



DOXCI

Document Cognitive Intelligence



Table of Contents

03 Executive Summary

04 The Problem We Solve

08 The Doxci Approach

12 Functional Architecture

16 Capabilities & Performance



Executive Summary

Organizations today face a critical challenge: turning the valuable information locked in their documents into usable data for their business systems. Traditional document processing solutions are complex, requiring extensive training data and creating bottlenecks between extraction and actual use of the data.

Doxci solves this challenge by offering a zero-training platform that extracts data from any document and delivers it directly to business systems. It transforms complex documents into structured, system-ready data, eliminating the traditional hurdles and long setup timelines of traditional document processing.

Key Capabilities

Instant Document Processing: Extract data from 27+ filetypes and 52+ languages - no training data needed

Direct System Integration: Get extracted data instantly into your business systems through flexible APIs and direct database connections

Simple Configuration: Define what you need in plain language using our scenario system

Universal Format Support: Process everything from handwritten forms to complex digital documents

Automated Data Pipelines: Create reliable flows from Emails, FTPs, and Online Drives directly to your operational systems at enterprise scale.

Impact

- Cut document processing time from days to minutes
- Eliminate manual data entry and validation
- Get data flowing directly into your business systems
- Scale document processing without adding complexity
- Process any document type through a single platform

Doxci transforms how organizations handle documents by making data extraction and integration simple, fast, and reliable. It bridges the gap between your documents and your business systems, turning document processing from a bottleneck into a business advantage.



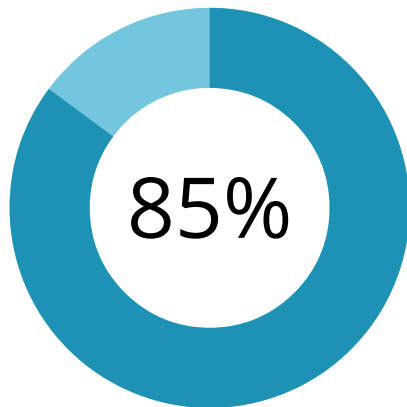
Part 1:

The Problem We Solve

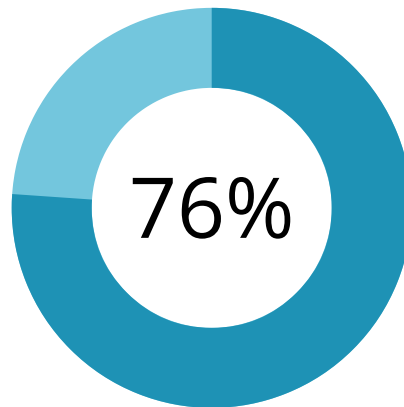


The Problem We Solve

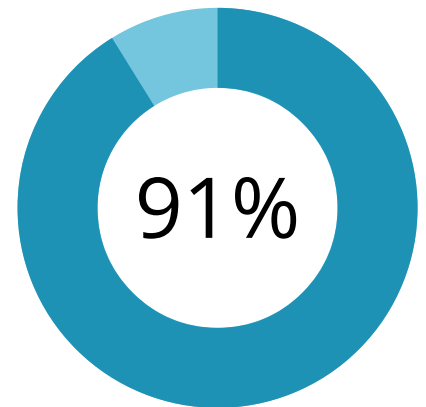
Bridging the Document-to-Data Gap



Of Vital Enterprise
Data is Locked in
Unstructured
Documents. [1]



Of Knowledge Workers
Spend **1-3 Hours a day**
Moving Data between
Systems. [2]



Of Surveyed
Companies Reported
Growth from Process
Automation. [3]

1. <https://researchworld.com/articles/possibilities-and-limitations-of-unstructured-data>

2. <https://zapier.com/blog/report-how-office-workers-spend-time/>

3. <https://camunda.com/blog/2024/01/state-of-process-orchestration-report-2024/>

Every organization sits on a goldmine of valuable information locked away in documents. Purchase orders, invoices, contracts, forms, and reports contain critical data that could drive automation, improve decision-making, and streamline operations. But there's a fundamental challenge: transforming this unstructured document information into structured, system-ready data.

