SISTEMA DE GESTIÓN DOCUMENTAL CENTRAL

Guía para la digitalización de documentos

Oficina del Principal Ejecutivo de Información de Puerto Rico
(Chief Information Officer)

Borrador de Trabajo
Versión 1.0 (2011-03-14)

Tabla de Contenido

INTRODUCCIÓN **1**

1. ADMINISTRACIÓN DE DOCUMENTOS 2
2. SISTEMA DE GESTIÓN DOCUMENTAL CENTRAL (SIGDOC) 4
3. CONTENIDO Y ALCANCE DE ESTA GUÍA 5

Propósito *5*
Base Legal *6*
Alcance *6*
Motivo *7*

GUÍA DE DIGITALIZACIÓN DE DOCUMENTOS **7**

1. PLANIFICACIÓN 8

Objetivos del Proyecto *8*
Pasos esenciales dentro del proceso de Digitalización *9*
Procesos y tareas de los cuatro pasos para la Digitalización de Documentos *12*

2. FUNCIONALIDADES DEL SISTEMA 17

Consideraciones *17*
Uso Práctico *18*

3. METODOLOGÍA 18
4. PROCESO DE DIGITALIZACIÓN DE DOCUMENTOS 21

Configuración del escáner *21*
Configuración básica *22*
Resolución y tamaño de la imagen *22*
Profundidad o intensidad de color *23*
Imagen en el monitor *27*
Tamaño y orientación del papel *28*
Compresión y tamaño del archivo *29*
Formato de los archivos *31*
Reconocimiento óptico de caracteres *35*
Proceso de escanear *36*
Control de calidad *37*

5. CLASIFICACIÓN DE DATOS 38

Metadatos *38*
Control de calidad *40*

6. REGISTRO DE AUDITORÍA (*AUDIT TRAIL*) Y PERÍODOS DE RETENCIÓN 40
7. MIGRACIÓN DE LOS DATOS AL SIGDOC 41
8. VERIFICACIÓN Y VALIDACIÓN DE CALIDAD 42

CONCLUSIÓN **44**

REFERENCIAS **45**

ANEJO I - DEFINICIONES **47**

ANEJO II – TABLA DE RESUMEN **51**

ANEJO III – TABLAS GUÍA **54**

A– Tamaños de papel *54*
B- Pros and Cons Digital Capture Devices *55*
C- Levels of scan, file format, uses and alterations *56*
D- Bit Depth *57*
E- Improving the lifespan of CD's and DVD's *58*
F- Improving the lifespan of Digital Linear Tape (DLT) *59*

F- Electronic Access Project Scanning and File Format Matrix *60*

Listado de Diagramas

FIGURA 1 - PROCESO DE DIGITALIZACIÓN 11
FIGURA 2 - CONTROL DE CALIDAD 43

Listado de Tablas

TABLA 1 - ESTÁNDARES UTILIZADOS EN LA GUÍA DE DIGITALIZACIÓN DE DOCUMENTOS 18
TABLA 2 - TIPO DE DOCUMENTO Y PROFUNDIDAD DE COLOR 25
TABLA 3 - PROFUNDIDAD DE COLOR Y RESOLUCIÓN SEGÚN LOS TIPOS DE DOCUMENTOS 26

INTRODUCCIÓN

La Oficina del Principal Ejecutivo de Información de Puerto Rico (CIO, por sus siglas en inglés) fue creada por virtud de la Orden Ejecutiva OE-2009-009, firmada por el Gobernador de Puerto Rico el 24 de febrero de 2009. La creación del cargo y de la Oficina se presentó como respuesta a la necesidad de lidiar con "la actual incompatibilidad de las tecnologias de información y comunicación ("TIC") utilizadas por múltiples entidades del Gobierno para prestar sus servicios". Se planteó además como respuesta a la necesidad de "desarrollar e implementar una agenda integrada de gobierno electrónico que redunde en el mejoramiento de la prestación de servicios gubernamentales y la facilitación del acceso e interacción de los ciudadanos con el Gobierno".[1]

A la nueva entidad se le asignó entre sus funciones primordiales el "establecer e implementar los planes estratégicos, las políticas, estándares y la arquitectura integrada de TIC del Gobierno"; "reducir redundancia y duplicación de costos de las operaciones, proyectos y esfuerzos de TIC del Gobierno"; "dirigir los esfuerzos de reingeniería de procesos gubernamentales y de TIC para conseguir las eficiencias necesarias en el servicio al ciudadano"; y "funcionar como promotor de la implementación de disciplina en las mejores prácticas en el manejo de proyectos y publicar guías y directrices a tales efectos".[2] Además, la orden ejecutiva mandató que, como parte del apoyo tecnológico que la Oficina del CIO brindará a las agencias del Gobierno, la misma "deberá identificar las áreas del Gobierno con necesidades apremiantes; asignar labores a los funcionarios de su Oficina; preparar manuales e informes técnicos; y desarrollar una infraestructura técnica y de tecnología adecuada para lograr dicho propósito".[3]

En el descargue de sus deberes y prerrogativas, la Oficina del CIO ha preparado esta **Guía de Digitalización**[4] dirigida principalmente a proveerle a las entidades gubernamentales los lineamientos necesarios para la digitalización de documentos públicos y su conservación en formato electrónico a través de su ciclo vital bajo un **Sistema de Gestión Documental Central** (SIGDOC).

Ante el incremento en el número de entidades gubernamentales que desean digitalizar los documentos y conservarlos en formato electrónico, la Oficina del CIO establece los parámetros mandatorios que deben ponerse en práctica en toda iniciativa de digitalización. El Gobierno de Puerto Rico está exhortando el que se conserven los documentos oficiales, que no son permanentes, en formato electrónico. Esto va a permitir un acceso más eficiente, promoviendo la transparencia en los procesos a la vez que disminuye el costo de almacenaje.

Manteniendo una visión global del gobierno, la Oficina del Principal Ejecutivo de Información se ha propuesto la creación del Sistema de Gestión Documental Central para que las entidades gubernamentales puedan depositar sus récords (series documentales), salvaguardando la integridad,

---

[1] Boletín Administrativo Núm.: OE-2009-009, Orden Ejecutiva del Gobernador del Estado Libre Asociado de Puerto Rico para establecer la Oficina del Principal Ejecutivo de Información del Gobierno de Puerto Rico ("Chief Information Officer"), 24 de febrero de 2009.

[2] Boletín Administrativo Núm.: OE-2009-009, Sección 2da, incisos D, F, G, y H.

[3] Boletín Administrativo Núm.: OE-2009-009, Sección 5ta.

[4] Esta Guía toma como punto de partida la desarrollada en el 2007 por el Programa de Administración de Documentos de la Rama Ejecutiva del Gobierno de Puerto Rico, adscrito a la Administración de Servicios Generales (ASG).

autenticidad, fiabilidad y el uso práctico[5] de los mismos. Es decir, este archivo central de gestión documental, de forma centralizada, va a permitir el acceso a la información, tanto por las entidades gubernamentales como por el ciudadano.

Por tal razón, las entidades gubernamentales que desean entrar en una iniciativa de digitalización deben formar parte de este repositorio para beneficio del gobierno y de la ciudadanía que se sirve de él.

El propósito de esta *Guía para la Digitalización del Sistema de Gestión Documental Central* (en adelante, "Guía") es ofrecer a las entidades gubernamentales unos lineamientos, así como requisitos mínimos en los equipos y sistemas que van a ser utilizados, para facilitar la transición de documentos de papel a digital y su conservación a través de su ciclo vital desde un solo repositorio.

1. ADMINISTRACIÓN DE DOCUMENTOS

El Programa de Administración de Documentos de la Administración de Servicios Generales (ASG) del Gobierno de Puerto Rico es el encargado de gestionar una sana administración de los documentos públicos para la Rama Ejecutiva del Gobierno de Puerto Rico. Su propósito es el de establecer un "programa sistemático de **conservación de documentos,** que en consideración a su valor histórico, legal, administrativo o informativo merecen preservarse por mucho más tiempo, y la eliminación de aquellos documentos que no solamente carecen de valor permanente, sino que además han perdido ya toda su utilidad administrativa."[6]

Se define como documento "todo papel, libro, folleto, fotografía, fotocopia, película, microforma, cinta magnetofónica, mapa, dibujo, plano, cinta magnética, disco, video cinta o cualquier otro material leído por máquina y cualquier otro material informativo independientemente de su forma o características físicas."[7]

Durante las pasadas dos décadas, la creación y el manejo de los documentos (*Record Management*) han sido impactados de forma significativa por las innovaciones en las computadoras y en la tecnología de las comunicaciones. Esto trajo como consecuencia que los documentos fueran creados y manejados en un formato electrónico, sustituyendo así a la generación de documentos en papel. La era electrónica estaba emergiendo y con ello se comenzaban a establecer nuevos cambios y retos en el manejo de los documentos.

Actualmente, más del noventa y cinco (95) por ciento de los documentos que se generan en el Gobierno de Puerto Rico se inician de forma electrónica. Es por esta razón que resulta importante poder conservar estos documentos durante todo el ciclo de vida, que va desde su inicio hasta la eventual disposición de estos. Existen diversas formas mediante las cuales se puede iniciar un documento electrónico. Una de ellas es a través de la **digitalización**.

Cuando hablamos de digitalización nos referimos al proceso de convertir documentos físicos (texto, mapas, planos, manuscritos, video, entre otros) conservándolos en formato digital y procesado por computadoras. Este proceso va a sustituir, mediante la captura[8], los documentos físicos que no son de

---

[5] Estos son los cuatro factores que definen el *recordkeeping* electrónico. El *recordkeeping* electrónico son los procesos y procedimientos que se crean para garantizar la validez de los documentos a través de su ciclo vital.

[6] Ley número 5 del Gobierno de Puerto Rico aprobada el 8 de diciembre de 1955, según enmendada.

[7] Estado Libre Asociado de Puerto Rico - Administración de Servicios Generales. Reglamento 4284 - Reglamento para la Administración de Documentos Públicos en la Rama Ejecutiva.

[8] La captura de un documento significa: Establecer una relación entre el récord, su creador y el contexto del negocio que lo originó, coloca el récord y sus relaciones dentro de un sistema de récords y lo vincula a otros. [ISO 15489-1:2001]

valor permanente transformándolos en documentos "oficiales" electrónicos con el propósito de conservarlos a través del ciclo vital. Como los documentos físicos son los oficiales, en el momento de la captura (digitalización) se establecen los criterios para oficializar los documentos electrónicos para su conservación garantizando la veracidad, integridad, autenticidad y uso práctico se establece mediante la implantación de un proceso de recordkeeping electrónico.

El Programa de Administración de Documentos, al amparo de la Ley 5, faculta a las dependencias a utilizar algún medio electrónico para reproducir los documentos por el período de retención correspondiente y disponer del original[9]. Con esto se indica que se puede hacer la sustitución de la información física por la digital. Con este propósito en mente, la presente Guía estará fundamentada en los aspectos regulatorios que rigen a las entidades, estándares internacionales o mejores prácticas, para facilitar la transición de oficializar las copias digitalizadas en sustitución de los documentos oficiales. Es decir, su propósito fundamental es proveer los mecanismos eficientes y seguros que permitan capturar las imágenes, presentarlas, conservarlas por el tiempo que sea necesario y disponer de ellas cuando se acabe su período de retención.

Debido a la diferencia en los tipos de documentos y el propósito por el cual se decide digitalizarlos, se tienen que tomar diferentes enfoques para manejar la captura[10] y la conversión a formatos digitales. Teniendo esa realidad en cuenta, se diseñó esta Guía para que proveyera un marco de referencia lo suficientemente amplio como para suplir las necesidades de las diferentes entidades en cualquier proyecto de digitalización. Es necesario señalar que la intención de esta Guía no es pasar juicio valorativo sobre las diferentes iniciativas de digitalización, sino que únicamente va a proveer los estándares necesarios para manejar el proyecto de forma exitosa y en cumplimiento con las leyes y regulaciones establecidas.

Debe entenderse que la iniciativa del Sistema de Gestión Documental Central incluye unas recomendaciones y unos lineamientos a seguir cuando las entidades de gobierno ponen en ejecución iniciativas de digitalización. Por ejemplo, la utilización del procedimiento de Reconocimiento Óptico de Caracteres (OCR, por sus siglas en inglés) al digitalizar documentos de texto **es de carácter mandatorio**. Aunque hace algunos años se utilizaba con precaución por razones técnicas y por el espacio requerido, el procedimiento de OCR se ha ido generalizando cada vez más y se ha convertido en un procedimiento estándar. Hoy día resulta relativamente fácil incorporarlo en los proyectos de digitalización. De hecho, a partir de abril de 2010, la Oficina del Contralor de Puerto Rico le requiere a las agencias del gobierno que se utilice "OCR en la Digitalización de Contratos, Escrituras y Documentos Relacionados", al enviar copias a la Oficina del Contralor del Estado Libre Asociado de Puerto Rico.[11]

Debido a los constantes cambios en la tecnología y los nuevos requerimientos que se van creando para mantener la integridad, autenticidad, fiabilidad y uso práctico de la información, se requiere su revisión constante. No obstante, lo que aquí se recoge cumple con las necesidades del momento.

---

[9] Ley número 5, de 8 de diciembre de 1955, según enmendada. Artículo 4-A (1) y (2) "autorizar los jefes de sus dependencias bajo sus respectivas jurisdicciones a reproducir por medio del proceso de microfotografía, fotocopia,…, o cualquier otro método de reproducción electrónica, aquellos documentos públicos que se encuentren bajo su custodia que merezcan ser conservados, así reproducido… Los originales de los documentos reproducidos…podrán ser destruidos en conformidad con lo dispuesto en esta ley…".

[10] La captura de un documento significa: Establecer una relación entre el récord, su creador y el contexto del negocio que lo originó, coloca el récord y sus relaciones dentro de un sistema de récords y lo vincula a otros. [ISO 15489-1:2001]

[11] La Carta Circular de la Oficina del Contralor de Puerto Rico especifica que "a partir del 30 de abril de 2010, el Programa de Registro de Contratos, no validará ningún contrato digitalizado que no esté bajo la aplicación de OCR, y el contrato no podrá ser registrado hasta tanto sea digitalizado con dicha aplicación." (Carta Circular OC-IO-27, 31 de marzo de 2010).

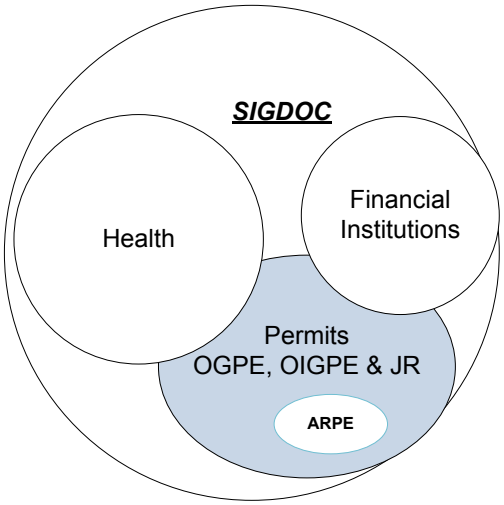
2. SISTEMA DE GESTIÓN DOCUMENTAL CENTRAL (SIGDOC)

La Oficina del CIO de Puerto Rico ha desarrollado y está en vías de implantar un Sistema de Gestión Documental Central (SIGDOC) con el propósito de capturar digitalmente la información que por separado generan las agencias y hacerla accesible en formato digital en forma centralizada. Este sistema va a permitir que los documentos públicos puedan ser conservados en formato electrónico a través de su período de retención promoviendo una sana administración y transparencia en los procesos.

Los documentos interagenciales que se conserven dentro de este sistema o plataforma centralizada de manejo de contenido van a seguir los lineamientos presentados en la Política de Récords Electrónicos y las guías que se desarrollen en esta oficina relacionados al manejo de los documentos electrónicos.

Las agencias que deseen digitalizar los documentos tienen que desarrollar los procedimientos más específicos que permitan el manejo de los documentos comenzando desde su creación y/o captura hasta depositarlos en el SIGDOC. A su vez, todos los procedimientos que se desarrollen para el manejo de los récords electrónicos tendrán que estar en armonía con la Política de Documentos Públicos de la Rama Ejecutiva del Gobierno de Puerto Rico para que permitan una transformación ordenada de papel a digital.

Uno de los pasos esenciales dentro de esta Guía es el de establecer el Esquema de Taxonomía para la clasificación de los documentos (o unidad documental) que se van a digitalizar. Ese esquema va a aplicar a una unidad (agencia) dentro de un esquema más amplio (SIGDOC) que habrá de establecerse para el conjunto de entidades del Gobierno de Puerto Rico.



