

*L'applicazione della tecnologia OCR
in ambito contabile.
Opportunità e vantaggi.*

OCR

Optical Character Recognition

White Paper



Wolters Kluwer

Introduzione

Lo scopo di questo White Paper è aiutare a comprendere cos'è l'**OCR (Optical Character Recognition)** e come si può **applicare alla tenuta della contabilità**, illustrandone i vantaggi che apporta al processo di tenuta della contabilità nel suo complesso.

Nel corso della lettura emergerà chiaramente che riconoscimento OCR apporta notevoli benefici a tale processo e ne incrementa la produttività e la profittabilità.

Tramite il riconoscimento OCR è infatti possibile **digitalizzare l'intero processo**: dalla consegna dei documenti fino alla loro conservazione, rendendo più semplice anche la consultazione degli stessi.

Chiuderemo illustrando come operano concretamente le soluzioni Riconoscimento OCR applicate alla contabilità, con una particolare attenzione alla soluzione Wolters Kluwer in quanto contiene diversi elementi innovativi, quali ad esempio: l'integrazione nel sistema contabile, senza alcuna necessità di esportare tabelle e dati contabili; la proposizione del riconoscimento come servizio dove l'operatore di Studio non deve intervenire nel processo di riconoscimento; l'unificazione per il cliente e per l'operatore dello Studio dei momenti di gestione, consegna ed elaborazione contabile dei documenti indipendentemente dal formato digitale.

Contenuti

Cos'è l'OCR	6
L'OCR applicato alla contabilità	7
Il flusso della fattura PDF alla contabilità	8
Acquisizioni dati da PDF	9
Il processo di riconoscimento PDF in Studio	10
Possibili ostacoli e obiezioni all'adozione	11
Vantaggi	13
La soluzione Wolters Kluwer	14
Conclusioni	16
Glossario	17



Cos'è l'OCR

Il riconoscimento ottico dei caratteri, noto anche come OCR, è il processo di **conversione dei documenti cartacei** scansionati in **file elettronici di ricerca**.

In molte applicazioni per ufficio, come l'elaborazione delle fatture, permette di risparmiare tempo e ridurre i costi in termini di conversione dei documenti cartacei in documenti elettronici.

Ad esempio, l'inserimento di informazioni sui documenti in un database, noto anche come codifica a campo, è spesso una proposta molto costosa.

Utilizzando file convertiti elettronicamente, invece della carta originale, il processo di codifica di campo può essere automatizzato.

Il campo dell'OCR precede anche l'avvento dei computer, con brevetti originali datati fino al 1929. Naturalmente, gran parte dei progressi significativi in OCR sono direttamente correlati all'era informatica e di solito, più avanzato è il sistema OCR, più risorse di calcolo (ad es. processore più veloce) sono necessarie per utilizzarlo.

L'OCR e la relativa disciplina ICR (**Intelligent Character Recognition**), stanno cambiando il modo con cui vengono gestiti i documenti.

In molte applicazioni, tra cui librerie legali, contabili, bancarie, digitali, assicurazioni, backup remoti e gestione dei record, l'OCR automatizza il modo con cui le aziende elaborano i file.

L'OCR si presta direttamente all'estrazione dei dati e riduce i costi associati all'elaborazione di moduli.

Distribuito all'interno di un flusso di lavoro di acquisizione dati, l'OCR comporta molti vantaggi per gli utenti finali, tra cui la significativa riduzione del costo totale di archiviazione - un risparmio che di per sé fornisce una giustificazione al sostenimento dei costi per il software preposto alla funzione.

Parlando di OCR generalmente si utilizzano termini come **'mappatura del documento'** o riconoscimento **'free-form o semantico'**.

Nel primo caso si intende **'mappare'**, cioè definire la posizione o le coordinate ove è presente l'informazione/dato da recuperare.

Il vantaggio in questo caso è di una maggior precisione e quindi alte performance del OCR; di contro è necessario eseguire la fase di mappatura sui vari tipi di documenti (es. fatture con diverso formato) ed è consigliabile lavorare su documenti digitali nativi (il documento digitale frutto di scansione non rispetta sempre le medesime posizioni/coordinate, quindi mancato riconoscimento).

Nel secondo caso **'free-form o semantico'** si utilizzano parole chiave o termini (un piccolo vocabolario) che risultano presenti nei documenti (es. partita IVA, nr. documento, data doc., aliquota, imponibile etc.) e permettono di catturare il dato utile presente nelle vicinanze del termine ricercato. Questa tecnologia permette di trattare documenti (fatture) con formato diverso.

