Healthcare and Lifesciences

Automated Document Processing & Data Extraction

Introduction

Business Need

- Many patients/physicians submit paper or PDF format forms in a large volume, creating issues for healthcare companies
- Document intake and the associated data entry into the current system is creating a backlog of paper documents and reducing the effectiveness of services and leading to customer service and data quality issues and errors
- If there are issues with the form submission, communication with the submitter can be difficult and create delayed turnaround times
- There is currently a lack of reporting and monitoring capabilities for submitted forms with little to no ability for ad-hoc reports

Targeted Outcomes

- Automate data extraction from forms to eliminate manual intake and data entry of paper/PDF applications and documents such as:
  - Clinical Trials
  - Physician/Clinical Notes
  - Patient Applications & Surveys
- Reduce environmental impact/carbon footprint by eliminating large paper consumption
- Reduce staff effort for document ingest and data entry for the IDs, Insurance Cards, Invoices, Order Forms, and more
  - Improve processing time for employees by over 30%
  - Eliminate document backlogs
  - One customer experienced over 1,000 employee hours saved in one month (equivalent to 13 full time employees for 2 weeks)
- Modernize communication between form submitters and reviewers to speed up verification and update workflows
- Gain insight into product and processing dynamics leveraging analytics, dashboards, and reports

Brief Solution Overview

The suggested solution offering will automate the intake, data extraction, and quality assurance processes for two form types. Documents can be scanned in bulk to a cloud storage system and then automatically processes where the data will be extracted into a cloud database. A web form provides document ingest and another form provides data verification and quality control. All services are cloud based with no need for costly hardware or maintenance.

Advantages of Hosting in Google Cloud

- *No per user license fees* - ongoing cost is fixed and is much lower than other Cloud vendors such as Salesforce, Amazon, and Microsoft
- Industry best document processing and AI tools
- Highly secured and safe - HITRUST Compliant
- Easy to add new document types or workflows over time, keeping costs low

Current Users of Document AI Systems

- ADD healthcare quals

Google Document AI Overview

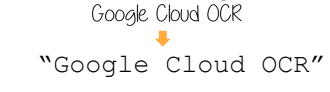
Core Document Processing Functions for Any Document Type

Print OCR

Core text extraction with Industry leading accuracy

>200 supported languages

Handwriting OCR



Handwriting and signature recognition

>50 supported languages

Form Extraction

Extract data from documents

Includes complex extraction:

[Form fields]
[Checkboxes]
[Signatures]

Image Tools

Clean up tools for imaged documents and more

Splitting
Alignment
Artifact removal
Contrast enhancement

Document AI Helps Your Team Operate More Efficiently

- Increase worker productivity while streamlining document intake and data entry
  - Reduce processing time by >30%
- Automate validation of documents/data to create more efficient compliance workflows, reduce guesswork, and keep data accurate
- Classify and compile documents and data extraction results into reports to more quickly help the people you serve
- *An example of how Document AI can create a data table from a clinical intake document is shown to the right:*

5 tables found: 1 2 3 4 5

Table 2

Information about the trial

Project identification UiO	444325
Project name/description	Ruflexin
Number of trial subjects	150
Planned inclusion date for first trial subje...	1/23/23
Planned end-date for last trial subject	1/23/26
Planned date for close-out of trial site	3/23/26

Typical Document Processing Workflow

Upload Inputs
Ingest documents for multiple sources

- **Docs**
  - PDF
  - DocX
  - HTML
  - TXT
  - JPG
  - TIFF

Document Prep
Clean images & get text

- **Preprocessing Tools**
  Splitting
  Quality Scoring
  Alignment
- **Digitization**
  Print OCR
  Handwriting OCR
  Language identification
  Specialized Language Models

Classification
Identify document type

- **General Docs**
  Unclassified
- **Specialized Docs**
  Google's Pre-trained Docs
- **Custom Doc Classifier**
  Train your own classifier

