XCENTER DIGITAL

Optical Character Recognition

Best Practices

NOTICE: The information within this document is proprietary to XCENTER DIGITAL and is intended for use by employees, business partners and customers. It may not be copied, duplicated, used in any way, or passed to any third party without the prior written consent of XCENTER DIGITAL.

Optical Character Recognition (OCR) – Overview

Although the scanning of paper-based invoices isn't considered e-invoicing, it is the natural first step for many organisations. Using OCR software, the data can be moved from a paper-based format to a digital format that can be entered in the Accounts Payable system. Outside of EDI and invoice portals, OCR has been a predominant tool of choice to enable the digitization of invoices.
Optical Character Recognition (OCR) technology is a hardware/software tool that takes a paper document, usually an invoice, scans then "reads" it and turns it into metadata that can be used to populate fields in a database.

From there the invoice can be brought into an electronic workflow for processing. Using OCR software, the data can be moved from a paper-based format to a digital format that can be entered in the AP system. OCR is the electronic conversion (through scanning) of invoices without extractable data (either paper or image files) into data that can be integrated directly (as an EDI or XML file) into a buyers Accounts Payable finance system for payment.

Whilst OCR solutions enables organisations to automate their AP processes to a certain extent, there are restrictions that are inherent to OCR technology, and which limit its impact beyond achieving a semi-automated state, where human intervention and errors are part and parcel of the technology in question. After all, we speak of "recognition" and not "extraction" when referring to OCR. Fundamentally, OCR solutions are all based on a similar probabilistic technology and methodology. For instance, the number "1" vs. lowercase letter "L", the number "0" vs. uppercase O, and so on. The latter is mitigated to some extent by the use of dictionaries (for example, "INVOICE" is more likely than "1NV0lCE"), but unfortunately invoice data such as the invoice number or the shipping reference, is usually not to be found in an OCR dictionary.

The challenge gets even more difficult when using OCR for invoice line item extraction. These inherent limitations of OCR result in varying accuracy recognition rates, which invariably requires human operators to check the results produced by OCR. Inaccuracies require manual intervention, leading to errors, long invoice processing time, and low percentage of "touchless" invoices or processed "straight-through".

Optical Character Recognition (OCR) - Best Practices

Situational analysis

OCR tools perform best with good quality typed documents. Therefore, if the scan quality is poor or the invoice is stained, wrinkled or the text is weakly printed then OCR will fail to successfully digitize such invoices. OCR tools also typically fail to read handwritten invoices, and therefore fonts that resemble handwriting create many errors during the OCR process. AP teams need to undertake a detailed analysis of the printed source material.

Specifics such as paper quality, characteristics such as language, font, and layout, and graphical elements are critical features that may affect the quality of data capture. This will provide information that will help to determine whether the data capture can be easily improved or not. For example, certain kind of historical documents may lack the lexical data that is required for OCR data entry and this can pose challenges. Image rich documents may need special measures to render them OCR compatible or may need improvised OCR data capture.

OCR process workflow

It cannot be emphasized enough that having a well-delineated process workflow will determine the success or failure of your OCR data entry effort. A well-charted flow will ensure that your data capture and conversion succeeds as per your expectations.

OCR quality check procedures

OCR data capture projects require quality assurance procedures in order to ensure a certain level of control. A quality assurance program will safeguard that the OCR results are on track and that the goals are achieved within the defined time.

Part of the procedure consists of the AP team carrying out a comprehensive review of the complete captured data set or a sectional review if historical reasons dictate this approach – for instance in the case of a spin-off. QA procedures also include tracking and modification of OCR errors. Ensure that your QA program is rigorously worked out and implemented and that the quality standards are adequately communicated to all staff involved.

