

# Unleash the Power of the Digital Mailroom

Automate business correspondence  
to improve efficiency  
and communication

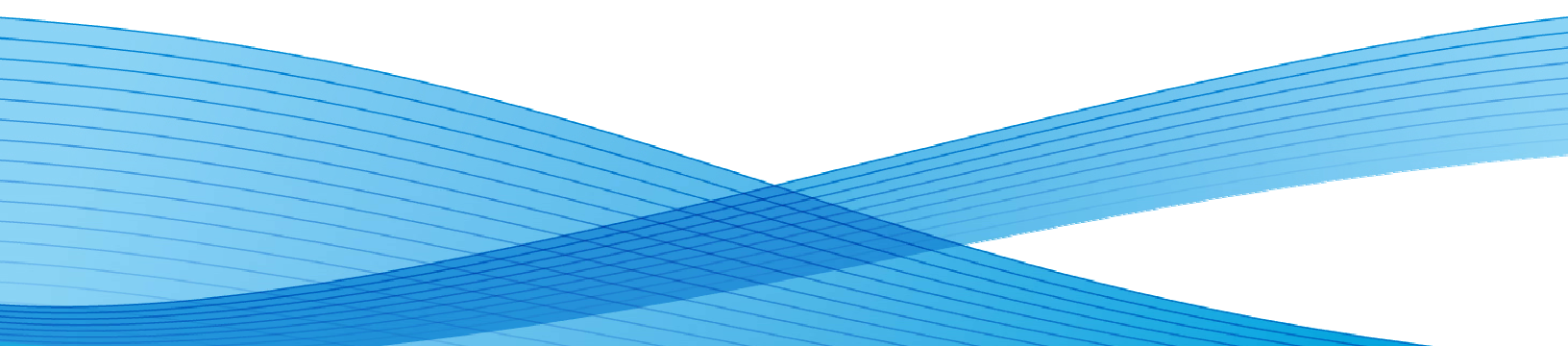
## Table of Contents

- 2 It all starts in the Digital Mailroom
- 3 Smarter Document Technologies
- 3 Classification
- 4 Text Categorization
- 5 Image Categorization
- 5 Wordspotting
- 6 Multimodal Categorization
- 6 Advanced Text Analysis
- 7 Smarter Document Assessment
- 7 Closing the Loop

January, 2008

Francois Ragnet  
Managing Principal,  
Technology Innovation  
Xerox Global Services

Xerox Corporation



# The Power of the Digital Mailroom

## Introduction

Every organization is challenged by a daily flood of business correspondence – mail, faxes, overnight letters, e-mail with attachments, Web forms, etc. These are vital communications with clients, supply chain partners and other stakeholders that flow into, through, and out of your organization. The ability to process this information quickly, accurately, securely, and efficiently is a key to your organization's cash flow, client acquisition, retention and ultimately, its profitability.

In most cases, processing the electronic correspondence is pretty straightforward, provided you have IT systems, policies and procedures in place to route it and protect it appropriately. On the other hand, processing the paper-based correspondence is a labor-intensive, manual screening procedure fraught with errors. Letters are sorted, routed, delivered, opened, and read to determine content, and then forwarded to the appropriate people/departments for the required action.

But what if there was a way to automate the initial handling and screening process and determine the document type (whether new contract, renewal, purchase order, invoice or letter) and extract the important content of each document with minimal human intervention? And then forward the actionable information directly into databases that would self-update and respond with outbound communications based on predetermined rules?

Sounds like science fiction? But it's not. It's automated business correspondence, a breakthrough that turns the mailroom into a hub for integrated paper and digital document workflows.

This paper shows how Xerox Global Services is reengineering the mailroom to reduce client response time and cost, improve communications and quality, while assuring regulatory compliance. We do this through Smarter Document Process and Technologies, a unique collection of information workflow practices and software from Xerox.

Combined with a fine-grained analysis of your documents and business processes, they can be combined and adapted to radically transform some of the most challenging business processes and, in this case, unleash the power of the Digital Mailroom.

## It all starts in the Digital Mailroom

A Digital Mailroom captures specific business critical information from both paper (incoming mail, faxes and other) and electronic sources (e-mails, web form data) and classifies that information according to its format or content determined by preestablished business rules.