Anche documenti digitali frutto di scansione non richiedono un'uniformità delle coordinate del dato ricercato, nonostante sia però richiesta una buona qualità di lettura del documento (300dpi).



Per arrivare ad una corretta individuazione del dato può essere necessario utilizzare sia la mappatura del documento che il riconoscimento semantico: quest'ultimo infatti potrebbe individuare più volte la stessa parola chiave e rispettivi valori. In tal caso un operatore deve eseguire la mappatura per identificare la posizione nella quale si trova il valore rilevante.

L'OCR applicato alla contabilità

La tenuta della contabilità IVA si basa sulla raccolta dei documenti emessi e ricevuti dall'azienda o dal Professionista e consegnati allo Studio. Da qui parte l'attività dello Studio che va dal mettere in ordine i documenti ricevuti dai clienti, al Data Entry contabile, fino all'archiviazione del documento del cartaceo.

L'OCR estrae dalla fattura i dati utili per il Data Entry, ad esempio la Partita IVA, il numero e la data del documento, il totale, l'imponibile distinto per aliquota IVA: tutte informazioni che l'operatore deve trovare sul singolo documento e poi digitare manualmente.

I sistemi OCR applicati alle fatture utilizzano di norma sia la tecnologia "**free form**" che la mappatura per risolvere i casi di più parole chiave presenti sul PDF, ad esempio nel caso in cui sul PDF di una fattura sono presenti più date (la data della fattura, la data dell'ordine, la data del DDT) e l'operatore sceglie quella corretta



Il flusso dalla fattura PDF alla contabilità



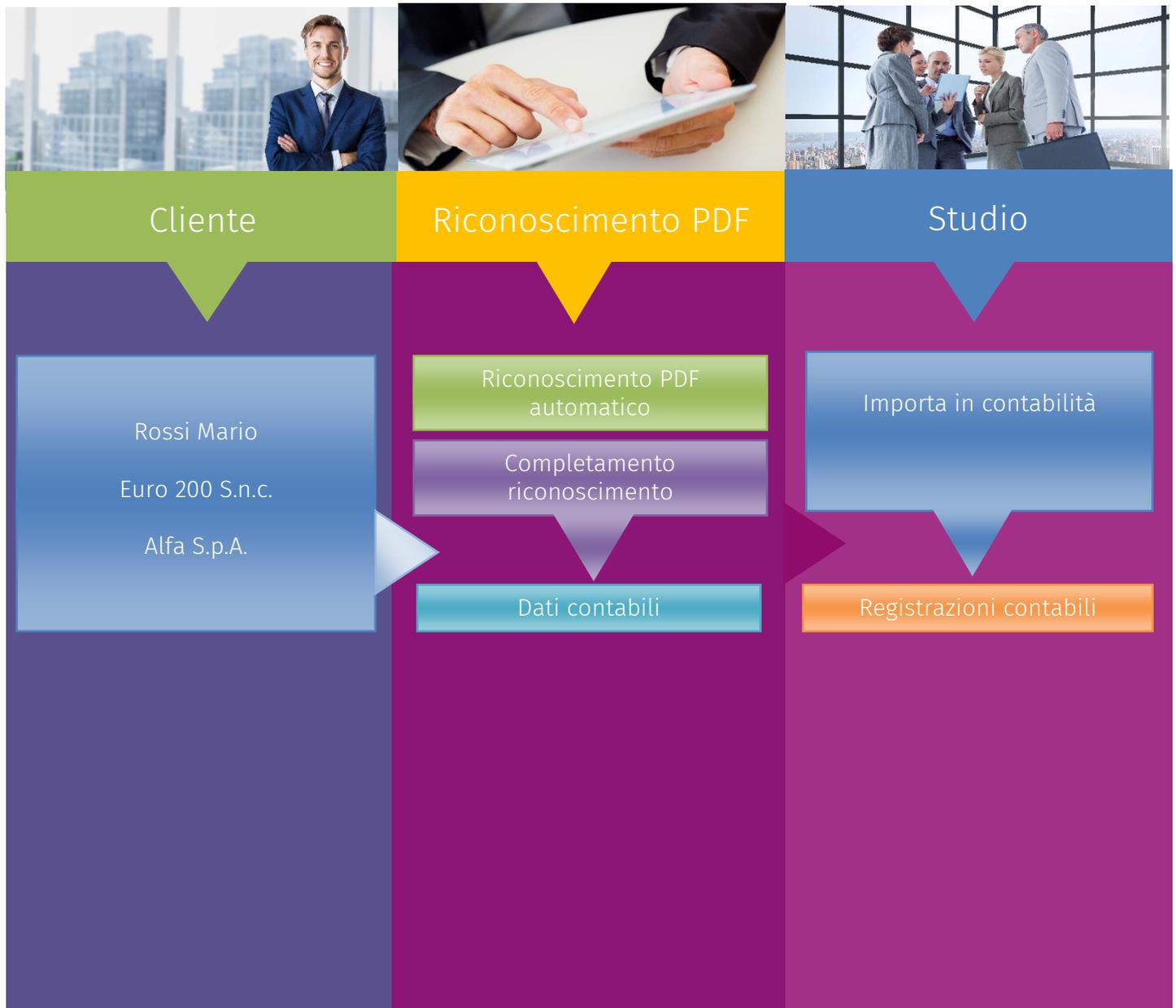
Il flusso dalla fattura alla contabilità coinvolge tre attori: il cliente, il riconoscimento PDF, lo Studio.

