
Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

*Issued by the EPA Chief Information Officer,
Pursuant to Delegation 1-19, dated 07/07/2005*

Digitization (Scanning) Procedure

1. PURPOSE

To establish EPA's approach for creating digitized versions of Agency paper, microfilm and/or microfiche documents, including: 1) criteria for determining whether agency documents and records in paper and microform format are good candidates for digitization and 2) steps for digitizing those documents and records.

2. SCOPE

This procedure covers digitization efforts across the agency and applies to all EPA programs, regions, labs, and offices. It is to be used in conjunction with the Digitization Standard issued under the information directives framework in the implementation of digitization programs. Owners of existing systems and applications currently digitizing documents within the scope of their operating authority (e.g., the Superfund Enterprise Management System, the Federal Docket Management System, the Correspondence Management System, etc.) are strongly encouraged to evaluate and compare the procedure to their current practices and adopt the procedures (immediately or in a phased approach), but only if doing so does not jeopardize the status of their established standard business practice(s).

Once digitized and validated, the resulting electronic versions of the documents/records shall be managed according to the appropriate information directives for agency records (<https://www.epa.gov/irmpoli8/current-information-directives>).

This procedure is effective immediately upon execution by the CIO for date-forward digitization operations.

3. AUDIENCE

The audience includes all employees, contractors, grantees and other agents who digitize agency documents and records through internal operations, and managers or other officials that administer EPA programs with digitization operations.

4. BACKGROUND

Over the last several decades the federal government has increasingly recognized the importance of sharing information assets to meet mission needs and making them accessible to stakeholders and the public. Several influences within the federal government are uniting to spur digitization, which facilitates information sharing. Increased need for transparency and accessibility to information (e.g., Presidential Memoranda and

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

Directives, Freedom of Information Act [FOIA] and E-discovery), changing human_resource management needs (e.g., teleworking) and space consolidation are significant drivers for digitization. Other drivers include the National Archives and Records Administration (NARA) and the Office of Management and Budget (OMB) Memorandum M-19-21, with the following goal:

- Goal 1.1, Requiring that all permanent electronic records be managed electronically by December 31, 2019, to the fullest extent possible for eventual transfer and accessioning to NARA in an electronic format
- Goal 1.2, Requiring all permanent records in federal agencies be managed in an electronic format with appropriate metadata by December 31, 2022
- Goal 1.3, Requiring all temporary records in federal agencies to be managed in electronic format or stored in commercial records storage facilities by December 31, 2022
- Goal 2.4, Requiring NARA to no longer accept transfers of permanent or temporary records in analog formats (hardcopy, microfilm and microfiche) and only accept records in electronic format and with appropriate metadata, after December 31, 2022

A significant benefit to digitization is the reduction in paper-based documents and records processes, which may include cost reductions and increased accessibility. The procedure:

- Supports the move from paper to electronic records;
- Integrates and standardizes the digitization process as part of the records life cycle while leveraging existing agency investments in the EPA Enterprise Architecture (EA), Enterprise Content Management (ECM), and Enterprise Information Management (EIM);
- Serves as a framework into which additional program-specific processes and workflows can be incorporated; and
- Establishes the framework to address the requirements of the National Archives and Records Administration (NARA) and the Office of Management and Budget (OMB) Memorandum, M-19-21, under which the agency must operate.

Digitizing analog audio and video recording records requires specialized procedures that are specific to the subject media, and are beyond the scope of this document. See NARA Guidance at <http://www.archives.gov/preservation/formats/video-playback-digitize.html>, and <http://www.archives.gov/preservation/formats/audio-toc.html>.

5. AUTHORITY

- Clinger-Cohen Act (also known as Information Technology Management Reform Act of 1996) (Pub.L. 104-106, Division E)
- Paperwork Reduction Act of 1980, as amended by the Paperwork Reduction Act of 1995 (44 U.S.C. Chapter 35)
- Government Paperwork Elimination Act of 1998 (Pub. L. 105-277, Title XVII)
- Executive Order – Making Open and Machine Readable the New