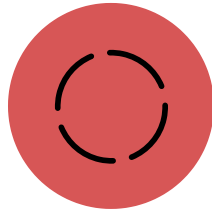
The Problem We Solve

The Hidden Costs of Trapped Data



Process Delays

Days or weeks spent manually transferring data from documents to systems



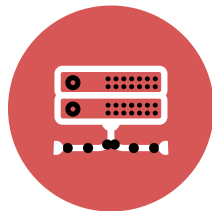
Lost Opportunities

Valuable data going unused because it's too time-consuming to extract



Resource Drain

Skilled employees spending hours on manual data entry instead of strategic work



Integration Complexity

Technical teams building and maintaining complex document processing pipelines



Scaling Limitations

Document processing becoming a bottleneck as business volumes grow



The Problem We Solve

Why Traditional Solutions Fall Short

Conventional approaches to document processing create as many problems as they solve:

- Require extensive training data and constant maintenance
- Support limited document types and languages
- Create technical debt through complex integration requirements
- Need specialized expertise to configure and maintain
- Struggle with variations in document layout and format

Moreover, companies often lack the necessary skill sets and technology understanding to address these challenges effectively, leaving them reliant on outdated methods or external vendors for suboptimal solutions.

The Integration Challenge

Even when data can be extracted, organizations face another hurdle: getting that data into their operational systems. Most solutions:

- Output data in formats that need additional processing
- Require custom coding for each new document type
- Create bottlenecks between extraction and system updates
- Need manual validation and data transformation
- Lack standardization across document sources



Part 2:

The Doxci Approach



The Doxci Approach

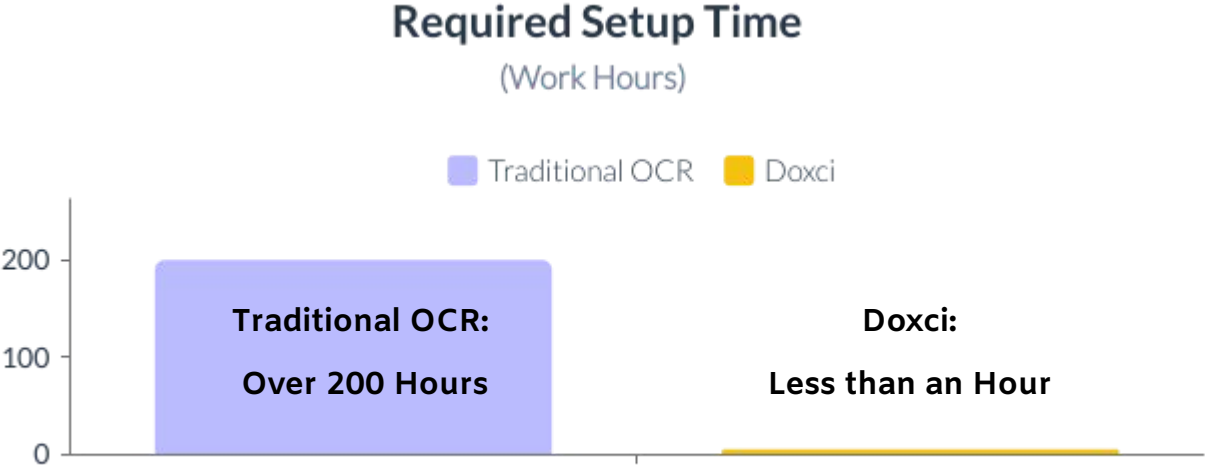
Doxci revolutionizes document processing by eliminating the traditional barriers between documents and business systems. Through our innovative scenario-based architecture, we've created a solution that makes document processing accessible, reliable, and instantly valuable for any organization.