- 1) Il **cliente** rende disponibile il documento emesso o ricevuto in formato PDF, nei casi di fattura su carta tramite scansione rispettando i requisiti tecnici richiesti dal fornitore del servizio di riconoscimento PDF.
- 2) Il servizio di **riconoscimento PDF** elabora quanto ricevuto e invia allo Studio il PDF accompagnato dai dati estratti in un formato fruibile dal sistema contabile. Il riconoscimento PDF avviene in due fasi:
 - **Automatica**, in cui il motore OCR esegue le ricerche per parole chiave ed estrae le informazioni richieste
 - **Manuale**, in caso di ambiguità o mancanza di quadratura tra i dati estratti e il PDF. In questo caso il motore OCR richiede l'intervento di un operatore che fornisce le indicazioni necessarie per estrarre i dati; l'operatore interviene anche per posizionamento, indica cioè al motore le coordinate relative a quel dato per quello specifico documento (fattura) in modo possa essere rilevato automaticamente nelle prossime elaborazioni (autoapprendimento). Grazie ad esso, la fase di importazione contabile e il servizio Riconoscimento PDF diventano via via più efficienti.
- 3) **In Studio** l'operatore contabile completa la registrazione contabile con la possibilità di aprire e visualizzare il PDF. In questa fase l'uso del doppio monitor, pur non essendo indispensabile, rende più veloce l'attività dell'operatore.

Per cogliere appieno i vantaggi del riconoscimento PDF è bene chiudere il processo con **la conservazione a norma dei documenti digitalizzati**, in questo caso le fatture, eliminando i costi della carta.

Acquisizione dati da PDF

Il flusso di acquisizione dei dati contabili in PDF è così raffigurabile:



Completata la registrazione contabile, il PDF è consultabile dalla registrazione contabile eliminando i tempi e i **costi relativi alla ricerca** della fattura cartacea.

Il processo di riconoscimento PDF in Studio

Aumentare produttività e margini del servizio di tenuta contabilità

Il motivo che spinge lo Studio ad avviare il processo riconoscimento PDF è aumentare la produttività e profittabilità della tenuta contabile, riducendo il tempo dedicato al Data Entry. Il flusso dal PDF alla registrazione contabile permette di abbattere i costi dell'intero processo di tenuta contabile e **non solo quelli più visibili legati al Data Entry.**

Per quantificare i vantaggi del sistema OCR è quindi opportuno quantificare il costo dell'intero ciclo di vita del documento contabile: dalla sua emissione, alla consegna in Studio fino alla conservazione per dieci anni.

Il costo medio totale della gestione del documento fattura in Studio è stimato tra i **7 e i 10 euro**; per inciso, il costo aziendale arriva fino a **60 Euro**, perché la gestione del documento comporta anche la riconciliazione della fattura con i documenti di trasporti e gli ordini.



Come avviare il processo di riconoscimento PDF

Per avviare con successo il processo in Studio occorre **individuare i primi clienti** con cui iniziare e fare esperienza in Studio e mettere a punto il flusso.

