

2026 年 7 月 14 日

TOPPAN 株式会社

株式会社 NTT データ

株式会社 Fides Policy & Strategy Consulting

TOPPAN、NTT データ、FPSC、デジタル庁「ガバメント AI 源内」向け 大規模データセット整備を推進

政府共通データの AI-Ready 化と AI 向けデータセット標準の検討を通じ、
行政業務における AI 活用の高度化を支援

TOPPAN ホールディングスのグループ会社である TOPPAN 株式会社(本社:東京都文京区、代表取締役社長:野口 晴彦、以下 TOPPAN)、株式会社 NTT データ(本社:東京都江東区、代表取締役社長:鈴木 正範、以下 NTT データ)、株式会社 Fides Policy & Strategy Consulting(本社:東京都渋谷区、代表取締役:清水 庸介、以下 FPSC)の 3 社は、デジタル庁が推進する「ガバメント AI のための大規模データセットに係る調査・収集・加工等事業」(以下、本事業)において、それぞれの専門性を活かし、「ガバメント AI 源内」(※)で利用される大規模データセットの整備とその AI-Ready 化(AI が理解しやすいように、品質を担保して構造化すること)、および AI 向けデータセット標準の検討を実施します。

本事業は、政府が推進する「ガバメント AI 源内」の活用に向けて、政府横断的に有用なデータ(以下、政府共通データ)や各省庁・各部局の特定業務で利用するデータ(以下、ドメイン特化データ)を収集・加工し、AI が利用しやすい形式へ整備するとともに、AI 向けデータセット標準の検討を行うものです。

TOPPAN はデジタル庁の本事業を受託し、NTT データ、FPSC の両社と連携して、約 1,800 万ページ相当(推計)の政府共通データ・ドメイン特化データを対象に、AI-Ready 化を実施し、「ガバメント AI 源内」の利活用を支えるデータ基盤の整備に取り組みます。

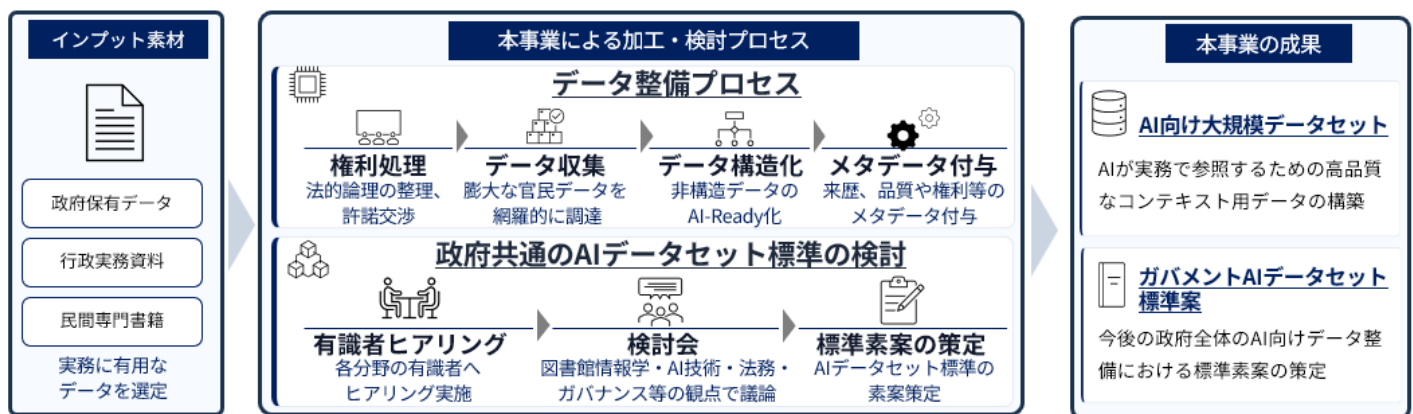


図: 本事業のイメージ

■ 本事業の背景

日本では人口減少や少子高齢化の進展に伴い、行政サービスの維持・高度化に向けた業務効率化が求められています。その中で AI への期待が高まっており、デジタル庁では、政府による AI 活用を社会実装の起点とする「ガバメント AI 源内」の取り組みを推進しています。政府職員が安全かつ適切に AI を活用できる環境として「ガバメント AI 源内」を整備し、2026 年度には約 18 万人の政府職員を対象とした大規模実証が進められています。

