

UNIVERSIDAD POLITÉCNICA DE VALENCIA
Escuela Técnica Superior de Ingeniería Informática
Licenciatura en Documentación



DOE-55. PLAN DE MEJORA DE LA DOCUMENTACIÓN EN
UNA ORGANIZACIÓN

Desarrollo de una Metodología para el Análisis de Requisitos
de la Documentación Física en el contexto del Diseño de un
Proyecto Global de Digitalización

Proyecto Fin de Carrera
Presentado por: María José Gimeno Sanjuán
Dirigido por: José María Torralba Martínez

Valencia, 27 de Septiembre de 2011

Tabla de Contenidos

RESUMEN.....	1
CAPÍTULO 1. INTRODUCCIÓN	4
1.1. PREÁMBULO.....	4
1.2. OBJETO	4
1.2. OBJETIVOS	4
1.3. ANTECEDENTES.....	5
1.4. METODOLOGÍA.....	7
1.5. PLAN DEL TRABAJO.....	7
1.6. ESTRUCTURA DOCUMENTAL DEL PROYECTO FINAL DE CARRERA	9
1.7. APLICACIÓN DE LOS CONOCIMIENTOS ADQUIRIDOS.....	12
CAPÍTULO 2. DESARROLLO DE UNA METODOLOGÍA PARA EL ANÁLISIS DE REQUISITOS DE LA DOCUMENTACIÓN FÍSICA EN EL CONTEXTO DEL DISEÑO DE UN PROYECTO GLOBAL DE DIGITALIZACIÓN	14
2.1. ANÁLISIS CUANTITATIVO Y ESTRUCTURAL	17
2.1.1. Local de Archivo	25
2.1.2. Muestreo Cuantitativo.....	25
2.1.2.1. Muestreo Aleatorio Simple	27
2.1.2.2. Muestreo Sistemático	27
2.1.2.3. Muestreo Estratificado	27
2.1.2.4. Muestreo por Conglomerados	28
2.1.2.5. Tamaño de la Muestra.....	29
2.1.3. Sistema de Archivo	30
2.2. ANÁLISIS FÍSICO	34
2.2.1. Evaluación de los Atributos Físicos de los Documentos.....	35
2.2.1.1. Tipos de documento según el modo de representación de su contenido.....	36
2.2.1.2. Presencia y significado del color en el documento.....	37
2.2.1.3. Atributos que afectan a la legibilidad del documento.....	38
2.2.1.4. Formato del soporte documental.....	39
2.2.1.4.1. Tipología de Soportes Documentales: Características y formato	40
PAPEL	40
TRANSPARENCIAS	41
PELÍCULAS FOTOGRÁFICAS	42
FOTOGRAFÍAS	43
MICROFILM.....	44
2.2.1.4.2. Formatos normalizados /especiales	44
2.2.1.4.3. Formatos simples /encuadernados	44
2.2.1.4.4. Formatos antiguos y/o frágiles	45
2.2.2. Caracterización Objetiva de las Variables de Calidad.....	45

2.2.2.1. Umbral de Resolución.....	45
2.2.2.2. Profundidad de Bits.....	45
2.2.2.3. Compresión.....	47
2.2.2.4. Escáner utilizado.....	48
2.2.3. <i>Determinación del Patrón de Referencia de las Variables de Calidad mediante su correlación con los Atributos Físicos analizados.....</i>	<i>51</i>
2.2.3.1. Patrón de Referencia de la Resolución basado en los atributos Tipo de Documento, Legibilidad y Color.....	51
2.2.3.1.1. Resolución de referencia para documentos textuales impresos.....	51
2.2.3.1.2. Resolución de referencia para documentos textuales manuscritos.....	53
2.2.3.1.3. Resolución de referencia para documentos gráficos a media tinta.....	54
2.2.3.1.4. Resolución de referencia para documentos gráficos de tono continuo.....	55
2.2.3.2. Patrón de Referencia del Modo de Escaneo /Profundidad de Bits basado en los atributos Tipo de Documento y Color.....	56
2.2.3.4. Patrón de Referencia de la Técnica y Nivel de Compresión basado en los atributos Tipo de Documento y Profundidad de Bits.....	57
2.2.3.5. Patrón de Referencia para la selección del Tipo de Escáner necesario basado en el atributo Formato.....	60
2.3. ANÁLISIS DE CONTENIDO.....	61
2.3.1. <i>Definición y contextualización de la indización como proceso de apoyo a la recuperación de fondos digitalizados.....</i>	<i>61</i>
2.3.2. <i>Puntos de análisis de la documentación física para la extracción de requisitos relacionados con la indización.....</i>	<i>63</i>
2.3.2.1. Nivel de Asignación Requerido.....	64
2.3.2.2. Determinación de las Fuentes de Información.....	66
2.3.2.3. Establecimiento de Herramientas de Control de Vocabulario.....	67
2.3.3. <i>Automatismos.....</i>	<i>69</i>
2.3.3.1. Procesamiento por Lotes.....	70
2.3.3.2. Aprovechamiento de Ficheros Cliente.....	72
2.3.3.3. Aplicación de Técnicas de Reconocimiento Óptico.....	73
2.4. REQUISITOS RELACIONADOS CON EL PRODUCTO DIGITALIZADO ENTREGADO.....	75
2.4.1. <i>Formato del Fichero de Imagen.....</i>	<i>75</i>
2.4.2. <i>Formato de la Base de Datos con los Atributos de Indización.....</i>	<i>77</i>
CAPÍTULO 3. CONCLUSIONES.....	78
BIBLIOGRAFÍA Y NORMATIVA TÉCNICA.....	79
NORMATIVA TÉCNICA.....	79
BIBLIOGRAFÍA.....	79
ANEXO I. ANÁLISIS DE LOS REQUISITOS EXPRESADOS EN EL ESQUEMA NACIONAL DE INTEROPERABILIDAD.....	84
ANEXO II. GLOSARIO DE SIGLAS Y ACRÓNIMOS.....	92

Índice de Tablas

Tabla 1: <i>Tipos de documento según la forma de representación de su contenido</i>	36
Tabla 2: <i>Características del Papel</i>	40
Tabla 3: <i>Formato Normalizado Papel ISO/DIN</i>	41
Tabla 4: <i>Profundidad de Bits</i>	47
Tabla 5: <i>Tipos de Escáner</i>	48
Tabla 6: <i>Patrón de referencia de la resolución para fotografías como documentos gráficos de tono continuo</i>	56
Tabla 7: <i>Patrón de referencia de la profundidad de bits</i>	57
Tabla 8: <i>Atributos de las técnicas de compresión más comunes</i>	60
Tabla 9: <i>Patrón de referencia para la selección del tipo de escáner necesario</i>	60
Tabla 10: <i>Tipos de Metadatos</i>	62
Tabla 11: <i>Modelo de Metadatos propuesto</i>	68
Tabla 12: <i>Formatos de los Archivos de Imagen según el modo de acceso previsto</i>	75
Tabla 13: <i>Características de los formatos de los Archivos de Imagen más comunes</i>	76

Índice de Cuadros

Cuadro 1: <i>Metodología para la recogida de datos cuantitativos y estructurales para el dimensionamiento del proyecto de digitalización</i>	23
Cuadro 3: <i>Ejemplo de recogida de datos referentes al Local de Archivo</i>	25
Cuadro 4: <i>Metodología para la determinación de los parámetros de conversión</i>	34
Cuadro 5: <i>Recogida de Datos referentes al Contenido del Documento</i>	66
Cuadro 6: <i>Ejemplo de asignación de metadatos en varios niveles de agrupación documental</i>	71

Índice de Fórmulas

Fórmula 1: <i>Cálculo del tamaño de la muestra</i>	30
Fórmula 2: <i>Valores más comunes de k y sus niveles de confianza</i>	30
Fórmula 3: <i>Patrón de referencia de la resolución para el escaneado bitonal de textos impresos</i>	52
Fórmula 4: <i>Patrón de referencia de la resolución para el escaneado a color o en escala de grises de textos impresos</i>	52
Fórmula 5: <i>Patrón de referencia de la resolución para el escaneado bitonal de textos manuscritos</i> ...	53
Fórmula 6: <i>Patrón de referencia de la resolución para el escaneado a color o en escala de grises de textos manuscritos</i>	54

RESUMEN

La *Licenciatura en Documentación* impartida por la *Universidad Politécnica de Valencia* capacita, supuestamente, a profesionales para el desarrollo y aplicación de habilidades relativas al análisis de la documentación y los procedimientos que la afectan. Sin embargo, el desarrollo de un nuevo sector de negocio alrededor de los proyectos de digitalización de documentos no ha sido recogido como una disciplina atribuible al desempeño del profesional de la documentación.

El presente Proyecto Final de Carrera pretende, por tanto, elaborar una metodología de trabajo que permita al documentalista participar en la elaboración tanto del proyecto técnico como del presupuesto en el entorno de un proyecto de digitalización. Es decir, formar parte del análisis de requisitos como información de respaldo para el análisis de costes necesario para la preparación del presupuesto y como patrón de referencia para el plan de calidad del producto a entregar¹.

El análisis de los requisitos relativos a la configuración de un proyecto de digitalización externalizado se ha dividido en cuatro bloques temáticos: análisis cuantitativo y estructural, análisis físico, análisis de contenido y análisis de los formatos de entrega. Cada uno de estos estudios se corresponde con una parte de la metodología global, objetivo del presente Proyecto Final de Carrera.

El *Análisis Cuantitativo y Estructural* recoge las variables asociadas a las tareas de recogida e inventariado de la documentación como son las características del local

¹ Project Management Institute. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*. Electronic Imaging Services, 2000. "Project Cost Management" y "Project Quality Management" (Pág. 190).

Project Management Body of Knowledge, es un estándar en la Administración de proyectos desarrollado por el *Project Management Institute (PMI)*. La misma comprende dos grandes secciones, la primera sobre los procesos y contextos de un proyecto, la segunda sobre las áreas de conocimiento específico para la gestión de un proyecto [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/PMBOK>].

de archivo, la previsión de volúmenes a procesar (mediante la aplicación de técnicas estadísticas) y el sistema de archivo (clasificación, ordenación y codificación) que habrá que reproducir y respetar mientras los fondos se encuentren en las instalaciones de la empresa digitalizadora.

El *Análisis Físico* refleja los aspectos necesarios para la determinación de los parámetros de conversión de un documento en papel a un archivo digital. Para ello se propone una metodología basada en la evaluación de los atributos físicos del documento para cuantificar su incidencia en las variables de calidad ligadas al proceso de digitalización y así, en última instancia, determinar un patrón de referencia válido. Entre los atributos físicos del documento se destaca el tipo de documento (según el modo de representación de su contenido), la presencia y significado del color, el grado de legibilidad y el formato del soporte. En cuanto a las variables de calidad, se contemplan como tales la resolución, la profundidad de bits, la compresión y el tipo de escáner.

En el apartado dedicado al *Análisis de Contenido* se define y contextualiza la indización como un proceso de apoyo a la recuperación de los fondos digitalizados y se introducen los puntos de análisis de la documentación física para la extracción de requisitos relacionados con la misma. Se estudia el nivel de asignación requerido para la caracterización de las unidades documentales (expediente, documento compuesto, documento simple), las fuentes de información de dónde extraer los índices y la aplicación de herramientas de control del vocabulario para disminuir la incertidumbre asociada a las búsquedas. Asimismo, se presentan los criterios a valorar para la aplicación de ciertos automatismos que, mediante el aprovechamiento de las sinergias del proceso, pueden repercutir en una disminución del coste asociado al proceso: procesamiento por lotes, aprovechamiento de ficheros cliente y/o técnicas de reconocimiento óptico.

Por último, el *Análisis de los Formatos de Entrega* recoge las principales características y aplicaciones de los distintos formatos de imagen existentes en el mercado. También, aunque tangencialmente, se tratan aspectos relativos al formato de la base de datos resultante con los atributos de indización. El tratamiento de este

apartado es más superficial que los anteriores en tanto que se le otorga un carácter introductorio, ya que presenta aspectos más ligados a la formación de un Ingeniero en Informática que a un Documentalista.

Cerrando el Proyecto Final de Carrera se refiere un *Anexo* con el análisis de los requisitos expresados en el *Esquema Nacional de Interoperabilidad*, aprobado por el Real Decreto 4/2010, de 8 de enero, y su normativa técnica de desarrollo que infiere aspectos tratados en los apartados referidos anteriormente constituyendo una línea de formación e información para los profesionales de la Documentación vinculados a proyectos de digitalización.

CAPÍTULO 1. INTRODUCCIÓN

1.1. Preámbulo

Actualmente, el desarrollo de medios técnicos que facilitan la edición e impresión de documentos ha desembocado en un crecimiento exponencial del volumen de papel que genera cualquier organización. Como es obvio, no toda la documentación que se imprime y fotocopia es vital para el mantenimiento de la actividad, ni siquiera es siempre relevante más allá de la actividad concreta para la que fue generada y, por tanto, podría destruirse o/y reciclarse.

Sin embargo, determinadas series documentales, por su antigüedad, su no disposición en formato electrónico, por la presencia de elementos manuscritos de validación, como la firma, que le otorgan un valor legal probatorio, o por cualquier otro motivo... se conservan indefinidamente en formato papel con los consecuentes problemas de espacio, dificultades para la localización, etc.

La conversión de estos documentos a un formato digital se presenta, por parte de las empresas proveedoras de servicios, como una solución factible a estos problemas siendo indispensable tener en cuenta las características específicas de la documentación y cómo éstas podrían influir en el posterior desarrollo del proyecto.

