

smart.CAPTURE

Die skalierbare Input-Management-Lösung

smart.CAPTURE automatisiert schnell, sicher und komfortabel die Erfassung und Indizierung ihrer digitalen Dokumente zur Weiterleitung an E-Mail-Empfänger sowie DMS-/Archiv-, ERP-, Fibu- und Workflowsysteme.



smart.CAPTURE

Mit dem System **smart.CAPTURE** bietet die DIGIDOK GmbH eine hochleistungsfähige Standard-Software zur rationellen Erfassung gescannter Papierdokumente und elektronisch empfangener PDF-Dateien an. **smart.CAPTURE** automatisiert die Heftung, das Auslesen von Textinhalten zur Indizierung der Dokumente und sorgt für die zuverlässige Weiterleitung in elektronischer Form an E-Mail-Empfänger, Workflowsystem-Nutzer sowie an DMS-/Archiv- und ERP-/FiBu-Systeme. Austauschbare Erkennungsverfahren für Barcode, OCR, ICR oder OMR von technologisch führenden Herstellern kombiniert mit eigenentwickelten Klassifikations- und Freiform-Extraktions-Modulen garantieren optimale Erfassungsqualität und hohe Investitionssicherheit. **smart.CAPTURE** basiert auf einem Hardwareneutralen und beliebig skalierbaren Java-Server und lässt sich problemlos auf allen gängigen Betriebssystemen wie u.a. MS Windows, UNIX, LINUX oder Mac OS X ausführen. Die Software gewährleistet durch den Einsatz von HTTP- und XML-Kommunikation größtmögliche Flexibilität und unterstützt alle gängigen relationalen Datenbanksysteme. Im Java-Umfeld kann die Funktionalität des CAPTURE-Servers über Java-Klassen erweitert werden; für Windows-Systeme stehen COM-Objekte und Schnittstellen für VB-Scripte zur Verfügung.

Posteingang, Rechnungen, Lieferscheine, Verträge, Kundenakten, eingehende Formulare und handschriftlich ausgefüllte Bestellungen, Zählerstände oder Arbeitsscheine – ein heterogener Dokumentenstapel, der zur wirtschaftlichen Verarbeitung zunächst mit Scannern digitalisiert werden muss. **smart.CAPTURE** sorgt für die lückenlose Erfassung und Indizierung der gescannten Dokumente – wahlweise manuell, halb- oder vollautomatisch.

Das Capture-System übernimmt digitale Dokumente in allen gängigen Dateiformaten aus dem Filesystem. Auch per E-Mail erhaltene PDF-Anlagen oder via Internet transferierte Scan-Images aus dezentral organisierten Anwendersystemen können eingespeist werden. Die Dokumente werden entweder interaktiv mit dem Heftungsclient oder automatisch anhand von Barcodes oder anderen eindeutigen Merkmalen separiert. Bei der manuellen Indizierung gibt der Sachbearbeiter die Indexdaten zu den einzelnen Belegen per Tastatur

in die jeweils zugeordnete Indexmaske ein oder lässt die textuellen Inhalte mittels interaktiver Zonen-OCR-Erkennung gezielt auslesen.

