IBM Document Capture Update

Putting Content to Work
ECM UserNet 2011

Disclaimer
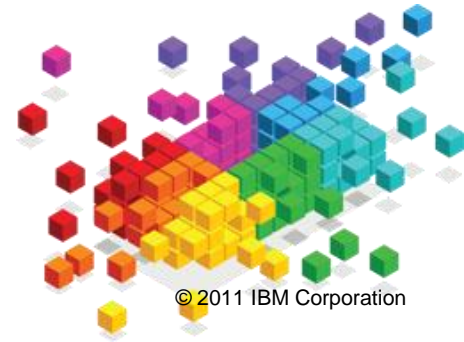
Putting Content to Work
ECM UserNet 2011

Agenda

- The Importance of Document Capture

- IBM Datacap Capture Update

- IBM FileNet Capture Update
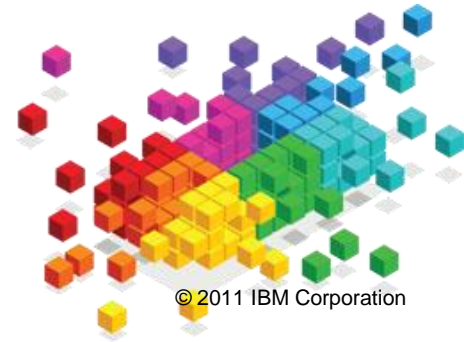
The Challenges Customers Face…

How much time is wasted manually keying data and indexes?

Blue Cross Blue Shield of Arizona:

Doubled their Data Entry productivity for Medical Claims and Enrollment processes

Excel Transportation Services
Reduced FTEs by over 75% by processing 85% of invoices automatically

How can content be fed quickly from remote locations?

A Transportation firm operating in more than 40 countries:

Reduced shipping request cycle time by 2 days

A Wholesale Mortgage Lender with 13 locations:

Reduced copy/overnight mail costs by $30,000/month in one location alone

*What if you could **extract information** from content automatically from any location?*

Putting Content to Work
ECM UserNet 2011

How do Customers Achieve their ROI goals?

- Reduce cost of transporting paper to a central location
  - Scan documents in remote locations – branches, stores, offices, etc.
  - Savings can be more than $1M annually
  - Key capability - Distributed Capture

- Reduce data entry labor costs
  - Extract data from documents without manual keying
  - Potential to reduce data entry staff up to 90%
  - Large organizations can have hundreds of employees performing data entry
  - Key capabilities – Rules, Advanced Data Extraction

- Reduce cost of document capture
  - Reduce paper sorting and document preparation
  - Potential to reduce capture staff up to 50%
  - Key capabilities – Rules, Advanced Data Extraction

- Standardize
  - Single vendor ECM and Capture solution
  - Replace obsolete or costly legacy capture systems
  - Reduce license fees, support and maintenance costs
  - Eliminate volume-based pricing

Benefits of Document Imaging & Capture

- Better **customer/vendor service** and communications

- **Reduced time and resources** required to manage paper and related business processes

- **Improved cash flow**, reduced transaction and paper costs while growing the business

- **Improved collaboration** as documents can be immediately accessed and shared around the world

- **Elimination of lost files**

- Secure and reliable **backup and disaster recovery**

- **And overall** Return On Investment
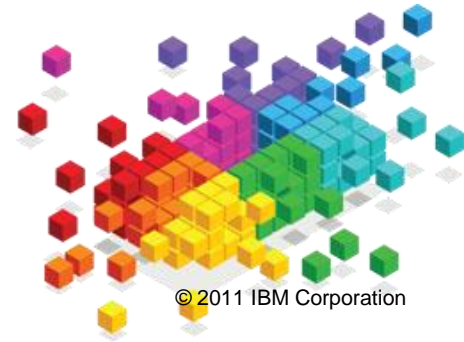
Components of Enterprise Capture

Central and Adhoc

Central Mallroom

Field

Field

Remote and Distributed

Fax

eMail

Central

CAPTURE

Collect

Process

Transform

ERP ECM BPM BI

Central and Adhoc

Department

Branch

Remote and Distributed
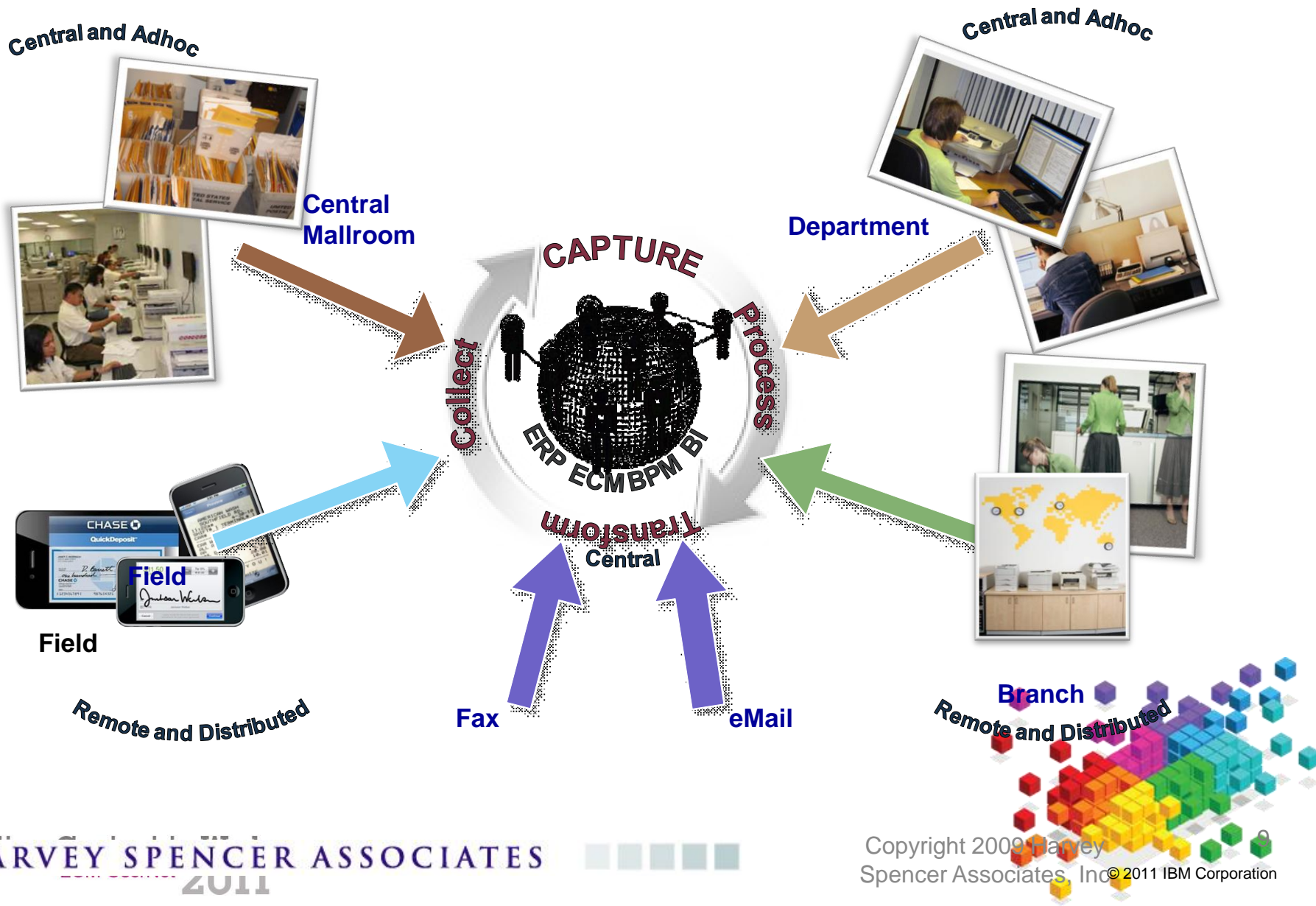
Datacap Document Capture Update

Putting Content to Work
ECM UserNet 2011
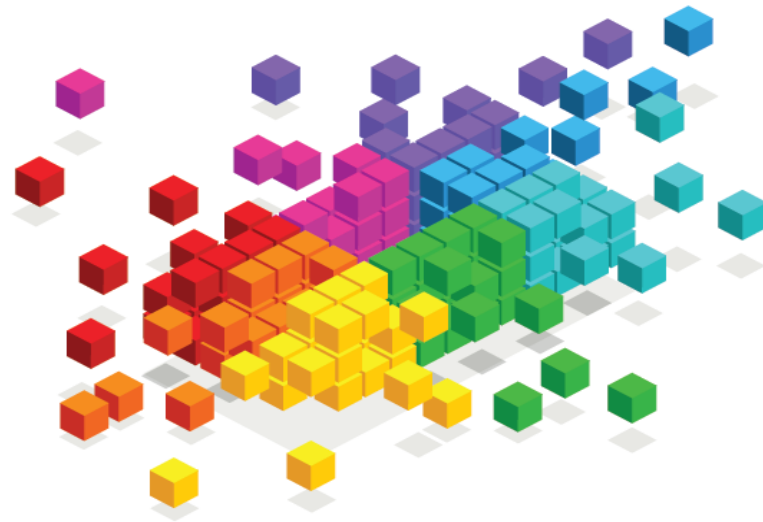
About Datacap