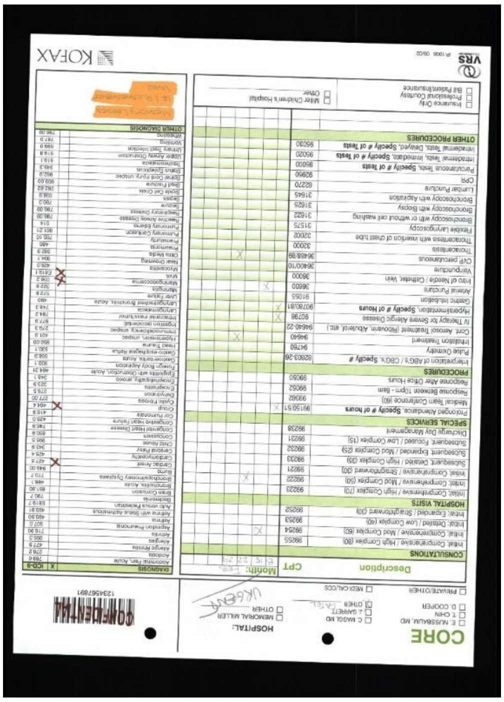
Quality of master vendor data

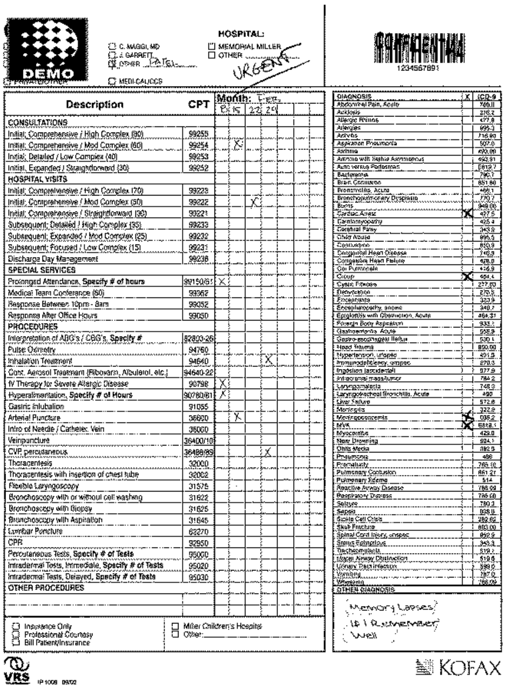
OCR solutions do not improve quality, accessibility and reliability of master data since there is a process gap between creation and digitization of invoices. It therefore helps to ensure accurate and up to date master vendor data.

Possible methods of processing

a) In-house mailroom and small scan workplaces
b) Outsourcing (PO boxes need to be set up)
c) Combination of both methods (PO boxes + centralized scanning for the majority of documents coming from vendors, local scan workplaces with desktop-class scanners for documents received locally on branches – faster processing but still good quality)

Equipment - Production scanner(s), image enhancement technology (VRS) – examples:





DEMO

○ C. MAGGI, MD
○ J. GARRETT
☒ OTHER PATEL
○ MEDI-CAL/CCS

HOSPITAL:
□ MEMORIAL MILLER
□ OTHER

URGENT

Description	CPT	Month: Feb. 8	15	22	29			
CONSULTATIONS								
Initial; Comprehensive / High Complex (80)	99255							
Initial; Comprehensive / Mod Complex (60)	99254		X					
Initial; Detailed / Low Complex (40)	99253							
Initial, Expanded / Straightforward (30)	99252							
HOSPITAL VISITS								
Initial; Comprehensive / High Complex (70)	99223							
Initial; Comprehensive / Mod Complex (50)	99222			X				
Initial; Comprehensive / Straightforward (30)	99221							
Subsequent; Detailed / High Complex (35)	99233							
Subsequent; Expanded / Mod Complex (25)	99232							
Subsequent; Focused / Low Complex (15)	99231							
Discharge Day Management	99238							
SPECIAL SERVICES								
Prolonged Attendance, **Specify # of hours**	99150/51	X						
Medical Team Conference (60)	99362							
Response Between 10pm - 8am	99052							
Response After Office Hours	99050							
PROCEDURES								
Interpretation of ABG's / CBG's, **Specify #**	82803-26							
Pulse Oximetry	94760							
Inhalation Treatment	94640				X			
Cont. Aerosol Treatment (Ribovarin, Albuterol, etc.)	94640-22							
IV Therapy for Severe Allergic Disease	90796	X						
Hyperalimentation, **Specify # of Hours**	90780/81	X						
Gastric Intubation	91055							
Arterial Puncture	36600		X					
Intro of Needle / Catheter, Vein	36000							
Veinpuncture	36400/10							
CVP, percutaneous	36488/89				X			
Thoracentesis	32000							
Thoracentesis with insertion of chest tube	32002							
Flexible Laryngoscopy	31575							
Bronchoscopy with or without cell washing	31622							
Bronchoscopy with Biopsy	31625							
Bronchoscopy with Aspiration	31645							
Lumbar Puncture	62270							
CPR	92950							
Percutaneous Tests, **Specify # of Tests**	95000							
Intradermal Tests, Immediate, **Specify # of Tests**	95020							
Intradermal Tests, Delayed, **Specify # of Tests**	95030							
OTHER PROCEDURES								

