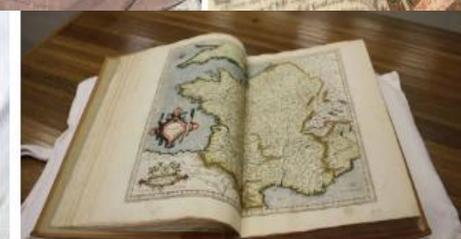


KBR  Koester de tijd
Protégeons le temps

KBR

- Bibliothèque royale de Belgique
- Bibliothèque nationale
- Musée (manuscrits médiévaux)
- 5 millions de livres
- 3 millions d'objets patrimoniaux (monnaies, manuscrits, lettres, dessins, partitions, reserve précieuse, etc)



Rétro-catalogage

- **Idée :**
 - Capture de la page de titre
 - Utilisation de l'IA pour la reconnaissance automatique des champs bibliographiques clés (auteur, titre, éditeur, année, lieu, etc.)
- **Aller au-delà des règles 'codées en dur'**
 - Alors que d'autres projets s'appuient sur un ensemble de règles rigides, notre approche visait à utiliser la puissance de l'IA pour obtenir des résultats supérieurs et plus rapides.
 - Il n'y a pas de codeurs dans notre équipe, mais le paysage actuel de l'IA offre une gamme d'outils conviviaux. Aucune expertise en matière de codage n'est requise.



Power Apps and Power Automate



The low code platform that spans Microsoft 365, Azure, Dynamics 365, and standalone apps.



Power BI
Business analytics



Power Apps
App development



Power Automate
Process automation



Power Pages
External-facing websites



Data connectors



AI Builder



Dataverse

Une plateforme de programmation allégée développée par Microsoft, permettant aux utilisateurs de créer des applications professionnelles personnalisées

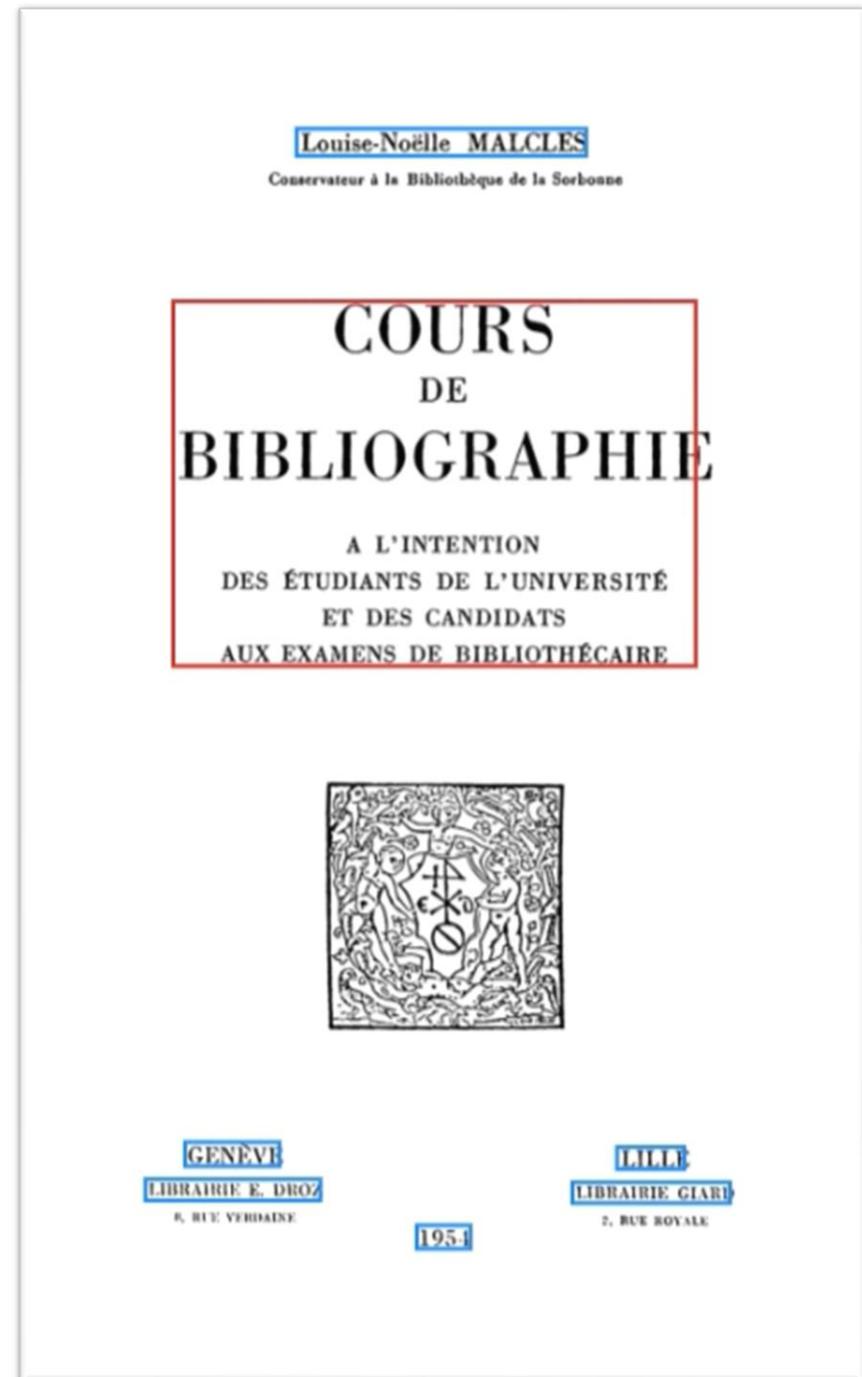
- Création d'application avec une interface visuelle, sans connaissance approfondies de codage
- Connection à l'ensemble des services Microsoft, y compris:
 - SharePoint
 - Microsoft 365
 - SQL Server
 - ...
- Ajout de capacités IA pour améliorer les processus
- Créé pour les développeurs professionnels et citoyens (le cas de KBR) pour une création et un déploiement rapide.