- 22 years as ISV

- Headquartered in Tarrytown, NY

- Leadership by co-founders (CEO & CTO)

- Dedicated to developing document capture solutions

- First-mover technical innovation

- Focused on demanding capture

- IBM FileNet ISV Partner since 1994

- PartnerWorld member since 2004

- Acquired by IBM August 2010

- Completed Transfer of Trade April 1, 2011
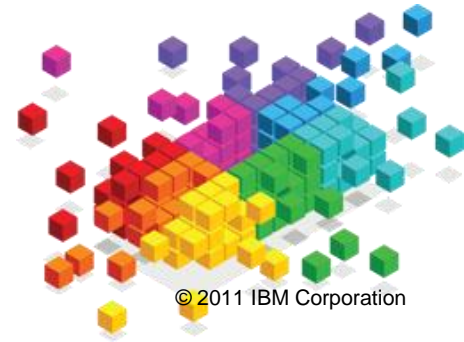
What Does Datacap Offer IBM ECM Customers?

Datacap offers a comprehensive Document Capture portfolio, offers unique customer value, and enhances IBM's Enterprise Content Management offerings

- Datacap brings new capabilities to IBM's already extensive set of ECM solutions to help companies make better decisions faster by managing content, optimizing processes and enabling compliance through agile ECM solutions and advanced case management.

- The Datacap acquisition expands IBM's leadership position in Enterprise Content Management (ECM) by offering:
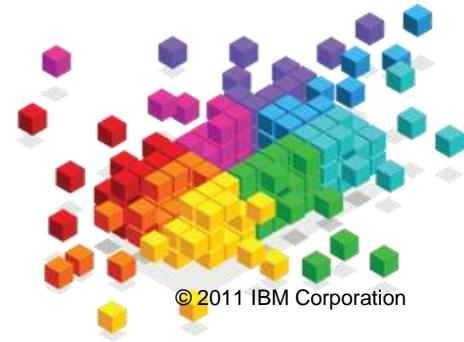 - A suite of enterprise capture solutions – centralized, and distributed thin client – that enhance current IBM ECM offerings (FileNet and CM8) and support non-IBM repositories from EMC, Open Text, Oracle, Microsoft and others.
  - Coverage for distributed and departmental capture needs with capture for smaller deployments such as Microsoft SharePoint.
  - The industry's most sophisticated technology for document capture, in process capture, and high speed data extraction, enabling organizations to extract meaningful information from unstructured content for better, faster decisions and significantly reduce the labor and paper costs of the capture process.

- Together, IBM and Datacap provide industry-leading Document Capture and ECM capabilities.

Datacap's Record of Capture Innovation

1994 – Integrated Capture Workflow

2006 – Capture Web Services

2008 – Flex Agile Implementation

1989 - PC-based Forms Processing

2000 – *Web-based Capture*

1988

2009-10 Rulerunner Quattro FastDoc Capture eDoc Actions

1992 – 1st PC ICR Solution

2002 – Invoice Processing

1998 – Healthcare Claims Processing

2007 – Document ID with Text Analytics
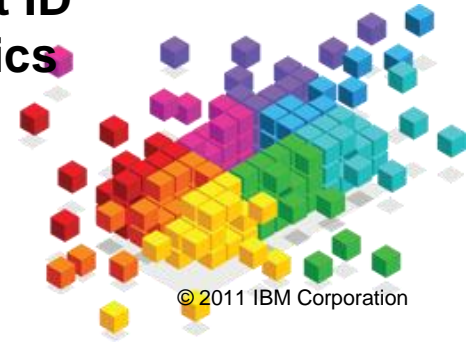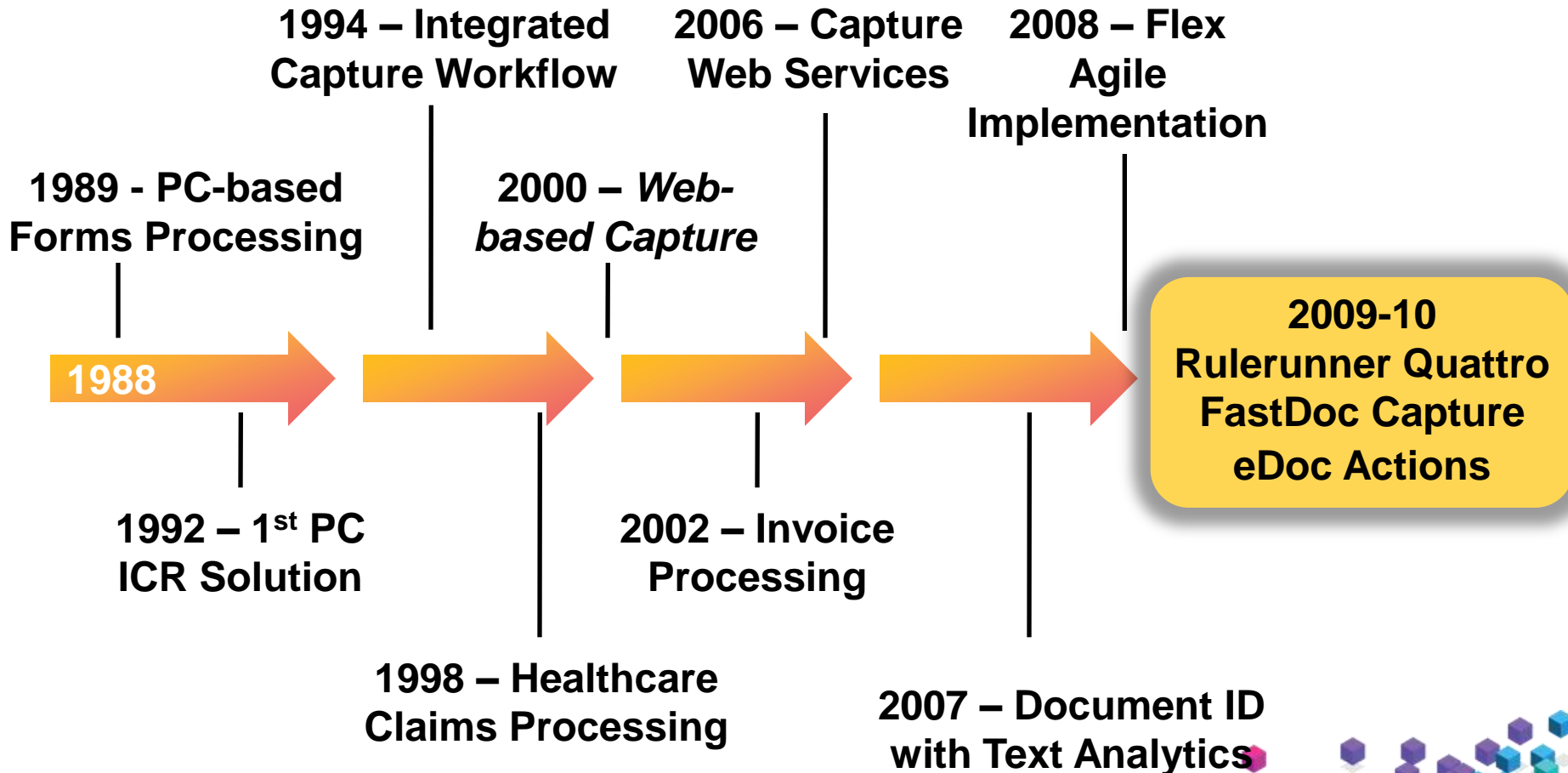
Problem: Traditional Capture Applications Are Inflexible
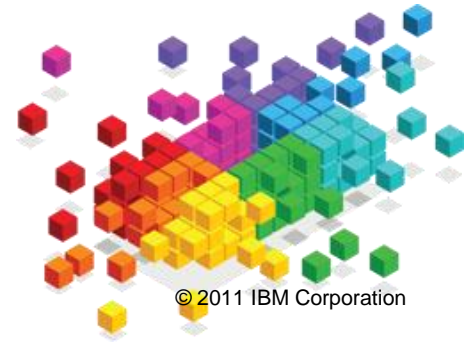
Variability is the challenge…