Information can be extracted and, based on the business rules, exported to a database, used to start an automated workflow, or produce an outbound correspondence (a confirmation letter or e-mail for example). At the same time it can populate back-office Enterprise Content Management, Enterprise Resource Planning, Customer Relationship Management, Records Management or legacy systems.

# The Power of the Digital Mailroom

The Digital Mailroom becomes the bridge between hard copy and electronic documents, enabling common processes. The benefits are:

- Time: improved processing speed;
- Cost: less process duplication and manual intervention;
- Quality: the right information gets delivered to the right people;
- Regulatory compliance: business records and personally identifiable information is treated appropriately, and
- Security: information is captured and processed at a centralized gateway for greater control.

## Smarter Document Technologies: The core of the Digital Mailroom

To achieve the vision of automated business correspondence you need a unique set of technologies that can scan, extract content, understand context, and integrate with a variety of enterprise management systems. We call them Smarter Document technologies.

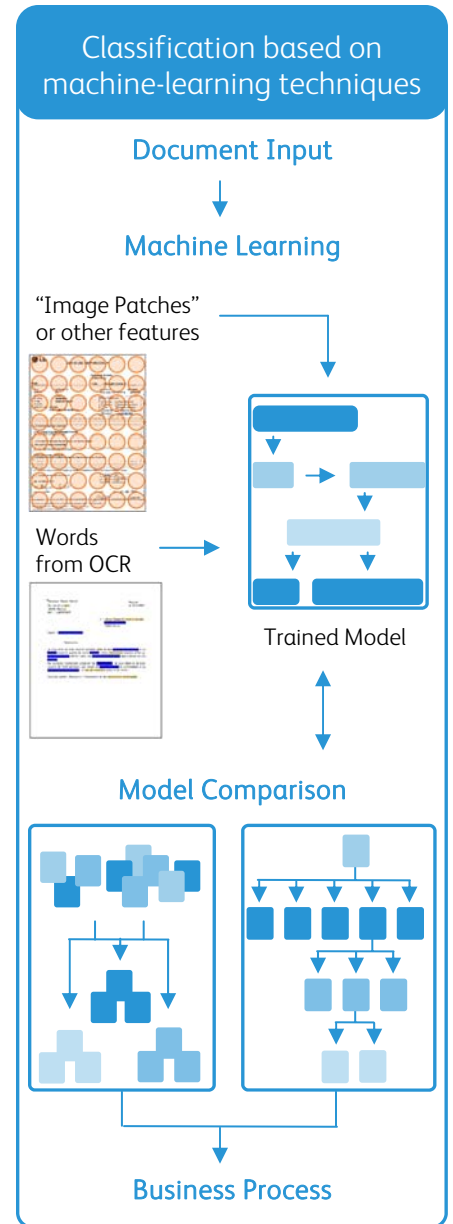
For years, Xerox has focused their research and development efforts on finding new and better ways to work with information. Our approach has been to concentrate our efforts on documents, which we see as inextricably linked to information-related business processes. Documents are the containers for almost all information; they are the essence of all information workflow.

But documents on their own are not that smart. Our goal has been to bring intelligence to the document and the process to make it easier to extract the relevant information. We build business rules to automatically route it to the appropriate people and/or databases. And encrypt it with security protections to make sure that only those who are authorized to access it can do so. In addition, we are innovating in the scanning arena, bringing new capabilities to convert hard-copy information to useful digital information. Going well beyond traditional optical character recognition (OCR), we have broken new ground in ways to identify content and context to make processing paper-based information simpler, faster, and more accurate, and accelerating business processes downstream.

Smarter Document Technologies are uniquely differentiated components coming out of this world-class research. Developed within Xerox's state-of-the-art R&D research labs, they can be customized and combined to meet some of the most advanced client's business processes. Below are some of the technologies that are enabling the new Digital Mailroom workflows.

## Classification

Document Genre Classification is the action of identifying a document type. This is a necessary step in most imaging and mailroom projects and it is also one that is labor and cost intensive. It requires trained staff with subject matter knowledge, decision making abilities, and an understanding of the many business processes that correspondence can impact.



# The Power of the Digital Mailroom

Rule-based systems are another option which rely on a user manually “teaching” the system the most characteristic features of the various document types. Such manual training process is time-consuming, error-prone and inflexible.

Document Genre Classification technologies are designed based on direct observation of work practices in the mailroom.

