- 機能概要
  - 任意のユーザーに対してアカウント発行、メールアドレスおよびパスワードの変更を行う機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ メールアドレス、パスワード
      - 形式
        - テキスト形式
      - データ詳細
        - ユーザーのログイン情報
  - 出力
    - ◇ なし
- 機能詳細
  - Veda ログインの際の認証機能として、メールアドレスとパスワードで認証を行いログインする機能を提供する。
  - メールアドレスにセキュリティコードを送付する 2 段階認証による認証機能を提供する。
    - ◇ 利用するソフトウェア
      - Cloud Identity Platform
    - ◇ 利用するアルゴリズム
      - なし

### 【FN031\_2】 アカウント管理機能

- 機能概要
  - UI 上からユーザーの追加・削除、権限管理を行う機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ なし
  - 出力
    - ◇ なし
- 機能詳細

- システム管理者がユーザー管理を行う。ユーザーの追加、削除、権限の変更を行う。本システムでは、権限はシステム管理者、一般ユーザーのみとする。
- ◇ 利用するソフトウェア
  - React/TypeScript
- ◇ 利用するアルゴリズム
  - なし

### 【FN031\_3】データセット読み込み機能

- 機能概要
  - 本機能は、EBPM Tools における共通機能として提供する。ユーザー(=原課レベル)は、EBPM Tools 画面訪問時に、利用するデータセットをリストから選択して読み込み、EBPM Tools とデータの紐づけを行う。データセットにはユースケースで利用するコンテンツ情報が紐づけられている。
  - 選択するデータセットは【FN036】データセット管理機能にて作成した EBPM 用データセットのみ表示される。利用できるコンテンツは【FN004】コンテンツ管理の公開ステータスで「可視化」がオンであるものに限る。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ 【DT901】最終構造化データ
      - 形式
        - JSON 形式、 GeoJSON 形式
      - データ詳細
        - 【FN036】データセット管理機能にて作成した EBPM 用データセットのみ表示される。利用できるコンテンツは【FN004】コンテンツ管理の公開ステータスで「可視化」がオンであるものに限る。
  - 出力
    - ◇ なし
- 機能詳細
  - 本機能は、EBPM Tools と利用するデータの紐づけを行う機能である。EBPM Tools 画面の初回訪問時、および利用するデータに対して更新が入った場合、ユーザーは本機能を利用してデータセットの紐づけを行う必要がある。
  - ◇ 利用するソフトウェア
    - React/Type Script
  - ◇ 利用するアルゴリズム
    - なし

#### 【FN031\_4】 設定 API 機能

- 機能概要
  - 本機能は、ユースケース+ユーザーの組み合わせで設定情報を Veda API サーバー内に保存する機能である。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ なし
  - 出力
    - ◇ なし
- 機能詳細
  - 本機能は、【FN031\_3】データセット読み込み機能、【FN028】グラフカスタマイズ機能、【FN029】マップカスタマイズ機能によって設定したカスタマイズ情報をユーザーごとに保存し、呼び出すものである。初期設定後、画面再訪問時に自動で読み込みがなされるため、ユーザーによる都度の読み込みは不要である。
  - 取り込んだデータセット、グラフィックアウトの変更、色の変更、グラフ種類の変更などユーザーレベルで実施するカスタマイズはオペレーションの都度 DB に保存される。
    - ◇ 利用するソフトウェア
      - React/Type Script
    - ◇ 利用するアルゴリズム
      - なし

#### 【FN031\_5】 データ参照機能

- 機能概要
  - 本機能は、EBPM Tools における共通機能として提供する。EBPM Tools 上からアプリ上で利用する Defined Data の情報を参照する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ なし
  - 出力
    - ◇ なし
- 機能詳細

- 本機能は、EBPM Tools で利用する Defined Data のデータの中身をモダル上で参照できる機能である。複数の Defined Data を取り込む場合、タブによってデータの切り替えを行う。
  - ◇ 利用するソフトウェア
    - React/Type Script
  - ◇ 利用するアルゴリズム
    - なし

#### 【FN032】 問い合わせ文入出力画面

- 機能概要
  - ユーザーがお問い合わせ内容を入力し、入力内容に対して【FN034】チャット生成機能での生成結果を出力する機能。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ 【FN034】チャット生成機能に則る
  - 出力
    - ◇ 【FN034】チャット生成機能に則る
  - 利用するライブラリ
    - React/Type Script
- 機能詳細
  - 【FN034】チャット生成機能を起動するためのインプットとして問い合わせ文を入力し、検索結果を出典と共に受け取るための画面。

#### 【FN033】 ベクトルデータ作成機能（ベクトル解析）

- 機能概要
  - 【FN032】問い合わせ文入出力画面からユーザーが入力したお問い合わせ文、および、【DT901】最終構造化データに対して、【FN20-3】ベクトル解析を実行する機能。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ 【FN020\_3】ベクトル解析に準ずる。
  - 出力
    - ◇ 【FN020\_3】ベクトル解析に準ずる。

- 利用するライブラリ
  - ◇ React/TypeScript
- 利用するアルゴリズム
  - ◇ なし
- 機能詳細
  - 【DT901】最終構造化データに対し、ベクトル解析を実行する機能。チャット機能起動時に、ユーザーがベクトルデータ作成を実行する。

#### 【FN034】チャット生成機能

- 機能概要

【FN032】問い合わせ文入出力画面において、ユーザーからの問い合わせ入力内容を元に関連性のあるデータを【FN033】ベクトルデータ作成機能で生成したベクトルデータから検索・取得し、質問と取得したデータを埋め込んだプロンプトを作成することで、LLM によってユーザーへの返答文（自然言語）を出力する機能。
- フローチャート

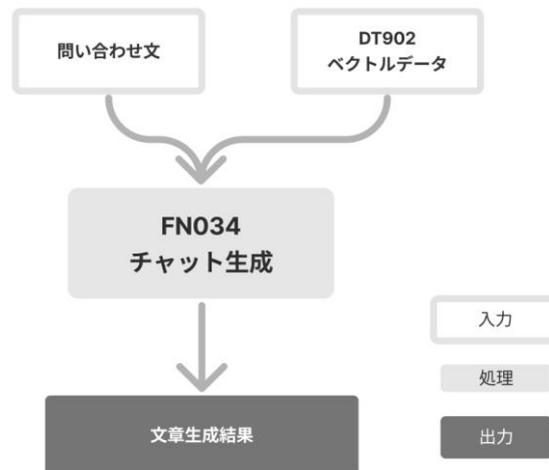


図 4-20 チャット生成フロー

- データ仕様
  - 入力
    - ◇ 【DT902】ベクトルデータ
      - 内容
        - 【FN020\_3】ベクトル解析によって生成されたベクトルデータ
      - 形式

- JSON 形式
- ◇ ユーザー入力データ
  - 内容
    - ユーザーが【FN032】 お問い合わせ文入出力画面に対して入力したテキストデータ
  - 形式
    - JSON 形式
- 出力
  - ◇ 文章生成結果
    - 内容
      - ユーザーからの入力（クエリ）に対して、ベクトルデータから生成した回答文
    - 形式
      - JSON 形式
    - データ詳細
      - ユーザーからの入力（クエリ）を取得し、【AL02】 ベクトル解析によってクエリをベクトル化し、【DT902】 ベクトルデータから最も類似度の高い情報を抽出し、【AL01】 LLM によってユーザーのクエリに対する自然な回答を生成し出展名と資料名、URL を含めたデータ。
- 機能詳細
  - チャット生成
    - ◇ LLM がユーザーの生成した【DT901】 最終構造化データをチャットから情報検索できるように、RAG（Retrieval-Augmented Generation）による回答生成を行えるようにする。
    - ◇ ユーザーからの入力（クエリ）を取得し、【AL02】 ベクトル解析によってクエリをベクトル化し、【DT902】 ベクトルデータから最も類似度の高い情報を抽出し、【AL01】 LLM によってユーザーのクエリに対する自然な回答を生成する。
  - 利用するライブラリ
    - ◇ 【SL003】 AWS Bedrock API
    - ◇ 【SL004】 Langchain
    - ◇ 【SL005】 Faiss
  - 利用するアルゴリズム
    - ◇ 【AL04】 チャット生成

【FN035】 テンプレート管理機能

- 機能概要
  - テンプレート管理機能では、データ生成に必要な処理を一連処理として実行するため、各オペレーターをシーケンシャルに設定したワークフローテンプレート、個別のオペレーター起動に必要なパラメータ情報を設定する個別テンプレートを作成・実行する機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ なし
  - 出力
    - ◇ なし
- 機能詳細
  - テンプレート管理機能では、以下のテンプレートを管理する。
    - ◇ ワークフローテンプレート
      - 一連の処理を一気通貫で実行するためにオペレーターをグルーピングしたテンプレートを指す。
    - ◇ 構造化テンプレート
      - 【FN007】 データ構造化処理 (UI) で利用するテンプレート。非構造化データを構造化する際に利用する。
      - テンプレートには、カラム名、タイプ、ポジション (抽出するキーワードが存在する文書内の章など)、単位、キーワード、自由文、出典名作成情報を設定することができる。
      - 本テンプレートは【DT001】 原票データ (PDF、PNG 形式)、【DT002】 テキストデータ (DOCX、XLSX 形式) のデータ構造化処理で利用する。
    - ◇ 結合前処理テンプレート
      - 【FN010】 データ結合前処理実行機能 (UI) で利用するテンプレート。資料名作成、データクレンジング、ジオコーディング、秘匿化処理を実行する。
    - ◇ テキストマッチングテンプレート
      - 【FN011】 データ結合実行機能 (テキストマッチング) (UI) で利用するテンプレート。データ結合の際のメインコンテンツ、結合コンテンツ、キーカラム、結合カラム、自由文が定義されている。
    - ◇ 空間結合テンプレート
      - 【FN013】 データ結合実行機能 (空間結合) (UI) で利用するテンプレート。空間結合の際のメインコンテンツ、結合コンテンツ、結合方式、検索上限範囲 (最近傍結合のみ) の情報が定義されている。
    - ◇ クロス集計テンプレート

- 【FN012】 データ集計実行機能（クロス集計）（UI）で利用するテンプレート。メインコンテンツ、データ集計の際の集計方式（合計/平均、カウント）、集計単位、集計対象の情報が定義されている。
- ◇ 空間集計テンプレート
  - 【FN014】 データ集計実行機能（空間集計）（UI）で利用するテンプレート。メインコンテンツ、集計コンテンツ、データ集計の際の集計方式（合計/平均、カウント）、集計単位、集計対象の情報が定義されている。
- 【SC014】 テンプレート一覧画面より、各テンプレートの削除、編集、実行ができる。
- 利用するソフトウェア
  - ◇ React/TypeScript
- 利用するアルゴリズム
  - ◇ なし

#### 【FN036】 データセット管理機能

- 機能概要
  - 本機能では、コンテンツをユースケースごとにグルーピングを行い、データセットとして管理する。
  - データセットは、新規作成、編集、削除が可能である。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ なし
  - 出力
    - ◇ なし
- 機能詳細
  - データセット管理画面では、コンテンツの公開ステータス「可視化」がオンであるコンテンツのみ紐づけ可能である。
  - 紐づくコンテンツの公開ステータスが1つでもオンの状態であるとき、データセットの削除はできないものとする。
  - コンテンツを紐づけ後、メタデータを入力する。以下はデータセットで設定するメタデータである。G 空間情報センターに公開される際には、データセットのメタデータとして公開する。
  - データ仕様書の情報をファイルアップロードにて紐づける。データ仕様書には、項目リスト、コードリスト、ER 図を含む。
  - データセットは、コンテンツと同様に公開ステータスを保持する。公開ステータスのオン/オフはオープンデータの公開に影響があるため、システム管理者のみ操作可能である。コンテンツの公開ステータスとデータセットの公開ステータスの関係性【FN019】 データ配信機能の 4-2-3. 開発機能の詳細要件にて詳細を記載する。

表 4-8 データセット用メタデータ

ID	入力項目名称	デフォルト入力値	内容
1	管理 ID	ProjectLinks{ID となる数字}_{年度}	● G 空間情報センターで管理するデータセットを一意で判断するための ID。
2	タイトル（データセット名称）		● G 空間情報センターで管理するデータセットを一意で判断するための名称。
3	説明	データの作成にあたっては、紙で保管された報告書をスキャンして PDF 化し、LINKS Veda を使用しました。 LINKS Veda によるデータ作成の精度は紙資料の保存状態や記入状況に依存し、データの完全性や正確性を保証するものではありません。 <a href="https://www.mlit.go.jp/links/">https://www.mlit.go.jp/links/</a> 本データは政府標準利用規約（第 2.0 版）に準拠しています。また、クリエイティブ・コモンズ・ライセンスの表示 4.0 国際と互換性があるとともに、利用者が Open Data Commons による ODC BY 又は ODbL での利用を希望する場合に、それを妨げるものではありません。	● G 空間情報センターに公開するデータに関する詳細説明。Project LINKS では、左記を固定とする。
4	キーワード		● G 空間情報センターでデータセット検索を補助するためのキーワード情報
5	テーマ分類		● G 空間情報センターでデータセット検索を補助するためのタグで保持する情報
6	対象地域	日本全国	● データ対象地域。Project LINKS では、左記を固定とする。
7	対象期間		● 公開するデータ対象期間を示す。
8	提供者	国土交通省 総合政策局 情報政策課	● データ提供者を示す。

			Project LINKS では、左記を固定とする。
9	連絡先情報		● データ提供者の連絡先情報を示す。
10	作成者	国土交通省 総合政策局 情報政策課	● データ作成者を示す。 Project LINKS では、左記を固定とする。
11	公開日		● G 空間情報センターへのデータセット公開日を示す。
12	最終更新日		● 作成されたデータセットの最終更新日を示す。
13	更新頻度	不定期	● 更新頻度を示す。Project LINKS では、左記を固定とする。
14	言語	ja	● データの言語を示す。 Project LINKS では、左記を固定とする。
15	公開範囲	制限付き公開	● データの公開範囲を示す。 Project LINKS では、左記を固定とする。
16	公開条件		● データの公開条件を示す。
17	ライセンス	政府標準利用規約（第 2.0 版）	● データのライセンスを示す。Project LINKS では、左記を固定とする。
18	利用規約	<a href="https://www.mlit.go.jp/links/terms-of-use.html">https://www.mlit.go.jp/links/terms-of-use.html</a>	● データの利用規約を示す。 Project LINKS では、左記を固定とする。
19	バージョン	ver1.0	● データのバージョンを示す。データ更新の度にバージョンを上げて対応する。 ● データセットを変更した場合：ver 番号を更新する。 （例：1.0→2.0） ● データセット内のコンテンツを更新した場合：枝番を更新する。（例：1.0→1.1）
20	タイプ	CSV、JSON、GeoJSON	● G 空間情報センターに公開

			する際のデータセットのファイル形式を示す。
21	エンコーディング	UTF-8	● G 空間情報センターに公開するデータのエンコーディング。UTF-8 BOM 付のファイルで公開される。
22	来歴情報		● G 空間情報センターへの来歴情報。空欄とする。
23	品質評価	正確性、完全性、一貫性	● LINKS Veda で作成したデータの品質評価を示す。 Project LINKS では、左記を固定とする。
24	データ品質（品質測定結果）	統計的な分析を行うのに十分な品質。 ただし、データ型の統一などデータクレンジング処理を行っておりますが、紙資料の保存状態や記入状況によっては、正確な記載となっていない場合がございます。	● Project LINKS では、左記を固定とする。
25	制約	利用規約による	● Project LINKS では、左記を固定とする。
26	有償無償区分*	無償	● Project LINKS では、左記を固定とする。
27	災害時区分*	無償提供	● Project LINKS では、左記を固定とする。
28	価格情報	無償	● Project LINKS では、左記を固定とする。
29	使用許諾	Project LINKS の利用規約に従って、どなたでも、複製、公衆送信、翻訳・変形等の翻案等、自由に利用できます。 商用利用も可能です。 ( <a href="https://www.mlit.go.jp/links/terms-of-use.html">https://www.mlit.go.jp/links/terms-of-use.html</a> )	● Project LINKS では、左記を固定とする。
30	準拠する標準	<a href="https://www.w3.org/TR/vocab-dcat-3/">https://www.w3.org/TR/vocab-dcat-3/</a>	● Project LINKS では、左記を固定とする。
31	関連ドキュメント	<a href="https://www.mlit.go.jp/links/">https://www.mlit.go.jp/links/</a>	● Project LINKS では、左記を固定とする。

- 利用するソフトウェア
  - ◇ React/TypeScript
- 利用するアルゴリズム
  - ◇ なし

### 【FN037】秘匿処理機能

- 機能概要
  - 本機能では【FN019】データ配信機能のデータ公開に向けて、数値情報の階層化・偏差値化、新規コードの追加、住所の秘匿化を行い、元の情報を特定できないように秘匿化する機能を提供する。
  - 本機能は、【FN010】データ結合前処理実行機能（UI）から実行するものとする。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ 【DT901】最終構造化データ
      - 内容
        - 【DT206】テキスト結合済構造化データ、【DT207】クロス集計済構造化データ、【DT208】空間結合済データ、【DT209】地域集計済構造化データのうち、最終的に出力されるデータ。
      - 形式
        - JSON形式、GeoJSON形式
      - データ詳細
        - D002 正規化ローデータの集計・加工機能の処理を完了し、【FN018】にてデータ承認を行うことができる状態のデータ。
  - 出力
    - ◇ 【DT901】最終構造化データ
      - 内容
        - 【DT206】テキスト結合済構造化データ、【DT207】クロス集計済構造化データ、【DT208】空間結合済データ、【DT209】地域集計済構造化データのうち、最終的に出力されるデータ。
      - 形式
        - JSON形式、GeoJSON形式
      - データ詳細
        - D002 正規化ローデータの集計・加工機能の処理を完了し、【FN018】にてデータ承認を行うことができる状態のデータ。
- 機能詳細
  - 秘匿処理機能

- ◇ 本機能ではオープンデータ公開において、直接的な一般公開が望ましくない情報を秘匿化するための機能として、階層化・偏差値化、新規コードの追加、住所の秘匿化機能を提供する。
- ◇ 階層化、偏差値化：対象項目が数値である場合に、該当項目を等量分類で手動または自動的に2から10階層までに分ける。また、偏差値化を行うことで同様に、元の情報が確認できない状態で分布がわかるようにする。
- ◇ 新規コードの追加：一意となるIDを新規に生成するため、uuidを生成する。
- ◇ 住所の秘匿化機能：住所情報から市区町村名称抽出し、付与する。
- 利用するソフトウェア
  - ◇ 【SL009】Pandas
- 利用するアルゴリズム
  - ◇ なし

### 【FN038】英語カラム化機能

- 機能概要
  - ・ 本機能では【FN019】データ配信機能のデータ公開に向けて、【AL01】LLM(Large Language Model)を用いて日本語カラムを英語カラムへと変換する機能を提供する。
- フローチャート
  - なし
- データ仕様
  - 入力
    - ◇ 【DT901】最終構造化データ 日本語カラム一覧
      - 内容
        - 【DT206】テキスト結合済構造化データ、【DT207】クロス集計済構造化データ、【DT208】空間結合済データ、【DT209】地域集計済構造化データのうち、最終的に出力されるデータ。
      - 形式
        - JSON形式、GeoJSON形式
      - データ詳細
        - D002 正規化ローデータの集計・加工機能の処理を完了し、【FN018】にてデータ承認を行うことができる状態のデータのうち、カラム一覧だけ取り出したデータ。
  - 出力
    - 【DT901】最終構造化データ 英語カラム一覧
    - 形式
      - JSON形式、GeoJSON形式
    - データ詳細
      - D002 正規化ローデータの集計・加工機能の処理を完了し、【FN018】にてデータ承認を行うことができる状態のデータのうち、カラム一覧だけ取り出したデータ。

- 機能詳細
  - 英語カラム化機能
    - ◇ オープンデータ公開に向けて、【DT901】最終構造化データをデータベースアプリケーション等へ読み込みがしやすくなるように、カラム情報を英語カラムへと変換する。英訳にあたって、【DT901】最終構造化データのカラム情報を規定のプロンプトにいれ、【AL01】LLM(Large Language Model) によって英訳を実施する。
  - 利用するライブラリ
    - ◇ 【SL003】 AWS Bedrock API
    - ◇ 【SL004】 Langchain
    - ◇ 【SL005】 Faiss
  - 利用するアルゴリズム
    - ◇ 【AL01】 LLM(Large Language Model)

## 4-3. アルゴリズム

表 4-9 利用するアルゴリズム一覧

ID	アルゴリズムを利用した機能	名称	説明
AL01	【FN020_2】	LLM (Large Language Model)	深層学習アーキテクチャ (Transformer) に大規模なデータを学習することで、質問応答・翻訳・コード作成など様々なタスクを可能にしたモデル。
AL02	【FN020_2】	構造データ生成モデル	LLM によって、異なるタイプの非構造データからドキュメントを対象に、特定のキーワードやフレーズをもとに構造データ化を可能にしたモデル。
AL03	【FN020_3】	ベクトル解析	LLM が入力テキストの意味と文脈を理解し、適切な応答を生成可能とするため、テキストデータを高次元空間上における数値ベクトルに変換するアルゴリズム。
AL04	【FN034】	チャット生成	RAG(Retrieval-Augmented Generation)により、LLM がユーザーの生成した最終構造化データから情報検索、回答を行えるようにする機能。

## 4-3-1. 利用したアルゴリズム

表 4-10 利用するアルゴリズム一覧

ID	アルゴリズム を利用した機能	名称	説明
AL01	【FN020_2】	LLM (Large Language Model)	大規模なデータを学習することで、質問応答・翻訳・コード作成など様々なタスクを可能にしたモデル。
AL03	【FN020_3】	ベクトル解析	LLM が入力テキストの意味と文脈を理解し、適切な応答を生成可能とするため、テキストデータを高次元空間上における数値ベクトルに変換するアルゴリズム。

## AL01 【FN020\_2】 LLM (Large Language Model)

- Transformer と呼ばれる深層ニューラルネットワークアーキテクチャを採用した自然言語処理モデルで、膨大な量の文書やインターネット上のテキストにおける質問応答、要約作成、翻訳、感情分析などを学習済。シーケンス内の全ての要素間の関係を同時に考慮して、高度な自然言語理解により、入力した情報から汎用的に LLM は投入したデータから、特定のドメインにおける専門知識を反映して出力できる。
- 【AL03】 ベクトル解析によって文章からベクトル化・Self Attention 複数トークンの情報を混合したベクトルを計算することで、トークン間の依存関係の強さを定量化し、トークン間の依存関係を調節することで、主語、動詞、目的語等の文章構造を理解・生成。
- 出力指示: 「事業所名」、「住所」などのキー項目を抽出し、そのキー項目の右側や下側にある対象項目を推定して、抽出。レイアウト構造を解析し項目抽出を行い、テキストの内容だけでなく、テキストの配置やページ上の位置といったレイアウト情報も利用文書内の特定の位置で情報を読み取る Markdown、HTML、XML、Json など様々な形式のテーブル情報を理解。

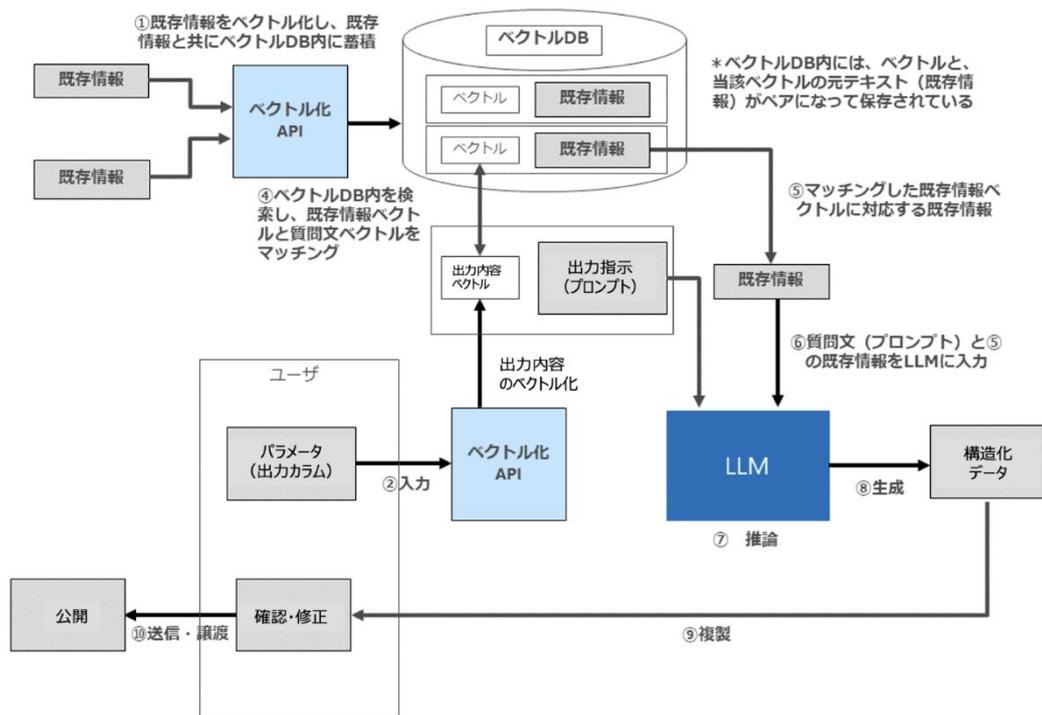


図 4-21 LLM データ処理フロー

### AL03 【FN020\_3】 ベクトル解析

- LLM が入力テキストの意味と文脈を理解し、適切な応答を生成可能とするため、テキストデータを高次元空間上における数値ベクトル (Embedding) に変換するアルゴリズム。

①トークン化: 入力テキストをトークン (単語やサブワード) に分割し、テキストを一連のトークンのシーケンスとする。

入力例: "海外からのインバウンドを推進する"の場合  
["海外", "から", "の", "インバウンド", "を", "推進", "する"]

②トークンの数値化: 各トークンを一意の数値 ID に変換し、テキストは数値のシーケンスとして表現。

入力例: "海外からのインバウンドを推進する"の場合  
[1234, 567, 89, 4321, 101, 2345, 678]

③エンベディングレイヤー: 数値 ID のシーケンスを高次元ベクトル空間に埋め込み、各トークン ID は事前に学習されたエンベディングマトリックスを使用して、固定長のベクトルにマッピング。

入力例: "海外からのインバウンドを推進する"の場合  
[ [0.12, -0.45, ..., 0.34],  
[-0.22, 0.33, ..., -0.14],  
[0.05, 0.19, ..., 0.27],  
[0.36, -0.12, ..., 0.18],  
[-0.15, 0.27, ..., 0.22],  
[0.20, -0.29, ..., -0.08],  
[-0.07, 0.40, ..., 0.15] ]

④コンテキストエンコーディング: Embeddings をモデルに入力し、文脈を考慮したエンコーディングを実施。Transformer モデルのコアコンポーネントである Self Attention メカニズムを使用して、入力されたテキスト内で各単語 (トークン) が他の単語にどの程度注目すべきかを以下のように計算することで、文脈を考慮した意味理解が可能となる。

A. クエリ (Query)、キー (Key)、バリュー (Value) の計算

各単語 (トークン) について、クエリベクトル、キーベクトル、バリューベクトルを計算。

Query ( $Q$ ): 他の単語に対して「どのような情報を探しているか」を表現するベクトル。

Key ( $K$ ): 自分が「どのような情報を持っているか」を表現するベクトル。

Value (V): 自分が「どの情報を提供するか」を表現するベクトル。

#### B. Attention Scores の計算

ある単語が他の単語にどれだけ注目すべきかを示す Attention Score を計算。クエリとキーのドット積の計算から得られる。

$$\text{Attention Score} = Q \cdot K^T$$

#### C. スコアの正規化 (Softmax)

Attention Scores は Softmax 関数を使って正規化し、各単語に対する注目度 (重み) を計算。

$$\text{Attention Weights} = \text{Softmax} \left( \frac{Q \cdot K^T}{\sqrt{d_k}} \right)$$

$d_k$  : キーベクトルの次元数 (スコアのスケールングに使用)

#### D. コンテキストの集約

各単語のバリューベクトルと Attention Weights を掛け合わせ、重み付きのバリューベクトルを計算します。これらをすべての単語について集約することで、その単語の新しい表現が得られます。

$$\text{Output} = \sum (\text{Attention Weights} \times V)$$

⑤出力生成: 最終的に、コンテキストを考慮したエンコーディングに基づき、次のトークンを予測し、文脈に沿ったテキストを出力。

## 4-3-2. 開発したアルゴリズム

表 4-11 開発するアルゴリズム

ID	アルゴリズム を利用した機能	名称	説明
AL02	【FN020_2】	構造データ生成モデル	LLM によって、異なるタイプの非構造データからドキュメントを対象に、特定のキーワードやフレーズをもとに構造データ化を可能にしたモデル
AL04	【FN034】	チャット生成	RAG(Retrieval-Augmented Generation)により、LLM がユーザーの生成した最終構造化データから情報検索、回答を行えるようにする機能。

## AL02 【FN020\_2】 構造データ生成モデル

- 非構造データを対象とし、対象データの単語や文といった自然言語の情報を、その単語や文の意味を表現するベクトル空間にマッピング (embedding) し、大規模な文書集合から関連性の高い情報の検索を可能とする。
- 非構造データが画像データの場合には、OCR を適用し、1)画像の前処理 (リサイズ、回転、トリミング、ノイズの除去等) により、文字認識に適した状態に加工後、画像内の文字のセグメンテーションを実施。2)抽出結果から抽出した文字形状やサイズなどの特徴を取得し、機械学習により文字認識モデルから文字を識別。3)認識された文字にスペル修正やフォーマット調整後、テキストデータとして出力。
- ユーザーが生成したい構造データ構成 (カラム構成、データ型等の詳細条件) を指定すると、LLM に投入するプロンプトを自動生成し、プロンプトに従い、embeddings データから構造データを自動生成する。追加オプション機能により、ユーザーはプロンプトを拡張することができる。たとえば、integer (整数型) を 1 円単位とするか、100 万単位とするか、住所のハイフンの有無、氏名の間の半角スペースの有無などは標準プロンプトを拡張・制御が可能。
- 生成されるデータの再現性を確保するため、サンプリング等のデータ処理において seed および fingerprint 等のパラメータを使用することで、類似結果が生成されるように制御する。(ただし、並列計算や API のハードウェア性能によって理論上、完全な再現は不可能である。)
- データバリデーションのため、生成したデータがユーザーの指定したコンテンツ設定にしたがって、自動評価する仕組みを設ける。ハルシネーションを防ぐため、対象データソースごとに適切に生成されたかを定量的に示す信頼度 confidence (定義域: 0-1.0) を算出する。信頼度はアンサンプルで一つの構造データ化処理に 3 つ LLM エージェントによって出力された結果をクラスタリング・距離計算して算出される。対象データソースにおいて出力スキーマで指定した情報がある場合には 1.0 に近くなり、妥当な結果

であることを示す。逆に 0.0 に近いほど出力スキーマで指定した情報が対象データソース内になく、LLM によるハルシネーションが起きている可能性が高いことを示す。また、エージェントでは実行するプロンプト条件が変更され、通常 (default) に加え、下記のテーブル抽出型 ("table") とリスト抽出型 ("list") に分けて実行される。これにより、対象データソースの文章構造が特異なために抽出に失敗している場合に confidence が低下し、失敗している可能性があることを検出することが可能となる。

"table": ""For tables, pay extra attention to

- Identifying all tables in the document.
- Extracting headers and all data rows accurately.
- Interpreting merged cells and complex table structures correctly.
- Preserving relationships between table headers and data."

"list": ""For lists, pay extra attention to:

- Identifying all types of lists (bulleted, numbered, or otherwise).
- Capturing the hierarchy and structure of nested lists.
- Preserving the relationship between list items and any associated descriptions."

入力ファイル	類似度	正答有無
20231023110654.pdf	0.198855296	TRUE
20231023113228.pdf	0.184428781	FALSE
20231023131937.pdf	0.247462884	TRUE
20231023110523.pdf	0.249867022	TRUE
20231023114625.pdf	0.207226872	TRUE
20231023131843.pdf	0.318044901	TRUE
20231023115707.pdf	0.190660208	FALSE
20231023120019.pdf	0.208662719	TRUE
20231023114236.pdf	0.221941501	TRUE
20231023134253.pdf	0.269986451	TRUE
20231023114408.pdf	0.239943519	TRUE

図 4-22 データ生成における類似度

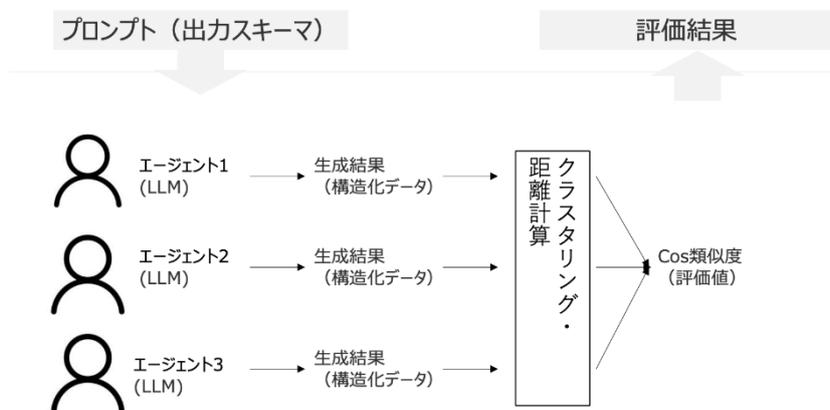


図 4-23 類似度算出方法

## AL04 【FN034】 チャット生成

LLM がユーザーの生成した最終構造化データをチャットから情報検索できるように、RAG (Retrieval-Augmented Generation) による回答生成を行えるようにする機能。

LLM はインターネット上に存在するドキュメントデータをクロール等により大量に収集し、学習データとして機械学習にかけたモデルであるため、デフォルトの状態ではインターネットに公開されていない省内の情報を加味したテキスト生成を行うことができない。そのため、ユーザーが生成した最終構造化データを加味した回答を可能とするため、ベクトル DB の生成、および、RAG (Retrieval-Augmented Generation) による検索拡張生成技術を取り入れた処理アルゴリズムを構築する。

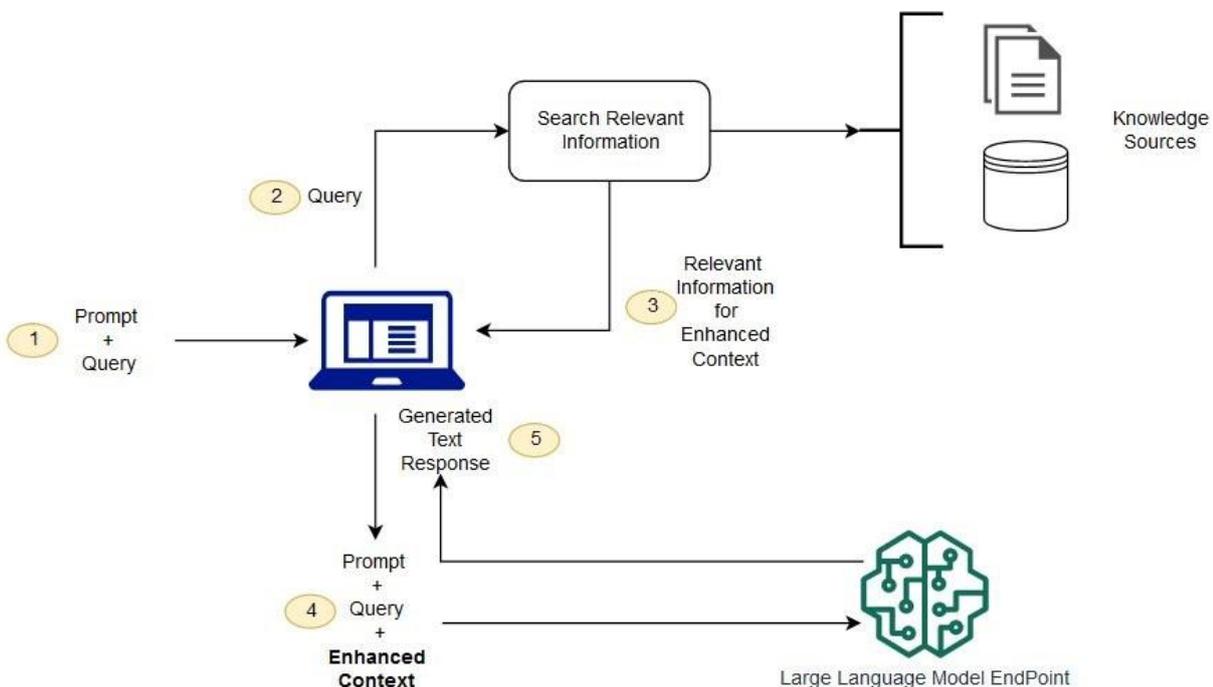


図 4-24 RAG (Retrieval-Augmented Generation) フロー図

(出典：[https://docs.aws.amazon.com/ja\\_jp/sagemaker/latest/dg/jumpstart-foundation-models-customize-rag.html](https://docs.aws.amazon.com/ja_jp/sagemaker/latest/dg/jumpstart-foundation-models-customize-rag.html))

処理の流れ：

- ①ベクトル DB の生成：クエリとデータベース内のドキュメントは、同じエンコーダ（通常は Transformer ベースのモデル）を使用してエンコードされ、ベクトルに変換される。
- ②クエリの生成：ユーザーからの入力（クエリ）を取得
- ③情報の検索：クエリを元に、事前に用意されたベクトル DB から関連情報を検索し、関連性の高い情報を取得。クエリのベクトルとデータベース内のドキュメントのベクトルを比較し、最も類似度の高い情報を抽出。
- ④生成モデルによる回答生成：取得した情報を入力として、ユーザーのクエリに対する自然な回答を生成。出典情報を返すことができる。

## 4-4. データインタフェース

### 4-4-1. ファイル入力インタフェース

表 4-12 入力データ一覧

ID	データを利用する機能	データ名	データの利用目的
DT001	FN006/FN007	原票データ (PDF、PNG 形式)	EBPM 及びオープンデータ化を実施するにあたり、データを結合するメインソースとして活用。
DT002	FN006/FN007	テキストデータ (DOCX、XLSX 形式)	EBPM 及びオープンデータ化を実施するにあたり、データを結合するメインソースとして活用。
DT003	FN006/FN007	地図データ (GeoJSON 形式、SHP 形式、CSV 形式)	メインソースデータと結合し、地図上で描画に活用。
DT004	FN006/FN007	テーブルデータ (CSV 形式)	メインソースデータと結合し、データ間の関係性や傾向等の発展的なデータ分析やグラフ等で視覚化するために活用。

#### 【DT001】原票データ

データを結合するメインソースとなる PDF 形式や PNG 形式のデータ。OCR を実行しテキスト化したのち構造データ化を行う。【DT001】原票データとしては、報告書や申請書、調査書などが想定される。最終的なデータの可視化を行う際に必要になる、名称や日時、座標、日時などの情報を含む。

#### 【DT002】テキストデータ

データを結合するメインソースとなる DOCX 形式、XLSX 形式のデータ。【AL01】LLM や【AL02】構造データ生成モデルを用いて構造データ化を行う。【DT002】テキストデータとしては、報告書や申請書、調査書などが想定される。最終的なデータの可視化を行う際に必要になる、名称や日時、座標、日時などの情報を含む。なお、XLSX 形式ですでに構造データ化されているデータもテキストデータとして入力される。

#### 【DT003】地図データ

メインソースデータと結合し、地図上で描画するための GeoJSON 形式、SHP 形式、CSV 形式のデータ。構造データ化されたデータのうち、ジオメトリを含んだデータが該当する。【DT003】地図データとして、国土

地理院の地図情報などが想定される。空間最終的にデータを地図上で表示させる際に必要になる情報を含む。

表 4-13 地図データのイメージ

code	subcode	year	Name	Value1	Value2	Value3	Value3	geometry
3001	411	7.8		2	0	88.22	1003	略
3001	411	7.4		2	0	67.67	1003	略
3001	411	7.5		2	0	55.27	1003	略
3003	Null	14.1		Null	1	Null	Null	略
3001	413	8		2	0	39.21	1003	略
:	:	:		:	:	:	:	:
任意	任意	任意		任意	任意	任意	任意	必須
数値	数値	数値		数値	ブーリアン	数値	カテゴリ	ジオメトリ

## 【DT004】テーブルデータ

メインソースデータと結合し、データ間の関係性や傾向等の発展的なデータ分析やグラフ等で視覚化するための情報を含む、CSV形式のデータ。【DT004】テーブルデータとしては、申請情報や統計調査などが想定される。

表 4-14 テーブルデータのイメージ

code	subcode	year	Name	Value1	Value2	Value3	Value3	geometry
3001	411	7.8		2	0	88.22	1003	21
3001	411	7.4		2	0	67.67	1003	21
3001	411	7.5		2	0	55.27	1003	21
3003	Null	14.1		Null	1	Null	Null	21
3001	413	8		2	0	39.21	1003	21
:	:	:		:	:	:	:	:
任意	任意	任意		任意	任意	任意	任意	任意
数値	数値	数値		数値	ブーリアン	数値	カテゴリ	数値

## 4-4-2. ファイル出力インターフェース

表 4-15 出力データ一覧

ID	データを出力する機能	データ名	形式
DT901	FN024, FN025, FN026, FN027	最終構造化データ	<u>CSV, JSON, GeoJSON</u>
DT902	FN20-3	ベクトルデータ	- (ベクトル DB にインポート)
DT903	FN024, FN025, FN026, FN027	オープンデータ	<u>JSON, GeoJSON</u>
DT904	FN003	コンテンツスキーマデータ	JSON
DT905	FN003	コンテンツデータ	JSON
DT906	FN030	属性情報出力データ	CSV
DT907	FN030	出力レポートデータ	PDF

## 【DT901】最終構造化データ

- 本インターフェースを利用する機能：【FN024】 【FN025】 【FN026】 【FN027】

【DT206】テキスト結合済構造化データ、【DT207】クロス集計済構造化データ、【DT208】空間結合済データ、【DT209】地域集計済構造化データのうち、最終的に出力されるデータ。

表 4-16 【DT901】最終構造化データの例

番号	フィールド名	データ型	項目	入力例	備考	出典
1	ID	整数型	ID			DT001/DT002
2	ファイル名	テキスト型	名称			DT001/DT002
3	PAGE	テキスト型	日時			DT001/DT002
4	TEXT	テキスト型	住所			DT001/DT002/DT003
5	GEOMETRY	ジオメトリ型	WKTフォーマット			DT001/DT002/DT003
6	INT_VALUE	整数型	値			DT001/DT002

7	TEXT_V ALUE	文字列	値			DT001/DT 002/DT00 3/DT004
8	BOOL_ VALUE	Bool	値			DT001/DT 002/DT00 3/DT004
9	REAL_V ALUE	小数点	値	...	...	DT001/DT 002/DT00 3/DT004
...	...	...	...	...	...	...
10	TEXT_V ALUE	文字列	値			DT001/DT 002/DT00 3/DT004
11	MATCH	Bool	完全一致の 有無			
12	SIMILA RITY	小数点	類似度		0-1.0	
13	OTHER _ONE	整数型	結合先 ID			DT001/DT 002
14	ADDRE SS2	テキス ト型	住所 2			DT001/DT 002/DT00 3
15	AREA	テキス ト型	地域			
...	...	...	...	...	...	...