PLUNGER LIFT SYSTEMS
PO Box 9423
Midland, TX 79708

INVOICE
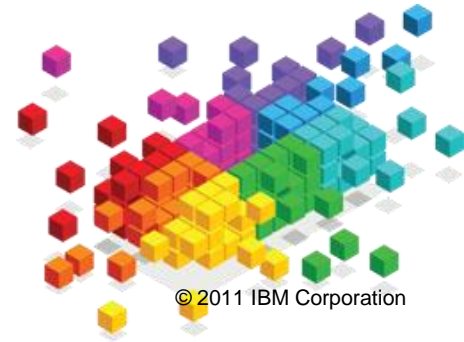
- Variable requirements
- Variable documents
- "Unreliability" of recognition
- Changing mix of formats – paper, image, fax, electronic
- Exceptions, exceptions, exceptions…

Stinger Wellhead
PO Box 1039
Allen, TX 75013-0026

Customer:
Name   HC Technologies
Address   4903 Business Parkway
City   Rochester   State
Phone   (507) 555-1212

TOTAL DUE   554.68

Make all checks payable to Plunger Lift Systems
If you have any questions concerning this invoice, contact (432) 699-1200

Thank you for your business!
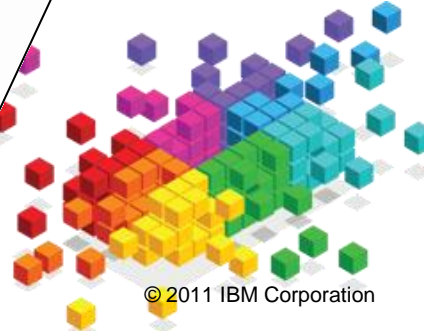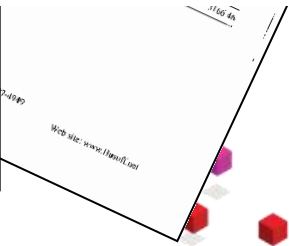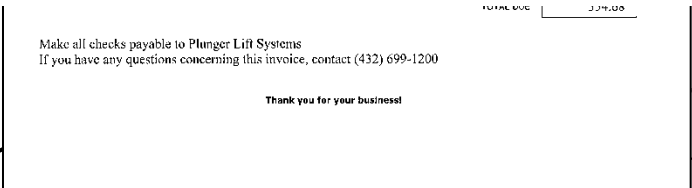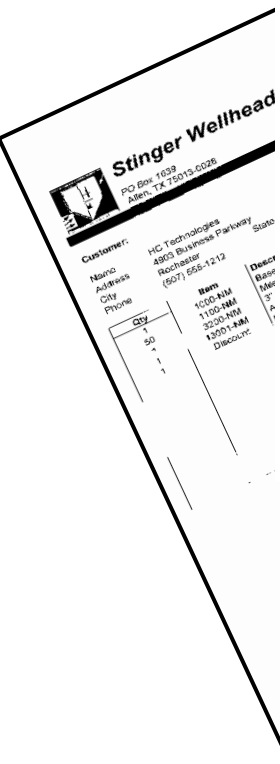
The Datacap Difference – Procedural Rules

- A Datacap rule consists of a series of actions, governed by logic
- An action is a single event for which the answer is true or false – i.e. Is field all numeric?
- A rule executes one action after another until it encounters a false value at which point another rule may be triggered – i.e. if field is NOT all numeric, another rules presents to the verify operator a Failed Validation.
- Procedural rules enable Datacap to deliver effective process orchestration for flexible – and scalable – document capture…
  - Rules logic excels in handling variability in real time
  - Rules empower customers for real time exception handling
  - Rules are not just check box options – they are reusable code snippets…
  - Rules are extensible – easy to create in .NET, VB Script, etc.
  - Rules offer capture capabilities as enterprise services

Datacap Taskmaster's Enterprise Capture Platform
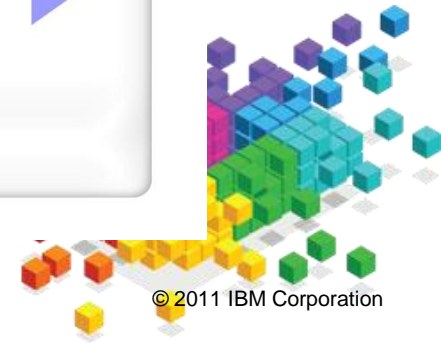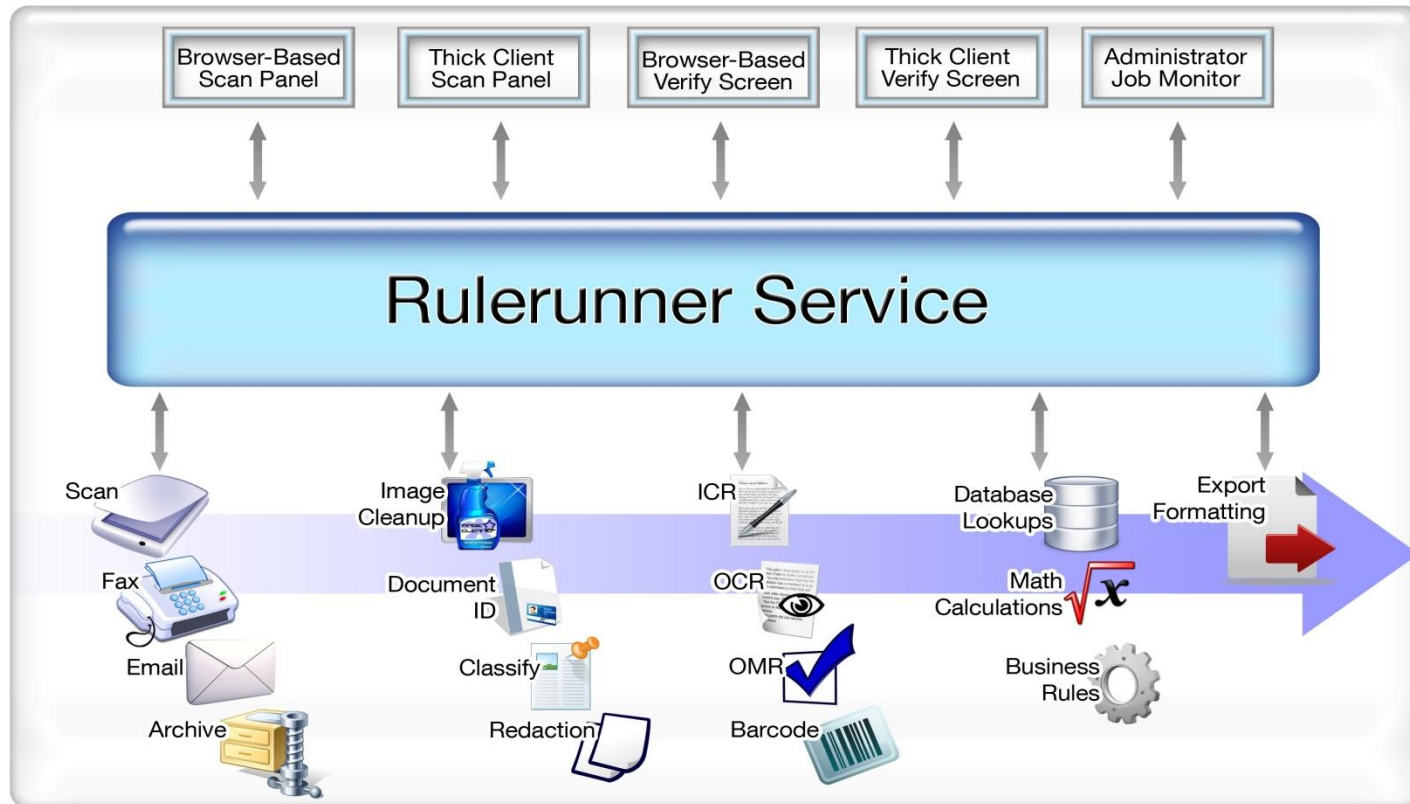
IBM Datacap Taskmaster Capture

- IBM's Strategic Capture Platform

  – Replacing FileNet Capture ADR and supplementing FileNet Capture for existing customers

  – Full integration with IBM CM8 and P8 ECM repositories

  – End to end, enterprise wide capture

- Scanned documents as well as electronic documents

  – emails, email attachments, Office documents converted to images and processed like images

- Integration with existing FileNet Capture implementations

- Connectors for third-party repositories and generic output
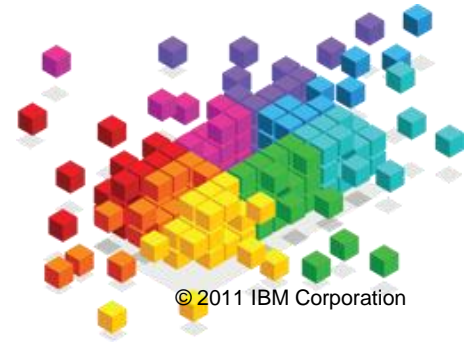
IBM Datacap Taskmaster V 8.0.1 Overview – Now Available