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

- Default for Government Information, May 9, 2013, Presidential Memorandum: Managing Government Records, November 28, 2011
- Presidential Memorandum: Building a 21st Century Digital Government, May 23, 2012
 - Management of Federal Information Resources (OMB Circular No. A-130)
 - Open Government Directive M10-06, December 8, 2009
 - OMB Memorandum M-13-13, Open Data Policy – Managing Information as an Asset, May 9, 2013
 - United States vs. Russo, 480 F.2d 1228, 1239 (6th Cir. 1973)
 - CIO 2130.1: Section 508: Accessible Electronic and Information Technology, February 20, 2014
 - NARA/OMB Memorandum, M-19-21: Transition to Electronic Records, June 8, 2019

6. PROCEDURE

Implementing the procedure at each dedicated scanning location requires, at a minimum,

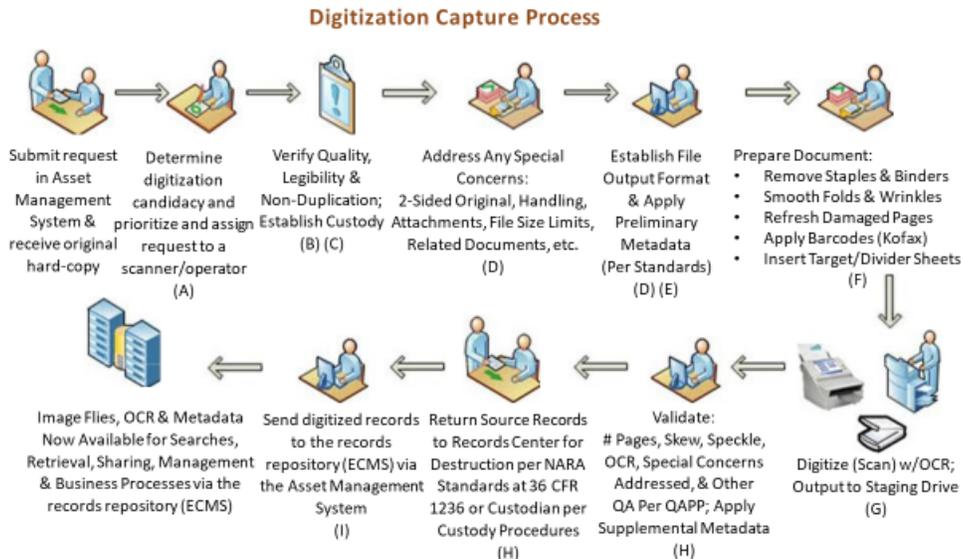
- 1) Developing a Standard Operating Procedure (SOP) that incorporates the core digitization capture process steps outlined in Sections 6.A. through 6.I below;
- 2) Including any necessary location-specific process adjustments; and
- 3) Accommodating the specific hardware and software to be used.

Production, documentation and implementation of the SOPs must include all stakeholders, such as the Agency Records Officer, regional Records Liaison Officers, information technology (IT) staff, *etc.*

The organization shall determine what level of sign-off is required to permit any variances from the procedure at a local site. This authority must be documented in the local SOPs to ensure reproducibility, defensibility, and non-repudiation of EPA records in court proceedings. Figure 1 graphically represents the core process steps that must be followed for the digitization process.

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1



There are several steps that must be followed to achieve the desired results when implementing a digitization process. These steps are described below:

A. Determine Records Status and Digitization Candidacy

- Is the document a record or a potential record?
 - Digitization of paper documents shall typically be limited to paper and microform versions of EPA records
 - In situations of litigation, discovery, subpoena or FOIA requests, responsive documents may not be limited to actual records as defined by the Federal Records Act. If in doubt about records status, the document custodian shall consult with the appropriate subject matter experts, attorneys and records officer(s)
- Should the document be digitized?
 - Some factors to consider before incurring the expense of digitization:
 1. Is a definitive “born digital” (records which are generated electronically) version of the document (and related metadata) readily available from the source that will satisfy legal and records requirements? If so, the electronic version shall be validated and used in lieu of digitizing a paper version of the same document.
 2. Is the document likely to be used periodically for EPA business purposes in the next three to five years? If so, the document should be digitized and managed according to the appropriate agency guidance.
 3. Is it now, or likely to be, requested under FOIA, litigation, discovery or subpoena? If so, digitize the document.
 4. Does a digitized version of the document already exist in an agency content management system or other repository? If so, either watermark the document as a “COPY” prior to digitizing it, or do not digitized the document