1.2. Objeto

El objetivo de este proyecto es, por tanto, establecer un modelo para la toma de requisitos documentales, es decir, un modelo válido que permita recoger las condiciones o necesidades documentales del usuario de la documentación, para un diseño óptimo de un proyecto de digitalización en el contexto de una empresa proveedora de los servicios referidos de digitalización.

1.2. Objetivos

En consonancia con el objeto del mismo, el presente Proyecto se ha enfocado en el desarrollo de una pauta para la observación activa de la condición física de los

documentos originales y cómo esto puede influir en el planteamiento del proceso de conversión digital.

En la medida en que los requisitos sean medibles, comprobables, sin ambigüedades ni contradicciones² posibilitarán el establecimiento objetivo de ciertos parámetros del proyecto. Con el fin de desarrollar una metodología que cumpla estas condiciones, los objetivos principales a satisfacer serán los siguientes:

- ✓ Establecer una metodología de recogida y análisis de las informaciones recabadas para su conversión a requisitos;
- ✓ Desarrollar un modelo de toma de datos relativos a los atributos físicos de los documentos a digitalizar.

1.3. Antecedentes

El principal motivo para iniciar la redacción de este trabajo ha sido la inclinación personal por los temas relativos a la digitalización por haber sido una materia que, aún siendo objeto de debate y desarrollo en los foros profesionales, solamente fue tratado tangencialmente durante la Licenciatura de Documentación.

Asimismo, por la orientación profesional de la autora, su labor se viene desarrollando en los últimos años en el ámbito de la definición y diseño de proyectos de digitalización, como fase previa a su ejecución en la que se analizan las variables con un impacto previsible en el desarrollo del mismo.

² *“Success of software development depends on the quality of the requirements specification. Moreover, good – sufficiently complete, consistent, traceable, and testable – requirements are a prerequisite for later activities of the development Project”.* Hasan Kitapci, Barry W. Boehm. *Using a Hybrid Method for Formalizing Informal Stakeholder Requirements Inputs*. Presentado en Fourth International Workshop on Comparative Evaluation in Requirements Engineering, 2006 (CERE '06). [Consultado el 20-08-2011 en http://www.di.unipi.it/CERE06/P5%20CERE%2006%20Kitapci_Hasan]

Se entiende por proyecto “*una empresa o actividad única que tiene un principio y un final, que está conducida por personas y que pretende alcanzar los objetivos establecidos dentro de los parámetros de coste, calendario y calidad*”³, por lo que están orientados a un objetivo, tienen un carácter único o singular, son finitos y resulta necesario planificarlos.

En analogía con lo arriba expuesto puede concluirse que el ciclo vital de un proyecto contempla las fases de análisis y definición, diseño, planificación y ejecución. En concreto, *la fase de análisis y definición*⁴, coincidente con el origen del proyecto, consiste en el análisis del problema o necesidad que se pretende resolver para circunscribirlo en el contexto en cual está presente. Esto es, pues el análisis de sus límites y alcance tras el que se formulan convenientemente los objetivos generales y específicos que se van a perseguir para acotarlo y orientarlo adecuadamente.

Aplicando esta metodología al presente Proyecto Final de Carrera podemos inferir que, en nuestro caso, el problema a analizar será el impacto de las características de los atributos físicos de la documentación en el proceso productivo de un proyecto de digitalización externalizado, siendo el objetivo general desarrollar una metodología que permita chequearlos y analizarlos convenientemente con vistas a determinar su influencia sobre la cadena de producción contribuyendo a ajustar al máximo el presupuesto a ofertar al cliente.

El texto de referencia básico para la determinación de estos requisitos físicos de la documentación han sido las “Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos”⁵, por varios motivos: por ser un trabajo que pretende dar las

³ Dinsmore, 1984. Citado por Ernest Abadal Falgueras en *Gestión de Proyectos en Información y Documentación*. Ediciones TREA, S.L., 2004 (Pág. 18).

⁴ Abadal Falgueras, Ernest. *Op. Cit.* (Pág. 22)

⁵ Grupo de Expertos del International Federation of Library Associations and Institutions (IFLA) e International Council of Archives (ICA) invitados por la UNESCO. *Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en*

pautas básicas a los profesionales del gremio de la documentación sobre las acciones a emprender para la digitalización de documentos en papel, por provenir de fuentes oficiales con gran prestigio internacional y por introducir la importancia de los atributos de los documentos originales.

Este documento ha sido, por tanto, el cabo del que se ha tirado para rescatar el resto de la bibliografía y configurar el trabajo que presentamos como Proyecto Final de Carrera.

1.4. Metodología

La metodología de trabajo utilizada para el desarrollo del presente Proyecto de Final de Carrera, como sería obvio, por otro lado, tratándose de un trabajo desarrollado por estudiante de la Licenciatura de Documentación, ha sido la búsqueda bibliográfica y el análisis de fuentes de información autorizadas en Documentación, pasadas por el tamiz de la experiencia profesional, para su concreción en los distintos apartados del capítulo 2 *“Desarrollo de una metodología para el análisis de requisitos de la documentación física en el contexto del diseño de un proyecto global de digitalización”*.

1.5. Plan del Trabajo

El plan de trabajo ha sido el siguiente:

- ✓ Selección de las palabras clave pertinentes a la temática del proyecto:
 - Consulta del término “digitalización” en el Tesoro de Biblioteconomía y Documentación⁶ del Instituto de Estudios Documentales sobre

bibliotecas y archivos, 2002 (Pág. 24). [Consultado el 29-03-2011 en www.mcu.es/archivos/docs/pautas_digitalizacion.pdf]

⁶ Mochón Bezares, Gonzalo y Sorlí Rojo, Ángela. Tesoro de Biblioteconomía y Documentación. Instituto de Estudios Documentales sobre Ciencia y Tecnología (antes CINDOC). Consejo Superior de Investigaciones Científicas (CSIC).

Ciencia y Tecnología (IEDCYT)⁷ del Consejo Superior de Investigaciones Científicas (CSIC)⁸.

- Concreción del mapa conceptual asociado a la temática del proyecto mediante la selección de las palabras clave relativas al concepto “digitalización”: *digitalización, tratamiento de imágenes, escáneres, imágenes electrónicas, formatos digitales*.
- ✓ Búsqueda bibliográfica, mediante el uso de las palabras clave seleccionadas, relativa a proyectos de digitalización en fuentes de prestigio en el ámbito de la Documentación (repositorios documentales, revistas especializadas, referencias del Ministerio de Cultura, asociaciones profesionales, etc.);
- ✓ Elaboración de fichas de lectura relativas a cada una de las fuentes. En cada ficha se reseñaron los siguientes datos:
 - Reseña bibliográfica de la fuente (título, autor, datos de la edición y datos de acceso);
 - Resumen del contenido;

[Consultado el 27/02/2011 en http://thes.cindoc.csic.es/index_BIBLIO_esp.php]

⁷ *Instituto de Estudios Documentales sobre Ciencia y Tecnología* constituido a partir del antiguo Centro de Información y Documentación Científica (CINDOC). Es un organismo del Consejo Superior de Investigaciones Científicas (CSIC), cuya principal línea de actuación es analizar, recopilar, difundir y potenciar la información científica en todas las áreas del conocimiento [Consultado el 25-09-2011 en <http://www.cindoc.csic.es>].

⁸ El *Consejo Superior de Investigaciones Científicas* es una agencia estatal adscrita al Ministerio de Ciencia e Innovación, a través de la Secretaría de Estado de Investigación cuyo objetivo fundamental es desarrollar y promover investigaciones en beneficio del progreso científico y tecnológico, para lo cual está abierta a la colaboración con entidades españolas y extranjeras. Tiene carácter multidisciplinar y realiza investigaciones avanzadas en todas las áreas científicas gracias a sus más de 130 centros distribuidos en todas las Comunidades Autónomas de España [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/CSIC>].

- Asignación de nuevas palabras clave (con un carácter más específico que las seleccionadas inicialmente).

- ✓ Análisis de los datos recogidos en las fichas de lectura: se “desprecian” las fuentes cuyo contenido no coincida con el núcleo temático del proyecto, es decir, se desecha el “ruido documental” generado por el uso de palabras clave de tipo genérico. Los resultados pertinentes se utilizan para “retroalimentar” la búsqueda bibliográfica. Se presta atención a las citas y bibliografía de estas fuentes;

- ✓ Diseño de la estructura general del Proyecto Final de Carrera estableciendo los principales apartados del capítulo 2 “*Desarrollo de una metodología para el análisis de requisitos de la documentación física en el contexto del diseño de un proyecto global de digitalización*” en epígrafes coincidentes con los temas recurrentes en la bibliografía;

- ✓ Búsqueda bibliográfica para completar los aspectos definiciones, desarrollo de siglas, recuperación de la fuente primaria relativa a una cita, etc.

- ✓ Redacción, edición y corrección de errores.

1.6. Estructura Documental del Proyecto Final de Carrera

Además de los apartados que dan estructura al documento como la introducción, bibliografía y conclusiones, el Proyecto Final de Carrera se estructura en torno a los siguientes núcleos temáticos:

- ✓ **Análisis Cuantitativo y Estructural:** En las *Directrices para proyectos de digitalización de colecciones y fondos de dominio público (...)*⁹ se refiere que “*si el material que se ha considerado candidato para la digitalización tiene*

⁹ Grupo de Expertos de IFLA e ICA. Op Cit. Sección 1 Selección, pág. 38 y Sección 4 Presupuesto, Pág. 90.

*lagunas en su catalogación o en la descripción de sus datos (...) será necesario considerar si los costes adicionales de este trabajo pueden incluirse en el presupuesto total del proyecto de digitalización". Se considera, por tanto, que existe una necesidad de determinar la magnitud y el coste asociado a las tareas de recogida e inventariado, sin obviar la planificación operativa de las mismas. En este apartado se recogen las variables que se han considerado de interés para la caracterización del esfuerzo requerido para su concreción como las características de las unidades de instalación, su localización y forma de ordenación. Se ha considerado oportuno titular el bloque temático como *Análisis Cuantitativo y Estructural*¹⁰ en tanto que supone un estudio de las magnitudes físicas asociadas a determinado fondo o colección y la distribución de sus partes.*

- ✓ **Análisis Físico:** La norma *UNE-ISO 15489 Información y Documentación. Gestión de Documentos*¹¹ considera que la conversión digital, en tanto que

¹⁰ Real Academia Española. Diccionario de la Lengua Española. 22ª edición, 2001. [Consultado el 20/09/2011 en www.rae.es]:

- *Análisis Cuantitativo: Método que se ocupa del análisis de las dimensiones de las magnitudes físicas, y que permite establecer directamente relaciones entre las que intervienen en un proceso, sin necesidad de realizar un análisis completo y detallado.*
- *Estructura: Distribución de las partes del cuerpo o de otra cosa.*

¹¹ UNE-ISO 15489 - Información y Documentación. Gestión de Documentos. 4.3.7.1. Decisiones en materia de almacenamiento de documentos. Págs. 118 - 119. [Consultada en Revista Española de Documentación Científica, nº 28, Vol. 1, 2005).

Una Norma Española (UNE), son un conjunto de normas tecnológicas creadas por los Comités Técnicos de Normalización (CTN), de los que forman parte todas las entidades y agentes implicados e interesados en los trabajos del comité. Por regla general estos comités suelen estar formados por AENOR, fabricantes, consumidores y usuarios, administración, laboratorios y centros de investigación [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Una_Norma_Española].

Organización Internacional de Normalización o ISO (del griego "isos", "igual"), nacida tras la Segunda Guerra Mundial (23 de febrero de 1947), es el organismo encargado de promover el desarrollo de normas internacionales de fabricación, comercio y comunicación para todas las ramas industriales a

proceso documental, debe contemplar como requisito *“las características físicas de los documentos”*. En este sentido, se ha considerado pertinente titular este apartado como *Análisis Físico en la medida que atañe a la descripción de las características físicas de la documentación (formatos, estado físico de conservación y tipo documental)* con vistas a determinar su incidencia en el diseño del proceso de digitalización.

- ✓ **Análisis de Contenido:** De acuerdo a la norma *ISO 5963:1985 sobre metodología de indización*, *“la indización es el proceso de describir o representar el contenido temático de un recurso de información. Este proceso da como resultado un índice de términos de indización que será utilizado como herramienta de búsqueda y acceso al contenido de recursos en sistemas de recuperación de información”¹²*. De modo que la identificación del tipo de documentación que integra cada serie documental de cara a la configuración de un esquema de acceso en un entorno de software,

excepción de la eléctrica y la electrónica. Su función principal es la de buscar la estandarización de normas de productos y seguridad para las empresas u organizaciones a nivel internacional. La ISO es una red de los institutos de normas nacionales de 160 países, sobre la base de un miembro por país, con una Secretaría Central en Ginebra (Suiza) que coordina el sistema. La Organización Internacional de Normalización (ISO), con sede en Ginebra, está compuesta por delegaciones gubernamentales y no gubernamentales subdivididos en una serie de subcomités encargados de desarrollar las guías que contribuirán al mejoramiento ambiental. Las normas desarrolladas por ISO son voluntarias, comprendiendo que ISO es un organismo no gubernamental y no depende de ningún otro organismo internacional, por lo tanto, no tiene autoridad para imponer sus normas a ningún país. El contenido de los estándares está protegido por derechos de copyright y para acceder ellos el público corriente debe comprar cada documento, que se valoran en francos suizos (CHF). Está compuesta por representantes de los organismos de normalización nacionales, que produce normas internacionales industriales y comerciales. Dichas normas se conocen como normas ISO y su finalidad es la coordinación de las normas nacionales, en consonancia con el Acta Final de la Organización Mundial del Comercio, con el propósito de facilitar el comercio, el intercambio de información y contribuir con normas comunes al desarrollo y a la transferencia de tecnologías [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Organización_Internacional_para_la_Estandarización].