- Production Document Capture
  - Central (corporate or mailroom-type) or full-function
  - Web-based distributed deployments

- Use rules to orchestrate every step in the capture process, including:
  - Input images from multiple sources including scanners, MFDs, folders
  - Process images
    - Image enhancement – deskew, remove lines, noise, rotate, etc.
    - Document classification (including Fingerprints to identify), page separation
    - Extracts data using OCR, ICR, OMR, barcode 1D and 2D technologies
    - Validates data including lookups (and retrievals)
  - Intuitive User Interface for data entry, edit and verification
    - Learns from human operator handling of exceptions
    - Thin and thick clients have same functionality
  - Export to select back-end ECM repositories and ERP/databases

- Administration
  - Web-based monitoring and administration
  - Extensive Reporting

- Simple User-based licensing – no per-click charges
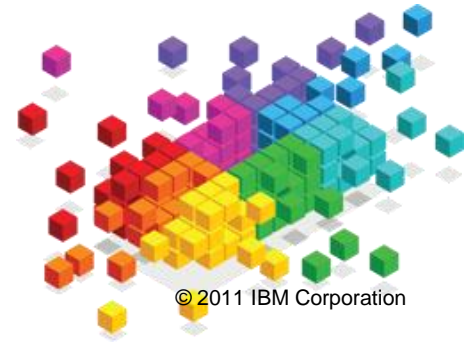
Taskmaster's Distributed Capture Advantages

- Rapid ROI
  - Dramatic reduction in document shipping costs
  - Faster turnaround
  - Reduced hardware & support costs, leverage existing MFDs
  - No software installed on client

- "Virtual Capture"
  - Documents can be scanned in one location and verified at any location
  - Administration from any location
  - Centralized administration further reduces TCO

- Competitive Points
  - Replacement - Existing distributed systems are very hardware dependent
  - Datacap is more advanced
    - Scan, Verify & Administration
    - Advanced UI – Advanced OCR Correction & Fingerprint Creation
    - Reduced bandwidth – Single Pass Verify
    - Licensing Model – Authorized UVU, Employee UVU

- Help drive Advanced Case Management!
  - Capture close to the source
  - Capture at any point in the process using Rulerunner web services
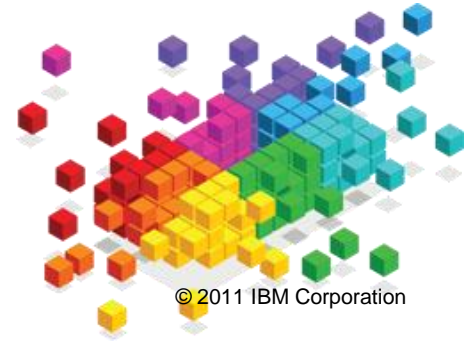
Key Industries / Applications

Government
- Tax Returns
- Property Taxes
- Motor Vehicles / Public Safety

Transportation / Logistics
- Commercial Invoices
- Customs Documents
- Brokers Orders

Healthcare
- EOBs
- Medical Records / EMR / EHR
- Healthcare Claims

Corporate AP (Invoices)
- Header Level Recognition
- Line Item Detail Extraction
- ECM / ERP Integration

Distributed Scanning
- Large # of satellite office
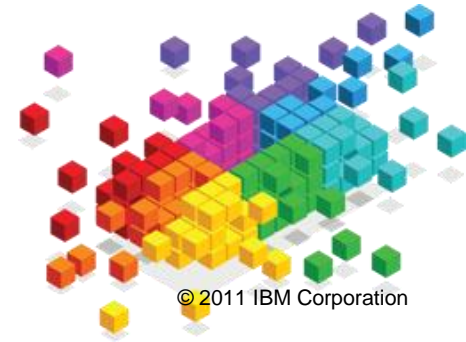- Remote Scan, Process Centrally
- Improved Provisioning

Financial Services & Insurance
- Loan Documents
- Account Service Requests
- Applications & Enrollment

Datacap Taskmaster for Accounts Payable

- Preconfigured application

- Captures, verifies and routes without manual data entry

- Locate and extract data including header and line item detail

- Learns new invoice types from operator

- Accurately captures all line items, even multi-page

- Complex validation rules on dates, math, lookups, data types, etc.

- Look up vendors, add line items, locate line items, calculate missing values

- Aids three-way match with Purchase Order Line item Reconciliation

- Send to operator for handling exceptions
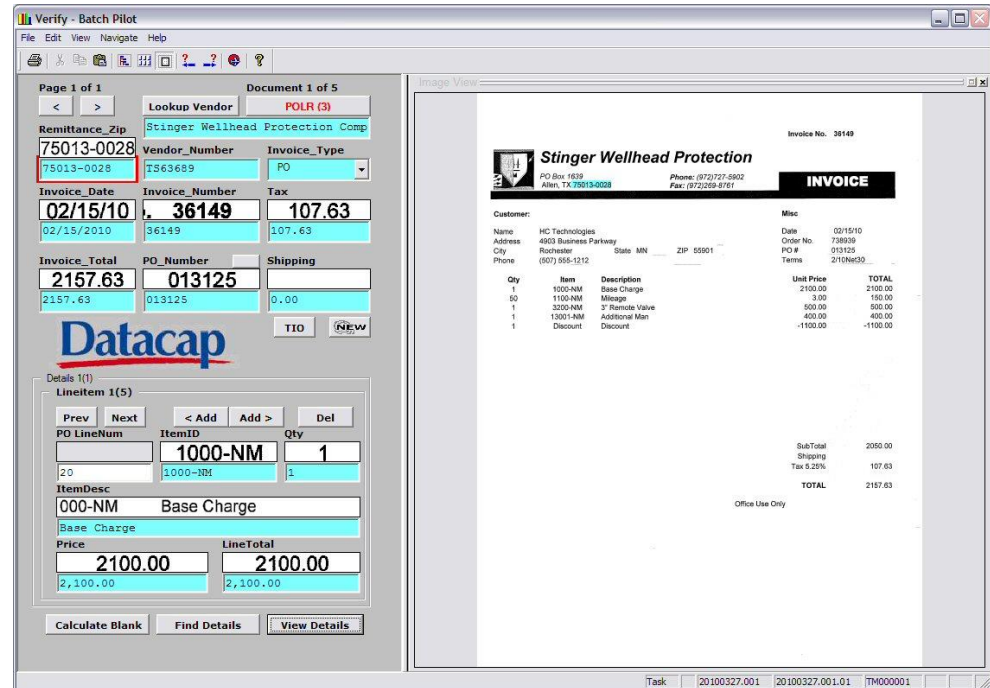
Taskmaster Accounts Payable Capture Advantages

- No preproduction set-up required

- Adapts to new invoice layouts on-the-fly learning the first time

- Single page, multi-page, attachments

- Line item capture out of the box

- POLR – Purchase Order Line Item Reconciliation - streamlines 3 way match downstream

- Thick and thin client architecture and user interfaces

- Fingerprint Service accommodates tens of thousands of vendors

- Pricing model by user - NOT pages/documents scanned or processed

- ROI in 6 – 12 months

- Many years experience in AP automation

- Easily extensible to new document types and add-on applications, i.e. sales orders, remittances, etc.