Desde el punto de vista técnico, los documentos van a ser capturados electrónicamente mediante un escáner, se clasificarán y se conservarán dentro de un Sistema de Manejo de Récords Central (SIGDOC). El sistema va a permitir la administración de la gestión documental electrónica, o recordkeeping electrónico para facilitar el manejar esos documentos de forma efectiva y a la vez proteger su integridad y autenticidad durante todo su ciclo vital. La utilización del esquema de taxonomía para su clasificación va a ser crucial para acceder la información durante todo su tiempo reglamentario de forma segura y en cumplimiento con la Política.

Como partimos de la premisa de que el Sistema de Recordkeeping Electrónico va a ser implantado por SIGDOC, cada entidad gubernamental es responsable de establecer el esquema de clasificación de acuerdo al tipo de documento que se está digitalizando.

De esta forma todas las entidades gubernamentales van a ir depositando los documentos para su retención y su eventual disposición siguiendo los lineamientos de la Política de Documentos.

3. CONTENIDO Y ALCANCE DE ESTA GUÍA

Esta Guía aborda la digitalización de documentos públicos para su conservación **dentro de un Sistema de Gestión Documental Central**, provisto por la Oficina del CIO, durante todo su período de retención.

Los aspectos que se abordarán en las secciones a continuación son los siguientes:

- Preparación del material para escaneo (Proceso de Planificación)
- Configuración del escáner desde el punto de vista del tipo de documento, resolución y calidad de imagen, control de calidad y compresión para textos, fotografías, mapas, material gráfico y documentos de valor histórico (Proceso de escaneo)
- Clasificación y metadatos (Proceso de Entrada de Datos)
- Formato de los documentos de acuerdo a su utilización y Preservación a través del Sistema de Gestión Documental (Proceso de Conservación de los datos)

Igualmente, es importante tener presente lo que no se incluye en esta Guía:

- Digitalización de vídeo y voz
- Recomendación de programas de manejo de contenido
- Recomendaciones de iniciativas de digitalización
- Arquitectura y plataforma de la red
- Digitalización de documentos en tránsito[12]

La prioridad de esta Guía es desarrollar una estrategia para cualquier iniciativa de digitalización en las entidades gubernamentales de forma efectiva y eficiente en donde se salvaguarde de forma centralizada la integridad, autenticidad, fiabilidad y uso práctico de esos documentos a través de todo el ciclo de vida. El propósito es asistir en el cumplimiento de la misión administrativa y programática de conservar los documentos públicos en armonía con los requisitos de las leyes, estatutos, regulaciones, estándares y mejores prácticas, para promover el acceso de la información entre la Rama Ejecutiva del Gobierno de Puerto Rico, instituciones públicas y privadas y el público en general.

PROPÓSITO

El propósito es el de asegurar que los récords públicos que van a ser mantenidos en formato electrónico, reemplazando el récord original, van a permanecer disponibles al público, entidades gubernamentales a través de todo su ciclo vital a través de un Sistema de Gestión Documental Central.

Esta guía provee los estándares que deben ser observados por las entidades gubernamentales cuando se lleva a cabo una iniciativa de digitalización de los documentos, especialmente cuando esa digitalización va dirigida a disponer de los récords públicos originales.

[12] Se define como *documentos en tránsito* aquellos que se escanean para completar un proceso y no para su conservación. Para propósitos de esta Guía se estará trabajando con aquellos documentos que se escanean para su conservación a través de su ciclo de vida.

BASE LEGAL

Esta Guía toma como punto de partida las siguientes leyes y reglamentos relacionados con la gestión documental:

1. la Ley Núm. 5, de 8 de diciembre de 1955, según enmendada, conocida como Ley de Administración de Documentos Públicos;
2. el Reglamento Núm. 4284, de 15 de noviembre de 1991, según enmendado, conocido como Reglamento para la Administración de Documentos Públicos de la Rama Ejecutiva del Estado Libre Asociado de Puerto Rico;[13]
3. la Ley Núm. 148, de 8 de agosto de 2006, conocida como Ley de Transacciones Electrónicas, según enmendada, entre otras, por la Ley Núm. 155, de 25 de octubre de 2010.[14]
4. la Ley Núm. 151, de 22 de junio de 2004, conocida como Ley de Gobierno Electrónico. Esta Ley faculta a la Oficina de Gerencia y Presupuesto a desarrollar lo guías, manuales y procedimientos para hacer valer la Ley.[15]

ALCANCE

1. Esta Guía provee los estándares para que las entidades gubernamentales puedan conservar su caudal documental a través de un Sistema de Gestión Documental Central.
2. Esta Guía es aplicable a todos los administradores de documentos públicos de la Rama Ejecutiva y Municipal y a los Departamentos de Sistemas de Información que son los responsables de proveer la infraestructura para llevar a cabo esta transición.
3. Esta Guía provee los estándares y mejores prácticas para todas las iniciativas de digitalización del gobierno de Puerto Rico.
4. Esta Guía provee los requerimientos mínimos para la digitalización de los documentos públicos con el propósito de disponer del original en formato físico.

---

[13] Una versión revisada de este Reglamento se encuentra en las etapas finales de aprobación y se espera que sea presentado ante el Departamento de Estado próximamente.

[14] La Ley Núm 155 de 2010 también derogó la Ley Núm. 359 de 16 de septiembre de 2004, conocida como Ley de Firmas Electrónicas de Puerto Rico.

[15] La Ley Núm. 151 establece como política pública la incorporación de las tecnologías de información a los procedimientos gubernamentales, a la prestación de servicios y a la difusión de información, mediante una estrategia enfocada en el ciudadano, orientada a la obtención de logros y que fomente activamente la innovación. También dispone que la Oficina de Gerencia y Presupuesto (OGP), al implantar la política pública, tendrá la encomienda de promover un acercamiento coordinado a las cuestiones que plantea la sociedad de la información y de facilitar que el acceso a la información y los servicios gubernamentales se ofrezca de manera armonizada con las disposiciones aplicables relativas a, entre otras, la protección de la privacidad, seguridad, políticas de disponibilidad de información y garantías de acceso a personas con impedimentos. Asimismo, la OGP tendrá a su cargo la evaluación y actualización de las Guías emitidas por el Comité del Gobernador sobre Sistemas de Información que rigen la adquisición e implantación de los sistemas, equipos y programas de información tecnológica para los organismos de la Rama Ejecutiva del Gobierno de Puerto Rico.

MOTIVO

En las últimas dos décadas, la Tecnología de Información y el acceso a los documentos públicos en formato electrónico ha tomado una relevancia mayor. Esto se debe a que la mayoría de los documentos que se generan o se reciben y los datos que se transfieren, se crean y conservan a través de medios electrónicos. Ante este nuevo paradigma, la Oficina del CIO es consciente de la necesidad de hacer una transición sistemática y ordenada de conservar los récords públicos de un formato físico a uno que sea digital.

Esta Guía establece los estándares que permiten viabilizar esa transición. La Oficina del CIO se dio a la tarea de crear el SIGDOC que permita que las entidades gubernamentales puedan conservar sus documentos públicos por el tiempo requerido bajo una misma plataforma. Para ello ha creado la taxonomía de clasificación de documentos que define los metadatos mínimos que va a ser mandatorio para los documentos que se van a conservar. Esto va a permitir que se pueda hacer búsquedas federadas de la información que ahí reside permitiendo un acceso más ágil y efectivo de la información.

Los requerimientos para la captura de la información, por parte de las entidades gubernamentales, y conservarlo dentro del SIGDOC por su tiempo reglamentario van a ser esbozados dentro de esta Guía. No es la intención del CIO de obligar a las entidades gubernamentales a que hagan la transición de papel a digital. Sin embargo, de querer las entidades hacer esa transición, deberán de seguir los guías que aquí se presentan.

GUÍA DE DIGITALIZACIÓN DE DOCUMENTOS

Esta sección de Digitalización de Documentos está dividida en las siguientes ocho secciones: (1) Planificación, (2) Funcionalidades del Sistema, (3) Metodología, (4) Proceso de Digitalización, (5) Clasificación de Datos, (6) Registro de Auditoría, (7) Migración al SIGDOC , y, (8) Validación y Verificación de Calidad. Cada una de estas áreas presenta un detalle de los aspectos más críticos y relevantes presentados, así como los requisitos mínimos a cumplir para garantizar que el proceso de la digitalización de documentos sea uno exitoso.

La Guía cubre todos los aspectos que se dan dentro de una iniciativa de digitalización sin limitarse al aspecto tecnológico. El aspecto tecnológico se da en función del alcance del proyecto y está condicionado por cuál es la finalidad última de ese caudal documental que se está digitalizando.

Tomando como premisa que las iniciativas de digitalización en las entidades gubernamentales van dirigidas a la conservación de los documentos en formato digital a través de su ciclo vital, los aspectos necesarios para el *recordkeeping* electrónico están incorporados en la presente Guía.

Como es el SIGDOC el que custodio de lo documentos que están siendo conservados, todos los elementos del *recordkeeping* electrónico recae en la Oficina del CIO. La responsabilidad de las entidades gubernamentales que entren en esta iniciativa es asegurarse que los procesos de captura, entrada de los metadatros y asignación de los períodos de retención se hagan de acuerdo a los lineamientos esbozados en esta guía.

1. PLANIFICACIÓN

Antes de entrar en una iniciativa de digitalización, las dependencias[16] tienen que atender varios aspectos para garantizar el éxito del proyecto. Entre los aspectos que se tienen que tomar en consideración se encuentran los siguientes:

- Planificación – tener los objetivos claros del proyecto
- Establecer los requerimientos para la adquisición de equipos y programas para llevar a cabo la iniciativa de digitalización
- Establecer la metodología para la captura de los documentos (captura)
- Establecer controles para medir la calidad del producto (imágenes)
- Establecer los mecanismos para la clasificación de los datos (metadatos)
- Migración al SIGDOC[17] (conservación y disposición)

El Proceso de Planificación es la parte vital de cualquier iniciativa de digitalización. Es en este proceso donde se definen el propósito, la metodología, los equipos y programas para la digitalización, el recurso humano interno y externo y el uso práctico de la información que se va a digitalizar. **La solución de escaneo a ser implantada dependerá del alcance del proyecto**. Es decir, que le tecnología de digitalización a ser adquirida para estos fines tiene que ir dirigida a cumplir con los objetivos del alcance.

Hay que tener en cuenta que la digitalización de los documentos no se da dentro de un vacío, sino que corresponde a un factor dentro de muchos que interaccionan para permitir el manejo efectivo de la información a través de todo su ciclo de vida dentro del SIGDOC. Entre los factores que interaccionan y se deben tomar en cuenta se encuentran los siguientes puntos:

- Definir el propósito y tener una visión clara del alcance del proyecto
- Establecer la metodología de trabajo
- Identificar los equipos y programas de captura a ser utilizados que permitan la entrada de metadatos
- Definir los recursos humanos internos y externos a ser utilizados
- Implantación del sistema (equipos/programas)
- Migración de la información digitalizada al SIGDOC
- Adiestramiento adecuado a los usuarios

OBJETIVOS DEL PROYECTO

Una vez se comienza el proyecto de digitalización y se determina el alcance del proyecto se define los objetivos específicos del proyecto basado en las siguientes preguntas:

- ¿Cuáles son los tipos de documentos que se van a estar digitalizando?

[16] Cuando hablamos de las entidades nos referimos al universo de Agencias, Municipio, Corporaciones y Juntas que caen bajo la jurisdicción del Programa de Administración de Documentos Públicos de la Rama Ejecutiva del Gobierno de Puerto Rico.

[17] Es importante señalar que al conservarse los documentos en el SIGDOC, todos los aspectos relacionados a la conservación y eventual disposición de los documentos van a ser responsabilidad de la Oficina del CIO. No obstante, en la captura de los datos, las agencias tienen que entrar los metadatos que establecen los períodos de retención.

- Documentos administrativos y/o fiscales
- Documentos de valor permanente
- Documentos de valor histórico[18]
- Documentos críticos y de valor legal
- Documentos / Récords Vitales
- Documentos no críticos

❖ ¿Cuánto es el volumen a ser digitalizado y el tiempo requerido?[19]

❖ ¿Cuál es la calidad de los documentos a ser digitalizados?[20]

- Están encuadernados
- Están en manuscrito
- Tipo de tinta
- Estado en que se encuentran
- Reproducción y/u original
- Tipo de tono
- Fotos (a color / blanco y negro)
- Mapas (estado de los mapas)

❖ ¿Se está trabajando con datos sensitivos?

❖ ¿Cuánto es el tiempo de retención de esos documentos?

Estas preguntas son necesarias para establecer los objetivos porque de ello depende cómo finalmente se configurará el equipo de digitalización (captura) y la interface de estos documentos al SIGDOC que es donde estarán conservando la información. Una vez definidos todos estos aspectos se comenzará con el proceso de captura[21]. A pesar de que existen diferentes formas de capturar la información, esta Guía únicamente se circunscribirá a la **digitalización por medio de un escáner**.

PASOS ESENCIALES DENTRO DEL PROCESO DE DIGITALIZACIÓN

Existen cuatro pasos esenciales dentro del Proceso de Digitalización. Estos se resumen en:

1. Preparación y depuración (*Prepping*) de los documentos a ser escaneados
   a. Limpiar los documentos (quitar las grapas, ponerlos en orden, alinearlos, entre otros)
   b. Fotocopiar aquellos que estén deteriorados
   c. Establecer niveles de clasificación de los datos de acuerdo al Esquema de Taxonomía de la Información[22]

[18] Todo documento que el Archivo General de Puerto Rico reclame y determine como tal.

[19] En muchos de los casos se utilizan servicios de Centros de Digitalización para la conversión a gran escala de documentos digitales.

[20] Refiérase al ANSI/AIIM TR33-1998.

[21] La captura, "establece una relación entre el récord, su creador y el contexto del negocio que lo originó, coloca el récord y sus relaciones dentro de un sistema de récords y lo vincula a otros." [ISO 15489-1:2001] Ese proceso puede hacerse por varias formas como lo son: digitalización de documentos, transferencias y radicaciones electrónicas, correo electrónicos entre otros.

[22] Una forma de hacerlo es a través de establecer sistemas de códigos de barra para identificar la forma en que se clasifican los datos.

   d. Ajustar los papeles y prepararlos para ser escaneados[23]
   e. Preparar los grupos de documentos a ser escaneados
   f. Primer Nivel de Control de Calidad (la entidad gubernamental establece los parámetros de esta prueba de calidad)
2. Escaneo de los documentos[24]
   a. Calibración y configuración del escáner
      i. Resolución y tamaño de la imagen
      ii. Profundidad o intensidad de color de la imagen (bitonal, tonalidades de gris, color)
      iii. Imagen en el monitor
      iv. Tamaño y orientación del papel
      v. Compresión y tamaño del archivo
      vi. Formato del archivo
   b. Segundo nivel de Control de Calidad – verificación de imagen con el original
3. Entrada de Datos
   a. Definición de los componentes o campos para la clasificación del documento
      i. Establecer los Metadatos de Acuerdo del Esquema de Taxonomía de clasificación de datos
      ii. Tercer Nivel de Control de Calidad para verificar la veracidad en la clasificación de las imágenes en el Sistema de Manejo de Contenido / Récord.
      iii. Verificación y validación de los resultados de las pruebas de calidad para la entrada de datos
4. Recompaginar los documentos escaneados

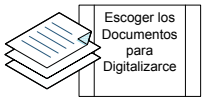
   El proceso de recompaginación de los documentos tiene que ser definido antes de comenzar el proceso de digitalización. En el caso de los documentos que vayan a ser digitalizados con el objetivo de reemplazar el original en formato físico, tiene que haberse establecido un proceso para la disposición adecuada de estos documentos.
   Para aquellos documentos que son de retención permanente, las agencias tienen que definir lo siguiente:
   a. Cómo se van a recompaginar
   b. Que van hacer con la copia original

[23] Luego de prepararlos, se pueden pasar a un *paper joggler*. Estas máquinas permiten que, por medio de un sistema de vibración, se elimine la estática a las hojas de los documentos; alinea los bordes para mejor agarre y los hace más sueltos para que en el momento que pasen por un escáner no se atasque la hoja.

[24] Dentro del proceso de escanear los documentos, se tiene que llevar un proceso de control de calidad por el cual aquellos documentos que están dentro de los estándares definidos se tienen que re-escanear.