○ Insurance Only
○ Professional Courtesy
○ Bill Patient/Insurance

□ Miller Children's Hospital
□ Other:

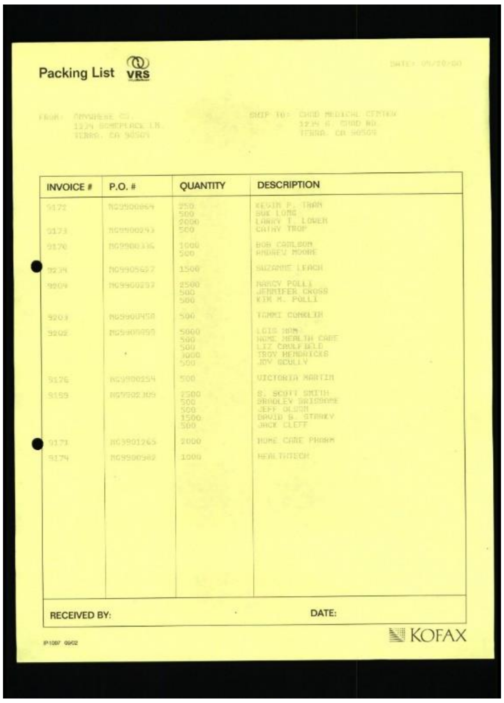
VRS IP 1008 08/02

1234567891

DIAGNOSIS	X	ICD-9
Abdominal Pain, Acute		789.0
Acidosis		276.2
Allergic Rhinitis		477.8
Allergies		995.3
Arthritis		716.90
Aspiration Pneumonia		507.0
Asthma		493.90
Asthma with Status Asthmaticus		493.91
Auto versus Pedestrian		E814.7
Bacteremia		790.7
Brain Contusion		851.80
Bronchiolitis, Acute		466.1
Bronchopulmonary Dysplasia		770.7
Burns		949.00
Cardiac Arrest	X	427.5
Cardiomyopathy		425.4
Cerebral Palsy		343.9
Child Abuse		995.5
Concussion		850.9
Congenital Heart Disease		746.9
Congestive Heart Failure		428.0
Cor Pulmonale		416.9
Croup	X	464.4
Cystic Fibrosis		277.00
Dehydration		276.5
Encephalitis		323.9
Encephalopathy, unspec		348.3
Epiglottitis with Obstruction, Acute		464.31
Foreign Body Aspiration		933.1
Gastroenteritis, Acute		558.9
Gastro-esophageal Reflux		530.1
Head Trauma		850.00
Hypertension, unspec		401.9
Immunodeficiency, unspec		279.3
Ingestion (accidental)		977.9
Intracranial mass/tumor		784.2
Laryngomalacia		748.3
Laryngotracheal Bronchitis, Acute		490
Liver Failure		572.8
Meningitis		322.9
Meningococcemia	X	036.2
MVA	X	E818.1
Myocarditis		429.0
Near Drowning		994.1
Otitis Media		382.9
Pneumonia		486
Prematurity		765.10
Pulmonary Contusion		861.21
Pulmonary Edema		514
Reactive Airway Disease		786.09
Respiratory Distress		786.09
Seizure		780.3
Sepsis		038.9
Sickle Cell Crisis		282.62
Skull Fracture		803.00
Spinal Cord Injury, unspec		952.9
Status Epilepticus		345.3
Tracheomalacia		519.1
Upper Airway Obstruction		519.8
Urinary Tract Infection		599.0
Vomiting		787.0
Wheezing		786.09
OTHER DIAGNOSIS		