IBM Datacap Taskmaster Medical Claims Capture

- Pre-packaged application to capture process CMS 1500 medical claim forms and UB-04 Institutional claim forms
  - CMS 1500 is HCFA standard health insurance claim form used by a non-institutional providers or suppliers to bill Medicare carriers and durable medical equipment regional carriers (aka "Professional")
  - UB-04 is the HCFA standard health insurance claim form used by hospitals, nursing homes, hospice, home health care providers (aka "Institutional")
- Thin Web and thick Windows clients
- Claim scanning, verification and application administration and reporting
  - Validations match diagnosis and CPT codes
  - HIPAA compliant 837 EDI output
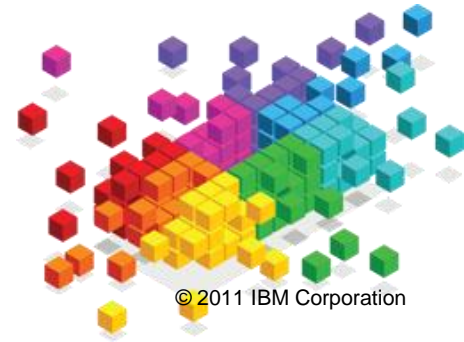 - Chargeable component

IBM Datacap Taskmaster Enterprise Expansion Options

- IBM Datacap Rulerunner Enterprise – enterprise scalability through virtualization

- Connectors for eMail and Electronic Documents

  - Access mail server(s) via Internet Message Access Protocol (IMAP), which is supported by IBM Lotus Domino, Microsoft Exchange Server, Novell GroupWise, and other mail servers

  - Provides ability to convert Microsoft Word, Excel, Outlook, PDF, and multipage TIFF files to single page TIFF files for capture processing

  - Supports extraction of ZIP archives

- Connector for Fax

- Connectors for non-IBM repositories: EMC Documentum, Microsoft SharePoint and OpenText LiveLink
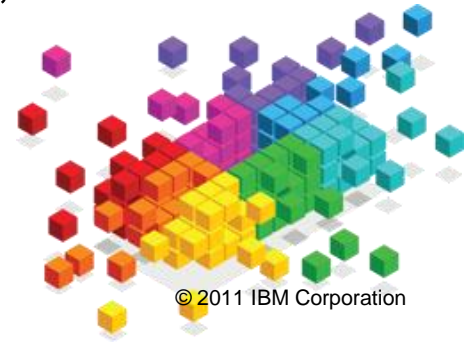
Support for IBM ECM and FileNet Capture

- IBM FileNet P8 and IBM Content Manager (CM8)
- IBM FileNet Image Services:
  - Included support for uploading images and associated index data to FileNet IS Server
  - Tested with Image Services 4.0 and 4.1, FileNet IBM Desktop 4.0 and 4.0.2
  - Uses IBM FileNet IDM Desktop Client
- IBM FileNet Capture Professional:
  - Support for FileNet Capture enables running a Taskmaster Capture application within the FileNet Capture Professional environment
  - Add automation and thin client distributed capture to existing IBM FileNet Capture installations
  - Requires FileNet Capture Professional 5.2.1 or above

What Does this mean for FileNet Capture customers?

Extend your current FileNet Capture – Don't Rip & Replace

- IBM continues to invest in FileNet Capture
  - FileNet Capture release (v 5.2.1) has been announced with significantly expanded features:
    - Mixed version support
    - Improved UNICODE support and language support
    - PDF/A compliance
    - DB/2 support

- With Datacap, do more inside of FileNet Capture
  - Works in your Capture Paths to extend your system
  - ICR/OCR-assisted Forms Processing
    - Advanced text recognition and extraction
    - Document classification and page separation
  - Validation Rules
    - database lookups, checksums, cross-field checks
  - Already integrated to FileNet Capture and used by many FileNet Capture customers.

- Extend your reach
  - In-Process Capture
    - Ad-hoc capture of documents driven by the business process
    - Real time and cost effective way to capture at the source (distributed, front-office, or web)
 - Additional devices – e.g. IBML scanners
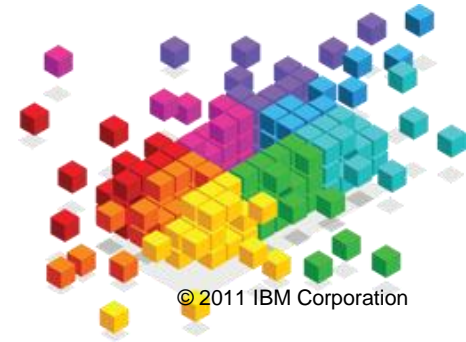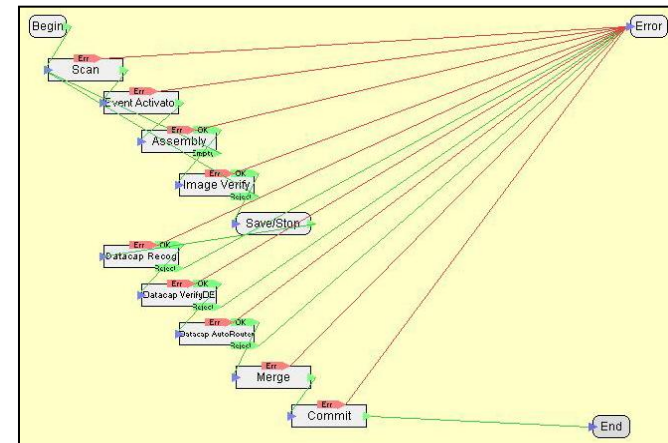
FastDoc Capture Key Features and Functionality

- Capture images from a scanner, multifunction device, or from a local or network folder containing previously scanned images

  - TWAIN driver

- Document separation via visual manipulation, fixed number of pages, separator sheets, or automatic

- Support for multiple document types in a batch

- Automatic document identification using Taskmaster Fingerprint technology

  - Builds library

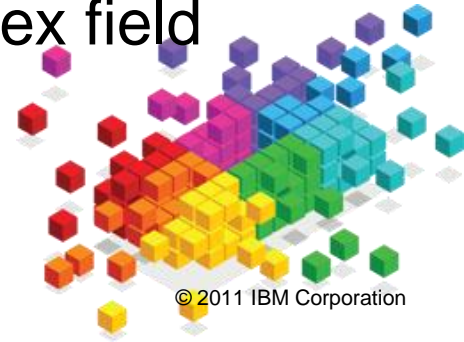- Full page OCR, bar code recognition
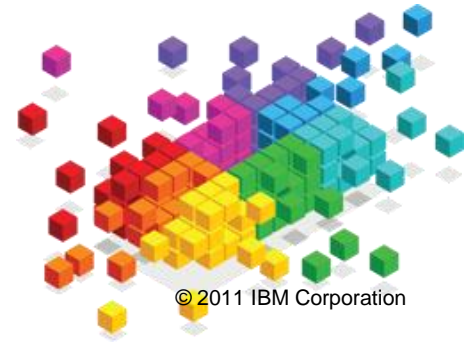
FastDoc Capture Features and Functionality

- Automatic indexing

  - Use image zones in saved Fingerprint

  - Search for a label, go right/left/up/down

  - Match an expression

- Automatic build of verification panel

- Basic validations: required, match data type, min/max length, match lookup

- Database value retrieval

- Pre-built integration to Microsoft SharePoint or generic file export for any ECM, imaging, ERP or other system

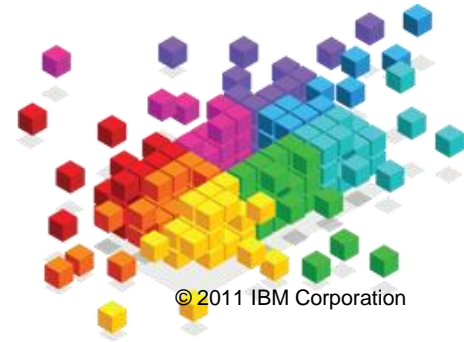 - Output image as PDF (searchable) or TIFF, index data as text or XML
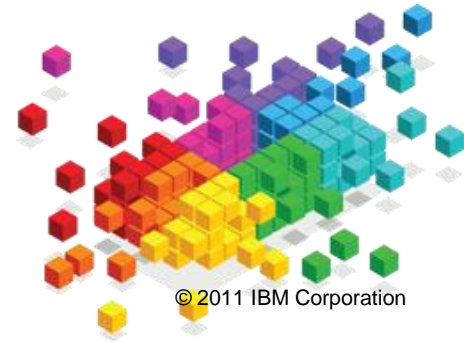
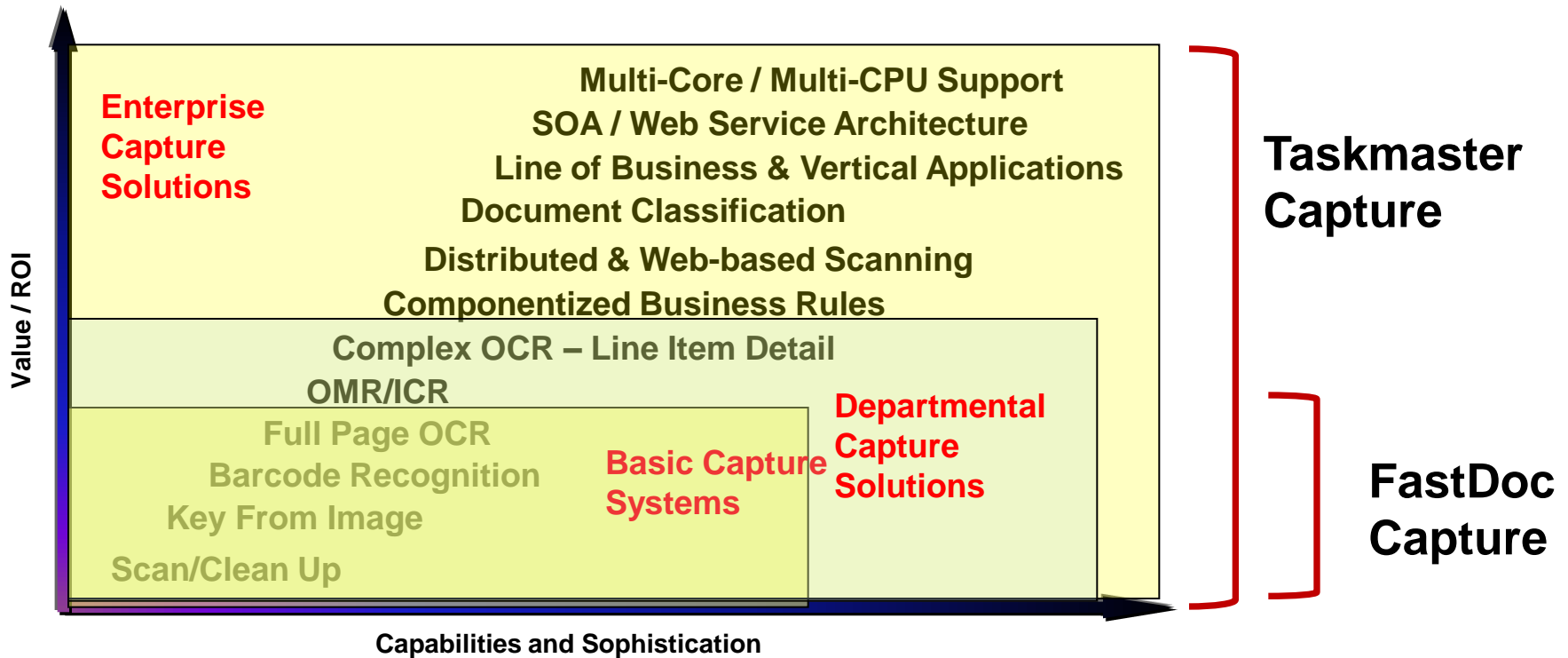# New Features FastDoc Capture 1.2