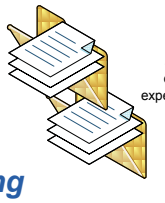
Figura 1 - Proceso de Digitalización

Proceso de Digitalización



Escoger los Documentos para Digitalizarce

Los documentos son escogidos por el cliente. La depuracion esta a cargo del cliente



Se preparan los documentos por expedientes o unidades documentales

Proceso de Preparación / *Prepping*

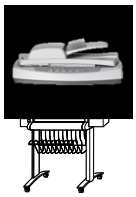
Se preparan los Documentos para ser escaneados

Proceso de Preparación:
Los documentos deben ser preparados de la siguiente forma:
1. Agruparlos en unidades y series documentales
2. Prepararlos de acuerdo a su clasificación.
   En algunos casos se insertan páginas con códigos de barras.
3. Dejarlos listo para ser escaneados.

Primer Nivel de Control de Calidad

Los supervisores del proceso por parte de la agencia tiene que

Proceso de Escaneo



Proceso de Escaneo:

- Se configura los escáner de acuerdo al tipo de documento (Refiérase a la Guía de Digitalización)
- Los documentos/planos/dibujos son escaneados.
- Los documentos escaneados son convertidos en TIFF/ PDF o JPEG
- Crear un *Master Image* en TIFF o PDF para ser desplazados en medios alternos como el Internet. La calidad de los archivos maestro debe mantenerse intacta.

Control de Calidad en procesar las Imagenes.

Segundo Nivel de Control de Calidad.

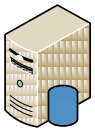
1. Ajustar las imágenes correctamente
2. Eliminar las Páginas en blanco
3. Documentar errores en las imágenes.

El operados hace el segundo nivel de control de calidad de las imágenes a ser escaneadas.
En el caso de planos o de documentos de alta precisión, se hace un 100% en control de calidad de las imágenes.

Proceso de Entrada de Meta Datos

Entrada de Datos (Manual / Automatizada)

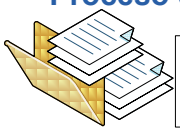
Proceso de Control de Calidad:
Las imágenes escaneadas son clasificadas usando el Esquema de Taxonomia de clasificacion de Metadatos.
Una vez entrado los metadatos, se migra al SIGDOC de la Oficina del CIO.



Tercer nivel de Control de Calidad

1. Se hace un muestra, no menor de un 25% para determinar la validad de la información entrada.
2. Documentar el Proceso

Proceso de Recompaginación



Recompaginar los expediente

Los documentos son recompaginados bajo la supervision de un agente autorizado.
Los empleados involucrados en el proceso de digitalizacion de expedientes, no pueden ser parte de la disposicion de los mismos..

PROCESOS Y TAREAS DE LOS CUATRO PASOS PARA LA DIGITALIZACIÓN DE DOCUMENTOS

Proceso	Tareas	Control de Calidad	Resultado Esperado	Responsabilidad
Proceso de Preparación (*prepping*)	❖ Planificación – tener los objetivos claros del proyecto ❖ Establecer la metodología para la captura de los documentos (captura) ❖ Establecer controles para medir la calidad del producto (imágenes) ❖ Establecer los mecanismos para la clasificación de los datos (metadatos) ❖ Establecer un método de indexación para la búsqueda efectiva de los documentos y que se puedan presentar como unidades documentales ❖ Establecer los mecanismos para mantener un registro de auditoría para los documentos a través de su ciclo vital ❖ Establecer el resguardo (tiempo que se van a guardar los documentos) ○ Conservación de la información a través de su ciclo vital[25] ○ Resguardo (*backup*) / archivo (corto / mediano / largo plazo) ○ Plan de Migración de Datos para evitar obsolescencia en los equipos ❖ Determinar la forma de disponer los documentos	Preparación y depuración (***Prepping***) de los documentos a ser escaneados 1. Limpiar los documentos (quitar las grapas, ponerlos en orden, alinearlos, entre otros) 2. Fotocopiar aquellos que estén deteriorados 3. Establecer niveles de clasificación de los datos 4. Ajustar los papeles y prepararlos para ser escaneados 5. Preparar los grupos de documentos a ser escaneados	Documentos listos para ser escaneados.	Oficial autorizado de la Agencia Equipo de Digitalización

[25] Es fundamental que los procedimientos implantados en cualquier momento durante el ciclo vital del documento de cualquier pieza de información puedan ser determinados de forma precisa. De lo contrario, compromete su integridad y pierde su fiabilidad – ISO/PDTR 15801.

Proceso	Tareas	Control de Calidad	Resultado Esperado	Responsabilidad
Proceso de escaneo	❖ Configuración de los escáneres 1. Resolución y tamaño de la imagen 2. Profundidad o intensidad de color de la imagen Tamaño y orientación del papel 3. Compresión y tamaño del archivo 4. Formato del archivo ❖ Escanear los documentos de acuerdo a la configuración establecida.	Verificar la calidad de las imágenes cotejando: ❖ la imagen digitalizada contra la original ❖ la legibilidad del documento ❖ la veracidad del documento (Si recogió los detalles de la imagen) ❖ la precisión en las dimensiones de la imagen vs. la original ❖ la densidad de las áreas ❖ la fidelidad de los colores ❖ la Digitalización de todas las áreas de la imagen (características de la imagen) % de cotejo de las imágenes de acuerdo al tipo de documento. ❖ El Control de Calidad de los documentos fiscales / administrativos debe ser mayor de un 20% del total escaneado. ❖ El Control de Calidad de los documentos categorizados como sensitivos, cuasi-permanentes, permanente e históricos, debe ser del 100% de las imágenes.	1. Crear un *Master Image* en TIFF o PDF para ser desplazados en medios alternos derivados como JPEG / PDF (comprimidos) / GIF el Internet. 2. Crear una lista de los resultados de las imágenes escaneadas. 3. El Archivo Maestro (*master copy*) debe ser inalterado.	Responsabilidad del Equipo de Digitalización Las agencias deberán crear un mecanismo para auditar periódicamente el proceso.

Proceso	Tareas	Control de Calidad	Resultado Esperado	Responsabilidad
Oficialización de los documentos	❖ Establecer una metodología para comparar el contenido, contexto y estructura del récord con el que se utilizó originalmente de modo que se puedan identificar errores, corregirlos y validar el récord. ❖ Seguir el proceso para la oficialización de los datos. Esos procesos deberán tener en cuenta los siguientes aspectos: o capturar el documento como récord público; o establecer el metadato o la clasificación; o conservar el documento por el tiempo que sea requerido; o salvaguardar la veracidad del documento; y o disponer del documento una vez acabado su período de retención. ❖ categorizar los documentos para preparar el inventario; ❖ establecer el período de retención; ❖ registrar cuándo se originó o capturó la información (*date and time stamp*); ❖ preparar el inventario de documentos; ❖ mantener el control de la disposición de documentos; ❖ audita el proceso de disponer documentos			

Proceso	Tareas	Control de Calidad	Resultado Esperado	Responsabilidad
Proceso de Entrada de Metadatos (clasificación)	Previo a comenzar este proceso entrada de datos, se debe haber determinado cómo van a migrar los documentos capturados al SIGDOC. Esto puede hacerse de tres formas: 1. Entrando la data directamente al SIGDOC 2. Configurando un sistema como el SIGDOC localmente y enviando la base de datos por medio de un Interface **Configuración del Sistema de Captura de Metadatos.** La configuración del Sistema de Captura que va a permitir la entrada de Metadatos debe tener en consideración los siguientes aspectos: 1. Método de Captura de la imagen. 2. Esquemas de Metadatos para la clasificación de las series documentales. 3. Métodos para el Control de Calidad que permita verificar las imágenes escaneadas. Entrada de Datos 1. Separa la entrada de datos en grupos (*batch*) para poder sacar una muestra para el control de calidad. 2. Comparar los datos contra el expediente. 3. Corregir los errores	1. Se hace un muestra, no menor de un 25% para determinar la validad de la información entrada. 2. Documentar el Proceso	Imágenes escaneadas y debidamente clasificadas.	Responsabilidad del equipo de digitalización bajo la supervisión de un oficial autorizado de la agencia.

Proceso	Tareas	Control de Calidad	Resultado Esperado	Responsabilidad
Proceso de Re-compaginar	El método mediante el cual se van a recompaginar los documentos ya digitalizados debe ser establecido por las entidades gubernamentales. El equipo de digitalización ***NO*** está autorizado a disponer de ningún expediente.			

No se puede perder de perspectiva que la relevancia y sensitividad de los documentos van a determinar qué tipo de preparación se le debe asignar. Por lo tanto, el cuidado y eficacia en la preparación de los documentos para ser digitalizados en gran medida ayuda en el resultado de la calidad en las imágenes. Por ejemplo, el cuidado y especificaciones de escaneo en los documentos de carácter histórico no van a ser lo mismo que un documento fiscal.

No obstante, hay que reconocer que el éxito de esta implantación proviene fundamentalmente de los usuarios que van a servirse del sistema. Por lo tanto, el aspecto operacional y de Recursos Humanos es fundamental para garantizar su éxito. Factores como el adiestramiento adecuado, la comprensión y aplicabilidad de la Política de Documentos juegan una parte crítica en la Implantación del Sistema.

2. FUNCIONALIDADES DEL SISTEMA

En la creación y producción de los documentos se tienen que considerar el equipo y los programas (Hardware/Software) y el volumen de los documentos a ser digitalizados. Si se está hablando de series documentales voluminosas, entonces se deben tener en cuenta factores como:

- El escáner debe ser apropiado para el volumen de documentos
- Permitir la configuración basada en los criterios que se definieron para cada tipo de documento
- Configuración mínima de la computadora asignada a la digitalización
- Crear los niveles de calidad para cada etapa del proceso de digitalización
- El tipo de resguardo que se va a utilizar para almacenar la información
- Programación que maneje la información (estructurada / no estructurada)[26]
- Sistema de entrada de metadatos siguiendo el Esquema de Taxonomía de clasificación de Documentos
- Capacidad de digitalización (Imagen y reconocimiento óptico de caracteres)

CONSIDERACIONES

En el caso de que existan muchos documentos para entrar en el sistema, una alternativa sería contratar los servicios de un *Centro de Digitalización* o también se puede habilitar un área para escanear y clasificar los documentos/récords. Independientemente del proceso que se utilice, las áreas de trabajo y la operación van a ser impactadas.

En el caso que se determine digitalizar documentos para integrarlos como parte de un sistema preestablecido, tendrán que seguir la política y los procedimientos establecidos para el manejo de los documentos. Si estos documentos se oficializan como récords y se van a conservar a través de su período de retención, el sistema debe contemplar el *Recordkeeping*[27] electrónico.

Todas estas consideraciones deben ser resueltas durante la etapa de planificación.

---

[26] Se define información estructurada aquella que proviene de tablas como Excel o de Bases de Datos. La no estructura se define aquella información que se entra sin estructura fija como lo es el procesador de palabra.

[27] Cuando hablamos de recordkeeping electrónico nos referimos las guías y procedimientos que permitan el manejo efectivo de los récords que se generan de forma electrónica. Ese mecanismo debe estar amparado en la Política

USO PRÁCTICO

El uso práctico de un documento se determina por la necesidad de acceso y la rapidez que se requiere de éste. Es por esta razón que la clasificación implementada y los metadatos asignados van a ser cruciales para acceder los documentos eficientemente. Como el objetivo primordial del uso práctico en un documento es su accesibilidad de forma segura, la idea de crear un SIGDOC proporcionaría un medio centralizado para hacer búsquedas efectivas, bien sean simples o federadas.

3. METODOLOGÍA

En la preparación de esta Guía se han utilizado como referencia un marco regulatorio (leyes, regulaciones, estatutos), estándares y mejores prácticas según aplican a la digitalización y al manejo de documentos electrónicos, además de aquellas que aplican al Programa de Administración de Documentos de la Rama Ejecutiva del Estado Libre Asociado de Puerto Rico.

En el proceso de desarrollarlas se verificaron varias guías existentes para utilizarlas como modelo de la Guía.[28] En realidad, lo que se pretendía era tomar el conocimiento práctico de entidades que habían desarrollado políticas similares para llevar a cabo iniciativas de digitalización. Obviamente, cada Guía enfocaba los aspectos relevantes dependiendo de cuál era el alcance y sus estrategias de digitalización.

Tabla 1 - Estándares utilizados en la Guía de Digitalización de Documentos

Estándares utilizados en la Guía de Digitalización de Documentos	
DoD 5015.2	DoD 5015.02-STD, "Electronic Records Management Software Applications Design Criteria Standard."
ISO 10196:2003	Document Imaging Recommendation for the creation of original records Recommended minimal font dimension for scanning (p. 11) – Scanned Resolution
ISO 15489	Information and documentation - Records Management

[28] North Carolina ECHO, Western States Digital Standards Group Digital Imaging Best Practices, Queensland State Archives: Guidelines for Digitisation, Digital Preservation Coalition, Universidad de Cornell / Departamento de Preservación y Conservación, US National Archives and Records Administration (NARA), entre otros.

Estándares utilizados en la Guía de Digitalización de Documentos	
ISO PDTR 15801	Electronic Imaging 1. Information Capture (p. 10) 2. Scanning process and metadata (p. 11) 3. Quality Control 4. Checking scanner performance (p. 16) 5. Data Capture (p. 17) 6. Migration (p. 17) 7. Indexing (p. 18) 8. Authenticated output procedure (p. 19) 9. Document Retention (p. 21) 10. Audit Trail (pp. 34-36) 11. Migration and conversion (p. 37) 12. Information Capture (p. 38)
ISO PDF 1.7 ISO 19005-1:2005(E)	Electronic Document File Format for Long Term Preservation 1. Metadata Scheme (p. 14) PDA/A Retention Scheme
ANSI /AIIM MS52-1991	Recommended Practice for the Requirements and Characteristics of Document Intended for Optical Scanning 1. Paper Size (p. 2) 2. Paper Orientation (p. 3) 3. OCR (Printed Document) (p. 4) 4. OCR – Recommendations (p. 5)

Estándares utilizados en la Guía de Digitalización de Documentos	
ANSI/AIIM TR15-1997	Planning Considerations Addressing Preparation of documents for Image Capture 1. Recommended National Standards for Scanning (pp. 1-2) 2. Planning Considerations (p. 3) 3. Physical characteristics of the paper document (p. 4) 4. Paper orientation (p. 5) 5. Paper size (p. 5) 6. Quality (contrast) 5.1.5 (p. 7) 7. Document Characteristics a. Binding (p. 7) b. Opacity (document handling) (p. 7) c. Color (p. 8) d. Embossing (p. 8) 8. Audit Trail 9. Document Preparation (pp. 10-11)
AIIM TR26 – 1993	Resolution as it Relates to Photographic and Electronic Imaging 1. Resolution (Contrast / Granularity) (pp. 3-5) 2. Quality Index for Digital Imaging (p. 12) 3. Digital Resolution vs. Image size (p. 13)
ANSI/AIIM TR33-1998	Selecting an Appropriate Image Compression Method to Match User Requirements 1. Description of Documents (p. 5) a. Types of Documents b. Document Content (p. 6) 2. Rare or Historical documents (p. 8) 3. Legal and Critical Documents (p. 8) 4. Non Critical Documents (p. 9) 5. Image Presentation (p. 10) 6. Table of compression methodology based on type of documents (pp. 11-12)

Estándares utilizados en la Guía de Digitalización de Documentos	
International Standards Organization *Information and Documentation – Records Management – Part 1: General (ISO 15489:1)*	Define los requerimientos del manejo estructurado de los récords en forma general (*high level*). La parte uno enfatiza la importancia y los beneficios del manejo apropiado de los récords, sus consideraciones reglamentarias y la importancia de asignar responsabilidad para asegurar su manejo efectivo. Entre los aspectos relevantes está definir claramente en el diseño, y luego en la implantación, los procesos de captura, manejo, acceso, conservación y disposición de la información. Fundamentalmente, el objetivo de este estándar es establecer un sistema que permita que los récords se manejen de forma confiable, estén disponibles cuando se requieran a través de su ciclo de vida y que evidencie o registre una actividad en la dependencia.
International Standards Organization *Information and Documentation – Records Management – Part 2: Guidelines 15489:2*	En esta sección se describe y se explica en detalle los requerimientos generales expuestos en la primera parte. Es decir, su propósito es ser utilizado como herramienta en la implantación de los procesos y controles de *recordkeeping* prestando particular atención a lo relacionado con el manejo de datos, la implantación de medidas de seguridad y el desarrollo de un esquema de clasificación de la información. Todas estas características, enmarcadas dentro de guías de auditoría y programas de adiestramiento, sirven para promover e implementar efectivamente el manejo de los récords dentro de la dependencia.

4. PROCESO DE DIGITALIZACIÓN DE DOCUMENTOS

CONFIGURACIÓN DEL ESCÁNER

Una vez el alcance del proyecto esté claramente definido y se hayan adquirido el equipo y los programas necesarios para llevar a cabo la digitalización, es cuando se comienzan a configurar los equipos de acuerdo a los documentos que vayan a ser digitalizados.