I clienti da cui partire possono essere:

I clienti che già consegnano le fatture in formato PDF, come allegato ad una mail

I clienti che utilizzano applicazioni gestionali in grado di generare il PDF della fattura emessa e che già ricevono dai loro fornitori le fatture in formato PDF; le fatture ricevute su carta sono riconducibili al formato PDF tramite la scansione, che deve rispettare i requisiti minimi richiesti dal servizio riconoscimento PDF. In una prima fase lo Studio può accollarsi l'onere della scansione dei documenti consegnati su carta.

I clienti che ricevono su carta un numero significativo di fatture di acquisto: lo Studio potrebbe ottenere dei benefici in termini di efficienza anche dal semplice ordinamento per data delle fatture.



Possibili ostacoli e obiezioni all'adozione del riconoscimento

“Nessuno dei miei clienti eseguirebbe la scansione”

Nel contesto dello Studio professionale la scansione dei documenti è vista come un ostacolo all'adozione di un sistema OCR.

È spesso un falso problema:

- Lo sviluppo e la diffusione della tecnologia informatica hanno reso possibile l'invio e la ricezione delle fatture come allegato di una mail in formato PDF.
- Le società di utilities (es. Servizi Telefonici, Telepass) rendono disponibili le fatture in formato PDF nell'area riservata dei clienti.

La scansione può quindi diventare una attività residuale per la maggior parte delle contabilità gestite in Studio.

Chi controlla eventuali errori del riconoscimento PDF?

Il riconoscimento PDF per portare le informazioni in contabilità è fornito da aziende specializzate nella fornitura di soluzioni per la gestione documentale che producono dati integrabili nei sistemi gestionali. La fase di OCR "free-form" o semantica è gestita automaticamente dai motori OCR.

La fase di controllo e correzione per "posizionamento", che è necessaria per risolvere i casi in cui più informazioni riportate nel documento corrispondono al dato cercato, è invece affidata ad un operatore, che potrebbe essere il cliente, l'operatore di Studio o l'operatore del servizio.

Wolters Kluwer, ad esempio, ha stretto un accordo con una società che esegue tramite suoi operatori la correzione e l'interpretazione dei dati ambigui.



Possibili ostacoli e obiezioni all'adozione del riconoscimento

L'assegnazione del numero di protocollo alle fatture di acquisto

L'apposizione manuale del protocollo richiederebbe di stampare i PDF ricevuto in formato digitale, riducendo quindi i vantaggi del riconoscimento PDF.

Per evitare l'apposizione manuale è indispensabile procedere **alla conservazione digitale dei documenti**, aggiungendo i vantaggi della conservazione digitale a quelli del riconoscimento OCR.

Nella fattispecie, diverse risoluzioni ministeriali hanno aperto la strada alla gestione automatica del numero di protocollo.

Le condizioni poste sono:

- 1) la garanzia di univocità del documento;
- 2) la stampa di un registro delle corrispondenze tra il protocollo assegnato e il criterio di univocità individuato

"La mia impiegata impiega un minuto a registrare un documento"

Attenzione! Il tempo di registrazione non è l'unico elemento di costo. L'intero processo di tenuta della contabilità comporta:

- un "prima" della registrazione: ricevere classificare e ordinare i documenti ricevuti
- un "dopo" la registrazione: riporre in archivio i documenti, ricercare un documento per un controllo, restituire i documenti al cliente.

"Come posso registrare senza vedere la fattura?"

Il problema è superabile con la soluzione hardware ormai a basso costo del doppio monitor o di schermo "wide", in modo da avere contemporaneamente aperti il programma di prima nota e il PDF in fase di registrazione