¹² Grupo de Expertos de IFLA e ICA. Op.cit. Sección 7 Preservación del Contenido Digital, Pág. 74.

corresponde con tareas relativas al análisis de contenido, título del presente apartado.

- ✓ **Análisis de los Formatos de Entrega:** En las referidas *Directrices para proyectos de digitalización de colecciones y fondos de dominio público (...)* se introduce la necesidad de que el “*formato de fichero de los ficheros de imagen del máster de archivo debe ser (...) interoperable (...) y el esquema de metadatos no debe exigir ningún software específico para ser inteligible (...)*”. Dicha afirmación introduce el requisito de interoperabilidad de los ficheros resultantes de la digitalización constituyentes del producto final a entregar al cliente. Así pues, en este apartado se estudian las características de los distintos formatos de entrega en base a una combinación de variables técnicas relativas a la infraestructura informática y los usos previstos para la documentación digitalizada. Se analiza, en definitiva, las variables técnicas que influirán en el modo de entrega de la documentación digitalizada.

En cada apartado del capítulo se sintetizan los datos necesarios para el establecimiento de estos requisitos y se dan nociones para predecir su impacto sobre la configuración del proyecto.

1.7. Aplicación de los conocimientos adquiridos

A continuación se referencian aquéllas asignaturas de la *Licenciatura en Documentación* de la *Universidad Politécnica de Valencia* cuyos contenidos han tenido mayor aplicación en el desarrollo del presente Proyecto de Final de Carrera.

“*Técnicas Documentales Aplicadas a la Investigación*” y “*Normativa Técnica*” han sido fundamentales para el desarrollo de estrategias de búsqueda bibliográfica utilizando fuentes autorizadas.

Los conocimientos adquiridos en “*Técnicas de Indización y Resumen en Documentación Científica*” se han aplicado a la elaboración de fichas de lectura, determinantes en el plan de trabajo utilizado. Tangencialmente, se han utilizado en

el apartado referente al “Análisis Estructural”, en lo relativo a la selección de criterios de indización.

Por su parte, “*Bases de Datos Documentales*” ha contribuido a ampliar la visión de la alumna sobre la importancia de una correcta asignación de atributos de indización en el funcionamiento de las diferentes utilidades de búsqueda asociadas a una base de datos.

Las nociones de “*Estadística*” han permitido incluir las técnicas relativas al muestreo y determinación del tamaño de la muestra del apartado “Herramientas”.

Y por último, aunque no se encuentra implementada en el plan formativo de la *Licenciatura en Documentación*, se considera importante reseñar la formación recibida en “*Archivística*” en la Diplomatura en Biblioteconomía como base educativa fundamental.

CAPÍTULO 2. DESARROLLO DE UNA METODOLOGÍA PARA EL ANÁLISIS DE REQUISITOS DE LA DOCUMENTACIÓN FÍSICA EN EL CONTEXTO DEL DISEÑO DE UN PROYECTO GLOBAL DE DIGITALIZACIÓN

La base metodológica utilizada como punto de partida para el desarrollo de este capítulo se corresponde con el llamado *método analítico*¹³, consistente en la desmembración de un todo en sus partes para observar sus causas, naturaleza y efectos. En el caso específico de este proyecto, el núcleo lo constituye el “proyecto de digitalización” como concepto de estudio y los elementos a analizar coincidirán con los requisitos de la documentación física.

Por otro lado, relacionándose el análisis de requisitos con la etapa inicial de cualquier proyecto (sea del sector que sea: informática, construcción, documentación, etc.) en la que se deben estudiar tanto las peculiaridades del cliente como los requerimientos concretos (como los que aquí se plantean), se ha recurrido a la consulta de antecedentes referentes al diseño de proyectos, de cualquier ámbito, a fuentes identificables con “literatura gris” (como protocolos de trabajo de empresas privadas, textos de pliegos de prescripciones técnicas de concursos públicos) y a la experiencia profesional.

En este sentido, tanto el apartado 2.1. *Análisis Cuantitativo y Estructural*, como los subsiguientes 2.2 *Análisis Físico*, 2.3 *Análisis de Contenido* y 2.4 *Análisis de los Requisitos de Entrega*, corresponden con la etapa de descripción e identificación de los recursos que intervendrán en el presupuesto del proyecto, significando los requisitos que condicionarán la fase de estimación de la cantidad necesaria de cada uno de ellos¹⁴ (por ejemplo, personal y tiempo dedicado a las tareas de recogida, inventariado, digitalización, etc.).

¹³ Ruiz Limón, Ramón. *Historia y evolución del pensamiento científico*, 2007. [Consultado el 20/09/2011 en www.eumed.net/libros/2007a/257/]

¹⁴ Torralba Martínez, José María. *Introducción al Presupuesto de Proyectos Informáticos y Telemáticos: Costes para el proveedor*. Departamento de Organización de Empresas, Economía

Así pues, en analogía con los proyectos de construcción, *“para obtener un buen presupuesto de un proyecto es preciso realizar (...) las mediciones de las distintas unidades de obra que lo componen (...)”*¹⁵. El presente Proyecto Final de Carrera pretende, por tanto, sintetizar estas “unidades” agrupadas, tal y como se referencia a continuación, alrededor de cuatro bloques temáticos, coincidentes con los distintos apartados del capítulo. Tal y como se ha apuntado en la introducción, cada uno de ellos corresponde con una fase o componente de la metodología que se pretende desarrollar en este Proyecto.

2.1. ANÁLISIS CUANTITATIVO Y ESTRUCTURAL

2.1.1. Local de Archivo

2.1.2. Muestreo Cuantitativo

2.1.3. Sistema de Archivo

2.2. ANÁLISIS FÍSICO

2.2.1. Evaluación de los Atributos Físicos de los Documentos

2.2.2. Caracterización Objetiva de las Variables de Calidad

2.2.3. Determinación de un Patrón de Referencia mediante la correlación entre Variables de Calidad y Atributos Físicos

2.3. ANÁLISIS DE CONTENIDO

2.3.1. Definición y contextualización de la indización como proceso de apoyo a la recuperación de fondos digitalizados

2.3.2. Puntos de análisis de la documentación física para la extracción de requisitos relacionados con la indización

2.3.3. Automatismos

Financiera y Contabilidad. Facultad de Informática. Universidad Politécnica de Valencia, Editorial de la UPV, 2003 (Pág. 28 – 29).

¹⁵ Torralba Martínez, José María (et. al.). El Proceso presupuestario en proyectos de construcción. Departamento de Organización de Empresas, Economía Financiera y Contabilidad. Escuela Superior de Ingeniería del Diseño. Universidad Politécnica de Valencia. Editorial de la UPV, 2003 (Pág. 34).

2.4. ANÁLISIS DE LOS FORMATOS DE ENTREGA

2.4.1. Formato del Fichero de Imagen

2.4.2. Formato de la Base de Datos con los Atributos de Indización

2.1. ANÁLISIS CUANTITATIVO Y ESTRUCTURAL

El *Análisis Cuantitativo y Estructural* responde a la necesidad de determinar la magnitud y el coste asociado a las tareas de recogida e inventariado. Se recogen, por tanto, las variables de interés para su planificación operativa, tales como las características de las unidades de instalación, su localización y forma de ordenación.

La mayoría de fuentes consultadas en el ámbito de la Documentación omiten este análisis. Esto es debido a que la mayor parte de la bibliografía referente a digitalización de documentos se orienta a proyectos insourcing de las administraciones públicas

La norma *UNE-ISO 15489 Información y Documentación. Gestión de Documentos* sí contempla, en su apartado 4.3.9.4. *Transferencia de la custodia o propiedad de los documentos* que, “en determinadas circunstancias, la responsabilidad de la custodia o la propiedad de los documentos se transfiere de la organización que los ha creado a otra (...). Cuando esto ocurre, se identifican los documentos que deberían ser transferidos, se retiran de los sistemas de gestión de documentos existentes y se transfieren físicamente¹⁶”. Existe por tanto, una normativa aceptada, con consideración de estándar, que legitima la realización de estos procesos de forma externalizada.

De acuerdo con la observación de las licitaciones públicas de contratos de digitalización los proyectos de externalización oscilan entre dos escenarios posibles: 1) Recogida y traslado de las unidades documentales a las dependencias del proveedor ó 2) Instalación en dependencias del cliente de los medios necesarios para su procesamiento in situ.

La elección entre una u otra modalidad suele decantarla el propio cliente quien, en función de aspectos como el nivel de confidencialidad de la documentación, el control que pretenda efectuar sobre el proceso o la intencionalidad de asumir su

¹⁶UNE-ISO 15489. Op. Cit. Apartado 3. Términos y definiciones.

ejecución, en el futuro, por sus propios medios, podría considerar oportuno realizar el proyecto en sus propias instalaciones.

Sin embargo, en caso de organizaciones con gran acumulación de papel, la solución suele ser una externalización 'pura' consistente en la delegación total de los fondos documentales y los procesos que de su digitalización se deriven en una empresa especializada. Probablemente esto sea debido a que, tal y como indica Emilio Navarro del Peso en su *Manual de Outsourcing Informático*¹⁷, “pocas estrategias de negocio pueden aportar tantas ventajas como el outsourcing” siendo éstas, principalmente, “la flexibilidad en la prestación y en el coste del servicio” y la “reducción de costes” en gastos directos de formación, máquinas, licencias de software, o indirectos de infraestructuras, supervisión y selección.

Así pues, la elección de un planteamiento u otro influye indefectiblemente en la asignación de medios. Las instalaciones en cliente se ven limitadas por las peculiaridades del entorno de trabajo en el que vaya a reproducirse, siempre a menor escala, la cadena de producción y por tanto están sujetas a una casuística diversa.

Por este motivo, se ha considerado que, de cara al presente proyecto, resulta más interesante, en tanto que pueden establecerse una serie de líneas generales de trabajo más homogéneas, presentar un proyecto total de externalización que implique también la recogida de los fondos.

Siguiendo con lo expuesto en la norma *UNE-ISO 15489 Información y Documentación. Gestión de Documentos*, “El objetivo del documento, su formato físico y su uso y valor determinarán las características del equipamiento, así como los servicios requeridos para gestionar el documento durante todo el tiempo que sea

¹⁷ Navarro del Peso, Emilio. *Manual de Outsourcing Informático: Análisis y Contratación*. Ediciones Díaz de Santos, S.A. 2ª ed., 2003. Págs. 25 - 29.

necesario¹⁸”. Extrapolando esta afirmación a las condiciones de recogida de la documentación y su posterior custodia (como mínimo durante el tiempo de proceso) en la empresa digitalizadora deberían tenerse en cuenta, de forma previa a la configuración del mismo “a) El volumen y tasa de crecimiento de los documentos (...)”¹⁹.

En este sentido, el Análisis Cuantitativo y Estructural del fondo responde a esta necesidad de “contabilizar” los volúmenes documentales a procesar e identificarlos preliminarmente. Por tanto, el primer paso debería constituirlo el control de la cantidad de unidades documentales que se trasladan a la empresa digitalizadora, lo cual implicará, necesariamente, realizar un inventario.

Esta información ha de ser recogida siempre de forma previa al igual que aquélla que pudiera originar un fallo de planificación operativa en la recogida. Es más, es información necesaria para el dimensionamiento del proceso con repercusión en el análisis de costes previo al inicio de cualquier proyecto.

La toma de datos expresada no es una tarea arbitraria sino que responde a la necesidad de delimitar, al máximo, la variabilidad que pueda afectar a una previsión de costes asociada al diseño de determinado proceso operativo.

Valga un ejemplo, una empresa del sector de la gestión documental consigue un contrato con una empresa que pretende externalizar la documentación que satura sus oficinas. De acuerdo con las notas tomadas por el comercial en los primeros contactos se conviene que uno de los operarios se desplace a las instalaciones del cliente con una furgoneta para recoger la documentación albergada en las estanterías del departamento de Administración. Según las informaciones facilitadas por el cliente la documentación se agrupa en forma de expedientes ubicados en 250 Archivadores Definitivos (AZ's) formando series documentales:

¹⁸UNE-ISO 15489. Op. Cit. 4.3.7.1. *Decisiones en materia de almacenamiento de documentos*. Págs. 118 - 119.

¹⁹ Véase nota 11.

- ✓ 100 AZ correspondientes a la serie Bancos
- ✓ 150 AZ correspondientes a la serie Contabilidad
- ✓ 20 AZ correspondientes a la serie Clientes
- ✓ 30 AZ correspondientes a la serie Proveedores

Según lo convenido, el operario designado se desplaza a las instalaciones del cliente con la furgoneta, cajas para la ubicación de los AZ y una carretilla para realizar la carga. Una vez allí se encuentra con una serie de eventos no previstos en la planificación:

- ✓ Los AZ's que conforman cada colección no están ubicados en un único punto, sino que resulta necesario recorrer la distancia existente entre el primer piso, donde se ubica el departamento de Administración, y el sótano, donde se encuentra el Archivo de la empresa con la evidente pérdida de tiempo para el operario.
- ✓ El edificio donde se ubican las oficinas es antiguo y la puerta del ascensor es estrecha por lo que no resulta posible subir la carretilla y la carga ha de realizarse en grupos de dos cajas.
- ✓ Las dimensiones de los AZ's superan las previstas por lo que no resulta posible encajarlos todos y algunos viajarán sueltos en la furgoneta con el riesgo consecuente para la pérdida de la unidad documental de los expedientes.