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

- and utilize the existing version.
5. Was the document/record previously requested under FOIA and provided as a response in hard copy at least two times before (the eFOIA Amendments require providing materials in an “electronic reading room,” e.g., FOIAonline, after the third request)? If so, digitize the record/document.
 6. Is it an essential record? Essential Records are records that are critical to the continuing operations of the agency in a continuity of operations situation. If so, digitize the document.
 7. Is it related to a high-profile agency activity? If so, digitize the document.
 8. Is it now, or likely to be, part of a Department of Justice (DOJ) referral package supporting an enforcement action? If so, digitize the document.
 9. Is it now, or likely to be part of an Administrative Record (i.e., records that are the basis for any administrative action, including rulemaking)? If so, digitize the document.
 10. Is it needed for collaborative activities/sharing? If so, digitize the document to facilitate ease in digital transfer/sharing.
 11. Is it needed by remote users or others in the field? If so, digitize the document.
 12. Will its digital availability improve productivity or improve/streamline business processes? If so, digitize the document to realize these benefits.
- The above list is not intended to be exhaustive, but can serve to inform a digitization decision
 - Development of a decision matrix or other documentation is recommended, which could be used to develop a standard approach, capture the most frequent rationale for the decision to digitize, create “rules of thumb” or decision tools for practitioners in specific areas, and justify the costs, which will be incurred by the program or region

B. Assess Quality and Context, and Determine Whether Duplicate Documents Exist

- Determine the quality of the document for scanning purposes, and address any issues, including but not necessarily limited to the following:
 - Are all pages present and in a condition ready for scanning?
 - Are any pages oversized, requiring scanning on a different machine?
 - Are any special scanner or software settings required to scan the document (e.g., two-sided original, contrast/brightness enhancements, etc.)?
 - Are there any attachments to the document that also need to be digitized and associated?
 - Are there any other known related documents (digitized or otherwise) that need to be captured and/or associated with the subject document?
 - Is there a risk of damage to the document from the scanning process?
- Determine whether the document is already available in electronic format and is a duplicate of the document to be digitized.
 - If so, a justification for digitizing the document again shall be developed and submitted to the Records Liaison Officer (RLO) or Information Management Official (IMO) for approval
 - For the purposes of this document, the record custodian/officer must use

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

experience and comparative analysis to determine whether one digitized version is a duplicate of another through visual inspection, review of the content, and other characteristics that are necessary to ensure the full content of the original version of a document is captured in any copy or duplicate thereof (e.g., the presence of signatures, marginalia, etc.), that lend record value over and above a version of the document without those attributes

C. Establish and Track Custody

Chain of custody for paper records during the scanning process is critical to establishing that the designated paper record is indeed what was digitized, that it was used for comparative purposes during the quality review of the scanned images, and that the paper original was returned to the records center/repository or custodian upon completion of the digitization process. Typically, chain of custody procedures for paper records management systems are already in place and shall apply to the digitization process; however, if a system for tracking custody of paper records/documents during the digitization process is not established, the RLO, IMO and other stakeholders shall devise and establish a means to track custody. Documents not already listed in a tracking system for digitizing paper records should be logged in, and then custody assigned to the scanning technician prior to proceeding with the capture processes. This establishes a chain of custody and identifies a responsible party/custodian in case access to the hard copy records is needed by others during the capture process, or if the materials are not returned to the records center/original custodian at the completion of the digitization process.

D. Identify Special Handling Concerns and Requirements

Special handling may be needed regarding the following:

- Identifying anomalies or unique aspects of the document to be digitized in order to prepare the equipment and software settings
- Determining the desired file output format – will be dependent upon the hardware and software used, document content, NARA or other requirements, and any downstream processing required to conform to the Document Digitization (Scanning) Standard and others, as appropriate
- Specifying any metadata/indexing elements that are needed to ensure that the relationships of the subject document to any related documents or attachments are captured during the cataloguing process, as required for conformance with other standards such as CIO Standards: EIM Minimum Metadata Standards, etc.
- Ensuring that embedded metadata in certain documents (e.g., exchangeable image file (Exif) or Extensible Metadata Platform [XMP] data in TIFF files) are not removed (although the data can also be copied into the external metadata/indexing schema to expose it to the widest variety of search tools)
- Ensuring that documents exceeding 250 pages (or 500 megabytes in digital file size, after scanning) shall be broken down into smaller pieces (and catalogued accordingly) since large file sizes become a factor when transmitting them across the Internet by email or saving to storage media for file sharing, such as CD-ROM, with limited storage capacity. Large format documents such as maps and large photographs, particularly color versions,