- Automatic document separation using Fingerprint identification
  - Increases automation
  - Speeds document processing
- Configurable Scan settings
  - Delete images
  - Image enhancement
  - Match Fingerprints
  - Perform recognition
  - Fast Mode
  - Recognition language
- Capture input of PDF and multi-page TIFF
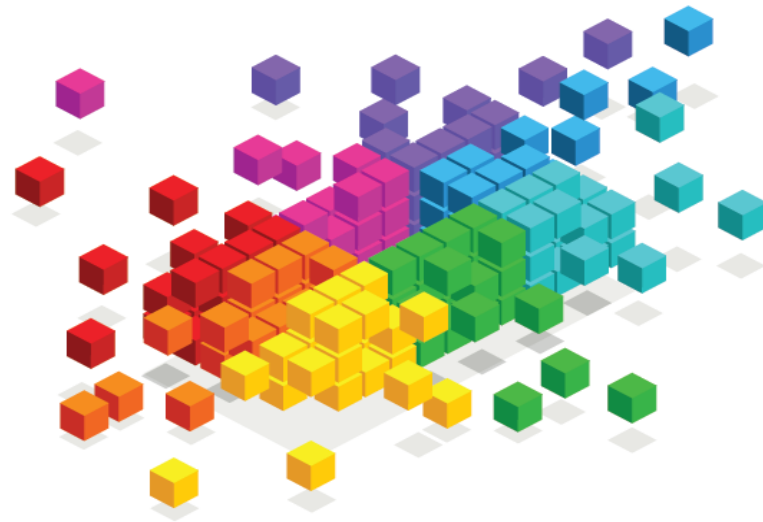- Recognition confidence level controllable by index field

Datacap Capture Capabilities Scale to the Enterprise

Value / ROI

Enterprise Capture Solutions

Multi-Core / Multi-CPU Support
SOA / Web Service Architecture
Line of Business & Vertical Applications
Document Classification
Distributed & Web-based Scanning
Componentized Business Rules
Complex OCR – Line Item Detail
OMR/ICR
Full Page OCR
Barcode Recognition
Key From Image
Scan/Clean Up

Basic Capture Systems

Departmental Capture Solutions

Capabilities and Sophistication

Taskmaster Capture

FastDoc Capture

IBM FileNet Capture 5.2.1 Update
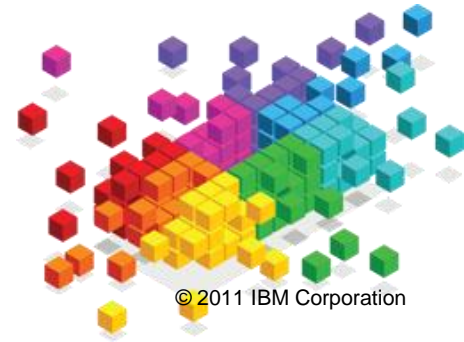
What's New in Capture Professional 5.2.1

- Mixed version support
- Asian language support
- DB2 support
- Dual-streaming support
- PDF/A compliance
- Scan to color PDF and JPEG 2000
- Batch export and import

- Improved logging of errors
- Performance & Configuration enhancements
- Reporting supported in MS SQL 2008
- On-line help improvements
- Currency: toolkits, OSes, DB

FileNet Capture ADR is no longer available, to be replaced with IBM Datacap Taskmaster Capture
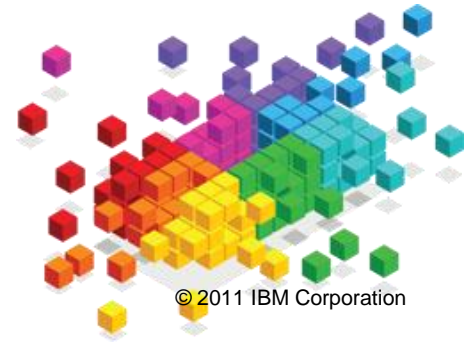
Support of Mixed Versions

- Provides more flexibility and reduce cost of upgrades
  - Ex: upgrade Scan stations for newly-supported driver/OS
  - Ex: upgrade non hardware/OS-dependent stations first
  - Ex: no big bang required; move users/applications, over time

- Components of Capture 5.2 and 5.2.1 can be mixed in a deployment
  - As long as the shared resources are supported by both (see table)

- Batches created by Capture 5.2 can be processed by Capture 5.2.1

- Batches created by Capture 5.2.1 can be processed by Capture 5.2
  - As long as the functionality is supported by 5.2 (see table)
    - 5.2.1 Capture Path cannot include functionality not supported by 5.2
    - 5.2 batch processed with 5.2.1 functionality cannot be sent back to a 5.2 station

- Hardware/software requirements for each version may be different
 (refer to Capture HW/SW Requirement Guide)

Asian Language Support

- Full-text OCR to PDF and Zonal OCR now support
  - Traditional and Simplified Chinese, Korean, Japanese

- Available in base Capture Professional; not as an add-on

- Functionality added to existing technology; same configuration steps

- Supported on TIFF, JPEG, JPEG2000 color, gray-scale and B&W

- Appropriate character set be must selected for all components
  - Client OS, shared or local repository and back-end repository

- 300+ DPI best

- One language at a time

- Latin characters recognized in Asian text
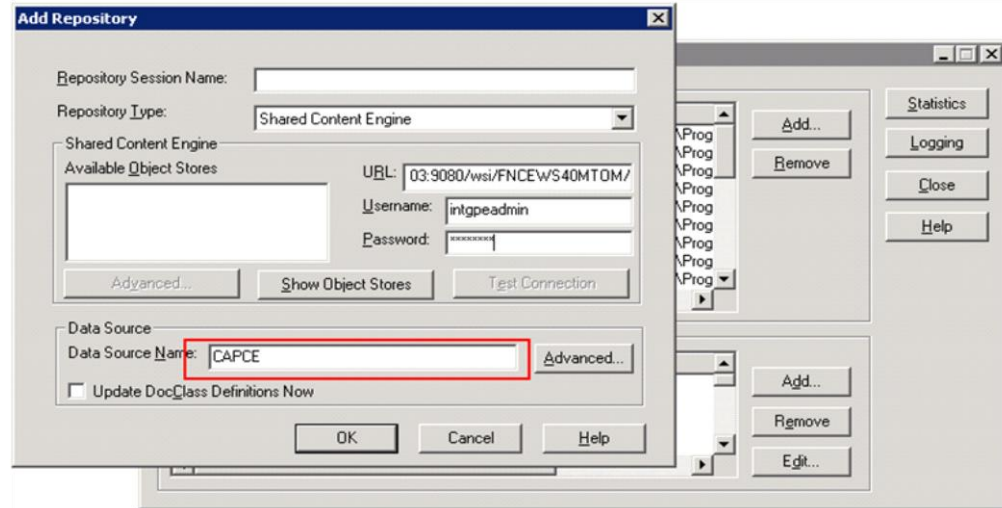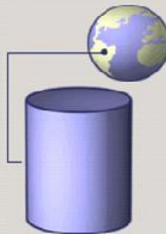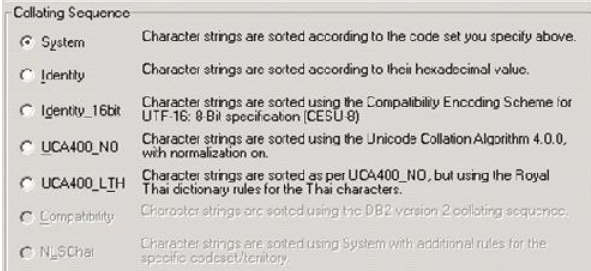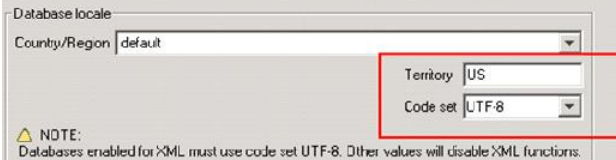
- No spell checking and verification
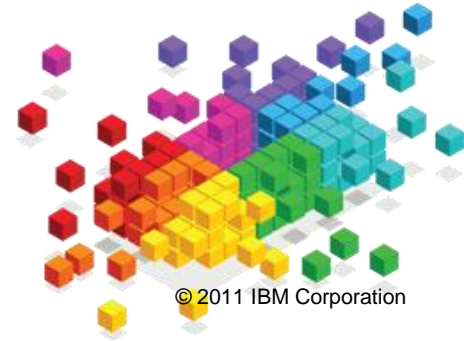
DB2 Support