## 【DT902】ベクトルデータ

- 本インターフェースを利用する機能： 【FN020\_3】ベクトル解析

【DT901】最終構造化データをベクトル解析によって、高次元空間上における数値ベクトルに変換したデータ。意味的に近い単語は互いに近い位置に配置され、遠い単語は遠くに配置される。【AL01】LLM において、本データによって入力テキストの意味と文脈を理解し、適切な応答を生成可能とするために作成される。

記載例：

ID ベクトル

1234 [0.12, -0.45, ..., 0.34]

567 [-0.22, 0.33, ..., -0.14]

89 [0.05, 0.19, ..., 0.27]

4321 [0.36, -0.12, ..., 0.18]

101 [-0.15, 0.27, ..., 0.22]

2345 [0.20, -0.29, ..., -0.08]

678 [-0.07, 0.40, ..., 0.15]

#### 【DT903】 オープンデータ

- 本インターフェースを利用する機能：【FN024】 【FN025】 【FN026】 【FN027】

【DT206】 テキスト結合済構造化データ、【DT207】 クロス集計済構造化データ、【DT208】 空間結合済データ、【DT209】 地域集計済構造化データのうち、公開用に最終的に出力されるデータ。【DT901】 最終構造化データから、非公開とすべきカラムを除外したデータ。

#### 【DT904】 コンテンツスキーマデータ

- 本インターフェースを利用する機能：【FN003】

【FN020\_2】 データ構造化処理、【FN021】 データ結合前処理（データクレンジング）、【FN022】 データ結合前処理（ジオコーディング）、【FN023】 データ結合前処理（座標系統一）、【FN024】 データ結合処理（テキストマッチング）、【FN025】 データ集計処理（クロス集計）、【FN026】 データ結合処理（空間結合）、【FN027】 データ集計処理（空間集計）の処理によって作成されたコンテンツのデータ定義情報を保持したデータ。

#### 【DT905】 コンテンツデータ

- 本インターフェースを利用する機能：【FN003】

【FN020\_2】 データ構造化処理、【FN021】 データ結合前処理（データクレンジング）、【FN022】 データ結合前処理（ジオコーディング）、【FN023】 データ結合前処理（座標系統一）、【FN024】 データ結合処理（テキストマッチング）、【FN025】 データ集計処理（クロス集計）、【FN026】 データ結合処理（空間結合）、【FN027】 データ集計処理（空間集計）の処理によって作成されたコンテンツのレコード情報を保持したデータ。

## 4-4-3. 内部連携インターフェース

表 4-17 内部連携 IF 一覧

内部連携 IF	内部連携 IF 名	ID	機能名	説明
IF201	データ登録 (CMS 連携)	FN004	アセット管理機能	● 本機能では、アップロードされたファイルの単位（ローデータ）であるアセットを管理する機能を提供する。
		FN006	データ登録機能	● 本機能は、アセットメニューから呼び出す。任意の静的ファイルをローカルから選択し、アセットとして新規アップロード、編集、削除する機能を提供する。
IF202	データ構造化 (CMS 連携)	FN003	コンテンツ管理機能	● 本機能では、D001、D002 の処理実行によって作成されるコンテンツ情報を管理する機能を提供する。
		FN005	アイテム管理機能	● 本機能は、CMS で管理するデータの最小単位で、コンテンツに則って作成されたデータのレコードを管理する機能を提供する。（Veda 内でのアクセスは行わない。）
IF203	データ構造化	FN020_1	OCR 処理	● OCR 処理を行い、画像データから文章を抽出し、テキストデータに変換する機能を提供する。
		FN020_2	データ構造化処理	● 構造データ生成モデルを利用し、LLM によって自然言語情報といった非構造データを構造データに変換する機能を提供する。
IF204	ベクトル解析	FN020_3	ベクトル解析	● ユーザーがチャットから登録したデータの内容を問い合わせができるように、ベクトル化処理を実施する機能を提供する。
IF205	データクレンジング	FN021	データ結合前処理（データクレンジング）	<ul style="list-style-type: none"> <li>● データ結合前にユーザーが指定したカラムを対象に表記ゆれの修正やデータクレンジング処理（スペースの削除や全角・半角の削除、テキストの置換処理など）を行う。</li> <li>● 表記ゆれの修正には、番地表記、市区町村名に含まれる小書き文字表記、数字表記の統一が含まれる。</li> </ul>
		FN022	データ結合前処理（ジオコーディング）	● 住所情報を保持するカラムに対してジオコーディング（座標付与）を実行し、新規に緯度経度のカラムを追加する機能を提供する。

				<ul style="list-style-type: none"> <li>● 後続の空間結合処理等で必要なカラム(緯度経度)を追加する。</li> </ul>
		FN023	データ結合前処理(座標系統一)	<ul style="list-style-type: none"> <li>● 登録データでは、複数の座標系が混在している状態である。後続でデータ結合するための内部処理としてWGS84に座標系統一処理を実行する。</li> </ul>
IF206	データ結合処理(テキストマッチング)	FN024	データ結合処理(テキストマッチング)	<ul style="list-style-type: none"> <li>● テキストマッチングにおいて、ユーザーが完全一致と部分一致を選択し、部分一致の際の類似度の閾値を設定することで、複数テーブルデータを属性結合する機能を提供する。</li> <li>● 結合対象データにおいて共通するキーとなる文字列を対象に一致を確認する。文字列が完全に一致する場合(完全一致)と、類似度を設けて部分的に一致する場合(部分一致)の2パターンを設ける。後者については、AI自然言語処理(Natural Language Processing)を用いて文字列間の類似度(ベクトル)を算出し、ユーザーが類似度に基づいて制御可能なものとする。・ユーザーは閾値、および完全一致、部分一致を選択し、処理の再実行もできる。</li> <li>● 任意のテーブルデータに対しテキストマッチングによるインデキシング処理を行う機能。この機能には特定のワードをキーとした結合、除外、確率計算等が含まれる。</li> <li>● テキストマッチングを利用した属性結合において、結合前後における総件数の表示、結合時に集約処理を行うかどうかを選択する機能を提供する。</li> </ul>
IF207	データ集計処理(クロス集計)	FN025	データ集計処理(クロス集計)	<ul style="list-style-type: none"> <li>● テキストマッチングで結合したデータをユーザーが指定した任意のカラムをキーに要約統計量(合計、カウント、平均)を算出する機能を提供する。</li> </ul>
IF208	データ結合処理(空間結合)	FN026	データ結合処理(空間結合)	<ul style="list-style-type: none"> <li>● 複数データを選択し、ポリゴンとポイントにおける交差結合をする機能を提供する。位置精度によるずれを防ぐため、最近傍結合も考慮する。</li> </ul>
IF209	データ結合処理(空間集計)	FN027	データ結合処理(空間集計)	<ul style="list-style-type: none"> <li>● ユーザーが指定した条件にのっとり(都道府県単位等)、ユーザーがアップロードした地域ポリゴンデータを結合し、地域単位で集計し、新規地図データとして保存する機能を提供する。この際、ポ</li> </ul>

				リゴンとポリゴンの交差判定を行い、複数のポリゴンにまたがる場合には割合が多いポリゴンへ集計されることとする。
IF210	チャット生成	FN033	ベクトルデータ作成機能 (ベクトル解析)	● Defined データに対して、FN20-3 を実行する機能。
		FN034	チャット生成機能	● FN032 のユーザーからのお問い合わせ入力内容を元に関連性のあるデータを FN034 で生成したベクトルデータから検索・取得し、質問と取得したデータを埋め込んだプロンプトを作成し、LLM によって FN032 で出力するユーザーへの返答文（自然言語）を生成する機能。
IF211	サジェッション機能	FN020_4	データ構造化前サジェッション機能	● データソースから出力スキーマ候補を提案し、ユーザーがデータ構造化処理の入力を支援する機能を設ける。
IF212	秘匿化処理	FN037	秘匿化処理	● データ公開に向けて、数値情報の階層化・偏差値化、新規コードの追加、住所の秘匿化を行い、元の情報を特定できないように秘匿化する機能を提供する。
IF213	英語カラム化	FN038	英語カラム化	● データ公開に向けて、LLM を用いて日本語カラムを英語カラムへと変換する機能を提供する。
IF214	メタデータ生成	FN039	RDF 生成	● データ仕様書のカタログ情報・データセット情報・データサービス情報・公開ファイル（配信情報）から RDF ファイル（ttl 形式）を作成する機能を提供する。

## 【IF201】データ登録（CMS 連携）

- 本インターフェースを利用する機能
  - 【FN004】アセット管理機能
  - 【FN006】データ登録機能
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST

- メソッド
  - POST
- パス
  - /projects/{projectId}/assets
- リクエストパラメータ

表 4-18 リクエストパラメータ

パラメータ	説明	値	必須
projectId	プロジェクト ID	string	○
file	アセットファイル	string(\$binary)	○
skipDecompression	解凍するかのオプション	boolean	

- リクエストパラメータモデル

図 4-25 リクエストパラメータモデル

```
{
  "projectId": "7341778-cc1d-4108-b294-8857ddb81",
  "file": "https://veda.domain/.....docx"
  "skipDecompression": false
}
```

- レスポンス

表 4-19 レスポンス(200)

パラメータ	説明	値	必須
id	ID	string	○
projectId	プロジェクト ID	string	○
name	アセット名	string	○
url	URL	string	○
contentType	コンテンツ型	string	○
previewType	プレビュー型	string	○
totalSize	トータルサイズ	number	○
archiveExtraction ステータス	抽出ステータス	string	○
ファイル名	ファイル名	string	○
file.size	ファイルサイズ	number	○
file.contentType	コンテンツ型	string	○

file.path	ファイルパス	string	○
file.children[]	ファイル子要素	string	○
createdAt	作成日	date	○
updatedAt	更新日	date	○

- レスポンスモデル

表 4-20 レスポンス(200)モデル

```
{
  "id": "7341778-cc1d-4108-b294-8857ddb81",
  "projectId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "name": "ファイル 1 ",
  "url": "https://veda.domain/.....",
  "contentType": "string",
  "previewType": "image",
  "totalSize": 100,
  "archiveExtraction ステータス": "pending",
  "file": {
    "name": "ファイル 1 ",
    "size": 100,
    "contentType": "string",
    "path": "https://veda.domain/.....",
    "children": []
  },
  "createdAt": "2024-09-18T00:59:07.492Z",
  "updatedAt": "2024-09-18T00:59:07.492Z"
}
```

## 【IF202】データ構造化 (CMS 連携)

- 本インターフェースを利用する機能
  - 【FN003】コンテンツ管理機能
  - 【FN005】アイテム管理機能
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド

- POST

- パス

- /projects/{projectIdOrAlias}/models

- リクエストパラメータ

表 4-21 リクエストパラメータ

パラメータ	説明	値	必須
projectIdOrAlias	プロジェクト ID	string	○
page	ページング	integer	
perPage	ページごとのアイテム数	integer	
name	名前	string	
description	概要	string	
key	キー	string	

- リクエストパラメータモデル

表 4-22 リクエストパラメータモデル

```
{
  "projectIdOrAlias": "7341778-cc1d-4108-b294-8857ddb81",
  "page": 1,
  "perPage": 50,
  "name": "モデル 1",
  "description": "モデル",
  "key": "model1"
}
```

- レスポンス

表 4-23 レスポンス(200)

パラメータ	説明	値	必須
id	名前	string	○
projectId	プロジェクト ID	string	○
schemald	コンテンツ ID	string	○
schema.id	コンテンツ ID	string	○
schema.projectId	プロジェクト ID	string	○
schema.fields[].id	フィールド ID	string	○
schema.fields[].type	フィールドの型	number	○

schema.fields[].key	フィールドのキー	string	○
schema.fields[].required	必須フラグ	boolean	○
schema.fields[].multiple	複数フラグ	boolean	○
schema.TitleField	コンテンツタイトル	string	○
schema.createdAt	コンテンツ作成日	date	○
metadataSchemald	メタデータ ID	string	○
MetadataSchema.id	メタデータ ID	string	○
MetadataSchema.projectId	メタデータプロジェクト ID	string	○
MetadataSchema.fields[].id	メタデータフィールド ID	string	○
MetadataSchema.fields[].type	メタデータフィールド型	string	○
MetadataSchema.fields[].key	メタデータフィールドのキー	string	○
MetadataSchema.fields[].required	メタデータフィールドの必須フラグ	boolean	○
MetadataSchema.fields[].multiple	メタデータフィールドの複数フラグ	boolean	○
MetadataSchema.TitleField	メタデータタイトル	string	○
MetadataSchema.createdAt	メタデータ作成日	date	○
name	名前	string	○
description	概要	string	○
public	公開フラグ	boolean	○
key	キー	string	○
createdAt	作成日	date	○
updatedAt	更新日	date	○
lastModified	最終修正日	date	○

- レスポンスモデル

表 4-24 レスポンス(200)モデル

```
{
  "id": "7341778-cc1d-4108-b294-8857ddb81",
  "projectId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "schemald": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "schema": {
```

```
"id": "7341778-cc1d-4108-b294-8857ddb81",
"projectId": "7341778-cc1d-4108-b294-8857ddb81",
"fields": [
  {
    "id": "7341778-cc1d-4108-b294-8857ddb81",
    "type": "text",
    "key": "text1",
    "required": true,
    "multiple": true,
  }
],
"TitleField": "title",
"createdAt": "2024-09-18T01:09:47.471Z"
},
"metadataSchemald": "7341778-cc1d-4108-b294-8857ddb81",
"metadataSchema": {
  "id": "7341778-cc1d-4108-b294-8857ddb81",
  "projectId": "7341778-cc1d-4108-b294-8857ddb81",
  "fields": [
    {
      "id": "7341778-cc1d-4108-b294-8857ddb81",
      "type": "text",
      "key": "text1",
      "required": true,
      "multiple": true,
    }
  ],
  "TitleField": "title",
  "createdAt": "2024-09-18T01:09:47.471Z"
},
"name": "name",
"description": "desc",
"public": true,
"key": "name1",
"createdAt": "2024-09-18T00:59:07.492Z",
```

```

"updatedAt": "2024-09-18T00:59:07.492Z"
"lastModified": "2024-09-18T00:59:07.492Z"
}

```

## 【IF203】データ構造化機能

- 本インターフェースを利用する機能
  - 【FN020\_1】OCR 処理
  - 【FN020\_2】データ構造化処理
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/structure
  
- リクエストパラメータ

表 4-25 リクエストパラメータ

パラメータ	説明	値	必須
mode	構造化で初期または更新リクエスト判定	"create"   "update"	○
files	アセットファイル配列	array	○
files.id	アセットファイル ID	string	○
files.url	アセットファイル URL	string	○
type_output	出力型指定（複数行：array）	"array"   "object"	
schema	構造化で作成されるコンテンツ	object	○
schema.type	コンテンツデータ型	"object"	○
schema.properties	コンテンツの属性一覧（取得するカラム情報）	object	○
schema.properties.[カラム名].type	カラム型	string	○
schema.properties.[カラム名].title	カラム名	string	○
schema.properties.[カラム名].description	カラム説明文	string	
schema.properties.[カラム名].position	ファイル内の章だてなど	string	

schema.properties.[カラム名].unit	ファイル内の章だてなど	string	
schema.properties.[カラム名].keyword	カラムの値に対するキーワード	string	
prompt	抽出指定文	string	○
genSourceName	出典情報配列	array	○
genSourceName.type	出典情報対象型	"column"   "text"	○
genSourceName.target	出典情報または指定カラム名	string	○
apiEndpoint	処理完了後にデータを送るエンドポイント	string	○

- リクエストパラメータモデル

表 4-26 リクエストパラメータモデル

```

{
  "mode": "create",
  "type_output*": "object",
  "files": [
    { "id": "01j7awhy0j3ffesxdm4krhyn0h", "url": "https://veda.domain/.....docx" },
    { "id": "01j7awhy0j3ffesxdm4krhyn1h", "url": "https://veda.domain/.....docx" },
    { "id": "01j7awhy0j3ffesxdm4krhyn2h", "url": "https://veda.domain/.....docx" }
  ],
  "schema": {
    "type": "object",
    "properties": { // ユーザーが出力コンテンツで指定
      "GDP": { "title": "GDP", "type": "string", "description": "GDP of city" },
      "area": { "title": "area", "type": "string", "description": "City 's area information" },
      "city-name": { "title": "city-name", "type": "string", "description": "Name of the city" },
      "population": { "title": "population", "type": "string", "description": "Population information
of the city" },
    }
  },
  "genSourceName": [ // 出典情報となるカラムを指定、ユーザーの入力情報
    {
      "type": "column",
      "target": "city-name",
    },
    {
      "type": "text",

```

```

    "target": " city 's data",
  }
],
"prompt": "この DOCX の詳細を抽出してください。",
"apiEndpoint": "https://veda.domain/callback"
}

```

- レスポンス

表 4-27 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
ticketId	チケット ID	string	○
message	メッセージ	string	
responseType	返却タイプ (json か geojson)	"json"   "geojson"	○
schema	構造化処理によって作成されたデータコンテンツ	object	○
schema.type	データ型	"object"	○
schema.propeprties	構造化処理によって作成された主データ	object	○
schema.propeprties.key	データキー	object	○
schema.propeprties.key.name	データ名	string	○
schema.propeprties.key.type	データ型	string	○
schema.propeprties.key._src_id	出典 ID	string	○
schema.propeprties.key._src_name	出典名	string	○
schema.propeprties.key._src_url	出典 URL	string	○

表 4-28 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
ticketId	チケット ID	string	○
message	メッセージ	string	○
responseType	返却タイプ (json か geojson)	"json"   "geojson"	○
schema	処理完了データのコンテンツ	object	○

schema.type	コンテンツデータ型	"object"	○
schema.properties	コンテンツ属性	object	○
schema.[カラム名].name	カラム名	string	○
Schema.[カラム名].type	値型	string	○

表 4-29 レスポンス(コールバック)

パラメータ	説明	値	必須
ticketId	チケット ID	string	○
fileId	ファイル ID	string	○
data	処理結果データ	object	○
data._src_id	出典 ID	string	○
data._src_name	出典名	string	○
data._src_url	出典 URL	string	○
data.<項目>	コンテンツの項目	string	○
dataConfidence	データの評価	object	○
dataConfidence.<項目>	データの評価値	number	○

- レスポンスモデル

表 4-30 レスポンス(200)モデル

```

{
  "ステータス": "ok",
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "message": "処理が成功しました。",
  "responseType": "json";
  "schema": {
    "type": "object",
    "propeprties": {
      "GDP": { "name": "GDP", "type": "string"},
      "area": { "name": "area", "type": "string" },
      "city-name": { "name": "city-name", "type": "string" },
      "population": { "name": "population", "type": "string" },
      "_src_id": { "name": "出典 ID", "type": "string" },
      "_src_name": { "name": "出典名", "type": "string" },
      "_src_url": { "name": "出典 URL", "type": "string" },
    }
  }
}

```

```
}

```

表 4-31 レスポンス(400)モデル

```
{
  "ステータス": "error",
  "message": "処理が失敗しました。"
}
```

表 4-32 レスポンス(コールバック)モデル

```
{
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "fileId": "01j7awhy0j3ffesxdm4krhyn0h",
  "data": {
    "GDP": "Accounts for roughly 30% of Japan's overall economy (in 2022)",
    "area": "2,191 square kilometers",
    "city-name": "Tokyo",
    "population": "14.2 million (as of 2023)",
    "_src_id": "01j7awhy0j3ffesxdm4krhyn0h",
    "_src_url": "https://veda.domain/.....docx",
    "_src_name": "Tokyo city 's data"
  },
  "dataConfidence": {
    "GDP": 1.0,
    "area": 1.0,
    "city-name": 1.0,
    "population": 1.0
  }
}
```

表 4-33 レスポンス(コールバック)モデル (1 ファイル=複数行)

```
{
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "fileId": "01j7awhy0j3ffesxdm4krhyn0h",
  "data": [
    {
      "GDP": "Accounts for roughly 30% of Japan's overall economy (in 2022)",
      "area": "2,191 square kilometers",

```

```
"city-name": "Tokyo",
"population": "14.2 million (as of 2023)",
"_src_id": "01j7awhy0j3ffesxdm4krhyn0h",
"_src_url": "https://veda.domain/.....docx",
"_src_name": "Japan 's data"
},
{
  "GDP": "Contributes significantly to the regional economy, with a strong presence in tourism,
culture, and technology sectors.",
  "area": "827.83 square kilometers",
  "city-name": "Kyoto",
  "population": "Approximately 1.46 million (as of 2023)",
  "_src_id": "01j7awhy0j3ffesxdm4krhyn0h",
  "_src_url": "https://veda.domain/.....docx",
  "_src_name": "Japan 's data"
},
],
"dataConfidence": [
  {
    "GDP": 1.0,
    "area": 1.0,
    "city-name": 1.0,
    "population": 1.0
  },
  {
    "GDP": 1.0,
    "area": 1.0,
    "city-name": 1.0,
    "population": 1.0
  }
]
}
```

## 【IF204】ベクトル解析

- 本インターフェースを利用する機能
  - 【FN020\_3】ベクトル解析
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/vectorize
- リクエストパラメータ

表 4-34 リクエストパラメータ

パラメータ	説明	値	必須
id	チャット API で呼び出す ID	string	○
input	JSON か GeoJSON の URL	string	○
apiEndpoint	処理完了後にデータを送るエンドポイント	string	○

- リクエストパラメータモデル

表 4-35 リクエストパラメータモデル

```

{
  "id": "01j7awhy0j3ffesxdm4krhyn0h",
  "input": "https://veda.domain/.....json",
  "apiEndpoint": "https://veda.domain/callback"
}

```

- レスポンス

表 4-36 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
ticketId	チケット ID	string	○
message	メッセージ	string	

表 4-37 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

表 4-38 レスポンス(コールバック)

パラメータ	説明	値	必須
ticketId	チケット ID	string	○
data	ベクトルデータ 情報	object	○

- レスポンスモデル

表 4-39 レスポンス(200)モデル

```
{
  "ステータス": "ok",
  "ticketId": " f7341778-cc1d-4108-b294-ff88857ddb81",
  "message": "処理が成功しました。",
  "responseType": "json";
}
```

表 4-40 レスポンス(400)モデル

```
{
  "ステータス": "error",
  "message": "失敗しました。"
}
```

表 4-41 レスポンス(コールバック)モデル

```
{
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "schema": {
    "type": "object",
    "properties": {
      "_id": { "title": "_id", "type": "string", "description": "UUID identifier for the document" },
      "text": { "title": "text", "type": "string", "description": "Data for vectorization" },
      "metadata": { "title": "metadata", "type": "object", "properties": {
        "file_id": { "title": "file_id", "type": "string", "description": "Unique identifier for the file" },
        "file_url": { "title": "file_url", "type": "string", "description": "URL of the file" } }
    }
  }
}
```

```

    }
  }
},
"data": [
  {
    "_id": "f7341778-cc1d-4108",
    "text": "{\"事故等名\":\"旅客船プリンス宗谷衝突(岸壁)\",\"事故等番号\":\"2008 函第 18 号\",\"発生場所\":\"北海道 泊港フェリー岸壁(利尻島) おしどまり\",\"発生年月日時刻\":\"平成 20 年 8 月 23 日 14 時 25 分ころ\", \"_src_id\":\"001\", \"_src_url\":\"https://veda.domain/linkdms-files/input/.....pdf\", \"_src_name\":\"旅客船プリンス宗谷衝突(岸壁)2008 函第 18 号\", \"_document_name\":\"事故報告\"}",
    "metadata": { "file_id": "001", "file_url": "https://veda.domain/linkdms-files/input/.....pdf" }
  },
  ...
]
}

```

#### 【IF205】データクレンジング

- 本インターフェースを利用する機能
  - 【FN021】データ結合前処理（データクレンジング）
  - 【FN022】データ結合前処理（ジオコーディング）
  - 【FN023】データ結合前処理（座標系統一）
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/preprocess
- リクエストパラメータ

表 4-42 リクエストパラメータ

パラメータ	説明	値	必須
input	データ URL	string	○
inputType	データ型	"json" "shapefile" "geojson" "csv"	○
cleansing	事前処理を行うためのオペレーションデ	object	

	ータ		
cleasing.type	オペレーション型	"replace" "normalize" "missing"	○
cleasing.fields	対象カラム	string[ ]	
cleasing.target	対象データ	string	
cleasing.replace	置き換えデータ	string	
cleasing.condition	削除データ	"comma"   ""   ""   ""	
normalizeCrs	座標系統一	boolean	
geocoding	ジオコーディング処理を行うためのカラム指定データ	object	
geocoding.fields	対象カラム	string[ ]	
documentName	資料名	string	○
apiEndpoint	処理完了後にデータを送るエンドポイント	string	○

- リクエストパラメータモデル

表 4-43 リクエストパラメータモデル

```
{
  "input": " https://veda.domain/.....json",
  "cleasing": [
    {
      "type": "normalize",
      "target": "浄水",
    },
    {
      "type": "replace",
      "fields": "住所",
      "target": "浄水",
      "replace": "浄水 12"
    }
  ],
  "normizeCrs": false,
  "geocoding": {
    "fields": ["住所"],
  },
}
```

```
"apiEndpoint": " https://veda.domain/callback"
}
```

- レスポンス

表 4-44 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
ticketId	チケット ID	string	○
message	メッセージ	string	
responseType	返却タイプ ( json か geojson)	"json"   "geojson"	○

表 4-45 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

表 4-46 レスポンス(コールバック)

パラメータ	説明	値	必須
ticketId	チケット ID	string	○
schema	コンテンツ	object	○
schema.type	データ型	string	○
schema.properties	属性	object	○
schema.properties. [カラム名].name	カラム名	string	○
schema.properties. [カラム名].type	カラム値型	string	○
data	処理結果	array	○

- レスポンスモデル

表 4-47 レスポンス(200)モデル

```
{
  "ステータス": "ok",
  "ticketId": " f7341778-cc1d-4108-b294-ff88857ddb81",
  "message": "処理が成功しました。",
  "responseType": "json";
}
```

表 4-48 レスポンス(400)モデル

```
{
  "ステータス": "error",
  "message": "...
}
```

表 4-49 レスポンス(コールバック)モデル

```
{
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "schema": {
    "type": "array",
    "properties": {
      "lat": {"name": "lat", "type": "number"},
      "lng": {"name": "lng", "type": "number"},
      "世帯コード": {"name": "世帯コード", "type": "number"},
      "住定異動年月日": {"name": "住定異動年月日", "type": "string"},
      "住所": {"name": "住所", "type": "string"},
      "性別": {"name": "性別", "type": "number"},
      "生年月日": {"name": "生年月日", "type": "string"},
    }
  },
  "data": [
    {
      "lat": 35.8710126,
      "lng": 139.6732722,
      "世帯コード": 43,
      "住定異動年月日": "1979/3/4",
      "住所": "浄水 12 町原山 309 番地 4",
      "性別": 1,
      "生年月日": "1999/11/1"
    }
  ]
}
```

## 【IF206】データ結合処理（テキストマッチング）

- 本インターフェースを利用する機能
  - 【FN024】データ結合処理（テキストマッチング）
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/text\_match
- リクエストパラメータ

表 4-50 リクエストパラメータ

パラメータ	説明	値	必須
inputLeft	対象データ URL	string	○
inputRight	対象データ URL	string	○
where	対象カラム指定データ	object	○
where. leftField	対象カラム	string	○
where. rightField	対象カラム	string	○
threshold	閾値	number	○
apiEndpoint	処理完了後にデータを送るエンドポイント	string	○

- リクエストパラメータモデル

表 4-51 リクエストパラメータモデル

```
{
  "inputLeft": "https://veda.domain/1.....json",
  "inputRight": "https://veda.domain/2.....json",
  "where": [
    {
      "leftField": "申請者 ID",
      "rightField": "申請者 ID"
    }
  ],
  "threshold": 0.6,
}
```

```

"apiEndpoint": "https://veda.domain/callback"
}

```

- レスポンス

表 4-52 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
ticketId	チケット ID	string	○
message	メッセージ	string	
responseType	返却タイプ	"json"	○

表 4-52 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

表 4-53 レスポンス(コールバック)

パラメータ	説明	値	必須
ticketId	チケット ID	string	○
schema	コンテンツ	object	○
schema.type	データ型	string	○
schema.properties	属性	object	○
schema.properties.[カラム名].name	カラム名	string	○
schema.properties.[カラム名].type	カラム値型	string	○
data	処理結果	object	○

- レスポンスモデル

表 4-54 レスポンス(200)モデル

```

{
  "ステータス": "ok",
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "message": "処理が成功しました。",
  "responseType": "json";
}

```

表 4-55 レスポンス(400)モデル

```
{
  "ステータス": "error",
  "message": "...
}
```

表 2 レスポンス(コールバック)モデル

```
{
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "schema": {
    "type": "object",
    "properties": {
      "match_rate": {
        "name": "match_rate",
        "type": "number"
      },
      "matched_data": {
        "name": "matched_data",
        "type": "array",
        "properties": {
          "番号": { "name": "番号", "type": "number" },
          "事故受付番号": { "name": "事故受付番号", "type": "string" },
          "事故等区分": { "name": "事故等区分", "type": "string" },
          "状態": { "name": "状態", "type": "string" },
          "申請者 ID": { "name": "申請者 ID", "type": "string" },
          "許可等番号": { "name": "許可等番号", "type": "string" },
          "飛行許可が必要な理由": { "name": "飛行許可が必要な理由", "type": "string" },
          "_src_id": { "name": "_src_id", "type": "string" },
          "_src_url": { "name": "_src_url", "type": "string" },
          "_src_name": { "name": "_src_name", "type": "string" },
          "_document_name": { "name": "_document_name", "type": "string" },
          "許可・承認の有無": { "name": "許可・承認の有無", "type": "string" },
          "型式_input2": { "name": "型式_input2", "type": "string" },
          "発生日時_input2": { "name": "発生日時_input2", "type": "string" },
          "発生場所_input2": { "name": "発生場所_input2", "type": "string" },
          "報告の概要_input2": { "name": "報告の概要_input2", "type": "string" },
          "損壊した物件の種類_input2": { "name": "損壊した物件の種類_input2", "type": "string" },
          "報告日_input2": { "name": "報告日_input2", "type": "string" },

```

```

    "名寄せ元情報_申請者 ID_input2": { "name": "名寄せ元情報_申請者 ID_input2", "type":
"string" },
    "input2_flag": { "name": "input2_flag", "type": "number" },
    "input1_flag": { "name": "input1_flag", "type": "number" }
  }
}
},
"data": {
  "match_rate": 0.5,
  "matched_data": [
    {
      "番号": 1,
      "事故受付番号": "I2407000025",
      "事故等区分": "2",
      "状態": "1",
      "申請者 ID": "IDQ928819",
      "許可等番号": "東空運航第 6129 号",
      "飛行許可が必要な理由": "DID、夜間、目視外、30m",
      "_src_id": "001",
      "_src_url": "https://veda.domain/linkdms-files/input/....pdf",
      "_src_name": "MATRICE300RTK",
      "_document_name": "事故報告",
      "許可・承認の有無": null,
      "型式_input2": "MATRICE300RTK",
      "発生日時_input2": "2024/7/21 14:00",
      "発生場所_input2": "秋田県湯沢市",
      "報告の概要_input2": "送電鉄塔撮影の為テストフライト実施中に操縦装置と映像伝送の電波確認
を実施し、良好だった為撮影を開始した。(離陸場所から撮影場所までは約 800m)¥n 鉄塔の下部を撮影中
に、無人航空機を肉眼で捉えられなくなり通信可能な範囲から逸脱させてしまった為、操縦装置と映像伝送
が遮断され操作不能となった。¥n 自動帰還装置 GO-HOME をすぐに発動させたが、離着陸地点に自動で帰
還せずに、その場でホバリングし続け、バッテリーがなくなり鉄塔付近に自動着陸し墜落したと思われる。
墜落した日から現在まで機体の捜索を行っているが、現在も機体を回収できていない。",
      "損壊した物件の種類_input2": "その他",
      "報告日_input2": "2024/7/24",
      "名寄せ元情報_申請者 ID_input2": "IDQ928819",
      "input2_flag": 1,
      "input1_flag": 1
    }
  ]
}

```

```
},
{
  "番号": 2,
  "事故受付番号": "I2407000024",
  "事故等区分": "1",
  "状態": "1",
  "申請者 ID": "YKN778709",
  "許可等番号": "東空運航第 39177 号",
  "飛行許可が必要な理由": null,
  "_src_id": "002",
  "_src_url": "https://veda.domain/linkdms-files/input/....pdf",
  "_src_name": "FAZERR(31)",
  "_document_name": "事故報告",
  "許可・承認の有無": null,
  "型式_input2": "FAZERR(31)",
  "発生日時_input2": "2024/7/20 9:10",
  "発生場所_input2": "大分県玖珠郡玖珠町大字戸畑 7188",
  "報告の概要_input2": "電柱間の支線(ワイヤー)の接触",
  "損壊した物件の種類_input2": "その他",
  "報告日_input2": "2024/7/24",
  "名寄せ元情報_申請者 ID_input2": "YKN778709",
  "input2_flag": 1,
  "input1_flag": 1
},
]
}
}
```

【IF207】データ集計処理（クロス集計）

- 本インターフェースを利用する機能
  - 【FN025】データ集計処理（クロス集計）
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST

- メソッド
  - POST
- パス
  - /api/cross
- リクエストパラメータ

表 4-57 リクエストパラメータ

パラメータ	説明	値	必須
input	データ URL	string	○
keyFields	主カラム	string[]	○
fields	集計対象	object	○
fields.name	集計対象カラム名	string	○
fields.sum	合計	boolean	○
fields.avg	平均	boolean	○
fields.cnt	カウント	boolean	○
apiEndpoint	処理完了後にデータを送るエンドポイント	string	○

- リクエストパラメータモデル

表 4-58 リクエストパラメータモデル

```
{
  "input": " https://veda.domain/.....json",
  "keyFields": ["Gender", "Country"],
  "fields": [
    {
      "name": "Occupation",
      "sum": false,
      "avg": false,
      "cnt": true
    },
    {
      "name": "YearsOfExperience",
      "sum": false,
      "avg": false,
      "cnt": true
    },
    {
      "name": "Age",
```

```

    "sum": true,
    "avg": true,
    "cnt": false
  },
  {
    "name": "Salary",
    "sum": true,
    "avg": true,
    "cnt": false
  }
],
"apiEndpoint": "https://veda.domain/callback"
}

```

- レスポンス

表 4-59 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
ticketId	チケット ID	string	○
message	メッセージ	string	○
responseType	返却タイプ	"json"   "geojson"	○

表 4-60 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

表 4-61 レスポンス(コールバック)

パラメータ	説明	値	必須
ticketId	チケット ID	string	○
schema	コンテンツ	object	○
schema.type	データ型	string	○
schema.properties	属性	object	○
schema.properties.[カラム名].name	カラム名	string	○
schema.properties.[カラム名].type	カラムデータ型	string	○
data	処理結果	object	○

● レスポンスモデル

表 4-62 レスポンス(200)モデル

```
{  
  "ステータス": "ok",  
  "ticketId": " f7341778-cc1d-4108-b294-ff88857ddb81",  
  "message": "処理が成功しました。",  
  "responseType": "json";  
}
```

表 4-63 レスポンス(400)モデル

```
{  
  "ステータス": "error",  
  "message": "失敗しました。"  
}
```

表 4-64 レスポンス(コールバック)モデル

```
{  
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",  
  "schema": {  
    "type": "object",  
    "properties": {  
      "countData": {  
        "name": "countData",  
        "type": "array",  
        "properties": {  
          "Gender": {  
            "name": "Gender",  
            "type": "string"  
          },  
          "Country": {  
            "name": "Country",  
            "type": "string"  
          },  
          "Age_sum": {  
            "name": "Age_sum",  
            "type": "number"  
          },  
        },  
      },  
    },  
  },  
}
```

```
"Age_avg": {
  "name": "Age_avg",
  "type": "number"
},
"Salary_sum": {
  "name": "Salary_sum",
  "type": "number"
},
"Salary_avg": {
  "name": "Salary_avg",
  "type": "number"
}
}
},
"crossTabData": {
  "name": "crossTabData",
  "type": "array",
  "properties": {
    "Gender": {
      "name": "Gender",
      "type": "string"
    },
    "Country": {
      "name": "Country",
      "type": "string"
    },
    "Artist_10 years": {
      "name": "Artist_10 years",
      "type": "number"
    },
    "Artist_2 years": {
      "name": "Artist_2 years",
      "type": "number"
    },
    "Doctor_10 years": {
      "name": "Doctor_10 years",
      "type": "number"
    }
  },
}
```

```
"Doctor_2 years": {
  "name": "Doctor_2 years",
  "type": "number"
},
"Doctor_8 years": {
  "name": "Doctor_8 years",
  "type": "number"
},
"Engineer_10 years": {
  "name": "Engineer_10 years",
  "type": "number"
},
"Engineer_2 years": {
  "name": "Engineer_2 years",
  "type": "number"
},
"Engineer_8 years": {
  "name": "Engineer_8 years",
  "type": "number"
},
"Total": {
  "name": "Total",
  "type": "number"
}
}
}
}
},
"data": {
  "countData": [
    {
      "Gender": "Female",
      "Country": "Angola",
      "Age_sum": 82,
      "Age_avg": 41,
      "Salary_sum": 158333,
      "Salary_avg": 79166.5
    }
  ],
}
```

```
{
  "Gender": "Male",
  "Country": "Angola",
  "Age_sum": 91,
  "Age_avg": 45.5,
  "Salary_sum": 81768,
  "Salary_avg": 40884
},
{
  "Gender": "Male",
  "Country": "Armenia",
  "Age_sum": 81,
  "Age_avg": 27,
  "Salary_sum": 198676,
  "Salary_avg": 66225.33333333333
},
...
],
"crossTabData": [
  {
    "Gender": "Female",
    "Country": "Angola",
    "Artist_10 years": 1,
    "Artist_2 years": 0,
    "Doctor_10 years": 0,
    "Doctor_2 years": 0,
    "Doctor_8 years": 0,
    "Engineer_10 years": 0,
    "Engineer_2 years": 1,
    "Engineer_8 years": 0,
    "Total": 2
  },
  {
    "Gender": "Female",
    "Country": "Armenia",
    "Artist_10 years": 0,
    "Artist_2 years": 0,
    "Doctor_10 years": 0,
```

```
"Doctor_2 years": 1,
"Doctor_8 years": 1,
"Engineer_10 years": 1,
"Engineer_2 years": 2,
"Engineer_8 years": 0,
"Total": 5
},
...
{
  "Gender": "Total",
  "Country": "",
  "Artist_10 years": 2,
  "Artist_2 years": 3,
  "Doctor_10 years": 3,
  "Doctor_2 years": 3,
  "Doctor_8 years": 2,
  "Engineer_10 years": 1,
  "Engineer_2 years": 5,
  "Engineer_8 years": 1,
  "Total": 20
}
]
}
}
```