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

that when digitized exceed 500 Mb should be considered exempt from this size limitation, represented by a “target” sheet if part of a document, and shared using another form of portable storage media such as DVD-ROM or secure digital card

- Ensuring that PDF output is set to “Optimize for Web and Mobile” to allow preliminary viewing of files delivered through networks or the Internet before the entire file has been downloaded, and to “Optimize Scanned Documents,” to make them more useful by converting the contents into searchable text, correcting page orientation, and reducing file size

E. Perform Preliminary Cataloguing and Indexing of Document

- Perform preliminary/initial cataloguing/indexing of the document according to applicable standards and conveniences as needed to establish primary key data elements for subsequent retrieval and additional cataloguing as needed to conform to agency metadata standards for digitized records (see sections 7 and 11 below for more details)
- Index additional metadata elements during the scanning process or after the document has been scanned and its image used for indexing purposes instead of the hard copy original (at which point, the original should be returned to the records center/repository/custodian and custody transferred accordingly)

F. Prepare Document for Digitization

- Remove binders, staples and paper clips
- Review document to verify completeness, legibility and “scan-ability” to ensure that the imaged file is representative of the original paper materials and will scan smoothly
- Make photocopies of original documents/pages, if necessary, that are fragile or of a non-standard size or thickness which may hinder or be irreparably damaged upon movement through the scanner mechanisms, or that may not yield a legible image due to bleed-through
- Insert target sheets, as needed, to create a place-holder for oversized or “special handling” pages such as maps and charts; the same metadata requirements described in Section 6E above are applicable to the target sheet as to the subject document
- Insert separator sheets to identify breaks in lengthy documents (>250 pages) or attachments (“children”) to a “parent” document, or to identify when a new document begins during batch scanning operations
- Determine if the original document has single- or double-sided printing.
- Count the number of pages to be scanned in the original document for quality assurance (QA) purposes after scanning to be sure that all necessary pages were captured
- Assemble materials that will be batch-processed according to scanner and software requirements
- Apply any bar codes (taking care not to obscure any important content) with identifying data on the first/cover page of a document (e.g., a document ID as primary key indexing information), or as required by the digitization software, to

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

link the image file with related documents and the document metadata captured in Section E above when it is released to the local storage environment

G. Digitize (Scan) the Document

- Using agency approved software, scan and apply optical character recognition (OCR) to the subject document(s) if there is textual content in the documents
- Outputting format, resolution, and destinations are dependent upon the source documents, whether the output requires additional processing by Enterprise Digitization Software, and the target storage environment of the content, with this subsequent processing performed according to appropriate SOPs for these activities
- Outputting of scanned images shall ideally be stored temporarily in a local staging storage environment to better facilitate additional processing, quality control, and finalization prior to release to the production storage environment

H. Verify, Perform Quality Assurance (QA) and Validate

- Perform quality control and assurance throughout the overall digitization process. Practitioners at each location must establish the controls to ensure quality measures are integrated into the overall process
- Verify visually that the scanned images match the hard copy originals (e.g., number of pages, skew, speckle, etc.); perform other QA activities as required to conform to applicable Quality Assurance Project Plan(s) (QAPP) for digitized records, or as required by CIO 2106: Quality Policy: Procedure for Quality Policy. Ideally, the person performing QA activities shall be someone other than the person who digitized the document to achieve independent review/validation of QA parameters
- Confirm that related documents (parent, child, attachments, appendices, etc.) were also digitized, or are in the digitization queue, and are appropriately associated to ensure completeness of future search/retrieval efforts
- Confirm the accuracy of indexing/metadata, and complete any additional cataloguing/indexing of the subject document(s), using the appropriate content management tools, user interfaces, etc.
- Return the hard-copy original document to the records center or custodian pursuant to the organization's records custody procedures once all processing is completed

I. Release to Production Environment

- Upon completion of Sections A – H above, as well as any additional steps required by system- or program-specific SOPs, the images and related metadata may be released from the staging storage environment to the production/enterprise storage environment
- Additional processing may occur later, but any major alterations or changes (e.g., broad changes to metadata elements, addition/removal of pages, etc.) must be documented as part of the audit trail (the documentation record of the process used) for capture and management of the document