- IBM DB2 9.7 supported for Shared Repository and Statistics database

- Set up DB2 for the UTF-8 character set

- Configured the same as other RDBMS using ODBC
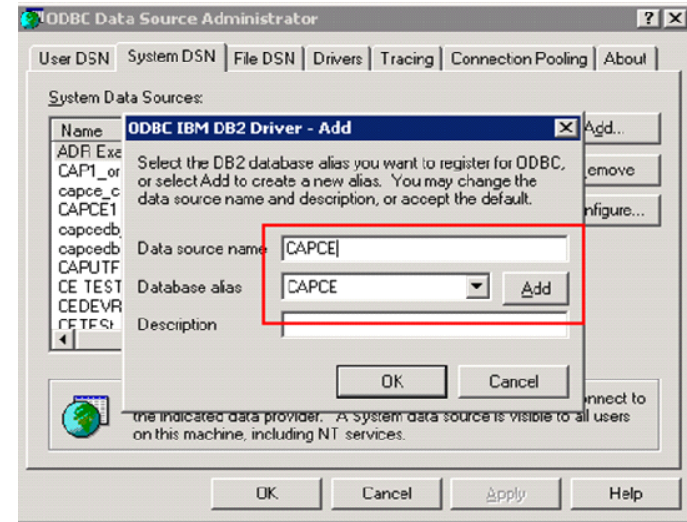
Dual-streaming

- Some scanners can produce 2 images out of a single scanned page

- 2 "cameras"; each set separately
  - Ex: b&w/300 dpi and color/200 dpi

- Pages of each type are scanned to the same batch in paired sequences
  - Ex: 001.TIF, 001.JPG; 002.TIF, 002.JPG, etc.

- Use when needing certain pages in color/gray scale and others in b&w

- Faster than scanning twice same document with different scan settings, or further separating

- Configured in Scan settings

- Depends on scanner functionality and driver

- Batch can then be processed manually or using a custom "worker bee"
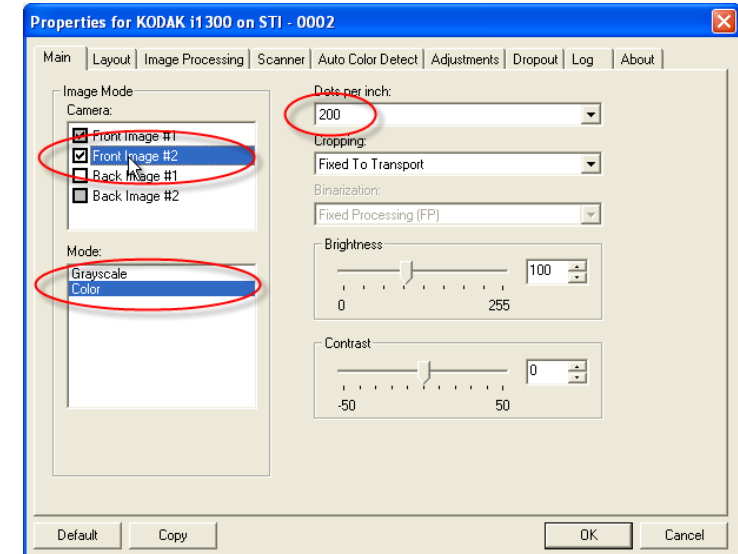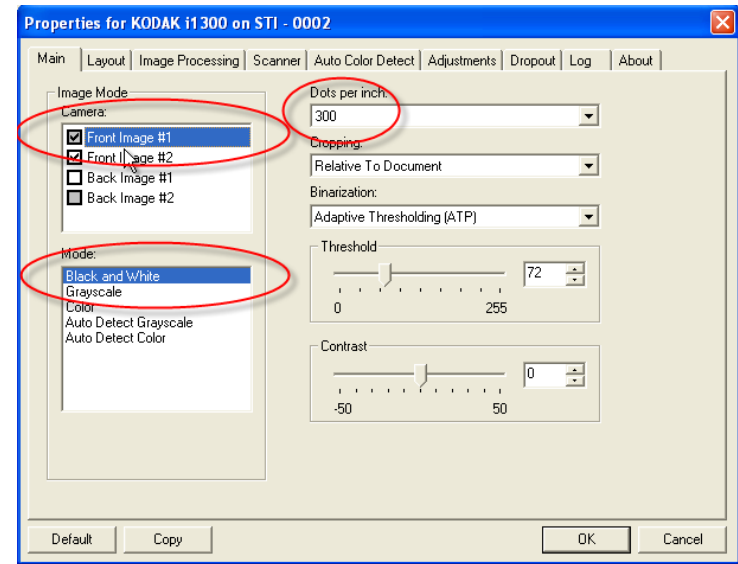
PDF/A Compliance

- Scan and OCR to the PDF/A file format

- Standard (ISO 19005-1) for long-term archiving of electronic documents

- Based on Adobe PDF 1.4 specification

- Preserves visual appearance over time, technology generations and operating environments

- Document self-contained and can be reproduced with fidelity independent of output devices.
    - Embeds color maps (scanned images)
    - Embeds font definitions (OCR)

- Selected in OCR settings

- Displays in Acrobat

WorkspaceView

Local Repository:/MyFolders/Batch 18/Document 1/Document 1.pdf

1 / 1   129%

You are viewing this document in PDF/A mode.

第 3 章 関 連 性

で、我々は関連性を想定がもつひとつの特性として
性をその発話が明示的に表す想定の関連性と同一の
も、聞き手はただ単に発話によって表される想定を
と一般的に言うならば、個人はただ単に自分達の居
だけではない。どちらの場合にも、心的労力を必要
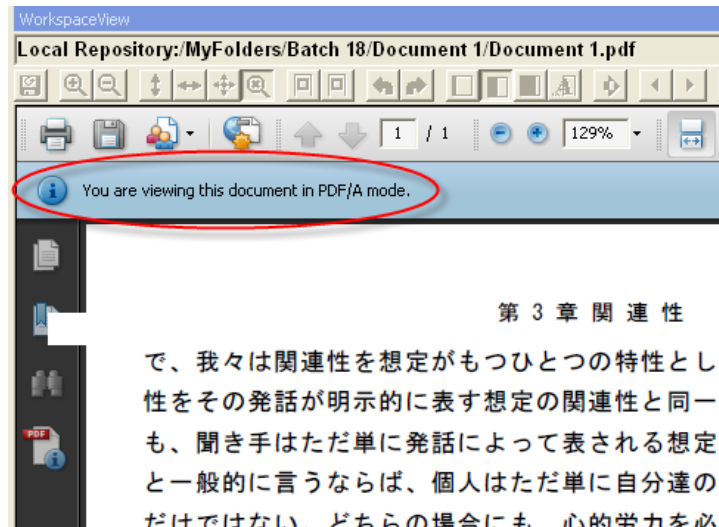
Scan to Color PDF, PDF/A and JPEG 2000