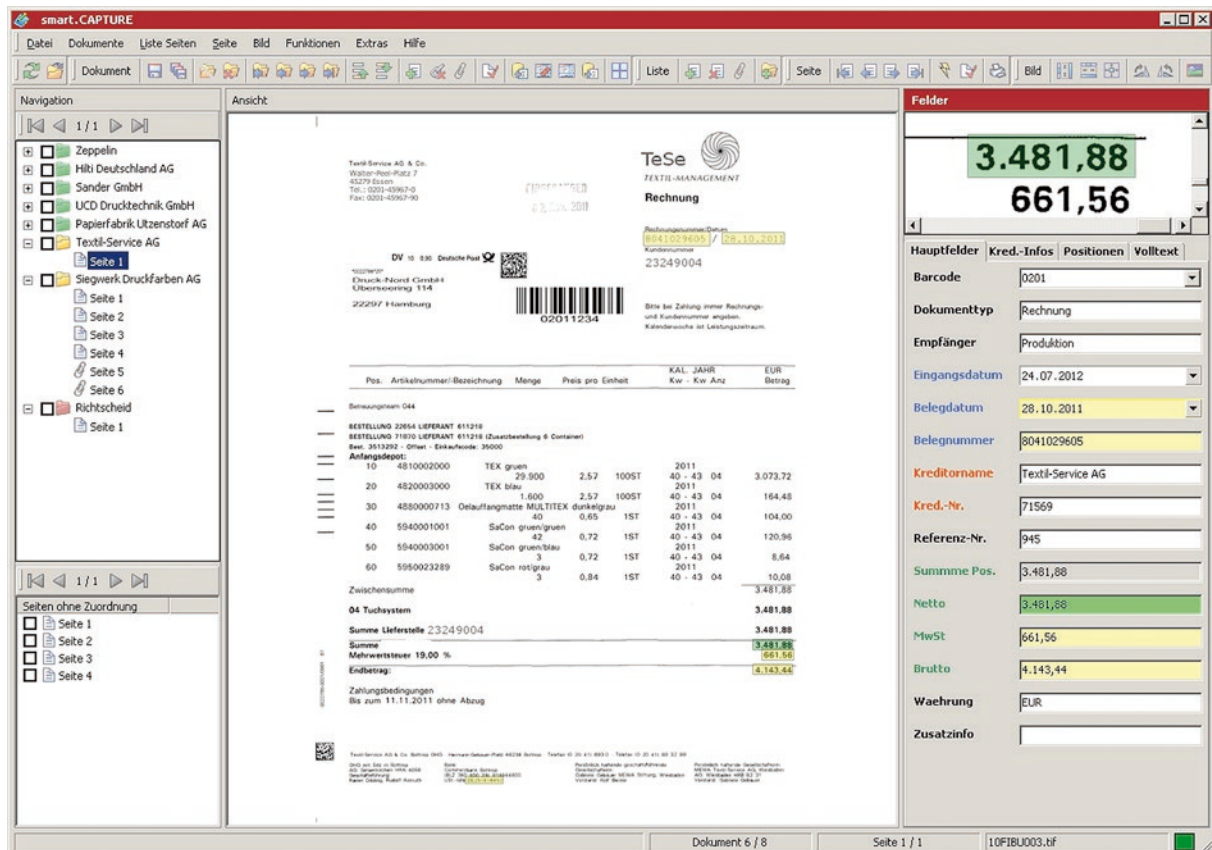
Automatische Indizierung – schnell, sicher und komfortabel

Für hohe Belegaufkommen empfiehlt sich dagegen die automatische Indexgenerierung durch Batchverarbeitung ausgewählter Verzeichnisse mit den digitalisierten Dokumenten. **smart.CAPTURE** liest die Imagedateien in einem selbsttätig ablaufenden Prozess nacheinander ein und extrahiert dabei die dokumentspezifischen Indexinformationen.

Die Kernfunktion von **smart.CAPTURE** ist die automatische Klassifikation der Dokumente und die Extraktion von Indexbegriffen wie z.B. Belegnummer, Betrags- oder Datumsfelder usw. von unterschiedlichen Dokumentpositionen. Dabei können die Daten z.B. bei teilstrukturierten Belegen wie Rechnungen, Gutschriften, Zahlungsavise, Aufträge usw. sowohl aus dem Kopf- und Fußteil als auch aus den Positionszeilen automatisch vom System erkannt und gelesen werden. Diese freiformbasierte Datenextraktion mit ihrer hohen Erkennungsquote wird u.a. durch die mitgelieferte und vom Anwender leicht zu erweiternde Wissensbasis aus Regeln und Schlüsselwörtern möglich. Damit findet der Extraktionsservice z.B. auch in unbekanntem Rechnungen sofort und zuverlässig alle buchungsrelevanten Daten – ohne Templatedefinitionen und Anlernschritte. Individuelle Indexfelder und ihre variablen Belegpositionen sowie alle Felder im Positionsteil von Rechnungen sind mit geringem Aufwand interaktiv zu definieren und die Erkennungsergebnisse können vom Anwender sofort überprüft werden. Feldspezifische formale, inhaltliche und rechnerische Prüfungen garantieren, dass nur Dokumente mit sicher erkannten Indexdaten automatisch weiterverarbeitet werden. Dokumente mit nicht zweifelsfrei entschlüsselten Zeichen oder nicht plausiblen Werten werden zur manuellen Nachbearbeitung mit dem **smart.CAPTURE**-Client zurückgestellt.

