

Paper to Pixel: Understanding the Different Types of Document Digitization for Federal Government Agencies

The United States Federal government is undergoing a digital transformation, and the National Archives and Records Administration (NARA) issued two crucial memorandums, [M-19-21](#) and its update [M-23-07](#), directing federal agencies to digitize all paper and analog records. This is an effort to modernize government operations and eliminate additional manual resources, management attention, and security investments.

In light of this, **Federal Agencies have been looking for an optimal digitization solution that meets their organization's unique needs.**

Now, since digitization is a broad spectrum, agencies often have difficulty in evaluating the digitization needs that suit them most. We are here to help, and this blog identifies what digitization is, what the key types of digitization are, the benefits of digitization, and how agencies can address their digitization needs.

What is Document Digitization?

Document digitization is simply the process of converting physical documents or analog files into digital files that can be stored, accessed, and managed electronically.

Types of Document Digitization

Digitization can take place at different levels of complexity. Let's discuss 3 types:

1. Basic Document Digitization: Scanning

2. Enhanced Document Digitization: Optical Character Recognition

3. Advanced Document Digitization: Intelligent Document Recognition

The most basic is Scanning documents to create a digital image of a document. Others include Optical Character Recognition, which is a more advanced form of digitization, making the document searchable and editable. The most advanced type of digitization includes using machine learning and artificial intelligence to understand the content and context of the documents.

Let's discuss each of these types briefly, which would make it easier for Federal Agencies to understand what complexity of digitization is required for their unique needs.

1. Basic Document Digitization: Scanning

This is a straightforward method of digitization. A scanner is used to create a digital document image, and the file is typically stored in a PDF, TIFF, or JPEG format. However, there are some key considerations in the scanning process. To perform effective digitization through scanning, organizations should put in place robust pre-scanning steps and procedures. This involves properly sorting and organizing the physical documents based on different parameters, ensuring compliance with regulatory standards for scanning. According to NARA guidelines, there are digitizing specifications for each type of document such as Textual documents, Photographs, Maps and Plans, Graphic Records, and comments on derivative files. These specifications include resolution and image size, file format, Aimpoint variability, tonal range, scanner calibration, etc. Files must meet all the specifications to support archival preservation and continued use. Below are a few specifications from NARA to guide agencies through basic digitization.

Basic Document Digitization: Scanning

NARA Digital Imaging Requirements for Text Documents

	Specification
Pixel Depth	8 bit grayscale or 24 bit color ; RGB mode
Resolution	Minimum 300 dpi ; recommended 600 dpi
Hardware Standard	Minimum 1000 pg/hour throughput
Skew Standard	3 degrees or less
Speckle standard	5% or less
QA	Conform to a formal agency-level plan
Scanner Calibration	Kodak grey scale
File Format	TIFF (recommended)

 **EIGENNET** | © 2023 Eigennet, LLC. All Rights Reserved. | eigennet.com | doc2data.ai

2. Enhanced Document Digitization: Optical Character Recognition

Enhanced document digitization goes beyond basic digitization by incorporating additional processing and functionalities to make the documents more interactive. Enhanced digitization often involves OCR to extract text from scanned documents, making the content searchable and editable. This type of digitization focuses on capturing the textual content of documents. OCR is used to convert the initially scanned images into machine-readable text. The resulting digital files are typically in formats like PDF, DOC, or TXT. Furthermore, text-based digitization enables document searchability, indexing (the process of assigning keywords to documents to facilitate storage, search, and retrieval), and editing of the document's content. Enhanced document digitization allows for **text search, annotations, hyperlink integration, and basic workflow automation.**

3. Advanced Document Digitization: Intelligent Document Recognition (IDR)

Intelligent document recognition leverages Artificial Intelligence (AI) and Machine Learning techniques to extract, analyze, and interpret the content of documents. The primary differentiating factor that sets intelligent document digitization apart is its ability to focus on capturing not only the visual representation of the document but also the underlying data and meaning. Intelligent document recognition combines advanced OCR, natural language processing, and machine learning algorithms to extract information, identify the various entities (named entity recognition) and categorize the content. For example, an IDR system can distinguish between invoices, forms, letters, contracts, etc. Identifying and extracting specific information from a document becomes easier, such as extracting an invoice number, date, vendor name, total amount, etc. This extracted data can then easily populate a database or be fed into an Enterprise Resource Planning (ERP) system that makes further automation possible.

In addition to the above factors, federal agencies use intelligent document digitization to make their documents compliant with a wide range of regulatory standards. There is a multitude of compliance categories, such as:

- Data Privacy Compliances like The Health Insurance Portability and Accountability Act (HIPAA) and Family Educational Rights and Privacy Act (FERPA)
- Accessibility Compliance such as Section 508
- Security Compliances like The Federal Information Security Management Act (FISMA)
- Preservation Compliance such as The National Archives and Records Administration (NARA)
- Other compliances such as Personally Identifiable Information (PII) compliance according to the Privacy Act of 1974