- Capture 5.2.1 supports scanning to
  - Color PDF and PDF/A
  - JPEG2000

- JPEG2000 is ISO standard (ISO/IEC 15444)
  - 20% more compressed than JPEG
  - Lossless compression 20% of uncompressed
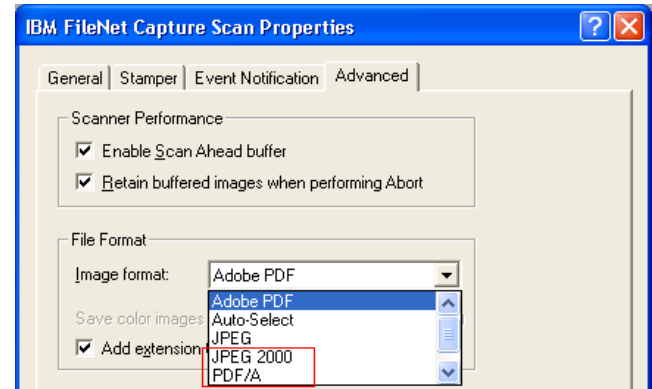 - Better picture fidelity (less artifacts)

IBM FileNet Capture Scan Properties

General | Stamper | Event Notification | Advanced

Scanner Performance
- ☑ Enable Scan Ahead buffer
- ☑ Retain buffered images when performing Abort

File Format

Image format: Adobe PDF
- Adobe PDF
- Auto-Select
- JPEG
- JPEG 2000
- PDF/A

☑ Add extension

Image Encoding Table

File Format		Scanner settings					
		Black&White		Grayscale		Color	
		Bit depth	Compression	Bit depth	Compression	Bit depth	Compression
	Adobe PDF	1	Group4	8	JPEG	24	JPEG
	Auto-Select	1	Group4	8	JPEG	24	JPEG
	JPEG	N/A	N/A	8	JPEG	24	JPEG
	JPEG 2000	N/A	N/A	N/A	N/A	24	JPEG 2000
	TIFF Group3	1	Group3	8	JPEG	24	JPEG
	TIFF Group4 or Color	1	Group4	8	JPEG	24	JPEG
	Auto-Select, w/dual streaming, JPEG	1	Group4	8	JPEG	24	JPEG
	Auto-Select, w/dual streaming, JPEG 2000	1	Group4	N/A	N/A	24	JPEG 2000

Note: JPEG: Sequential JPEG

Putting Content to Work
ECM UserNet 2011

49
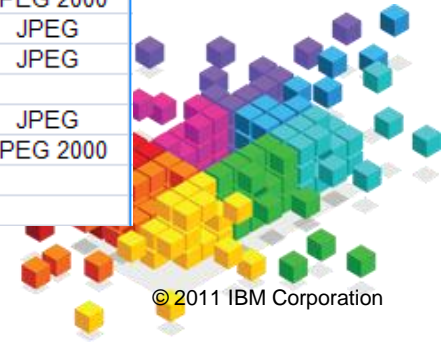
Exporting and Importing Batches

- Two new components: "Batch Exporter" and "Batch Importer"

- Batch Exporter
  - Exports an entire batch to the file system
  - All images written to a directory tree
  - Directory tree named after batch and documents
  - Optionally export XML-encoded attributes

- Batch Importer
  - Imports into Capture an entire batch:
    - Images, documents, and attributes

- Can be used to
  - Generate a large corpus of sample forms to be used with ADR T-Studio
  - Integrate with other applications
  - Commit content to multiple back-end repositories
 - Extend Capture path over the WAN in "decoupled mode"
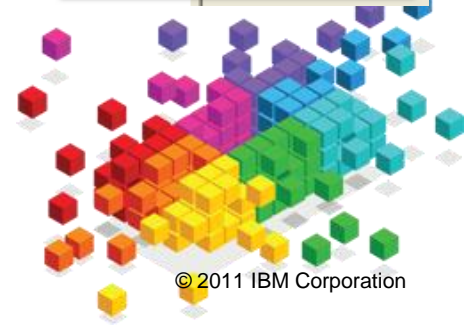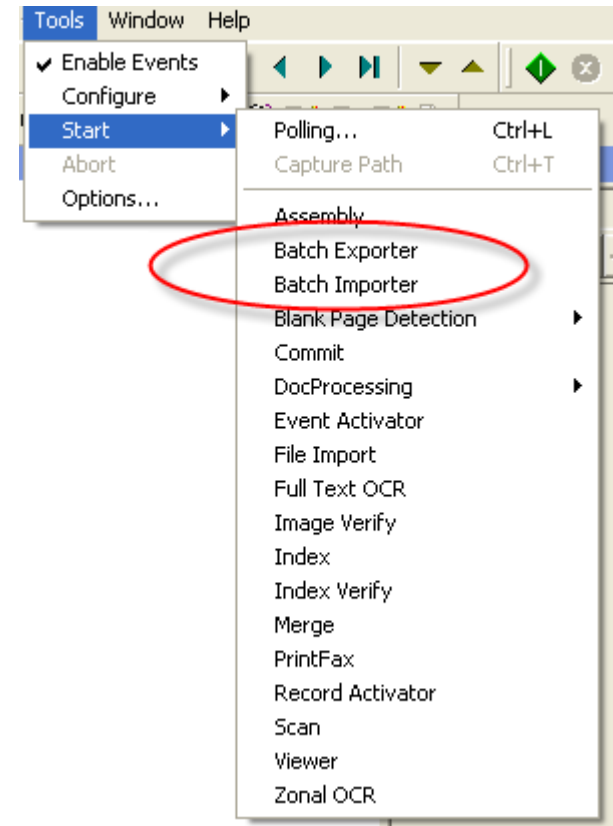
Summary

- Document Capture drives significant return on investment with enterprise standardization on single-vendor, simply priced solutions

- IBM Datacap Taskmaster Capture significantly extends IBM's leadership in Imaging and Capture

- Customers' existing FileNet Capture investments are protected with Datacap and new FileNet Capture release

For more information on Datacap, visit www.datacap.com

Backup Slides
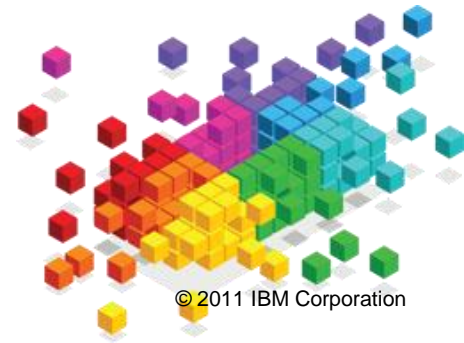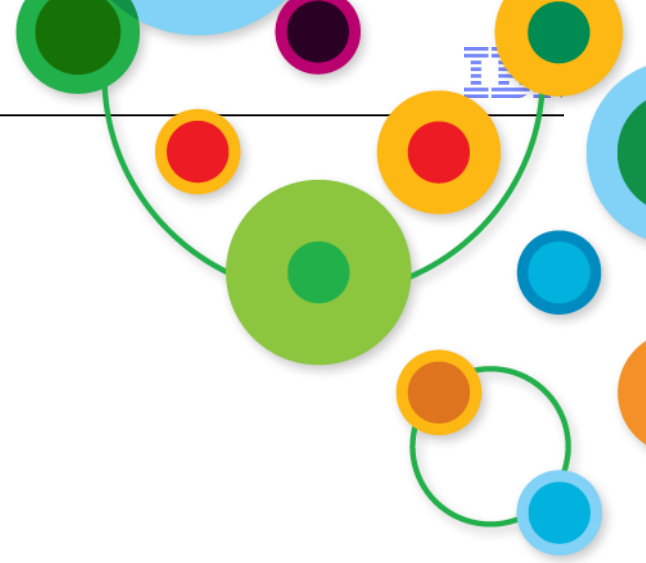
Feature	5.2	5.2.1
Support of Asian Languages for Zonal OCR and Full-text OCR to PDF	No	Yes
PDF/A	No	Yes
Dual-streaming	No	Yes
Scan to color PDF & JPEG 2000	Yes*	Yes
Batch Export & Import	No	Yes
Configuration and Perf. Improvements	No	Yes
Shared repository & statistics on DB2	No	Yes
Shared statistics on SQL 2008	No	Yes
Shared repository & statistics on Oracle 9i.2	Yes	No
Shared repository & statistics on SQL 2000	Yes	No
Records Manager 4.0.0-4.0.1	Yes	No
IDM Desktop 4.0.1 (Image Services)	Yes	No
Doc Processing on 32-bit Windows Vista, 7 and 2008	No	Yes
Scan on 32-bit Windows 7 and 2008	No	Yes
Zonal OCR and Full-text OCR to PDF on 32-bit Windows Vista, 7 and 2008	No	Yes
Report Viewer on 32-bit Windows 7 and 2008	No	Yes
RightFax on 32-bit Windows 7 and 2008	No	Yes

*: 5.2 FP2