**【IF208】 データ結合処理（空間結合）**

- 本インターフェースを利用する機能
  - **【FN026】 データ結合処理（空間結合）**
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/spatial\_join
- リクエストパラメータ

表 4-65 リクエストパラメータ

パラメータ	説明	値	必須
inputLeft	対象データ URL	string	○
inputRight	対象データ URL	string	○
op	オペレーション	"intersects" "nearest"	○
distance	距離 ( nearest で必要)	number	
apiEndpoint	処理完了後にデータを送るエンドポイント	string	○

- リクエストパラメータモデル

表 4-66 リクエストパラメータモデル

```
{
  "inputLeft": "https://veda.domain/1.....geojson",
  "inputRight": "https://veda.domain/2.....geojson",
  "op": "intersects",
  "apiEndpoint": "https://veda.domain/callback"
}
```

- レスポンス

表 4-67 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
ticketId	チケット ID	string	○
message	メッセージ	string	○
responseType	返却タイプ	"geojson"	○

表 4-68 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

表 4-69 レスポンス(コールバック)

パラメータ	説明	値	必須
ticketId	チケット ID	string	○
schema	コンテンツ	object	○

schema.type	データ型	string	○
schema.properties	属性	object	○
schema.properties.[カラム名].name	カラム名	string	○
schema.properties.[カラム名].type	カラムデータ 型	string	○
data	処理結果	object(geojson)	○

- レスポンスモデル

表 4-70 レスポンス(200)モデル

```

{
  "ステータス": "ok",
  "ticketId": " f7341778-cc1d-4108-b294-ff88857ddb81",
  "message": "処理が成功しました。",
  "responseType": "geojson"
}

```

表 4-71 レスポンス(400)モデル

```

{
  "ステータス": "error",
  "message": "失敗しました。"
}

```

表 4-72 レスポンス(コールバック)モデル

```

{
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "schema": {
    "type": "object",
    "properties": {
      "name": { "name": "name", "type": "string", },
      "tourism": { "name": "tourism", "type": "number" },
    }
  },
  "data": {
    "type": "FeatureCollection",
    "features": [
      {
        "id": "0",

```

```
"type": "Feature",
"properties": {
  "name": null,
  "tourism": "hotel"
},
"geometry": {
  "type": "MultiPolygon",
  "coordinates": [
    [
      [
        [102.0, 2.0, 10.0],
        [103.0, 2.0, 15.0],
        [103.0, 3.0, 12.0],
        [102.0, 3.0, 11.0],
        [102.0, 2.0, 10.0]
      ]
    ],
    [
      [
        [100.0, 0.0, 5.0],
        [101.0, 0.0, 6.0],
        [101.0, 1.0, 8.0],
        [100.0, 1.0, 7.0],
        [100.0, 0.0, 5.0]
      ]
    ],
    [
      [100.2, 0.2, 3.0],
      [100.8, 0.2, 4.0],
      [100.8, 0.8, 2.0],
      [100.2, 0.8, 1.0],
      [100.2, 0.2, 3.0]
    ]
  ]
}
}
```

```

    },
  ]
}
}

```

## 【IF209】 データ結合処理（空間集計）

- 本インターフェースを利用する機能
  - 【FN027】 データ結合処理（空間集計）
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/spatial\_aggregate
- リクエストパラメータ

表 4-73 リクエストパラメータ

パラメータ	説明	値	必須
inputLeft	対象データ URL	string	○
inputRight	対象データ URL	string	○
fields	集計対象	object	○
fields.name	集計対象カラム名	string	○
fields.sum	合計	boolean	○
fields.avg	平均	boolean	○
fields.cnt	カウント	boolean	○
apiEndpoint	処理完了後にデータを送るエンドポイント	string	○

- リクエストパラメータモデル

表 4-74 リクエストパラメータモデル

```

{
  "inputLeft": "https://veda.domain/1.....geojson",

```

```



```

- レスポンス

表 4-75 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
ticketId	チケット ID	string	○
message	メッセージ	string	
responseType	返却タイプ	"geojson"	○

表 4-76 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

表 4-77 レスポンス(コールバック)

パラメータ	説明	値	必須
ticketId	チケット ID	string	○
schema	コンテンツ	object	○
schema.type	データ型	string	○
schema.properties	属性	object	○
schema.properties.[カラム名].name	カラム名	string	○
schema.properties.[カラム名].type	カラムデータ型	string	○
data	処理結果	object(geojson)	○

● レスポンスモデル

表 4-78 レスポンス(200)モデル

```
{
  "ステータス": "ok",
  "ticketId": " f7341778-cc1d-4108-b294-ff88857ddb81",
  "message": "処理が成功しました。",
  "responseType": "geojson"
}
```

表 4-79 レスポンス(400)モデル

```
{
  "ステータス": "error",
  "message": "失敗しました。"
}
```

表 4-80 レスポンス(コールバック)モデル

```
{
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "schema": {
    "type": "object",
    "properties": {
      "都道府県": { "name": "都道府県", "type": "string", },
      "希望者数_sum": { "name": "希望者数_sum", "type": "number" },
      "希望者数_mean": { "name": "希望者数_mean", "type": "number" }
    }
  },
  "data": {
    "type": "FeatureCollection",
    "features": [
      {
        "id": "0",
        "type": "Feature",
        "properties": {
          "都道府県": "東京都",
          "希望者数_sum": 1215478,
          "希望者数_mean": 2114
        }
      },
    ],
  },
}
```

```
    "geometry": {
      "type": "Point",
      "coordinates": [...]
    }
  },
  {
    "id": "1",
    "type": "Feature",
    "properties": {
      "都道府県": "千葉県",
      "希望者数_sum": 5512,
      "希望者数_mean": 24
    },
    "geometry": {
      "type": "Point",
      "coordinates": [...]
    }
  },
]
}
```

#### 【IF210】 チャット生成

- 本インターフェースを利用する機能
  - 【FN034】 チャット生成機能
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/chat
- リクエストパラメータ

表 4-81 リクエストパラメータ

パラメータ	説明	値	必須
targetId	コンテンツ ID	string	○
prompt	問い合わせ	string	○
category	UC	string	○

- リクエストパラメータモデル

表 4-82 リクエストパラメータモデル

```
{
  "targetId": "01j441nj6hzzb441qdd26hg8q1",
  "prompt": "昨年の DMO 観光の KPI を教えてください",
  "category": "DMO 観光"
}
```

- レスポンス

表 4-83 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
message	メッセージ	string	○

表 4-84 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

- レスポンスモデル

表 4-85 レスポンス(200)モデル

```
{
  "ステータス": "ok",
  "answer": "回答文...¥n{資料名}{出典}",
  "message": "成功しました。"
}
```

表 4-86 レスポンス(400)モデル

```
{
  "ステータス": "error",
  "message": "..."}

```

## 【IF211】 サジェッション

- 本インターフェースを利用する機能
  - 【FN020\_4】 データ構造化前サジェッション機能
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/suggest-schema-structure
- リクエストパラメータ

表 4-87 リクエストパラメータ

パラメータ	説明	値	必須
input	サジェッション対象ファイル URL	string	○

- リクエストパラメータモデル

表 4-88 リクエストパラメータモデル

```
{
  "input": " https://veda.domain/.....pdf"}

```

- レスポンス

表 4-89 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
data	生成スキーマ	JSON schema	○
message	メッセージ	string	○

表 4-90 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

- レスポンスモデル

表 4-91 レスポンス(200)モデル

```

{
  "status": "ok",
  "data": {
    "type": "object",
    "properties": {
      "事故種類": {
        "type": "string",
        "description": "事故の種類を示す。"
      },
      "発生日時": {
        "type": "string",
        "description": "事故が発生した日時。"
      },
      "発生場所": {
        "type": "string",
        "description": "事故が発生した場所の詳細な説明。"
      },
      "事故調査の経過": {
        "type": "string",
        "description": "事故調査の進行状況や実施された手順。"
      },
      "船種船名": {
        "type": "string",
        "description": "事故に関与した船舶の種類と名前。"
      },
      "総トン数": {
        "type": "number",
        "description": "船舶の総トン数。"
      }
    }
  }
}

```

```

    },
    "船舶番号": {
      "type": "string",
      "description": "船舶の登録番号。"
    },
    },
    "船舶所有者等": {
      "type": "string",
      "description": "船舶の所有者に関する情報。"
    },
    },
    "L×B×D": {
      "type": "string",
      "description": "船舶の長さ、幅、深さの寸法。"
    },
    },
    ...
  }
}

```

表 4-92 レスポンス(400)モデル

```

{
  "ステータス": "error",
  "message": "処理に失敗しました。"
}

```

## 【IF212】 秘匿化処理

- 本インターフェースを利用する機能
  - 【FN037】 秘匿化処理機能
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/masking-data
- リクエストパラメータ

表 4-93 リクエストパラメータ

パラメータ	説明	値	必須
Input	インプットデータ URL	string	○
op	オペレーション一覧	array	○
op.type	オペレーション型	"masking_id" "masking_address" "ranking" "deviation_val"	○
op.field	対象カラム	string	○
op.max_rank	(ランク化) 最大数値	number	○
op.rankRanges	ランク化詳細設定	array	
op.rankRanges.min	各ランク値の最小値	number	
op.rankRanges.max	各ランク値の最大値	number	
op.prefix	(秘匿化 ID) 生成 ID 先頭 文字列	string	
apiEndpoint	処理完了後にデータを送 るエンドポイント	string	○

- リクエストパラメータモデル

表 4-94 リクエストパラメータモデル

```

{
  "input": " https://veda.domain/.....json",
  "op": [
    {
      "type": "masking_id",
      "field": "事業所番号"
    },
    {
      "type": "masking_address",
      "fields": "住所"
    },
    {
      "type": "ranking",
      "fields": "従業員数",
      "max_rank": 6
    },
    {

```

```

    "type": "deviation_val",
    "fields": "従業員数"
  }
],
"apiEndpoint": " https://veda.domain/callback"
}

```

- レスポンス

表 4-95 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
ticketId	チケット ID	string	○
message	メッセージ	string	
responseType	返却タイプ ( json か geojson)	"json"   "geojson"	○

表 4-96 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

- レスポンスモデル

表 4-97 レスポンス(200)モデル

```

{
  "ステータス": "ok",
  "ticketId": " f7341778-cc1d-4108-b294-ff88857ddb81",
  "message": "処理が成功しました。",
  "responseType": "json";
}

```

表 4-98 レスポンス(400)モデル

```

{
  "ステータス": "error",
  "message": "..."
}

```

表 4-99 レスポンス(コールバック)モデル

```

{
  "ticketId": "f7341778-cc1d-4108-b294-ff88857ddb81",
  "schema": {
    "type": "array",
    "properties": {
      "事業所名": {"name": "事業所名", "type": "string"},
      "事業所番号": {"name": "事業所番号", "type": "string"},
      "事業所番号 ID": {"name": "事業所番号 ID", "type": "string"},
      "住所": {"name": "住所", "type": "string"},
      "住所_秘匿化": {"name": "住所_秘匿化", "type": "string"}
    }
  },
  "data": [
    {
      "事業所名": "自動車事業所東京",
      "事業所番号": "001AA",
      "事業所番号 ID": "1a799f14-268a-4101-98b0-6694fba7f7a4",
      "住所": "東京都昭島市朝日町 3 - 5 - 5",
      "住所_秘匿化": "東京都昭島市朝日町",
      "従業員数": 32,
      "従業員数_ランク": 3,
      "従業員数_偏差値": 39
    },
    ...
  ]
}

```

## 【IF213】英語カラム化

- 本インターフェースを利用する機能
  - 【FN038】英語カラム化機能
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST

- メソッド
  - POST
- パス
  - /api/column-japanese-to-english
- リクエストパラメータ

表 4-100 リクエストパラメータ

パラメータ	説明	値	必須
columns	カラム情報一覧	array	○
columns.jp_name	日本語カラム名	string	○
columns.type	カラムのデータ型	string	
columns.description	カラムの説明文	string	

- リクエストパラメータモデル

表 4-101 リクエストパラメータモデル

```

{
  "columns": [
    {
      "jp_name": "必須",
      "description": "データの説明には値が必須ですか?"
    },
    {
      "jp_name": "内容",
      "description": ""
    },
    {
      "jp_name": "住所",
      "description": ""
    },
    {
      "jp_name": "人数",
      "description": ""
    },
    {
      "jp_name": "会社名"
    },
    {
      "jp_name": "銀行名"
    }
  ]
}

```

```

    }
  ]
}

```

- レスポンス

表 4-102 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
data	結果データ	array	○
data.jp_name	元カラム名 (日本語)	string	○
data.en_name	英語カラム名	string	○

表 4-103 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	"error"	○
message	メッセージ	string	○

- レスポンスモデル

表 4-104 レスポンス(200)モデル

```

{
  "data": [
    {
      "en_name": "required",
      "jp_name": "必須"
    },
    {
      "en_name": "content",
      "jp_name": "内容"
    },
    {
      "en_name": "location",
      "jp_name": "住所"
    },
    {
      "en_name": "number_of_people",
      "jp_name": "人数"
    }
  ]
}

```

```

    },
    {
      "en_name": "company_name",
      "jp_name": "会社名"
    },
    {
      "en_name": "bank_name",
      "jp_name": "銀行名"
    }
  ],
  "status": "ok"
}

```

表 4-105 レスポンス(400)モデル

```

{
  "ステータス": "error",
  "message": "..."
}

```

## 【IF214】メタデータ生成

- 本インターフェースを利用する機能
  - 【FN039】RDF 作成機能
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/rdf-create
- リクエストパラメータ

表 4-106 リクエストパラメータ

パラメータ	説明	値	必須
input	インプットエクセル URL	string	○

- リクエストパラメータモデル

表 4-107 リクエストパラメータモデル

```
{
  "input": " https://veda.domain/.....xlsx"
}
```

- レスポンス

表 4-108 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	“ok”	○
data	ttl 出力データ	string	○

表 4-109 レスポンス(400)

パラメータ	説明	値	必須
ステータス	状態	“error”	○
message	メッセージ	string	○

- レスポンスモデル

表 4-110 レスポンス(200)モデル

```
{
  "status": "ok",
  "data": "@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<https://drive.google.com/drive/folders/1A4_cFGs43GbcFfjJHdKvrHxYxZw-Aiiq?usp=drive_link>    a
dcat:Distribution ;
  dc:description "船舶事故等調査報告書から抽出した事故情報と当該事故発生時の海上分布予報の情報を結合したデータ" ;
  dc:format "Geojson" ;
  dc:identifier      <https://drive.google.com/drive/folders/1A4_cFGs43GbcFfjJHdKvrHxYxZw-Aiiq?usp=drive_link> ;
  dc:language "ja" ;
  dc:rights <https://www.mlit.go.jp/links/terms-of-use.html> ;
```

```
dc:title "02_船舶事故等調査報告書.geojson" ;
dcat:accessRights "配信中" ;
dcat:accessService "Google ドライブ" ;
dcat:accessURL "https://drive.google.com/drive/folders/1ssSwh1SpiwFiT3l-
qAfBX4zUn7KZaz_t?usp=drive_link" ;
dcat:byteSize "3M" ;
dcat:conformsTo <https://www.w3.org/TR/vocab-dcat-3/> ;
dcat:downloadURL "https://drive.google.com/file/d/12gF28camcaPuXDJAI06WAM-E6AV7N_y-
/view?usp=drive_link" ;
dcat:isReferencedBy <https://www.mlit.go.jp/links/> ;
dcat:issued "2024-11-23T00:00:00+00:00"^^xsd:dateTime ;
dcat:license "政府標準利用規約（第 2.0 版）" ;
dcat:mediaType "text/csv" ;
dcat:modified "2024-11-22T00:00:00+00:00"^^xsd:dateTime ;
dcat:temporalResolution "2024-11-23/2025-11-22" .
}
```

表 4-111 レスポンス(400)モデル

```
{
  "ステータス": "error",
  "message": "エクセルファイル形式が異常で処理できません。"
}
```

## 4-4-4. 外部連携インターフェース

表 4-112 外部連携 IF 一覧

外部連携 IF	外部連携 IF 名	ID	機能名	説明
IF301	G 空間情報センター連携（データセット登録・更新・削除）	FN019	データ配信機能	● 本機能では、G 空間情報センターAPI を利用しデータセット、リソースを配信する。
		FN036	データセット管理機能	● 本機能では、G 空間情報センターへのデータセット配信を制御する。
IF302	G 空間情報センター連携（リソース登録・更新・削除）	FN019	データ配信機能	● 本機能では、G 空間情報センターAPI を利用しデータセット、リソースを配信する。
		FN003	コンテンツ管理機能	● 本機能では、G 空間情報センターへのコンテンツ配信を制御する。

## 【IF301】 G 空間情報センター連携（データセット登録・更新・削除）

- 本インターフェースを利用する機能
  - 【FN019】 データ配信機能
  - 【FN036】 データセット管理機能
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/3/action/package\_create
- リクエストパラメータ

表 4-113 リクエストパラメータ

パラメータ	説明	値	必須
name	名前	string	○
title	タイトル	string	-
notes	説明	string	-
private	公開か非公開化	boolean	○
area	地理的範囲	string	-
author	作成者	string	-

maintainer	メンテナー	string	-
author_email	作成者の email	string	-
maintainer_email	メンテナーの email	string	-
license_id	ライセンス ID	string	-
owner_org	データセットを所有する 組織の ID	string	○

- リクエストパラメータモデル

表 4-114 リクエストパラメータモデル

```
{
  name: veda-{datasetName},
  title: veda-{datasetName},
  notes: {metadata md content},
  private: true,
  area: 北海道_札幌市,
  author: {username},
  maintainer: {username},
  author_email: {useremail},
  maintainer_email: {useremail},
  license_id: "CC-BY",
  owner_org: "links-veda"
}
```

- レスポンス

表 4-115 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
データ.id	データセットの ID	string	○

## 【IF302】 G 空間情報センター連携（リソース登録・更新・削除）

- 本インターフェースを利用する機能
  - 【FN019】 データ配信機能
  - 【FN003】 コンテンツ管理機能
- プロトコル
  - HTTPS
- アーキテクチャ
  - REST
- メソッド
  - POST
- パス
  - /api/3/action/resource\_create
- リクエストパラメータ

表 4-116 リクエストパラメータ

パラメータ	説明	値	必須
package_id	データセットの ID	string	○
url	url	string	○
format	説明	string	-
name	公開か非公開化	string	-

- リクエストパラメータモデル

表 4-117 リクエストパラメータモデル

```
{
  url: {assetUrl},
  format: zip,
  name: {contentName},
  package_id: {packageId}
}
```

- レスポンス

表 4-118 レスポンス(200)

パラメータ	説明	値	必須
ステータス	状態	"ok"	○
データ.id	リソースの ID	string	○

## 4-5. 実証に用いたデータ

## 4-5-1. 活用したデータ一覧

表 4-119 原課から提供が想定されるデータ一覧

ID	活用データ	形態	データ形式		利用想定
			IN	OUT	
DT101	報告書データ	【DT001】原票データ	pdf, png 等	xlsx,csv	事業情報、実績評価等を行うためのメインデータとして利用。
DT102		【DT002】テキストデータ	docx,xlsx ,csv 等	xlsx,csv	
DT103		【DT004】テーブルデータ	csv 等	csv	
DT104	統計データ	【DT001】原票データ	pdf, png 等	xlsx,csv	【DT101】報告書データに結合し、発展的な分析を行うために利用。
DT105		【DT002】テキストデータ	docx,xlsx ,csv 等	xlsx,csv	
DT106		【DT004】テーブルデータ	csv 等	csv	
DT107	時系列データ	【DT004】テーブルデータ	csv, json 等	csv, json	【DT101】報告書データに結合し、発展的な分析を行うために利用。
DT108	地点情報データ (ポイントデータ)	【DT003】地図データ	shp,	shp	事故発生地点、事務所所在地等で地図にプロット、空間処理を行う際の基礎データとして利用。

DT109	路線情報データ (ポリラインデータ)	【DT003】地図データ	shp	shp	航路ルート・道路ルート等で集計を行う際の基礎データとして利用。
DT110	地域情報データ (ポリゴンデータ)	【DT003】地図データ	shp	shp	都道府県、海域、管轄区域等で集計を行う際の基礎データとして利用。

## 1) 【DT101】報告書データ (原票データ)

報告書データのうち PDF 形式や PNG 形式である画像データ。【DT001】原票データとなるため、画像データで保存され、【FN020\_1】OCR 処理を施すことで、テキストデータ化できるデータが該当する。文章、表に加え、チェックボックスや丸囲みなどが入る場合もある。

## 船舶事故調査報告書

平成26年12月4日  
 運輸安全委員会 (海事専門部会) 議決  
 委員 庄司邦昭 (部会長)  
 委員 小須田 敏  
 委員 根本美奈

事故種類	衝突
発生日時	平成26年6月24日 19時45分ごろ
発生場所	長崎県長崎市大壱島南方沖 <small>長崎市所在の大壱島大瀬灯台から真方位180°4.5海里 (M) 付近</small> <small>(概位 北緯32°48.2′ 東経129°32.6′)</small>
事故調査の経過	平成26年6月25日、本事故の調査を担当する主管調査官 (長崎事務所) ほか1人の地方事故調査官を指名した。 原因関係者から意見聴取を行った。
事実情報	<p>船種船名、総トン数        船舶番号、船舶所有者等        L×B×D、船質        機関、出力、進水等</p> <p>A 貨物船 正輝丸、198トン        128596、有限会社松崎海運        51.89m×10.50m×5.30m、銅        ディーゼル機関、661kW、平成2年6月26日</p> <p>B 漁船 第十八恵比須丸、19トン        NS2-10539 (漁船登録番号)、株式会社勉水産        19.15m (Lr) × 4.62m × 1.78m、FRP        ディーゼル機関、736kW、平成5年11月6日  <small>第2000-17321号 (船舶検査簿の番号)</small></p>

図 4-26 報告書データ (png 形式、pdf 形式) の例

## 2) 【DT102】 報告書データ (テキストデータ)

報告書データのうち、DOCX 形式や XLSX 形式のテキストデータ。【D002】 テキストデータとなり、構造化されていないため、直接集計や分析ができない状態のデータが該当する。

## 第3章 基本理念と将来像

### 3.1 登録事業者基本情報

事業者名	株式会社 山田観光開発
代表者名	山田 太郎
所在地	〒123-4567 ○○県△△市中央通り2-15-8 △△ビル5F
電話番号	0123-45-6789
FAX番号	0123-45-6780
設立年月	平成12年4月
従業員数	48名(うちパート・アルバイト 15名)
事業種別	宿泊業・飲食業・交通業・旅行業・物販業・その他
営業区域	△△市内・□□町内・◇◇村内・地域全域
DMO会員	正会員・準会員・非会員

図 4-27 報告書データ (テキストデータ) の例

## 3) 【DT103】 報告書データ (テーブルデータ)

報告書データのうち CSV 形式などで構造データ化されているデータ。【DT004】 テーブルデータとなっているため、集計や分析ができる状態で保存されているものが該当する。

コード	局名	支局名	業種別	現行事業者名
110000077	北海道	札幌	一般	
110000350	北海道	札幌	特積み	
110000547	北海道	札幌	一般	
110000607	北海道	札幌	一般	
110000539	北海道	札幌	一般	
110000796	北海道	札幌	一般	
110000927	北海道	札幌	特積み	
110000979	北海道	札幌	特積み	
120000192	北海道	旭川	特積み	
130000151	北海道	函館	特積み	
130000049	北海道	函館	一般	
130000088	北海道	函館	一般	
130000183	北海道	函館	一般	
140000040	北海道	室蘭	一般	
140000291	北海道	室蘭	一般	
160000190	北海道	帯広	一般	
160000214	北海道	帯広	一般	

図 4-28 報告書データ (テーブルデータ) の例

4) 【DT104】統計データ（原票データ）

統計データのうち PDF 形式や PNG 形式である画像データ。【DT001】原票データとなるため、【FN020\_1】OCR 処理を施すことで、テーブルデータに変換され、集計や分析ができるフォーマットが該当する。

令和6年度 地域別産業統計調査報告書

作成日：2025年2月19日 | 調査期間：2024年4月～2025年3月 | サンプル数：N=12,450

■ 表1 地域別・産業分類別 事業所数および従業者数

地域	地方	都道府県	第一次産業			第二次産業			第三次産業			合計	
			事業所数	従業者数 (千人)	構成比 (%)	事業所数	従業者数 (千人)	構成比 (%)	事業所数	従業者数 (千人)	構成比 (%)	事業所数	従業者数 (千人)
北海道・東北	北海道		4,218	89.3	12.4	8,432	245.6	34.2	15,890	383.1	53.4	28,540	718.0
	青森県		2,145	42.1	14.8	4,567	112.3	39.5	5,678	138.0	45.7	12,390	284.4
	宮城県		1,890	35.7	8.2	6,234	189.4	43.7	10,456	288.3	48.1	18,580	433.4
	小計		8,253	167.1	11.5	19,233	547.3	37.6	32,024	721.4	49.6	59,510	1,435.8
関東	東京都		1,234	28.9	2.1	32,456	1,245.8	42.5	78,901	1,623.7	55.4	112,591	2,896.4
	神奈川県		987	21.3	3.0	18,765	678.2	44.8	25,432	789.5	52.2	45,184	1,489.0
	埼玉県		1,567	33.4	4.5	14,321	534.6	45.3	18,765	592.0	50.2	34,653	1,168.0
	千葉県		1,890	38.7	5.2	12,456	467.3	43.1	16,789	561.0	51.7	31,135	1,067.0
	小計		5,678	122.3	3.4	77,998	2,925.9	43.8	139,887	3,566.2	52.8	223,563	6,614.4
中部	愛知県		2,345	48.9	5.8	22,567	898.4	48.2	28,901	849.7	46.0	53,813	1,789.0
	静岡県		3,456	67.8	9.2	15,678	534.2	45.6	18,234	528.0	45.2	37,368	1,138.0
	小計		5,801	116.7	7.3	38,245	1,424.6	47.0	47,135	1,377.7	45.7	91,181	2,919.0
近畿	大阪府		1,123	24.5	2.8	24,567	934.2	44.1	34,567	1,123.3	53.1	60,257	2,082.0
	京都府		1,678	34.2	6.1	8,901	312.4	42.3	14,567	388.4	51.6	25,146	727.0
	小計		2,801	58.7	4.1	33,468	1,246.6	43.5	49,134	1,583.7	52.4	85,403	2,809.0
九州・沖縄	福岡県		2,567	52.3	7.4	14,567	478.9	42.8	22,345	556.8	49.8	39,479	1,088.0
	沖縄県		1,890	38.9	11.2	5,678	167.3	38.5	9,876	218.8	58.3	17,444	425.0
	小計		4,457	91.2	8.9	20,245	646.2	41.2	32,221	775.6	49.9	56,923	1,513.0
	全国合計		26,990	556.0	5.8	189,189	6,798.6	43.5	380,481	7,944.6	50.7	516,580	15,291.2

■ 表2 主要経済指標（前年度比較）

指標	単位	令和5年度	令和6年度	増減	前年比(%)	全国順位
域内総生産(GDP)	兆円	548.2	562.7	+14.5	102.6	—
一人当たり所得	万円	318.4	326.1	+7.7	102.4	—
完全失業率	%	2.6	2.4	-0.2	92.3	—
有効求人倍率	倍	1.28	1.35	+0.07	105.5	—
消費者物価指数	2020=100	105.8	108.3	+2.5	102.4	—
鉱工業生産指数	2020=100	98.7	101.2	+2.5	102.5	—
新設住宅着工戸数	万戸	81.9	78.4	-3.5	95.7	—
貿易収支	兆円	-6.8	-3.2	+3.6	—	—

※ 本データは全てダミーデータであり、実在の統計値とは一切関係ありません。  
 ※ 「構成比」は各地域における産業別従業者数の割合を示す。四捨五入のため合計が100%にならない場合がある。  
 ※ 出典：国土統計局「令和6年度地域別産業統計調査」（サンプルデータ）

図 4-29 統計データ（原票データ）の例

5) 【DT105】統計データ（テキストデータ）

統計データのうち、DOCX 形式や XLSX 形式のテキストデータ。【D002】テキストデータとなり、調査統計や報告書などテキストデータで保存されているものが該当する。

地方区分	都市名	人口			世帯		年齢別構成			人口移動				
		総人口	男女別		世帯数	平均	構成比(%)		指数	社会移動(千人)			自然	
		(千人)	男性	女性	(千世帯)	世帯人員	年少	生産年齢	老年	高齢化指数	転入	転出	社会増減	(千人)
北海道・東北	札幌市	1,975	926	1,049	982	2.01	10.8	68.2	29.0	268.5	+48.3	-41.7	+6.6	-8.2
	仙台市	1,097	534	563	523	2.10	11.5	62.1	26.4	229.6	+32.1	-28.9	+3.2	-3.1
	北海道・東北 小計	3,072	1,460	1,612	1,505	—	—	—	—	—	+80.4	-70.6	+9.8	-11.3
関東	さいたま市	1,348	667	673	589	2.27	12.8	63.5	23.7	185.2	+38.5	-31.2	+7.3	-1.4
	千葉市	982	487	495	434	2.26	11.9	62.0	26.1	219.3	+22.4	-20.1	+2.3	-2.8
	東京23区	9,784	4,821	4,963	5,234	1.87	11.2	66.8	22.0	196.4	+312.5	-245.8	+66.7	-5.3
	横浜市	3,749	1,862	1,887	1,694	2.21	12.1	63.1	24.8	205.0	+89.2	-78.4	+10.8	-6.1
	川崎市	1,542	785	757	741	2.08	13.0	66.2	20.8	160.0	+51.3	-42.1	+9.2	+0.4
	関東 小計	17,397	8,622	8,775	8,692	—	—	—	—	—	+513.9	-417.6	+96.3	-15.2
中部	名古屋市	2,321	1,148	1,173	1,087	2.14	11.7	62.5	25.8	228.5	+52.8	-49.3	+3.5	-4.7
	静岡市	689	335	354	294	2.34	11.0	59.8	29.2	265.5	+12.4	-14.8	-2.4	-4.1
	中部 小計	3,010	1,483	1,527	1,381	—	—	—	—	—	+65.2	-64.1	+1.1	-8.8
近畿	京都市	1,464	700	764	712	2.06	10.5	61.3	28.2	268.6	+38.7	-40.2	-1.5	-5.0
	大阪市	2,765	1,367	1,398	1,456	1.90	10.9	63.4	25.7	235.8	+78.4	-69.1	+9.3	-7.8
	神戸市	1,512	718	794	698	2.17	11.0	60.8	28.2	256.4	+28.9	-32.4	-3.5	-5.9

図 4-30 統計データ（テキストデータ）の例

6) 【DT106】 統計データ（テーブルデータ）

統計データのうちや CSV 形式などで構造データ化されているデータ。【DT004】 テーブルデータとなるため、集計や分析ができる状態で保存されているものが該当する。

1 令和2年国勢調査 小地域集計 (総務省統計局)														
2 第16-1表 男女、従業地・通学地別就業者・通学者数 (15歳以上) - 町丁・字等														
3														
4 男女	市区町村	町丁字	地域階層	秘匿処理	秘匿先情報	合算地域	都道府県	市区町村	大字・町	字・丁目	総数 (常)	自市区町村	自宅で従	自宅外の
5 総数	1101	-	1				北海道	札幌市中央区			113331	77901	7778	70123
6 総数	1101	200	2	秘匿地域	1000		北海道	札幌市中宮ケ丘 (番地)	X	X	X	X		
7 総数	1101	300	2				北海道	札幌市中宮丸山	-	-	-	-		
8 総数	1101	400	2				北海道	札幌市中宮丸山西町 (番地)	-	-	-	-		
9 総数	1101	1000	2	合算地域あり		200	北海道	札幌市中宮の森			25	20	6	14
10 総数	1101	1100	2				北海道	札幌市中宮盤溪			49	34	14	20
11 総数	1101	1201	3				北海道	札幌市中宮の森一条			2078	1364	177	1187
12 総数	1101	120101	4				北海道	札幌市中宮の森一丁目			47	37	2	35
13 総数	1101	120102	4				北海道	札幌市中宮の森二丁目			42	36	4	32
14 総数	1101	120103	4				北海道	札幌市中宮の森三丁目			97	66	5	61
15 総数	1101	120104	4				北海道	札幌市中宮の森四丁目			78	53	2	51
16 総数	1101	120105	4				北海道	札幌市中宮の森五丁目			60	35	4	31
17 総数	1101	120106	4				北海道	札幌市中宮の森六丁目			349	239	24	215
18 総数	1101	120107	4				北海道	札幌市中宮の森七丁目			128	84	17	67
19 総数	1101	120108	4				北海道	札幌市中宮の森八丁目			120	72	9	63
20 総数	1101	120109	4				北海道	札幌市中宮の森九丁目			183	103	8	95
21 総数	1101	120110	4				北海道	札幌市中宮の森十丁目			293	186	41	145

図 4-31 統計データ（テーブルデータ）の例

7) 【DT107】 気象データ（テーブルデータ）

CSV 形式や JSON 形式で保存されている気象データが該当する。事故が起きた日時の気象情報などを示すときに利用される。

東京	東京	東京	東京	東京	東京	東京	東京	東京	東京	東京							
気温(°C)	気温(°C)	気温(°C)	降水量(m)	降水量(m)	降水量(m)	降水量(m)	風速(m/s)	風速(m/s)	風速(m/s)	風速(m/s)	風速(m/s)	風速(m/s)	天気	天気	天気	視程(km)	視程(km)
品質情報		均質番号	現象なし		品質情報		均質番号	品質情報		品質情報		均質番号	品質情報		均質番号	品質情報	
7	8	1	2	0	8	1	2.8	8	北北東	8	1		0	1			0
6.7	8	1	4	0	8	1	1.9	8	北北西	8	1		0	1			0
6	8	1	4	0	8	1	2	8	北北西	8	1	10	8	1	5		8
5.4	8	1	3.5	0	8	1	4	8	北北西	8	1		0	1			0
4.4	8	1	2	0	8	1	4.2	8	北西	8	1		0	1			0
3.8	8	1	0	0	8	1	5.8	8	北西	8	1	10	8	1	8		8
4.2	8	1	0	0	8	1	5.3	8	北西	8	1		0	1			0
5	8	1	0	0	8	1	5.5	8	北西	8	1		0	1			0
6.3	8	1	0	0	8	1	3.3	8	北北西	8	1	4	8	1	25		8
8.7	8	1	0	1	8	1	2.3	8	北北西	8	1		0	1			0
12.2	8	1	0	1	8	1	5.6	8	北西	8	1		0	1			0
14.5	8	1	0	1	8	1	5	8	北西	8	1	1	8	1	30		8
16	8	1	0	1	8	1	4.4	8	北西	8	1		0	1			0
16	8	1	0	1	8	1	3.2	8	北北西	8	1		0	1			0
15.7	8	1	0	1	8	1	2.4	8	北	8	1	2	8	1	40		8
14.6	8	1	0	1	8	1	2.5	8	北北西	8	1		0	1			0
13.2	8	1	0	1	8	1	3.2	8	北西	8	1		0	1			0
11.4	8	1	0	1	8	1	0.6	8	西北西	8	1	3	8	1	30		8

図 4-32 気象データ（テーブルデータ）の例

8) 【DT108】 地点情報データ

SHP 形式等でポイントのジオメトリ情報を保有するデータ。【DT003】 地図データとなるため、GIS ソフト等での読み込みが可能なテーブルデータが該当する。

	P11_001	P11_002	P11_003_1	P11_004_1	P11_004_2	P11_004_3	P11_004_4
1	阿佐谷南一	1,1,1,1,2,3,1		阿02,阿05,洪66,...	NULL	NULL	NULL
2	幡ヶ谷原町	1,1,1,1,1,2,1		洪63,洪66,洪67,...	NULL	NULL	NULL
3	大宮町	1,1,1,1,1,1		高45,新02,高45,...	NULL	NULL	NULL
4	西永福	1,1,1,1,1,1,1,1		高45,新02,高45,...	NULL	NULL	NULL
5	天沼	1,1,1,1		阿02,阿05,荻15,...	NULL	NULL	NULL
6	笹塚二	1,1,1,2,1		洪66,洪67,宿44,...	NULL	NULL	NULL
7	南台三	1,1,1,1		洪63,宿41,宿45,...	NULL	NULL	NULL
8	杉並都税事務所	1,1,1,1,1		阿02,阿05,荻15,...	NULL	NULL	NULL
9	富士高	1,1,1,1		洪63,宿41,宿45,...	NULL	NULL	NULL
10	新山小	1,1,1,1		洪63,宿41,宿45,...	NULL	NULL	NULL
11	南中野地域セン...	1,1,1,1,1,1,1		洪63,宿32,宿33,...	NULL	NULL	NULL
12	代田橋	1,2,2,1		洪66,洪66,宿91,...	NULL	NULL	NULL
13	笹塚中	1,1,1,3,1		洪63,宿41,宿45,...	NULL	NULL	NULL
14	富士見丘学園	1,1,1		洪67,洪68,中81	NULL	NULL	NULL
15	笹塚三	1,1,1		洪67,洪68,中81	NULL	NULL	NULL
16	阿佐ヶ谷南三	1,1,1,1,1		阿02,阿05,荻15,...	NULL	NULL	NULL
17	阿佐ヶ谷駅	1,1,2,3,1		洪66,荻15,洪66,...	NULL	NULL	NULL
18	荻窪公園アパート前	1,1,1,1,1		阿02,阿05,荻15,...	NULL	NULL	NULL

図 4-33 地点情報データの例

9) 【DT109】 路線情報データ

SHP 形式等でポリラインのジオメトリ情報を保有するデータ。【DT003】 地図データとなるため、GIS ソフト等での読み込みが可能なテーブルデータが該当する。

	N09_005	N09_006	N09_007	N09_008	N09_009	N09_010	N09_011	N09_012	N09_013	N09_014	N09_015
1	001	東京～八丈島	東京～八丈島	302		00101	13002	NULL	東京	13015	三宅島
2	001	東京～八丈島	東京～八丈島	302		00102	13015	三宅島	三宅島	13016	御蔵島
3	001	東京～八丈島	東京～八丈島	302		00103	13016	御蔵島	御蔵島	13017	八丈島
4	003	東京～大島～神...	東京～神津島	207		00301	13002	NULL	東京	14001	NULL
5	003	東京～大島～神...	東京～神津島	207		00302	13002	NULL	東京	13005	大島
6	003	東京～大島～神...	東京～神津島	207		00303	14001	NULL	横浜	13005	大島
7	003	東京～大島～神...	東京～神津島	207		00304	13005	大島	大島	13006	利島
8	003	東京～大島～神...	東京～神津島	207		00305	13006	利島	利島	13009	新島
9	003	東京～大島～神...	東京～神津島	207		00306	13009	新島	新島	13011	式根島
10	003	東京～大島～神...	東京～神津島	207		00307	13011	式根島	式根島	13013	神津島
11	004	熱海～大島	熱海～大島	51		00401	22001	NULL	熱海	13005	大島
12	005	大島～下田	下田～大島	43		00501	22015	NULL	下田	13005	大島
13	006	神津島～熱海	神津島～熱海	112		00601	22001	NULL	熱海	13011	式根島
14	006	神津島～熱海	神津島～熱海	112		00602	22001	NULL	熱海	13009	新島
15	006	神津島～熱海	神津島～熱海	112		00603	13009	新島	新島	13013	神津島
16	007	名古屋～仙台～...	名古屋～苫小牧	1330		00701	23001	NULL	名古屋	04017	NULL

図 4-34 路線情報データの例

## 10) 【DT110】地域情報データ

SHP 形式等でポリゴンのジオメトリ情報を保有するデータ。【DT003】地図データとなるため、GIS ソフト等での読み込みが可能なテーブルデータが該当する。

	N03_001	N03_002	N03_003	N03_004	N03_005	N03_007
1	東京都	NULL	NULL	所属未定地	NULL	13000
2	東京都	NULL	NULL	所属未定地	NULL	13000
3	東京都	NULL	NULL	所属未定地	NULL	13000
4	東京都	NULL	NULL	所属未定地	NULL	13000
5	東京都	NULL	NULL	所属未定地	NULL	13000
6	東京都	NULL	NULL	所属未定地	NULL	13000
7	東京都	NULL	NULL	所属未定地	NULL	13000
8	東京都	NULL	NULL	所属未定地	NULL	13000
9	東京都	NULL	NULL	所属未定地	NULL	13000
10	東京都	NULL	NULL	所属未定地	NULL	13000
11	東京都	NULL	NULL	所属未定地	NULL	13000
12	東京都	NULL	NULL	所属未定地	NULL	13000
13	東京都	NULL	NULL	所属未定地	NULL	13000
14	東京都	NULL	NULL	所属未定地	NULL	13000

図 4-35 地域情報データの例

## 4-5-2. 生成・変換したデータ一覧

表 4-120 生成・変換したデータ一覧

ID	システムに入力するデータ (データ形式)	用途	処理内容	活用データ (データ形式)
DT201	正規化地図データテーブル (CSV形式)	視覚化 (地図描画)	<ul style="list-style-type: none"> <li>● メインソースデータと結合し、地図上で描画に活用。</li> </ul>	正規化地図データテーブル (CSV形式)
DT202	正規化テーブルデータ (CSV形式)	分析・視覚化	<ul style="list-style-type: none"> <li>● メインソースデータと結合し、データ間の関係性や傾向等の発展的なデータ分析やグラフ等で視覚化するために活用。</li> </ul>	正規化テーブルデータ (CSV形式)
DT203	OCR 処理済テキストデータ (JSON形式)	メインデータソース	<ul style="list-style-type: none"> <li>● EBPM 及びオープンデータ化を実施するにあたり、データを結合するメインソースとして活用。</li> </ul>	OCR 処理済テキストデータ (CSV形式)
DT204	構造データ化済テキストデータ (CSV形式)	メインデータソース	<ul style="list-style-type: none"> <li>● EBPM 及びオープンデータ化を実施するにあたり、データを結合するメインソースとして活用。</li> </ul>	構造データ化済テキストデータ (CSV形式)
DT205	テキスト結合済構造化データ (CSV形式)	メインデータソース	<ul style="list-style-type: none"> <li>● EBPM 及びオープンデータ化を実施するにあたり、データを結合するメインソースとして活用。</li> </ul>	テキスト結合済構造化データ (CSV形式)

DT206	空間結合済データ (CSV 形式)	メインデータソース	● EBPM 及びオープンデータ化を実施するにあたり、データを結合するメインソースとして活用。	空間結合済データ (CSV 形式)
DT207	地域集計済構造化データ (CSV 形式)	メインデータソース	● EBPM 及びオープンデータ化を実施するにあたり、データを結合するメインソースとして活用。	地域集計済構造化データ (CSV 形式)
DT208	クロス集計済構造化データ (CSV 形式)	メインデータソース	● EBPM 及びオープンデータ化を実施するにあたり、データを結合するメインソースとして活用。	クロス集計済構造化データ (CSV 形式)

## 【DT201】正規化地図データ

【DT205】正規化テーブルデータならびに【DT204】地図データを【FN022】データ結合前処理（ジオコーディング）または【FN023】データ結合前処理（座標系統一）によって正規化した GeoJSON 形式のデータ。

【DT205】クロス集計済データに対して、geometry カラムを有する。

表 4-121 【DT201】正規化地図データテーブルの例

番号	フィールド名	データ型	項目	入力例	備考	出典
1	ID	整数型	ID		元データにあった seq 値	DT001/DT002
2	DOCUMENT_NAME	テキスト型	資料名		【FN021】で付与	DT001/DT002
3	SOURCE_NAME	テキスト型	出典名		【FN020_2】で付与	DT001/DT002
4	GEOMETRY	ジオメトリ型	WKT フォーマット		元データにあったジオメトリ、または、ジオコーディングで付与	DT001/DT002/DT003
5	INT_VALUE	整数型	値		元データにあった値	DT001/DT002

6	TEXT_VAL UE	文字列	値		元データに あった値	DT001/DT00 2/DT003/DT 004
7	BOOL_VA LUE	Bool	値		元データに あった値	DT001/DT00 2/DT003/DT 004
8	REAL_VAL UE	小数点	値	...	元データに あった値	DT001/DT00 2/DT003/DT 004
9	...	...	...	...	...	...
10	TEXT_VAL UE	文字列	値		元データに あった値	DT001/DT00 2/DT003/DT 004
...						