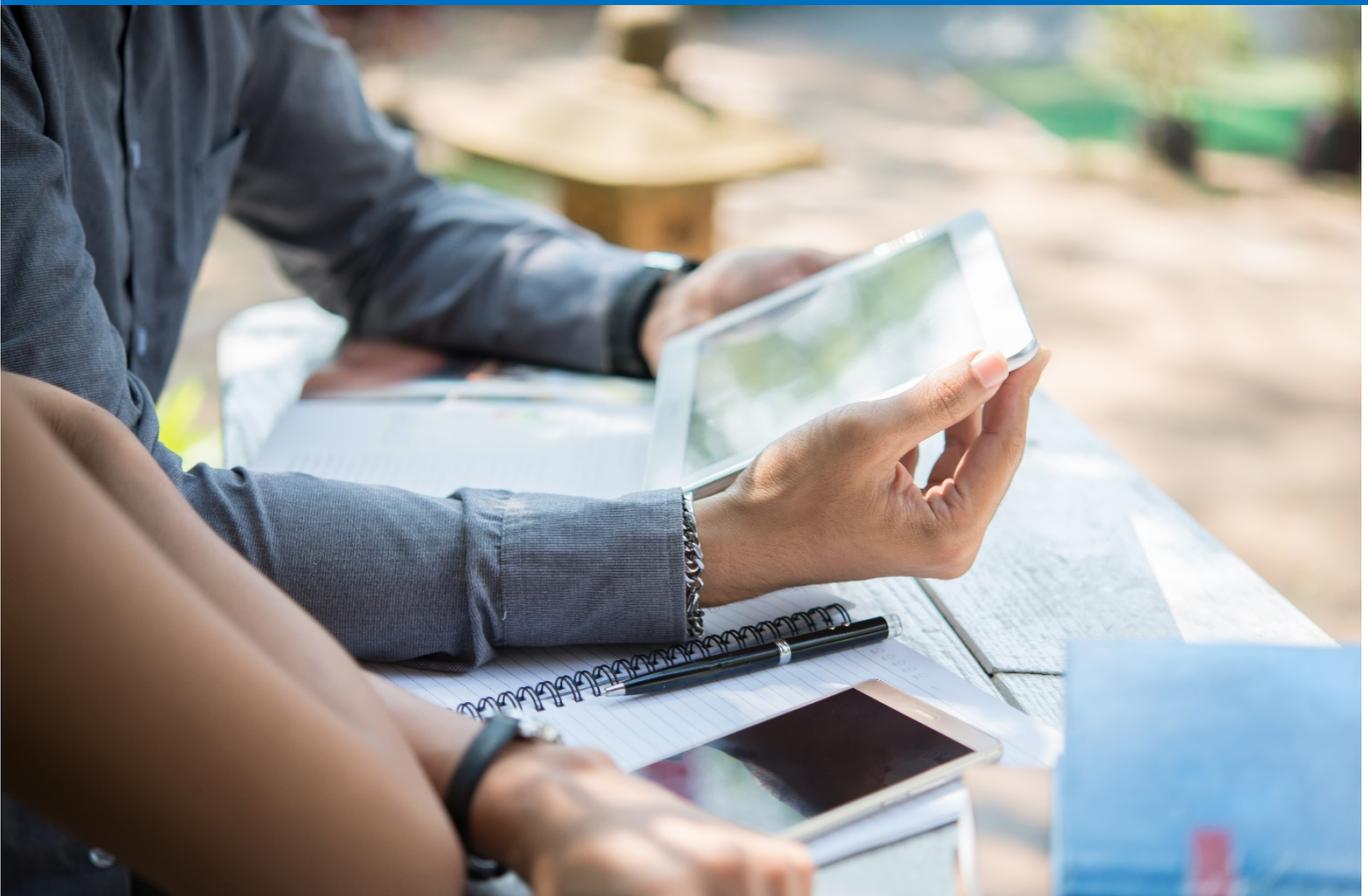


Vantaggi

I vantaggi dell'applicazione di un flusso di fatturazione PDF alla contabilità variano in relazione a come il processo viene organizzato e sono pienamente comprensibili solo analizzando il flusso nella sua **completezza** e non solo limitatamente all'attività di inserimento dei dati.

Rispetto al **processo cartaceo** si eliminano i seguenti passaggi:

- **il tempo per dividere le fatture tra attive e passive**, che altrimenti dipende dalla diligenza del singolo cliente (è raro ricevere una busta con il mese già perfettamente ordinato);
- **il tempo per l'attività di ordinamento delle fatture in base alla data del documento**, che può essere eseguita automaticamente dalla procedura contabile;
- **il tempo per cercare e inserire la partita IVA del cliente/fornitore**, il numero, la data del documento e il suo totale. L'integrazione con il sistema contabile consente di avere già precompilato il dato Partita IVA, la contabilità seleziona l'anagrafica del cliente o fornitore oppure propone la creazione di una nuova anagrafica senza che l'operatore debba cercare la partita IVA sul documento, digitarla, avviare la ricerca;
- **il tempo per scrivere manualmente il numero di protocollo**, se integrato con la Conservazione digitale a norma.