Spitzentechnologie inklusive

Im Lieferumfang von **smart.CAPTURE** ist ein integrierter Hochleistungs-OCR-Dienst vom technologischen Marktführer **ABBYY** sowie ein universeller Barcode-Er-



kennungsservice für ein- und zweidimensionale Codes enthalten. Zusätzliche Erkennungsdienste wie z.B. **ICR** für die Erkennung handgeschriebener Schrift, hier vorzugsweise Ziffern (z.B. für Arbeitsscheine usw.) oder **OMR** für das Auslesen von Markierungs-codes (z.B. bei Fragebögen und Formularen) können problemlos eingebunden werden. Mit den automatisch ausgelesenen Indexdaten lassen sich Zugriffe auf externe Datenbanken realisieren, um z.B. weitere Indexdaten aus Kunden- oder Bestelldatenbanken zu selektieren oder um die extrahierten Daten automatisch abzugleichen. Zusätzlich wird der gesamte Textinhalt des Dokuments für spätere Volltextrecherchen extrahiert und in einem eigenen Feld abgelegt.

Übergabe der gelesenen Dokumentdaten zur Weiterverarbeitung

Nach der Indizierung exportiert **smart.CAPTURE** die aufbereiteten und validierten Dokumente, Indexdaten und weitere Attribute – wie z.B. eine digitale Signatur zur Gewährleistung der Revisions-sicherheit – über vorgefertigte Exportservices in ein angeschlossenes Workflow-, ERP-, FiBu- oder ECM-System. Dabei können die gewünschten Zieldateiformate, meist XML und TIFF oder PDF/A, durch integrierte Konvertierungsdienste angepasst werden. Neben Bild- und Indexdaten werden

optional auch Buchungssätze im Importformat der jeweiligen Finanzbuchhaltung (z.B. CSV, XML oder IDoc für SAP und EDIFACT-Dateien) generiert. Optional bietet **smart.CAPTURE** flexible Funktionen zur **elektronischen Postverteilung**. Dabei werden geheftete und indizierte Dokumente, auf Basis der erkannten Indexbegriffe und gesteuert durch individuelle Scripte, in den elektronischen Postkorb des zuständigen Sachbearbeiters übertragen. Unterstützt werden Microsoft Exchange Server, Lotus Notes und alle MAPI- oder SMTP-basierten Mail-systeme.

Einfache Konfiguration

Für jeden Dokumenttyp lassen sich schnell eigene Indexmasken, Erkennungsszenarien oder Exportservices definieren. Die Software kann mit dem **Administrations-client** für vielfältige Indizierungs- und Heftungsaufgaben konfiguriert werden. Das Anlegen, Zuordnen und die Verwaltung der Indexfelder, Masken oder Belegzonen für die Zeichenerkennung geschieht über hierarchisch gegliederte Verzeichnisse. Für jeden Dokumenttyp werden die einzelnen Verarbeitungsschritte wie der benötigte **Importservice** (Filesystem, FTP, externe Anwendung oder Scanapplikation), der **Dokumentenservice** inklusive der präferierten OCR- oder Barcode-Engine oder der abschließende **Output-Service** (Archivimport,

Mailversand, SAP-Import usw.) individuell per Konfiguration festgelegt. Die häufigsten Heftungs- und Indizierungsszenarien sind bereits implementiert und werden vom IT-Administrator bei Bedarf nur noch aktiviert. Diese ausgelieferten Szenarien lassen sich einfach an geänderte Aufgabenstellungen anpassen. Einzurichtende Verarbeitungsszenarien können mit sofortiger Ergebniskontrolle interaktiv am Bildschirm getestet werden.

Hochleistung durch Skalierung und Lastverteilung

Der auf Performanz ausgelegte **CAPTURE-Server** ist im Bedarfsfall **beliebig skalierbar**. Rechenintensive Prozesse, z.B. die OCR-Erkennung können ausgelagert und auf mehrere Prozessoren verteilt werden. Die in den Capture-Server integrierte Lastverteilung sorgt dabei automatisch für die nötige Synchronisation der verteilten Prozesse. Die Speicherintensiven Imagedateien der gescannten Belege werden außerhalb der Datenbank im Filesystem gespeichert.