Existen varios elementos que se tienen que considerar previo a hacer la captura de papel a imagen. Entre estos factores se encuentran:[29]

- Tamaño y peso
- Orientación en el tipo de retrato y tipo de paisaje (*Portrait / Landscape*)
- Calidad y estado físico de los documentos
- Rasgos únicos
- Encuadernados
- Nivel de opacidad
- Color

[29] Refiérase a las características físicas de los documentos en el estándar ANSI/AIIM TR15 – 1997.

- Elaboración / adornos (*embossing*)

Dependiendo de cómo se defina cada elemento es la forma en que se configurará el/los escáner(es).

CONFIGURACIÓN BÁSICA

Es importante señalar que la configuración de los equipos depende del formato y la expectativa del proyecto que se esté desarrollando. Por lo tanto, ese proceso de configuración se llevará a cabo cada vez que se inicie un proyecto de digitalización.

A pesar de que existe una variedad de proyectos e iniciativas de digitalización, ciertamente todas tienen que configurar los mismos parámetros para los escáneres y las estaciones de trabajo. Estas son:

- Resolución y tamaño de la imagen
- Profundidad / intensidad de colores
- Imagen en el monitor
- Orientación del papel
- Formato de la imagen (TIFF/PDF/JPEG/JPEG2000)
- Compresión y tamaño del archivo
- Reconocimiento óptico de caracteres (“OCR”, por sus siglas en inglés)

Los componentes que permiten un buen manejo del contraste y de la nitidez y/o claridad de la imagen son parte del mecanismo interno que utiliza el escáner para manejar efectivamente las imágenes.

RESOLUCIÓN Y TAMAÑO DE LA IMAGEN

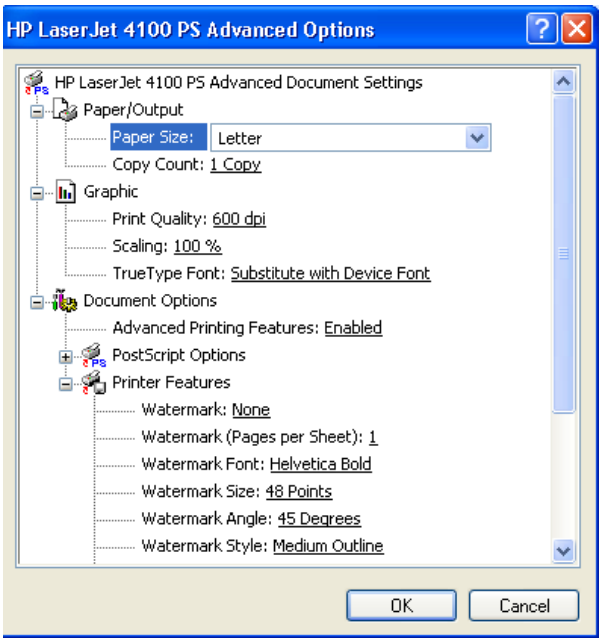
La resolución de una imagen se determina por la densidad de píxeles que se organizan para conformar la imagen. Este término se utiliza mayormente cuando se habla de imagen y no de impresión de la imagen. Como la resolución no se puede incrementar una vez haya sido digitalizada, es fundamental que se tome este factor en consideración a la hora de determinar la resolución. Sin embargo, para propósitos de calidad en la resolución de un documento que se desea preservar, existen varios factores que se tienen que tomar en consideración:

- Tiempo estipulado para la retención del documento
- Uso práctico por usuarios
- Publicación en otros medios, como el Internet
- Equipo disponible para acceder las imágenes

La forma en que se define el tamaño de una imagen puede ser medida de varias maneras: mediante la definición del número de puntos que existe por pulgada o DPI (Dots Per Inch), la concentración de píxeles por pulgada o PPI[30] (Pixel Per Inch) o el número de líneas por pulgada o LPI (Line Per Inch). Sin

[30] Un *elemento de una imagen* o *píxel* se define como células pequeñas de un color que se organizan en patrones para formar una imagen. A mayor concentración de píxel, mayor va a ser la resolución de la imagen. Definición de (NCECHO).

embargo, para propósitos de esta Guía[31] se utilizarán ambas medidas, DPI (Dot per Inch) o PPI (Pixel per Inch), para hacer referencia a la resolución y/o al tamaño de una imagen en la pantalla. A pesar de que ambos términos, DPI y PPI, se usan para medir el tamaño de la imagen, el DPI se utiliza mayormente para referirse a un documento, bien sea gráfico o de texto, y su impresión. El término PPI se utiliza mayormente para efectos de una imagen porque hace referencia a su resolución en la pantalla.



Este punto es importante cuando se digitaliza porque los factores de resolución e impresión tienen que tomarse en consideración a la hora de determinar la configuración final en un escáner. Dado que la resolución en la impresión se mide típicamente en DPI, el imprimir un texto a 600 DPI no implicaría un esfuerzo en la impresión pero, al mostrar esa imagen a tal resolución en pantalla, resultaría muy cargado. Por ende, se optará por tomar la regla de escanear a la mitad de la resolución que se utilizaría en la impresión. Es decir, a mayor resolución mayor va a ser el tamaño de los archivos porque requerirá una densidad mayor de DPI para la imagen.

PROFUNDIDAD O INTENSIDAD DE COLOR

Otro factor a considerar es el color o tonalidad de la imagen. Este color puede ir desde blanco y negro, usado mayormente en texto, hasta imagen a color. Existen tres tonalidades que se utilizan de acuerdo a la intensidad del *bit* (*BIT-Depth* en adelante, “profundidad de color”). Estas son: bi-tonal, escala de gris y color real.

[31] En muchos documentos el PPI y el DPI se utilizan como sinónimos. El PPI se establece mayormente cuando se habla de resolución en la impresión, mientras que le DPI se utiliza para la resolución en la pantalla.

A su vez, cada una de estas categorías se subdivide en otras subcategorías dependiendo de cómo se desea la profundidad o intensidad de los colores del documento a ser digitalizado.

Cuando se habla de la intensidad del *bit*[32] ello se refiere a la medida que se utiliza para medir el color o las tonalidades de gris que existen dentro de una imagen.

Bitonal (1-bit)

Se representa con dos colores, blanco o negro. Un color es el que va en el trasfondo mientras que el otro va en el primer plano (*foreground*). Usualmente se utiliza para documentos que son texto y que no contienen imágenes o formas y/o sombras con diferentes tonalidades de gris.

Paleta de Colores (4-*bit*: 16 colores / escalas de gris y 8-*bit*: 256 colores / escala de gris)

Esta imagen se reconoce porque a cada píxel en la imagen se le asigna un valor que se relaciona a un color específico en la paleta de colores. Es decir, la correlación es de uno a uno. Este *BIT-Depth* no debe ser utilizado en imágenes de color, como son las fotos, porque la resolución sería muy pobre.

La escala de gris de una imagen 8-*bit* es utilizada en textos que contienen alguna imagen en blanco y negro como son fotos en blanco y negro, ilustraciones a medio tono, y cualquier otra ilustración que no conlleve colores.

Color (16-bit: 65536 colores / escala de gris)

Se utiliza para imágenes que contengan más de 256 diferentes tonalidades de gris. Estas imágenes pueden ser imágenes médicas (*record imaging*) que tengan que incrementar el tamaño.

El color 16-bit puede ser interpretado como un híbrido entre la p*aleta de colores* y el c*olor real* que se da en una profundidad mayor de *BIT-Depth*. El color 16-bit se utiliza mayormente en vídeos y aplicaciones que conllevan animación y que la resolución y nitidez de la proyección en pantalla no tiene que ser tan definida.

Color real (24-bit: 16.7 millones de colores and 48-bit: billones de colores)

A la hora de digitalizar imágenes y fotografías en pleno color (full color), se deben utilizar imágenes de 24-bit. Las intensidades de color real se recomiendan para cualquier material con color donde el color **transmita** información esencial. Para fotografías a color, la profundidad de color mínima recomendada es de 24-bit (color real). El color 32-bit es un color 24-bit junto con un canal 8-bit adicional que provee 256 niveles de transparencia y se utiliza mayormente para vídeo digital y aplicaciones animadas.

---

[32] Un *bit* es la unidad más pequeña de información en una computadora la cual se representa en dos valores, ya sea 0 ó 1. Por su parte, *BIT-Depth* se refiere al número de *bits* necesarios para describir un color en un píxel. Mientras mayor sea el *BIT-Depth*, mayor será la profundidad en el color o la tonalidad de gris que se representa en un píxel. Esto es lo que se conoce como *resolución*.

Tabla 2 - Tipo de Documento y Profundidad de color

Tipo de Documento y Profundidad de color	
Tipo de Documento	**Profundidad de color**
Texto en blanco y negro solamente	1-bit bi-tonal
Texto con un poco de color	8-bit color
Texto con sombras grises	8-bit tonalidad gris
Dibujos a color/ Presentaciones/ Gráficas	8-bit color
Fotografías en blanco y negro	8-bit tonalidad gris
Fotografías a color	24-bit color real

Ref: Queensland State Archives: Guideline for the Digitisation of Paper Records

La configuración en la tonalidad del escáner depende del tipo de documento que se esté escaneando. Es importante señalar que a mayor escala, mayor va a ser el tamaño de la imagen que se digitalice y mayor será el tiempo que va a requerir el escanear y cargar la imagen en pantalla para su utilización.

Tabla 3 - Profundidad de color y Resolución según los tipos de documentos

Profundidad de color y Resolución según los tipos de documentos				
Escanear	**Profundidad de color**	**Resolución**	**Tipos de Documentos**	**Uso**
Bi-tonal Blanco y negro sólo texto Hasta A3	1 BIT	300 dpi (Mínimo) 400 dpi (se recomienda cuando se va a utilizar OCR)	Incluye los siguiente: 1. Texto con ilustraciones o fotografía en blanco y negro. 2. Documentos administrativos fiscales. 3. Expedientes que se digitalizan para la conservación por un tiempo reglamentario.[33] 4. Manuscritos o copias de pobre legibilidad.	Se utiliza en texto que no contiene imagen con color o de tonalidades grises. Cada píxel es blanco o negro. Usualmente se procesa más rápido.
Escala de gris (*grayscale*) (Texto con algún color o tonalidades de gris, mapas, etc.) Más de A3	8 BIT	300 dpi mínimo 400 (se recomienda cuando se requiere OCR)	1. Texto con ilustraciones o fotografía con tonalidades grises. 2. Documentos que son bitonales, pero con el tiempo se desgastan y adquieren un color que no es su original. Ejemplo de esto podría ser los documentos o fotos en sepia, mapas hechos en *blueprints*, periódicos viejos, entre otros.	Cada píxel puede variar de 1 a 256 tonalidades de gris. Se utiliza el texto que contiene fotos en blanco y negro.

[33] Por lo general, se refiere a las series documentales en donde cada expediente se digitaliza en un solo documento en formato PDF o TIFF.

Profundidad de color y Resolución según los tipos de documentos				
Escanear	**Profundidad de color**	**Resolución**	**Tipos de Documentos**	**Uso**
Color real (Gráfica, fotos a color, texto, opúsculo (*brochure*), etc.)	8 BIT a 24 BIT Dependiendo de la calidad en los colores necesaria	6”x4” - 600 ppi 7”x5” - 450 ppi 9”x6” - 300 ppi	1. Fotos o ilustraciones a colores. 2. Documentos que contienen colores 3. *Brochures,* presentaciones oficiales a color que se requiere preservar.	La medida de 8-bit se utiliza para definir un píxel que va desde 1 a 256 tonalidades de gris o de color. El 24-bit, por consiguiente, se utiliza mayormente para definir tonalidades de color que van desde 1 a 16.8 millones de posibilidades. Esta escala se utiliza usualmente cuando se trabaja con fotos o imágenes a color.

Referencias utilizadas: (1) Queensland State Archives: Guideline for the Digitisation of Paper Records y North Carolina ECHO. (2) Tutorial de Digitalización de Imágenes - © 2000-2003 Biblioteca de la Universidad de Cornell / Departamento de Preservación y Conservación. © Biblioteca de la Universidad de Cornell / Departamento de Preservación y Conservación

IMAGEN EN EL MONITOR

La resolución es un factor importante en la digitalización de los documentos. Es importante recordar que, a mayor resolución, mayor va a ser el tamaño del documento y por lo tanto el tiempo de accederlo en pantalla. Otro factor que debe ser considerado es el almacenaje de los datos.

Para los documentos que requieran ser conservados por un período reglamentario es recomendable que haga una copia maestra con los requisitos de configuración que se requiera de acuerdo al tipo de documentos. De ser requerido que estos documentos sean accedidos regularmente por medio de la Red (interna o externa) o por el Internet, se puede utilizar un medio alterno de acceso para disminuir el tamaño de la imagen reduciendo los requisitos de configuración.

Típicamente, aquellos documentos que se digitalizan para hacerlos disponibles, bien sea en el Internet o en línea (*on-line*), van a requerir una resolución menor para que el acceso sea más ágil y eficiente. Como regla general, en la digitalización de un documento se debe utilizar la mejor resolución y/o profundidad de color que se determina para el tipo de documento, de modo que se tome ventaja de nuevas tecnologías. Digitalizar los documentos en una profundidad de color por debajo de lo recomendado puede resultar en que la imagen no sea igual a la original.

Sin embargo, digitalizar un documento a una resolución mayor de la recomendada sólo se justifica si esa imagen va a ser utilizada para ser agrandada o para capturar una imagen que requiera de mucho detalle. De lo contrario, resultaría muy oneroso en cuanto al espacio y tiempo requeridos en proyectar la imagen en la pantalla.

La resolución del monitor puede variar con el tipo de monitor y tarjeta de vídeo que se utilicen. Esta resolución puede alterar la calidad de la imagen digitalizada que se proyecta en pantalla. Las diferentes resoluciones de monitor son:

Imagen en el monitor		
Monitor	**Núm. de píxeles x Núm. de líneas**	**Calidad**
VGA	640 x 480	Baja
SVGA	800 x 600	Mediana
XGA	1024 x 768	Alta
SXGA	1280 x 1024	Alta
SXGA	1440 x 900	Alta

Ref.: North Carolina ECHO

Los monitores deben estar calibrados para obtener mejores resultados. Los paquetes de software, tales como *Adobe Photoshop*, suelen incluir una herramienta básica para la calibración del monitor.

TAMAÑO Y ORIENTACIÓN DEL PAPEL

A la hora de digitalizar documentos se debe establecer la orientación del papel.[34] La orientación que mayormente se utiliza es la del modo vertical (*portrait*) en la cual la mayoría de los documentos 8.5” x 11” se imprime. En estos documentos la línea de impresión es paralela al lado más corto del documento. Otra forma de orientación en los documentos es la forma paisajista (*landscape*). En esta forma, la horizontal, la línea de impresión es paralela al lado más largo de la página.[35]

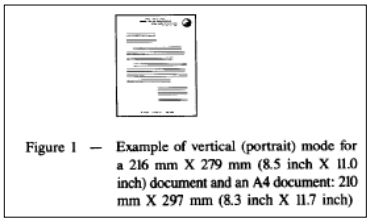


Figure 1 — Example of vertical (portrait) mode for a 216 mm X 279 mm (8.5 inch X 11.0 inch) document and an A4 document: 210 mm X 297 mm (8.3 inch X 11.7 inch)

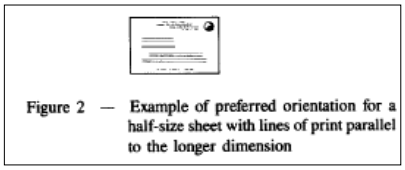


Figure 2 — Example of preferred orientation for a half-size sheet with lines of print parallel to the longer dimension

[34] Para determinar la orientación puede referirse al estándar ANSI/AIIM TR2.

[35] ANSI/AIIM MS52-1991, página 3.

Los tamaños estándares de papel más comunes se describen en la tabla siguiente. La mayoría de los escáneres están preparados para aceptar esta configuración sin problemas. A modo de referencia, se ha incluido en el anejo de la Guía una tabla más detallada de los tamaños de papel existentes. Sin embargo, existen tamaños de papel que son irregulares, como pueden ser los cheques y diferentes formatos de órdenes de compras, facturas y recibos de pagos. Las técnicas que se utilizan para reducir estos documentos y capturar la información relevante están definidas en el estándar ANSI/AIIM MS32.