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

7. ROLES AND RESPONSIBILITIES

The roles and responsibilities of EPA employees with respect to digitization of agency documents and records include:

The Chief Information Officer (CIO):

- Lead agency-wide implementation of the procedures as part of the overall framework of Information Directives
- Facilitate the process for appropriate business organizations to incorporate the Digitization Procedure into their organization and operations
- Manage the Senior Advisory Council process to update the procedures and associated policies, and propose new information policies and procedures
- Authorize formal information calls for updates or reviews of the procedures, as appropriate
- Grant waivers to selected provisions of the procedures for sufficient cause, or delegate waiver authority

Senior Advisory Council (SAC):

- Advise and assist CIO in developing and implementing the agency's quality and information goals and policies
- Review updates to the Digitization Procedure and associated policies, and proposed new information policies and procedures
- Review any progress reports provided and addresses success as well as agency-wide challenges for the effective implementation of the procedures
- Endorse enterprise-wide information investments, coordinates with Agency Investment Oversight Boards, as appropriate

Senior Information Officials (SIOs):

- Implement the Digitization Procedure within their organizations
- Apprise the SAC of major information management issues within their offices
- Ensure that the information technology used and managed by their organizations support their business needs and mission and helps to achieve strategic goals
- Ensure Enterprise Architecture compliance of solution architectures
- Review, concur, and advise on waivers to the procedure, typically through participation on the Information Technology Operations Planning Technology Committee (IOPC)

Information Management Officials (IMOs):

- Review, concur on or approve acquisition of digitization technologies to ensure compliance with the procedures
- Review agency digitization policy documents, as appropriate. Proposals to set new standards and procedures should be submitted to the appropriate group under the IT/IM Governance model, following the most current CIO Policy Review

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

procedures

- Ensure that staff and contractors are aware of the procedures, and related technical specifications and standard operating procedures
- Ensure that employees, senior environmental enrollees, and contractors are aware of their responsibilities regarding digitization
- Review and/or certify compliance with the procedures and other agency digitization policies and procedures, as appropriate

Records Liaison Officers:

- Participate in the development and maintenance of digitization standards and procedures, as appropriate, for relevant programs, regional offices, laboratories, etc.
- Support and implement standards, technical specifications, the procedures, and standard operating procedures (SOPs) for their organizations
 - Work with records, document and content owners/generators to plan and manage the life cycle of the digitized materials
 - Assist personnel to implement the procedures and related SOPs
 - Coordinate with the Information Management Official and provide outreach, support, and technical assistance as appropriate to ensure the proper implementation of the procedures

Records Custodians:

- Use the procedures to capture agency-owned digitized information
- If serving as a Contracting Officer Representative, you may instruct contractors to employ the procedures in digitization operations

All EPA employees and agents:

- Use the procedures to capture agency-owned digitized information in content repositories

8. RELATED INFORMATION

- CIO 2105.0: Policy and Program Requirements for the Mandatory Agency-Wide Quality System, May 5, 2000
- CIO 2106.0: Quality Policy, October 20, 2008
- CIO 2106-P-01.0, Procedure for Quality Policy, October 20, 2008
- CIO 2122.1, Enterprise Architecture Policy, December 21, 2017
- CIO 2122-P-03.0, Information Technology Infrastructure Standard Procedure, October 1, 2010
- CIO 2133.0: Data Standards, June 28, 2007
- CIO 2155.4: Interim Records Management Policy, August 22, 2018
- CIO 2131.0: National Geospatial Data Policy, August 24, 2005