There is essentially an escalating three-step process to determine the appropriate action:

1. Identify the document type visually at a glance – contract, purchase order, invoice etc.
2. Spot one or two key words – renew, cancel, shipment date, etc.
3. Read the entire document to determine content and context

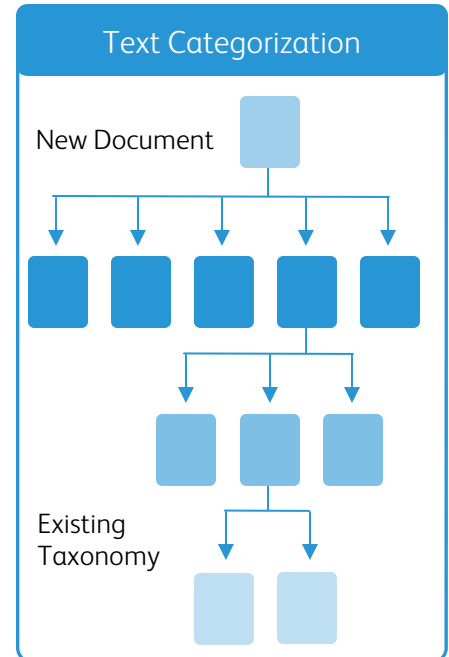
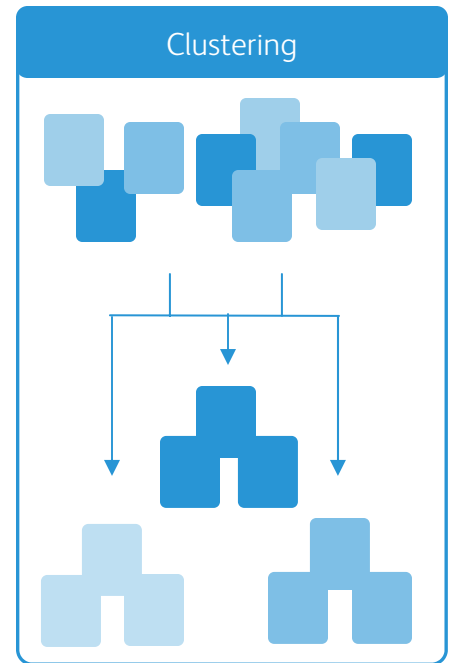
Xerox Smarter Document classification technologies use Advanced Machine Learning techniques to replicate and automate this process with little or no human intervention required: given a set of “training” documents, the system is able to automatically “learn” the features, either visual or textual, which are representative of each document set. Once that model is understood, new incoming documents can be evaluated against this model, and automatically recognized. Now correspondence can be quickly classified into a set of categories or ‘types’ such as new contract, renewal, purchase order, invoice, etc.

These technologies present a number of benefits. First, they are much faster to set up and less subjective than a human, as the system automatically “learns” what characterizes certain sets of documents. Second, they are agile, and can adapt to evolutions in the customer’s business, as new processes are implemented and old ones are retired. Third, their state-of-the-art performance can be adjusted for optimal quality or speed – and can achieve usually a quality above what a human can do, in a fraction of the time.

## Text Categorization

In the next step of the Digital Mailroom workflow, a document is scanned and classified after off-the-shelf OCR has been applied. The software “reads” the actual content of the document, decides how it should be classified and routes it to the right destination – whether an e-mail address or another document management system - automatically.

This classification technology was developed using patented linguistic analysis technologies and machine learning algorithms and involves two essential, tightly integrated, processes: clustering and categorization. Categorization relies on a known list of document types (usually client or process driven), while clustering “learns” new categories by automatically grouping documents on a common topic.



# The Power of the Digital Mailroom

## Image Categorization

Accessing the textual content of a document is not always possible or practical. The Xerox Smarter Document R&D team has developed multiple technologies for image classification, with state-of-the-art performance, which alleviate the need for OCR.

Image Categorization is a general-purpose image classification engine that can adapt to many different image types. This technology is capable of categorizing multiple everyday image content types, including buildings, airplanes, books, faces and documents. Using a combination of expertise in image processing, computer vision, and machine learning, the computer is “trained” to map the key features, including the subject, texture, structure, layout and colors of an object.