Lo descrito hasta este punto evidenciaría falta de planificación en el ámbito operativo de la empresa, que escaparía al propósito de este proyecto. Se trata de problemas fáciles de predecir por un cargo intermedio con experiencia y recursos. Sin embargo, del supuesto planteado podrían desprenderse otra serie de errores de planificación relativos a la determinación de los datos a reseñar sobre las mismas para el mantenimiento de la trazabilidad:

- ✓ Las series documentales enumeradas por el cliente no son fácilmente identificables por las anotaciones en el lomo de los AZ.
- ✓ Se desconoce el número total de expedientes asociados a cada colección.
- ✓ Exteriormente (en la carpetilla) no se identifica sistema de denominación de los expedientes (código único, título, etc.).

Tomando de nuevo el referente de la norma *UNE-ISO 15489 - Información y Documentación. Gestión de Documentos*, se considera adecuado introducir en este punto el concepto de *registro* como medio para “formalizar la incorporación de un documento en un sistema de gestión de documentos²⁰” como solución a lo expuesto anteriormente. Pero para la implantación de un verdadero registro que aporte valor operativo al proceso resulta imprescindible determinar previamente el nivel de agrupación documental (hoja, expediente, unidad contenedora, etc.) que es necesario reflejar en él (siguiendo el ejemplo anterior, la documentación debería registrarse a nivel de expediente). Además deberán acotarse los índices a registrar, las fuentes de información de las que extraerlos y asignar un identificador único a cada agrupación o unidad documental que permita caracterizarla unívocamente dentro del registro.

La no observancia de estos factores puede originar un caos documental en el ámbito de un proyecto de digitalización por los siguientes motivos:

- 1) Cada serie documental produce expedientes similares entre sí pero diferentes respecto a los de las otras colecciones. La incapacidad de identificar la serie de determinados AZ inhabilita posibles sinergias dentro del proceso de digitalización como sería el procesamiento ordenado por serie con la repetición de tareas similares (ordenación de determinados tipos documentales, expurgo de algunos formatos, etc.). Este hecho obligará a la

²⁰ UNE-ISO 15489. Op. Cit. 4.3.3. Registro. Págs. 114 – 115.

intervención de un superior para el examen de las unidades dudosas y la posterior determinación de la serie documental.

- 2) La unidad digitalizable es el expediente, no el AZ contenedor, por lo que sin una identificación única del mismo no podrán nombrarse los ficheros resultantes de forma significativa (si no es teniendo que 'leer' la documentación) ni podrá llevarse un control efectivo sobre las unidades procesadas respecto a las pendientes de procesar.
- 3) Esta falta de control sobre el proceso, en un entorno de producción multi-proyecto puede desembocar en una situación de pérdida total de la trazabilidad en la que, aún conociendo el número de unidades procesadas (porque los datos referentes a las mismas se almacenen puntualmente en el sistema que sostiene la cadena de digitalización) no se conozca, si no es en base a estimaciones, el volumen de proyecto pendiente de ejecutar.

La parte productiva de una empresa privada, que pretenda ser medianamente competitiva, no puede verse demorada por tareas que excedan la repetición de una serie de automatismos (como tener que leer la documentación relativa a un expediente para asignarle un identificador o realizar un estudio de las series documentales para determinar la pertenencia de determinado fondo).

A continuación se propone la metodología para la determinación de los requisitos necesarios para la correcta planificación de un proyecto de digitalización basada en la recogida de valores cuantitativos y estructurales a nivel de fondo o colección:

En el cuadro nº. 1 se presentan las etapas de la Metodología:

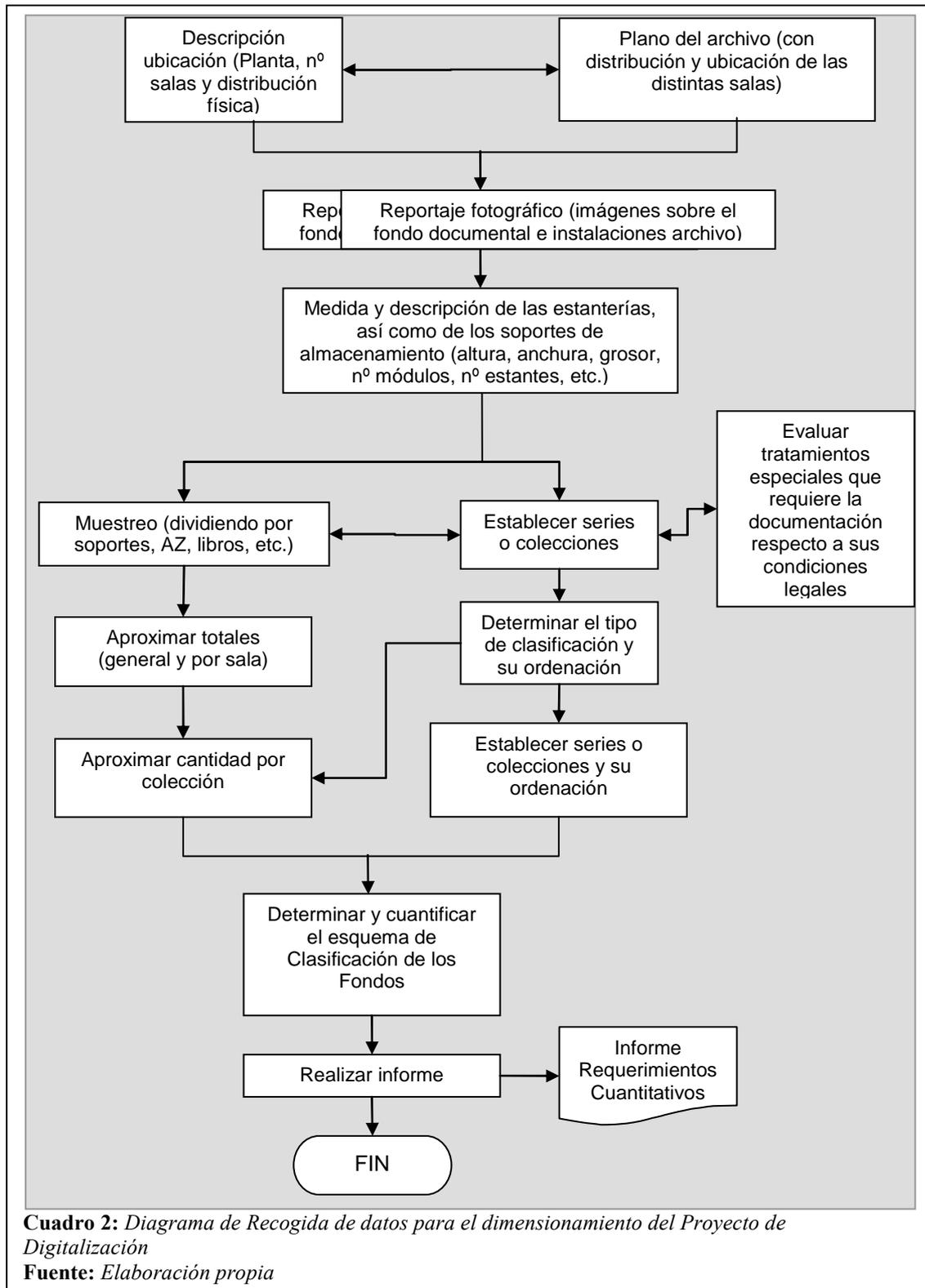
<p>Metodología para la recogida de datos cuantitativos y estructurales para el dimensionamiento del proyecto de digitalización</p> <ul style="list-style-type: none">✓ Local de archivo:<ul style="list-style-type: none">○ Requisitos de recogida✓ Muestreo cuantitativo:<ul style="list-style-type: none">○ Previsión de volúmenes a procesar

- ✓ Sistema de Archivo:
 - Clasificación
 - Ordenación
 - Codificación

Cuadro 1: *Metodología para la recogida de datos cuantitativos y estructurales para el dimensionamiento del proyecto de digitalización*

Fuente: *Elaboración propia*

Desarrollando esta metodología, a continuación se plantea, en forma de diagrama de flujo, la propuesta de actuación presentada en el Proyecto Final de Carrera para la recogida de los datos necesarios para el dimensionamiento del proceso:



2.1.1. Local de Archivo

En este punto se pretende definir, por un lado, las características físicas del local de archivo y las unidades de instalación (datos indispensables para calcular el coste de recogida y el espacio necesario en el almacén para ubicar la documentación) y, por otro, aproximar el volumen de unidades a procesar en función de éstas.

Un ejemplo de recogida de datos referentes al local de archivo sería el siguiente:

Cliente: GESTORÍA ADMINISTRATIVA. Avda. Tarongers, nº 120 - Bajo, Valencia

Características del Local de Archivo:

Sótano, accesible con montacargas de 3 metros ancho X 5 profundo

2 salas contiguas con salida directa al ascensor, puertas de 1,5 m ancho

Sala 1: 3 estanterías, 2 pasillos de 1 metro entre ellas

Sala 2: 1 estantería, la sala es espaciosa, sin problemas de accesibilidad

Unidad de instalación: 4 estanterías fijas de 6 filas con 3 módulos de 95 cm de largo

Soporte de almacenamiento: archivadores definitivos

20 cm alto X 9,5 cm ancho X 15 cm fondo

Cuadro 3: *Ejemplo de recogida de datos referentes al Local de Archivo*

Fuente: *Elaboración propia*

Este supuesto anecdótico es suficiente para denotar la necesidad de complementar estas informaciones con otro tipo de documentación de carácter iconográfico como el plano del archivo y/o fotografías del mismo.

2.1.2. Muestreo Cuantitativo²¹

En la inmensa mayoría de los casos la recogida de requisitos referente a determinado proyecto de digitalización difícilmente podrá ser implementada sobre el total de la documentación. Recordemos que la “unidad” de proceso es la página, constituyendo éste un universo demasiado grande para una revisión total unitaria,

²¹ El contenido de este apartado es una sintetización del manual elaborado por Marín Fernández, Josefa. *Estadística aplicada a las Ciencias de la Documentación*. DM, 2008.

aún enfrentándonos a una colección de unidades documentales relativamente pequeña.

De acuerdo con lo expuesto, el documentalista se halla ante la difícil situación de “seleccionar” las unidades (libros, expedientes, archivadores, etc.) que tomará como muestra para la recogida de los requisitos necesarios para la correcta definición del proyecto. Esta decisión debe fundamentarse sobre una base estadística que minimice el impacto que una selección subjetiva (habitualmente inconsciente) que “desvirtúe” la fiabilidad de los datos obtenidos.

Por otro lado, nos encontramos ante el dilema de determinar el tamaño mínimo de la muestra para garantizar que sea suficientemente representativa respecto del total de unidades documentales.

Intuitivamente, los conceptos introducidos pueden asimilarse a ciencia estadística. Más concretamente el documentalista necesitará conocer el apartado de la estadística que se ocupa de la inducción, que permite hacer generalizaciones de una población²² a partir de una muestra extraída de la misma (estadística inductiva o inferencial).

A través de una muestra representativa de unidades documentales resultará posible obtener información similar a la que se conseguiría del estudio de toda la colección evitando las dificultades derivadas de un exceso de costos. Previamente²³, habremos determinado las características o requisitos (“estadísticos”) sobre los que se pretende recabar los datos (“parámetros”). En resumen, lo que se pretende es obtener información sobre determinada población documental a partir de una muestra.

²² Una población es un conjunto de individuos con propiedades comunes, sobre los que realiza una investigación de tipo estadístico.

²³ Véase apartado 2.2.1 *Evaluación de los Atributos Físicos de los Documentos* del presente Proyecto Final de Carrera.

Mediante un tipo de muestreo adecuado sería posible evitar la selección de una muestra parcial o sesgada que indujese al documentalista a una toma de datos poco representativa y, por tanto, no válida para la valoración de requisitos necesaria para la configuración de un proyecto de digitalización.

Existen cuatro **tipos de muestreos de probabilidad**: aleatorio, sistemático, estratificado y por conglomerados. En todo caso, a partir de muestras obtenidas por cualquier de estos tipos de muestreo es posible realizar inferencias estadísticas. Los muestreos de probabilidad se basan en que todo individuo de la población debe tener una probabilidad conocida de ser incluida en la muestra.

2.1.2.1. Muestreo Aleatorio Simple

Se denomina muestreo aleatorio simple a aquél que cumple las condiciones de que cada elemento de la población tiene la misma probabilidad de ser elegido y que las observaciones se realizan con reemplazamiento. Es decir, que la población es idéntica en todas las observaciones por lo que deberá estar claramente definida antes de que la muestra sea elegida. Se aplica a poblaciones homogéneas.

Si el tamaño de la población (N) es grande respecto al tamaño de la muestra (n) es casi indiferente realizar el muestreo con o sin reemplazamiento.

2.1.2.2. Muestreo Sistemático

Cuando los miembros de la población a estudiar están ordenados se puede utilizar el muestreo sistemático. Para elegir esta muestra se recorre la población tomando cada k -ésimo individuo, empezando por un caso escogido aleatoriamente entre los primeros k individuos. Para conocer el valor de k se divide la población (N) por el tamaño de la muestra que se desea (n).

2.1.2.3. Muestreo Estratificado

El muestreo aleatorio simple se suele utilizar cuando se asume que los individuos de la población son homogéneos respecto a las características a estudiar. Por el

contrario, cuando se dispone de información sobre la población se debe tener presente a la hora de seleccionar la muestra, esto es lo que ocurre en el caso del muestreo estratificado.

