事例No.	3109
公表年度	R6
団体の属性	一般市
団体名	茨城県つくば市

事例区分	DX

タグ	・（行政改革）業務の効率化 ・クラウド

事例種類	AI-OCR

事例内容・タイトル

AI-OCRによる行政文書の読取・データ化

出典

自治体におけるAIの利用に関するワーキンググループ（第2回）③

本格導入 17

AI-OCRによる行政文書の読取・データ化（茨城県つくば市）

導入分野：情報化・ICT

AIの機能：文字認識

クラウド

非公開データ

■事業概要：

- つくば市は、2018年12月～2019年３月、町田市、横浜市、福岡市、郡山市、市川市とともに、AI-OCRの正読率検証の共同研究を行った。2019年12月からLGWAN-ASPのAI-OCRツールを本格導入した。
- 読み取りの精度は95%程度。読み取れない画像については、その画像のみがピックアップされて画面に表示される。それを人が見てテキストを入力すると、CSVファイルに反映されるようになっている。

■効果：

- 2018年度の共同研究において、つくば市の帳票（20帳票）の正読率は文字単位で93.41%、項目単位で89.11%。行政分野で使われる申請書等の文字判読について、約９割の読取効果を確認した。

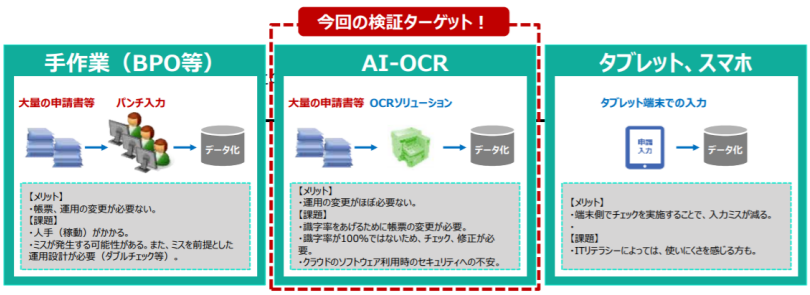
■取り扱う主なデータと共有・提供の範囲：

- 2018年度の共同研究においては、取り扱い予定にかかわらず、帳票の特徴にバリエーションを持たせた。
  - 市長へのたより・要望書（ご意見）　、資格審査処理表【物品納入・役務の提供・印刷請負等】、法人市民税領収済通知書、給与支払報告書（個人別明細書）、出生届、死亡届、つくば市体育施設使用許可申請書、図書館カード申込書
- 実運用において取り扱う帳票は、幼児保育無償化にかかる書類や償却資産申告書などとし、次年度に向けて拡大予定である。
- 本格導入にあたり、AI-OCRには基幹系からのみアクセスできるようにした。LGWAN-ASPへの接続については、LGWAN端末データと基幹系データを分離するために、特定通信を使用している。
- データの処理はLGWAN-ASP内で行われる。民間のデータを学習データとして構築されたAIがLGWAN-ASP上に展開されている。LGWAN-ASPの中のAIは追加で学習しないが充分な読取効果を得ている。

■今後の更なる展開に向けた課題・示唆等：

- 2019年の時点では、パブリッククラウド上でAI-OCRサービスを展開する事業者のサービスとも比較したが、AI-OCRで読み込むデータには個人情報が含まれる可能性があるため、パブリッククラウド上でデータの処理を行うサービスは適当ではないと判断した。

AI-OCRを導入した業務



出所）つくば市プレス資料（つくば市政策イノベーション部情報政策課、2019年３月）

スケジュール

- 2018年12月～2019年3月：事業者と共同研究実施。
- 2019年4月～2019年8月：導入対象とする業務の検討、庁内調整。
- 2019年12月～：資産税課の償却資産申告書・幼児保育課の無償化関係資料等数課で順次導入。
- 2020年12月時点：　税・福祉・窓口を中心に導入に向けた準備を進め、24課まで利用が広がっている。

事例のポイント

- 特定個人情報保護評価書の記載を要する事務において、個人番号欄を含む書類をマスキングせずに読み取る場合は、評価書に利用する旨の記載をしなければならないことに留意が必要。また、クラウド型のAI-OCRを利用する場合は、サービス自体も安全管理措置等の要件を満たしていることも必要となる。
- オンプレで構築しようとすると億単位のコストがかかるとのことで、LGWAN-ASPのサービスを使用することとなった。導入したLGWAN-ASPは年間定額だが、基本使用量を超えると、項目単位で従量課金される。