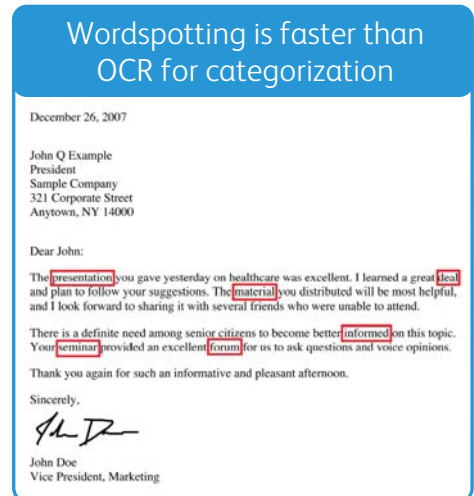
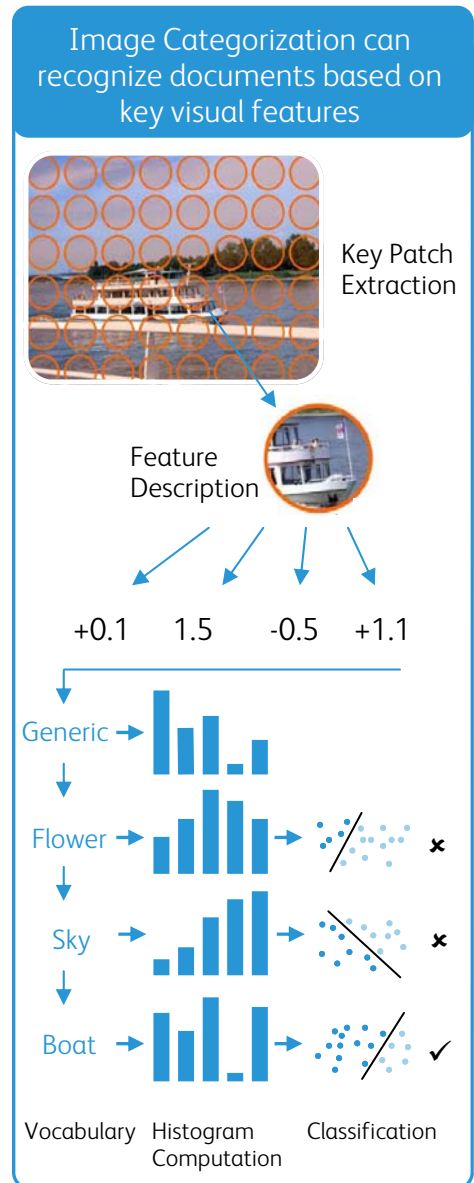
The system has proven state-of-the-art performance in a large number of scenarios such as consumer photography annotation, document object recognition, etc. The technology is accurate, fast and can effectively cope with hundreds of categories and the variations usually encountered in imaging scenarios.

Classifying documents is a very simple task for software that can recognize photos. Image Categorization can indeed recognize specific structured document types thus avoiding OCR altogether, or limiting the need for it. Similar to an operator taking a quick look at the document image and recognizing a well-known document type at a glance, Image Categorization can, in a split second, recognize a number of pre-defined document types based on their structure and form with very high confidence, and route them to the right business process. Even better, it can also recognize “visual classes” of semi-structured documents to guide post-processing. For example, it can recognize typewritten letters from handwritten letters, or guide the OCR to the zones where the relevant information is located in a known form.

## Wordspotting

Wordspotting is an image-based categorization technology that scans through documents looking for a few specific words. This is at only a fraction of the cost and time that OCR and text categorization requires. These words can be sufficient indication of a category, e.g. change in address, change in bank account, or contract termination. But most importantly the technology can be used to spot documents that need to be processed with high urgency (where key words to spot might be “cancel,”) or other indicators of client dissatisfaction so that those documents that impact client relationships are handled quickly.

Currently available for typed documents, this technology will also be available soon for handwritten documents, a breakthrough that will enable an entirely new flow of information to be handled automatically by the Digital Mailroom.



# The Power of the Digital Mailroom

## Multimodal Categorization

In the Digital Mailroom, Smarter Document image categorization technologies complement text categorization technologies in a very powerful way. Multimodal Categorization combines the results of multiple categorization engines (image, full-text or even single-word based) for further confidence in the results.

Let's take an example of an insurance claim. A dossier, consisting of multiple documents, is received by an insurance company, and might contain a claim letter and photos of the damage. Even this simple case is a challenge for textual or image-based classification alone. Indeed, image recognition might recognize a letter and a photo, but this combination can occur in many different document types. Text categorization alone might be able to tell from the words that it is related to a claim, but will not be able to tell whether it's a house incident, a car accident, or something else.