Data Extraction
Get simple & compound data

- **Specialized Models**
  Schematization
- **Text Entities**
  Strings
  Numbers
  Dates
  Addresses
  Phone #'s
- **Spatial Entities**
  Tables
  Signatures
  Checkboxes
  Bar Codes
  Stamps
- **Customized Entity Extraction**

Data Entity Relationships
Identify & link data entities

- **General E-R Model**
- **Vertical E-R Models**
  - Medical
  - Legal
  - Banking...
- **AutoML E-R**

Repository
The System of Record for unstructured data

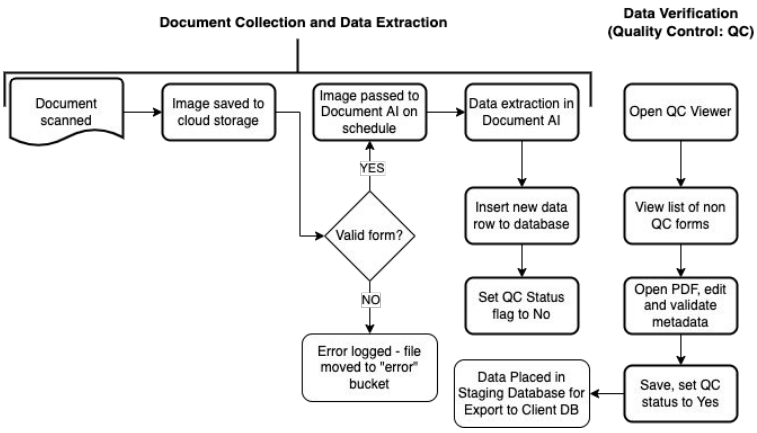
- **CMS**
- **Database**
- **Knowledge Graph**

Try Document AI Now!

Suggested Solution Architecture and Project Structure

- The Project will focus on the following main activities:
  - Create Document AI data extraction models
  - Build staging database for extracted data and document metadata management
  - Integrate with Microsoft SQL for data handoff
  - Configure web portal for document search and verification including a Document Inspector
  - Build suggestion engine for rapid review
- Possible Future Solution Expansions
  - Analytics & Dashboarding
    - Monitoring
  - Add Other Document Types and Workflows
    - Invoices
    - Return Forms
    - Compliance Documents
  - Document Translation Features

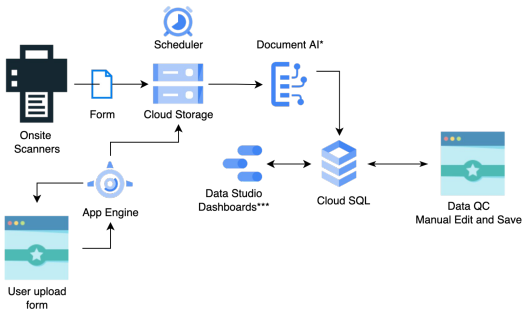
Proposed Workflow Chart



Task/Activities	Discovery and Definition				
	Week 1	Week 2-3	Week 3-6	Week 6-7	Final Week
Discovery					
Form analysis					
Database definition					
Dataflow definition					
Build					
Custom form parser					
Stand up cloud environment					
Configure Cloud SQL Database					
Data import from DocAI to Database					
Deploy web page form upload, document management, and data verification					
Automated scan to ingest scheduler					
Testing					
Dashboard creation and sign off					
User Acceptance Testing					
Support and Training					
End user training					
Handoff					

Suggested Solution Architecture

- Two custom Document AI processors
- Document ingest page for manual upload
- Document Inspector (data verification) Page
- Two operational dashboards for monitoring & reporting
- *A simple diagram of the suggested architecture is below:*



Suggested Project Timeline / Resources

6 - 8 weeks:

- Based on reusable Solution accelerators
- Main Solution components are already built
- Project Resources
  - Engagement/Project Manager
  - Solution Architect
  - Technical Architect
  - Core Developers