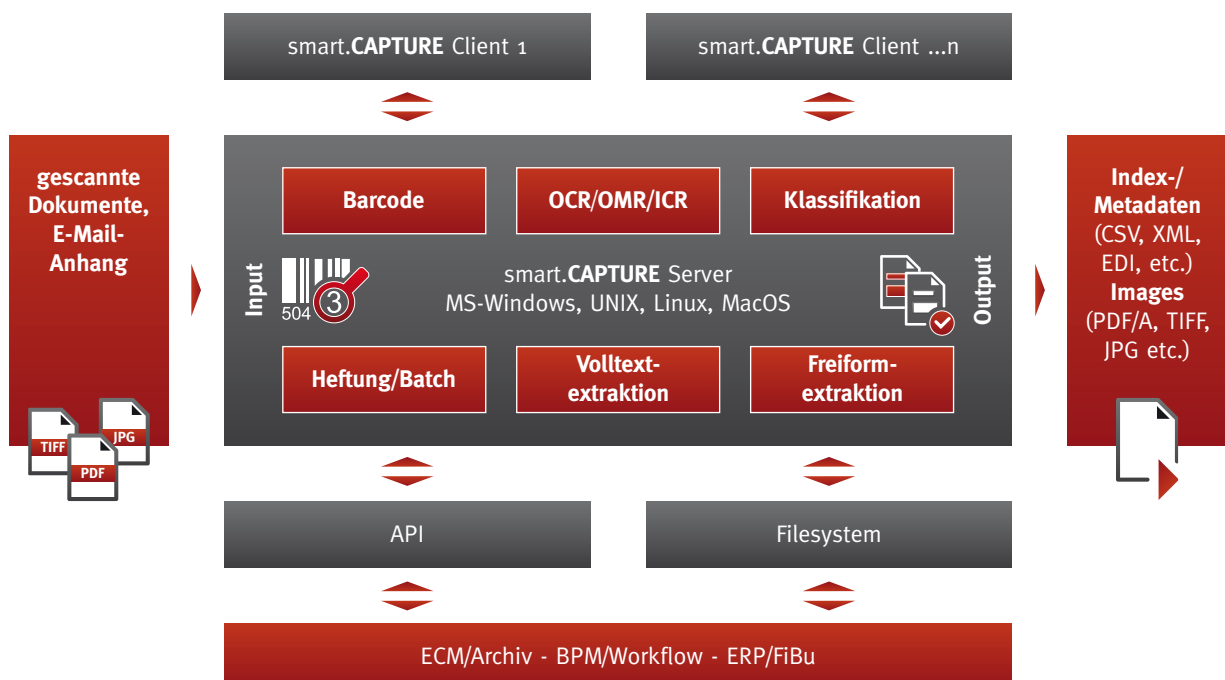
Offene Architektur und austauschbare Erkennungsdienste

Sämtliche Systemkomponenten können auf verschiede-

nen, auch räumlich entfernten Rechnern installiert und via LAN, Intranet oder Internet gesteuert und betrieben werden. Durch einfach zu integrierende Plug-Ins oder Java-Klassen lässt sich **smart.CAPTURE** jederzeit an neue Erfordernisse und Technologien anpassen. Alle wichtigen Elemente wie z.B. die Erkennungsdienste sind leicht auszutauschen und zu ergänzen. Die komplette Konfiguration eines CAPTURE-Servers kann als **XML**-Datei exportiert, editiert und wieder importiert werden.

smart.CAPTURE ist für eine große und **dauerhafte Auslastung** konzipiert und bietet durch seine moderne skalierbare Struktur eine **hohe Investitionssicherheit** für alle Unternehmen und Organisationen, die auf ein digitales und fortschrittliches Dokumentenmanagement setzen.

Das Erfassungssystem lässt sich einfach in bestehende IT-Infrastrukturen integrieren, egal, welche Hardwarekonfiguration oder Softwaresysteme bislang eingesetzt werden. Mit **smart.CAPTURE** profitieren Sie als Anwender von der langjährigen Praxiserfahrung ambitionierter Programmentwickler und Softwareingenieure mit dem Schwerpunkt DMS/ECM. **DIGIDOK smart.CAPTURE** ist daher das optimale Input-Management-System per API integriert oder als vorgelagerte Lösung für alle elektronischen Archiv-, ERP/FiBu- und Workflow-Systeme.