Tamaños estándares de papel		
Tamaño	**Formato**	**Ejemplo de su uso**
Tamaño carta	8.5 x 11	Cartas, memos, informes administrativos, correspondencia, copias, documentos oficiales
Tamaño legal	8.5 x 14	Contratos, gráficas, informes
Tabloide	11 x 17	Periódico
Tamaño medio	8.5 x 5.5	Talonario de cheques, recibo de compra, facturas
A4	8.3 x 11.7	
Tres tamaños (*three part paper*)	8.5 x 3.5 en algunos casos	Talonario de nómina, recibos

Ref: ANSI/AIIM MS-52

COMPRESIÓN Y TAMAÑO DEL ARCHIVO

Existen diferentes métodos para la compresión de documentos que pueden variar de acuerdo a la aplicación que se utilice. Se puede definir la compresión como el método de reducir la cantidad de datos necesarios para resguardar o transmitir la reproducción de una imagen.[36] Idealmente, cuando la data se comprime usando un algoritmo que maneja eficientemente la codificación de la información de la imagen y su reconstrucción sin perder datos, se puede decir que es fiel al original. Este proceso se conoce como *lossless*. Ahora bien, cuando el algoritmo de compresión elimina selectivamente información de la imagen, el proceso se llama *lossy*. En este proceso, la reconstrucción de la imagen será imprecisa porque hay pérdida de información[37].

[36] Refiérase a ANSI/AIIM TR33-1998 “Selecting an Appropriate Image Compression Method to Match User Requirements”.

[37] El método *lossy* se utiliza para diseminar información en tránsito.

Esta Guía va a estar dirigida a los documentos que se digitalizan para propósitos de conservarlos durante todo su ciclo de vida. Por lo tanto, el método de compresión de los documentos que van a conservarse tiene que permitir la conservación exacta de la imagen sin pérdida de información o degradación de resolución. Esto se debe a que, al comprimir un archivo y haber pérdida de información, la resolución y la calidad de la imagen se afectan, por lo que se comprometería la integridad del documento.

El número de píxeles que utiliza una imagen es lo que determina el tamaño de un archivo (*file*). La intensidad en cada color (*color depth*) de la imagen va a tener un efecto multiplicador en el número de píxeles necesarios para componer la imagen. Por lo tanto, si la imagen se comprime, el número de píxeles va a ser menor, lo cual afectaría la calidad de la imagen.

Si la imagen que va a ser digitalizada tiene como fin diseminarla a través de medios electrónicos en donde no se utilice para reemplazar la original, entonces comprimir la foto usando el método *lossy* sería una opción porque al archivo ser menor en el número de *bits*, permitiría proyectar la imagen en pantalla más rápidamente.

En la tabla siguiente[38] se presenta el total de *bits* que comprendería una imagen no comprimida por cada tipo de *BIT-Depth*. Por lo tanto, es importante establecer un método para la compresión que, aunque reduzca el tamaño de los archivos, no pierda información de la imagen y no comprometa su calidad.

Profundidad de color, resolución y tamaño del archivo			
Profundidad de color	**Resolución (ppi)**	**Total de bits**	**Tamaño del archivo no comprimido (Mb)**
1 bit bi-tonal	**300**	**8,700,867**	**1.04**
1 bit bi-tonal	600	34,803,468	4.15
8 bit gris o color	300	69,606,936	8.30
8 bit gris o color	600	278,427,744	34.00
24 bit color	300	208,820,808	24.89
24 bit color	600	835,283,232	101.96

Ref: Queensland State Archives: Guideline for the Digitisation of Paper Records

Las ventajas en comprimir la información es que se reduce el tamaño de espacio, el tiempo de resguardar la información y, por consiguiente, la información puede ser distribuida a través de la Red/Internet más rápida y eficientemente. En resumen, es recomendable aplicar algún tipo de compresión a las imágenes que van a ser digitalizadas porque permite acceder y resguardarlas más

[38] Queensland State Archives: Guideline for the Digitisation of Paper Records.

eficientemente. Si se desea comprimir sin perder información tendría que utilizarse un método como *lossless*.[39]

Sin embargo, para aquellos documentos que no van a ser conservados, o que son transitorios, se podría utilizar el método *lossy*[40]. El tamaño del archivo es menor, lo cual permite que se despliegue en la pantalla más rápido. La particularidad de este método es que la imagen luce casi igual a la original, pero los valores del número de píxeles varían. Es decir, se pierde información en el proceso.

En resumen, seleccionar una metodología de compresión depende de la aplicación que se esté utilizando. Es necesario que se tomen en consideración los requerimientos de los usuarios, cuál es el objetivo final y cuáles son sus rasgos particulares.

Se recomienda hacer la copia maestra sin compresión. Los otros métodos alternos de imagen, que proviene de esta copia maestra, se puede comprimir para ser desplegado a medios como el Internet[41].

FORMATO DE LOS ARCHIVOS

La configuración de las imágenes, relacionada con el formato, resolución y compresión, depende de sus respectivos períodos de retención. Típicamente, cuando se decide emprender una iniciativa de digitalización, ésta se basa ya sea en la necesidad de acceder los récords de forma electrónica, establecer sistemas mecanizados en los cuales los documentos se generen electrónicamente o economizar el espacio físico. Cualesquiera que sean las razones para entrar en estas iniciativas, la información tiene que ser clasificada (metadatos), resguardada (*storage*) por el tiempo requerido, protegida y de fácil acceso cuando sea requerida.

Existen varios formatos en los que se puede guardar un documento. Para convertir un documento en medio digital se tienen que establecer unos criterios tomando en consideración si el documento es texto o imagen, al igual que el período por el cual debe retenerse. Utilizando como base los criterios presentados por North Carolina ECHO (Exploring Cultural Heritage Online), NARA y Departamento de Preservación y Conservación de la Universidad de Cornell, estos se pueden resumir en la siguiente tabla:

[39] *Lossless data compression* consiste en un algoritmo de compresión que permite que la data comprimida se reconstruya de la data original sin pérdida de información (*as is*). Esta compresión contrasta con el *Lossy data compression* que no permite que la data original se reproduzca exactamente en la data comprimida.

[40] ANSI/AIIM TR33-1998 “Selecting an Appropriate Image Compression Method to Match User Requirements”.

[41] Favor de hacer referencia al anejo F de ELECTRONIC ACCESS PROJECT SCANNING AND FILE FORMAT MATRIX (NARA).,

Tipo de formato y su uso			
Formato	**Escaneo imagen para sustituir original**	**Digitalización para documentos que no van a ser conservados**	**Metadatos**
TEXTO (documentos originados por medio electrónico – procesadores de palabras y se van a escanear)	300 dpi bi-tonal / escala de gris 400 dpi bi-tonal en OCR / escala de gris TIFF - 8-bit depth / 300 dpi PDF - 300 dpi Compresión sin alterar el original.	La resolución del archivo es de 300 dpi sin alterar el tamaño de la imagen TIFF/PDF/JPEG2000 JPEG Estos documentos pueden ser comprimidos	Esquema de Taxonomía para la Clasificación de Metadatos provisto por la Oficina del CIO
FOTOGRAFÍAS	600 dpi, 24 bit color TIFF / PFD 600 Blanco y Negro 8 bit escala de gris. (300 dpi) La compresión no puede alterar el original.	8 bit escala de gris, 24 bit color. JPEG 200-300 dpi File resolution 200-300 dpi Se recomienda método para comprimir el archivo	Esquema de Taxonomía para la Clasificación de Metadatos provisto por la Oficina del CIO
DOCUMENTOS (manuscritos) Material de valor permanente	400 - 600 dpi TIFF / PDF Color < 24 bit 300 mínimo en blanco y negro y 8 bit escala de gris La compresión no puede alterar el original.	Resolución de 300 dpi sin alterar la imagen. Se recomienda método para comprimir el archivo.	Esquema de Taxonomía para la Clasificación de Metadatos provisto por la Oficina del CIO
MAPAS, DIBUJOS, BI-TONAL Material de valor permanente	300 dpi a resolución no menor de 300 dpi si son de valor permanente y de mucho detalle. 300 - 400 color, bit depth 24, puede ser TIFF/ JPEG/ JPEG2000 / PDF La compresión no puede alterar el original.	8 bit escala de gris, 24 bit color JPEG (resolución alta) Resolución 300 dpi sin alterar la imagen La compresión no puede alterar el original.	Esquema de Taxonomía para la Clasificación de Metadatos provisto por la Oficina del CIO
OBJETOS	Uso de cámara digital de 300 - 600 dpi. TIFF/PDF/JPEG/JPEG2000 Color, bit-depth 24. La compresión no puede alterar el original.	8 bit escala de gris, 24 bit color. JPEG (resolución alta). Se recomienda resolución de 300 dpi sin alterar la imagen La compresión no puede alterar el original.	Esquema de Taxonomía para la Clasificación de Metadatos provisto por la Oficina del CIO

El formato a utilizarse se determina dependiendo del tipo de documento, si es texto o gráfico, y el tiempo que se tiene que conservar. Para propósitos de esta Guía se está asumiendo que los esfuerzos de digitalización van dirigidos a la conservación del documento a través de su ciclo de vida.

Aunque existen otras variedades de formatos para archivar digitalmente los documentos, para efectos de esta Guía usaremos los siguientes los siguientes cinco formatos que son los que comúnmente se utilizan para conservar imágenes digitalizadas y que se han convertido *de facto* en los formatos estándares.

Se pueden definir los formatos[42] de la siguiente forma:

PDF - Formato de Documento Portátil (Portable Document Format) es un formato digital para representar documentos. Los archivos PDF pueden ser creados originalmente en forma PDF, convertidos de otros formatos electrónicos, o digitalizados de papel, microforma, u otro formato de copia dura (hard copy). Los negocios, gobiernos, bibliotecas, archivos y otras instituciones e individuos alrededor del mundo usan PDF para representar cuerpos considerables de información importante. Mucha de esta información debe ser mantenida por períodos de tiempo sustanciales, o incluso permanentemente. Estos archivos PDF deben mantenerse utilizables y accesibles a través de múltiples generaciones de tecnología. El uso futuro de estos objetos, y el acceso a ellos, depende de mantener su apariencia visual al igual que sus propiedades de orden mayor tales como, la organización lógica de las páginas, secciones y párrafos, el flujo de texto recuperable por máquina a partir de un orden de lectura natural, y una variedad de metadatos administrativos, de preservación y descriptivos.[43]

El propósito principal de la primera parte del estándar ISO 19005 es definir un formato de archivo basado en PDF, conocido como PDF/A, el cual provee un mecanismo para representar documentos electrónicos en una manera que preserve su apariencia visual a través del tiempo, independientemente de las herramientas y sistemas utilizados para crear, almacenar o presentar los archivos.[44]

Un propósito secundario de esta parte del estándar ISO 19005 es proveer un marco de referencia para registrar el contexto y la historia de los documentos electrónicos en los metadatos dentro de sus archivos correspondientes.[45]

Otro propósito es definir un marco de referencia para representar la estructura lógica y otra información semántica de los documentos electrónicos dentro de sus archivos correspondientes.[46]

Para propósitos de preservación, esta declaración de mejores prácticas provee recomendaciones para procesos de captura o conversión de documentos al formato PDF/A para asegurar que los archivos correspondientes resultantes retengan su calidad e integridad como récords. Instituciones de archivística y otras organizaciones con requisitos de preservación a largo plazo deben alentar el uso del Nivel A.[47]

TIFF (*Tagged Image File Format*) Este tipo de formato digital denominado “formato de archivo de imagen con etiquetas”, por sus siglas en inglés TIFF (*"Tagged Image File Format"*) contiene, además de los datos de la imagen propiamente dicha, "etiquetas" en las que se archiva información relacionada con las características de la imagen que sirven para su tratamiento[48] posterior. Una de las características que hace este formato relevante es que se genera sin comprimir la imagen y sin perder información de la

---

[42] Las definiciones son tomadas con texto añadido de North Carolina ECHO.

[43] ISO 19005:1.

[44] ISO 19005:1. Actualmente, este estándar está siendo reevaluado para que se incluya que no puede estar basado en tecnología propietaria. Sin embargo, es el que se utiliza actualmente para la conservación de récords a largo plazo.

[45] ISO 19005:1.

[46] ISO 19005:1.

[47] ISO 19005:1.

[48] http://es.wikipedia.org/wiki/TIFF - Definición de TIFF provista por Wikipedia.

imagen original. Comúnmente, se utiliza en la mayoría de los programas empresariales de manejo de contenido porque, al igual que el PDF, está diseñado para trabajar en cualquier plataforma sin necesidad de convertirlo a otro formato. Estas etiquetas describen el formato[49], tipo de compresión y características del archivo de las imágenes almacenadas.

JPEG (*Joint Photographic Experts Group*) Este formato se utiliza mayormente para la representación de imágenes que no son textos. Por la forma en que comprime al archivo para hacerlo más liviano y poder diseminarlo con mayor rapidez a través de las diferentes plataformas, puede perder píxeles por lo que redundaría en una pérdida de resolución.

GIF (*Graphic Image File Format*) Al igual que JPEG, este formato utiliza un mecanismo para comprimir y reducir el tamaño de la imagen, lo cual compromete la calidad de ésta. Se utiliza mayormente para distribuir y diseminar información a través de redes como el Internet. Un problema con este tipo de formato es que es propietario.

JPEG 2000[50] Este formato no está relacionado al JPEG original. Su particularidad es que se basa en un algoritmo que permite comprimir la información sin pérdida de píxeles, por lo tanto no compromete la calidad de la resolución de la imagen. Una particularidad de este formato es que permite añadir información al archivo a través de códigos. Además, permite comunicar y estructurar a través de códigos.

La siguiente tabla utiliza como base los criterios de North Carolina ECHO:

Imágenes, formato, uso y fiabilidad			
Para qué se escanea	**Formato**[51]	**Uso**	**Fiabilidad**
Imagen del original[52]	TIFF / PDF	Conservar la imagen a largo plazo.	No altera la imagen ni la comprime.
Digitalización para acceso	JPEG o JPEG2000	Presentar en pantalla o para impresión.	Puede ser alterada cuando se disemina en otras plataformas, como el Internet.

[49] Binarias (blanco y negro), adecuadas para textos, por ejemplo. Niveles de gris, adecuados para imágenes de tonos continuos, como fotos en blanco y negro. Paleta de colores, adecuada para almacenar diseños gráficos con un número limitado de colores. Color real, adecuado para almacenar imágenes de tono continuo, como fotos en color.

[50] Para mayor información acerca del **BSI's *Guide to the practical implementation of JPEG 2000*** refiérase a: http://www.jpeg.org/jpeg2000/.

[51] ISO 15801: The Policy Document should contain details of the approved data file formats that may be used for each information type.

All information stored on a computer system requires software for retrieval and display. This software is subject to change, either by the implementation of new releases, or by changes to operating systems and/or hardware. By implementing a policy of approved storage formats, the necessary data migration or alternative procedures can be implemented satisfactorily to ensure long-term retrieval of the stored information. Where compression techniques are available, policy on their use should be documented. Where multiple versions of a document may be stored, a policy is required which ensures that all relevant versions are stored, and their relationship maintained. The Policy Document should contain details of policy on the storage of versions of documents.

[52] La configuración que conlleva este tipo de disposición para escanear páginas es la que se debe utilizar para aquellas imágenes digitales que van a sustituir la copia física.

Imágenes, formato, uso y fiabilidad			
Para qué se escanea	**Formato[51]**	**Uso**	**Fiabilidad**
Thumbnail	JPEG	Desplegar en pantalla.	Se toma de imágenes que han sido reducidas. En algunos de los casos compromete la resolución y/o la calidad de la imagen.

Para propósitos de esta Guía se están tomando en consideración los archivos de datos y no los de vídeo y voz. Estos últimos requieren sus propias guías para su conservación. El interés de esta Guía reside mayormente en desarrollar estrategias para la digitalización de documentos.

El formato que últimamente se utilice para digitalizar un documento va a depender de su propósito final:

- Conservación por el ciclo de vida del documento.
- Diseminación de la información y rápido acceso. (No se crea con el propósito de conservación y/o preservación).
- *Thumbnail*[53], o imagen reducida, para rápido y fácil acceso. Primordialmente se utiliza para fines de la Web.

Cuando se determina que los documentos van a ser conservados a través de su ciclo de vida se tienen que considerar los aspectos de resguardo de la información y la seguridad de los datos.