Modèles IA Power Platform (préconstruit + personnalisé)

 <p>Invoice processing</p> <p>Extract information from invoices</p> <p>PREBUILT MODEL</p>	 <p>Text recognition</p> <p>Extract all the text in photos and PDF documents (OCR)</p> <p>PREBUILT MODEL</p>	 <p>Receipt processing</p> <p>Extract information from receipts</p> <p>PREBUILT MODEL</p>	 <p>Identity document reader</p> <p>Extract information from identity documents</p> <p>PREBUILT MODEL</p>	 <p>Business card reader</p> <p>Extract information from business cards</p> <p>PREBUILT MODEL</p>	 <p>Document processing</p> <p>Extract custom information from documents</p> <p>CUSTOM MODEL</p>
 <p>Sentiment analysis</p> <p>Detect positive, negative, or neutral sentiment in text data</p> <p>PREBUILT MODEL</p>	 <p>Category classification</p> <p>Classify customer feedback into predefined categories</p> <p>PREBUILT MODEL PREVIEW</p>	 <p>Entity extraction</p> <p>Extract key elements from text, and classifies them into predefined categories</p> <p>PREBUILT MODEL</p>	 <p>Key phrase extraction</p> <p>Extract most relevant words and phrases from text</p> <p>PREBUILT MODEL</p>	 <p>Language detection</p> <p>Detect the predominant language of a text document</p> <p>PREBUILT MODEL</p>	 <p>Text translation</p> <p>Detect and translate more than 90 supported languages</p> <p>PREBUILT MODEL</p>
 <p>Category classification</p> <p>Classify texts into custom categories</p> <p>CUSTOM MODEL</p>	 <p>Entity extraction</p> <p>Extract custom entities from your text</p> <p>CUSTOM MODEL</p>	 <p>Prediction</p> <p>Predict future outcomes from historical data</p> <p>CUSTOM MODEL</p>	 <p>Object detection</p> <p>Detect custom objects in images</p> <p>CUSTOM MODEL</p>	 <p>Image description</p> <p>Generate description of an image</p> <p>PREBUILT MODEL PREVIEW</p>	

Modèle personnalisé 'Document processing'

- Basé sur des images
 - Détection de métadonnées prédéfinies sur (les images de) documents
 - Un ensemble minimal de 5 exemples peut suffire
 - L'IA crée des règles en utilisant diverses variables, telles que :
 - Structure générale de la page de titre
 - Placement du texte
 - Taille de la police
 - Contenu du texte (par exemple : mention d'édition)





October more de wind !.... hoe wild toch woedt
Uz. U Doet.
: Woude Schroomen., Vedle