El muestreo estratificado es aquel en que los elementos de la población se dividen en clases o estratos (como distintas series documentales o colecciones). Para elegir la muestra se asigna una cuota de miembros a cada clase y después se escogen los elementos por muestro aleatorio simple dentro de cada clase.

Con el muestreo estratificado lo que se pretende es que la muestra sea altamente representativa de la población pues cuando en ésta existen clases o categorías podemos pensar que los miembros de una clase están más relacionados entre sí que con la población vista como un todo. Lo que se intenta es, pues, aumentar la eficacia del diseño del muestreo.

Por ejemplo, para el estudio de la resolución necesaria para representar la copia digital de los documentos producidos por determinada entidad puede resultar interesante analizar por separado los volúmenes producidos por medios manuales y mecánicos, puesto que se estima que, dentro de cada uno de estos grupos, puede haber cierta homogeneidad lo que facilitará la medición de los detalles espaciales finos por el trazo o por el tamaño del carácter mecanoscrito más pequeño (según el caso). Así, si la colección - población está compuesta de un 55% de volúmenes manuscritos y un 45% de volúmenes editados con la ayuda de medios mecánicos, se tomaría una muestra que contuviese también esos mismos porcentajes de ambos.

2.1.2.4. Muestreo por Conglomerados

Puede ocurrir que el tamaño de la población no se conozca, con lo que no sería posible aplicar los muestreos anteriores. En estos casos puede recurrirse a un muestreo por conglomerados. Para realizar este tipo de muestreo partimos de considerar que los elementos de la población se encuentran agrupados en

conglomerados de los que sí conocemos el número. Lo que hacemos es dividir la población en conglomerados y seleccionar entre éstos.

Nótese que la estratificación y el conglomerado responden a planteamientos opuestos. En el muestreo por conglomerados lo que se pretende es que las diferencias entre los estratos sean grandes y que internamente sean homogéneos. En los conglomerados se prevé que existan pocas diferencias entre ellos y que sean muy heterogéneos internamente, es decir, que recojan toda la variabilidad de la población dentro de cada uno.

Por ejemplo, siguiendo lo planteado para el estudio de la resolución en el apartado anterior, ahora nos encontramos ante una colección que se encuentra dividida en expedientes que contienen, cada uno de ellos, documentos manuscritos y editados a máquina. Es decir, que cada expediente recoge toda la variabilidad de la población por lo que resulta necesario seleccionar solo alguno de estos expedientes para la realización del estudio. Por el contrario, en el ejemplo anterior, la población se dividía en volúmenes conteniendo cada uno un solo tipo de documento lo que hacía necesario estratificar la muestra.

2.1.2.5. Tamaño de la Muestra

De forma preliminar a la recogida de datos, será necesario determinar el **tamaño de la muestra**.

En Estadística el tamaño de la muestra es el número de sujetos que componen la muestra extraída de una población, necesarios para que los datos obtenidos sean representativos de la población. El tamaño de una muestra es el número de individuos que contiene.

Una fórmula muy extendida que orienta sobre el cálculo del tamaño de la muestra para datos globales es la siguiente:

$$n = \frac{(k^2) * N * p * q}{(e^2 * (N-1)) + (k^2) * p * q}$$

Fórmula 1: Cálculo del tamaño de la muestra

Fuente: Marín Fernández, Josefa. *Estadística aplicada a las Ciencias de la Documentación*. DM, 2008.

Donde,

N: es el tamaño de la población o universo.

k: es una constante que depende del nivel de confianza que asignemos. El nivel de confianza indica la probabilidad de que los resultados sean ciertos: un 95,5 % de confianza es lo mismo que decir que nos podemos equivocar con una probabilidad del 4,5%. Los valores k más utilizados y sus niveles de confianza son:

k	1,15	1,28	1,44	1,65	1,96	2	2,58
Nivel de confianza (%)	75	80	85	90	95	95,5	99

Fórmula 2: Valores más comunes de k y sus niveles de confianza

Fuente: Marín Fernández, Josefa. *Estadística aplicada a las Ciencias de la Documentación*. DM, 2008.

(Por tanto si pretendemos obtener un nivel de confianza del 95% necesitamos poner en la fórmula k=1,96)

e: es el error muestral deseado. El error muestral es la diferencia que puede haber entre el resultado que obtenemos de una muestra de la población y el que obtendríamos del total de ella.

p: proporción de individuos que poseen en la población la característica de estudio. Este dato es generalmente desconocido y se suele suponer que p=q=0,5 que es la opción más segura.

q: proporción de individuos que no poseen esa característica, es decir, es 1-p.

n: tamaño de la muestra.

2.1.3. Sistema de Archivo

Los archivos son el resultado de la tarea específica de clasificar que engloba tres actividades, de las cuáles se destacan las dos que, según el juicio de la autora, tienen huella en el fondo documental ya conformado (es decir, clasificado):

- La **clasificación** propiamente dicha, en el sentido concreto de constituir unas determinadas clases dentro del archivo, que pueden responder a una realidad o a una convención. Los distintos métodos de clasificación se diferencian en base a tres elementos a considerar para construir los distintos sistemas clasificatorios posibles: las *acciones* (clasificación funcional), las *estructuras* (clasificación orgánica) y los *asuntos* (clasificación por materias).
- La operación de poner en orden las unidades dentro de las series o grupos constituidos. Esta operación es conocida como **ordenación**.

Dentro de cada colección o serie documental, la *clasificación* tendrá su reflejo en la disposición de los fondos en el archivo, es decir, en el criterio utilizado para la agrupación de los mismos. Por su parte, la *ordenación* se evidenciará en el modo en que se colocan las unidades documentales dentro de cada serie o colección. Por ejemplo, dentro de la serie “Licencias de actividad” de un Ayuntamiento los archivadores podrían colocarse ordenados según la fecha de inicio de tramitación del expediente o alfabéticamente por el nombre del establecimiento.

Para facilitar la identificación y agrupación de los expedientes dentro de cada serie documental o colección, resulta necesario asignarle un **código** que, dependiendo del sistema elegido para su conformación, puede resultar más o menos significativo. Habitualmente, el código se encontrará reflejado en el exterior de la unidad contenedora del expediente (carpetilla, sobre, archivador, etc.). La codificación de las unidades documentales puede ser alfabética, numérica o alfanumérica.

En resumen, la determinación de los criterios de clasificación y ordenación, así como el sistema de codificación de los fondos de determinada serie o colección se considera indispensable para la correcta gestión de sus fondos en cuanto a su identificación y el mantenimiento de la trazabilidad durante su procesado. Asimismo, se considera que este sería el estadio directamente anterior al establecimiento de un esquema válido de indización.

Entre los métodos válidos para recabar esta información, se consideran los siguientes²⁴:

- Reuniones y entrevistas con el personal responsable de la documentación;
- Cuestionarios al personal implicado;
- Muestreo.

De acuerdo con lo expresado por el Grupo de Expertos de IFLA e ICA²⁵, resulta recomendable que los ficheros de imagen resultantes de la digitalización se organicen reproduciendo la organización física inicial de los fondos en papel.

Es decir, habría que, en la medida de lo posible, extrapolar los principios fundamentales de la organización física, tomados en la fase de análisis cuantitativo, a la jerarquía lógico de un disco de ordenador.

De igual modo, el sistema de denominación de los ficheros de imagen debe permitir reflejar las relaciones lógicas existentes entre ellas bien mediante la utilización del esquema de numeración del sistema de catalogación física en que se enmarcaban

²⁴ Cubells y M^a José; París, Lidón. *Como implantar un Sistema de Gestión Documental y un sistema archivístico en una institución*. Universitat Jaume I, curso 2006/2007 (Módulo II. Elaboración del cuadro de clasificación, Pág. 19).

²⁵ Grupo de Expertos de IFLA e ICA. Op.cit.

International Federation of Library Associations and Institutions (IFLA) es una organización independiente, internacional, no gubernamental, sin fines de lucro fundada en Edimburgo, Escocia, en 1927. Sus objetivos principales son: promover estándares de provisión y prestación de servicios bibliotecarios y de información, favorecer una mayor comprensión del valor de los buenos servicios bibliotecarios y de información y representar los intereses de sus miembros en el mundo [Consultado el 25-09-2011 en <http://www.ifla.org>].

International Council of Archives (ICA) es una organización profesional de la comunidad de archivos, dedicada a promover la conservación, desarrollo y utilización del patrimonio mundial de los archivos. Reúne a gestores de archivos nacionales, asociaciones profesionales de archiveros, archivos locales y regionales y archivos de otras organizaciones así como archiveros a título individual. CIA es una organización neutra no gubernamental, financiada por sus miembros [Consultado el 25-09-2011 en <http://www.ica.org>].

inicialmente, bien a través de un sistema creado expresamente. Ambos procedimientos son válidos y debe elegirse el que mejor convenga a una colección o grupo determinados de documentos originales.

Las consideraciones referentes al formato de archivo se recogen en el apartado relativo a los *Requisitos de Entrega*. En cualquier caso, el documentalista deberá tener nociones suficientes para asesorar al propietario de la documentación en caso de que así se requiera, sin embargo la toma de decisiones en lo relativo a este apartado y el referido a los requisitos de entrega debería ser responsabilidad del Departamento de Informática de la organización (si existiese) o de su homólogo en la empresa de digitalización (si el propietario de la documentación no dispusiese de él).

2.2. ANÁLISIS FÍSICO

Retomando lo reseñado en el epígrafe 2.1. *Análisis Cuantitativo y Estructural* respecto a las indicaciones en materia de almacenamiento de documentos expuestas por la norma *UNE-ISO 15489 Información y Documentación. Gestión de Documentos*, y considerando la conversión digital como un proceso documental, un requisito a considerar serán “d) Las características físicas²⁶” de los documentos.

Dicho de otro modo, para la correcta parametrización de los outputs de un proceso de digitalización resulta imprescindible recabar información precisa acerca de las características de la documentación física que afectarán a la calidad de la imagen escaneada. La metodología para definir estos requisitos de conversión atenderá a los siguientes aspectos:

Metodología para la determinación de los parámetros de conversión

- ✓ Evaluación de los atributos físicos del documento:
 - Tipo de documento según el modo de representación de su contenido
 - Presencia y significado del color
 - Legibilidad
 - Formato del soporte;

- ✓ Caracterización objetiva de las variables de calidad relevantes del producto digitalizado:
 - Resolución
 - Modo de Escaneo /Profundidad de Bits
 - Compresión
 - Escáner utilizado

- ✓ Determinación del patrón de referencia mediante la correlación entre variables de calidad y atributos físicos

Cuadro 4: Metodología para la determinación de los parámetros de conversión.

²⁶ UNE-ISO 15489 - Información y Documentación. Gestión de Documentos. 4.3.7.1. Decisiones en materia de almacenamiento de documentos. Op. Cit.

Fuente: *Elaboración propia fundamentada en lo expuesto en el apartado 3 “Conversión. Patrón de referencia para la captura digital” del Tutorial de Digitalización de Imágenes de la Universidad de Cornell²⁷ y las Directrices para Proyectos de Digitalización Publicadas por el Ministerio de Cultura²⁸*

Estos parámetros serán los que, tras su confirmación mediante pruebas y evaluaciones, se utilizarán para calibrar el escáner con el cual se procederá a digitalizar los documentos en papel. Siendo así, de la correcta recogida de los requisitos documentales dependerá, en gran medida, una buena parte del proyecto de digitalización.

El objetivo de este análisis será, por tanto, la selección del estándar de imagen que satisfaga los requisitos funcionales y de calidad de la documentación.

2.2.1. Evaluación de los Atributos Físicos de los Documentos

Tal y como se viene argumentando en el texto del presente Proyecto Final de Carrera (PFC), se considera que en la captura de imágenes digitales deben tenerse en cuenta los atributos de los documentos fuente en sí mismos. Intuitivamente resulta previsible que características diferenciales evidentes como la presencia del color, las dimensiones físicas, etc. tendrán una repercusión inevitable sobre la parametrización del proyecto y, por tanto, sobre la calibración de los distintos procesos técnicos involucrados en la conversión digital.

Siendo estos atributos de los documentos definitorios, en cierta medida, del proceso a seguir, en caso de colecciones documentales heterogéneas será necesario separar documentos de acuerdo con sus características (según sean documentos a color o en blanco y negro, según su tipografía o tamaño, etc.). Con esto se evitará una reconfiguración recurrente del escáner o equipo de captura y obteniendo un ahorro sustancial en el tiempo de procesamiento digital.

²⁷ Departamento de Investigación de la Universidad de Cornell. *Tutorial de Digitalización de Imágenes Llevando la Teoría a la Práctica*. [Consultado 21-03-2011 en www.library.cornell.edu/preservation/tutorial-spanish/contents.html].

²⁸ Grupo de Expertos de IFLA e ICA. Op. Cit.

Por el contrario, es muy probable que ni siquiera haya que separar los materiales por alguna característica particular porque éstos sean muy similares entre sí.

En cualquier caso, las valoraciones de las características de los documentos originales a digitalizar y de la calidad de las imágenes digitales deberían basarse en la comparación entre ambas unidades (documento original - imagen digital). El objetivo “no es capturar una imagen con la mayor calidad posible, sino comparar el proceso de conversión con el contenido informativo del original”²⁹.

En los epígrafes subsiguientes se exponen los atributos que permitirán agrupar los documentos originales según su incidencia sobre las variables objetivas a caracterizar. Dicho de otro modo, se recogen las características de los documentos que influirán de forma indefectible sobre los requisitos técnicos (sobre la resolución, el modelo de escáner necesario, etc.).