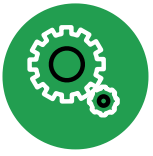
No Training Required

Unlike traditional solutions that require extensive sample documents and training time, Doxci works immediately with any document type. Our advanced AI understands document context from and structure from descriptions in plain language, enabling:

- Immediate processing of new document types
- Handling of variations in layout and format
- Support for multiple languages and document structures
- Consistent accuracy across document sources



The Doxci Approach



Scenario-Based Configuration

Doxci replaces complex technical setups with an intuitive scenario system. Rather than writing code or building complex rules, users simply define what information they need in business terms:

- Configure extraction targets using plain language
- Define business rules and validation criteria
- Set up data formatting and transformation rules
- Create standardized processing workflows

DOXCI

SCENARIO

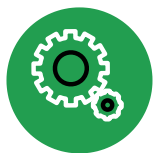
Details Targets API Config History

Costs

Type

Target Name	Type		
Description	The primary functional category that classifies the nature of the project or expe		
Choices	Power generation, Construction, Design-build, Software, ERP, Cybersecurity, Tc		
Type	Choice	Group	Expenditure
Selection	Required	Notes	Always select a value for th
Default	Default Value	Optional	No

The Doxci Approach

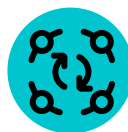


Direct System Integration

Instead of creating another data silo, Doxci connects directly with your business systems. Our platform:



Delivers extracted data in your system's native format



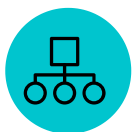
Supports real-time integration through APIs



Enables direct database connections



Maintains data consistency across systems



Automates end-to-end document processing workflows

Part 3:

Functional Architecture



Functional Architecture

Intelligent Document Processing Core

At the heart of Doxci is our zero-training document processing engine. It combines advanced computer vision, natural language processing, and machine learning to understand and extract information from any document format. The engine:

- Automatically identifies document structure and layout
- Understands contextual relationships between data elements
- Processes both digital and scanned documents
- Handles multiple languages and character sets
- Maintains accuracy across document variations

Scenario System

The scenario system is Doxci's unique approach to configuring document processing workflows. Rather than requiring complex rules or code, scenarios let users define their needs in business terms:

- Target specific data points using natural language descriptions
- Set business rules and validation criteria
- Define data transformations and formatting
- Create standardized processing templates
- Share and reuse scenarios across teams