Multimodal categorization will be able to combine the detection of a typed letter, the extraction of its characteristic vocabulary (e.g. characteristic of an insurance claim), with the presence of a photo, or even better, a type of photo (e.g. a wrecked car) to uniquely identify the dossier as an automobile claim.

## Advanced Text Analysis

Xerox technology can find simple index fields in a document, such as telephone, customer ID, Social Security/National Insurance number or other such simple patterns that are easily recognizable or located in a known region. That's the easy stuff.

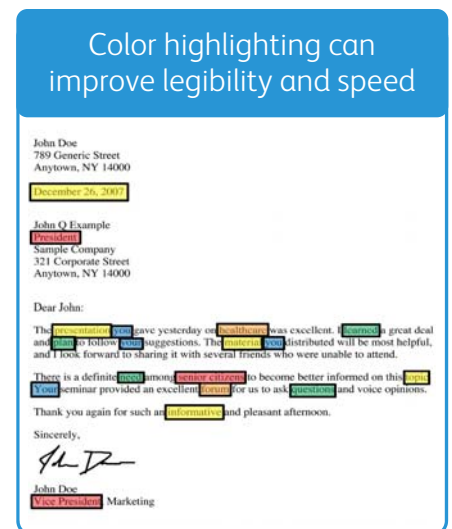
Some Smarter Document technologies can dive deep into the content and context of the correspondence. Data Extraction is the action of extracting additional information to index the document. These metadata fields are "parameters" required for the rest of the business process. As such, they typically depend on the document type, and are often extracted manually.

Advanced text analysis recognizes more advanced elements or concepts such as dates, addresses, company and person names, locations or organizations.

It becomes extremely powerful analyzing documents to detect facts and other more advanced relationships, such as temporal relationships (before, after a certain date), and other types of relations ("Company Y earned Z billions of Euros"). Co-references can even be resolved throughout a document – for example, advanced text analysis can identify people named in a document even when there are multiple iterations of the name, including personal pronouns – John Smith, Mr. Smith, he, his, etc.).

In the Digital Mailroom, advanced text analysis can enable the level of understanding of the correspondence required to be able to fully automate a specific business process. Some fields are common to most business processes, for example, name, address, town and ZIP code. These are easily recognizable and can be easily extended with customizable rules for specific metadata fields, such as client or phone number. This information can be cross-checked with a client database for further validation.

Some of these fields are dependent on the actual process and document type. For example, the type of contract and requested date for termination are



# The Power of the Digital Mailroom

important to qualify a cancellation. Or, a bank account number is required for a bank account change. Advanced text analysis can also extract this raw data from the document with good precision.

Once this information is collected, the business process can be executed automatically if all information is available and confidence is deemed high enough; otherwise, these fields can be presented to the operator for a much quicker rekeying of the information, assisted by color coded highlighting which draws important, predetermined data fields in the document.

## Smarter Document Assessment

All of these Digital Mailroom workflows involve technologies, but before they can be optimized inside of an actual mailroom environment, you have to comprehend all of the business processes they impact. It requires an understanding of the organization, direct observation of the people and their work practices, and a commitment to continual improvement. Only then can you reengineer the processes, write the business rules and integrate the technologies to a truly powerful Digital Mailroom –one that can be refined over time. This is typically done using well-established quality techniques such as Lean Six Sigma to identify the pain points, bottlenecks, and areas of opportunity.

Here's how it works:

Drilling down through multiple layers of processes enables the identification of the biggest areas of automation. For this example, out of over 130 categories used in a typical mailroom process, the top 3 categories represent over 60% of the overall correspondence (and over 90% of the registered mail), while the top 15 represent over 90%.

This fine-grained analysis lets our specialists understand your document processes inside and out, learn how to recognize the “intent” of each business process, and the key information that drives it. Then we can tailor our Smarter Document Technologies to reengineer your processes and look for ways to continually improve them.

## Closing the Loop: Outbound Communications

The Digital Mailroom is the hub for processing incoming correspondence. But just as important is the ability to communicate back to the sender to confirm receipt of the information, and provide logical next steps in the relationship.

Imagine a client renewing their subscription to a magazine. The card comes back in hard copy, is scanned and categorized and information such as client name and address, account number, subscription start and end dates is extracted. The client has extended the subscription for one year. However, the magazine wants to lock the client up for a much longer period of time and has very attractive discounts on three- and five-year subscriptions.