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

9. DEFINITIONS

- **Cataloguing:** The process of listing an information object for inclusion in a registry/repository as part of an inventory
- **Content:** The intellectual substance of a document, including text, data, symbols, numerals, images and sound (Society of American Archivists)
- **Content Management:** The capability to manage and track the location of, and relationships among, content within a repository. (AIIM International)
- **Content Repository:** Securely stores electronic content and associated metadata with management controls
- **Custodian:** A person who has custody; keeper; guardian; trusted with guarding or maintaining property
- **Document:** Information set down in any physical form or characteristic. A document may or may not meet the definition of a record. (DOD 5015.2-STD)
- **Enterprise:** An organization (or cross-organizational entity) supporting a defined business scope and mission. An enterprise includes interdependent resources (e.g., people, organizations, and information technology) that must coordinate their functions and share information in support of a common mission (or set of related missions)
- **Enterprise Content Management:** The strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists. (AIIM International, <https://www.aiim.org/What-is-ECM>)
- **Metadata:** Data describing stored data; that is, data describing the structure, data elements, interrelationships, and other characteristics of electronic records. (DOD 5015.2)
- **Senior Advisory Council (SAC):** The SAC consists of high-level managers from each EPA Region and program office, typically at the Deputy Assistant Administrator level. The SAC's primary focus is on addressing and resolving intra-agency cross-media, cross-program, and interdisciplinary information technology/information management and related policy issues
- **Quality Assurance (QA):** A management or oversight function that deals with setting policy and running an administrative system of management controls that cover planning, implementation, review, and maintenance to ensure products and services are meeting their intended use
- **Quality Control (QC):** The overall system of technical activities that measure the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer; operational techniques and activities that are used to fulfill requirements for quality
- **Record(s):** All recorded information, regardless of physical form or characteristics, made or received by an agency of the United States government under federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

of the government or because of the informational value of data in them (44 U.S.C. §3301)

- **Records Management:** The planning, controlling, directing, organizing, training, promoting and other managerial activities involved with respect to records creation, records maintenance and use, and records disposition in order to achieve adequate and proper documentation of the policies and transactions of the Federal Government and effective and economical management of agency operations. (36 CFR §1220.14)
- **Skew:** Pages that are not straight, not in line with others or turned to the side.
- **Speckle:** Small spots or marks on a document or image.
- **Standard:** Universally or widely accepted, agreed upon, or established means of determining what something should be. Major classifications of the term include: (1) Material or substance whose properties are known with a level of accuracy that is sufficient to allow its use as a physical reference in calibrating or measuring the same properties of another material or substance. (2) Concept, norm, or principle established by agreement, authority, or custom, and used generally as an example or model to compare or measure the quality or performance of a practice or procedure. (3) Written definition, limit, or rule approved and monitored for compliance by an authoritative agency (or professional or recognized body) as a minimum acceptable benchmark. This is the usual meaning of the plural term (standards) (<http://www.businessdictionary.com/definition/standard.html>)
- **Standard Operating Procedure (SOP):** Written procedure prescribed for repetitive use as a practice, in accordance with agreed upon specifications aimed at obtaining a desired outcome (<http://www.businessdictionary.com/definition/standard-operating-procedure-SOP.html>)
- **Taxonomy:** A hierarchical structure of information components, any part of which can be used to classify a content item in relation to other items in the structure (from “The Challenges of Building Enterprise Content Taxonomies and the Role of Classification Technologies in Maintaining their Effectiveness”, Reginald J. Twigg, PhD, IBM Corporation, 2007)
- **Web:** A system of Internet servers that support specially formatted documents. The documents are formatted in a markup language called HTML (Hypertext Markup Language) that supports links to other documents, as well as graphics, audio, and video files. (http://www.webopedia.com/TERM/W/World_Wide_Web.html)

10. WAIVERS

Consistent digitization procedures are critical to facilitating the exchange, use and integrity of the agency's unstructured information. For this reason, waivers to the procedures are rare and will be considered on a case by case basis.

Waiver Process: The agency's CIO may grant waivers to selected provisions of the procedures for sufficient cause. The CIO may delegate waiver authority to others, e.g., the Information Management Official (IMO).

Applications: Requests for waivers to specific provisions of the procedures must conform to the appropriate OMS-EI waiver procedures, and must contain 1) identification of the

Digitization (Scanning) Procedure

Directive No: CIO 2155-P-05.1

procedures provision; 2) a listing of reasons why the procedures cannot be applied or maintained; 3) an assessment of impacts resulting from non-compliance; and 4) a memorandum to the CIO originating at the Office Director level (or equivalent) responsible for the information in question, through the SIO or other senior manager.

Notification: The CIO will notify the requesting office in writing of the disposition of the waiver application within 60 days of receipt.

11. MATERIAL SUPERSEDED

- CIO 2155.P-05.0: Digitization (Scanning) Procedures, August 4, 2015
 - NARA/OMB Directive M-12-18: Managing Government Records, August 24, 2012
-

12. CONTACTS

For questions about this Procedure, please contact the U.S. EPA National Records Management Program, (202) 566-1494.

Vaughn Noga
Chief Information Officer and
Deputy Assistant Administrator for Environmental Information
U.S. Environmental Protection Agency