Book Status: Review needed - Call Number: III 65.695 A/13 - Stacks location: M-SLZ
Modified on: 10/10/2023 12:12 PM - Modified by: PrjRetroPP SA



October...
Monseigneur G. Waffelaert,
Couckelaere. li. Defoort. :,

Book Status: Review needed - Call Number: III 65.695 A/14 - Stacks location: M-SLZ
Modified on: 10/10/2023 12:12 PM - Modified by: PrjRetroPP SA



Twaalfde jaar. n° 1 oktober 1904. de mandelbie orgaan der west-vlaamsche bientelers. onder de bescherming van zijne doorluchtige hoogwaardigheid,
En Het. Eerevoorzitterschap Van Den Heer Baron Leon Peers.
October. :, Mitunten

Book Status: Review needed - Call Number: III 65.695 A/12 - Stacks location: M-SLZ
Modified on: 10/10/2023 12:12 PM - Modified by: PrjRetroPP SA

Workflow Application de Rétro-catalographie

Scan de la page de titre : Etudiants (2023) et workshop (2024)

SharePoint : Téléchargement du fichier (scan) dans SharePoint

Power Automate : Détecter les nouveaux fichiers → Power Automate débute le process

Modèle IA : Détection via IA du titre, auteur, éditeur...

Dataverse : Informations sauvegardées dans le dataverse

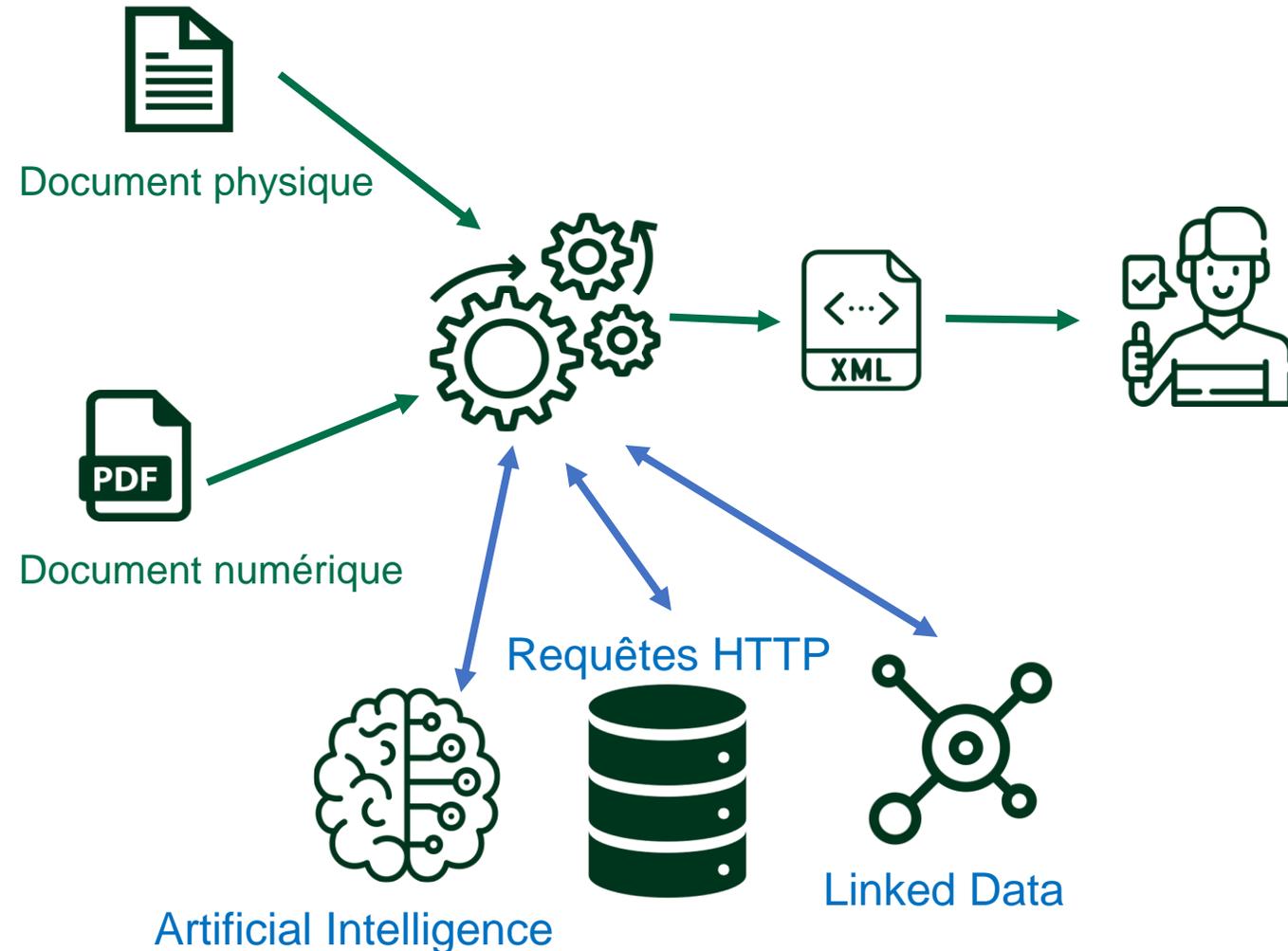
Power Apps : Métadonnées du dataverse visualisées dans l'application