RECONOCIMIENTO ÓPTICO DE CARACTERES

Cuando se escanea un texto lo que sucede es que se duplican los *bits* gráficos en una página virtual. Es decir, la imagen es una copia de la página que se está escaneando que no se puede alterar. Sin embargo, cuando se desea que esta página de texto tenga la habilidad de reconocer palabras, entonces se pasa a través de un programa de Reconocimiento Óptico de Caracteres (llamado en inglés *Optical Character Recognition*, en adelante, "OCR"). Dicho de otra forma, lo que hace este programa es que toma las formas individuales de la imagen y las convierte en un texto sin alterar la imagen original. Debido a que la conversión se hace a base de la imagen, si su resolución no está clara, se puede esperar que la conversión a texto contenga errores. Esto es frecuente en documentos que son viejos o que han estado expuestos a condiciones que han afectado su calidad.

Cuando se digitalizan documentos usando programas de OCR que permiten editar el documento mediante un procesador de palabras, el trabajo de editarlo puede ser intenso. Esto se debe a que, al hacer la conversión, muchas de las palabras no se definen claramente y hay que editarlas manualmente. En el caso de utilizar formatos tales como TIFF o PDF, que son sistemas abiertos que preservan el tipo de letra (*fonts*), imagen, gráfica y diseño (*layout*), no importa la plataforma y el programa de donde provenga la imagen ya que proveen la habilidad de ejecutar la función de OCR sin alterar la imagen que se está digitalizando. PDF y TIFF se están convirtiendo en el estándar para la digitalización de

[53] Ejemplo de un *Thumbnail* lo podemos apreciar en las fotos archivadas las cuales se presentan y una tamaño menos del original y mantienen su integridad. .

documentos, no sólo en las agencias gubernamentales, sino también internacionalmente, debido a su versatilidad y portabilidad.

El procedimiento de Reconocimiento Óptico de Caracteres se ha ido generalizando cada vez más. Hoy día resulta relativamente fácil incorporarlo en los proyectos de digitalización. De hecho, a partir de abril de 2010, la Oficina del Contralor de Puerto Rico le requiere a las agencias del gobierno que se utilice "OCR en la Digitalización de Contratos, Escrituras y Documentos Relacionados", al enviar copias a la Oficina del Contralor del Estado Libre Asociado de Puerto Rico.[54]

PROCESO DE ESCANEAR

Esta sección trata con el proceso de escanear los documentos. Es decir:

- Todo el proceso de la planificación se llevó a cabo
- Se adquirió el equipo (escáner(es), servidores, estaciones)
- Se estableció el equipo de trabajo
- Se determinó la interface para la migración de los documentos digitalizados al SIGDOC

Al comenzar con el proceso de escaneo de los documentos y una vez completados todos los asuntos relacionados con la planificación del proyecto, existen unos aspectos que se tienen que tomar en consideración. Estos son[55]:

- Hacer la preparación (*prepping*) de los documentos que van a ser escaneados. Se tiene que limpiar el documento, alinear el material y prepararlo en la forma en que será digitalizado.
- Primer Nivel de Control de Calidad para verificar el estado do los documentos a ser digitalizados. Este proceso se tiene que llevar a cabo por un personal asignado por la entidad gubernamental.
- Verificar el status del escáner.
- Configurar el escáner de acuerdo al tipo de documento (resolución, establecer los DPI, etc.).
- Ajustar el monitor para la resolución deseada.
- Escanear en la resolución mayor que es apropiada para el material a ser digitalizado y su uso posterior.
- Escanear de la manera más apropiada para evitar el *rescanning* y, por consiguiente, el desgaste de los documentos originales.
- Escanear siempre de la copia original.
- Segundo Control de Calidad - Cotejar la imagen contra la original.
- Entrada de los Metadatos
- Tercer Nivel de Control de Calidad – Calidad en la Clasificación de las imágenes
- Guardar el documento en el formato que se desea preservar (ej. TIFF, PDF, PDF/A).
- Migrar los datos al SIGDOC una vez completado el proceso de validación.

[54] Carta Circular OC-IO-27, 31 de marzo de 2010.

[55] Favor de hacer referencia a la Figura 1 de Proceso de Digitalización.

CONTROL DE CALIDAD

Previo a comenzar cualquier iniciativa de digitalización se tienen que establecer unos controles de calidad para cada uno de sus procesos. Esos controles de calidad deben ir acompañados de unos procedimientos que permitan medir la eficacia del proceso de modo que garantice la calidad deseada.

Algunos de los controles de calidad para comparar la imagen vs. el original pueden ser:[56]

- Seleccionar los documentos a ser digitalizados
- Cotejar la imagen digitalizada contra la original
    - La legibilidad del documento
    - Veracidad del documento (Si recogió los detalles de la imagen)
    - Precisión en las dimensiones de la imagen vs. la original
    - Densidad de las áreas
    - Fidelidad de los colores
    - Digitalización de todas las áreas de la imagen (características de la imagen)
- Cotejar la clasificación y entrada de metadatos de las imágenes siguiendo el Esquema de Taxonomía de clasificación de Metadatos que se detalla en esta Guía.
- Migrar los datos al SIGDOC. Establecer los mecanismos para verificar los datos.

Dependiendo del tipo de documento y sus características específicas es que se determinan los aspectos que van a ser relevantes en el control de calidad. Por ejemplo, si las escalas de colores son importantes en la definición de la imagen, entonces se tienen que establecer los criterios que van a permitir medir la intensidad de esas escalas.

Típicamente, la logística para digitalizar documentos es prepararlos en grupos (*batch*) para pasarlos por el proceso de escanear. Es en este proceso donde se incorporan los códigos de barra para clasificar los documentos. La selección de los documentos por parte del personal autorizado de la entidad gubernamental es el primer nivel de control de calidad.

El operador de escáner inicia el segundo nivel de control de calidad (cotejo de las imágenes) que es verificar la imagen que se reproduce contra el original. Un segundo paso en este control de calidad debe hacerse por un personal que no sea el que está envuelto en el proceso para asegurarse que sigue los niveles de calidad especificados.

La verificación en la calidad de la entrada de los metadatos en el Sistema de Manejo de Contenido constituye el tercer nivel de Control de Calidad. Los parámetros para establecer esta verificación pueden variar dependiendo de los sistemas.

Los resultados de las pruebas de control de calidad deben ser debidamente documentados. Una vez terminado el proceso de control de calidad para la parte de imagen, se podrá proceder con los demás procesos.

---

[56] Refiérase al ISO/PDTR 15801.

En resumen, se debe aplicar un estricto y minucioso control de calidad en las imágenes digitales, sobre todo si la intención es reemplazar el original por la imagen. Convertir series de documentos en formato digital para rápido acceso, reemplazar los archivos inactivos o digitalizar transacciones de negocios va a requerir de una metodología de control de calidad para lograr dichos objetivos de forma efectiva. Obviamente, esos niveles de calidad tienen que ir en conformidad con las realidades específicas de cada entidad.

5. CLASIFICACIÓN DE DATOS

METADATOS

Una definición informal de metadatos es "la información sobre la información" o cualquier dato asociado con un recurso que describa ese recurso en particular. Una definición más general, que es útil para las instituciones culturales, es "la información estructurada sobre cualquier recurso de información de cualquier tipo de medio o formato".[57] El aspecto esencial de un sistema de metadatos que describe un objeto es su habilidad para proveer un formato estructurado de información acerca de ese objeto.[58]

La definición especificada en el estándar ISO 15489:1 se refiere a metadatos como:

- **Metadatos *(metadata)*** Datos que describen el contexto, el contenido y la estructura de los documentos de archivo y su gestión a lo largo del tiempo.[59]

El propósito fundamental para crear un método de clasificar la información de los documentos que están digitalizándose es a través de un sistema en el que se cree la clasificación o la taxonomía de la información. El objetivo fundamental es poder encontrar la información de forma rápida, segura y eficiente. Para lograr este propósito se crea un sistema estandarizado para nombrar y clasificar los documentos.

Muchas de las iniciativas de digitalización utilizan esquemas de taxonomía para la clasificación de la información que ya están establecidos. Uno de estos esquemas es el *Dublin Core* que es la base que tomamos para definir nuestro esquema de taxonomía para el Gobierno de Puerto Rico.

Los metadatos o elementos que se utilizan para la conservación de los documentos durante su ciclo vital se dividen en cinco (5) grupos o clases. Estas grupos son: (1) administrativos, (2) descriptivos, (3) analíticos, (4) seguridad o derecho de propiedad y (5) recordkeeping. Todos los grupos o clases, a excepción del descriptivo, va a aplicar a todas las entidades gubernamentales. Los metadatos que pertenecen al grupo descriptivo son los que definen las series documentales dentro de cada entidad gubernamental. Por lo tanto, la rigurosidad con la cual se crean estos elementos o metadatos va a ser determinante en el acceso de estos documentos.

TIPO	OBJETIVO	ELEMENTOS

[57] Priscilla Caplan, "Metadata Fundamentals for All Librarians". Chicago; American Library Association, 2003. p. 3. Tomado de North Carolina ECHO, Sección 5.2.

[58] North Carolina ECHO, Sección 5.2.

[59] Proyecto UNE-ISO 15489-1.

TIPO	OBJETIVO	ELEMENTOS
Metadatos administrativos	Los metadatos administrativos son aquellos que proveen información relacionada a los récords, por ejemplo: cuando se crearon, modificaron, relación a otros récords. Incluye datos como: • la creación y el control de calidad; • persona que lo creó, modifico • requisitos de control de acceso y utilización • estructura en la cual fue creada	✓Número de Serie asignado ✓Medio Físico; formato (estructura) ✓Método de Captura ✓Fecha de Captura ✓Persona que lo creó ✓Fecha de Entrada ✓Fecha de la última modificación[60] ✓Persona que lo modificó ✓Lenguaje ✓Relación de este record con otros récords ✓Entrado por ✓Notas
Metadatos descriptivos	Los metadatos descriptivos son aquellos que describen las propiedades físicas o de propiedad intelectual, o el contenido de un(os) récord(s). Además, facilitan la gestión documental de las colecciones digitales tanto a corto como a largo plazo. Cada entidad gubernamental deberá proveer los metadatos descriptivos que definen las series documentales aplicables a sus actividades.	Algunos ***ejemplos*** podrían ser: ✓ Tipo de documento ✓ Número de expediente ✓ Fecha de Radicación ✓ Año de Radicación ✓ Oficina ✓ Status ✓ Fecha de Status ✓ Título ✓ Descripción **(hasta un máximo de 10 campos)**
Metadatos analíticos	Los metadatos analíticos son aquellos que proveen información acerca del contenido del documento. En algunas ocasiones se les conoce como "*subject metadata*".	✓ Temática ✓ Tesauro ✓ Temas/palabras claves ✓ Sinopsis; Tabla de Contenido ✓ Códigos de clasificación ✓ Otros elementos importantes (e.g., agencia a la que pertenece; enlaces a otros documentos relacionados, etc.)

[60] La mayoría de los sistemas de manejo de contenido proveen un registro de auditoría de todas las acciones que se han ejecutado para ese documento/récord.

TIPO	OBJETIVO	ELEMENTOS
Metadatos de Seguridad y Derecho de Propiedad	Son aquellos relacionados a la seguridad de los documentos o restricciones de índole legal, financiera y operacional.	✓ Restricciones en el uso ✓ Permiso (Control de Acceso) ✓ Reconocimientos ✓ Períodos de Retención ✓ Calidad de la clasificación ✓ Uso de Notas relevando responsabilidades
Metadatos de Recordkeeping	Los metadatos de *recordkeeping* son aquellos que tienen que ver directamente con el ciclo de vida del documento. Estos metadatos son mandatarios para la conservación de los documentos por un tiempo reglamentario. Las entidades gubernamentales que deseen conservar los documentos en formato electrónicos por el tiempo reglamentario, deberán utilizar estos elementos para: (1) la creación del Inventario y Plan de Retención y (2) la Disposición.	✓ Serie Documental ✓ Número de la Serie Documental ✓ Status de la Serie Documental ✓ Fecha de Inactivo ✓ Período de Retención ✓ Fecha de Disposición ✓ *Litigation Hold* (S/N) ✓ Validaciones en los procesos

CONTROL DE CALIDAD

Cuando se crean o importan documentos o archivos de datos, se debe asegurar que todos los metadatos relevantes también sean transferidos al SIGDOC. Se debe verificar que todos los metadatos hayan sido capturados, para asegurar que a todos los documentos y archivos de datos se les ha colocado la interpretación correcta.[61]

El contenido de la información de los metadatos puede ser revisado para asegurar que está completo y es apropiado. Es importante seguir el Esquema de Taxonomía de Metadatos provisto en esta Guía.[62]

6. REGISTRO DE AUDITORÍA (*AUDIT TRAIL*) Y PERÍODOS DE RETENCIÓN

Cuando se prepara la información para ser utilizada como evidencia, transacción y/o evento es necesario que se provea información adicional relacionada con el documento.[63] Esa información debe

[61] ISO 15801.
[62] ISO 15801.
[63] ISO/PDTR 15801.

incluir detalles tales como el tiempo y hora que se creó, el traslado de la información de un medio a otro, y la evidencia del control de la operación. Esto es lo que se conoce como Registro de Auditoría. Este registro permite rastrear el historial de todos los eventos y movimientos de la información a través de su ciclo de vida. Es fundamental que el Registro de Auditoría contenga suficiente información como para poder establecer la autenticidad de ese récord.

Los períodos de retención tienen que ser entrados previo a transferirse la información al SIGDOC. No se va a transferir ninguna información de las entidades documentales si no tiene los datos de recordkeeping debidamente entrados y validados.

El Reglamento 4284 para la Administración de Documentos Públicos de la Rama Ejecutiva establece, para los Administradores de Documentos de las Agencias, Municipios, Corporaciones y Juntas que estén adscritos al Programa de Administración de Documentos, las "normas para la retención selectiva de aquellos documentos de valor permanente y proporcionará ayudas a las dependencias en la aplicación de estas normas en los documentos bajo su custodia". Esta reglamentación tiene como objetivo la disponibilidad de los récords públicos por el período de retención asignado. Establecer los períodos de retención permite, entre otros aspectos:

- La disponibilidad de los récords durante su ciclo de vida
- Garantiza el cumplimiento con la Administración de Documentos especificado en los estatutos legales y reguladores del Gobierno de Puerto Rico
- La disponibilidad de los récords como evidencia
- Establece mecanismos para una sana administración de los documentos y promueve la transparencia

A esos efectos, y como exigencia reglamentaria, se requiere que a todos los documentos públicos, no importa si son físicos o electrónicos, se les asigne un período de retención. El período de retención es el mismo no importa el medio en el cual se esté conservando. De igual forma, el mecanismo que se utiliza para los documentos, bien sean físicos o electrónicos, es el mismo. Lo único que cambia es el medio.

7. MIGRACIÓN DE LOS DATOS AL SIGDOC

El SIGDOC tiene como propósito el de conservar los documentos de las entidades gubernamentales a través de su ciclo de las entidades gubernamentales. Debido a que es el SIGDOC la plataforma tecnológica responsable de la conservación de los documentos en formato electrónico por su tiempo reglamentario, tiene que atender todos los aspectos de cumplimientos del recordkeeping electrónico de los documentos que custodia.

Las entidades gubernamentales son responsables de la entrada de datos para la clasificación de los documentos, siguiendo el Esquema de Taxonomía de clasificación de Metadatos y la asignación de los períodos de retención que le aplique a las series documentales que está digitalizando.