## 【DT202】正規化テーブルデータ

【DT204】構造データ化済テキストデータならびに【DT004】テーブルデータを【FN021】データ結合前処理（データクレンジング）によって正規化した CSV 形式のデータ。

表 4-122 【DT202】正規化テーブルデータテーブルの例

番号	フィールド名	データ型	項目	入力例	備考	出典
1	ID	整数型	ID		元データに あった値	DT001/DT00 2
2	DOCUME NT_NAME	テキスト型	資料名		【FN021】で 付与	DT001/DT00 2
3	SOURCE_ NAME	テキスト型	出典名		【FN020_2 】で付与	DT001/DT00 2
4	INT_VALU E	整数型	値		元データに あった値	DT001/DT00 2
5	TEXT_VAL UE	文字列	値		元データに あった値	DT001/DT00 2/DT003/DT 004
6	BOOL_VA LUE	Bool	値		元データに あった値	DT001/DT00 2/DT003/DT

						004
7	REAL_VAL UE	小数点	値	…	…	DT001/DT002/DT003/DT004
…	…	…	…	…	…	…
10	TEXT_VAL UE	文字列	値		元データにあった値	DT001/DT002/DT003/DT004
…						

## 【DT203】 OCR 処理済テキストデータ

【DT001】 原票データに対し、OCR 処理を行いテキストデータ化したデータ。原票データ内で画像データ化されていた文字情報を含む。

## 【DT204】 構造化データ化済テーブルデータ

【DT203】 OCR 処理済テキストデータおよび 【DT002】 テキストデータを、CSV 形式に変換したデータ。

表 4-123 【DT204】 構造化データ化済テーブルデータの例

番号	フィールド名	データ型	項目	入力例	備考	出典
1	ID	整数型	ID		自動付与	DT001/DT002
2	DOCUMENT_NAME	テキスト型	出典名		【 FN020_2】 で付与	DT001/DT002
3	INT_VALUE	整数型	値		【 FN020_2】 で付与	DT001/DT002
4	TEXT_VALUE	文字列	値		【 FN020_2】 で付与	DT001/DT002/DT003/DT004
5	BOOLEAN_VALUE	Bool	値		【 FN020_2】 で付与	DT001/DT002/DT003/DT004
6	REAL_VALUE	小数点	値	…	【 FN020_2】 で付与	DT001/DT002

	LUE				2】で付与	02/DT003/ DT004
...	...	...	...	...	...	...
...	TEXT_VA LUE	文字列	値			DT001/DT0 02/DT003/ DT004
...	...	...	...	...	...	...

## 【DT205】クロス集計データ

【D202】正規化テーブルデータ、【DT206】テキスト結合データ、【DT207】空間結合済データ、【DT208】地域集計済データをクロス集計したデータ。正規化テーブルデータに含まれている情報を含む。

表 4-124 【DT205】正規化済構造データ化テーブルデータの例

番号	フィールド名	データ型	項目	入力例	備考	出典
1	ID	整数型	ID		自動付与	DT001/DT 002
2	AXIS1	テキスト型	日時		ユーザーが 集計の軸とし て指定	DT001/DT 002
3	AXIS2	テキスト型	住所		ユーザーが集 計の軸として 指定	DT001/DT 002/DT00 3
4	INT_VA LUE	整数型	値		ユーザーが指 定した集計対 象	DT001/DT 002
5	TEXT_V ALUE	文字列	値		ユーザーが指 定した集計対 象	DT001/DT 002/DT00 3/DT004
6	BOOL_ VALUE	Bool	値		ユーザーが指 定した集計対 象	DT001/DT 002/DT00 3/DT004
7	REAL_V ALUE	小数点	値	...	ユーザーが指 定した集計対 象	DT001/DT 002/DT00 3/DT004
...	...	...	...	...	...	...

## 【DT206】テキスト結合済データ

【DT202】正規化テーブルデータ、【DT205】クロス集計済データ、【DT201】正規化地図データうち、②種のデータを【FN024】データ結合機能（テキストマッチング）によってテキスト結合した構造化データ。

表 4-125 【DT206】テキスト結合済み構造データの例

番号	フィールド名	データ型	項目	入力例	備考	出典
1	ID	整数型	ID		自動付与	DT001/DT002
2	DOCUMENT_NAME	テキスト型	資料名		結合元ファイルの値。 【DT205】、 【DT208】を除く。	DT001/DT002
3	SOURCE_NAME	テキスト型	出典名		結合元ファイルの値。 【DT205】、 【DT208】を除く。	DT001/DT002/DT003
4	TEXT_VALUE	文字列	値		結合に用いたキーカラム	DT001/DT002
5	INT_VALUE	整数型	値		結合元ファイルのカラム	DT001/DT002/DT003/DT004
6	BOOL_VALUE	Bool	値		結合元ファイルのカラム	DT001/DT002/DT003/DT004
7	REAL_VALUE	小数点	値	...	結合元ファイルのカラム	DT001/DT002/DT003/DT004
...	...	...	...	...	...	...
...	TEXT_VALUE	文字列	値		結合対象ファイルのカラム	DT001/DT002/DT003/DT004
...	MATCH	Bool	完全一致の有無		結合対象ファイルのカラム	...

...	OTHER _ONE	整数型	結合先 ID		結合対象ファ イルのカラム	DT001/DT 002
...	ADDRE SS2	テキス ト型	住所 2		結合対象ファ イルのカラム	DT001/DT 002/DT00 3
...	...	...	...	...	...	...

## 【DT207】空間結合済データ

【DT201】正規化地図データ、【DT204】空間結合済データ、【DT208】地域集計済構造化データのうち2種類  
のファイルを対象に、空間結合を実施したデータ。

表 4-126 【DT207】空間結合済データの例

番号	フィールド名	データ型	項目	入力例	備考	出典
1	ID	整数型	ID		結合元ファ イルのカラ ム	DT001/DT 002
2	DOCUM ENT_N AME	テキス ト型	資料名		結合元ファ イルの値。 【DT205】 、【DT208】 を除く。	DT001/DT 002
3	SOURC E_NAM E	テキス ト型	出典名		結合元ファ イルの値。 【DT205】 、【DT208】 を除く。	DT001/DT 002
4	GEOME TRY	ジオメ トリ型	WKTフォー マット		結合元ファ イルのジオ メトリ	DT001/DT 002/DT00 3
5	INT_VA LUE	整数型	値		結合元ファ イルのカラ ム	DT001/DT 002
6	TEXT_V ALUE	文字列	値		結合元ファ イルのカラ	DT001/DT 002/DT00

					ム	3/DT004
7	BOOL_VALUE	Bool	値		結合元ファイルのカラム	DT001/DT002/DT003/DT004
8	REAL_VALUE	小数点	値	...	結合元ファイルのカラム	DT001/DT002/DT003/DT004
...	...	...	...	...	...	...
...	TEXT_VALUE	文字列	値		結合対象ファイルのカラム	DT001/DT002/DT003/DT004
...	...	...	...	...	...	...

## 【DT208】地域集計済構造化データ

【DT201】正規化地図データ、【DT204】空間結合済データ、【DT208】地域集計済構造化データ（いずれもポリゴンデータ）を集計元とし、集計対象となる【DT201】正規化地図データ、【DT204】空間結合済データ、【DT208】地域集計済構造化データを集計したデータ。集計元となる情報をすべて含む。

表 4-127 【D205】地域集計済構造化データの例

番号	フィールド名	データ型	項目	入力例	備考	出典
1	ID	整数型	ID		集計に用いた【DT201】のID	DT001/DT002
2	GEOMETRY	ジオメトリ型	WKTフォーマット		集計元ファイルのジオメトリ	DT001/DT002
3	AREA_NAME	テキスト型	名称		集計に用いた【DT201】のカラム	DT001/DT002
4	INT_VALUE	整数型	値		集計元ファイルのカラム	DT001/DT002

5	TEXT_V ALUE	文字列	値		集計元フ ァイルの カラム	DT001/DT 002/DT00 3/DT004
6	BOOL_ VALUE	Bool	値		集計元フ ァイルの カラム	DT001/DT 002/DT00 3/DT004
7	REAL_V ALUE	小数点	値	...	集計元フ ァイルの カラム	DT001/DT 002/DT00 3/DT004
...	...	...	...	...	...	...
...	TEXT_V ALUE	文字列	値		集計対象 ファイル のうちユ ーザーが 指定した 集計対象	DT001/DT 002/DT00 3/DT004
...	MATCH	Bool	完全一致の 有無		集計対象 ファイル のうちユ ーザーが 指定した 集計対象	
...	...	...	...	...	...	...

## 4-6. ユーザーインターフェース

## 4-6-1. 画面一覧

一般ユーザーは、1) PC 用画面にてオペレーションを実行する。2) CMS 画面はシステム管理者のみがアクセスするものとし、一般ユーザーへの公開はしない。

## 1) PC 用画面

表 4-128 PC 画面一覧

ID	連携 (ID)	画面名	画面説明	画面を表示した機能 (ID)
SC001	SC002	ログイン画面	<ul style="list-style-type: none"> <li>メールアドレス/パスワードを入力することでログインする画面を提供する。</li> </ul>	FN031_1
SC002	SC003,SC004,SC005,SC0007,SC014,SC016,SC017,SC018	ホーム画面	<ul style="list-style-type: none"> <li>アカウント管理画面、アセット一覧/詳細画面、コンテンツ一覧画面、オペレーター一覧画面、テンプレート一覧画面、処理状況一覧画面、EBPM Tools 一覧画面、チャットボット画面にアクセスする。</li> <li>ユーザーは本画面から各メニューに遷移する。</li> </ul>	FN031_2
SC003	SC002	アカウント管理画面	<ul style="list-style-type: none"> <li>アカウントの追加・削除、権限管理を行う。</li> </ul>	FN031_2
SC004	SC002	アセット一覧/詳細画面	<ul style="list-style-type: none"> <li>ユーザーがアップロードしたファイル(=アセット)を一覧で確認、検索できる画面を提供する。</li> <li>ユーザーはアセット画面から直接ファイルをアップロードする。</li> <li>ユーザーがアップロードしたアセットのファイルの中身をプレビューにて確認できる画面を提供する。</li> </ul>	FN004,FN006,FN015
SC005	SC002	コンテンツ一覧画面	<ul style="list-style-type: none"> <li>作成済みのコンテンツを一覧で確認、検索できる画面を提供する。</li> <li>一覧からコンテンツの削除、名称の変更を可能とする。</li> </ul>	FN015,FN016
SC006	SC005	コンテンツ詳細画面	<ul style="list-style-type: none"> <li>コンテンツへのインサート、カラム追加、コンテンツ内の値の更新ができる画面を提供する。</li> <li>コンテンツ詳細画面内の管理パネルより、オー</li> </ul>	FN018,FN019

			<p>プンデータ配信を可能とする。</p>	
SC007	SC002,SC008,SC009,SC010,SC011SC012,SC013,SC015	オペレーター一覧画面	<ul style="list-style-type: none"> <li>● 構造化処理画面、データ結合前処理画面、テキストマッチング画面、空間結合画面、空間結合画面に遷移する画面を提供する。</li> </ul>	
SC008	SC007	構造化処理画面	<ul style="list-style-type: none"> <li>● 原票データ、テキストデータの構造化処理を実行する画面を提供する。</li> </ul>	FN007,FN009,FN020_1,FN020_2
SC009	SC007	データ結合前処理画面	<ul style="list-style-type: none"> <li>● データクレンジング、ジオコーディングを実行する画面を提供する。</li> </ul>	FN010,FN013,FN017,FN021,FN022,FN023
SC010	SC007	テキストマッチング画面	<ul style="list-style-type: none"> <li>● 結合対象とするテーブル、結合方式（完全一致/部分一致）、部分一致の場合は類似度閾値、結合条件とするカラム名を指定し、データ結合処理（テキストマッチング）を実行する画面を提供する。</li> <li>● テーブルデータは、JSON 形式、GeoJSON 形式いずれも対象とする。</li> </ul>	FN011, FN017,FN024
SC011	SC007	空間結合画面	<ul style="list-style-type: none"> <li>● テーブルデータ（GeoJSON 形式のみ）に対し、結合方式、メインデータソース、結合データソース、結合条件を選択しデータを結合する画面を提供する。</li> </ul>	FN013,FN017,FN026
SC012	SC007	クロス集計画面	<ul style="list-style-type: none"> <li>● テーブルデータ（JSON 形式、GeoJSON 形式）に対し、ユーザーが任意のカラムを指定し、指定したカラムをキーとしデータ集計処理（クロス集計）を実行する画面を提供する。</li> </ul>	FN012.FN017,FN025
SC013	SC007	空間集計画面	<ul style="list-style-type: none"> <li>● テーブルデータ（GeoJSON 形式のみ）に対し、結合方式、メインデータソース、結合データソース、結合条件を選択しデータを結合する画面を提供する。</li> </ul>	FN012,FN014,FN027
SC014	SC007	テンプレート一覧画面	<ul style="list-style-type: none"> <li>● ユーザーが登録したテンプレートを一覧で確認できる画面を提供する。</li> </ul>	FN035

			<ul style="list-style-type: none"> <li>● 新規登録ボタンより詳細画面に遷移することで、新規にテンプレートを登録できる。</li> </ul>	
SC015	SC007	ワークフローテンプレート編集画面	<ul style="list-style-type: none"> <li>● 構造化処理を実行する際の抽出カラムと条件を定義したテンプレート、およびデータ結合前処理、結合処理、集計処理で利用する条件句を定義したテンプレートを登録する画面を提供する。</li> <li>● 定義したテンプレートは、各処理画面から呼び出し利用することができる。</li> </ul>	FN035
SC016	SC002	処理状況一覧画面	<ul style="list-style-type: none"> <li>● オペレーターでの実行結果を一覧、および実行したオペレーターの処理結果を確認できる画面を提供する。</li> </ul>	FN008
SC017	SC002	可視化アプリ一覧画面	<ul style="list-style-type: none"> <li>● 各ユースケースで作成する可視化アプリのURLを一覧形式で確認できる画面を提供する。</li> <li>● ユーザーは一覧から各ユースケースにアクセスできる。</li> <li>● 可視化アプリ一覧上でのアクセス制御は行わないものとする。</li> </ul>	-
SC018	SC002	チャットボット画面	<ul style="list-style-type: none"> <li>● 対話型のUIにて問い合わせ可能な画面を提供する。</li> </ul>	FN020_3, FN032, FN033, FN034
SC019	SC002	データセット管理画面	<ul style="list-style-type: none"> <li>● データセットの作成、公開、削除を行う画面を提供する。</li> </ul>	-

## 2) CMS 画面

表 4-129 管理者画面一覧

ID	連携 (ID)	画面名	画面説明	画面を表示した機能 (ID)
SC101	SC102	CMS ログイン画面	● メールアドレス/パスワードを入力することで CMS にログインする画面を提供する。	FN031
SC102	SC101,SC103,SC104,SC105,SC106	CMS ホーム画面	● アカウント設定、ワークスペースの作成・削除、プロジェクトの作成・削除をする画面を提供する。	FN001, FN002
SC103	SC102	CMS アセット画面	● Veda にてアセットとして登録されたデータを確認できる画面を提供する。	FN004
SC104	SC102	CMS コンテンツ画面	● Veda にてコンテンツとして登録されたデータを確認できる画面を提供する。	FN003
SC105	SC102	CMS アイテム画面	● Veda にてコンテンツとして登録されたデータのレコードを確認できる画面を提供する。	FN005
SC105	SC106	CMS アカウント管理画面	● CMS 上のアカウントの追加・削除、権限管理を行う画面を提供する。	-

4-6-2. 画面遷移図

1) PC 用画面

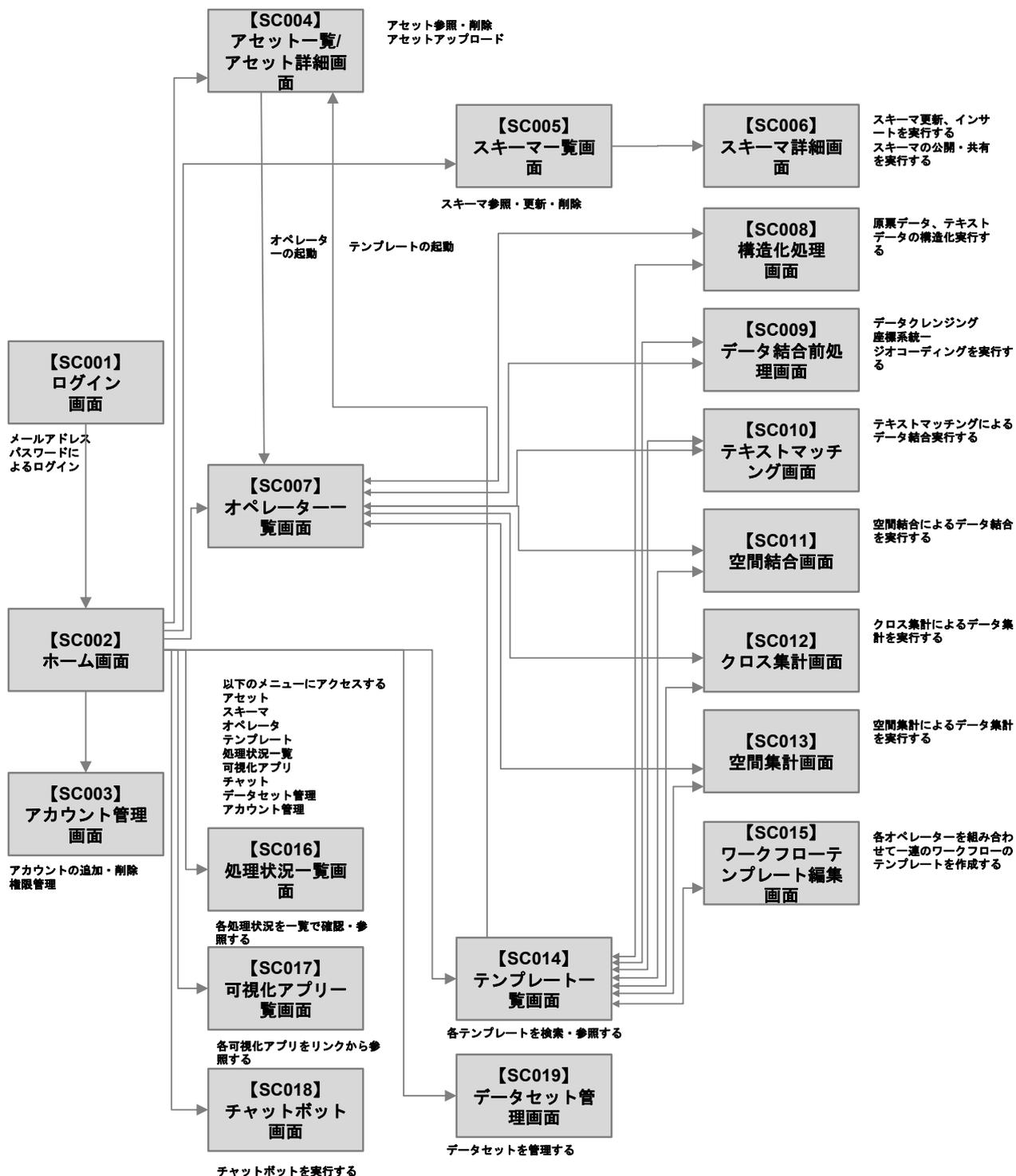


図 4-36 PC 画面遷移図

2) CMS 画面

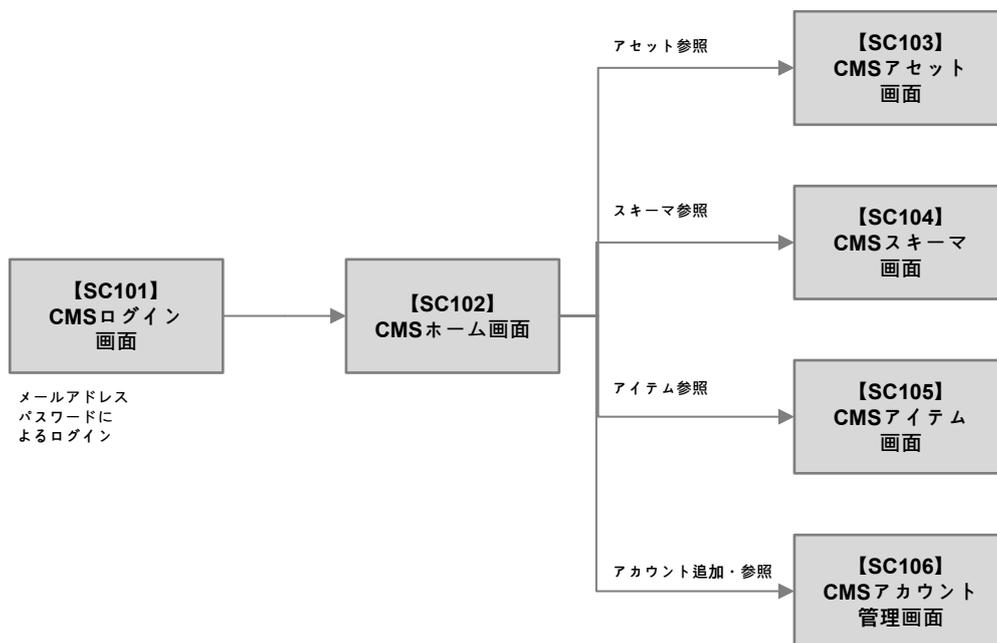


図 4-37 管理者画面遷移図

### 4-6-3. 各画面仕様詳細

#### 1) PC用画面

##### 1. 【SC001】ログイン画面

- 画面の目的・概要
  - メールアドレス、パスワードを入力しシステムにログインするための画面。
- 画面イメージ

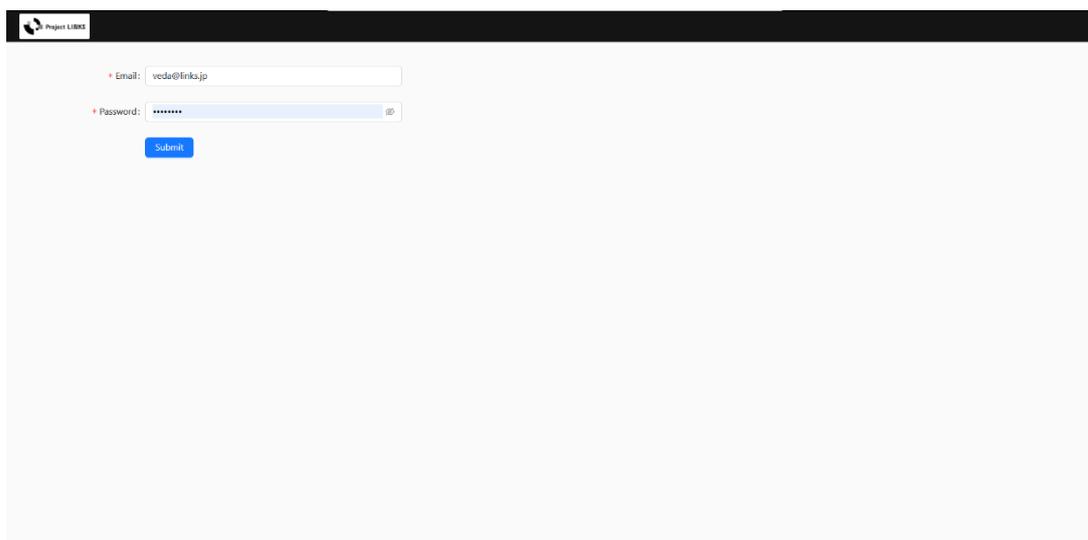


図 4-38 ログイン画面のイメージ

- 【SC002】ホーム画面
- 画面の目的・概要
  - 各オペレーターにアクセスするタイルを表示する。
  - 画面左部には、オペレーター、アセット、コンテンツ、EBPM Tools、テンプレート、処理状況一覧、チャット機能にアクセスするメニューを表示する。

● 画面イメージ

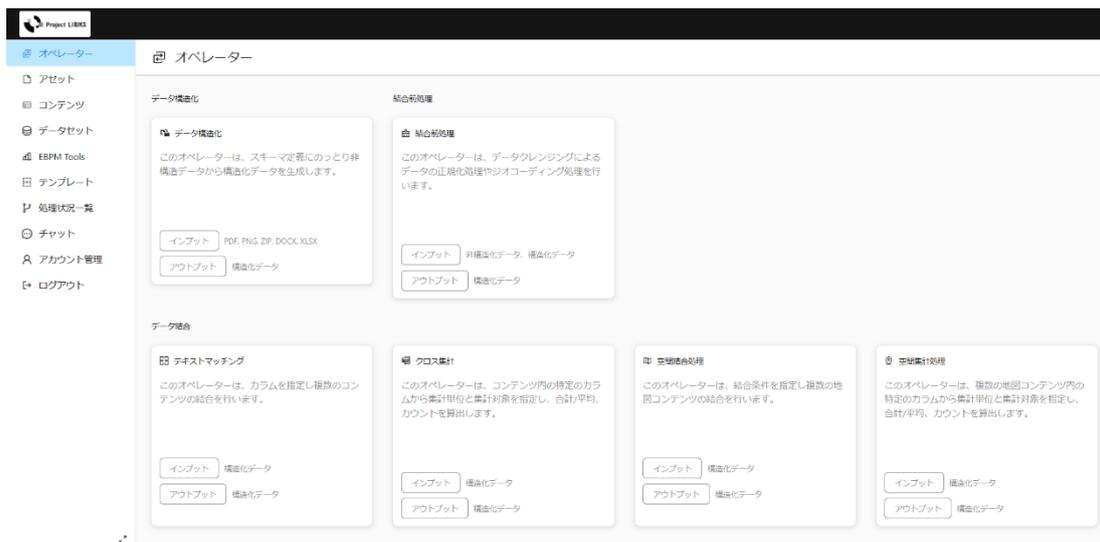


図 4-39 ホーム画面のイメージ

● 【SC003】 アカウント管理画面

● 画面の目的・概要

➤ アカウントの追加、削除、権限変更を行う。

● 画面イメージ

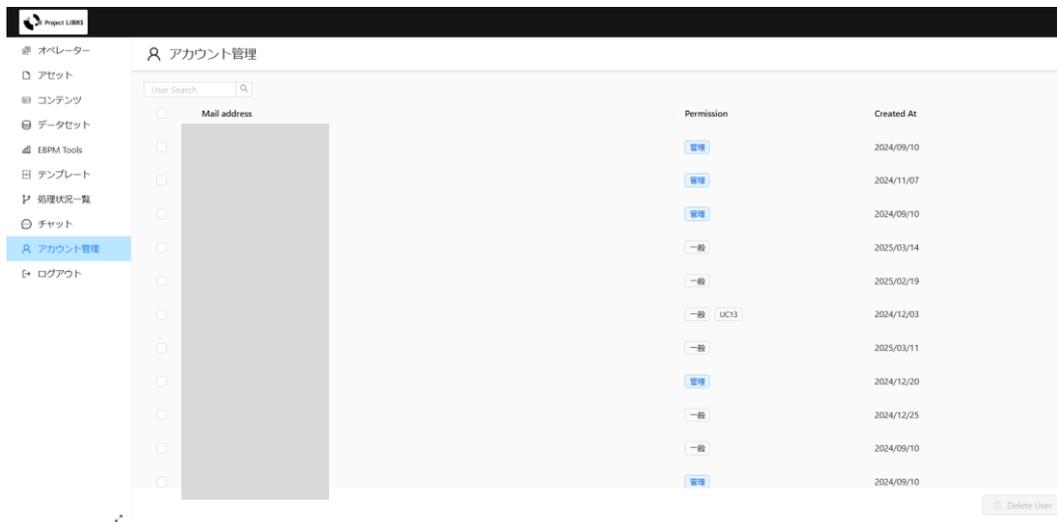


図 4-40 アカウント管理画面のイメージ

## 2. 【SC004】アセット一覧/詳細画面

### ● 画面の目的・概要

- ユーザーが登録したアセットの一覧を表示する。
- アップロードボタンよりアップロードするファイルを選択し、ローカルからファイルアップロードを実行できる。
- 左上部の検索ウィンドウ検索キーワードを指定することで、任意のアセットを検索できる。
- アセット選択をすることで、画面右部にアセットに含まれるファイルのプレビューを表示する。
- 

### ● 画面イメージ

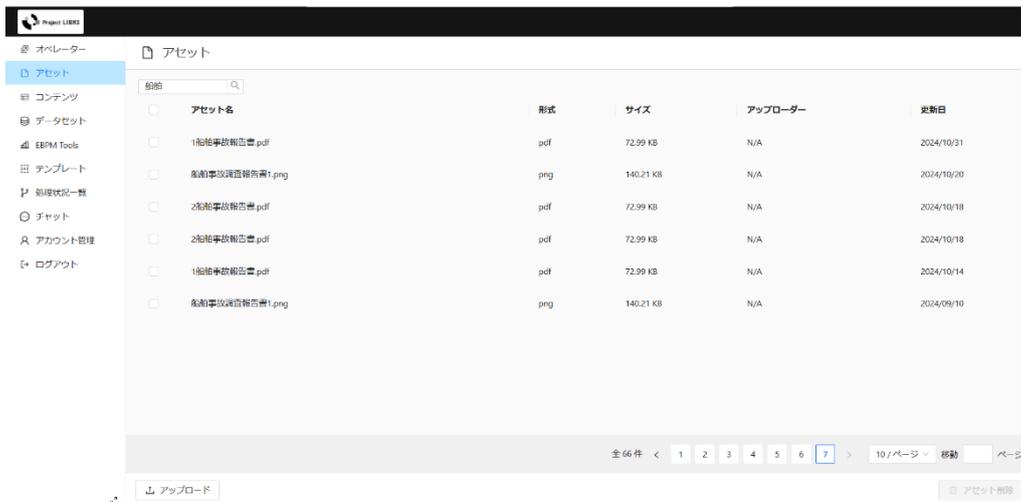


図 4-41 アセット画面のイメージ

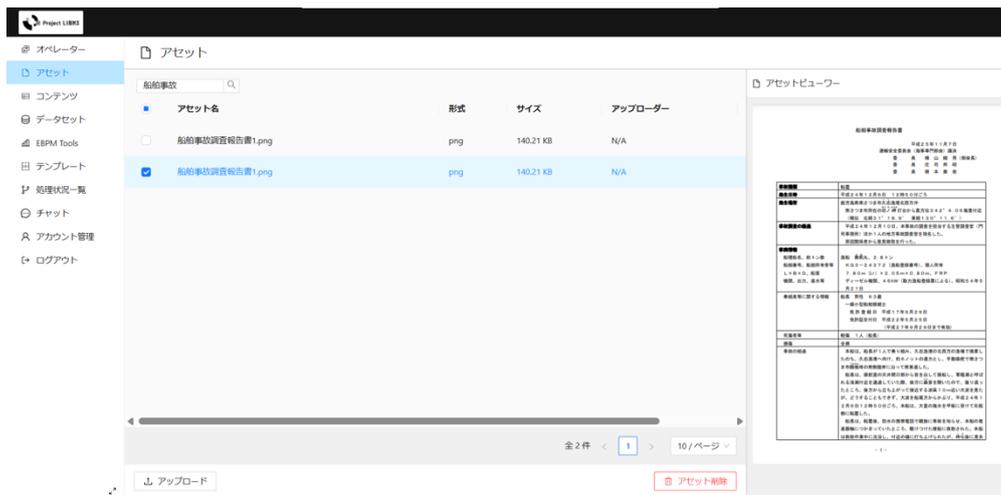


図 4-42 アセット画面のイメージ (アセット選択時：原票)

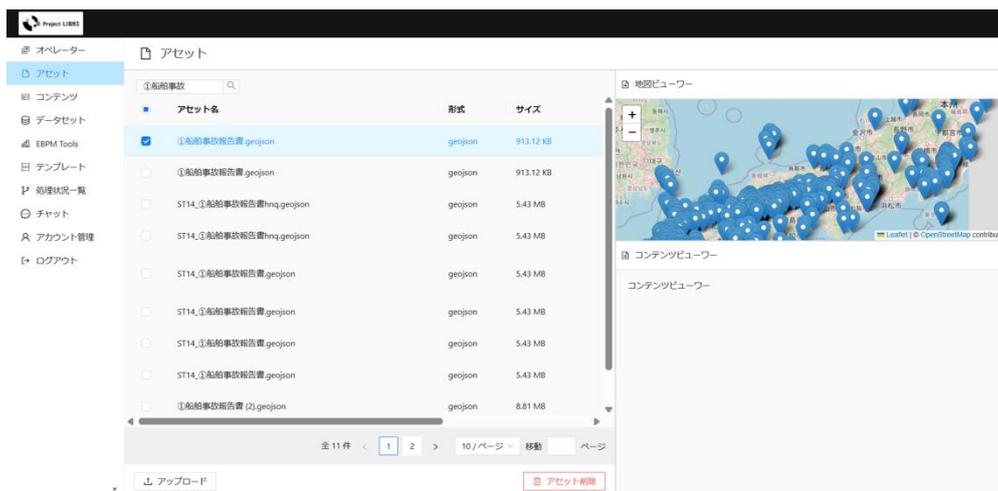


図 4-43 アセット画面のイメージ（アセット選択時：地図データ）

## 3. 【SC005】コンテンツ一覧画面

## ● 画面の目的・概要

- 作成済みのコンテンツを一覧で確認、検索できる画面を提供する。
- 一覧からコンテンツの削除、名称の変更を可能とする。左上部の検索ウィンドウよりコンテンツ名を入力し、登録済みのコンテンツを検索することができる。
- 編集ボタン  を押下することで、【SC006】コンテンツ詳細画面に遷移する。
- 管理パネルからは、コンテンツの公開情報、共有情報を管理する。
- コンテンツ選択をすることで、画面右部にアセットに含まれるコンテンツのプレビューを表示する。
- コンテンツ選択すると画面下部のコンテンツ編集、コンテンツ削除、オペレーター、テンプレートボタンがアクティブになり、コンテンツの削除、オペレーター画面の起動（【SC009】データ結合前処理、【SC010】テキストマッチング画面、【SC011】空間結合画面、【SC012】クロス集計、【SC013】空間集計画面）、テンプレートの紐づけを実施できる。
- コンテンツ編集を押下し、「名前の変更」、「コンテンツ更新」、「コンテンツの複製」をおこなう。「コンテンツ更新」を押下した場合、【SC006】コンテンツ詳細画面に遷移する。

## ● 画面イメージ

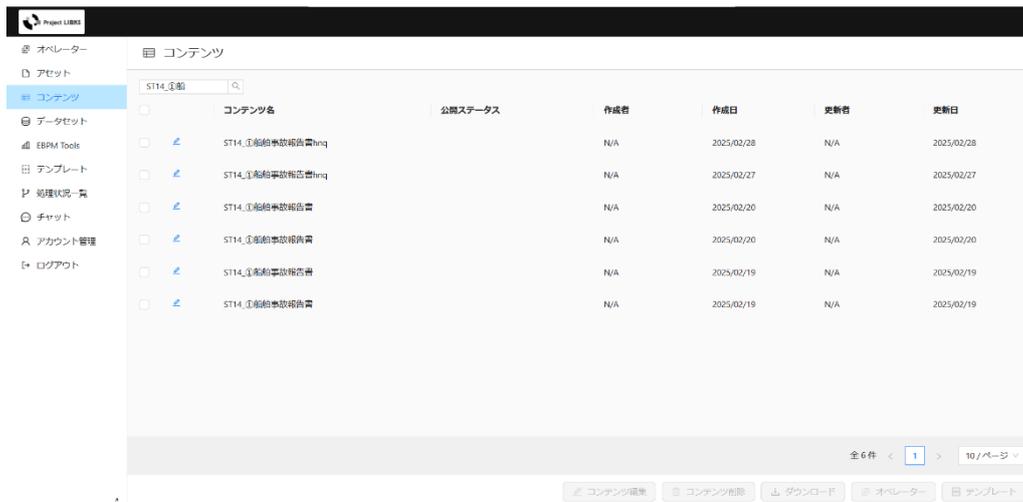


図 4-44 コンテンツ一覧画面のイメージ

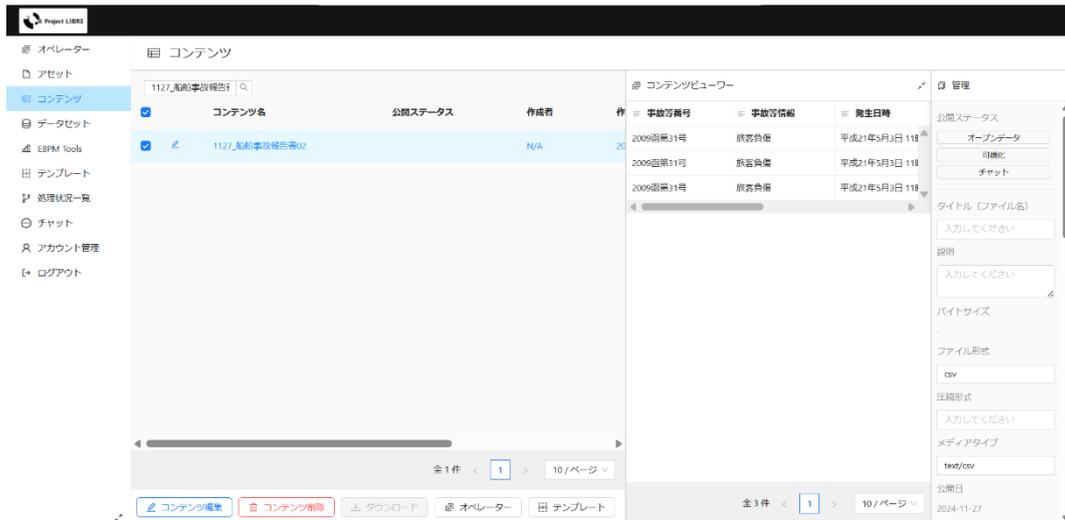


図 4-45 コンテンツ一覧画面のイメージ (コンテンツ選択時: JSON)