2.2.1.1. Tipos de documento según el modo de representación de su contenido

Tipo	Representación	Características	Ejemplos
Textuales	Impresos	Puramente textuales y/o con dibujos de líneas simples y bordes definidos, sin variación de tono	Documentos con texto y gráficos simples
	Manuscritos	Textuales cuya tipografía se caracteriza por su representación con bordes suaves. También tablas con tipografía muy pequeña y diagramas con acotaciones. Pueden haber sido elaborados manualmente o producidos por medio de máquina	Documentos escritos a mano o impresos cuya tipografía imite la escritura manuscrita
Gráficos	Media tinta	Materiales gráficos o fotográficos representados por una cuadrícula con un esquema de puntos o líneas de diferente tamaño y espaciadas regularmente.	Imágenes de prensa, cómics, grabados
	Tono continuo	Materiales gráficos o fotográficos representados mediante finas líneas con una variación sutil de tonos	Fotografías, acuarelas y ciertos dibujos
Combinado		Contienen dos o más de las categorías mencionadas anteriormente	Libros ilustrados

Tabla 1: *Tipos de documento según la forma de representación de su contenido.*

²⁹ Grupo de Expertos de IFLA e ICA. Op. Cit. Pág. 44

Fuente: *Elaboración propia basada en los tipos de documentos referenciados en la Introducción del Tutorial de Digitalización de Imágenes de la Universidad de Cornell*³⁰ y el manual *Digitalización de colecciones: texto e imagen editado por la Universidad Nacional Autónoma de México*³¹.

2.2.1.2. Presencia y significado del color en el documento

En el epígrafe anterior se caracterizaban las distintas tipologías documentales según el modo en que se representaba su contenido. Siendo las variables resultantes de esta clasificación válidas para su aplicación al contexto de la determinación del patrón de referencia de la conversión digital, resulta ineludible correlacionarlas con el uso del color en el documento original.

Por ejemplo, un documento de tono continuo como una fotografía puede hallarse representado a color o mediante sutiles variaciones de tono dentro de la gama de los grises (escala de grises). O un documento textual aparecer en blanco y negro o con algunas frases resaltadas a color. Es decir, que dentro de la primera agrupación en la que se englobarían documentos de un mismo tipo podrían diferenciarse subgrupos caracterizados por el uso del color en sus representaciones.

Pero antes de ahondar más en los aspectos sobre la presencia del color como atributo del documento original que pudiera afectar a la imagen resultante en un proceso de escaneado, es necesario introducir el concepto de “significado” del color desde dos vertientes: el color como símbolo y el color como recurso estilístico-artístico.

Podemos afirmar que el color ha sido utilizado como símbolo (y como tal, con un significado inseparable del código en que ha sido representado el documento original) cuando su uso responde a la intención de “comunicar”. Un ejemplo de esta funcionalidad sería la presencia del color en gráficos de barras en los que cada tono fuese referenciado con un significado único en la leyenda. La representación del color sería, por tanto, inherente al significado del contenido del documento.

³⁰ Departamento de Investigación de la Universidad de Cornell. Op. Cit.

³¹ Lara Pacheco, C.G., Castro Thompson, A., Ortiz Ancona, D., et al. Digitalización de colecciones: texto e imagen, 2008. Universidad Nacional Autónoma de México [consultado 03-03-2011 en <http://www.digitalizacion.unam.mx/index.html>]

Sin embargo, en otros casos, el uso del color no responde a la necesidad de “comunicar” sino a una necesidad estilística, como puede ser su uso en títulos u otros elementos utilizados para la maquetación de un texto. La representación de color podría ser perfectamente prescindible en estos casos puesto que no supone riesgo alguno para la representatividad del contenido del documento original. No obstante, hay que ser especialmente cuidadosos con no confundir esta función estilística con el valor artístico y/o histórico atribuible tanto a obras fotográficas y/o pictóricas como a manuscritos, incunables o, incluso, ejemplares impresos. En caso de duda, habrá de ser el responsable de los fondos quien determine la significación del color valorando, entre otros aspectos, el uso que se le pretende dar a la copia digital.

Resulta importante destacar que en caso de que no se considerase necesario reproducir el color en la copia digital de documentos gráficos, no sería una opción convertirlos a blanco y negro, puesto que se perdería parte del rango dinámico³² del documento y, por tanto, de su contenido.

2.2.1.3. Atributos que afectan a la legibilidad del documento

En los epígrafes anteriores se ha tenido en cuenta criterios tales como el tipo de documento y la presencia y significado del color para establecer distintas clasificaciones de los originales a digitalizar. En base al estado físico será posible establecer grupos de mayor a menor legibilidad.

A priori puede resultar difícil realizar una gradación de la legibilidad de un documento mediante una revisión visual por lo que es común analizar separadamente los factores que pueden afectar a la inteligibilidad de un documento:

³² “El rango dinámico es el rango de diferencia tonal entre la parte más clara y la más oscura de una imagen. Cuanto más alto sea el rango dinámico, se pueden potencialmente representar más matices, a pesar de que el rango dinámico no se correlaciona en forma automática con la cantidad de tonos reproducidos” Departamento de Investigación de la Universidad de Cornell. Op. Cit. 1. Terminología Básica.

- ✓ Rango dinámico (variación tonal, contraste).
- ✓ Estado de conservación (manchas, agujeros, arrugas, estado de las tintas, etc.).
- ✓ [En documentos textuales impresos] Altura de la letra minúscula más pequeña (tomar el valor en milímetros).
- ✓ [En documentos gráficos a media tinta y textuales manuscritos] Grosor de línea más fina que deba ser representada (tomar el valor en milímetros)

En el contexto de los documentos textuales será sumamente determinante tener en cuenta la necesidad de tratar separadamente aquellos documentos que pudiesen requerir de un aumento del rango dinámico ante el escaso contraste entre el fondo y el tono en que se encuentra representado el texto (por ejemplo, hojas de papel autocopiable, manchadas u oscurecidas). Esto se resolverá realizando un escaneado a color o en escala de grises, dependiendo del caso.

2.2.1.4. Formato del soporte documental

Siguiendo la categorización de la Digital Libraries Federation (DLF)³³, se ha elaborado la siguiente clasificación de atributos físicos relativos al formato sobre el que se haya “fijado” el contenido informativo del documento:

- ✓ Tipo de soporte (papel, película fotográfica, lienzo, etc.), grosor y textura,
- ✓ Formatos normalizados/ especiales,
- ✓ Formatos simples/ encuadernados.
- ✓ Formatos antiguos y/o frágiles,

³³ Citado en Lara Pacheco, C.G., Castro Thompson, A., Ortiz Ancona, D., et al. Op. Cit.

La *Digital Libraries Federation* es un consorcio de bibliotecas fundado en 1995 que promueve el empleo de tecnologías de información electrónica para ampliar las colecciones y servicios de las bibliotecas. A través de sus miembros, la DLF se dedica a la identificación de estándares para colecciones digitales y redes de acceso, la coordinación de actividades de investigación y desarrollo en el uso de tecnologías de la información, iniciar proyectos y servicios que las bibliotecas necesitan pero no pueden desarrollar individualmente [Consultado el 25-09-2011 en <http://www.clir.org/dlf.html>].

2.2.1.4.1. Tipología de Soportes Documentales: Características y formato

La documentación puede tipificarse sobre distintos soportes, siendo el más frecuente de ellos el papel, sobretodo de cara a un supuesto proceso de digitalización. Sin embargo, existen otros soportes susceptibles de ser digitalizados.

En este apartado se presenta una pequeña muestra de los soportes documentales que, conforme los escáneres disponibles en el mercado, pueden ser procesados de cara a obtener una copia digital así como una enumeración de aquellas variables características que pudiesen repercutir en la calidad de la misma.

PAPEL

Cualidad	Definición	Incidencia en la calidad de la copia digital
Peso – Gramaje	Peso en gramos por unidad de superficie (g/m ²)	Un alto peso o rigidez podrían dificultar la entrada del papel en un escáner de bandeja y/o con alimentación automática
Rigidez	Resistencia al plegado	
Color	Tono	Un color oscuro o un alto grado de transparencia suelen limitar el rango dinámico y dificultar la legibilidad de la imagen digital
Opacidad / Transparencia	Propiedad del papel que reduce o previene el paso de la luz a través de la hoja. Es lo contrario a la transparencia	

Tabla 2: Características del Papel

Fuente: Elaboración propia basada en www.wikipedia.es (18-09-2011.)

Entre los estándares reconocidos relativos a formatos normalizados de documentos en papel destaca la normativa europea ISO / DIN³⁴ (siendo DIN el equivalente alemán a las normas UNE españolas reconocidas por la International Standard Organization) expresada a continuación:

³⁴ El *Deutsches Institut für Normung* es el organismo nacional de normalización de Alemania. Elabora, en cooperación con el comercio, la industria, la ciencia, los consumidores e instituciones públicas, estándares técnicos (normas) para la racionalización y el aseguramiento de la calidad. El DIN representa los intereses alemanes en las organizaciones internacionales de normalización. El acrónimo DIN también ha sido interpretado como *Deutsche Industrie Norm* y *Das Ist Norm* [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/DIN>].

FORMATO NORMALIZADO ISO/DIN A÷E (mm.)					
	A-	B-	C-	D-	E-
4-0	1682 x 2378				
2-0	1189 x 1682				
-0	841 x 1189	1000 x 1414	917 x 1297		
-1	594 x 841	700 x 1000	648 x 917	545 x 779	
-2	420 x 594	500 x 700	458 x 648	385 x 545	
-3	297 x 420	350 x 500	324 x 458	272 x 385	400 x 560
-4	210 x 297	250 x 350	229 x 324	192 x 272	280 x 400
-5	148 x 210	175 x 250	162 x 229	136 x 192	200 x 280
-6	105 x 148	125 x 175	114 x 162	96 x 136	140 x 200
-7	74 x 105	88 x 125	81 x 114	68 x 96	
-8	52 x 74	62 x 88	57 x 81		
-9	37 x 52	44 x 62	40 x 57		
-10	26 x 37	31 x 44	28 x 40		

Tabla 3: Formato Normalizado Papel ISO/DIN
Fuente: Normativa ISO/DIN

Existen otros formatos³⁵ que, pese a no estar reconocidos por la normativa ISO/ DIN cuentan con un alto nivel de aceptación:

- ✓ Folio: 220x320 mm.
- ✓ Cuartilla: 220x160 mm.
- ✓ Octavo u octavilla: la mitad de una cuartilla
- ✓ Holandesa (en desuso): 220x280 mm.
- ✓ Carta: 279x216 mm.
- ✓ Oficio: 330x216 mm.
- ✓ Tabloid: 432x279 mm.

TRANSPARENCIAS

Las láminas de acetato se utilizan como soporte de impresión para documentos en formato transparencia. Este material presenta mayor rigidez que el papel y, como es obvio, un alto grado de transparencia por lo que deberá ser procesado con ayuda de un escáner especial.

³⁵ Departamento de Investigación de la Universidad de Cornell. Op. Cit.

En cuanto a los formatos, los más habituales son aquellos expuestos para el papel.

PELÍCULAS FOTOGRÁFICAS

Entre los formatos fotográficos digitalizables se encuentran las películas fotográficas, conocidas habitualmente como negativos. Los negativos, al igual que las transparencias, presentan mayor rigidez y transparencia que el papel. El formato de la película depende del tipo de cámara utilizado siendo válida la regla de que, a mayor calidad, mayor fotograma:

- ✓ Película de 135mm.: También se denominan de "paso universal" o "standard". Son las más empleadas en fotografía y cine. Tienen por aproximación 35 mm de diagonal (24 x 3mm) y perforaciones por ambos bordes. Se comercializan en carretes de 12, 24 o 36 exposiciones. Para profesionales se venden en latas de varios metros.
- ✓ Película de 120: Esta película tiene 6 cm de ancho. No tiene perforaciones laterales. Según la cámara que se utilice esta película puede ser:
 - negativos 6 x 6 de 12 exposiciones
 - negativos 6 x 7 de 10 exposiciones
 - negativos 6 x 4,5 de 16 exposiciones
 - negativos 6 x 9 de 8 exposiciones
- ✓ Película de 110mm.: Negativo pequeño, mide 16mm de ancho. Cada negativo de 13 x 17 lleva una perforación en el borde. Comercialmente se venden en cassetes de 12 a 24 exposiciones. Actualmente en desuso.

Existen diferentes tipos de películas de entre ellas destacamos las siguientes:

- ✓ Película en color: Esta película se encuentra formada por tres capas de emulsión, siendo cada una de ellas sensible a uno de los colores.

- ✓ Película en blanco y negro: Para imágenes en blanco y negro. En el negativo se observa inversión tonal: las áreas que se visualizan oscuras del negativo son las zonas claras y las zonas claras son las zonas oscuras.
- ✓ Película infrarroja: Con esta clase de películas se logran alteraciones cromáticas para crear efectos artificiales, un campo verde se convierte en blanco.

FOTOGRAFÍAS

La mayoría de los papeles fotográficos son de bromuro de plata, pudiéndose introducirse sobre papel blando o de color recubierto con resina. El papel fotográfico también es sensible a la luz como las películas fotográficas, estos grados de sensibilidad se fijan con valores. Lleva la emulsión sensible a la luz y según su superficie sensible, puede tener diferentes texturas y gradaciones de contraste.

Según la superficie del papel este puede ser mate o brillo.

Según la textura distinguimos:

- ✓ **El papel baritado:** Es un papel más económico. Lleva una capa de barita entre base y emulsión para que sea más blanquecino y proporciona mayor calidad, pero se necesitan tiempos mayores en los procesos químicos del positivado. Las copias realizadas con el papel baritado se mantienen inalterables en el tiempo.
- ✓ **El papel plastificado o resina:** Tiene una capa plastificada que evita que los elementos químicos sean absorbidos, con lo cual, este papel no se humedece mucho consiguiendo un secado más rápido y sin arrugas. Proporcionan una muy buena calidad.