Las entidades gubernamentales que van a digitalizar y transferir al SISDOC tienen que seguir los siguientes pasos:

1. El sistema que se utilice para la captura de los documentos debe permitir la entrada y verificación de los metadatos.

2. La migración se puede establecer de las siguietes tres formas:

   - Entrando la data directamente al SIGDOC
   - Configurando un sistema como el SIGDOC localmente y enviando la base de datos por medio de un Interface

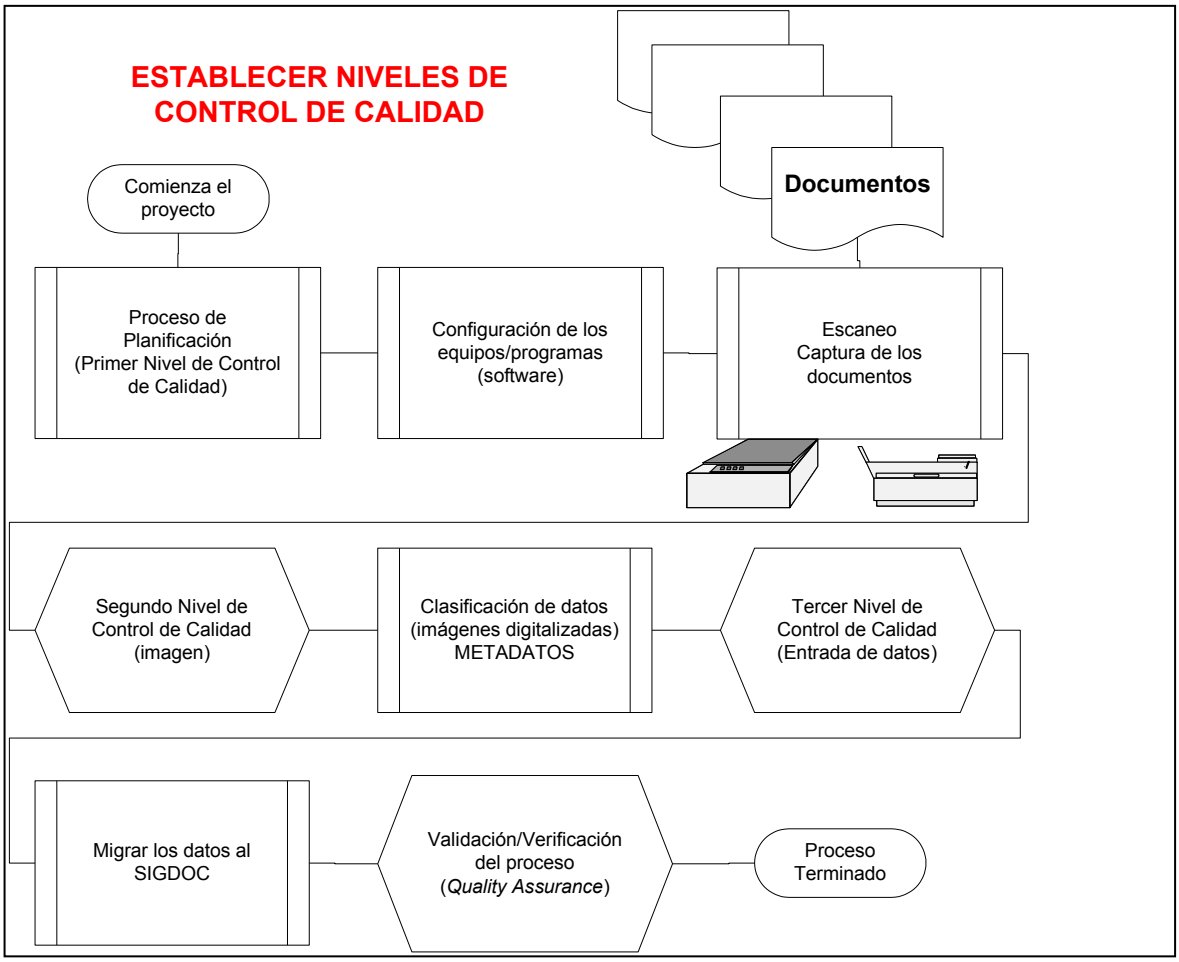
3. Establecer un mecanismo para la validación de los datos, una vez estén en el SIGDOC

4. Adiestramiento al personal.

8. VERIFICACIÓN Y VALIDACIÓN DE CALIDAD

Previo a comenzar cualquier iniciativa de digitalización es necesario establecer los niveles de calidad necesarios para cada proceso dentro del proyecto. Estos niveles deben establecerse durante la etapa de planificación. Una vez establecidos los controles de calidad, se debe establecer un mecanismo para la validación y verificación del proyecto. Ello se debe a que, cualquier proyecto que involucre la parte de digitalización para captura de los documentos y su eventual conservación a través de su ciclo de vida, va a estar enmarcado dentro de un contexto de tecnología de información. Por lo tanto, es necesario establecer controles técnicos y lógicos para verificar y validar la operación y el producto final.

Figura 2 - Control de Calidad



Es importante señalar que el tipo de proyecto va a determinar la rigurosidad de los controles que van a establecerse en el proyecto. Es decir, si los documentos a ser digitalizados van a reemplazar el original, los controles de calidad tienen que ser mucho más específicos, detallados y precisos que aquellos que se utilizan para diseminación de la información dentro de una ambiente de Internet, Intranet o cualquier otro medio cuyos datos son transitorios. Obviamente, hay que tomar en consideración factores tales como costo, volumen y accesibilidad a los documentos para establecer la tecnología más apropiada sin sacrificar la calidad de lo que se va a digitalizar.

Al establecer los controles de calidad para cada proceso, es importante que se determine cuáles son las variables o características que se utilizan para medir su cumplimiento a la hora de hacerse una auditoría. Ejemplos de las características o variables que se pueden cotejar son:

- Tipo de documento
- Resolución de la imagen
- Contraste
- Profundidad
- Calidad de la imagen

- Formato de archivo
- Niveles de compresión
- Calidad del metadato
- Resguardo de la información

CONCLUSIÓN

Con esta Guía, la Oficina del CIO establece los parámetros para la digitalización de documentos en las entidades gubernamentales con el propósito de conservarlos a través de su ciclo vital dentro de un SIGDOC. Es de vital importancia que las entidades entiendan la importancia de la accesibilidad de los documentos de una forma integrada mediante un archivo nacional central.

Ese SIGDOC va a permitir que la información se maneje de forma eficiente y ágil a través de un motor de búsqueda que permita acceder la información de forma simple o federada.

Uno de los factores importantes en esta plataforma es la implantación de un sistema de *recordkeeping* electrónico que garantice la confiabilidad, autenticidad, fiabilidad y uso práctico de los documentos que ahí residen. Por lo tanto, las entidades gubernamentales no tendrían la necesidad de establecer e implantar sistemas complejos de manejo de contenido que tengan el componente de *recordkeeping* electrónico para aquellos documentos que se digitalizan con el propósito de sustituir el original por el formato electrónico a través del ciclo vital.

REFERENCIAS

1. Reglamento 4284 - Reglamento para la Administración de Documentos Públicos en la Rama Ejecutiva. Estado Libre Asociado de Puerto Rico - Administración de Servicios Generales.
2. Ley número 5 del Gobierno de Puerto Rico aprobada el 8 de diciembre de 1955 y enmendada por las Leyes número 77 del 13 de junio de 1960 y 63 del 4 de junio de 1979.
3. Association Information Imaging Management (AIIM)
4. Association Record Management Administration (ARMA)
5. Council of Library and Information Resources, Preservation in the Age of Large Scale Digitization.
6. Digital Preservation Coalition, http://www.dpc.online.org/graphics/index.html .
7. National Archives and Records Administration, Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Image.
8. North Carolina ECHO, Guidelines for Digitization.
9. Queensland State Archives, Guideline for the Digitisation of Paper Records.
10. Western States Digital Standards Group, Digital Imaging Best Practices.
11. U.S. NATIONALARCHIVES AND RECORDS ADMINISTRATION- Electronic Access Project Scanning and File Format Matrix, http://www.archives.gov/preservation/technical/guidelines-matrix.pdf
12. Moving theory into practice Digital Imaging Tutorial - http://www.library.cornell.edu/preservation/tutorial/intro/intro-04.html © 2000-2003 Cornell University Library/Research Department
13. Anne R. Kenney and Oya Y. Rieger, Co-Chairs et al, Establishing a Central Depository for Preserving Digital Image Collections, March 2001, version 1
14. United Nations Organization, Archives and Records Management Section (ARMS). Glossary of Recordkeeping Terms. http://archives.un.org/unarms/en/unrecordsmgmt/unrecordsresources/glossaryofrecordkp
15. Universidad de Cornell / Centro de Conservación y Preservación Archivística
16. Estándare internacionales
    a) DoD 5015.2
    b) ISO 10196:2003
    c) ISO 15489
    d) ISO PDTR 15801
    e) ISO 19005-1:2005(E)
    f) ANSI /AIIM MS52-1991
    g) ANSI/AIIM TR15-1997
    h) AIIM TR26 – 1993

i) ANSI/AIIM TR33-1998

j) Information and Documentation – Records Management – Part 1: General (ISO 15489:1)

k) Information and Documentation – Records Management – Part 2: Guidelines 15489:2

ANEJO I - DEFINICIONES

Acceso *(access)* - Derecho, modo y medios de localizar, usar o recuperar información. [**ISO 15489-1:2001]**

Aplicación - En el contexto de esta Política se utiliza como sinónimo de *software*.

Aplicaciones ofimáticas - Los paquetes de *software* que llevan a cabo una variedad de funciones de manejo de oficina tales como procesadores de palabras, aplicaciones de publicación, hoja de cálculo, correo electrónico, transmisión de facsímil y recibo, imagen documental, reconocimiento de carácter óptico (OCR), flujo de trabajo y manejo de datos. Estas aplicaciones se utilizan generalmente para generar, convertir, transmitir o recibir documentos de negocio. **[DoD 5015.02-STD]**

Autenticidad - Un récord es auténtico cuando se puede probar que:

a. es el que se alega que es,

b. ha sido creado o enviado por la persona que se alega haberlo creado o enviado, y

c. ha sido creado o enviado en el momento que se alega. [**ISO 15489-1:2001]**

Bases de datos - En relación con récords electrónicos, es un conjunto de elementos de datos que contiene al menos un archivo o un grupo integrado de archivos que se hacen disponibles a varios usuarios. **[NARA 36 CFR Part 1234.2]**

Captura - Capturar un documento es el proceso por el cual se determina que un récord tiene que crearse o mantenerse por un determinado período de tiempo.

Captura de los récords - Establece una relación entre el récord, su creador y el contexto del negocio que lo originó, coloca el récord y sus relaciones dentro de un sistema de récords y lo vincula a otros. [**ISO 15489-1:2001]**

Ciclo de vida de los récords - Un concepto archivístico que describe el período de vida de un récord, desde su creación o recibo hasta su disposición final. El ciclo de vida de los récords se divide en las siguientes etapas: creación/recibo, mantenimiento y uso, retiro, disposición final y uso continuo. **[NARA Glossary]**

Contenido - La sustancia intelectual del documento, incluyendo texto, datos, símbolos, números, imagen y sonidos. **[The Society of American Archivists. *A Glossary of Archival and Records Terminology*]**

Contexto - Las circunstancias organizacionales, funcionales y operacionales que rodean la creación, recibo, almacenaje o utilización de un material y su relación con otros materiales. **[The Society of American Archivists. *A Glossary of Archival and Records Terminology*]**

Documento - La palabra documento comprenderá todo papel, libro, folleto, fotografía, fotocopia, película, microforma, cinta magnetofónica, mapa, dibujo, plano, cinta magnética, disco, videocinta o cualquier otro material leído por máquina y cualquier otro material informativo sin importar su forma o características físicas. El material bibliográfico, o de museo, adquirido para propósitos de exposición, consulta u otros relacionados y las publicaciones no están incluidos en la definición de la palabra "documento". **[Ley Núm. 5 del 8 de diciembre de 1955]**

Documentos de archivo *(récords)* - Información creada o recibida, conservada como información y prueba, por una organización o un individuo en el desarrollo de sus actividades o en virtud de sus obligaciones legales. [**ISO 15489-1:2001]**

Documento público - Todo documento que se origine, conserve o reciba en cualquier dependencia del Estado Libre Asociado de Puerto Rico de acuerdo con la ley o en relación con el manejo de los asuntos públicos y que de conformidad con lo dispuesto en la sec. 1002 de este título se haga conservar permanentemente o temporeramente como prueba de las transacciones o por su valor legal. **[Administración de Servicios Generales. *Reglamento Núm. 4284*]**

DPI (***Dots Per Inch***) – en escaneo, es una medida de resolución – el número de pixeles que el escáner puede fácilmente distinguir en cada línea horizontal y vertical de una imagen original.

Elementos de dato - Unidad de un dato por la cual se especifican los valores de admisión, definición, identificación y representación por medio de un conjunto de atributos. **[DoD 5015.02-STD]**

Estándares - Referencia o punto de referencia para establecer una práctica o calidad deseables.

Estructura - La manera en la que los elementos están organizados, interrelacionados y representados (*displayed*). **[The Society of American Archivists. *A Glossary of Archival and Records Terminology*]**

Fiabilidad - Un récord fiable es aquel cuyos contenidos pueden ser confiados como una representación completa y precisa de las transacciones, actividades o hechos que los atestiguan y en los cuales se pueden depender en el transcurso de transacciones o actividades subsiguientes. Los récords deben ser creados en el momento de la transacción o el incidente con el que se relacionan, o poco tiempo después, por los individuos que tienen conocimiento directo de los hechos o por instrumentos utilizados rutinariamente para llevar a cabo la transacción dentro del negocio. [**ISO 15489-1:2001]**

ICR (Intelligent Character Recognition) – es la traducción de texto hecho a manuscrito a caracteres que puedan ser leídos por una máquina.

Integridad - La integridad de un récord se refiere a que está completo e inalterado. Es necesario proteger al récord contra alteraciones no autorizadas. Las políticas y procedimientos para el manejo de los documentos públicos deben especificar qué añadiduras o anotaciones pueden hacerse a un récord luego de creado, bajo qué circunstancias se pueden autorizar tales añadiduras o anotaciones y quién tiene autorización para hacerlas. Cualquier anotación, adición o eliminación autorizada que se le efectúe a un récord debe ser indicado explícitamente y fácil de encontrar. [**ISO 15489-1:2001]**

ISO (*International Organization for Standardization***) -** Es una federación mundial de entidades para estándares nacionales. El trabajo de preparar estándares internacionales normalmente es llevado a cabo por los comités técnicos de ISO. Cada entidad afiliada que esté interesada en un tema por el cual se ha establecido un comité técnico, tiene el derecho de ser representado en ese comité. Organizaciones internacionales, gubernamentales o no, en coordinación con ISO, también forma parte del trabajo. ISO colabora estrechamente con la Comisión Electrotécnica Internacional (*International Electrotechnical Commission* (IEC) en todos los asuntos correspondientes a la estandarización electrotécnica. [**ISO 15489-1:2001]**

Metadatos *(metadata)* - Datos que describen el contexto, el contenido y la estructura de los documentos de archivo y su gestión a lo largo del tiempo. [**ISO 15489-1:2001]**

Migración *(migration)* - Acción de trasladar documentos de archivo de un sistema a otro, manteniendo la autenticidad, la integridad, la fiabilidad y la disponibilidad de los mismos. [**ISO 15489-1:2001]**

OCR (Reconocimiento Óptico de Caracteres) –transforman en texto una imagen escaneada. Identifican automáticamente símbolos o caracteres que pertenecen a un determinado alfabeto, a partir de una imagen para almacenarla en forma de datos con los que podremos interactuar mediante un programa de edición de texto o similar.

Período de retención - Período de tiempo fijado durante el cual deben retenerse los documentos luego del cual pueden disponerse. **[Administración de Servicios Generales. *Reglamento Núm. 4284*]**

Preservación *(preservation)* - Procesos y operaciones realizados para garantizar la permanencia intelectual y técnica de documentos de archivo auténticos a lo largo del tiempo. [**ISO 15489-1:2001]**

Récord - "*Recorded information, regardless of medium or characteristics, made or received by an organization in pursuance of legal obligations or in the transaction of business*." (ARMA, 2007). Para efectos de los documentos que caen en la categoría de documentos públicos y que son generados o creados a través de un medio electrónico, se les considerará récord electrónico.

***Recordkeeping* electrónico -** Refiérase a *Manejo de Documentos Públicos Electrónicos*.

Récord electrónico (*electronic record*) - Información grabada en una forma que requiere de una computadora u otra máquina para procesarla y que satisface la definición legal de un récord de acuerdo con la sección 3301 del Título 44 del Código de los Estados Unidos. **[DoD 5015.02-STD]**

Registro de auditoría - Un medio electrónico para rastrear la interacción con los récords dentro de un sistema electrónico para que cualquier acceso al récord dentro del sistema electrónico pueda quedar documentado mientras ocurre o posteriormente. Un registro de auditoría puede identificar acciones no autorizadas en relación con los récords, por ejemplo modificaciones, eliminaciones o adiciones. **[DoD 5015.02-STD]**

Re-validación - Re-confirmar el proceso de validación por una firma electrónica validada previamente. **[The National Archives. *Record Management Guidance for Agencies Implementing Electronic Signature Technologies*]**

Series documentales - Documentos que poseen características en común, que son archivados, usados, trasladados y dispuestos como unidad. **[Administración de Servicios Generales. *Reglamento Núm. 4284*]**

Serie de récords (expediente) - Unidades de archivo o documentos ordenados de acuerdo con un sistema de clasificación o mantenidos como una unidad pues resultan de la misma acumulación o proceso de clasificación, de la misma función o actividad; que tengan una forma particular; o por alguna otra relación que surja de su creación, recibo o uso. **[The Society of American Archivists. *A Glossary of Archival and Records Terminology*] [The National Archives. *Record Management Guidance for Agencies Implementing Electronic Signature Technologies*]**

Open System - a system with characteristics that comply with specified, publicly maintained, readily available standards and that therefore can be connected to other systems that comply with those same standards.

Transacción - La unidad más pequeña de una actividad de negocios.