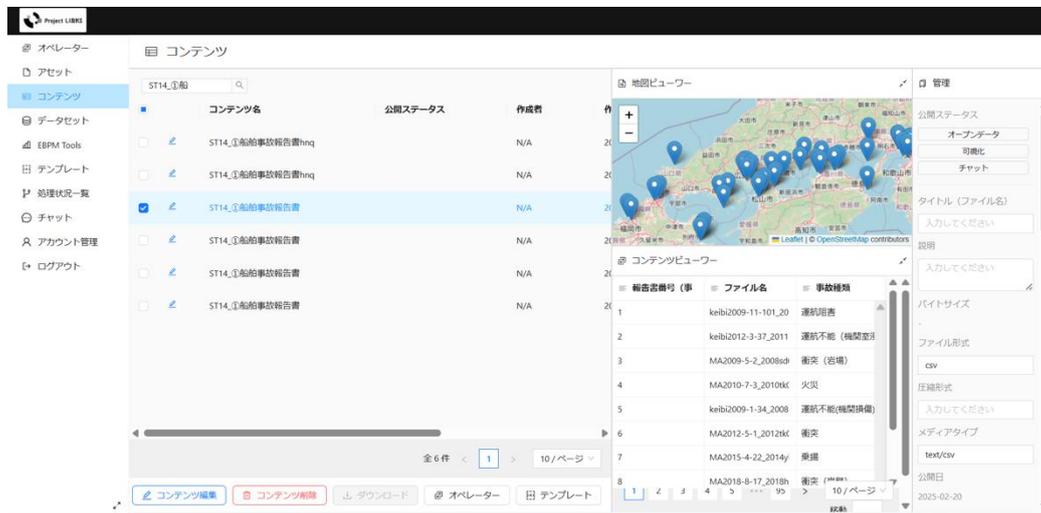


図 4-46 コンテンツ一覧画面のイメージ (コンテンツ選択時: GeoJSON)

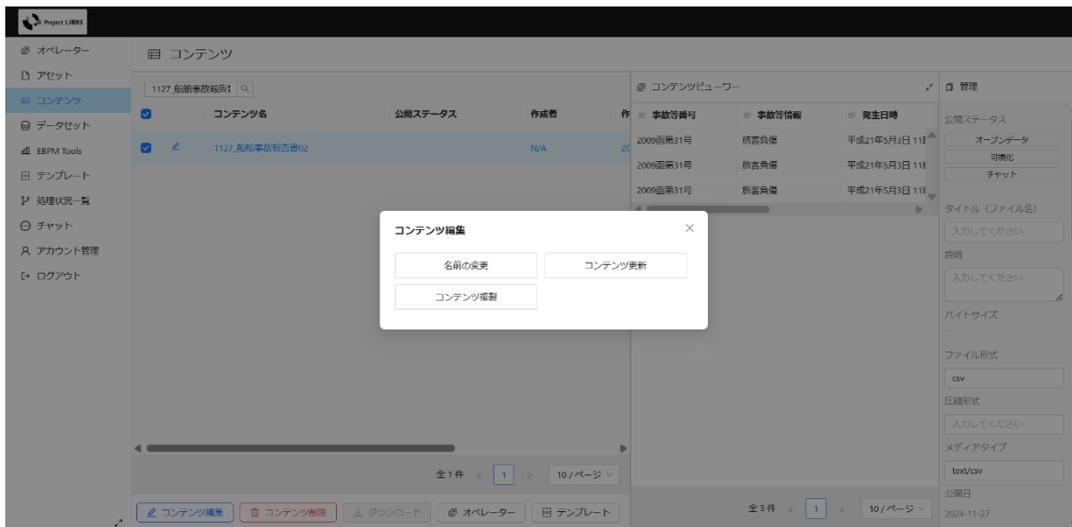


図 4-47 コンテツツ一覧画面のイメージ（コンテツツ編集ボタン押下時）

#### 4. 【SC006】 コンテツツ詳細画面

##### ● 画面の目的・概要

- 【SC005】 コンテツツ一覧画面より遷移し、コンテツツへのインサート、カラム追加、コンテツツ内の値の更新ができる画面を提供する。
- 【SC006】 コンテツツ詳細画面へ遷移すると、画面下部の行を追加、カラムを追加、保存ボタンがアクティブになり、コンテツツへの行追加、カラムの編集（更新、削除）、コンテツツ内の値編集を実施できる。
- コンテツツへの行追加は、追加するコンテツツを指定し、結合後のコンテツツプレビューを処理結果を確認から確認し、問題なければ OK を押下し、保存する。
- カラムの追加は、カラムを追加ボタンから実施する。
- カラムの削除、カラム名の変更はカラム名のツールチップから実施する。
- コンテツツ内の値の更新は、セルを押下し直接更新を実施する。

● 画面イメージ

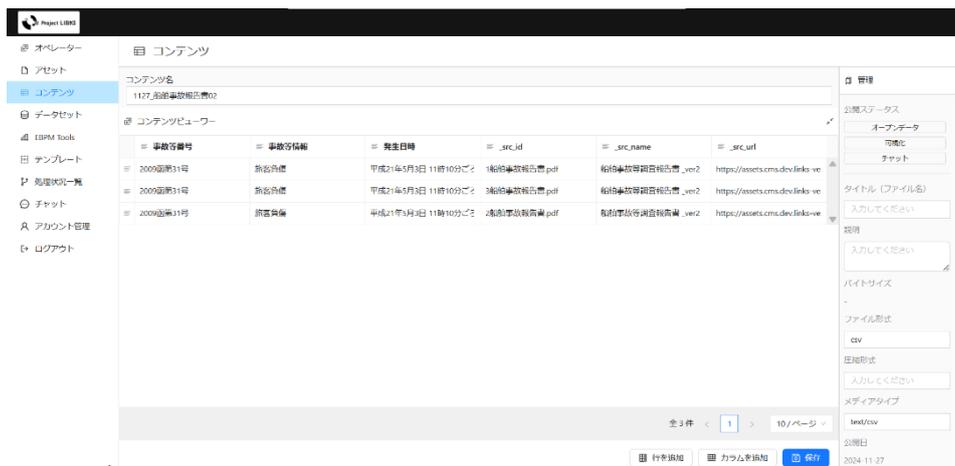


図 4-48 コンテンツ詳細画面のイメージ

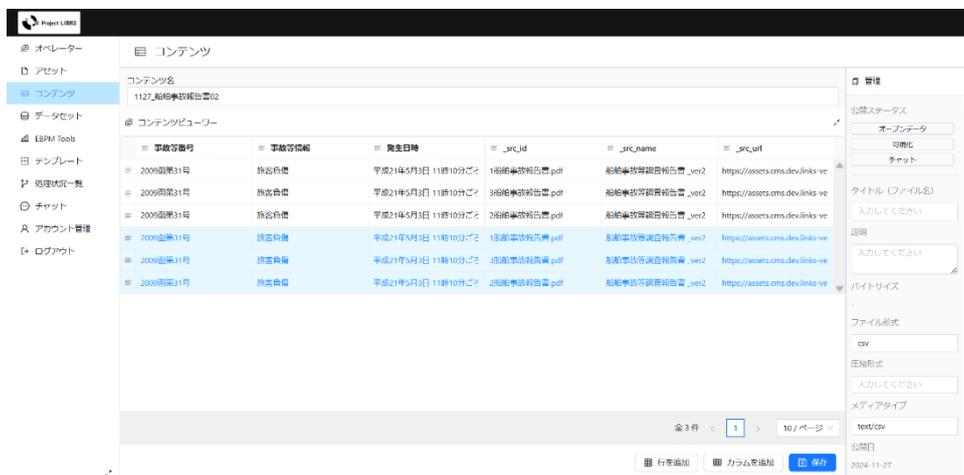


図 4-49 コンテンツ詳細画面のイメージ (行の追加)

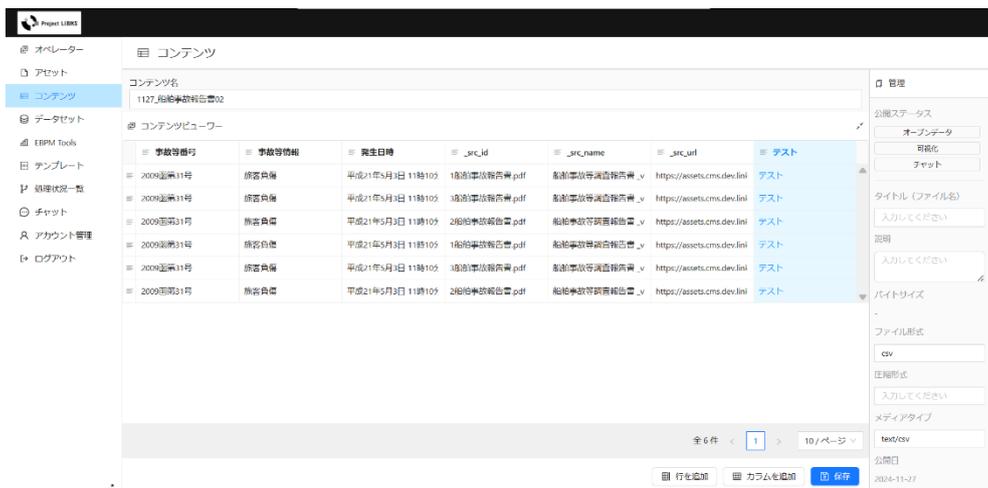


図 4-50 コンテンツ詳細画面のイメージ (カラム追加)

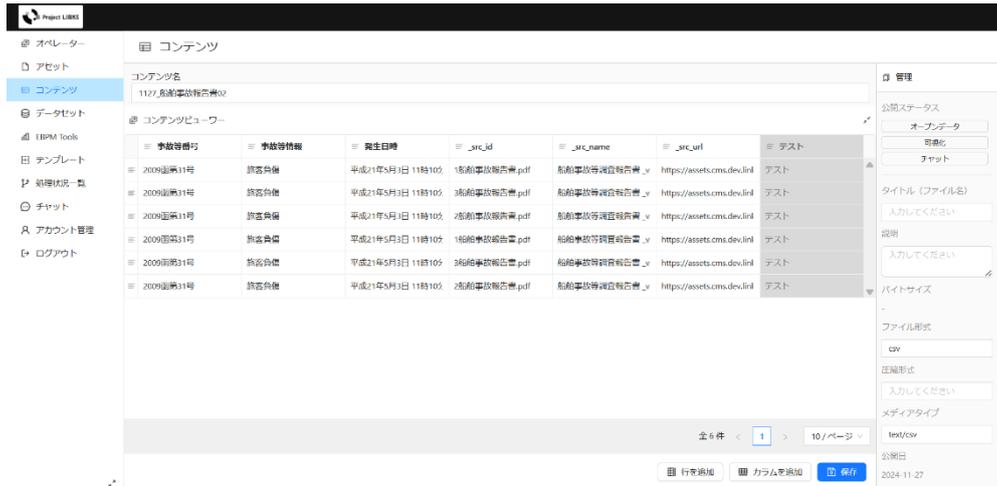


図 4-51 コンテンツ詳細画面のイメージ (カラムの削除・変更)

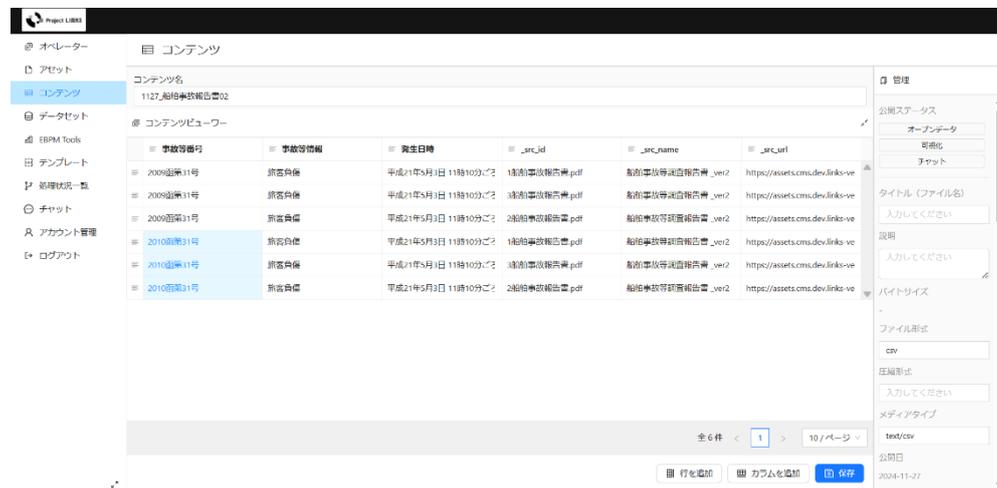


図 4-52 コンテンツ詳細画面のイメージ (値更新)

## 5. 【SC007】オペレーター一覧画面

### ● 画面の目的・概要

- オペレーターメニューから遷移する。
- 【SC008】構造化処理画面、【SC009】データ結合前処理画面、【SC010】テキストマッチング画面、【SC011】空間結合画面、【SC012】クロス集計画面、【SC013】空間集計画面を起動する画面を提供する。

## ● 画面イメージ



図 4-53 オペレーター一覧画面のイメージ

## 6. 【SC008】構造化処理画面

## ● 画面の目的・概要

- アセットのデータ構造化処理を実行する画面を提供する。対象となるアセットは、【DT001】原票データ（PDF、PNG 形式）、【DT002】テキストデータ（DOCX、XLSX 形式）である。アセット選択ボタンよりアセットを選択し、構造化処理を実行する。
- アセット選択後、アセットビューワーフィールドからファイルの中身をプレビューで参照することができる。ユーザーは設定タブより構造化に必要なパラメータを入力し構造化処理を実行する。（詳細は 4-2-3. 開発機能の詳細要件を参照する。）本パラメータをもとにプロンプトがシステム内で作成され、コンテンツデータが生成される。構造化するカラムを追加する場合には、カラムを追加ボタンよりカラムを追加する。
- データ構造化処理と同タイミングで、出典名の作成を必須で行う。（詳細は 4-2-3. 開発機能の詳細要件を参照する。）
- 処理実行後、処理結果タブに自動で遷移し、実行結果を確認できる。
- 内容に問題がなければ、保存を押下しコンテンツとして保存する。
- 実行パラメータをテンプレートとして保存する場合には、テンプレートとして保存を押下し保存する。

● 画面イメージ



図 4-54 構造化処理画面のイメージ (構造化処理実行前 1/2)

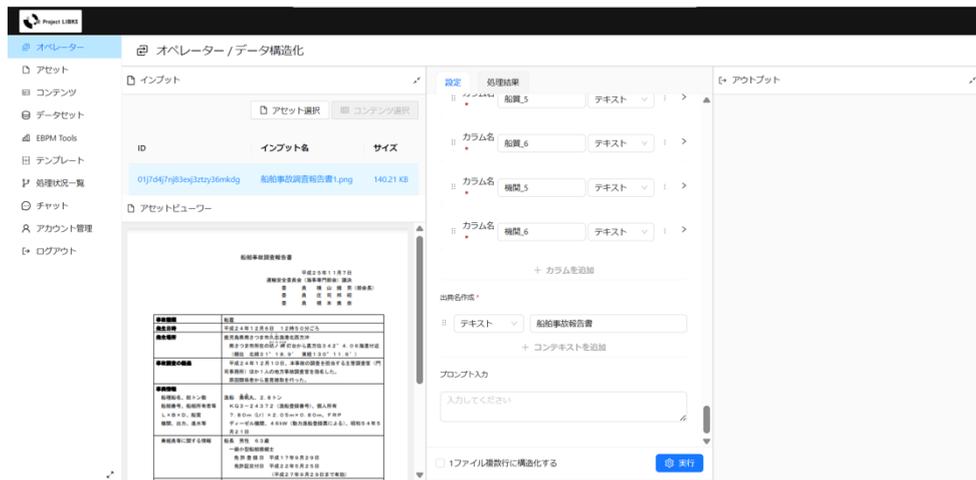


図 4-55 構造化処理画面のイメージ (構造化処理実行前 2/2)

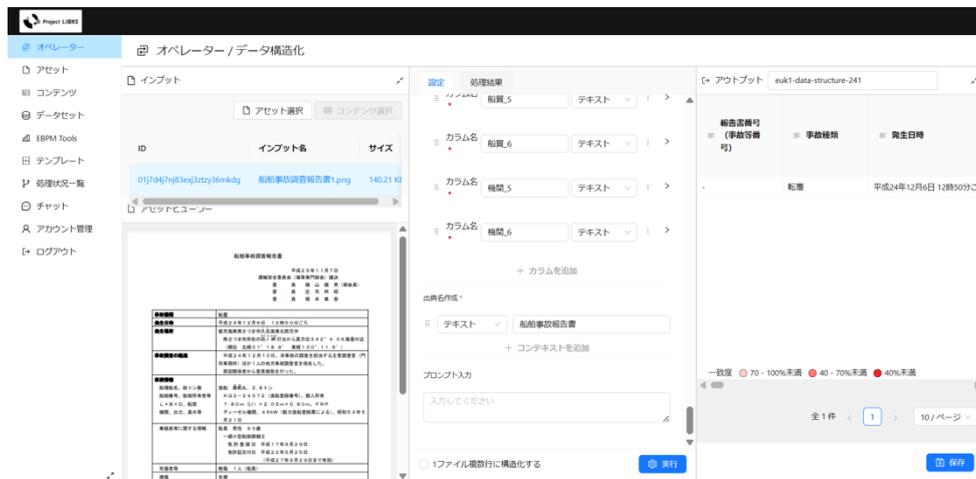


図 4-56 構造化処理画面のイメージ (構造化処理実行後)

## 7. 【SC009】データ結合前処理画面

## ● 画面の目的・概要

- 【DT003】地図データ、【DT004】テーブルデータ、【DT204】構造データ化済テキストデータに対し、データクレンジング、ジオコーディング、資料名作成、地図データに対し座標系統一の処理を実行する画面を提供する。
  - ◇ アセット選択：【DT003】地図データ、【DT004】テーブルデータ
  - ◇ コンテンツ選択：【DT204】構造データ化済テキストデータ
- 上記に則りアセット、もしくはコンテンツを選択後、本機能の処理オプションとして、資料名作成、データクレンジング、ジオコーディング、秘匿化処理を選択する。
- 本処理内では、【DT201】正規化地図データに対して【FN023】データ結合前処理（座標系統一）をシステム内で自動実行し座標系を WGS84 に統一する。
- 【DT003】地図データ、【DT004】テーブルデータの正規化処理をシステム内で自動実行し、【DT201】正規化地図データ、【DT202】正規化テーブルデータを作成する。
- 処理実行後、処理結果タブに自動で遷移し、実行結果を確認できる。
- 内容に問題がなければ、保存を押下しコンテンツとして保存する。
- 実行パラメータをテンプレートとして保存する場合には、テンプレートとして保存を押下し保存する。

## ● 画面イメージ

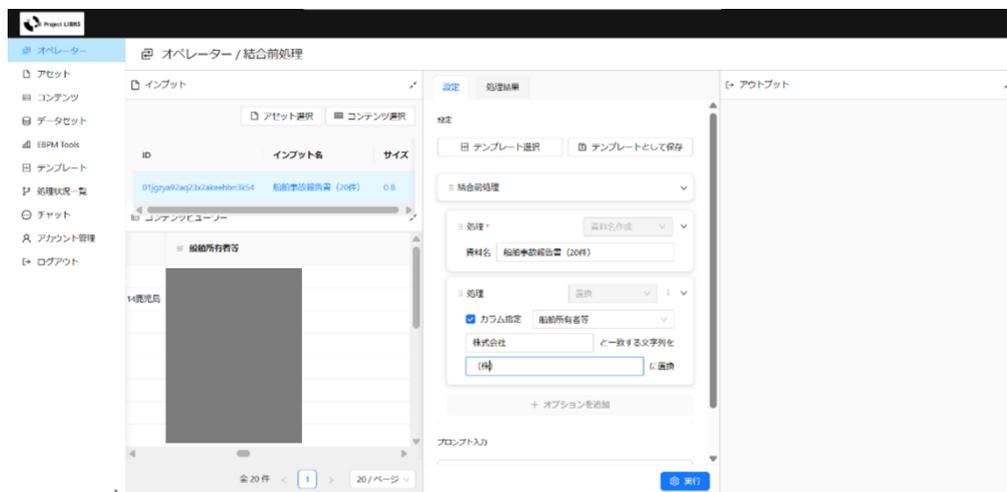


図 4-57 データ結合前処理の画面イメージ（処理実行前）

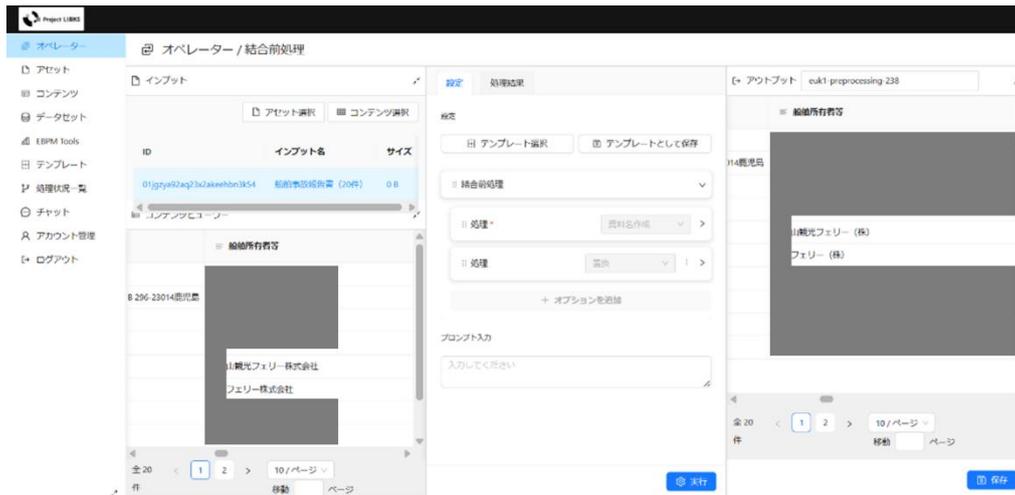


図 4-58 データ結合前処理の画面イメージ（処理実行後）

## 8. 【SC010】テキストマッチング画面

### ● 画面の目的・概要

- 【DT201】正規化地図データ、【DT202】正規化テーブルデータ、【DT206】データ結合済構造化データ、【DT207】空間結合済データ、【DT208】地域集計済構造化データ、【DT209】クロス集計済構造化データを対象にコンテンツ同士の結合処理を行う画面を提供する。
- コンテンツ選択から結合元とするコンテンツをメインコンテンツと結合するコンテンツ選択し、それぞれ結合するカラムをキーカラムとして指定し【FN024】データ結合処理（テキストマッチング）を押下し実行する。
- 本機能は 1 コンテンツごとの結合に対応しており、複数のコンテンツを結合しようとした場合、警告が起きる。
- 結合コンテンツで結合後残すカラムを指定したい場合には、結合カラムにカラム名を入力する。結合カラムを入力しない場合には、結合コンテンツに存在するカラムすべてを残す。
- 処理実行後、処理結果タブに自動で遷移し、実行結果を確認できる。
- 内容に問題がなければ、保存を押下しコンテンツとして保存する。
- 実行パラメータをテンプレートとして保存する場合には、テンプレートとして保存を押下し保存する。

## ● 画面イメージ



図 4-59 テキストマッチング画面のイメージ（処理実行前）

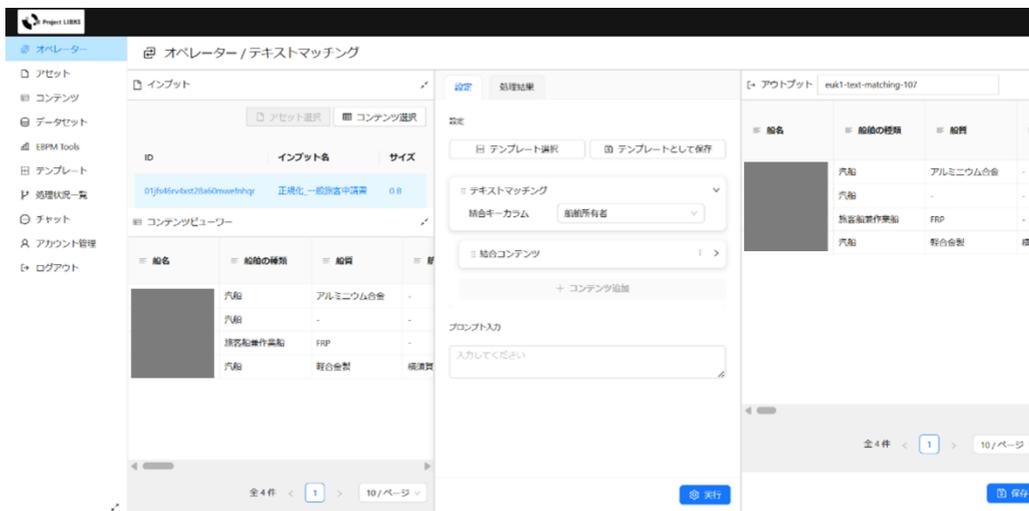


図 4-60 テキストマッチング画面のイメージ（処理実行後）

## 9. 【SC011】空間結合画面

## ● 画面の目的・概要

- 本画面では、【DT201】正規化地図データ、【DT207】空間結合済データ、【DT208】地域集計済構造化データに対し、【FN026】データ結合処理（空間結合）を行う。
- コンテンツ選択から結合元とするコンテンツをメインコンテンツと結合するコンテンツ、結合方式を選択し【FN026】データ結合機能（空間結合処理）を実行を押下し実行する。
- 本機能は 1 コンテンツごとの結合に対応しており、複数のコンテンツを結合しようとした場合、警告が起きる。
- 結合コンテンツを追加後、結合方式で交差結合、最近傍結合を選択する。最近傍結合の場合には検索上限範囲を指定する。
- 処理実行後、処理結果タブに自動で遷移し、実行結果を確認できる。

- 内容に問題がなければ、保存を押下しコンテンツとして保存する。
- 実行パラメータをテンプレートとして保存する場合には、テンプレートとして保存を押下し保存する。

● 画面イメージ

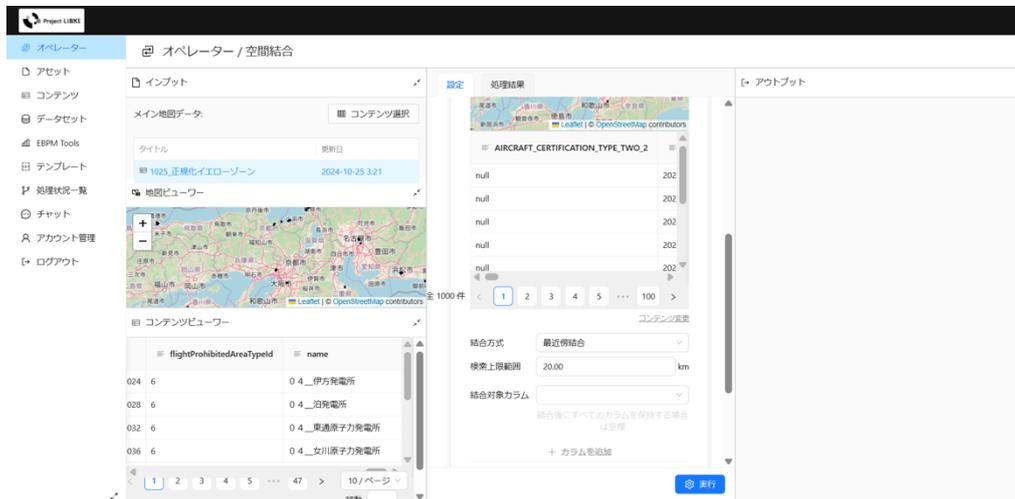


図 4-61 空間結合画面のイメージ（最近傍結合：処理実行前）

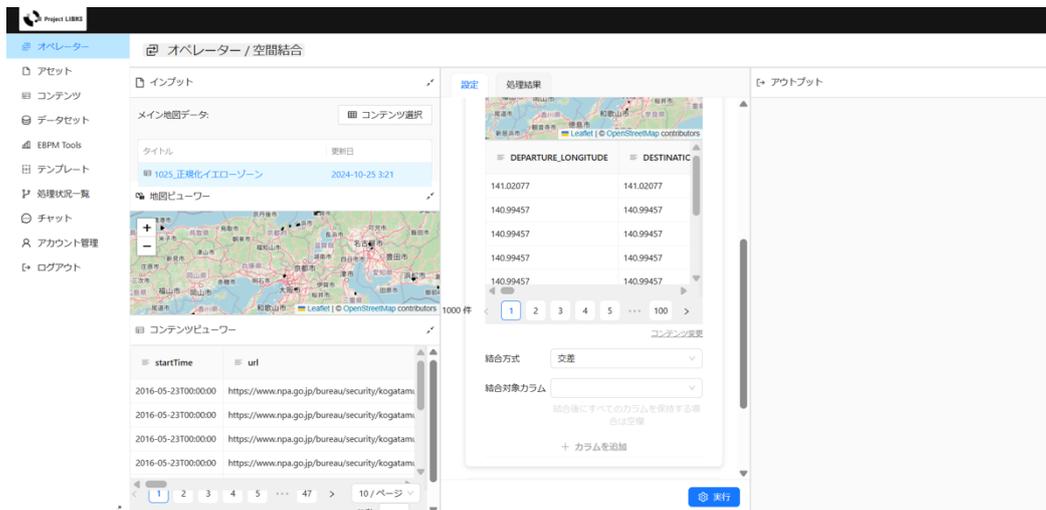


図 4-62 空間結合画面のイメージ（交差結合：処理実行前）

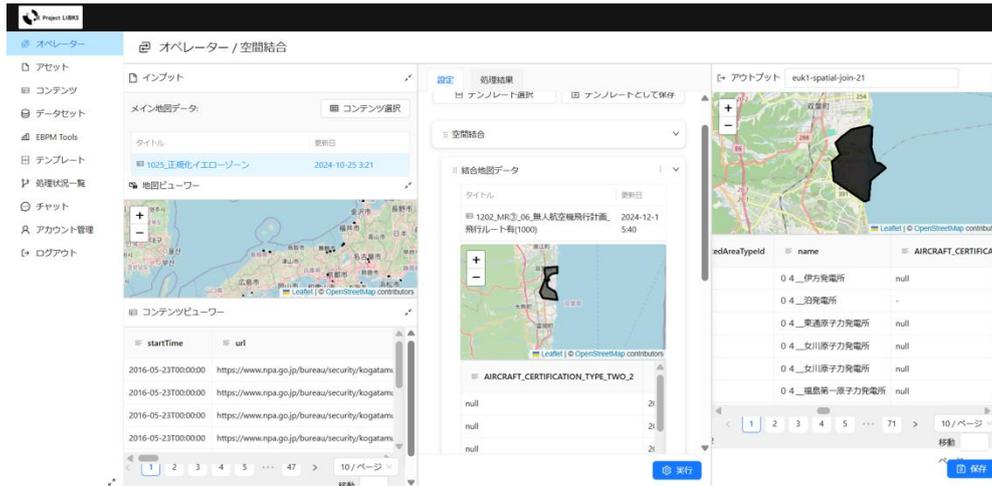


図 4-63 空間結合画面のイメージ (処理実行後)

## 10. 【SC012】クロス集計画面

### ● 画面の目的・概要

- 本画面では、【DT201】正規化地図データ、【DT202】正規化テーブルデータ、【DT206】データ結合済構造化データ、【DT207】空間結合済データ、【DT208】地域集計済構造化データ、【DT209】クロス集計済構造化データを対象に単一コンテンツのクロス集計処理を行う
- コンテンツ選択から集計するコンテンツを選択し、集計単位と集計対象と指定し、【FN025】データ集計処理（クロス集計）を実行する。
- 集計単位、集計対象はコンテンツのカラムから複数指定可能であり、集計対象の要約統計量（合計、平均、カウント）を集計できる。
- カラムを追加より集計単位、集計対象を追加可能である。
- 要約統計量で（合計、平均）を選択した場合、コンテンツで合計、平均カラムを作成する場合には合計、平均のチェックボックスをオンにする。
- 処理実行後、処理結果タブに自動で遷移し、実行結果を確認できる。
- 内容に問題がなければ、保存を押下しコンテンツとして保存する。
- 実行パラメータをテンプレートとして保存する場合には、テンプレートとして保存を押下し保存する。

● 画面イメージ



図 4-64 クロス集計画面のイメージ（合計/平均：処理実行前）

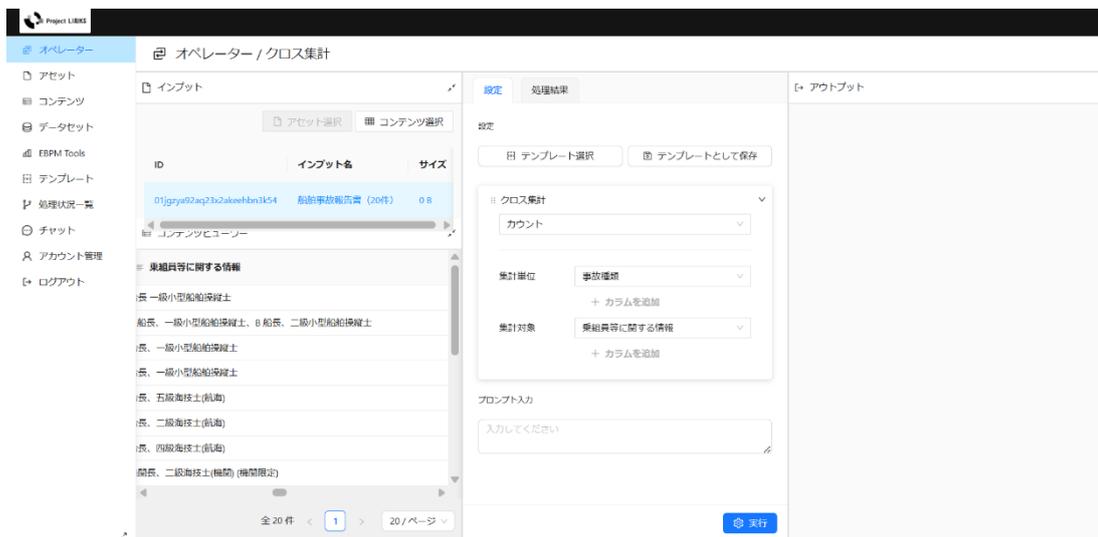


図 4-65 クロス集計画面のイメージ（カウント：処理実行前）

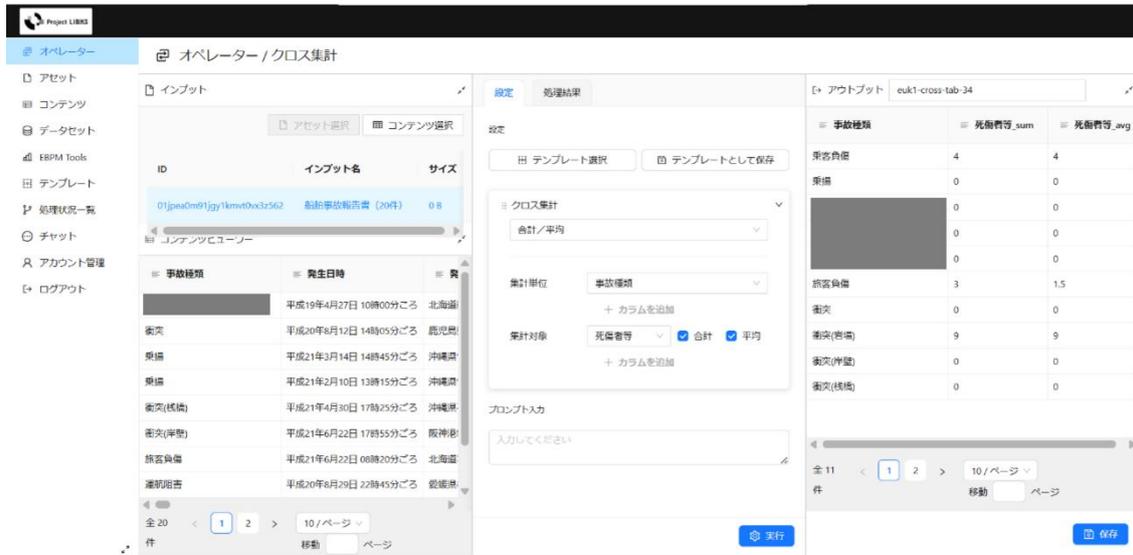


図 4-66 クロス集計画面のイメージ（処理実行後）

## 11. 【SC013】空間集計画面

### ● 画面の目的・概要

- 本画面は、【DT201】正規化地図データ、【DT207】空間結合済データ、【DT208】地域集計済構造化データを対象に空間集計を行う。
- コンテンツ選択から集計する地図データ（ポリゴンデータとポイントデータ）、集計単位と集計対象と指定し、【FN027】データ集計機能（空間集計処理）を実行する。
- 集計単位、集計対象はコンテンツのカラムから複数指定可能であり、集計対象の要約統計量（合計、平均、カウント）を集計できる。
- カラムを追加より集計単位、集計対象を追加可能である。
- 要約統計量で（合計、平均）を選択した場合、コンテンツで合計、平均カラムを作成する場合には合計、平均のチェックボックスをオンにする。
- 処理実行後、処理結果タブに自動で遷移し、実行結果を確認できる。
- 内容に問題がなければ、保存を押下しコンテンツとして保存する。
- 実行パラメータをテンプレートとして保存する場合には、テンプレートとして保存を押下し保存する。

● 画面イメージ

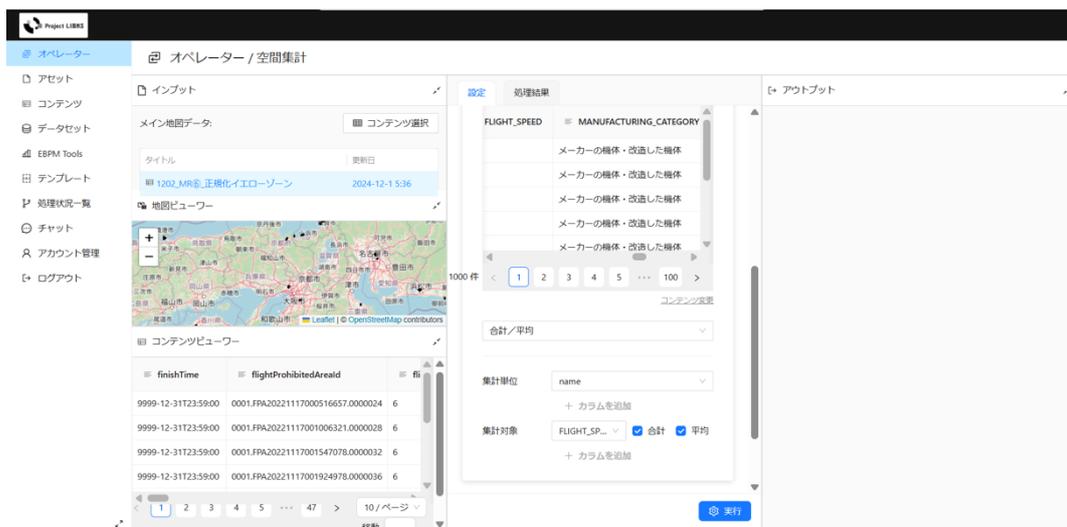


図 4-67 空間集計画面のイメージ (合計/平均：処理実行前)

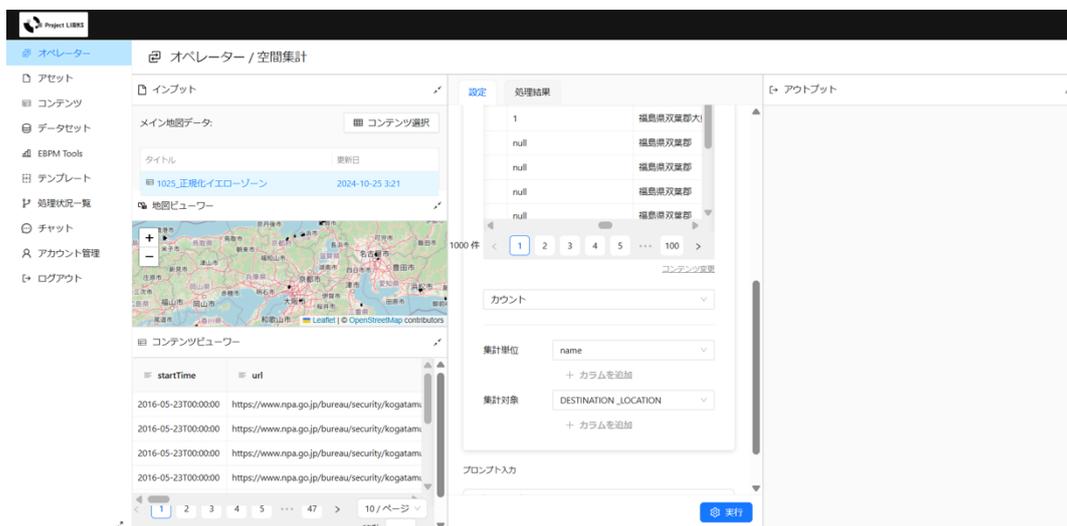


図 4-68 空間集計画面のイメージ (カウント：処理実行前)