: Validation et correction humaine manuelle (optionnel)

: Export CSV

Dernière étape : Import CSV dans notre SIGB

Power automate



- **Automatisation complète des processus, y compris :**

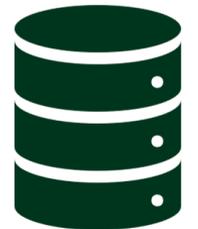
- OneDrive/SharePoint (et leurs applications mobiles avec scan de documents inclus)
- Intégration de l'IA avec AI-builder (modèles d'IA pré-entraînés ou entraînés sur mesure)
- Connectivité avec les bases de données
- Intégration avec FTP ou API pour l'interaction avec le SIGB

Requêtes HTTP

Utiliser les métadonnées détectées immédiatement dans la requête HTTP pour importer plus d'informations sur la notice, l'auteur ou l'œuvre.

- ISBN → requêtes HTTP dans des bases de données externes (il est aussi possible d'utiliser SRU ou API)
- ISNI → ISNI-SRU
- VIAF → requête HTTP dans le VIAF (à l'avenir, il sera sans doute possible d'utiliser WorldCat-entities, Share-VDE)

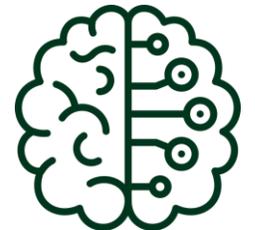
HTTP-queries



Indexation matière automatisée

- **L'application IA la plus utilisée dans le monde des bibliothèques**
 - Modèle personnalisé
 - Création d'un vocabulaire prédéfini, complété par des résumés de textes sur chaque sujet grâce à un fichier à deux colonnes (Excel basique/csv/Dataverse)
 1. Les termes de vocabulaire
 2. Les résumés
 - Alternative : grands modèles de langage comme ChatGPT
 - Contrôle limité des termes/vocabulaire utilisés
 - Utilise des données de formation très vastes = plus grande précision

Artificial Intelligence



Dernière étape

- **Générer des résultats avec les métadonnées collectées**

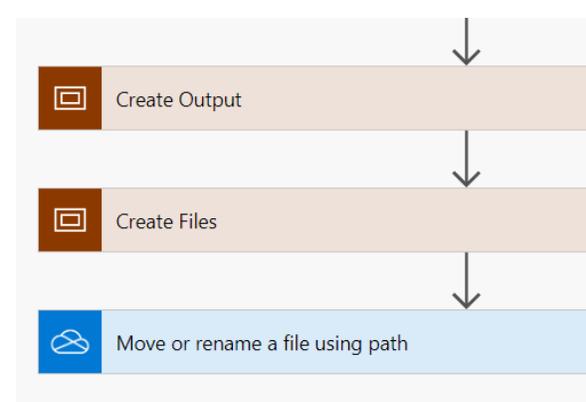
- Métadonnées extraites
- Métadonnées importées à partir des requêtes HTTP
- Sujets ou entités de la 4ème de couverture