「ガバメント AI 源内」の活用を拡大するためには、AI が参照・学習できる高品質なデータセットの整備が不可欠です。近年は AI の回答精度向上において、AI モデルそのものの性能向上に加えて、学習データや検索拡張生成(RAG)で参照するデータの品質が重要な要素になっています。そのため、データの収集だけでなく、メタデータ整備、データ構造の標準化が求められています。

さらに、経済安全保障やデータ主権の観点からも、国内の安全な環境下で、行政文書などの機微な専門知データを高品質な AI-Ready データへ変換し、適切に管理・活用できる基盤の整備が求められています。こうした背景を踏まえ、デジタル庁は「ガバメント AI 源内」のための大規模データセットの整備を推進しており、TOPPAN と NTT データ、FPSC は本事業を通じて、政府共通データの AI-Ready 化と AI 向けデータセット標準の検討に取り組みます。

■ 本事業の概要

本事業は、デジタル庁の仕様書に基づき、政府共通データおよびドメイン特化データの収集・加工、メタデータ整備、AI 向けデータセット標準の検討を実施します。

1) 権利処理

収集データの知的財産権等の利用許諾交渉を実施します。

2) 政府共通データ・ドメイン特化データの収集・加工

政府共通データ(閣議決定、国会会議録、白書等の政府横断的に有用なデータ)およびドメイン特化データ(各省庁・各部局の特定業務で利用する書籍等のデータ)を収集します。さらに、OCR やクリーニングなどの前処理を実施し、AI による検索・学習・推論に適した高品質なデータに変換します。

3) 収集したデータセットのメタデータの設計・作成・付与

本事業で収集したデータセットに対し、来歴情報、品質情報、権利情報、利用条件等のメタデータを設計・作成・付与します。これにより、AI および利用者がデータの信頼性や利用可能性を適切に判断できる環境を整備します。

4) 『「ガバメント AI 源内」データセット標準』の検討

有識者ヒアリングおよび検討会の運営を通じて、AI 向けデータセットの構造化、メタデータ仕様の在り方などを検討し、行政機関横断で活用可能な政府共通の AI 向けデータセット標準の素案策定を行います。

【各社の役割】

・TOPPAN

公共・金融分野を中心に培った BPO 事業の業務分析力とデータ加工・構造化技術を活かし、行政実務のユースケースに基づくデータ選定・収集、権利処理、OCR 処理、データ前処理を担当します。

また、行政文書を AI-Ready 化し、検索・推論に適したデータセットを整備します。

・NTT データ

行政機関向けデジタルアーカイブや生成 AI 導入支援で培った知見を活かし、メタデータの設計・作成・付与や、政府共通の AI 向けデータセット標準の検討を担当します。データの来歴や利用条件等のメタデータ整備を通じて、信頼性の高いデータ基盤の構築を支援します。

・FPSC

中央官庁における行政官としての法令事務経験・各省との協議経験を活かし、AI 向けデータ活用における法的論点の整理や権利処理スキームの構築支援を担当します。著作権をはじめとする知的財産権に関する検討や権利処理実務を支援し、適正かつ円滑なデータ利活用の実現に貢献します。

■ 今後の展開

3 社は、政府職員が利用する「ガバメント AI 源内」における政府共通データ基盤の整備を 2027 年 3 月までに実施します。

また、有識者ヒアリングや検討会を通じて、AI 向けデータセットの構造化やメタデータ仕様等に関する検討を進めるとともに、デジタル庁が進める「ガバメント AI 源内」の利活用を支えるデータ基盤の整備に貢献します。

3 社は今後も、データ収集・加工、AI 向けに最適なメタデータ整備、データ利活用に関する知見を活用し、政府・自治体・企業の AI 活用の推進を支援するとともに、デジタル社会の実現に貢献します。

※ 『「ガバメント AI 源内」：デジタル庁が内製開発で構築した生成 AI 利用環境』

* 本ニュースリリースに記載された商品・サービス名、会社名、団体名は各社の商標または登録商標です。

* 本ニュースリリースに記載された内容は発表日現在のものです。その後予告なしに変更されることがあります。

以 上