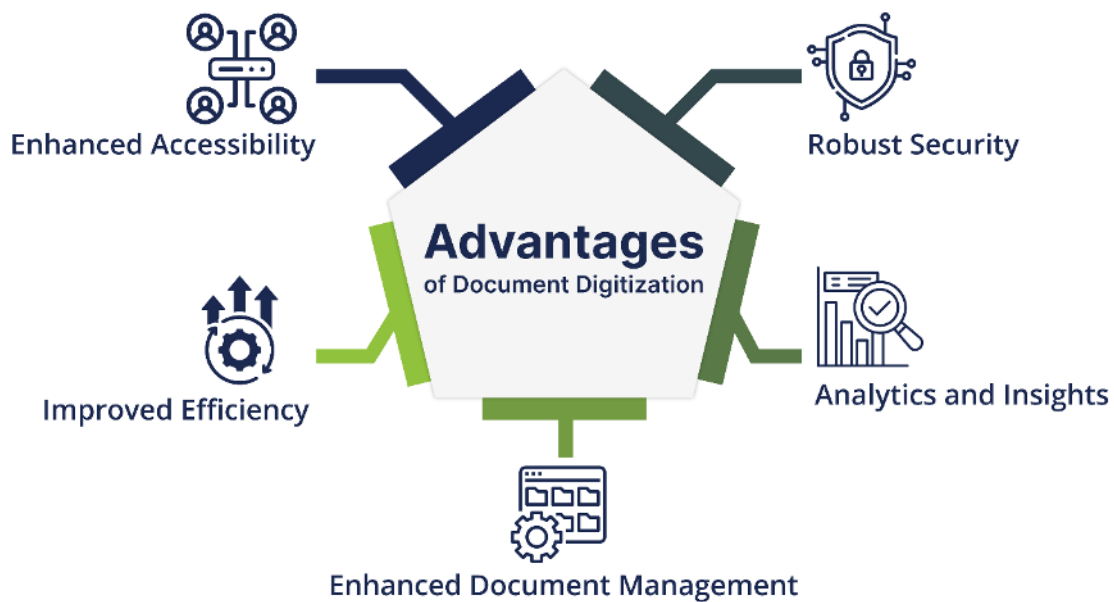
Intelligent Document Recognition paves the way for configuring any of these compliance standards as required by agencies.

Furthermore, intelligent document processing allows agencies and organizations to capture the metadata present in the documents which includes attributes like document title, author, date, keywords, and other descriptive details. Extracting the text within documents and metadata using intelligent document digitization provides efficient organization, categorization, and search and retrieval of documents.

It is easy for agencies to get overwhelmed by the web of digitization and compliance challenges. The primary issue to address is this: How can agencies ensure full compliance? With Eigennet's presence, agencies can be confident as we facilitate simplifying compliance challenges and guide organizations to choose the right digitization type that caters to their specific needs.

Advantages of Document Digitization

Digitizing documents offers numerous benefits to federal agencies and organizations dealing with a large volume of documents, contributing to their growth and operational efficiency. Let's look at some of the advantages.



Enhanced Accessibility: Federal Agencies dealing with historical records can easily transfer their digitized records to NARA and make them accessible through platforms like National Archives Catalog to relevant stakeholders.

Improved Efficiency: By streamlining workflows and facilitating easy access to documents, federal agencies can enhance collaboration and information sharing leading to improved efficiency and productivity. For example: For an agency like the Internal Revenue Service (IRS) digitizing tax records makes it easy for the IRS stakeholders to collaborate enabling the processing and managing of tax data more effectively.

Enhanced Records Management: With document digitization, agencies can efficiently organize, index, store, and access federal records. Digitized records can be made searchable and tagged with metadata facilitating enhanced records management. With proper digitization federal agencies can properly organize and manage their records in a centralized records repository paving the way for enhanced records organization and searchability.

Analytics and Insights: By digitizing documents, organizations can quickly extract valuable insights from the information contained within the documents. Patterns, trends, and correlations can be identified, enabling data-driven decision-making and optimization of operations. Federal agencies and organizations dealing with public records can analyze the digitized files to identify patterns and trends to make informed policymaking decisions.

Improved Security: With document digitization, sensitive documents can be encrypted, ensuring their confidentiality. Access controls can also be implemented, preventing individuals from accessing sensitive information. This heightened security helps federal agencies dealing with sensitive documents to safeguard their data and maintain compliance with regulations.

In conclusion, based on the organizational need, stakeholders can opt for different kinds of digitization techniques that will best suit the types of documents they possess. While some agencies and organizations start with basic document digitization i.e., converting physical documents to images, they will need to implement advanced digitization for full-fledged intelligent document processing to make their documents compliant or to gain advanced data-driven insights from the documents. In light of the government's push towards digitization, it is imperative for agencies to choose the right kind of digitization for their current and future needs.

Addressing Tailored Digitization Needs: Eigennet's Solution

Whether your agency seeks basic document digitization, enhanced digitization, or advanced digitization, Eigennet has tailored offerings that cater to your unique requirements.

When it comes to Basic document digitization, Eigennet goes above and beyond to provide the right hardware, expert personnel, and regulatory assistance to agencies.

If agencies deem Basic document digitization insufficient and foresee the need for enhanced or advanced digitization, Eigennet has just the right framework and software tool for it. [Doc2Data.ai](#) is Eigennet's proprietary tool that offers an unmatched automated document processing experience, transforming files into searchable and scalable data assets. Using state-of-the-art tools, Artificial Intelligence, and Intelligence Automation, it can seamlessly extract, convert, and systematize critical data from diverse documents, images, and formats. Customizability and Automated

processing are Doc2Data.ai's biggest strengths.

Eigennet has successfully completed projects and provided digitization assistance to reputed federal agencies such as the Library of Congress (LoC), the Internal Revenue Service (IRS), Heritage Abroad, and multiple other private entities. We take pride and gain confidence from the positive feedback and testimonials received from them. Take the step towards digitization with Eigennet as your strategic partner, and you can see the difference with enhanced data accessibility, scalable processing, and increased productivity- all backed by Eigennet's motto: Engineering the Mission with Innovation.