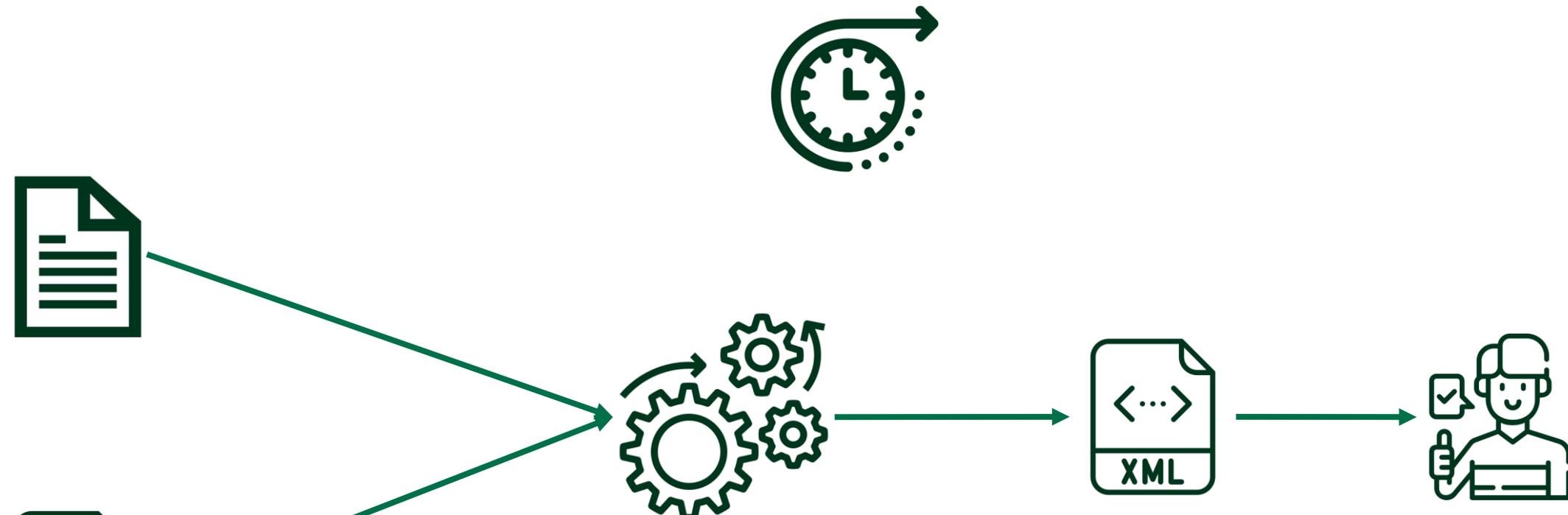
- **Options de production**

- Créer un fichier MARC 21 XML, BIBFRAME ...
- Envoyer le fichier au SIGB (via FTP ou API)
- Collecte d'informations dans Excel
- Renvoyer les informations à Power App pour correction, validation

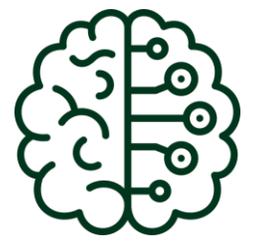
- **Workflow**

- Choix déterminé par les flux de travail internes et la structure organisationnelle
- Une solution unique et définitive n'est pas possible

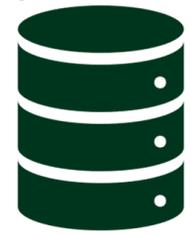




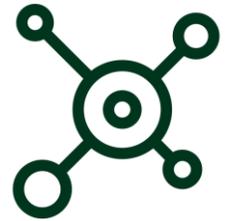
Artificial Intelligence



Requêtes HTTP



Linked Data



- Rétro-catalogage
- Livres contemporains (Dépôt Légal)
- Journaux et périodiques
- Documents patrimoniaux
 - Cartes
 - Lettres manuscrites
 - ...

Impact

- **Identification rapide, y compris dans différentes écritures (arabe, cyrillique, etc.)**
- **Les tâches du catalogueur passeront de la simple "saisie" à**
 - La gestion des sets de données liées et des vocabulaires contrôlés - essentiels pour l'IA et les procédures automatiques afin de garantir une identification précise.
 - Le contrôle des résultats des modèles d'IA, l'apport de corrections et de modifications et la participation aux processus de reconversion.
 - **En substance, l'accent sera mis sur des tâches de plus haut niveau, sur la supervision des processus et des bases de données, plutôt que sur le traitement de livres individuels.**

Questions ?

- Hannes Lowagie: hannes.lowagie@kbr.be
- Marie-Sophie Bercegeay : marie-sophie.bercegeay@kbr.be