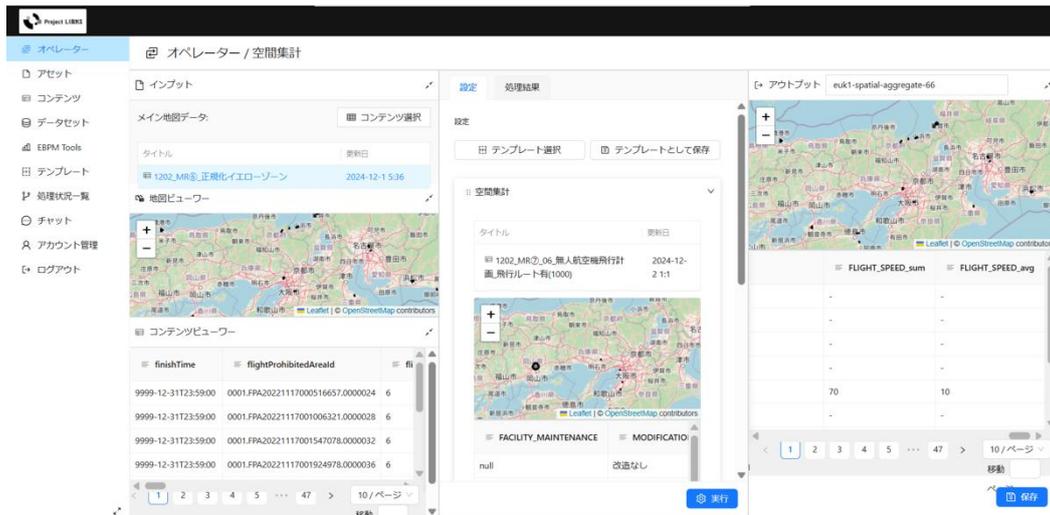


図 4-69 空間集計画面のイメージ (処理実行後)

## 12. 【SC014】テンプレート一覧画面

### ● 画面の目的・概要

- 【SC002】ホーム画面よりテンプレートメニューから遷移する。
- テンプレートには、ワークフローテンプレート、構造化テンプレート、結合前処理テンプレート、テキストマッチングテンプレート、空間結合テンプレート、クロス集計テンプレート、空間集計テンプレートが存在し、テンプレートを押下することでプレビューを確認できる。
- 削除ボタンを押下し、テンプレートの削除、テンプレートを編集でテンプレートの編集を実行できる。

### ● 画面イメージ

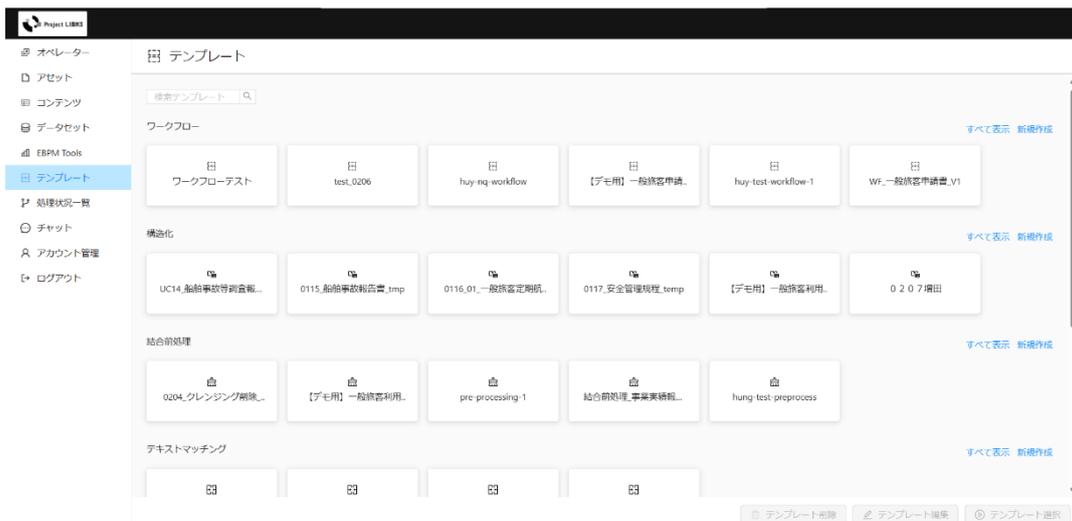


図 4-70 テンプレート一覧の画面イメージ

## 13. 【SC015】ワークフローテンプレート編集画面

## ● 画面の目的・概要

- 本画面は、【SC014】テンプレート一覧画面のワークフローテンプレートの新規作成ボタンから遷移する。処理を一気通貫で実行する際に必要なワークフローテンプレート作成・編集の際に利用する。なお、各オペレーターのテンプレート作成時には各オペレーター処理画面に遷移するものとする。
- アセットとして【DT001】原票データ、【DT002】テキストデータ、【DT203】OCR 処理済テキストデータを登録した場合には、構造化処理が必要となるため、ワークオペレーターの最初の処理として構造化を追加ボタンから構造化オペレーターを追加する。構造化処理が不要な場合には、オペレーターの追加ボタンを押下し、結合前処理、テキストマッチング、クロス集計、空間結合、空間集計から必要な処理を選択し、ワークフローに追加する。
- 本画面から実行した処理は、【SC016】処理状況一覧画面で確認が可能であり、対象の処理内容はテンプレート作成となる。処理ステータスが処理済となったことを確認後、開くを押下しテンプレートの中身の確認する。問題がなければ保存 テンプレートを押下し、ワークフローテンプレートに名前を付けて保存する。

## ● 画面イメージ

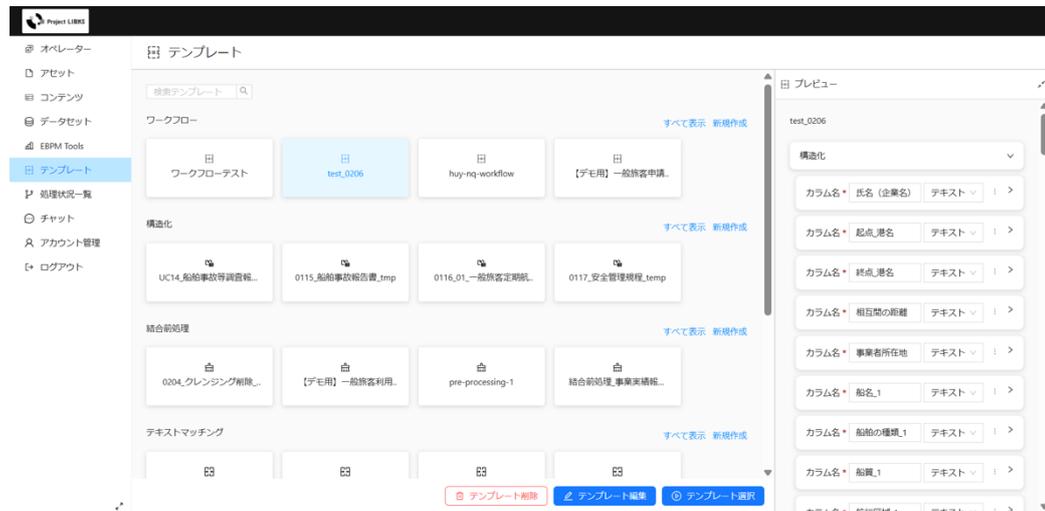


図 4-71 ワークフローテンプレート編集画面のイメージ（一覧から新規作成指定）

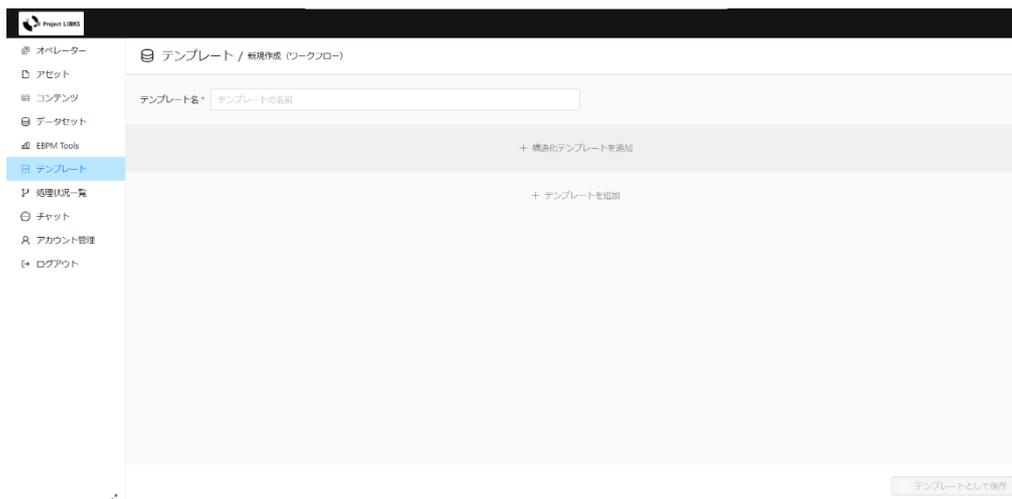


図 4-72 ワークフローテンプレート編集画面のイメージ



図 4-73 ワークフローテンプレート編集画面のイメージ (オペレーターの指定：構造化)

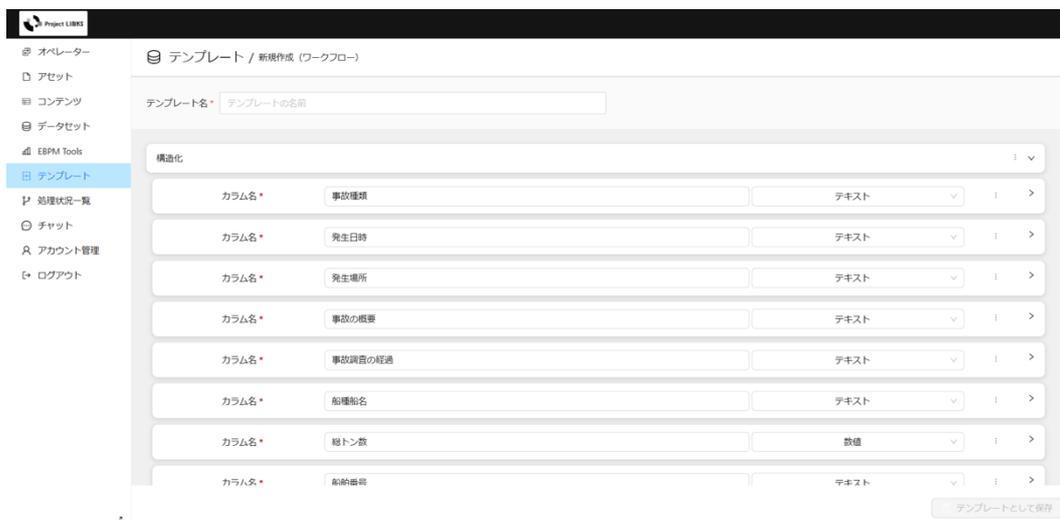


図 4-74 ワークフローテンプレート編集画面のイメージ (オペレーター指定後)



図 4-75 ワークフローテンプレート編集画面のイメージ（オペレーターの指定：構造化以外）

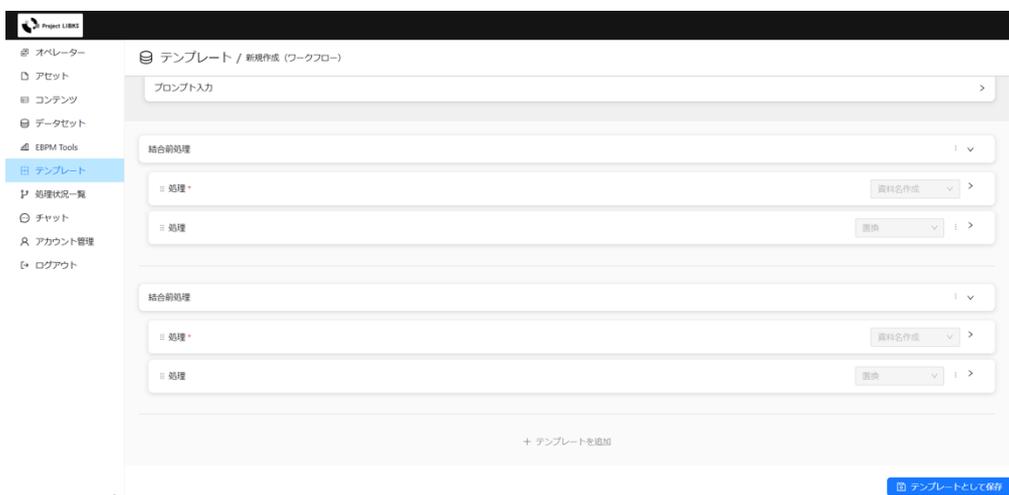


図 4-76 ワークフローテンプレート編集画面のイメージ（オペレーター追加後）



図 4-77 ワークフローテンプレート編集画面のイメージ（ワークフローテンプレートの保存）

## 14. 【SC016】 処理状況一覧画面

## ● 画面の目的・概要

- 処理状況一覧メニューから遷移する。本画面では、ユーザー自身が発行した処理結果のみ確認/編集できる。ファイル名、ファイル形式、処理内容、ステータス、実行者、実行日を一覧で確認できる。処理状況を押下することで、処理結果を確認できる。
- 処理状況を選択し、開くを押下することで、処理実行したオペレーターの詳細を開くことができる。

## ● 画面イメージ

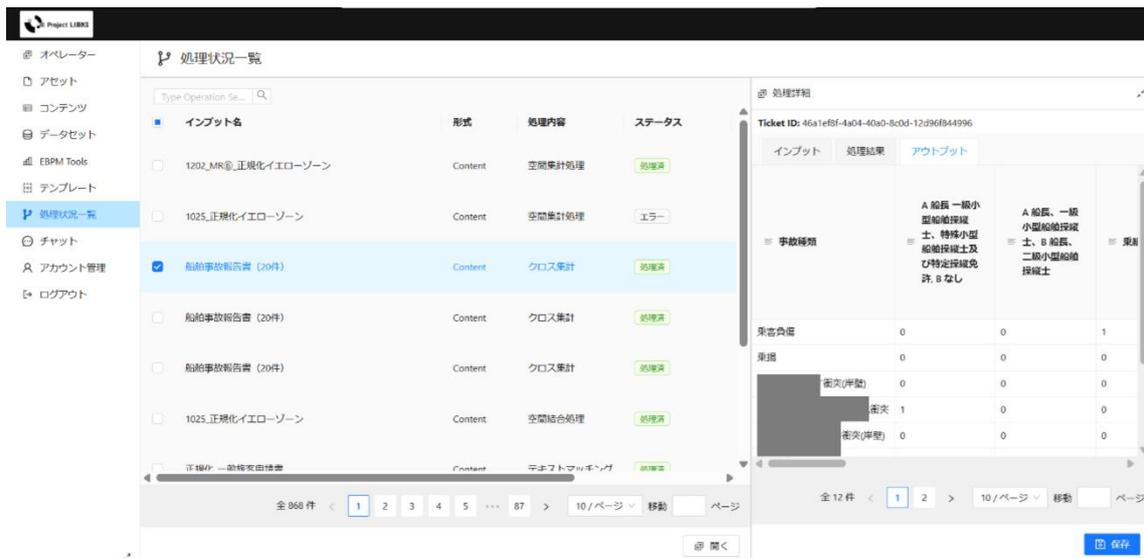


図 4-78 処理状況一覧の画面イメージ

## 15. 【SC017】 EBPM Tools 覧画面

## ● 画面の目的・概要

- EBPM Tools から遷移する。本画面では、ユースケースとして登録された EBPM Tools を一元的に管理する画面を提供する。
- LINKS Veda にログイン可能なユーザーは、EBPM Tools 一覧に表示されるすべてのユースケースにアクセスすることができる。

● 画面イメージ

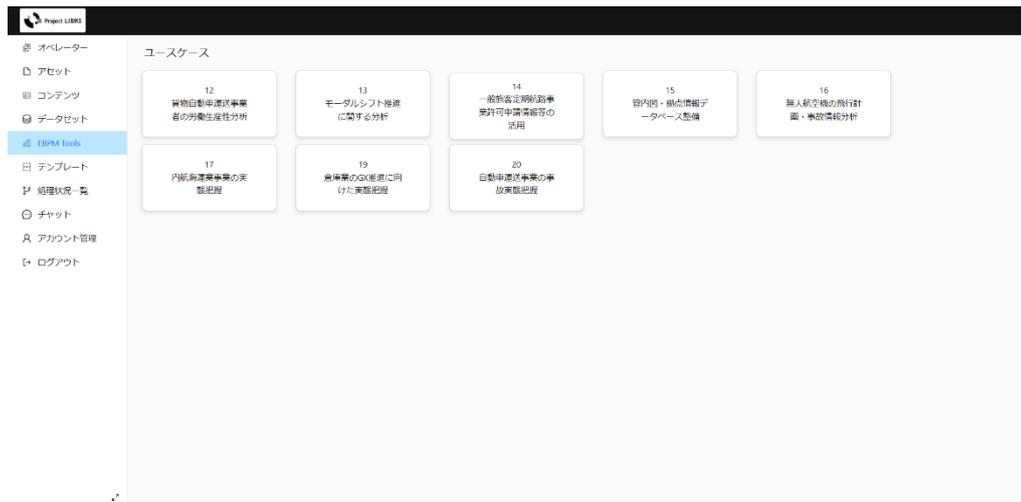


図 4-79 EBPM Tools 一覧画面のイメージ

16. 【SC018】 チャットボット画面

● 画面の目的・概要

- チャットメニューから遷移する。対話型にて問い合わせができる画面を提供する。
- 本画面からは、ユーザーがバクトルデータを選択しチャットを開始する。

● 画面イメージ



図 4-80 チャットボット画面イメージ

17. 【SC019】 データセット管理画面

● 画面の目的・概要

- データセットメニューから遷移する。LINKS Veda 内で管理するコンテンツ情報を同一の属性情報を持ったデータ群として管理するため、ユースケースの単位でデータセットとして管理する機能を提供する。

● 画面イメージ

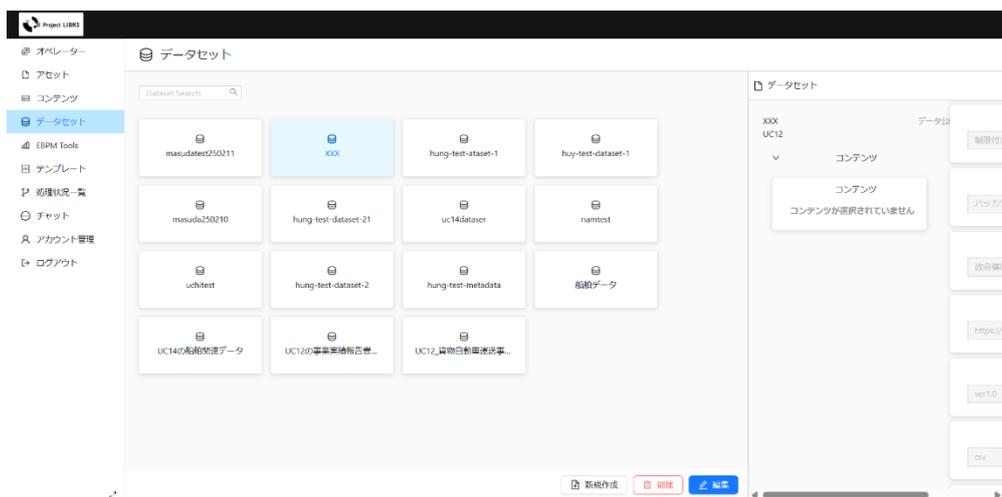


図 4-81 データセット管理画面イメージ (1/2)



図 4-82 データセットの管理画面イメージ (2/2)

2) CMS 画面

## 18. 【SC101】 CMS ログイン画面

- 画面の目的・概要
  - メールアドレス、パスワードを入力し CMS にログインするための画面。
  - 本画面は、システム管理者のみ参照可能であり一般ユーザーは参照しない。
- 画面イメージ
  - システム運用者のみログインするため、割愛する。

## 19. 【SC102】 CMS ホーム画面

- 画面の目的・概要
  - CMS ログイン後に表示される画面。CMS 上のメニューに遷移する。アカウント設定、ワークスペースの作成・削除、プロジェクトの作成・削除をする画面を提供する。
  - 本画面は、システム管理者のみ参照可能であり一般ユーザーは参照しない。
- 画面イメージ
  - ユーザーログインはないため、割愛する。

## 20. 【SC103】 CMS アセット画面

- 画面の目的・概要
  - Veda にてアセットとして登録されたデータを確認できる画面を提供する。
  - 本画面は、システム管理者のみ参照可能であり一般ユーザーは参照しない。
- 画面イメージ
  - ユーザーログインはないため、割愛する。

## 21. 【SC104】 CMS コンテンツ画面

- 画面の目的・概要
  - Veda にてコンテンツとして登録されたデータを確認できる画面を提供する。
  - 本画面は、システム管理者のみ参照可能であり一般ユーザーは参照しない。
- 画面イメージ
  - ユーザーログインはないため、割愛する。

## 22. 【SC105】 CMS アイテム画面

- 画面の目的・概要
  - Veda にてテーブルデータとして登録されたデータのレコードを確認できる画面を提供する。
  - 本画面は、システム管理者のみ参照可能であり一般ユーザーは参照しない。
- 画面イメージ
  - ユーザーログインはないため、割愛する。

## 23. 【SC106】 CMS アカウント管理画面

- 画面の目的・概要
  - CMS 上のアカウントの追加・削除、権限管理を行う画面を提供する。
- 画面イメージ
  - ユーザーログインはないため、割愛する。

## 4-7. 実証システムの利用手順

### 4-7-1. 実証システムの利用フロー

実証システムの利用フロー概要は以下の通りです。構造化するデータを準備し、システム内で構造化・加工を実行後、アプリケーションでの利用やオープンデータとして G 空間情報センターに公開する。

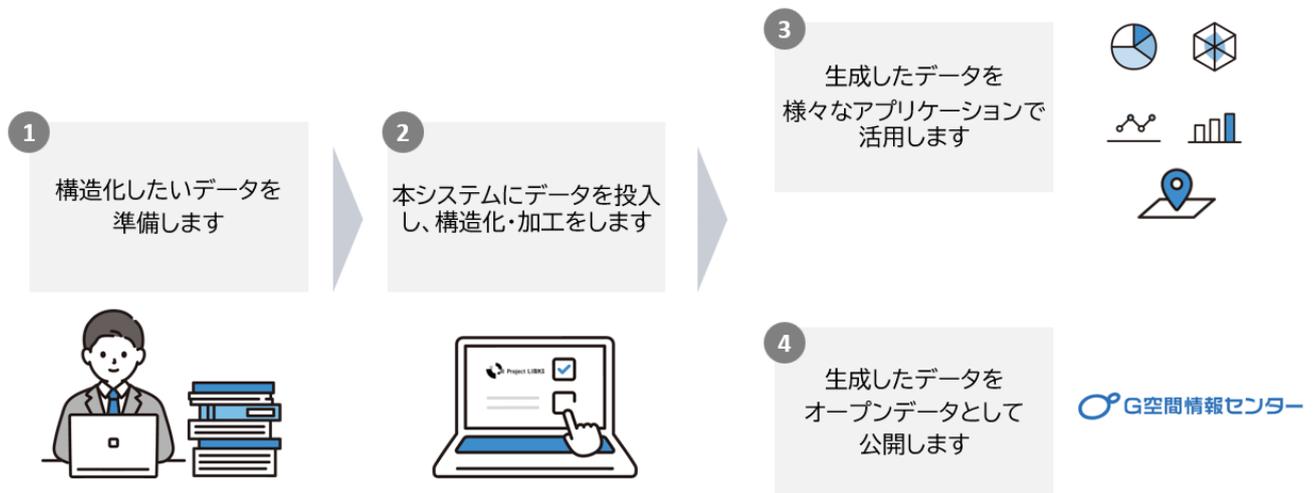


図 4-83 システム利用のフロー

## 4-7-2. 各画面操作方法

本システムの操作方法は、「LINKS Veda ユーザーマニュアル」として下記リンクにて公開している。操作マニュアルでは、本システムを利用するシステム概要、システム内で扱うデータ概要、本システムの操作手順等を記載した。

- 参照：[リンク](#)

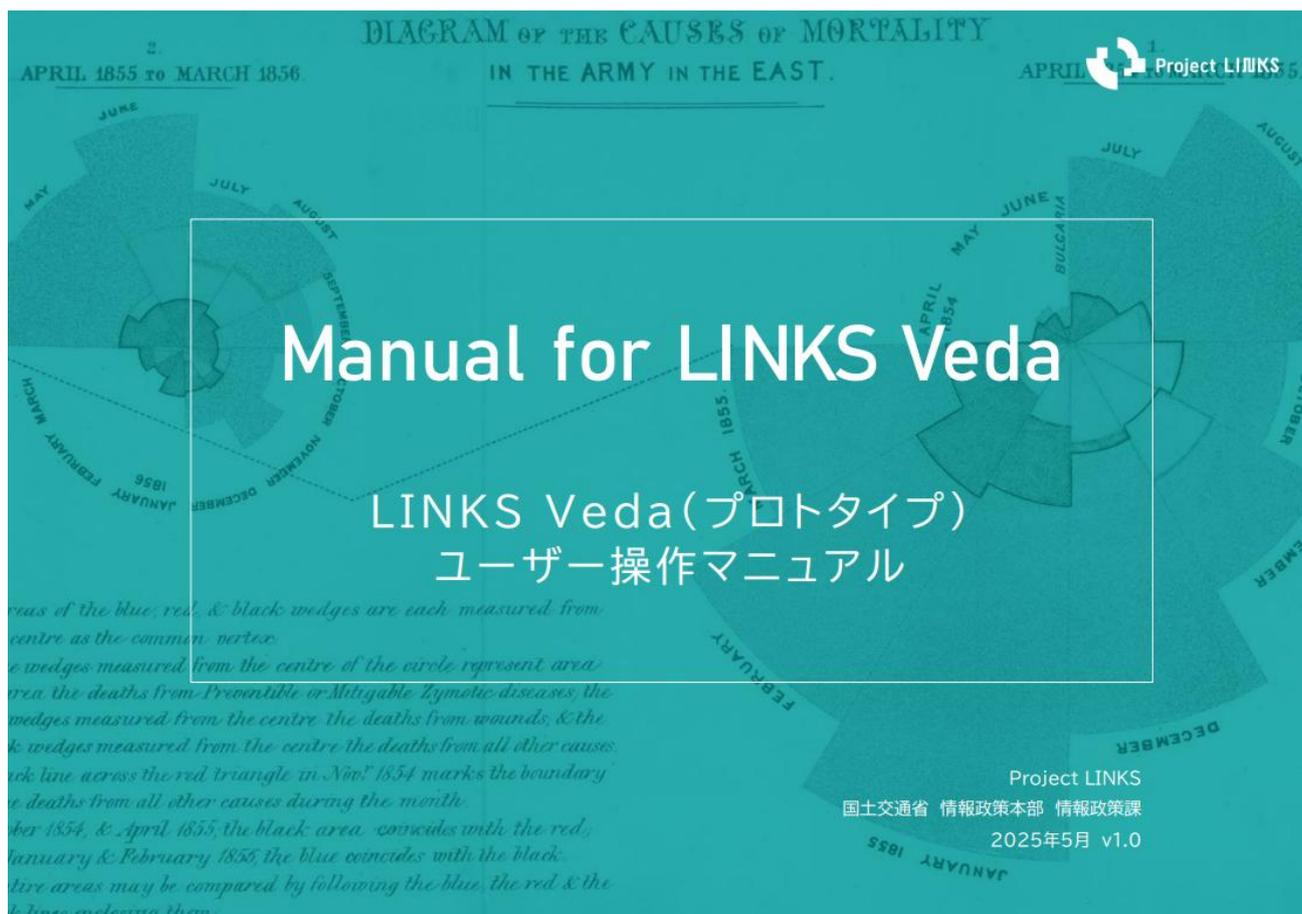


図 4-84 LINKS Veda ユーザーマニュアル

## 5. システムの非機能要件

### 5-1. 社会実装に向けた非機能要件

表 5-1 非機能要件一覧

カテゴリ	ID	項目	詳細
クラウド環境の構築	D011	本システムはクラウド環境で構築し、セキュリティ対策及び認証システムを備える。	<ul style="list-style-type: none"> <li>● Google Cloud Platform 上に構築し、セキュリティ対策及び Cloud Identity を利用した認証を行う。詳細の要件は「表 39：非機能要件一覧（クラウド環境構築に係る非機能要件）」に記載する。</li> </ul>
OSS 利用	D012	本システムは可能な限り OSS ライブラリ等を用いて構築する。	<ul style="list-style-type: none"> <li>● 利用するソフトウェア・ライブラリに記載の通り、基本的には OSS ライブラリを利用し実装をする。</li> <li>● 自然言語処理（NLP）ディープラーニング（AWS Claude API）についてはサーバーコスト等検討の結果有償 API を利用するものとする。</li> </ul>
OSS 提供	D013	開発成果は可能な限り OSS として提供する。	<ul style="list-style-type: none"> <li>● 本システムで利用したソースコードは OSS にて提供する。</li> </ul>
UI/UX への配慮	D014	本システムの主たるユーザーは国土交通省職員であることを鑑み、簡素かつ明快な UI/UX を設計する。	<ul style="list-style-type: none"> <li>● システム管理者が設定した情報をテンプレート化し、一般ユーザーはそれら呼び出すことで処理が自動的に実行される等、直感的に実行できる UI を設計する。</li> </ul>
スケーラビリティと拡張性	D015	システムは将来のデータ量の増加や新機能の追加に対応できるように設計する。	<ul style="list-style-type: none"> <li>● 将来のデータの増大を想定し、容量に余裕を持たせた設計を行う。</li> <li>● 将来的な機能の汎用化を見込み、個別のユースケースにおいても可能な限り統一性を持たせた設計とする。（以下、「表 39：非機能要件一覧（クラウド環境構築に係る非機能要件）」に記載の通り）</li> </ul>
データのバックアップ	D016	データの定期的なバックアップをする。	<ul style="list-style-type: none"> <li>● システム稼働開始後、データベースの日次バックアップの取得によりデータ復旧を可能とする。（Google Cloud にて自動で取得する。）（以下、「表 39：非機能要件一覧（クラウド環境構築に係る非機能要件）」に記載の通り）</li> </ul>

LINKS Veda は、省内データの正規化、結合・集計、公開までの自動化の機能的実証が目的である。以下の非機能要件は IPA 公開の非機能要求グレード 2018「システム基盤の非機能要求に関するグレード表」の「社会的影響がほとんどないシステム」を想定し、以下の非機能要件の達成を目指す。

表 5-2 非機能要件一覧（クラウド環境構築に係る非機能要件）

カテゴリ	ID	項目	詳細
可用性	NR001	稼働率	<ul style="list-style-type: none"> <li>● システムの稼働率 97%を目指す。：本システムは省内利用であるため、不特定多数の一般ユーザーへの影響はないものとし 97%（年間許容停止時間：10.96 日）を目指すものとする。</li> <li>● 外形監視における失敗率を検知する。（エラーコード、レスポンスの内容確認）</li> </ul>
	NR002	目標復旧水準（業務停止時）	<ul style="list-style-type: none"> <li>● 1 営業日前時点の日次バックアップを取得し、可能な限り迅速な復旧を行う。</li> </ul>
	NR003	耐障害性	<ul style="list-style-type: none"> <li>● システムを構成するサーバーを冗長化することで対応する。</li> </ul>
性能・拡張性	NR004	性能目標値	<ul style="list-style-type: none"> <li>● オンラインレスポンスは、データ容量にも左右されるが、データの構造化処理に対し、1 ファイルの処理当たり 2 分以内を目標値とする。</li> </ul>
	NR005	拡張性	<ul style="list-style-type: none"> <li>● 将来のデータの増大を想定し、容量に余裕を持たせた設計を行う。</li> <li>● 将来的な機能の汎用化を見込み、個別のユースケースにおいても可能な限り統一性を持たせた設計とする。</li> </ul>
運用・保守性	NR006	運用時間	<ul style="list-style-type: none"> <li>● 定時内 9 時 30 分～18 時 15 分の時間帯はサービスを提供する。夜間帯のシステムの運用は行うが、サービス提供の義務は負わない。</li> </ul>
	NR007	バックアップ	<ul style="list-style-type: none"> <li>● システム稼働開始後、データベースの日次バックアップの取得によりデータ復旧を可能とする。（Google Cloud にて自動で取得する。）</li> </ul>
	NR008	運用監視	<ul style="list-style-type: none"> <li>● 外形監視として HTTP リクエストを逐次投げることで、レスポンスを確認する。</li> </ul>
	NR009	マニュアル準備	<ul style="list-style-type: none"> <li>● システム運用開始までにユーザーマニュアルを整備し展開する。</li> </ul>
	NR010	構成管理	<ul style="list-style-type: none"> <li>● GitHub による構成管理を実施する。</li> </ul>
移行性	NR011	移行対象	<ul style="list-style-type: none"> <li>● システムの新規構築のため移行作業は発生しない。</li> </ul>
セキュリティ	NR012	アクセス・利用制限	<ul style="list-style-type: none"> <li>● Cloud Identity を利用することで認証を行い、認可に</li> </ul>

			<p>については実装することで不要なアクセス制限を担保する。メールアドレス、パスワードによる認証を行う。</p> <ul style="list-style-type: none"><li>● 上記に加えて、2段階認証を行う。</li></ul>
	NR013	データの秘匿	<ul style="list-style-type: none"><li>● システムへのアクセスを制御することで、データの安全性を担保する。</li></ul>
	NR014	不正追跡・監視	<ul style="list-style-type: none"><li>● ログイン/ログアウト履歴 (Cloud Identity)、ユーザーのアクセスログを取得し、不正な操作等を検出する。</li></ul>

## 6. 品質

### 6-1. 機能要件の品質担保

表 6-1 機能要件の品質担保方針

対象プロセス/ サブシステム	品質評価項目	目標値	期間の単位	アクティビティ
構造データ生成	UI	<ul style="list-style-type: none"> <li>● ユーザー操作（ノコード）</li> </ul>	2024年11月～ 2025年3月	<ul style="list-style-type: none"> <li>● 各機能単体テスト、結合テストによる検証</li> <li>● ユーザーによる有用性検証</li> </ul>
非構造データの自動 正規化機能構築	入出力対応	<ul style="list-style-type: none"> <li>● 入出力の一致</li> </ul>	2024年10月～ 2025年3月	<ul style="list-style-type: none"> <li>● 各機能単体テスト、結合テストによる検証</li> <li>● ユーザーによる有用性検証</li> </ul>
生成データ妥当性検証	実データ（手動作成）との比較	<ul style="list-style-type: none"> <li>● 文意の整合</li> </ul>	2024年10月～ 2025年3月	<ul style="list-style-type: none"> <li>● 実データとの比較による検証</li> <li>● ユーザーによる有用性検証</li> </ul>
正規化ローデータの 集計・加工機能	入出力対応	<ul style="list-style-type: none"> <li>● 入出力の一致</li> </ul>	2024年10月～ 2025年3月	<ul style="list-style-type: none"> <li>● 各機能単体テスト、結合テストによる検証</li> <li>● ユーザーによる有用性検証</li> </ul>
データ管理	登録、管理	<ul style="list-style-type: none"> <li>● 登録したデータが想定する表示、動作をする</li> </ul>	2024年11月～ 2025年3月	<ul style="list-style-type: none"> <li>● 各機能単体テスト、結合テストによる検証</li> <li>● ユーザーによる有用性検証</li> </ul>
データ配信	公開、表示、承認、動作	<ul style="list-style-type: none"> <li>● 一般公開したデータが想定する表示、動作をする</li> </ul>	2024年11月～ 2025年3月	<ul style="list-style-type: none"> <li>● 各機能単体テスト、結合テストによる検証</li> <li>● ユーザーによる有用性検証</li> </ul>

## LINKS24-03\_技術検証レポート\_データ管理システムのプロトタイプ開発

データ可視化	UI	<ul style="list-style-type: none"><li>● EBPM Tools に連携されたデータが正しく表示、動作する</li></ul>	2025年1月～ 2025年3月	<ul style="list-style-type: none"><li>● 各機能単体テスト、結合テストによる検証</li><li>● ユーザーによる有用性検証</li></ul>
--------	----	---	---------------------	--

## 6-2. 非機能要件の品質担保

表 6-2 非機能要件の品質担保方針

対象項目	品質評価項目	目標値	期間の単位	アクティビティ
LINKS Veda	稼働率	● 97%	2024年10月～ 2025年3月	● 運用テストによる 検証
	目標復旧水準	● -	2024年10月～ 2025年3月	● 運用テストによる 検証
	耐障害性	● -	2024年10月～ 2025年3月	● 運用テストによる 検証
	拡張性	● --	2024年10月～ 2025年3月	● 運用テストによる 検証
	運用時間	● 9~17時のサービス提供	2025年3月～	● 実運用での確認
	バックアップ	● 日次バックアップの取得	2024年10月～ 2025年3月	● 運用テストによる 検証
	運用監視	● -	2024年10月～ 2025年3月	● 運用テストによる 検証
	マニュアル準備	● -	2024年11月 ～2025年3月	● ユーザーによる有 用性検証
	構成管理	● -	2025年3月～	● 実運用での確認
	アクセス・利用制限	● -	2024年10月～ 2025年3月	● ユーザーによる有 用性検証
	データの秘匿	● -	2024年10月～ 2025年3月	● ユーザーによる有 用性検証
	不正追跡・監視	● -	2024年10月 ～2025年3月	● 運用テストによる 検証

## 7. 実証技術の機能要件の検証

---

### 7-1. 検証目的

本実証では、実証技術の機能検証として、EBPM 活用観点で整理したデータを LINKS Veda に Raw データとして投入し、想定通り正規化データが生成できることを確認する。

## 7-2. 検証ポイントと KPI

実証仮説の検証のため、以下の検証ポイントと KPI を設定し本システムを評価する。検証においてはシステム管理者<sup>\*注1</sup>向けと一般ユーザー向けの検証ポイントを分類し評価を行う。

\*注1：システム管理者の検証には、JV 内の検証も含める。

表 7-1 機能要件の検証ポイントと KPI

No.	検証メニュー	対象	検証ポイント	KPI
1	実証技術の検証：データ処理の正確性	システム管理者	今年度の整備対象となるすべての Raw データを投入し、Veda を介して想定通りに Defined Data (EBPM データ、オープンデータ) が生成できることを確認する。	シナリオ消化率：100%
2				シナリオ合格率：100%
3	実証技術の検証：構造データ生成技術	システム管理者	構造データ生成技術 (LLM) を用いて、非構造データから Defined Data が正しく生成されることを確認する。	自動評価 (出力スキーマの一致) 正解率 : 100%
4				人間評価 (内容の一致) 正解率 : 100%
5	実証技術の検証：チャット生成技術	システム管理者	チャット生成技術(RAG) を用いて、構造データ生成技術によって生成した Defined Data から正しく情報検索、回答できることを確認する。	正解率：100%

## 7-3. 検証方法

### 7-3-1. 実証技術の検証：データ処理の正確性

- 1) 検証目的
  - システムが仕様を満たしていることを確認するため。
- 2) KPI

表 7-2 データ処理の正確性の KPI

No.	評価指標・KPI	目標値	目標値の設定理由
1	シナリオ消化率	100%	<ul style="list-style-type: none"> <li>● システムの品質を測るために十分なテストが実施されていることを確認するため。</li> </ul>
2	シナリオ合格率	100%	<ul style="list-style-type: none"> <li>● システムを運用するにあたっての不具合が存在していないことを確認するため。</li> <li>● 運用で吸収可能な不具合（ユーザー負荷を鑑みてワークアラウンドが可能な範囲、運用への影響が低い不具合など）に関しては、その対応策が明確になっていること、および対応可否を事前に国土交通省情報政策課職員と合意の上で対応可否の判断を行う。</li> </ul>

## 3) 検証方法と検証シナリオ

- 検証の範囲は、各ユースケースで投入する Raw Data 全量を対象とする。
- ユースケース検証用に整理した処理フロー、プロンプトの内容と同等の処理を Veda で実行することで、システムを介してデータ登録、データ生成、データ配信まで一気通貫で実行できることを確認する。
- ユースケース検証用に事前に手動で作成したデータと Veda で生成したデータを比較し、差異がないことを確認する。

## 【検証に当たっての前提条件】

- 実行における LLM API コストとテスト期間が限られていることから、Raw Data 全量による検証は JV 内で実施するものとし、国土交通省情報政策課職員向けのテスト結果の報告にて確認とさせて頂く。
- 国土交通省情報政策課職員によって大量の PDF データを処理するリクエストがある場合には、事前に処理件数とタイミングをご相談頂く。