Anwendervorteile

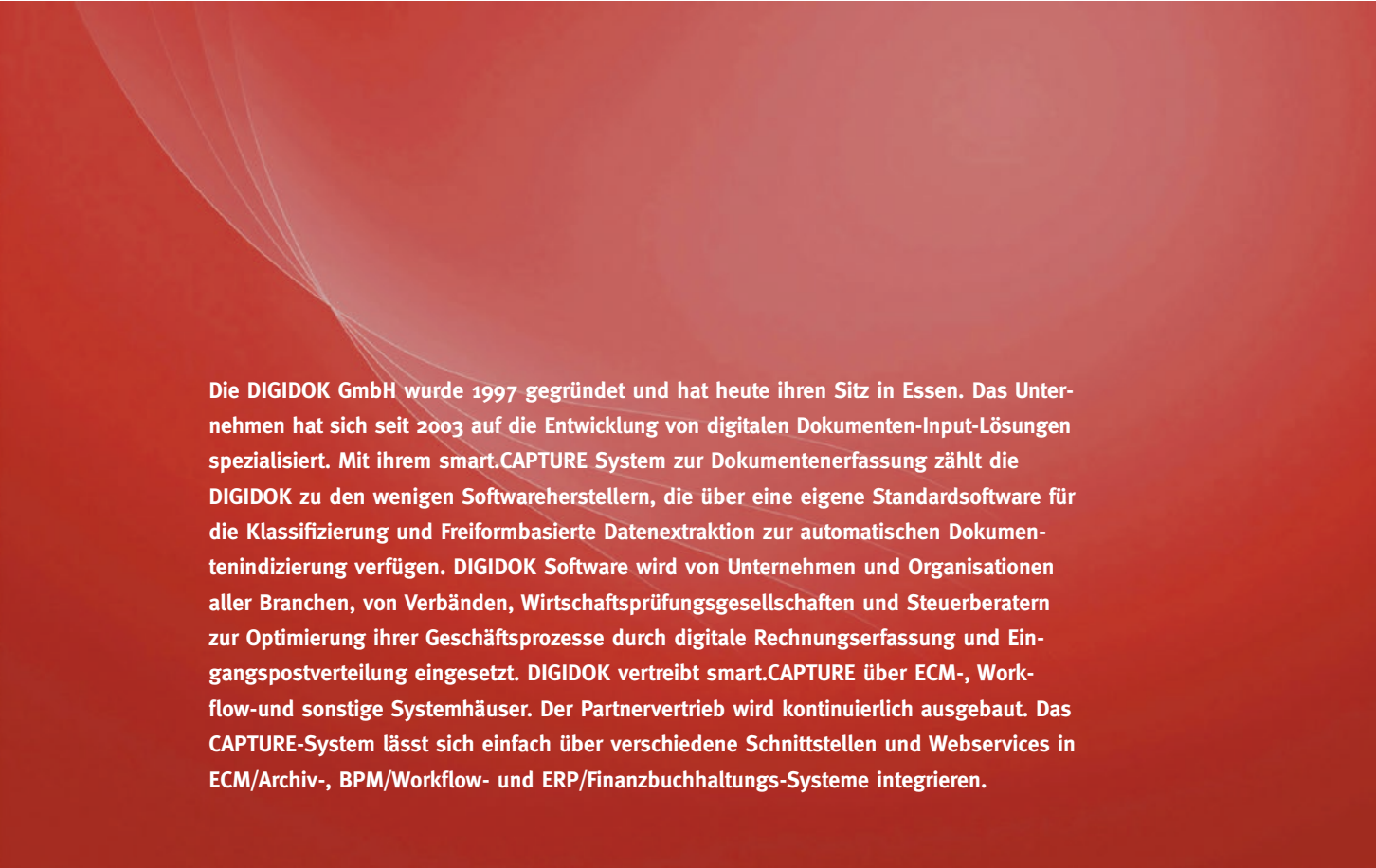
- Nutzung bereits vorhandener File-Server unterschiedlicher Systemplattformen
- Nutzung lizenzfreier oder schon vorhandener Datenbank-Server (PostgreSQL, MySQL, MS-SQL, Oracle, DB/2 u.a.)
- Geringere Kosten für Installation, Administration, Konfigurations- und Schulung
- Leicht integrierbar in vorhandene EMC-/DMS-, Workflow- und ERP-Systeme
- Einfaches Handling für die Anwender
- Vom Anwender leicht zu erweiternde Wissensbasis für die Freiform-Datenextraktion und Klassifikation
- Systemdurchsatz kann bei steigenden Dokumentenzahlen durch flexible Skalierung erhöht werden
- Aktualisierbare Erkennungstechnologien garantieren den Schutz der Investition
- Datenschutz und differenzierte Zugriffsrechte schon ab dem Dokumenteneingang in das Capture-System (und nicht erst nach der Archivierung)
- Beste Erkennungs-, Klassifikations- und Extraktionsergebnisse