Functional Architecture

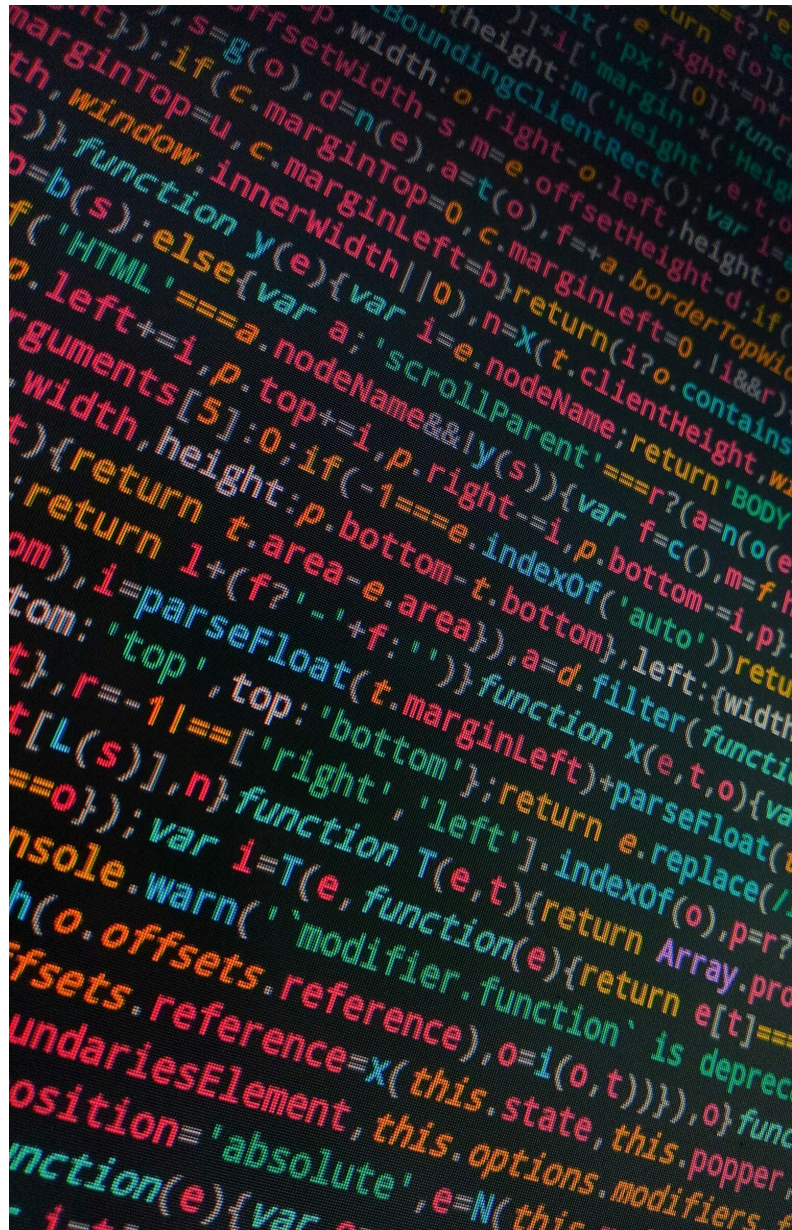
Integration Layer

Doxci's integration layer ensures extracted data flows seamlessly into business systems:

- RESTful APIs for real-time data access
- Direct database connections
- Webhook support for event-driven architectures*
- Custom data mapping and transformation**
- Secure data transmission and storage

*Webhooks automatically send real-time data updates to other systems whenever specific events occur, enabling efficient and event-driven workflows.

**Custom data mapping and transformation: Doxci adapts data to fit the exact structure and format required by your systems.



Functional Architecture

Processing Pipeline

Documents flow through Doxci's automated processing pipeline:

1. Document Ingestion

- Email integration
- FTP/SFTP support
- Cloud storage connections
- Batch upload capabilities

2. Document Analysis

- Format detection
- Structure identification
- Language recognition

3. Data Extraction

- Field identification
- Value extraction
- Validation checks

4. Data Delivery

- System integration
- Format transformation
- Delivery confirmation



Part 4:

Capabilities & Performance



Capabilities & Performance

Document Support

Doxci processes any document type through a unified platform:

- 27+ supported file formats including PDF, Word, Excel, Images
- Handwritten and printed text recognition
- Support for tables, forms, and unstructured documents
- Multi-page document processing
- 52+ languages supported out of the box

Data Extraction

Intelligent extraction capabilities handle complex document scenarios:

- Structured and unstructured data extraction
- Table and grid data capture
- Currency and numeric value processing
- Date and time standardization
- Automatic field type detection
- Contextual data validation



Capabilities & Performance

Business Logic

Built-in business rules engine ensures data accuracy:

- Configurable validation rules
- Data transformation and standardization
- Cross-field validation
- Custom calculations and formulas
- Error handling and exception management

Integration Features

IFlexible integration options connect with any business system:

- REST API access
- Direct database connections
- Webhook support
- Batch processing
- Real-time data delivery
- Secure data transmission



Capabilities & Performance

Enterprise Scale

Enterprise-ready features support large-scale deployments:

- High-volume document processing
- Multi-tenant architecture
- Role-based access control
- Audit logging and tracking
- SLA monitoring and reporting
- Disaster recovery support

Key Processing Metrics - Software

Metric	Capacity	Unit of Measure
Max Rate of Throughput (per instance)	1000	documents/minute
RAM Required per Instance	4	GB
Deployment Capability	Cloud (AWS, Azure, Google), On-Prem	N/A
Scenario Setup Time	20 min + 1 min per field	per scenario
Max Instances at Runtime	Unlimited	N/A

Capabilities & Performance

Key Processing Metrics - Logistics

Metric	Capacity	Unit of Measure
Standard Installation Manhours	10	hours
On-Prem Installation Manhours	480	Hours Per Environment
Time to Go-Live - Standard	2	Weeks
Time to Go-Live - Custom	Varies	Weeks
Minimum Specs for On-Prem Server	Windows Server 2016 - Intel i7 series - 4 cores - 4GB RAM	N/A

Contact Information

Issac Hicks - issac@doxci.ai

Austin Ambrozi - austin@doxci.ai

<https://doxci.ai>