表 7-3 検証シナリオ一覧

No.	検証方法	詳細
1	データ登録（貨物自動車運送事業実績報告書）	<ul style="list-style-type: none"> <li>● 正常にアセットとしてデータ登録できること</li> </ul>
2	データ構造化処理（貨物自動車運送事業実績報告書）	<ul style="list-style-type: none"> <li>● 定義したスキーマ通りに構造化処理できること</li> <li>● コンテンツとして登録できること。</li> </ul>
3	データ結合前処理（データクレンジング・ジオコーディング：貨物自動車運送事業実績報告書）	<ul style="list-style-type: none"> <li>● 処理が正常終了すること。</li> <li>● 処理結果が処理状況一覧から確認できること。</li> <li>● 指定した型（和暦の日付や異常な値の数値等）に合わせてデータクレンジング・ジオコーディングが実行できること。</li> </ul>
4	データ登録（行政ポリゴンデータ）	<ul style="list-style-type: none"> <li>● 正常にアセットとしてデータ登録できること。</li> </ul>
5	データ結合前処理（データクレンジング：行政ポリゴンデータ）	<ul style="list-style-type: none"> <li>● 処理が正常終了すること。</li> <li>● 処理結果が処理状況一覧から確認できること。</li> <li>● 入力した項目通りにデータクレンジングが実行できること。</li> </ul>
6	空間結合（貨物自動車運送事業実績報告書＋行政ポリゴンデータ）	<ul style="list-style-type: none"> <li>● 処理が正常終了すること。</li> <li>● 処理結果が処理状況一覧から確認できること。</li> <li>● 結合結果が想定通りであること。</li> </ul>
7	最終構造化データ確認	<ul style="list-style-type: none"> <li>● 生成結果がユースケース検証用に事前に作成したデータと一致すること。</li> </ul>

## 7-3-2. 実証技術の検証：構造データ生成技術

## 1) 検証目的

- LLM によるデータ構造化の精度検証を行う。
- LLM によって機械的に抽出ができない場合には、手動でのバリデーションによって想定通りのアウトプットを得られることを確認する。

## 2) KPI

表 7-4 構造データ生成技術の KPI

No.	評価指標・KPI	目標値	目標値の設定理由
1	自動評価 (出力スキーマの一致) 正解率	100%	<ul style="list-style-type: none"> <li>● データ構造化結果が指定した出力スキーマどおりになっているかを目視で確認し、出力スキーマの一致を評価する。</li> <li>● 評価を行う際は、データ構造化対象の UC のデータを対象とし、1 データセットにつき 50 レコードを対象とする。ただし、入力データに誤りや文章で取得できないなどの異常がある場合には評価対象外とする。</li> </ul>
2	人間評価 (内容の一致) 正解率	100%	<ul style="list-style-type: none"> <li>● データ構造化結果と元の資料を目視で確認し、内容の一致を後述の正解率の指標を用いて性能を評価する。</li> <li>● 評価を行う際は、データ構造化対象の UC のデータを対象とし、1 データセットにつき 50 レコードを対象とする。ただし、入力データに誤りや文章で取得できないなどの異常がある場合には評価対象外とする。</li> </ul>

<評価指標の計算方法>

$$\text{正解率 (accuracy)} = (TP + TN) / (TP + TN + FP + FN)$$

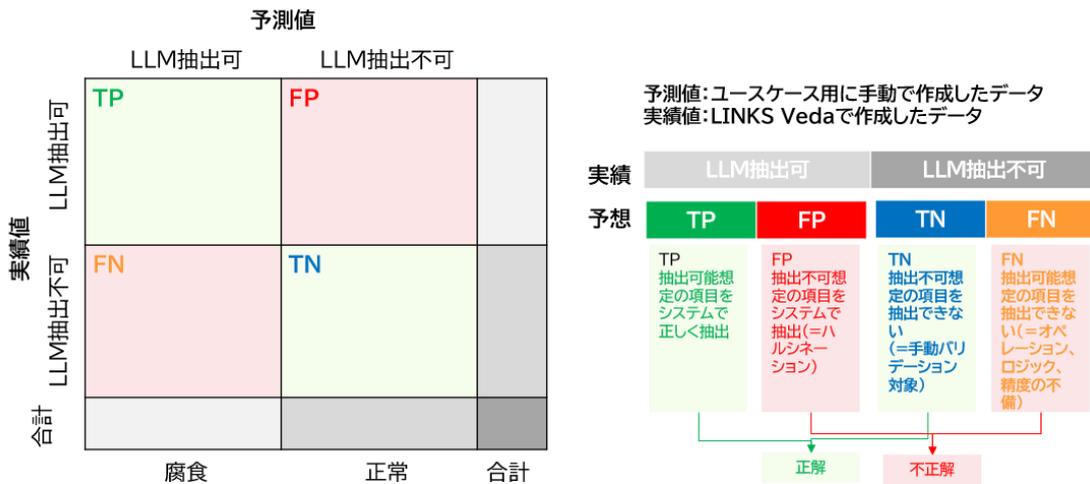


図 7-1 構造データ生成技術（評価指標説明）

3) 検証方法と検証シナリオ

- ユースケース用に手動で作成検証したデータ（予測値）と LINKS Veda を介して作成したデータ（実績値）双方を用いて、LLM によるデータの抽出可・抽出不可の項目の確かさを測るため、上記の評価指標を用いて正解率を計測する。
- 検証シナリオは II-3-①に示したデータ処理の正確性内のシナリオのデータ構造化を伴うデータにて包含する。
- 評価指標の計測と共に、各計測項目を以下の観点から分析する。
  - TP：こういったデータであれば精度高く抽出が可能か。
  - FP：抽出可能想定データの抽出できない場合の原因がどこにあるか。（オペレーションのミス、LLM の精度、内部ロジックのバグ等の切り分けを行う。）
  - TN：ユーザーが手動でバリデーションを行う必要のあるデータパターンはどのようなものか。
  - FN:ハルシネーションがどの程度発生しているか。

表 7-5 検証シナリオ一覧（構造データ生成技術の精度）

No.	検証方法	詳細
1	データ登録	● データ構造化処理が必要な Raw Data すべてを対象とする。
2	データ構造化処理	● 構造化処理を実行し、処理が正常終了することを確認する。
3	評価指標計測（正解率）	● 正解率 (accuracy) = $(TP + TN) / (TP + TN + FP + FN)$ <ul style="list-style-type: none"> <li>➢ TP：データ構造化処理対象データ全数に対し、抽出可能想定 の項目をシステムで正しく抽出できる場合。</li> </ul>

		<ul style="list-style-type: none"><li>➤ FP：データ構造化処理対象データ全数に対し、抽出可能想定 の項目にもかかわらず、誤った内容を抽出、もしくは何も抽出 されない場合。（=オペレーション、ロジック、精度の不備）</li><li>➤ TN：データ構造化処理対象データ全数に対し、空欄や該当の 記載がない等の抽出不可項目においては何も抽出されない場 合。（=手動バリデーション対象）</li><li>➤ FN：データ構造化処理対象データ全数に対し、空欄や該当の 記載がない等の抽出不可項目にもかかわらず、誤った内容を 抽出する場合。（=ハルシネーションの可能性はある）</li></ul>
4	計測項目ごとの考察	<ul style="list-style-type: none"><li>● 各計測項目として TP、FN、FP、TN を設け、それらの項目に分類 された原因を考察する。</li></ul>

## 7-3-3. 実証技術の検証：チャット生成技術

## 1) 検証目的

- チャット生成技術の精度検証を行う。

## 2) KPI

表 7-6 チャット生成技術の KPI

No.	評価指標・KPI	目標値	目標値の設定理由
1	正解率	100%	<ul style="list-style-type: none"> <li>● チャットの返答結果が、ユーザーが問い合わせした内容に合致する内容となっているかを目視で確認し、後述の正解率の指標を用いて性能を評価する。</li> <li>● 評価を行う際は、令和 4 年度政策評価チェックリスト、令和 5 年度国土交通省所管独立行政法人の各年度の業務実績評価、行政事業レビュー（令和 6 年度）等を対象とし、50 ケースを対象とする。</li> <li>● ただし、入力データに誤りや文章で取得できないなどの異常がある場合には評価対象外とする。</li> </ul>

&lt;評価指標の計算方法&gt;

$$\text{正解率 (accuracy)} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

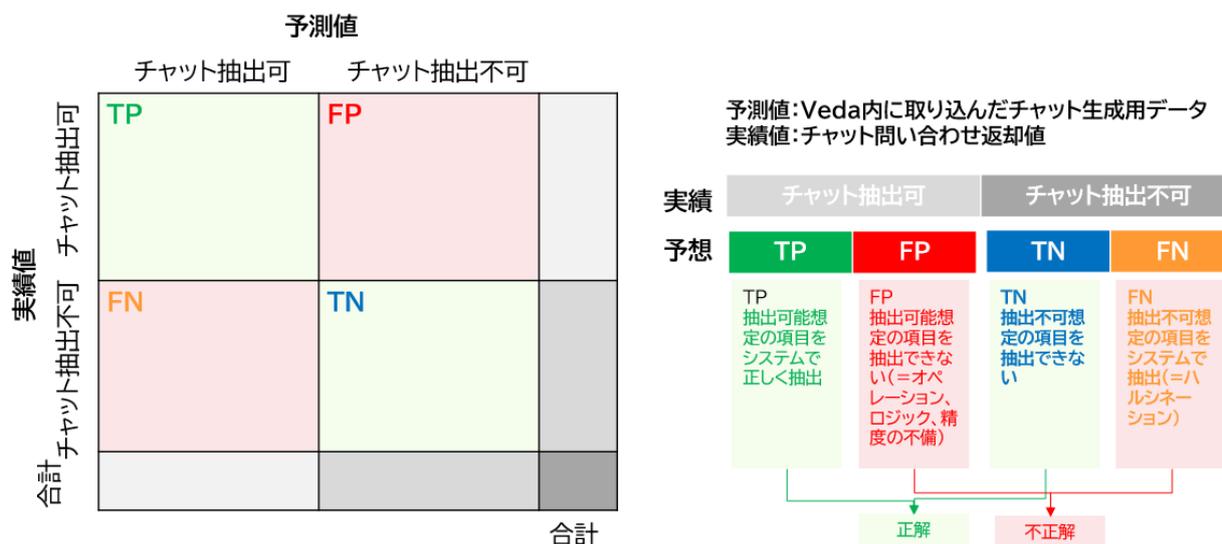


図 1 チャット生成技術（評価指標説明）

## 3) 検証方法と検証シナリオ

- チャットボットの返答内容の確かさを測るため、Veda 内に取り込んだチャット生成用データの内容と目視確認にて比較し、上記の評価指標を用いて正解率を計測する。
- 評価指標の計測と共に、各計測項目を以下の観点から分析する。
  - TP：こういったデータであれば精度高く抽出が可能か。
  - FP：抽出可能想定が抽出できない場合の原因がどこにあるか。（オペレーションのミス、LLM の精度、内部ロジックのバグ等の切り分けを行う。）
  - TN：ユーザーが手動でバリデーションを行う必要のあるデータパターンはどのようなものか。
  - FN:ハルシネーションがどの程度発生しているか。

## 【検証に当たっての前提条件】

- 評価を行う際は、令和 4 年度政策評価チェックリスト、令和 5 年度国土交通省所管独立行政法人の業務実績評価のデータを対象とし、30 ケースを対象とする。
- ただし、入力データに誤りや文章で取得できないなどの異常がある場合には評価対象外とする。

表 7-7 検証シナリオ一覧（構造データ生成技術の精度）

No.	検証方法	詳細
1	データ登録	<ul style="list-style-type: none"> <li>● チャットボットで利用するデータ：令和 4 年度政策評価チェックアップ評価書、令和 5 年度国土交通省所管独立行政法人を登録する。</li> </ul>
2	データ構造化処理	<ul style="list-style-type: none"> <li>● 構造化処理を実行し、処理が正常終了することを確認する。</li> </ul>
3	ベクトルデータ作成処理	<ul style="list-style-type: none"> <li>● 令和 4 年度政策評価チェックアップ評価書、令和 5 年度国土交通省所管独立行政法人の構造化データをベクトルデータ化する。</li> </ul>
4	チャット問い合わせ実行	<ul style="list-style-type: none"> <li>● チャットメニューより、問い合わせ処理を実行する。</li> </ul>
5	評価指標計測（正解率）	<ul style="list-style-type: none"> <li>● 正解率（accuracy） = <math>(TP + TN) / (TP + TN + FP + FN)</math> <ul style="list-style-type: none"> <li>➢ TP：チャット問い合わせに対し、投入した元データから返答した内容を正しく返却できる場合。</li> <li>➢ FP：チャット問い合わせに対し、返答可能想定項目を返答できない場合。（=ロジック、精度の不備）</li> <li>➢ TN：チャット問い合わせに対し、返答不可想定項目を抽出できないと返却できる場合。（=手動バリデーション対象）</li> <li>➢ FN：チャット問い合わせに対し、返答不可想定項目をシステムで抽出する場合（=ハルシネーションの可能性が有る）</li> </ul> </li> </ul>

6	計測項目ごとの考察	<ul style="list-style-type: none"><li>● 各計測項目として TP、FN、FP、TN を設け、それらの項目に分類された原因を考察する。</li></ul>
---	-----------	---

## 7-4. 検証結果

## 1) 実証技術の検証：データ処理の正確性

## ● 検証結果まとめ

- システム要件を満たすデータ処理機能の検証を実施した結果、全てのテストシナリオにおいて正常に動作することを確認した。
- Raw データの登録から構造化処理、データクレンジング、テキストマッチングまでの一連の処理フローが期待通りに機能し、最終構造化データが想定通りに生成されることを検証できた。

黄色セル：KPI 達成

青セル：KPI 未達

表 7-8 データ処理の正確性の検証結果

No.	評価指標・KPI	目標値	結果		示唆
			項目	評価	
1	シナリオ消化率 シナリオ合格率	100%	データ登録（貨物自動車運送事業実績報告書）	達成	● Raw データの登録機能は正常に動作し、アセットとして適切に登録できることを確認した。
2		100%	データ構造化処理（貨物自動車運送事業実績報告書）	達成	● 定義したスキーマに従った構造化処理が正しく実行され、コンテンツとして適切に登録されることを確認した。スキーマ定義に基づくデータ変換処理の精度が高いことが示された。
3		100%	データ結合前処理（データクレンジング・ジオコーディング：貨物自動車運送事業実績報告書）	達成	● 和暦の日付変換や異常値の数値処理など、指定した型に合わせたデータクレンジング・ジオコーディングが正確に実行されることを確認した。 ● 前処理の結果が処理状況一覧から適切に確認できる機能も実装されている。

4		100%	データ登録（行政区域ポリゴン）	達成	<ul style="list-style-type: none"> <li>● Raw データの登録機能は正常に動作し、アセットとして適切に登録できることを確認した。</li> </ul>
5		100%	データ結合前処理（データクレンジング：行政区域ポリゴン）	達成	<ul style="list-style-type: none"> <li>● プロンプト指定による柔軟なデータクレンジングが実行できることを確認した。</li> <li>● 処理結果の可視化機能も適切に動作しており、データの品質管理が容易である。</li> </ul>
6		100%	空間結合（貨物自動車運送事業実績報告書＋行政区域ポリゴン）	達成	<ul style="list-style-type: none"> <li>● 文書から抽出・ジオコーディングしたデータを、正常に空間結合ができ、結合結果が想定通りであることを確認した。</li> <li>● 複数ソースからの情報統合機能が高い精度で動作している。</li> </ul>
7		100%	最終構造化データ確認	達成	<ul style="list-style-type: none"> <li>● 生成された最終データがユースケース検証用に事前に作成したデータと一致することを確認した。</li> <li>● エンドツーエンドでのデータ処理パイプラインが正確に機能している。</li> </ul>

## 2) 実証技術の検証：構造データ生成技術

## ● 検証結果まとめ

- LLM によるデータ構造化の精度検証を実施した結果、出力スキーマの一致および内容の一致について高い精度を達成した。
- 機械的な抽出が難しいケースでも、適切なバリデーション処理により想定通りの出力を得られることを確認した。

黄色セル：KPI 達成

青セル：KPI 未達

表 7-9 構造データ生成技術の検証結果

No.	評価指標・KPI	目標値	結果		示唆
			項目	評価	
1	自動評価 (出力スキーマの一致) 正解率	100%	データ登録	達成	<ul style="list-style-type: none"> <li>● データ構造化処理が必要な Raw Data を全て適切に登録できることを確認した。</li> <li>● 多様なデータ形式への対応が可能であり、前処理段階でのデータ品質も確保されている。</li> </ul>
	人間評価 (内容の一致) 正解率		データ構造化処理	達成	
2		100%			

3		100%	評価指標計測（正解率）	未達	<ul style="list-style-type: none"><li>抽出可能な項目の正確な抽出（TP）と抽出不可項目の適切な処理（TN）は目標値を達成できなかったが、高い精度で実現されていることを確認した。</li><li>丸囲みや手書き文字の構造化において誤抽出（FP）が確認されたが、ハルシネーション（FN）の発生は最小限に抑えられている。</li></ul>
4		100%	計測項目ごとの考察	達成	<ul style="list-style-type: none"><li>各計測項目（TP/FP/TN/FN）の分析から、構造化データ抽出の精度向上要因と潜在的な課題を特定できた。</li><li>特に定型的な記述パターンからの抽出精度が高く、非定型的な記述に対しても適切な対応が可能であることを確認した。</li></ul>

## 3) 実証技術の検証：チャット生成技術

## ● 検証結果まとめ

- RAG を用いたチャット生成技術の精度検証を実施した結果、「独立行政法人の評価（令和5年度）」の一部カラムがデータ構造化処理においてデータ抽出に失敗したために、本データについてはチャット生成を検証できなかったが、想定した全てのユースケースにおいて高い応答精度を達成した。
- 構造化データを基にした情報検索と回答生成が正確に機能し、ユーザークエリに対して適切な情報提供が行えることを確認した。

黄色セル：KPI 達成	青セル：KPI 未達
-------------	------------

表 7-10 チャット生成技術の検証結果

No.	評価指標・KPI	目標値	結果		示唆
			項目	評価	
1	正解率	100%	データ登録	達成	<ul style="list-style-type: none"> <li>● チャットボットで利用するデータが正常に登録され、後続処理の基盤となるデータセットが適切に整備されていることを確認した。</li> <li>● 複数種類の文書（政策評価チェックリスト、業務実績評価等）の登録がエラーなく完了している。</li> </ul>
2		100%	データ構造化処理	未達	<ul style="list-style-type: none"> <li>● 「独立行政法人の評価（令和5年度）」を除き、登録されたデータから情報抽出を行う構造化処理が正確に実行され、チャット生成に必要な基礎データが作成されることを確認した。</li> <li>● 本データについては、①表が複数ページにまたいでいる箇所について、またいだ後のテキストがどの列の内容か認識できていない、②セルが閉じていない表の列の認識ができていない問題が生じている。</li> </ul>

3	100%	ベクトルデータ作成処理	達成	<ul style="list-style-type: none"> <li>● 構造化データから検索効率の高いベクトル表現への変換処理が正しく機能することを確認した。</li> <li>● 意味的な類似性に基づく検索を可能にする埋め込み処理が精度高く実行されている。</li> </ul>
4	100%	チャット問い合わせ実行	達成	<ul style="list-style-type: none"> <li>● ユーザーからの問い合わせに対して、25回中15回チャット回答に失敗していることを確認した。</li> <li>● チャット回答動作が不安定であり、実用的なユーザー体験において課題が残った。</li> </ul>
5	100%	評価指標計測（正解率）	未達	<ul style="list-style-type: none"> <li>● ユーザークエリに対する応答の正確性を評価した結果、正答率は50%（上述の回答失敗除く）と目標達成には至らなかった。</li> <li>● お問い合わせに対する正確な情報抽出ができておらず、回答が不十分または誤りとなる場合が多く確認された。</li> </ul>
6	100%	計測項目ごとの考察	達成	<ul style="list-style-type: none"> <li>● 正答以外では、お問い合わせに対して、別の情報を抽出し、FPの割合が高くなった。項目へのお問い合わせなど詳細または関係に関する内容ほど誤りが多く、確認された。</li> </ul>

## 8. 実証技術の非機能要件の検証

---

---

### 8-1. 検証目的

実証を行うにあたり、十分な処理能力（処理時間、処理件数）が担保されていることを確認する。

## 8-2. 検証ポイントと KPI

実証仮説の検証のため、以下の検証ポイントと KPI を設定し本システムを評価する。検証においてはシステム管理者<sup>\*注1</sup>向けと一般ユーザー向けの検証ポイントを分類し評価を行う。

\*注1：システム管理者の検証には、JV 内の検証も含める。

表 8-1 非機能要件の検証ポイントと KPI

No.	検証メニュー	対象	検証ポイント	KPI
1	データ処理効率の検証：データ処理時間	システム管理者	・インプットとして想定しうるデータパターンごとにデータを投入し、各機能の処理が完了するまでの時間を計測する。 データパターン：データ形式（PDF、DOCX、XLSX、GeoJSON、CSV）	・投入データを処理するのに要する時間 ・データの構造化処理に対し、1ファイルの処理当たり2分以内を目標値とする。（要件定義書より抜粋）
2	データ処理効率の検証：データ処理件数	システム管理者	・運用に耐えうるデータ件数の目安を計測する。	・スループット = 処理したデータ件数 / 測定時間

## 8-3. 検証結果

- 検証結果まとめ
  - OCR 処理、データ構造化、ハルシネーション検出を 1 セットとしたデータ処理機能の検証を実施した結果、正常に目標値を達成できることを確認した。
  - データ処理機能ではデータ構造化、ハルシネーション検出のために LLM を合計 3 回実行する必要がある、処理時間が目標値以内に達成することができた。
  - LLM の実行において AWS Bedrock API を用いているが、本 API では処理制限により、1 分間に実行できる回数が 20 回までしかできず、並列処理が正常に機能するかが重要である。複数モデルの活用による目標の処理時間に達成することを確認でき、並列処理が正常に機能することを確認できた。

黄色セル：KPI 達成	青セル：KPI 未達
-------------	------------

表 8-2 非機能要件の検証結果

検証内容	評価指標・KPI	目標値	結果		示唆
			項目	評価	
貨物自動車運送事業実績報告書のデータ処理	処理速度	1 ファイルの処理当たり 2 分以内	1 ファイルの処理当たり 48 秒	達成	● OCR 処理、データ構造化、ハルシネーション検出すべての処理が含まれた状態で正常に処理できることを確認できた。
	スループット	3 分あたりの処理可能なデータ件数 3 ファイル	3 分あたりの処理可能なデータ件数 10 ファイル	達成	● 並列処理が正常に動作していることを確認できた。

## 9. 原課向け検証会のチャット機能有用性検証

### 9-1. 検証目的

- 効果検証
  - アンケートを通じて本システムの有用性の評価を行う
- ユーザービリティ評価
  - 本システムの使い勝手に問題が無いかを確認する
- 業務効率化
  - 本システムが政策評価室の業務効率化に寄与するか検証を行う。

### 9-2. 被験者

本実証では、国土交通省における政策評価の実施を統括・支援する部門である国土交通省政策評価室の担当者をターゲットとしている。本実証実験では、検証期間内を設けて、本システムのチャット機能をユーザーに自由に使って頂く。加えて、ユーザーに該当する以下の方々にアンケートを行い、本システムの有用性を検証する。

対象ユーザー

分類	部署	人数
国土交通省	政策評価官室	1名

### 9-3. 検証方法

本システムのデモ環境を国土交通省政策評価官室担当者に開放し、以下の検証期間内で立会なしで個人の環境で検証頂く。検証期間後にアンケートを用いて検証を行う。

【検証期間】：4/1～4/18

#### ① 政策評価室向け

No.	検証ポイント	KPI	定量評価	定性評価
1	ユーザービリティに問題はないか	チャット問い合わせは容易でしたか	<ul style="list-style-type: none"> <li>● アンケートを実施</li> <li>● 選択肢は「とても満足」「満足」「やや満足」「どちらでもない」「やや不満」「不満」「とても不満」の7つで設定</li> <li>● 回答を集計し、各選択肢の選択率から評価（各設問で、過半数の「満足」以上回答を目標とする）</li> </ul>	<ul style="list-style-type: none"> <li>● アンケートの各設問に自由記入欄を設定</li> <li>● 定量アンケートの回答から対象を選定し、後日ヒアリングを実施</li> </ul>
2		回答内容が視覚的に分かりやすいか		
3		反応速度・レスポンスは十分か		
4	本システムが既存業務の業務効率化に寄与するか	回答内容が期待された内容になっているか		
5		本システムを導入することにより、政策評価のモニタリング業務内で役立てそうか		
6		業務効率化に役立てそうな場合、どの位の業務時間が短縮できそうか。 例：ある1つの政策の計画と実績の情報を抽出する場合に従来は30分程度かかっていたが、チャットを利用することで5分程度に改善できそうである。		

#### ● 検証実施方法

- 被験者に検証の趣旨を説明の上、配布したマニュアルと検証用シートをもとに検証期間内で自由に検証頂く。
  - ◇ 場所：被験者自身の作業場所
  - ◇ 機材：被験者のPCを利用し、Veda デモ環境にアクセスし実施頂く。（以下には国土交通省情

報政策課様の PC スペックを記載する。)

- CPU : AMD Ryzen (TM) 3 5400U
- メモリ : 32GB
- OS : Windows11
- ブラウザ : Microsoft Edge、Google Chrome
- 通信環境 : 国土交通省様省内 WiFi

● 事前準備

No.	項目	内容
1	被験者用デモ環境アカウントの発行	被験者である政策評価室ご担当者向けアカウントを1つ準備する。 ※アカウント発行用のメールアドレスを事前に受領する。
2	デモ環境データ登録、ベクトル化処理実行	被験者用アカウントが参照できるよう事前にデモ環境にチャット用データを登録、ベクトル化する。
3	検証関連資料の配布	マニュアル、検証シート、アンケートを被験者に配布する。
4	本実証実験に関する事前説明	被験者向けに、本実証実験の背景に関してご説明の場を設ける。

## 9-4. ヒアリング・アンケートの詳細

### 9-4-1. アジェンダの詳細

- 1. 事前説明（メール、個別会議の開催等）
  - ✧ 本実証実験でアプローチする課題や背景の説明（国土交通省情報政策課様よりご説明）
- 2. マニュアル、検証用シート、アンケートの配布
  - ✧ 利用マニュアル、チャットに入力する問い合わせ文のサンプルと回答結果サンプルを記載した検証用シート、アンケートを提供する。
- 3. 体験
  - ✧ 被験者の端末から Veda デモ環境にログインし、チャット機能から問い合わせを実行する。
  - ✧ 被験者が実行した問い合わせ文と回答結果を一覧に記載し記録をとる。
- 4. アンケート回答
  - ✧ 検証用シートと合わせて、アンケートに回答する。検証期間後に回収する。

## 9-5. 検証結果

チャット機能の検証では、政策評価室担当 1 名にご協力いただいた。本項目は、LINKS Veda チャット機能の有用性検証に関するアンケート結果を分析し、「使いやすさ」「業務効率化」「検証会」の 3 つの観点から、参加者からの意見を整理したものである。各カテゴリにおける良い意見、悪い意見、および今後の改善点について詳述する。

### 1. システムの「使いやすさ」について

No.	項目	思わ ない	やや思 わ ない	ど ち ら も な い	やや思 う	思 う	理由・改善点等
1	問い合わせ方法は容易でしたか？		○				問合せ方法そのものは容易でしたが、同一の質問内容であっても、質問文に含める部局名や政策分野を変更すると回答出力不可となる場合があり、どのような質問であればチャットボットが回答可能であるか推測することが難しかったです。

2	回答内容が視覚的に分かりやすいものでしたか？		○				指標の実績値の推移を表形式で出力するよう問合せた際、左から順に新しい年度で表にまとめられたため、左から順に古い年度でまとめた方が見やすいのではないかと感じました。
3	チャットのレスポンス速度（クリック後の反応・画面反映速度）は、満足できましたか？		○				回答に要する時間は、出力可能な場合・不可の場合共に数十秒～1分程度という実感で、回答が返ってくる場合はそこまで反応が遅いとは感じませんでした。一方で、1分程度要して出力不可となることもあったため、技術的に難しいかもしれませんが、出力不可の場合にはもう少し早めに反応があれば良いと感じました。

## 2. 業務効率化に関するご質問

No.	項目	思わない	やや思わない	どちらでもない	やや思う	思う	理由・改善点等
4	回答内容が期待された内容になっていましたか？	○					全体として回答出力不可となるが多かった印象です。また、回答が出力された場合でも、誤った内容が含まれていることがありました。
5	本システムを導入することにより、政策評価のモニタリング業務内で役立ってそうですか？	○					現時点のチャットボットの回答精度やレスポンス速度を考慮すると、チャットボットで情報を抽出するのに比べて、政策評価業務の担当者が、ホームページや各局のフォルダにおいて政策評価書

							を自ら探索した方が、短時間で正確性の高い情報を得られるのではないかと感じました。
--	--	--	--	--	--	--	--

6	業務効率化に役立 てそうな場合、ど の位の業務時間が 短縮できそうです か？	回答なし
---	--	------

## 3. 自由にお答えください

7	本システムを利用 した感想や、今後 期待することなど 自由に記載してく ださい	政策チェックアップ評価書では施策目標・業績指標ごとに担当部局や評価結果の情報が記載されているところ、各施策目標や業績指標に紐づくデータの構造化が上手くいっていないように感じました。チャットボットにより過去の評価結果や評価のベースとなっている施策の取組状況を横断的に探索するには、データ構造化や情報検索の精度をかなり向上させなければ難しいのではないかと考えます。
---	---	--

- 今回のアンケートでは、指標実績値の出力機能や応答速度に一定の評価が得られる一方、情報抽出の安定性やデータ構造化精度に課題が明らかになった。現状では手動による情報探索が短時間かつ正確である場合もあり、チャットボットの導入効果を最大限に発揮するにはさらなる改良が必要である。
- 実績値の年度並び順について、「新→旧」のデフォルト表示に対し「古→新」の方が視認性が高いとの意見が寄せられた。
- 応答時間は出力可否を問わず平均 30 秒～1 分程度を要し、特に出力不可時には長時間待機する必要があるため、UX 向上に向けて早期失敗応答やプログレス表示の導入が必要である。
- 全体として回答出力不可率の高さおよび出力された回答への誤情報混入が散見され、モデル精度向上および結果のクロスチェック機能強化が求められている。
- 政策チェックアップ評価書の施策目標・業績指標ごとのデータ構造化が十分に機能しておらず、横断的探索の実現には構造化精度の大幅な向上が不可欠である。
- 今後の改善案として、年度並び替え機能の実装、フェールファスト機能およびプログレス表示導入

による応答時間短縮、学習データ拡充とクロスチェック強化による回答精度向上、前処理パイプライン最適化による構造化精度改善、最適質問例やエラーケースをまとめたユーザーガイド整備を優先的に検討する必要がある。

## 10. 原課向け検証会の Veda 本体機能有用性検証

### 10-1. 検証目的

本実証では、国土交通省原課職員が LINKS Veda でデータ構造化からデータ正規化の処理を実行し、正規化済みのデータを活用する一連の流れを体験する。職員が実際にシステムを操作することで、システム利用のユーザービリティと業務効率化利用の価値検証を行う。

- 効果検証
  - アンケートを通じて本システムの有用性の評価を行う。
- ユーザービリティ評価
  - 本システムの使い勝手に問題が無いかを確認する。
- 業務効率化
  - 本システムが担当業務効率化に寄与する可能性があるか検証を行う。
- 業務高度化
  - 本システム担当業務高度化に寄与する可能性があるか検証を行う。

## 10-2. 検証ポイントと KPI

実証仮説の検証のため、以下の検証ポイントと KPI を設定し本システムを評価する。検証においてはシステム管理者<sup>\*注1</sup>向けと一般ユーザー向けの検証ポイントを分類し評価を行う。

\*注1：システム管理者の検証には、JV 内の検証も含める。

表 9-1 検証ポイントと KPI

No.	検証メニュー	対象	検証ポイント	KPI
1	有用性：ユーザービリティの検証	システム管理者 一般ユーザー	<ul style="list-style-type: none"> <li>・実証参加者を対象に LINKS Veda の各機能における使用感に関するアンケート・ヒアリング調査を実施する。</li> <li>・各評価項目の 5 段階のスコアで評価し、それらの算出に加えて、ユーザーフィードバックに対し優先度を設定し、システム改善に向けた対応および検討をする。</li> <li>・優先度は、難易度と影響度を元に国土交通省情報政策課職員と協議の上進める。</li> </ul>	・直感的にわかりやすい UI か
2				・各処理における設定条件、処理方法が分かりやすいか
3				・設定できる条件・使える機能は十分か
4				・抽出結果・処理結果がわかりやすいか
5				・反応速度・レスポンスは十分か
6	システムの価値検証：業務効率化・高度化	一般ユーザー	<ul style="list-style-type: none"> <li>・検証会終了後に「システムおよびシステムから出力されたデータ、可視化表現から得られたインサイト」について、参加者を対象にアンケートを行う。</li> <li>・アンケート結果から、システムで最も業務で役立てそうと思った内容（データ構造化、クレンジング、ジオコーディング、結合、集計等）のヒ</li> </ul>	<ul style="list-style-type: none"> <li>・システムを用いることによって明らかになった新たなインサイト（ユーザー自身がノーコードで構造データを作成することによって期待される成果、および改善点）</li> <li>・システムおよびシステムから出力されたデータの更なる活用に向けたアイデア</li> </ul>

			アリングを行うと同時に、活用アイデアや改善が必要な内容を抽出する。	
--	--	--	-----------------------------------	--

### 10-3. 被験者

本実証では、国土交通省における以下の原課を想定している。本実証実験では、本システム利用にあたってのデータの準備方法の紹介から、実際のユースケースで処理をしたデータを用いてデータ構造化からデータの生成処理を体験する。加えて、生成されたデータを用いた EBPM Tools をデモすることで、データがどのように可視化・活用されるかイメージを掴み、EBPM での活用可能性を検証する。検証会後にアンケートを収集し、結果をもとに本システムの価値を検証する。

表 9-2 対象ユーザー一覧

分類	部署	人数
国土交通省	航空局安全部無人航空機安全課	2名
	大臣官房総務課	2名
	住宅局 住宅総合整備課 住環境整備室	1名
	政策統括官付地理空間情報課	2名
	住宅局住宅戦略官付	1名
	不動産・建設経済局不動産課	2名
	海事局内航課	2名
	海事局安全政策課	2名
	物流・自動車局安全政策課	1名
	総合政策局モビリティサービス推進課	2名

## 10-4. 検証方法

本実証では、対面とオンラインのハイブリッド開催とし、実際にシステムを操作する場合には原則対面での参加とする。

なお、異なる原課担当が参加することから、検証データは「貨物自動車運送事業実績報告書」（10 ファイル）を利用するものとする。

本検証データは PDF 形式のデータであり、事業用自動車数・従業員数・輸送実績・事故件数などの情報を記載したデータである。

【検証期間】：3/14

表 9-3 検証方法一覧

No.	検証ポイント	KPI	定量評価	定性評価
1	ユーザービリティに問題はないか	直感的にわかりやすいUIか	<ul style="list-style-type: none"> <li>● アンケートを実施</li> <li>● 選択肢は「とても満足」「満足」「やや満足」「どちらでもない」「やや不満」「不満」「とても不満」の7つで設定</li> <li>● 回答を集計し、各選択肢の選択率から評価（各設問で、過半数の「満足」以上回答を目標とする）</li> </ul>	<ul style="list-style-type: none"> <li>● アンケートの各設問に自由記入欄を設定</li> <li>● 定量アンケートの回答から対象を選定し、後日ヒアリングを実施</li> </ul>
2		各処理における設定条件、処理方法がわかりやすいか		
3		設定できる条件・使える機能は十分か		
4		抽出結果・処理結果がわかりやすいか		
5		反応速度・レスポンスは十分か		
6		本システムを用いてデータ生成を自らで実行できそう・実行したいか		
7	本システムが既存業務の業務効率化・高度化に寄与するか	生成したデータの活用方法を具体的にイメージできたか		
8		本システムを用いて生成してみたいデータあるか		
9		本システムを用いることで、具体的にどのような業務の効率化・高度化ができそうか		

- 検証実施方法

- 本検証は対面とオンラインのハイブリッドでの開催とする。参加申し込みを受付、対面参加の被験者にのみ LINKS Veda アカウントを配布し、実際にシステム操作する。
- ◇ 場所：国土交通省会議室、teams オンラインのハイブリッド
- ◇ 機材：被験者の PC を利用し、LINKS Veda にアクセスし実施する。(以下には国土交通省職員様の PC スペックを記載する。)
  - CPU：AMD Ryzen (TM) 3 5400U
  - メモリ：32GB
  - OS：Windows11
  - ブラウザ：Microsoft Edge、Google Chrome
  - 通信環境：国土交通省様省内 WiFi

- 事前準備

表 9-4 事前準備内容

No.	項目	内容
1	検証会開催のアナウンスを実施	原課様向けに LINKS Veda 検証会の案内を送付。 対面参加、オンライン参加の被験者情報を収集する。
2	被験者用アカウントの発行	被験者である対面参加の各原課様ご担当者向けアカウントを1つ準備する。 ※アカウントはダミーアドレスで作成する。
3	検証関連資料の配布	マニュアル、アンケートを被験者に配布する。
4	検証会アジェンダ、オンライン開催用の teams 案内の送付	被験者向けに、検証会詳細の案内を送付する。

## 10-5. ヒアリング・アンケートの詳細

## 10-5-1. 検証項目と評価方法

表 9-5 検証シナリオ一覧

No.	検証方法	詳細
1	データ登録	<ul style="list-style-type: none"><li>● 非構造化データをアセットとしてデータ登録する。(実演のみ)</li></ul>
2	データ構造化処理	<ul style="list-style-type: none"><li>● スキーマを定義し、構造化処理をする。</li><li>● 構造化データコンテンツとして登録する。</li><li>● 処理実行時は事前に定義したテンプレートを利用する。</li></ul>
3	データ結合前処理 (データクレンジング)	<ul style="list-style-type: none"><li>● 型 (半角スペースを削除) を指定し、データクレンジングを実行する。</li><li>● 処理状況一覧からクレンジングの処理結果を確認する。</li></ul>
4	データ結合前処理 (ジオコーディング)	<ul style="list-style-type: none"><li>● 所在地を指定し、ジオコーディングを実行する。</li><li>● 処理状況一覧からジオコーディングの処理結果を確認する。</li></ul>
5	作成データの編集・確認	<ul style="list-style-type: none"><li>● コンテンツ詳細画面からフィールド内の値を更新する。</li></ul>
6	メタデータの入力 (コンテンツ)	<ul style="list-style-type: none"><li>● コンテンツ管理パネルからコンテンツ用メタデータを入力する。</li></ul>

## 10-6. 検証結果

アンケート協力ユーザーは合計 16 名。本項目は、LINKS Veda システムの有用性検証に関するアンケート結果を分析し、「使いやすさ」「業務効率化」「検証会」の 3 つの観点から、参加者からの意見を整理したものである。各カテゴリにおける良い意見、悪い意見、および今後の改善点について詳述する。

- システムの「使いやすさ」について(複数回答あり)

- アンケート結果

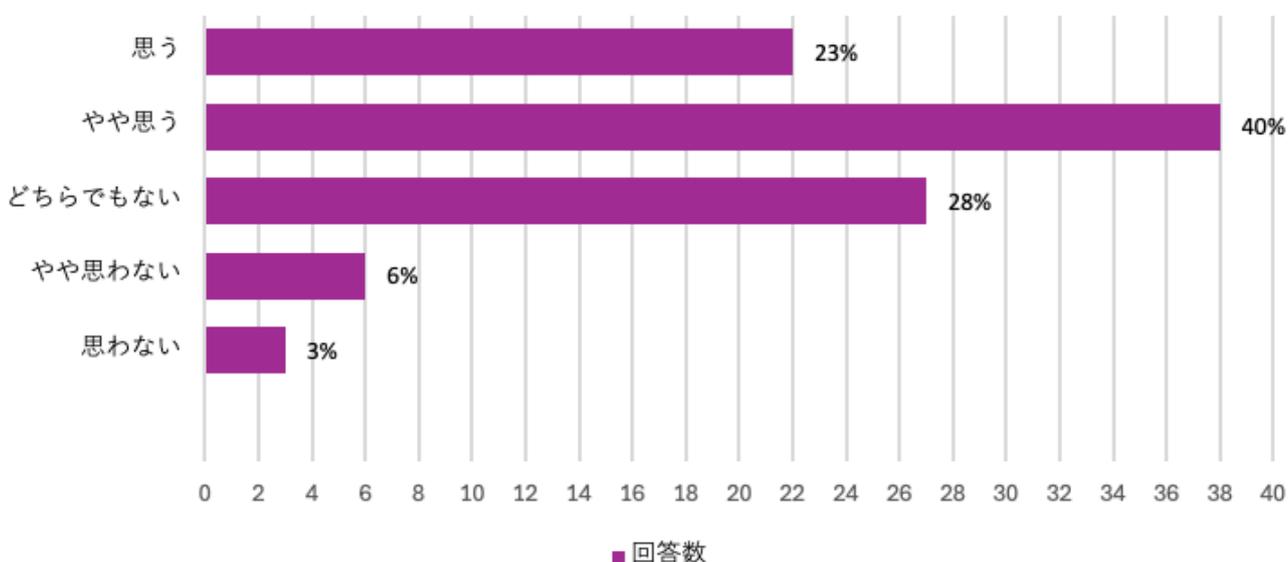


図 9-1 システムの「使いやすさ」についてのアンケート結果

- 全体を通じて、操作方法は直感的にわかりやすいという評価が半数からあり、各処理における設定条件や処理方法についても半数以上がわかりやすいとの意見を示した。特に抽出結果・処理結果の視覚的な把握のしやすさは約 4 分の 3 から評価されている。また、約 7 割が自らデータ生成を実行できそう・したいと回答している。
- 一方で、システムに搭載される用語の専門性が高いため、UI/UX の観点から工夫が必要であるとの指摘があった。設定できる条件・使える機能が十分ではないとの意見も 4 割強から示されている。また、画面サイズが小さい PC では表示が見切れる問題も報告されており、テスト環境のレスポンス速度についても約 3 割が満足していないことが明らかになった。
- 今後の改善点としては、専門用語の取り扱いを考慮した UI の改善、様々な画面サイズに対応できる表示の調整が必要である。また、ジオコーディング機能の操作性向上のためデータクレンジングの作業を自動化することや、処理速度の向上も求められている。自由記述からは、小型ラップトップでの操作性の問題や、オプション設定の影響が把握しづらいことも指摘されている。