Uso práctico - Un récord es utilizable cuando se puede localizar, recuperar, presentar e interpretar. Debe ser capaz de presentarse posteriormente ligado directamente a la actividad o transacción de negocio que lo produjo. Los vínculos contextuales de los récords deben portar la información necesaria para el entendimiento de las transacciones que los crearon y utilizaron. Debe ser posible identificar un

récord dentro del contexto de las funciones y actividades de negocio más amplias. Los vínculos entre récords que documentan una secuencia de actividades deben mantenerse. [**ISO 15489-1:2001]**

Validez - Para conservar un documento público en formato electrónico es fundamental que contenga las cuatro características que garantizan la validez (*trustworthiness*) de la información desde la perspectiva del manejo de los documentos. Estas características son **autenticidad**, **integridad**, **fiabilidad** y **uso práctico**

Además, es importante que el récord conserve su contenido, contexto y estructura. Es decir, que se conserve el **contenido** total de la información que relaciona al récord dentro del **contexto** en que fue creado y utilizado, conservando su formato o **estructura**. **[The National Archives. *Record Management Guidance for Agencies Implementing Electronic Signature Technologies*]**

Validación (conversión de datos) - Los procedimientos y métodos que se utilizan para verificar que el récord electrónico ha sido convertido de forma exitosa y que cualquier persona puede asegurar que la autenticidad se ha mantenido. De acuerdo con Pearce-Moses, validación es "el proceso de establecer veracidad" y va de la mano con la autenticación. Un manual debe describir los procedimientos para asegurar la calidad a través del proceso de conversión, incluyendo aquellos que prueban e informan resultados de exámenes y los errores y pasos que se han tomado para corregirlos. Los procedimientos de garantía de calidad (*quality assurance*) también deben enumerar los pasos necesarios para asegurar un ambiente seguro y controlado para el proceso de conversión. **[ANSI/ARMA 16-2007. *The Digital Record Conversion Process: Program Planning, Requirements, Procedures*.)]**

Verificación - Es un proceso de calidad que se utiliza para evaluar si un producto, servicio o sistema cumple o no con la regulación, especificación o condiciones impuestas al principio de una etapa de desarrollo. La verificación puede estar en desarrollo, a escala o en producción. Usualmente esto es un proceso interno. **[ISO 9000]**

ANEJO II – TABLA DE RESUMEN

Documentos o Series documentales	Resolución Copia Maestra	Profundidad	Compresión	OCR	Esquema de Metadatos	Sistema de Manejo de Contenido	Archivo Formato	Método alterno de compresión para desplegar en el Internet	Tecnología de conservación
TEXTO (documentos originados por medio electrónico – procesadores de palabras y se van a escanear)	300 DPI Mínima	Bitonal	Recomendable Ninguna La compresión no puede alterar el original.	Si 400 DPI recomendable	Dublin Core	DoD 5015.2	PDF/TIFF	150-200 DPI Dependiendo de la Calidad deseada y el ancho de banda disponible	**SIGDOC**
Documentos oficiales en Texto con ilustraciones en blanco y negro	300 DPI Mínima	Bitonal	Recomendable Ninguna La compresión no puede alterar el original.	Si 400 DPI recomendable	Dublin Core	DoD 5015.2	PDF/TIFF	Compresión dependiendo de la calidad que se desea y el ancho de banda disponible	SIGDOC
Documentos oficiales en Texto con ilustraciones	300 DPI Mínima	8-24bi-tonal Dependiendo de la precisión que se necesitas en los colores	Recomendable Ninguna La compresión no puede alterar el original.	Si 400 DPI recomendable	Dublin Core	DoD 5015.2	PDF/TIFF	Compresión dependiendo de la calidad que se desea y el ancho de banda disponible	SIGDOC

Documentos a Colores Brochure Presentaciones	6”x4” - 600 ppi 7”x5” - 450 ppi 9”x6” - 300 ppi	8-24bi-tonal Dependiendo de la precisión que se necesitas en los colores	Recomendable Ninguna La compresión no puede alterar el original.	En el caso de no ser imágenes	Dublin Core	DoD 5015.2	TIFF/ JPEG/ JPEG2000 / PDF	Compresión dependiendo de la calidad que se desea y el ancho de banda disponible	SIGDOC
FOTOGRAFÍAS	6”x4” - 600 ppi 7”x5” - 450 ppi 9”x6” - 300 ppi	8-24bi-tonal Dependiendo de la precisión que se necesitas en los colores	Recomendable Ninguna La compresión no puede alterar el original.	En el caso de no ser imágenes	Dublin Core	DoD 5015.2	TIFF/ JPEG/ JPEG2000 / PDF	Compresión dependiendo de la calidad que se desea y el ancho de banda disponible	SIGDOC
DOCUMENTOS (manuscritos) Material de valor permanente	300 mínimo 400 DPI recomendable para la copia maestra (*Master Copy*)	400 - 600 dpi TIFF / PDF Color < 24 bit 300 mínimo en blanco y negro y/o 8 bit escala de gris	Recomendable Ninguna La compresión no puede alterar el original.	N/A Si es Manuscrito	Dublin Core	DoD 5015.2	PDF/TIFF preferible O TIFF/ JPEG/ JPEG2000 / PDF	Compresión dependiendo de la calidad que se desea y el ancho de banda disponible	SIGDOC
MAPAS, DIBUJOS, BI-TONAL Material de valor permanente	300 dpi Mínima	Bitonal	Recomendable Ninguna La compresión no puede alterar el original.	N/A	Dublin Core	DoD 5015.2	PDF/TIFF preferible O TIFF/ JPEG/ JPEG2000 / PDF	Compresión dependiendo de la calidad que se desea y el ancho de banda disponible	SIGDOC

MAPAS, DIBUJOS, con colores	300 dpi Mínima	8-24bi-tonal Dependiendo de la precisión que se necesitas en los colores	Recomendable Ninguna La compresión no puede alterar el original.	N/A	Dublin Core	DoD 5015.2	PDF/TIFF preferible O TIFF/ JPEG/ JPEG2000 / PDF	Compresión dependiendo de la calidad que se desea y el ancho de banda disponible	SIGDOC

ANEJO III – TABLAS GUÍA

A– TAMAÑOS DE PAPEL

Tamaños de papel	
Tamaño	**Medidas**
A4	210x297 mm
A4L	297x210 mm
A5	148x210 mm
A5L	210x148 mm
A6	105x148 mm
A6L	148x105 mm
B4	257x364 mm
B4L	364x257 mm
B5	182x257 mm
B5L	257x182 mm
B6	128x182 mm
B6L	182 x 128 mm
Letter Portrait	8.5”x11.0” mm
Letter Landscape	11.00”x8.5” mm
Legal Portrait	8.5”x14.0” mm
Legal Landscape	14.0”x8.5” mm
US Executive	7.25”x10.5” mm
US Executive Legal	10.5”x7.25 mm
Double length	11.0”x17.0” mm
A3	297mm x 420 mm Molde del escáner

B- PROS AND CONS DIGITAL CAPTURE DEVICES

Pros and Cons of Digital Capture Devices		
Capture Device	**Pros**	**Cons**
Flatbed scanner	Highly addressable Inexpensive Many units can handle both transmission and reflection materials Flexible software drivers Most good up to 600 dpi of real resolution Low learning curve	Low productivity, frequent document handling Tendency toward streaking and color misregistration Prone to inflated marketing claims
Sheet-fed scanner	High productivity As good as or better than flatbed scanners Many automatic features	Unsuitable for fragile, bound, wrinkled, 3-D, or inflexible objects More expensive than flatbed scanners May not handle all sizes of documents
Drum scanner	Very high image quality High resolution Low noise High dynamic range Good tone/color fidelity Few artifacts Very flexible software drivers Variable sampling rate	Expensive Low productivity Frequent handling High operator skill level Handles limited document types; must be mountable on drum
Reprographic stand scanner	Very high image quality High resolution Low noise High dynamic range Good tone/color fidelity Few artifacts Flexible software drivers	Expensive High operator skill level Frequent document handling, although minimized impact on document.

Pros and Cons of Digital Capture Devices		
Capture Device	**Pros**	**Cons**
Camera	Can handle a variety of document/object types (3-D, bound, glass plates, non-flat, oversized) Unlimited field size User-controlled lighting Rapid capture for area arrays Non-contact capture May have interchangeable lenses Generally good image quality	Good models are expensive Limited sensor size Low productivity for linear array types Nonuniformity artifacts common Area array devices prone to low dynamic range due to flare Moderate skill level required
Film scanner	Highly productive for roll film Low flare/good dynamic range for linear arrays	Low productivity for sheet film or slides Potential for high flare in area-array devices Dust/scratch artifacts common Image quality characterization difficult due to lack of targets

C- LEVELS OF SCAN, FILE FORMAT, USES AND ALTERATIONS

Levels of scan, file format, uses and alterations			
Levels of Scan	**File Format**	**Used for**	**Alter?**
Master image	TIFF / PDF/A	Long-term storage or print	Do not alter, or resize, or compress
Access image	JPEG or JPEG2000	Screen display or print	Taken from the master, it is altered for presentation over the Web or other uses
Thumbnail	JPEG or GIF	Screen display	Taken from access, reduced size but not altered otherwise

D- BIT DEPTH

Bit depth		
Types of Scan	**Preferred Bit-Depth**	**This Means…**
bi-tonal	1 bit	each pixel is either black or white
grayscale	8 bit	each pixel can be 1 of 256 shades of gray
color	8 bit or 24 bit	8 bit: each pixel can be 1 of 256 shades of color or 24 bit: each pixel can be 1 of 16.8 million color possibilities

E- IMPROVING THE LIFESPAN OF CD'S AND DVD'S

Improving the lifespan of CDs and DVDs		
Always	**Avoid**	**Never**
Store media in controlled archival environment	Damage to the upper and lower surfaces and edges of the disk	Attach or fix anything to the surface of a disk
Store media in a jewel case or protective sleeve when not in use	Scratching and contact with surfaces that might result in grease deposits (e.g. human hands)	Write on any part of the disk other than the plastic area of the spindle
If using sleeves, use those that are of low-lint and acid-free archival quality	Exposing disks to direct sunlight	
Wear gloves when handling the master disks		

DAT Tape, DLT Tape, ZIP® and JAZ® drives

Tape, ZIP® and JAZ® drives are all magnetic media, and magnetic media is NOT recommended for long-term storage. Tape is, however, an excellent intermediate medium, particularly for transport of data and for backup.

F- IMPROVING THE LIFESPAN OF DIGITAL LINEAR TAPE (DLT)

Improving the lifespan of DLTs		
Always	**Avoid**	**Never**
Keep tape in its protective case when not in use	Placing the tapes near magnetic fields	Stack the tapes horizontally
Move tapes in their cases	Moving the tapes about	Put adhesive labels on the cartridge
Store the tapes in appropriate archival environment	Exposing disks to direct sunlight	Touch the surface of the tape
Store the tapes vertically		Put a tape that has been dropped in a drive without first visually inspecting it to make certain that the tape has not been dislodged or moved

F- ELECTRONIC ACCESS PROJECT SCANNING AND FILE FORMAT MATRIX

U.S. NATIONAL ARCHIVES AND RECORDS ADMINISTRATION- Electronic Access Project Scanning and File Format Matrix **January 1998**

All dots per inch (dpi) specified for scanning are equivalent dpi for the size of the original documents, unless otherwise specified. All scanning from photographic intermediates or microfilm shall be done at the dpi needed to achieve the cited resolution on the original documents. Due to inconsistent legibility of documents and the desire to image documents in a consistent manner, all scanning will be 8-bit greyscale or 24-bit color. The tone reproduction and aimpoint values are intended to provide good monitor representation of the images on an "average" computer monitor based on the visual evaluation of a selection and variety of monitors. Scanning and quality control workstation monitors shall be set to 24 bits (millions of colors), set to a gamma of 2.2, set to a color temperature of 6500°K, calibrated for color neutrality, and the contrast and brightness adjusted for the aimpoint values.

Scanning Aimpoints- Using the Kodak greyscale, the tonal controls of the scanner shall be adjusted to place step #19 at RGB levels of 8, 8, and 8 or at % black of 97%, to place step "M" at RGB levels of 105, 105, and 105 or at % black of 59%, and to place step "A" at RGB levels of 247, 247, and 247 or at % black of 3%. Using the Kodak color bars, the tonal controls of the scanner shall be adjusted to place single color black patch at RGB levels of 19, 19, and 19 or at % black of 93%, to place the gray background at RGB levels of 101, 101, and 101 or % black of 60%, and to place the white patch at RGB levels of 241, 241, and 241 or % black of 5%. Variability: no more than ± 3 RGB level variance from aimpoints and no more than 3 RGB level difference in the individual channels within a patch for RGB scanning and no more than ± 1% level variance from the aimpoints in % black for greyscale scanning.

TYPES OF DOCUMENTS

FILES	Textual	Photographs	Maps and Drawings	Graphic
MASTER FILES	**Scan Resolution:** 300 dpi for docs. UP TO 11"x17" or 187 sq. in. 200 dpi for docs. OVER 11"x17" or 187 sq. in.	**Scan Resolution:** adjust dpi to achieve 3,000 pixels across the long dimension adjust dpi to achieve 2,700 x 2,700 pixels for square images	**Scan Resolution:** 300 dpi for docs. UP TO 11"x17" or 187 sq. in. 200 dpi for docs. OVER 11"x17" or 187 sq. in.	**Scan Resolution:** adjust dpi to achieve 3,000 pixels across the long dimension for docs. UP TO 11"x17" or 187 sq. in. 200 dpi for docs. OVER 11"x17" or 187 sq. In.
	File Format and Resolution: TIFF •uncompressed •Intel (IBM) byte order •300 or 200 dpi at orig. size	**File Format and Resolution:** TIFF •uncompressed •Intel (IBM) byte order •10" across long dim. or 9"x9" at 300 dpi	**File Format and Resolution:** TIFF •uncompressed •Intel (IBM) byte order •300 or 200 dpi at orig. size	**File Format and Resolution:** TIFF •uncompressed •Intel (IBM) byte order •10" across long dim. at 300 dpi or 200 dpi at orig. size
ACCESS FILES	**File Format:** GIF for docs. UP TO 8.5"x14" or 119 sq. in. •4 bit/89a/interlaced for greyscale •8 bit/89a/interlaced/custom palette for color JPEG for docs. OVER 8.5"x14" or 119 sq. in. •8 bit/low quality compression for greyscale •24 bit/RGB/low quality compression for color	**File Format:** GIF •4 bit/89a/interlaced for greyscale •8 bit/89a/interlaced/custom palette for color	**File Format:** JPEG •8 bit/low quality compression for greyscale •24 bit/RGB/low quality compression for color	**File Format:** GIF for docs. UP TO 11"x17" or 187 sq. in. •4 bit/89a/interlaced for greyscale •8 bit/89a/interlaced/custom palette for color JPEG for docs. OVER 11"x17" or 187 sq. in. •8 bit/low quality compression for greyscale •24 bit/RGB/low quality compression for color
	File Resolution: for GIF files, resize to 90 dpi at original size and let GIF default to 72 dpi at larger than orig. size 90 dpi at orig. size for JPEG files	**File Resolution:** 600 pixels across the long dimension at 72 dpi 540x540 pixels for square images at 72 dpi	**File Resolution:** 10" (or 1,200 pixels) across the long dimension at 120 dpi 9"x9" (or 1,080x1,080 pixels) at 120 dpi for square images	**File Resolution:** GIF files- 600 pixels across long dimension or 540x540 pixels for square images at 72 dpi JPEG files- 10" (or 1,200 pixels) across the long dimension at 120 dpi
THUMBNAIL FILES	**File Format and Resolution:** GIF •200 pixels across the long dimension •72 dpi •4 bit/89a/interlaced for greyscale •8 bit/89a/interlaced/custom palette for color	**File Format and Resolution:** GIF •200 pixels across the long dimension •72 dpi •4 bit/89a/interlaced for greyscale •8 bit/89a/interlaced/custom palette for color	**File Format and Resolution:** GIF •200 pixels across the long dimension •72 dpi •4 bit/89a/interlaced for greyscale •8 bit/89a/interlaced/custom palette for color	**File Format and Resolution:** GIF •200 pixels across the long dimension •72 dpi •4 bit/89a/interlaced for greyscale •8 bit/89a/interlaced/custom palette for color

Chart compiled by Steven Puglis. National Archives at College Park, 8601 Adelphi Road, Room B810, College Park, MD, 20740, USA. Phone: 301-713-6712. FAX: 301-713-6921. Email: steven.puglis@arch2.nara.gov

http://www.archives.gov/preservation/technical/guidelines-matrix.pdf