Memory Lapses
If I Remember
Well

KOFAX

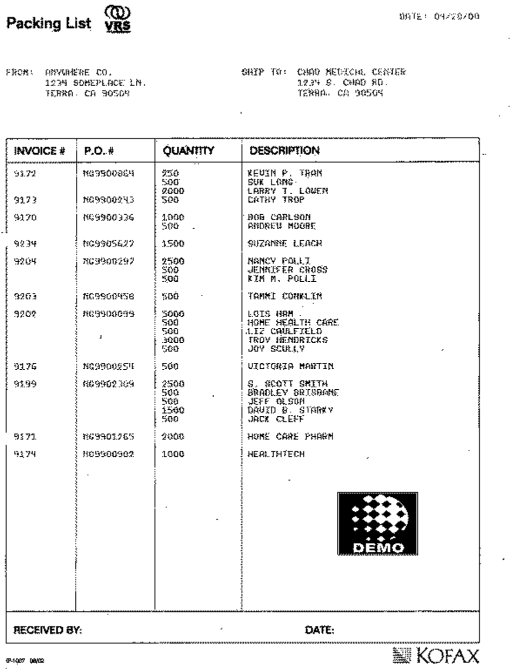


Packing List VRS

INVOICE #	P.O. #	QUANTITY	DESCRIPTION

RECEIVED BY: DATE:

KOFAX



Packing List VRS

DATE: 04/20/00

FROM: ANYWHERE CO.
1234 SOMEPLACE LN.
TERRA. CA 90509

SHIP TO: CHAD MEDICAL CENTER
1234 S. CHAD RD.
TERRA. CA 90509

INVOICE #	P.O. #	QUANTITY	DESCRIPTION
9172	NG9500864	250 500 2000	KEVIN P. TRAN SUK LONG LARRY T. LOWEN
9173	NG9900243	500	CATHY TROP
9170	NG9900336	1000 500	BOB CARLSON ANDREW MOORE
9234	NG9905627	1500	SUZANNE LEACH
9204	NG9900297	2500 500 500	NANCY POLLI JENNIFER CROSS KIM M. POLLI
9203	NG9900458	500	TAMMI CONKLIN
9202	NG9900099	5000 500 500 3000 500	LOIS HAM HOME HEALTH CARE LIZ CAULFIELD TROY HENDRICKS JOY SCULLY
9176	NG9900254	500	VICTORIA MARTIN
9199	NG9902309	2500 500 500 1500 500	S. SCOTT SMITH BRADLEY BRISBANE JEFF OLSON DAVID B. STARKY JACK CLEFF
9171	NG9901265	2000	HOME CARE PHARM
9174	NG5900902	1000	HEALTHTECH

DEMO

RECEIVED BY: DATE:

KOFAX

Background color removal, de-skew, de-speckle, automatic orientation, auto crop, edge clean-up, blank page removal, de-dithering…

** For small workplaces we recommend Kofax Express*

If small local scanning workplaces are not possible for budgetary reasons for example, we recommend to scan on MFP devices in full color and then use software binarization for image enhancement (explained later in "Email attachments" chapter)

Scanning parameters

- Vertical resolution: 300DPI
- Horizontal resolution: 300DPI
- Color: black and white
- Sizeautomatic cut according to the actual size of the document
- Output scanning: 1x multipage PDF file per document (incl appendixes if applicable)
- Image enhancement: Image enhancing techniques must be applied on document. Techniques such as: De-skew, Black border crop & remove, De-speckle / Noise reduction, Punch Hole fill, Blank page removal etc.

NOTICE: The information within this document is proprietary to XCENTER DIGITAL and is intended for use by employees, business partners and customers; it may not be copied, duplicated, used in any way, or passed to any third party without prior written consent.