El papel puede tener diferentes tonos, el blanco o blanco cálido (tonalidad un tanto amarillenta).

MICROFILM

Hay tres tipos de película: sales de plata, diazo y vesicular. En el pasado, el acetato de celulosa fue usado como base para el microfilm, pero desde mediados de 1950 casi todos los microfilm son de una base de poliéster, que contribuye a la durabilidad y longevidad de la película.

Según el formato, podemos distinguir:

- ✓ Rollo de 16mm
- ✓ Rollo de 35mm
- ✓ Microfichas

2.2.1.4.2. Formatos normalizados /especiales

Obviamente, existen otros estándares que habría que consultar y conocer en caso de colecciones cuyo soporte fuese distinto al papel.

No obstante, puede concluirse, en cuanto a los formatos especiales, que bajo esta denominación se englobarían todos aquellos documentos no recogidos bajo estándar normalizado o de facto. Será necesario en estos casos, tomar las medidas (alto x ancho) del documento.

2.2.1.4.3. Formatos simples /encuadernados

Otro aspecto a tener en cuenta en lo concerniente al formato de los documentos originales es su condición de documento simple o encuadernado.

Para los documentos encuadernados, típicamente formando libros, cuyo medio de unión no sea fácilmente reversible deberán tenerse en cuenta otros factores como el tamaño de los márgenes de cara al proceso de digitalización ya que en algunos casos puede resultar necesario desencuadernar completamente los ejemplares para poder representar enteramente el contenido (por ejemplo, en encuadernaciones antiguas resulta frecuente que se hayan cosido las páginas ocultando parte de la información en la parte central del libro o legajo).

Asimismo, tal y como se verá más adelante, la condición de documento encuadrado unido a la restricción de no poder alterar esta circunstancia, condicionará forzosamente el tipo de escáner a utilizar.

2.2.1.4.4. Formatos antiguos y/o frágiles

El requisito común de este tipo de documentación será, en cualquier caso, el especial cuidado exigido en su manipulación para minimizar al máximo cualquier riesgo de daño sobre el documento original.

2.2.2. Caracterización Objetiva de las Variables de Calidad

La calidad de la imagen durante la digitalización depende de la suma de resultados de la resolución aplicada al escaneo, la profundidad de bits de la imagen escaneada, los procesos de mejora y el nivel de compresión aplicada, el dispositivo de escaneo utilizado o técnicas usadas, y la preparación del operador del escáner.

2.2.2.1. Umbral de Resolución

El aumento de la resolución permite capturar detalles más precisos. Sin embargo, en algún punto, una mayor resolución no tendrá como resultado una ganancia evidente en la calidad de la imagen, sino un mayor tamaño de archivo. La clave es determinar la resolución necesaria para capturar todos los detalles importantes que están presentes en el documento fuente.

2.2.2.2. Profundidad de Bits

La Profundidad de Bits es determinada por la cantidad de bits utilizados para definir cada píxel. Cuanto mayor sea la profundidad de bits, tanto mayor será la cantidad de tonos (escala de grises o color) que puedan ser representados. Las imágenes digitales se pueden producir en blanco y negro (en forma bitonal), a escala de grises o a color.

- ✓ *Imagen bitonal* está representada por píxeles que constan de 1 bit cada uno, que pueden representar dos tonos (blanco y negro), utilizando los valores 0 para el negro y 1 para el blanco o viceversa. Se recomienda para los documentos que requieren altos niveles de contraste, como los textos impresos. Para su representación se utiliza sólo un bit.
- ✓ *Imagen en escala de grises*: está compuesta por píxeles representados por múltiples bits de información, que típicamente varían entre 2 a 8 bits o más. Se recomienda para documentos de tonos continuos, como fotografías en blanco y negro. Para su representación se utilizan 8 bits: cada punto puede estar dentro de 256 posibles tonos de gris, dando una imagen clara y buena. Es funcional para la mayoría de las imágenes que no tienen colores.
- ✓ *Imagen a color*: está típicamente representada por una profundidad de bits entre 8 y 24 o superior a ésta. Se recomienda para documentos con tonos continuos y de variados colores, como fotografías a color, diapositivas y láminas. Las representaciones a 8 bits (256 colores por punto) funcionan bien con dibujos e imágenes. En cuanto a los 24 bits, la gama se incrementa a 16,8 millones de colores adecuada para la representación de fotografías.

Cuando mayor sea la profundidad de bits, mayor será la cantidad de información recogida y, por ende, será mayor la fidelidad de la imagen con respecto al original. Una profundidad de 8 bits puede representar 256 colores o grados de gris, mientras que una de 24 bits captura más de 16 millones de colores o matices de grises, como ya se ha indicado. Obviamente, una mayor cantidad de información impacta en el tamaño de la imagen:

Profundidad de bits		
1 bit	2 colores	2^1
2 bits	4 colores	2^2
4 bits	16 colores	2^4
8 bits	256 colores	2^8
24 bits	16.777.216 colores	2^{24}

36 bits	68.719.476.736 colores	2^{36}
48 bits	281.474.976.710.656 colores	2^{48}

Tabla 4: Profundidad de Bits**Fuente:** Elaboración propia

2.2.2.3. Compresión

La compresión se utiliza para reducir el tamaño del archivo de imagen para su almacenamiento, procesamiento y transmisión. Las técnicas de compresión abrevian la cadena de código binario de una imagen sin comprimir, a una forma de abreviatura matemática, basada en complejos algoritmos. Existen técnicas de compresión estándar y otras patentadas³⁶.

Los sistemas de compresión también pueden caracterizarse como sin pérdida o con pérdida. Los sistemas sin pérdida, como ITU-T.6³⁷, abrevian el código binario sin desechar información, por lo que, cuando se "descomprime" la imagen, ésta es idéntica bit por bit al original. Los sistemas con pérdida, como JPEG³⁸, utilizan una manera de compensar o desechar la información menos importante, basada en un entendimiento de la percepción visual. Sin embargo, puede ser extremadamente difícil detectar los efectos de la compresión con pérdida, y la imagen puede considerarse "sin pérdida visual".

³⁶ Departamento de Investigación de la Universidad de Cornell. Op, Cit. 1 Terminología Básica.

³⁷ Conocido como CCITT Grupo 4, es un algoritmo de compresión de imágenes de dos dimensiones desarrollado por la ITU [Consultado el 25-09-2011 en <http://www.itu.int>].

³⁸ *Joint Photographic Experts Group*, es el nombre de un comité de expertos que creó un estándar de compresión y codificación de archivos de imágenes fijas. Además de ser un método de compresión, es a menudo considerado como un formato de archivo. JPEG/Exif es el formato de imagen más común utilizado por las cámaras fotográficas digitales y otros dispositivos de captura de imagen, junto con JPG/JFIF, que también es otro formato para el almacenamiento y la transmisión de imágenes fotográficas en la World Wide Web. Estas variaciones de formatos a menudo no se distinguen, y se llaman JPEG [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/JPG>].

2.2.2.4. Escáner utilizado

A continuación se presenta una tabla con los principales tipos de escáner y sus características más destacadas:

Tipo de Escáner	Características
 <p>Escáner de Bandeja</p>	<p>Los escáneres de alimentación múltiple permiten digitalizar grandes volúmenes de documentos. Cuentan con una bandeja en la que se depositan las hojas sueltas y el aparato las ingresa una por una hasta terminar; los documentos pueden ser digitalizados por uno o ambos lados. Se recomiendan para la digitalización de documentos como facturas, notas o documentos similares conformados por hojas sueltas.</p> <p>Sin embargo, también son de gran ayuda para digitalizar documentos como libros, revistas y similares, siempre y cuando se puedan desencuadernar. Aunque el escáner de alimentación múltiple tiene un costo relativamente bajo, su resolución es de apenas 600 dpi³⁹ en color o escala de grises.</p>
 <p>Escáner Plano (Flatbed)</p>	<p>Los escáneres de cama plana regularmente son de bajo costo y fácil operación. Son los más utilizados para cubrir necesidades personales y de pequeñas empresas, así como para procesar documentos, fotografías, libros y otros impresos, cuando es pequeña la colección por digitalizar. Es mejor contar con un escáner especializado si la intención es digitalizar una colección de tamaño considerable. El escáner de cama plana resulta de utilidad en los siguientes casos:</p> <ul style="list-style-type: none"> ✓ El ejemplar no se puede desencuadernar. ✓ Las hojas son demasiado gruesas, como diplomas, reconocimientos, etcétera. ✓ Las hojas sobrepasan las dimensiones de la bandeja de alimentación múltiple. <p>Los escáneres de cama plana pueden alcanzar resoluciones de 1600 dpi o superiores. Al seleccionar este tipo de equipos es importante observar las dimensiones del área de barrido. Los modelos poco sofisticados ofrecen dimensiones de alrededor de 8,5x11", pero los profesionales permiten escanear documentos de mayores dimensiones, como 12x18".</p>

³⁹ Dots Per Inch (o Puntos Por Pulgada -ppp-) es una unidad de medida para resoluciones de impresión, concretamente, el número de puntos individuales de tinta que una impresora o tóner puede producir en un espacio lineal de una pulgada [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Puntos_por_pulgada].

<p style="text-align: center;">Escáner Cenital</p> 	<p>Los escáneres aéreos son superiores en varios aspectos, empezando por el precio. Pero también pueden alcanzar resoluciones mayores a los 600 dpi. Este tipo de dispositivos está enfocado a la digitalización de libros y documentos antiguos o de gran tamaño.</p>
<p style="text-align: center;">Escáner de Tambor</p> 	<p>En el escáner de tambor los materiales son fijados a una especie, precisamente, de tambor o rodillo, el cual gira para que se efectúe la digitalización. Evidentemente, los materiales delicados están descartados para su captura en este tipo de escáneres.</p> <p>Este escáner es usado frecuentemente en el trabajo de pre prensa y en el diseño gráfico con materiales de uso corriente. También se recomienda para capturar materiales como transparencias o negativos de gran tamaño, a partir de los cuales puede obtener una digitalización de gran precisión, alcanzando resoluciones superiores a los 3.000 dpi. Sin embargo, además de requerir mayores conocimientos para su operación y mantenimiento, su costo es alto.</p>
<p style="text-align: center;">Escáner Grandes Formatos</p> 	<p>Los escáneres para grandes dimensiones fueron diseñados para procesar planos y cartas geográficas, además de otros materiales de gran tamaño en cuya captura interviene un sistema de rodillos. En ocasiones el material sufre a través de éstos algún desgarramiento o ruptura, razón por la cual no se recomienda este tipo de escáner para trabajar con material delicado e irremplazable.</p> <p>Estos escáneres alcanzan resoluciones de hasta 400 dpi y tienen una capacidad para digitalizar documentos de hasta 54 pulgadas. Tiene un costo medio en comparación con los escáneres anteriores.</p>
<p style="text-align: center;">Escáner de Ranura</p> 	<p>Los escáneres de ranura están diseñados para digitalizar transparencias como diapositivas, negativos y exposiciones de 35mm, entre otros formatos.</p> <p>Aunque es elevado el costo de este tipo de escáneres, esto se ve compensado con la calidad del producto final, pues alcanzan resoluciones de hasta 4,000 dpi. Generalmente estos escáneres producen una imagen más refinada y detallada que los de cama plana. Esto se debe a que las transparencias reciben mayores cantidades de luz y tienen una mejor definición que los impresos.</p>

<p>Escáner de Microfilm</p> 	<p>El escáner de microfilm fue concebido para la digitalización de rollos de película y microfichas. La calidad o resolución que brinda no es totalmente buena aunque alcanza los 1200 dpi. La definición de las imágenes depende directamente del estado en que se encuentre la película. Su funcionamiento es complicado y costoso y actualmente son pocas las empresas que lo fabrican.</p>
<p>Cámaras Digitales</p> 	<p>En términos generales, las cámaras más comunes en el mercado no cuentan todavía con suficiente resolución para generar archivos <i>master</i> o garantizar la calidad suficiente que las convierta en el formato que preserve un original análogo.</p> <p>No obstante, resultan prácticas para digitalizar algunas colecciones como carteles y mapas, así como obras de arte, material tridimensional o de gran tamaño, entre otras, que difícilmente tienen cabida en un escáner convencional. Al igual que los escáneres, cuanto mayor sea la resolución de la cámara, mejor será la calidad de la imagen.</p> <p>Una forma de mantener niveles de fidelidad aceptables en la digitalización, es el empleo de una cámara identificada como <i>digital scan back</i>. Se trata de una cámara que, provista de un conjunto de sensores, digitaliza una imagen en el anverso y el reverso.</p> <p>Sin embargo, para trabajar con una cámara <i>digital scan back</i> se requiere la asesoría de expertos en su manejo para determinar factores como iluminación, tipos de lentes a usar, etcétera. Por lo regular su costo rebasa el presupuesto de proyectos de mediana envergadura. Se recomienda, por lo tanto, considerar la viabilidad de trabajar con negativos o equipo menos costoso. Carteles, mapas, material tridimensional o de gran tamaño.</p>

Tabla 5: Tipos de Escáner

Fuente: Elaboración propia basada en el artículo *Digitalización de colecciones: texto e imagen*⁴⁰.

⁴⁰ Lara Pacheco, C.G., Castro Thompson, A., Ortiz Ancona, D., et al. Digitalización de colecciones: texto e imagen, 2008. Universidad Nacional Autónoma de México [consultado 03-03-2011 en <http://www.digitalizacion.unam.mx/index.html>].