Wouldn't it make sense then to have a CRM database programmed to generate a letter or postcard confirming the subscription renewal, but also cross-selling new services along with your renewal? By automatically combining transactional information with promotional messages, you can

### Activities on the Digital Mailroom Site (Percent of Correspondence)

25%	Contract Information
15%	Takeover
10%	Preparation
9%	Indexing
6%	“Bons Portage”
4%	Prepaid Contracts
4%	Exception Handling
3%	Change in Address or Bank
3%	Invoice Copy
2%	Scanning

### Identification of Key Documents for Specific Business Purposes



# The Power of the Digital Mailroom

create a more effective client communication, one that can drive revenue growth.

For years, businesses have sent out invoices and statements on a regular basis, and then sent out marketing materials on a “campaign basis.” Marketers realized the value of these “must read” documents and oftentimes enclosed generic marketing inserts in the mailing, but results were generally disappointing. Today, with full-color variable information text and graphics, high-impact personalized marketing offers can be added right onto the statement or invoice. These offers can be based on purchase history, known client preferences, strategic marketing initiatives, etc.

These trans-promo materials are part of the ongoing dialogue that you want to maintain with your clients and a key to building stronger and lasting client relationships.

Using powerful Digital Mailroom technologies at the front end of this dialogue allows a much finer integration into your business processes – allowing you to turn the daily flood of business correspondence into a source of client insight and a competitive advantage.

## Conclusion

Ninety percent of client communications is delivered through documents. Each day vital business correspondence flows into, through and out of your organization. It consumes an inordinate amount of time and expense. The risks associated with mishandled and lost communications can impact cash flow, client loyalty, compliance and litigation.

With so much at stake, perhaps it is time to take another look at a department that is often overlooked or ignored, the “mailroom.”

At Xerox, we believe that there is a tremendous opportunity to improve, automate, and accelerate the flow of information through the use of Smarter Document Technologies. The creation of a Digital Mailroom to process incoming correspondence, capture vital information and deliver it to the appropriate people and processes is the first step to achieving a vision of automated business correspondence.

It should be clear however, that the Digital Mailroom is not an “off the shelf” solution. You won’t find it in a box at your local software store. It is highly specialized Smarter Document Process and Technologies, combined with business process expertise applied to solve the unique challenges of very specific business requirements, such as yours.

# The Power of the Digital Mailroom

## About the Author, Francois Ragnet

Francois Ragnet is a passionate man, challenged by the gap between what is and what could be. At Xerox Global Services, Francois knows where things are heading. When he talks about, "The Future of Documents," he conjures a picture of the time when smart documents will energize companies as they move across the enterprise and perfectly integrated with all business processes. He foresees when 3D visualizations of complex information will be commonplace. And when office productivity will achieve new heights. To Francois, the future is now.



## Expertise

- Document intelligence and document management
- Novel technology transfer
- Innovative and wireless technologies
- Technical Evangelist

## Background

- 10+ years industry experience
- Program Manager, Xerox Research Center - Grenoble, France
- Researcher at the National Institute of Standards and Technologies - Gaithersburg, MD

Visit [www.xerox.com/thoughtleadership\\_Ragnet](http://www.xerox.com/thoughtleadership_Ragnet) for more information about Francois and to find the following:

Full Biography PDF

Ask The Expert

Request A Speaker

Articles and Links

Downloads

# The Power of the Digital Mailroom

## Xerox Global Services

Enterprise Document Services<sup>SM</sup> is our strategic approach to transform your documents to help reduce costs, achieve operational excellence and grow revenue. Through our unique combination of experience, resources and technology, we help integrate and optimize your office environment, high-volume print production and business processes. And it all begins by working with a proven partner.

We employ Smarter Document Management<sup>SM</sup> technologies to add intelligence and structure to both paper and digital documents and activate the content they contain. And we combine our extensive industry expertise with tools like Lean Six Sigma to create a powerful portfolio of services that adds real value to enterprises worldwide.

Xerox Global Services (XGS) delivers measurable results for companies in industries from aerospace and the financial services to healthcare, government and retail by looking at business challenges in a whole new way.

**For more information, visit [www.xerox.com/globalservices](http://www.xerox.com/globalservices) or [www.xerox.com/thoughtleadership](http://www.xerox.com/thoughtleadership) today.**