La soluzione Wolters Kluwer Italia

Come esempio di applicazione della tecnologia OCR alla contabilità, analizziamo sia la soluzione generalmente proposta agli studi che la soluzione proposta da Wolters Kluwer Italia, seguendo l'intero flusso dalla consegna dei PDF allo Studio fino alla loro conservazione.

ATTIVITÀ	FLUSSO ORDINARIO	FLUSSO WOLTERS KLUWER
Requisiti per integrazione tra sistema riconoscimento OCR e Contabilità	Le anagrafiche dei clienti e fornitori e parte o tutte le tabelle del sistema contabile, ad esempio le causali contabili o il piano dei conti, devono essere inviate al sistema di riconoscimento	Nessun requisito
Il cliente consegna i PDF allo Studio	Via mail o portale del fornitore del servizio di riconoscimento	Portale dei servizi on-line che lo Studio offre ai suoi clienti: webdesk un unico contenitore per tutti i servizi. Wolters Kluwer esegue l'invio in modo sicuro al fornitore del servizio di riconoscimento. In ogni momento lo Studio, per ogni contabilità e per ogni documento, apprende a che punto è l'attività di elaborazione. Oltre alla consegna di PDF, è possibile allegare fogli Excel con i dati delle fatture, fatture in formato FEPA (XML) firmato (es fatture PA) o non firmato. Massima adattabilità del sistema a cosa può fornire il cliente: modalità di consegna distinte tra vendite e acquisti, e all'interno dello stesso tipo di fatture.
Riconoscimento OCR semantico	Motore OCR del fornitore	Motore OCR del fornitore
Riconoscimento OCR mappatura per casi ambiguità	Operatore di Studio o cliente	Operatore del fornitore
Esito estrazione dati utili alla registrazione	Riempimento parziale della maschera di preregistrazione del sistema di riconoscimento	Restituzione del file XML in formato FEPA e del PDF elaborato sul portale webdesk

ATTIVITÀ	FLUSSO ORDINARIO	FLUSSO WOLTERS KLUWER
Completamento preregistrazione	L'operatore di Studio accede al portale del fornitore del servizio e nella maschera di preregistrazione completa la registrazione, utilizzando le tabelle importate dalla contabilità.	Quando l'operatore di Studio accede alla prima nota, il programma propone l'acquisizione delle fatture consegnate su webdesk . Qui può completare la registrazione in una griglia di preregistrazione all'interno della contabilità, utilizzando strumenti che consentono di completare automaticamente le registrazioni. L'operatore contabile opera sempre nell'ambito contabile che conosce e sempre nello stesso modo indipendentemente dal formato di consegna della fattura (XLS; PDF; XML FEPA)
Contabilizzazione	Importazione di file in contabilità	Un clic, perché si è già in contabilità.
Consultazione PDF	Disponibile sul portale del fornitore di servizio	Disponibile dalla prima nota e dalla scheda contabile
Conservazione digitale a norma	Può essere offerta o meno dal fornitore di servizio	Integrata con la soluzione di Conservazione Digitale, un unico punto di riferimento per i tutti i documenti conservati dallo Studio per conto dei clienti

Conclusioni

Trasformare la fattura PDF in registrazione contabile apporta notevoli vantaggi allo Studio, vantaggi che sono tanto più evidenti quanto più ci si sforza di andare oltre il solo Data Entry per vedere l'intero flusso.

... guardando all'intero flusso di lavoro si scoprono le attività a basso valore aggiunto che rendono improduttiva la contabilità, e i benefici apportati dal riconoscimento PDF.



Glossario

OCR

Optical Character Recognition, è una tecnologia che consente di tradurre i caratteri in testo digitale.

PDF

Portable Document Format, è un formato di file basato su un linguaggio di descrizione di pagina sviluppato da Adobe Systems nel 1993. I documenti sono visualizzabili in modo indipendente dall'hardware e dal software utilizzati per generarli.

ICR

Intelligent Character Recognizing (Riconoscimento intelligente di caratteri).



Per maggiori informazioni www.software.wolterskluwer.it

Follow us



I nomi e i marchi citati sono registrati dai rispettivi proprietari. Wolters Kluwer Italia si riserva la facoltà di modificare i prodotti senza preavviso. Giugno 2017

Wolters Kluwer Italia srl – Tutti i diritti riservati. Tutti i materiali informativi sono costantemente aggiornati.