Funktionsumfang (Auszug)

- Barcode-Erkennungsdienst (1- und 2-D) ist im Standardlieferungsumfang enthalten. Zusätzlich können neben OCR für Maschinschrift auch andere Erkennungsdienste wie z.B. ICR für Handschrift- und OMR für die Markierungserkennung eingebunden werden
- Standard-Heftungs- und -Indizierungsszenarien werden mit ausgeliefert und brauchen nur noch ausgewählt zu werden. Sie sind leicht an neue Aufgabenstellungen anzupassen
- Die Verarbeitung ist sowohl interaktiv für einzelne Dokumente als auch im Stapelverarbeitungsmodus ausführbar
- Wahlweise auch manuelle Erfassung, bei der die Daten per Tastatur, über Auswahllisten oder mittels interaktiver Zonen-OCR eingegeben werden
- Nur Dokumente mit sicher erkannten und plausiblen Indexdaten werden weiter verarbeitet
- Alle anderen Dokumente werden mit Fehlerhinweisen zur manuellen Nachbearbeitung zurückgestellt
- Verschiedene Output-Services übernehmen den Export der Dokumente, Indexdaten und Attribute. Optionale Umwandlung des Dateiformats z.B. von TIFF oder JPG in PDF/A

Systemeigenschaften

- JAVA basierter Dokumentenserver, lauffähig auf allen Hardware-Plattformen und unter den Betriebssystemen: LINUX, UNIX, MS-Windows oder Mac OS
- Ergonomisch optimierte Windows.NET-Clients für die Nachbearbeitung, Heftung und Administration
- Multi-tier-Architektur garantiert hohe Verarbeitungsgeschwindigkeit und unbegrenzte Skalierbarkeit
- Verarbeitung erfolgt auf dem Server, dadurch geringstmögliche Netz- und Arbeitsplatz-PC-Belastung
- Servergesteuerte Lastverteilung von parallel ausgeführten Diensten (OCR, Barcode, Konvertierungen/Rendering etc.) auf beliebig viele Rechner und Prozessoren
- Integrationsmöglichkeit zusätzlicher Recognition Engines (OCR, OMR, ICR, Barcode) von Drittherstellern
- Kompletter Konfigurationstransport bzw. -austausch zwischen zwei Servern als XML-Datei
- Rollenbasierte Verwaltung differenzierter Benutzerrechte für die Konfiguration, Dokumentenbearbeitung und den generellen Zugriff
- Dokumentenschutz durch Hashwertbildung, Verschlüsselung und Speicherung von Zugriffsrechten



Die DIGIDOK GmbH wurde 1997 gegründet und hat heute ihren Sitz in Essen. Das Unternehmen hat sich seit 2003 auf die Entwicklung von digitalen Dokumenten-Input-Lösungen spezialisiert. Mit ihrem smart.CAPTURE System zur Dokumentenerfassung zählt die DIGIDOK zu den wenigen Softwareherstellern, die über eine eigene Standardsoftware für die Klassifizierung und Freiformbasierte Datenextraktion zur automatischen Dokumentenindizierung verfügen. DIGIDOK Software wird von Unternehmen und Organisationen aller Branchen, von Verbänden, Wirtschaftsprüfungsgesellschaften und Steuerberatern zur Optimierung ihrer Geschäftsprozesse durch digitale Rechnungserfassung und Eingangspostverteilung eingesetzt. DIGIDOK vertreibt smart.CAPTURE über ECM-, Workflow- und sonstige Systemhäuser. Der Partnervertrieb wird kontinuierlich ausgebaut. Das CAPTURE-System lässt sich einfach über verschiedene Schnittstellen und Webservices in ECM/Archiv-, BPM/Workflow- und ERP/Finanzbuchhaltungs-Systeme integrieren.

DIGIDOK GmbH

Kruppstr. 82-100
45145 Essen

Tel. 0201.87919-50
Fax 0201.87919-59

info@digidok.com
www.digidok.com