● 業務効率化に関する意見

➤ アンケート結果

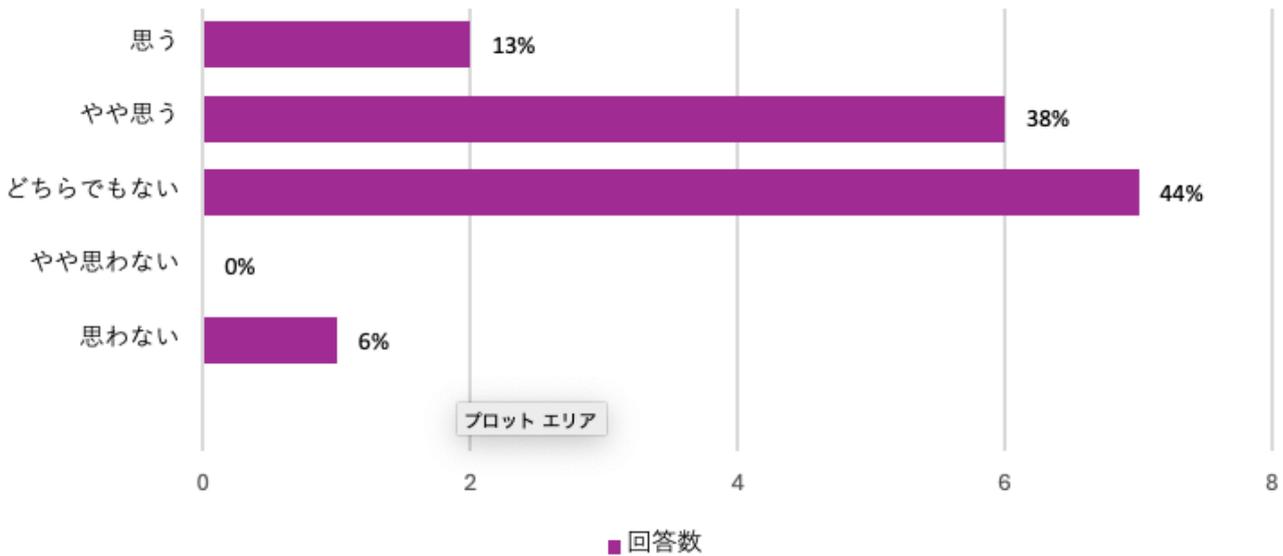


図 9-2 業務効率化に関するアンケート結果

- 生成したデータの活用方法を具体的にイメージできるとの評価が半数からあり、既存のデータ分析が容易になること、異なる情報システムに格納されたデータを自動集計できるようになることへの期待が示されている。
- 一方で、生成したデータの具体的な活用方法がまだ明確でないとの意見や、システムを用いて生成してみたいデータが明確でない状況も見られた。自由記述からは、レンタカー事業者からの報告書集計や、無人機の飛行申請管理などの具体的な活用例が挙げられている一方で、業務に占める比重が小さいため効率化効果が限定的との指摘もあった。
- 今後の改善点としては、データの具体的な活用事例や活用方法を明示することが求められている。また、既存の効率化プロジェクトとの重複を避けるための調整も必要とされている。

- 検証会への意見(複数回答あり)
  - アンケート結果

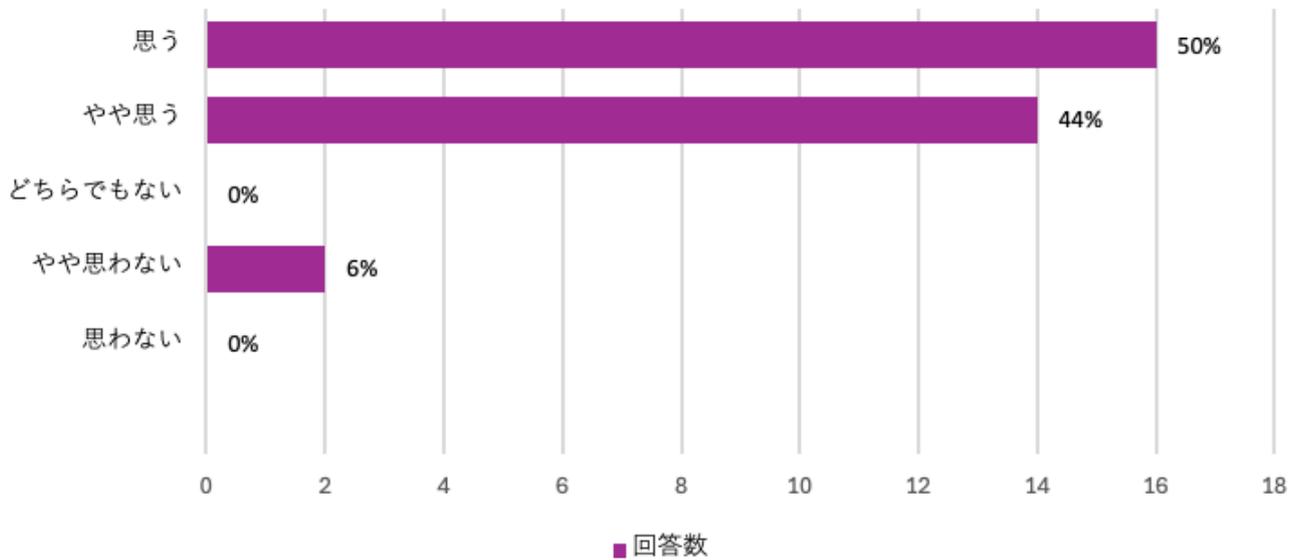


図 9-3 検証会についてのアンケート結果

- 当日の操作体験の説明はわかりやすいとの評価がほぼ全員から得られた。全体で示した上で説明したことが評価されている。
- 一方で、操作マニュアルについては約 6 割が分かりやすいと評価する一方、約 4 割が「どちらでもない」と回答している。自由記述では、特に、メリットのみの説明ではなく、最初から想定されるデメリットと思われる事項を全体で示したうえで説明するとわかりやすいとの声が上がっている。より単純かつ簡単な事例を用いた説明や、背景・問題意識・期待される効果を示すことの重要性も指摘されている。
- 今後の改善点としては、操作マニュアルの改善が必要である。また、Veda 環境では自前で非構造化データを用意できる人が想定されるため、非構造化データの提示方法を見直す必要性が指摘されている。加えて、データ集計のパターン別のマニュアルだけでなく簡単な動画の提供も有効との提案があった。

## 11. 有識者とのディスカッション

### 11-1. ディスカッションの目的

生成 AI を活用したデータの構造化手法について、国内外の技術動向や業種を問わず幅広い実務経験のある有識者として、以下の有識者とディスカッションする機会を設けた。

ディスカッションの主眼は、近年著しく進化している生成 AI の最先端技術の動向を正確に捉え、既存の実装事例がどの程度存在しているのかを確認することである。もし現時点で実装に至っていない分野や領域がある場合には、どのような形でアプローチし、具体的にどのような取組を行っていけばよいかといった実践的なアイデアやフィードバックを得ることを目的とした。

表 10-1 ディスカッションを行った有識者一覧

所属組織	所属部署	職種	氏名
アマゾンウェブサービス ジャパン合同会社	パブリックセクター技術 統括本部 スマートシティ /ヘルスケア担当	シニアソリューションア ーキテクト	今井 真宏
アマゾンウェブサービス ジャパン合同会社	パブリックセクター技術 統括本部 中央省庁担当	シニアソリューションア ーキテクト	日吉 康仁

### 11-2. ディスカッションから得られた結果

表 10-2 ディスカッションから得られた結果内容

調査項目	ヒアリング内容	ヒアリング結果	
1	モデル評価方法	<ul style="list-style-type: none"> <li>データ構造化した結果の評価において、オフライン評価（正解データに基づく評価）以外で、どのような評価方法や指標があるか。</li> </ul>	<p>評価方法は大きく二つが考えられ、このうち後者が該当すると考えられる。</p> <ul style="list-style-type: none"> <li>統計的スコア： LLM の出力に正解があり、コードベースで精度を計算できる場合</li> <li>モデルベーススコア： 決まった回答が存在しない非定型的なタスクにおける LLM の性能評価が必要な場合</li> </ul>
2		<ul style="list-style-type: none"> <li>本事業で開発している複数 LLM を用いて、ハルシネーション防止のための評価指標は適切か。</li> </ul>	<ul style="list-style-type: none"> <li>海外の研究論文において同じような手法が報告されていることをフィードバックいただけた。処理内容としても類似の処理が行われていることから、本事業で実装した手法には客観性がある手法であること</li> </ul>

			<p>を確認できた。</p> <ul style="list-style-type: none"> <li>● 出典： Manakul, Potsawee, Adian Liusie, and Mark JF Gales. 2023. Selfcheckgpt: Zero-resource black-box hallucination detection for generative large language models. In arXiv preprint arXiv:2303.08896.</li> <li>● <a href="https://github.com/potsawee/selfcheckgpt">https://github.com/potsawee/selfcheckgpt</a></li> </ul>
3	大量データ処理に向けた対策	<ul style="list-style-type: none"> <li>● Bedrock API には 1 分 20 リクエストといったスロットリングがあり、処理対象が多い場合にはどのような対策がとられているか？</li> <li>● また、取り扱うデータの性質上、国内リージョンだけで行う必要があり、クロスリージョン推論のような搭載機能を用いることができない。</li> </ul>	<p>以下のような方法で対策できる可能性があることが分かった。</p> <ul style="list-style-type: none"> <li>● 複数モデルの活用：クォーターは、リージョン毎、モデル毎に設定されているため、複数のモデルを使うことでスロットリングを回避</li> <li>● バッチ推論の使用：まとまったデータセットに対して実行することで、オンデマンドでない実行。ただし、非同期で実行されるため、バッチ処理とステータス管理の仕組みを作り込む必要がある。</li> <li>● プロビジョンドスループットの導入 (参考) Amazon Bedrock モデル推論 実践編 <a href="https://pages.awscloud.com/rs/112-TZM-766/images/AWS-Black-Belt_2024_Amazon-Bedrock-Model-Inference-b_0909_v1.pdf">https://pages.awscloud.com/rs/112-TZM-766/images/AWS-Black-Belt_2024_Amazon-Bedrock-Model-Inference-b_0909_v1.pdf</a></li> </ul> <p>Amazon Bedrock オンデマンドリソースを使用しているときの、"429 Throttling" エラーをトラブルシューティングする方法を教えてください。 <a href="https://repost.aws/ja/knowledge-center/bedrock-throttling-error">https://repost.aws/ja/knowledge-center/bedrock-throttling-error</a></p>
4	データ容量の多い入力データを対象にする場合	<ul style="list-style-type: none"> <li>● 読み込ませる資料が膨大であると、処理が実行できない場合がある。</li> </ul>	<ul style="list-style-type: none"> <li>● 基本的に生成 AI の処理では複雑なことを一度に行うことが苦手であるため、複数回に分けて処理を行うことや、生成 AI から</li> </ul>

	のデータ構造化の対応	<ul style="list-style-type: none"> <li>● 実行できたとしても、抜け漏れが入ってしまうことが多く、完全な結果を得られないことがある。</li> </ul>	python コードを生成して実行させるような進め方が考えられることをご示唆いただいた。
5	個別具体的、かつ、複雑なレイアウトのデータへのデータ構造化の対応	<ul style="list-style-type: none"> <li>● 入力データとして用いる原票には人間においては見やすいが、機械では読み取りにくいレイアウトとなっていることが少なくない。</li> <li>● たとえば、表においてヘッダ部分が複数列・複数行結合されている場合があり、資料の種類によって毎回異なることがある。</li> <li>● このような個別具体的なケースに対して、改善を図っていく必要がある。</li> </ul>	<ul style="list-style-type: none"> <li>● Few Shot Learning や Chain of Thought などのプロンプトエンジニアリングで改善できるかをまずは試すのが良い。</li> <li>● プロンプトだけで解決しない場合には、複数回に分けて処理を行うことや、生成 AI から python コードを生成して実行させる方法、ファインチューニングによって学習させていくことも考えられる。</li> </ul> <p>(参考) Chain of Thought によるプロンプト改善の例</p> <p>あなたはテーブルからデータを抽出し、構造化するスペシャリストです。データの行と列の関係を注意深く読み取り、データを抽出してテーブルを作成してください。また、テーブルのセルに複数のデータが存在している場合、一つ一つ読み取ってテーブルのセルとしてデータを出力してください。セル内にデータの意味づけがされている場合、カラムとみなして、テーブルを生成するようにしてください。マークダウンやワードのように見出しごとに箇条書きで記載されている場合も、テーブルとして読んでください。以下のステップに従ってください。</p> <p>Step1: マークダウンやワードのような見出しと箇条書きで表現されている部分があれば、テーブルとして構造を認識します。</p> <p>Step2: テーブルの列と行の関係を注意深く読み取ります。</p> <p>Step3: セル内に複数のデータが存在し、タグづけされているような場合、新規のカラムとして認識してください。</p> <p>Step4: 各行と列に該当する値を一つ一つ確認します。</p>

			<p>Step5: テーブルデータとして抽出し、文字起こしをしてください。</p> <p>Step1 から 5 をそれぞれ出力しながら処理してください。</p>
6	本開発において参考にできる事例	<ul style="list-style-type: none"> <li>● 本開発で行っているデータ構造化を中心として参考にできる生成 AI の実績・事例はあるか？</li> </ul>	<ul style="list-style-type: none"> <li>● 将来的な実用化を見据えた取組として、ヘルスケア領域における生成 AI の活用事例が特に参考になる。</li> <li>● たとえば、Preferred Networks (PFN) が公開している医師国家試験を解答するためのヘルスケアモデルは、医療系に特化した取組の代表例として注目度が高い。また、Llama のような汎用モデルを医療分野向けにカスタマイズし、手書きの間診票の解析など幅広い用途で応用が進んでいる。</li> </ul>

### 11-3. 成果と展望

#### 1) ヒアリング結果から、本開発に活かされた点

本ヒアリングによって、基本的な生成 AI を活用した開発方針や最新の開発ツール等について知見を得ることができ、データ構造化処理の改善に生かすことができた。

具体的には、本開発で扱うデータには 1 ファイルで 300 ページにわたる場合や、一度で数百ファイルを扱う場合があった。このような場合にご示唆をいただいた複数モデルを活用する方法によって、開発当初 200 ファイルの処理で 60 分程度要していた処理が、3 分強で処理できるようになるなど、大きな改善を確認することができた。また、複雑なレイアウトのデータにおいても、プロンプトの改善や複数回 LLM を実行するなどの方式を取り入れることで、改善を確認することができた。

#### 2) ヒアリング結果からの技術的な障壁、課題点

データデータ構造化の実装においては、未だ完全には対応しきれていない資料やファイルが残っており、現行の手法だけでは十分に対応しきれないケースが存在する。こうした未解決の領域については、ヒアリングにおいても示唆のあった生成 AI モデルのファインチューニングにより改善が見込まれるが、日本国内では事例が少ないため、さらなるリサーチとノウハウの蓄積が必要となっている。これらは今後の開発課題として取組むべき重要なポイントである。取り扱うデータの処理プロセスが複雑化しやすい以上、ワークフローの導入についても考慮していく必要がある。たとえば、今後は「dify」のようなオープンソースツールを組み合わせることで、汎用性を維持しながら拡張していくための発展的な開発が必要である。また、スケーラビリティやジョブ管理の観点では、生成 AI を含むデータ処理は、今後利用者が拡大していくことで一層、リソース負荷が高くなることが予想される。たとえば、AWS Step Functions のような

ジョブ管理機能付きワークフローを活用することで、インフラ管理の負担を軽減し、高負荷環境への対応を効率的に行うことも考慮して設計していく必要がある。

## 12. 成果と課題

### 12-1. 本実証で得られた成果

#### 12-1-1. AI を活用した自動データ生成における技術面での優位性

実証実験を通じて、以下のような AI を活用した自動データ生成の技術面での優位性が示された。

表 11-1121 AI を活用した自動データ生成における技術面での優位性

大項目	小項目	AI を活用した自動データ生成における技術面での優位性
データ準備	データ準備の容易さ	<ul style="list-style-type: none"> <li>一般的に、非構造データをシステムで取り込むには、事前に定義されたフォーマットに成形するなど人の手を介する作業が必要であるが、本システムでは、同じ様式のファイルを 1 フォルダにまとめるのみで準備を完了できる。</li> </ul>
システム・機能	多様なデータ形式の許容	<ul style="list-style-type: none"> <li>PDF、XLSX、DOCX、CSV、SHP、GeoJSON など様々なデータ形式を許容しており、ユーザーはそれらの違いを意識することなくデータ生成に必要な処理を行うことができる。</li> </ul>
	データ生成から可視化、オープンデータ化までオールインワンでの実装	<ul style="list-style-type: none"> <li>本システム内で、データ構造化、データクレンジング、データ加工（結合処理や集計処理）によるデータ生成処理のみでなく、EBPM での活用を想定した可視化アプリケーションでの利用が可能である。</li> <li>ユーザーが「データ公開」を操作するのみで、システム内で生成したデータを容易に G 空間情報センターにシームレスに公開できる。</li> </ul>
	秘匿化処理によるデータの安全性の担保	<ul style="list-style-type: none"> <li>本システムでは機密性の高い行政データを扱うことから、個人情報や法人等の権利利益にかかわる情報を統計化、匿名化することで秘匿性を担保している。</li> </ul>
	処理のテンプレート実装によるデータ品質の担保	<ul style="list-style-type: none"> <li>本システムでは、システム内の各処理をテンプレート化し、ユーザーが類似データの生成を容易かつ再帰的に実行できるようにすることで、データ品質を担保している。</li> </ul>
	ノーコード UI	<ul style="list-style-type: none"> <li>データ生成にかかる一連の処理を概念化したノーコード UI で開発したことから、プログラミングスキル、データサイエンススキルがないユーザーでもデータ構造化からデータバリデーション等の処理が可能である。</li> <li>ブラックボックス化や特定の技術担当者に依存することなく、システムを運用することが可能である。</li> </ul>

アルゴリズム	構造データ自動生成のための出力スキーマ作成支援	<ul style="list-style-type: none"> <li>● ユーザーが扱うデータソースや出力項目が多岐にわたる場合でも、LLM が自動的に候補をサジェストすることで、構造定義の手間を大幅に削減できる。</li> <li>● 表や箇条書きの見出しなどを読み取るプロンプト設計により、専門知識のないユーザーでも容易に適切なスキーマを構築可能。</li> </ul>
	OCR と LLM の連携による高精度なデータ抽出	<ul style="list-style-type: none"> <li>● 画像や PDF などの非構造データを OCR でテキスト化し、LLM を用いて構造データに変換する一連のパイプラインが整備されている。</li> <li>● 検索拡張 (RAG) などの技術を組み合わせることで、大規模文書からでも必要な情報を的確に取り出せる設計となっている。</li> </ul>
	自動バリデーションとクレンジングの実装	<ul style="list-style-type: none"> <li>● 生成されたデータに対して、ハルシネーションの検出や表記揺れの統一、異常値の除去などを自動的に行う仕組みを備えている。</li> <li>● 人間評価とのハイブリッド運用により、全自動化だけではカバーしきれない誤差や文脈上のミスを最小限に抑えつつ、ユーザーの負担を減らすことが可能。</li> </ul>
	並列処理やスケーラビリティへの配慮	<ul style="list-style-type: none"> <li>● 大量のファイルや高負荷環境でも対応できるよう、処理の分散化や複数モデルの併用を検討している。</li> <li>● API のレート制限を回避するために複数リージョンを使い分ける設計など、大規模運用に適したアーキテクチャを備えている。</li> </ul>
	チャット機能による対話型インターフェース	<ul style="list-style-type: none"> <li>● 最終的に生成された構造化データを基に、RAG を活用して自然言語での問い合わせに対応できる。</li> <li>● ユーザーはクエリを入力するだけで、背後で自動検索と LLM の回答生成が行われるため、直感的な操作で目的の情報を得やすい。</li> </ul>
	アップデートや拡張性の容易さ	<ul style="list-style-type: none"> <li>● 新たなデータが追加された場合でも、ベクトル DB の更新やモデルの差し替えで即座に処理精度を保てる。</li> <li>● 今後、OCR モデルや LLM のファインチューニングを行うことで、さらに精度を高める拡張性が確保されている。</li> </ul>
その他	オープンデータ化ガイドラインとユーザーマニュアルの整備	<ul style="list-style-type: none"> <li>● システムの操作方法を整理したユーザーマニュアルのみでなく、オープンデータ化にあたっての確認事項や留意事項をまとめたガイドラインを公開することで、誰でもデータ準備からデータ生成まで環境を実現。</li> </ul>

## 12-1-2. AI を活用した自動データ生成のビジネス面での優位性

実証実験を通じて、以下のような AI を活用した自動データ生成のビジネス面での優位性が示された。

表 11-2 AI を活用した自動データ生成におけるビジネス面での優位性

大項目	小項目	AI を活用した自動データ生成におけるビジネス面での優位性
業務効率化	コスト削減と生産性向上	<ul style="list-style-type: none"> <li>● 従来の手作業によるデータ入力・整理の大部分を自動化できるため、人的リソースを削減することができる。</li> <li>● デジタル化が進んでいない領域において、デジタル化を推進する基盤となりえる。</li> </ul>
	迅速かつデータに基づく意思決定の実現	<ul style="list-style-type: none"> <li>● データ構造化を自動化することで、情報取得から分析までのリードタイムが大幅に短縮される。</li> <li>● 長い待ち時間や煩雑な操作が不要になることで、統計資料の更新サイクルの短縮化と報告書の作成が短時間で完了できるようになる。</li> </ul>
業務高度化	スケーラビリティと横断連携	<ul style="list-style-type: none"> <li>● 大量の非構造データにも対応できる高いスケーラビリティを備えるため、業務量の増大や新規プロジェクトへの展開が容易。</li> <li>● 既存のデータやシステムとの連携により今後の部署横断的な連携強化が見込まれる。</li> </ul>
	差別化要因としての付加価値創出	<ul style="list-style-type: none"> <li>● AI を用いた高度なデータ処理を内製化・外販化することで、自社ソリューションの競争優位性を高められる。</li> <li>● データドリブンな施策展開を促進し、新たな施策を創出する基盤となる。</li> </ul>
セキュリティ	リスク管理とコンプライアンス対応	<ul style="list-style-type: none"> <li>● 自動評価やバリデーションを組み込んだフローにより、ヒューマンエラーを抑制し、品質管理を強化。</li> <li>● 大量の書類をデータとして取り出すことが可能になることで、対外的な規制・監査要件への対応もしやすくなる。</li> </ul>

## 12-1-3. AI を活用した自動データ生成の公共政策面での優位性

実証実験を通じて、以下のような AI を活用した自動データ生成の公共政策面での優位性が示された。

表 11-3 AI を利活用した自動データ生成における公共政策面での優位性

大項目	小項目	AI を利用した空き家推定の公共政策面での優位性
施策立案	施策立案の効率化	<ul style="list-style-type: none"> <li>● 大量の報告書・申請書などの行政文書を自動的に構造化し、分析に要する時間を大幅に短縮できる。</li> <li>● 正確かつ最新のデータを基に、施策立案に必要な情報収集や資料作成をスピーディに行えるようになる。</li> </ul>
情報公開	統計資料の迅速かつ正確な作成	<ul style="list-style-type: none"> <li>● 国や自治体で実施する各種調査・報告の集計を効率化し、結果公表までのリードタイムを短縮できる。</li> </ul>
	行政の透明化の促進	<ul style="list-style-type: none"> <li>● 構造化が行われたデータを公開しやすくなり、各種統計情報や報告書のオープンデータ化を迅速に進められる。</li> <li>● データ公開を進めることで、市民や企業、研究機関などによるデータ活用や行政の透明性が期待できる。</li> </ul>

## 12-2. 実証実験で得られた課題と対応策

実証実験における主な技術的課題は、以下のデータ構造化とチャットボット、アプリケーション構築上の課題に集約される。

表 11-4 実証実験で得られた課題と対応策

項目	課題	対応策
大規模ファイルのデータ処理	数百ページにおよぶ PDF や大量のテキストを扱う際、現行の API が一度に処理できるトークン数の上限を超える場合がある。 文書全体ではなく重要なページを自動判別して抽出する仕組みを導入しているが、依然として未対応ファイルの存在やリクエスト上限の問題が残っている。	重要ページ抽出アルゴリズムの導入 キーワード検索・Tf-idf 検索・ベクトル検索の 3 手法を組み合わせ、ページごとのスコアを統合することで、関連度の高いページを優先的に抽出する。これにより、数百ページに及ぶ PDF でもトークン数制限を超えない範囲で処理できるようになった。全体の約 1/4 のページ削減 (23.4%) を達成しつつ、重要情報の欠落を最小限に抑える効果が得られた。
データ構造化における実行時間の長さ	1 ページ処理あたり平均 44.5 秒程度を要し、OCR・LLM による構造化・自動評価の各ステップがユーザー体験を大きく左右している。 数十～数百ファイルを同時に処理するケースが想定される一方、LLM のレート制限により並列処理には制約がある。大量の行政文書を短時間で処理するためには、より高速かつスケラブルな仕組みが求められる。	複数リージョン・複数モデルの活用 AWS Bedrock の Haiku や Sonnet を併用することで、API のレート制限を分散し、同時処理リクエスト数を増やす試みを実施。処理能力を向上させ、大量ファイルの同時処理時にも全体の処理時間を短縮可能なことを確認した。 単純な文書処理には高速モデルを、複雑な文書には高精度モデルを使うなど、文書特性に応じた最適なモデルを選択する仕組みを検討。処理全体の効率化につながる可能性があるが、今後さらに最適な運用方法を模索していく必要がある。
画像判読の困難さ	手書き文字や印刷文字の OCR は一定の成果があるが、矢印や丸囲みなどの図形要素による意味づけを正確に読み取るのは難しい。 表と丸囲みの区別がつかないなど、レイアウト認識における誤認識が発生しやすい。 視覚的に示されている関係性を機械的に	画像前処理の導入 問題箇所を切り出して個別に OCR 処理を行うなどの対策を試み、一部精度向上を確認。また、丸囲み認識では囲み線が明瞭な場合は 80%以上の認識精度を確認。ただし、線が薄かったり四角囲みであったりと、形状が不明瞭な場合や表枠線や文字の重なり

	理解する手法の確立が課題となっている。	がある場合、矢印など図形のみから読み取れる情報の抽出には依然として課題が残る。OCR モデルのファインチューニングやユーザーが結果の妥当性を評価できるように信頼度スコアを設けるなどの改善が必要である。
訂正線やイレギュラーな記載への対応	行政文書では、手書きの訂正や追記が頻繁に行われるが、OCR は訂正前後の情報を区別できずに誤った値を取得する場合がある。 訂正線を引いて上書きされた数値など、どちらの情報を有効とみなすかを自動的に判断する仕組みが必要である。	NULL 値の優先的採用 訂正線による修正箇所など、誤抽出のリスクが高い項目は、あえて NULL として扱い、人間の確認を促すように試みたところ、誤った抽出を防ぐことができるようになった。しかし、訂正後の修正後の数字は位置づれ等により取得することができず、削除したままとなるため、根本的な解決策はできなかった。 今後の対策として、本対応に特化した画像認識モデルを構築し、OCR 前の段階でイレギュラーな記載を判別するなどの処理を土入していくことが考えられる。
チャット回答速度への対応	応答時間は出力可否を問わず平均 30 秒～1 分程度を要し、特に出力不可時には長時間待機する場合が生じていた。	早期失敗応答やプログレス表示等の機能を導入し、UX を向上する必要がある。
チャット回答品質への対応	全体として回答出力不可率の高さおよび出力された回答への誤情報混入が散見され、情報抽出の安定性やデータ構造化精度に課題が明らかになった。	学習データ拡充とクロスチェック強化による回答精度向上、前処理パイプライン最適化による構造化精度改善、最適質問例やエラーケースをまとめたユーザーガイド整備を優先的に検討する必要がある。

## 12-3. 今後の展望

LINKS Veda プロトタイプの開発を通じて得られた知見と技術的成果を踏まえ、今後さらなる発展が期待される方向性について述べる。国土交通分野の庁内情報を中心とした行政情報を構造データ化することで行政ビッグデータを構築し、さらにそのモデルをオープンに展開することで、行政 DX を単に省庁内に閉じたものにせず、全国や国際レベルの行政 DX 推進を実現することを目指している。以下では、重要な展望について詳述する。

### EBPM アプリケーションにおけるパフォーマンス最適化

本システムでは gzip 圧縮・MVT の採用・JSONL 等の行指向フォーマットの採用により、EBPM アプリケーションのパフォーマンス最適化を試みた。しかし依然としてデータ件数が増えると動作速度が低下することがある。また、JSONL のストリーミング処理を行う場合でも、1 行 1 行 JSON パースを行う必要があり処理効率が良いとは言えない状況にある。

こうした問題を解決するため、Parquet や GeoParquet などの列指向フォーマットを検討しうる。

列指向データフォーマットは、従来の行指向フォーマット（CSV や JSON）とは根本的に異なるデータ保存方式である。行指向では、各レコード（行）のすべてのフィールドが連続して保存されるのに対し、列指向では各フィールド（列）のすべての値が連続して保存される。例えば、行指向フォーマットでは「001,鈴木,34,東京」「002,田中,42,大阪」というように各行が連続して保存される。一方、列指向フォーマットでは「001,002,003」「鈴木,田中,佐藤」「34,42,28」「東京,大阪,福岡」というように各列ごとに値が連続して保存される。

この方式は、すべての列を必要とせず、特定の列のみを対象とすることが多いという、分析クエリの特徴に対し最大の利点を発揮する。例えば都道府県別の平均年齢を計算する場合、行指向では全レコードを読み込む必要があるが、列指向では「年齢」と「都道府県」の列だけを読み込めばよい。データ量が大きい場合、これにより大幅な I/O の削減が可能となる。また、列指向では同じデータ型の値が連続して並ぶため、型特有の効率的な圧縮が可能になる。例えば数値列は数値に特化した圧縮、文字列は文字列に適した圧縮を適用できる。

この列指向を採用したデータフォーマットとして Parquet がある。Parquet は、Apache Hadoop エコシステムのために開発された列指向のオープンソースデータフォーマットである。現在ではビッグデータ処理やデータウェアハウス、分析システムで広く採用されている。Parquet の主要な特徴としては、列指向による効率的な読み込み、スキーマ情報の保持、高い圧縮率等が挙げられる。また、GeoParquet は Parquet の拡張として位置付けられ、地理空間データを効率的に保存・処理するための仕様である。GeoJSON と同様の地理

情報を保持しつつ、Parquet の列指向構造の利点を活かしている。これにより大規模な地理データセットでも高速な空間クエリが可能になる。

このような列指向フォーマットの利点を活かすためには、それを効率的に処理できるエンジンが必要となる。Web ブラウザ環境では DuckDB-WASM が有力な選択肢となる。

DuckDB はアナリティクス処理に特化した組み込み型の列指向データベースエンジンであり、その特性は EBPM アプリケーションの要件と高い親和性がある。DuckDB-WASM はこの DuckDB を WebAssembly (WASM) にコンパイルしたもので、Web ブラウザ上で動作可能となっている。DuckDB-WASM を活用することで、サーバーとの通信なしにフロントエンド上で、Parquet をはじめとする様々なファイルやデータの処理が行え、SQL を用いてクエリを記述可能で、空間クエリにも対応可能である。

このアプローチにより、現行の JSONL ストリーミング処理と比較して、データ処理の効率性と応答性を大幅に向上させることが期待できる。

## 13. 用語集

用語	読み仮名	説明
CMS	しーえむえす	Contents Management System の略 DMS で登録したデータを保存するシステム。
DMS	でいーえむえす	Data Management System の略 非構造データ、構造データを登録・管理するシステムで、G 空間情報センター等外部システムへのデータ配信機能を含む。 原票等の非構造化データを構造化データへと変換、地図データと結合、描画する機能を保持する。
Defined Data	でいふあいんどでーた	正規化ローデータを結合・加工して構造データ化したデータであり、本システム内の処理が全て完了した状態のデータを指す。「最終構造化データ」とも表現する。
EBPM	いーびーぴーえむ	Evidence based policy making の略で、政策の企画を政策目的を明確化した上で合理的根拠（エビデンス）に基づくものにする。
EBPM Tools	いーびーぴーえむつーるず	Veda 内で作成したデータを活用したユースケースごとの可視化アプリケーションの総称。Veda のメニューからアクセスする。
GIS	じーあいえす	Geographic Information System の略で、地理情報システムとも呼ばれる。 地理的位置を手がかりに、位置に関する情報を持ったデータ（空間データ）を総合的に管理・加工し、視覚的に表示し、高度な分析や迅速な判断を可能にする技術。
LLM	えるえるえむ	Large Language Models の略で、非常に巨大なデータセットとディープラーニング技術を用いて構築された言語モデル（文章の並び方に確率を割り当てる確率モデル）を指す。
OCR	おーしーあーる	Optical Character Recognition/Reader（光学的文字認識）の略で、手書きや印刷された文字を、イメージスキャナやデジタルカメラによって読みとり、コンピュータが利用できるデジタルの文字コードに変換するシステムを指す。
Open Data	おーぶんでーた	誰でも許可されたルールの範囲内で自由に複製・加工や頒布などができるデータのことであり、DMS では一般の利用者が利用可能な G 空間情報センターにオープンデータとして公開する。「最終構造化データ」として作成されるデータの 1 つである。

OSS	おーえすえす	Open Source Software の略で、ソースコードが公開されたソフトウェアのこと。
Processed Data	ぷろせすどでーた	正規化処理によって作成された機械判読可能データを指す。ローデータに対して正規化処理を実行した状態でのデータであるため、「正規化ローデータ」とも表現する。
RAG	らぐ	Retrieval-Augmented Generation の略。生成ベースの NLP（自然言語処理）モデルに外部知識を取り込むことで、より正確で有用な回答を生成するための技術である。
Raw Data	ろーでーた	ユーザーがアップロードした何も手を加えていない状態のデータ（=アセット）を指す。非構造データ、構造データの 2 種類が存在する。
RDF	あーるでいーえふ	情報についての情報（メタデータ）を表記するための汎用的な手法を定めたデータ形式。機械判読がしやすい形式としてオープンデータの整備や公開に用いられる。
Re:Earth	りあーす	Eukarya が主導で進める OSS プロジェクト及びそのクラウドサービスを指す。本資料では「Re:Earth OSS」を OSS プロジェクトにおけるソースコードを指し、単に「Re:Earth」と称す場合は Eukarya が運営するクラウドサービスを指す。
原課	げんか	省庁で特定の政策を担当する局・課のこと。
公開 API	こうかいえーぴー あい	外部からアクセス可能な認証不要の API を指す。Veda では登録したデータが内部処理を経て承認者によって公開処理がなされると、そのコンテンツごとに紐づく URL が発行される。省内システム、外部システムでデータを利用する際には、この公開 API から配信された URL からデータを取得する。
項目リスト	こうもくりすと	Veda で定義するスキーマ情報、各カラムの説明、秘匿化の有無を示したものである。
構造化データ	こうぞうかでーた	構造化データとは、Excel などのように行と列を持ち、事前に定めた構造（=コンテンツとなる）に整形されたデータのこと。本システムでは、地図データ、テーブルデータが該当する。
非構造化データ	ひこうぞうか でーた	非構造化データとは、事前に整形されず元の形式のまま保存され、使用時まで処理されないデータのこと。本システムでは、原票データ、テキストデータが該当する。
交差結合	こうさけつごう	二つのテーブルデータのうち、一方の各行に対して、他方のテーブルデータのすべてを組み合わせる操作のこと。

最近傍結合	さいきんぼうけつごう	候補から一つずつオブジェクトを抜き出して最近傍を検索することを目的とする空間結合処理のこと。
正規化	せいきか	データを一定の規則に基づいて変形し、利用しやすくすること。
G 空間情報センター	じーくうかんじょうほうセンター	官民間わずさまざまな主体により整備・提供される 多様な地理空間情報を集約し、利用者がワンストップで検索・ダウンロードし利用できる、産学官の地理空間情報を扱うプラットフォーム。OSS の CKAN を用いて構築されている。
アカウント	あかうんと	データ管理者、登録者それぞれに1つずつ与えられるユーザーを指す。本システムでは原課ごとのアカウント管理とする。
アセット	あせっと	Veda に登録するファイルを指す。ファイル単体の場合、zip 化された複数のファイル群を指す。zip 化するアセットは同一のファイル種類、ファイル形式のみ対応し、異なる場合には同梱しないものとする。
インデキシング	いんできしんぐ	データから特定の要素を検索できるように、データにインデックスを付与すること。
オペレーター	おぺれーたー	Veda 内でデータ作成のために行うデータ構造化処理、データ結合前処理（データクレンジング・ジオコーディング）、データ結合処理（テキストマッチング）、クロス集計、空間結合処理、空間集計処理といった処理を指す。
コードリスト	こーどりすと	Veda で生成されるデータは、秘匿化処理のため一部コード化されている。各コードが示す内容を説明するための資料を指す。
コンテンツ	こんてんつ	なお、構造化処理によって作成された構造化データおよび結合・集約処理によって作成されたテーブルデータはそれぞれ新規コンテンツとして Veda 内で保持する。
スキーマ	すきーま	コンテンツのデータ構造を指す。ユーザーはデータ構造化処理実行時に生成するスキーマを指定し、データ抽出を行う。
アイテム	あいてむ	管理するデータの最小単位で、Veda の場合、コンテンツ内のレコードを指す。
ダッシュボード	だっしゅぼーど	本システムで処理したデータ（Defined Data）をグラフ、マップ形式で可視化表現を行うための GUI。ユースケースごとに複数のダッシュボードを保持する。
テキストマッチング	てきすとまっちんぐ	複数のデータ内にあるテキストのうち、一致する部分を見つける処理のこと。

データセット	でーたせっと	特定のルールによってコンテンツをまとめた集合体を指す。Veda内では、ユースケースごとにデータセットを作成し、G 空間情報センターに公開する際のデータセットの単位となる。
データモデル	でーたもでる	CMS 内でデータ型を管理する単位である。Veda に登録された構造データのスキーマ情報は CMS 内でデータモデルとして登録される。
テーブル	てーぶる	定義されたコンテンツに則り作成された構造データを指す。
テンプレート	てんぷれーと	データ構造化処理、データ結合前処理（データクレンジング・ジオコーディング）、データ結合処理（テキストマッチング）、クロス集計、空間結合処理、空間集計処理といったオペレーター情報と実行時に必要なパラメータ情報のセットを指す。テンプレートにはスキーマ情報（カラム名、カラムごとのデータ型）、抽出に必要な詳細条件（文書の章を示すポジション、単位、より詳細なプロンプトを記述する自由文記述項目の情報）を定義する。これらの情報を事前に定義しておくことで、データの登録者はテンプレートを選択することで任意のスキーマに則った構造化データを作成できる。
ノーコード	のーこーど	プログラミングをしないでシステムなどを開発する手法。
フィールド	ふいーるど	コンテンツ内の値を格納するための最小単位である。Excel のセルの単位を表す。
プロジェクト	ぷろじえくと	ワークスペースに複数作成可能で、データ管理の目的に応じて作成する。Veda では、1 つのみプロジェクトを作成する。
ベクトルデータ	べくとるでーた	文章や単語を数値ベクトルで表すことで、テキストデータである自然言語を機械が処理できるようにしたデータである。
ローコード	ろーこーど	可能な限りソースコードを書かずにアプリケーションを迅速に開発する手法。
ワークスペース	わーくすぺーす	複数のユーザーが同じワークスペースで作業できる場所のことを指す。Veda では、1 つのみワークスペースを作成する。
ワークフロー	わーくふろー	テンプレートとして登録した情報を一連の処理として流すためのテンプレート群を指す。過去に処理実績のあるデータであれば、ワークフローから実行することで、都度オペレーターを実行することなく一気通貫で処理できる。 ※補足 ただし、令和 6 年度の事業においては、単一の入力のみを想定した作りになっている。複数の入力を想定した実行対応は次年度に実施する。

データ構築基盤 LINKS Veda 開発  
技術検証レポート

2026年2月 発行

委託者：国土交通省 総合政策局 情報政策課

受託者：日建設計総合研究所・マイクロベース・ユーカリヤ・

ESRI ジャパン・三菱総合研究所