2.2.3. Determinación del Patrón de Referencia de las Variables de Calidad mediante su correlación con los Atributos Físicos analizados

A continuación se introducen una serie de fórmulas, estándares y recomendaciones técnicas para la determinación objetiva del patrón de referencia de las variables de calidad mediante la caracterización de los atributos físicos del documento:

2.2.3.1. Patrón de Referencia de la Resolución basado en los atributos Tipo de Documento, Legibilidad y Color

La resolución ha de ser la suficiente para capturar los más sutiles detalles requeridos por el grupo de documentos originales que se van a escanear. Siendo así, la determinación de su valor deberá regirse por consideraciones objetivas, basadas en la adopción de estándares reconocidos de acuerdo con los requisitos de la documentación recogidos durante el *Análisis Físico*.

2.2.3.1.1. Resolución de referencia para documentos textuales impresos

La fórmula de Índice de Calidad (QI) para textos impresos desarrollada por el Comité de Normas C10⁴¹ de la AIIM⁴² relaciona calidad (QI) con el tamaño del carácter (h) en mm y la resolución (dpi). Esta norma prevé los siguientes niveles de calidad de imagen:

- ✓ Apenas legible (3,0)
- ✓ Mínima (3,6)
- ✓ Buena (5,0)
- ✓ Excelente (8,0)

⁴¹ Comité de la AIMM (Association for Information and Image Management) responsable del establecimiento de procedimientos de control de calidad y estándares del documento origen y otras especificaciones en el ámbito de la conversión digital [consultado el 18-07-2011 en www.aiim.org].

⁴² Association for Information and Image Management, fundada en 1943 como la Asociación Nacional del Microfilm, es una organización internacional sin ánimo de lucro enfocada a estudiar y difundir los avances relacionados con la gestión documental, de contenidos y procesos de negocio [consultado el 18-07-2011 en www.aiim.org].

A continuación se refiere la fórmula para el escaneado bitonal que proporciona una generosa muestra en exceso para compensar errores de registro y reducción de calidad debido a que limita el umbral de la información a píxeles blancos y negros:

dpi=3QI/0,039h	Ejemplo: <i>Determinar la resolución necesaria para obtener un nivel de calidad excelente escaneando en blanco y negro textos cuya letra minúscula más pequeña tiene una altura de 1 mm.</i>
	$\text{dpi} = 3 \cdot 8 / 0,039 \cdot 1 \Rightarrow 24 / 0,039 \Rightarrow 615$ <p><i>La resolución necesaria sería de unos 600 dpi</i></p>

Fórmula 3: Patrón de referencia de la resolución para el escaneado bitonal de textos impresos⁴³

Fuente: Comité de Normas C10 de la AIIM adaptadas y perfeccionadas por la Universidad de Cornell⁴⁴

Como se ha referido anteriormente, la fórmula reseñada corresponde al escaneado bitonal, no obstante, algunos textos impresos requerirán ser escaneados en escala de grises o a color por motivos derivados del uso y significado del color o de la legibilidad del documento (según lo expuesto en el apartado 2.2.1 referente a la *Evaluación de los Atributos Físicos de los Documentos*).

Debido a que tanto en las imágenes a color como en escala de grises se combinan píxeles en distintos colores (asumiendo las variaciones sutiles de gris como “colores” diferentes), existe una segunda fórmula que determina un patrón de referencia de la resolución válido para ambos casos:

dpi=2QI/0,039h	Ejemplo: <i>Determinar la resolución necesaria para obtener un nivel de calidad excelente escaneando a color o en escala de grises textos cuya letra minúscula más pequeña tiene una altura de 1 mm.</i>
	$\text{dpi} = 2 \cdot 8 / 0,039 \cdot 1 \Rightarrow 16 / 0,039 \Rightarrow 410$ <p><i>La resolución necesaria sería de unos 400 dpi</i></p>

Fórmula 4: Patrón de referencia de la resolución para el escaneado a color o en escala de grises de textos impresos

Fuente: Comité de Normas C10 de la AIIM adaptadas y perfeccionadas por la Universidad de Cornell

⁴³ El valor 0,039, expresado tanto en esta fórmula como la referente al escaneado en color o escala de grises, responde a la necesidad de convertir a pulgadas el valor *h* para correlacionarlo con los *dpi*: 1 mm = 0,039 pulgadas; 1 pulgada = 25,4 mm (N. de la A.).

⁴⁴ Todas las fórmulas citadas en el apartado 2.2.3.1 *Patrón de Referencia de la Resolución (...)* del presente Proyecto Final de Carrera han sido extraídas de 3. Conversión. Tutorial de Digitalización de Imágenes *Llevando la Teoría a la Práctica*. Departamento de Investigación de la Universidad de Cornell [consultado 21-03-2011 en www.library.cornell.edu/preservation/tutorial-spanish/contents.html]

Puede observarse en la fórmula que, al contrario del escaneado bitonal, para el escaneado a color o en escala de grises no es necesario minimizar la incidencia y repercusión de los errores de registro.

2.2.3.1.2. Resolución de referencia para documentos textuales manuscritos

Los manuscritos y otro tipo de materiales gráficos basados en bordes definidos como mapas, bosquejos o grabados, no ofrecen una métrica fija equivalente a la de los textos impresos. En estas tipologías la mejor representación del detalle será el ancho de línea, trazo o la marca más fina que deba ser representada digitalmente.

Se establece una relación funcional entre QI, dpi y el ancho de trazo (w) determinado en mm. En este caso, QI se basa en la evaluación de la calidad que se relaciona con la cantidad de píxeles que cubren el trazo o índice de calidad del mismo⁴⁵:

- ✓ Entre mala y inaceptable (<1,0)
- ✓ Cuestionable (a confirmar según la calidad en pantalla) (1,0)
- ✓ Buena (1,5)
- ✓ Excelente (2,0)

Para el escaneado bitonal, la fórmula a aplicar será la siguiente:

dpi=1,5QI/0.039w	Ejemplo: <i>Determinar la resolución necesaria para obtener un nivel de calidad excelente escaneando en blanco y negro un texto manuscrito cuyo trazo más fino mide 0,1mm.</i>
	$\text{dpi} = 1,5 \cdot 2 / 0,039 \cdot 0,1 \Rightarrow 3 / 0,0039 \Rightarrow 769$ <i>La resolución necesaria sería de unos 800 dpi</i>

Fórmula 5: Patrón de referencia de la resolución para el escaneado bitonal de textos manuscritos

Fuente: Universidad de Cornell

⁴⁵ 3. Conversión. Tutorial de Digitalización de Imágenes *Llevando la Teoría a la Práctica*. Departamento de Investigación de la Universidad de Cornell [consultado 21-03-2011 en www.library.cornell.edu/preservation/tutorial-spanish/contents.html]

Para el escaneado en escala de grises / color la fórmula de calidad basada en el trazo será:

dpi=Ql/0.039w	Ejemplo: <i>Determinar la resolución necesaria para obtener un nivel de calidad excelente escaneando a color o en escala de grises un texto manuscrito cuyo trazo más fino mide 0,1mm.</i>
	$\text{dpi} = 2/0,039 \cdot 0,1 \Rightarrow 2/0,0039 \Rightarrow 512$ <i>La resolución necesaria sería de unos 500 dpi</i>

Fórmula 6: Patrón de referencia de la resolución para el escaneado a color o en escala de grises de textos manuscritos

Fuente: Universidad de Cornell

No obstante, no debe obviarse que muchos de los documentos que se engloban bajo esta categoría exhiben características que van más allá de una simple representación en base a bordes definidos y que, por tanto, la resolución no será el único factor determinante en la calidad de la imagen.

Así el proceso de digitalización exigirá ser especialmente cuidadosos en la evaluación de los atributos físicos, estudiar los estándares y/o recomendaciones técnicas que pudiesen haber surgido en el contexto profesional y/o de la institución a la que pertenece la documentación y, por último, realizar cuántas pruebas de digitalización con distintas parametrizaciones fuesen necesarias para determinar el patrón de referencia adecuado.

2.2.3.1.3. Resolución de referencia para documentos gráficos a media tinta

Las medias tintas son particularmente difíciles de capturar en forma digital, ya que la plantilla de la media tinta y la cuadrícula de la imagen digital con frecuencia entran en conflicto generando imágenes distorsionadas.

Según el Departamento de Investigación de la Universidad de Cornell⁴⁶, una captura de 8 bits a 400 dpi suele considerarse suficiente. Sin embargo, se recomienda realizar varias pruebas oscilando estos valores hasta obtener la combinación más eficiente.

⁴⁶ 3. Conversión. Tutorial de Digitalización de Imágenes *Llevando la Teoría a la Práctica*. Departamento de Investigación de la Universidad de Cornell [consultado 21-03-2011 en www.library.cornell.edu/preservation/tutorial-spanish/contents.html]

2.2.3.1.4. Resolución de referencia para documentos gráficos de tono continuo

La resolución para fotografías y otros documentos de tono continuo es difícil de determinar ya que no hay una métrica fija evidente para medir el detalle. Los detalles se pueden definir como partes de escala relativamente pequeña en un documento, pero esta valoración también puede ser muy subjetiva. En el nivel granular, el medio fotográfico se caracteriza por grupos aleatorios de tamaño y forma irregular, que pueden prácticamente no tener sentido o ser difíciles de distinguir del ruido del fondo.

Lo importante es recordar que la reproducción del tono y del color es tan importante como la resolución, sino más, para determinar la calidad de la imagen. Por tanto, es recomendable efectuar varias pruebas variando estos indicadores hasta obtener el nivel de calidad deseado.

A continuación se presentan, a modo de referencia, los requisitos de calidad en la digitalización para fotografías publicados por el *Arxiu del Regne de València*:

Tipo de Material	Formato	Resolución
Opacos Blanco y Negro	10 x 15 cm.	300 dpi
	15 x 20 cm.	
	20 x 24 cm.	
	24 x 30 cm.	
Opacos Color	10 x 15 cm.	300 dpi
	15 x 20 cm.	
	20 x 24 cm.	
	24 x 30 cm.	
Negativos Blanco y Negro	Paso universal de 35 mm.	2.600 dpi
Negativos Blanco y Negro	Formato medio, 4 x 5,5 cm.	1.800 dpi
Negativos Blanco y Negro	Formato medio, 6 x 7 cm.	1.200 dpi
Negativos y Diapositivas Color	Paso universal de 35 mm.	2.600 dpi
Negativos y Diapositivas Color	Formato medio, 4 x 5,5 cm.	1.800 dpi
Negativos y Diapositivas Color	Formato medio, 6 x 7 cm.	1.200 dpi
Placas	10 x 15 cm	800 dpi

Tipo de Material	Formato	Resolución
Placas	15 x 20 cm	600 dpi
Placas	20 x 25 cm	500 dpi

Tabla 6: Patrón de referencia de la resolución para fotografías como documentos gráficos de tono continuo
Fuente: Artículo publicado por el Arxiu del Regne de València⁴⁷

2.2.3.2. Patrón de Referencia del Modo de Escaneo /Profundidad de Bits basado en los atributos Tipo de Documento y Color

La profundidad de bits, siguiendo la metodología propuesta al inicio del apartado 2.2 *Análisis Físico* del presente Proyecto Final de Carrera, deberá estar en consonancia con las características de los documentos originales y con el nivel de calidad necesario para transferir su contenido informativo. De este modo, las decisiones sobre la profundidad de los bits deben tener en cuenta tanto el aspecto físico del documento, o de sus partes, como el supuesto valor informativo del uso del color en el mismo. Es decir, que, más allá del resultado del análisis físico a realizar por el documentalista, el responsable de la documentación (u organismo autorizado) deberá determinar la “necesidad” o no de representar el color tal y como aparece en el documento original.

Obviando estas cuestiones y siguiendo lo expuesto podemos establecer el siguiente cuadro sinóptico del patrón de referencia de la profundidad de bits de acuerdo con los atributos tipo de documento y color presentes en el documento original:

Tipo de Documento	Color	Legibilidad	Modo Escaneo	de	Profundidad de Bits	de
Textual	B/N	Contraste Alto	Bitonal		1 bit	
Textual	B/N	Contraste Bajo Anotaciones manuscritas	Escala de Grises		8 bits	
Gráfico – Media Tinta	B/N	-	Escala de Grises		8 bits	
Gráfico – Media Tinta	Color	-	Color		24 bits	
Gráfico – Tono Continuo	B/N	-	Escala de Grises		8 bits	

⁴⁷ Giménez Chornet, V., Campuzano Juan, F. *Recomendaciones técnicas para la captura y digitalización de fondos de archivos*. Arxiu del Regne de València, 2005. Consultado el 29-03-2011 en [dglab.cult.gva.es/ArxiuRegne/documents/Recdigitalizacion.pdf].

Tipo de Documento	Color	Legibilidad	Modo de Escaneo	Profundidad de Bits
Gráfico – Tono Continuo	Color	-	Color	24 bits

Tabla 7: Patrón de referencia de la profundidad de bits

Fuente: Elaboración propia basada en el artículo *Digitalización de colecciones: texto e imagen*⁴⁸ y en las *Directrices para proyectos de digitalización de colecciones y fondos de dominio publico*⁴⁹.

Para la determinación de la profundidad de bits asociada al modo de escaneo se ha tomado como referente lo expuesto en las citadas *Directrices para proyectos de digitalización (...)* “El mejor nivel de escala de grises es de ocho bits por píxel (...) se considera un buen nivel de color 24 bits por píxel” considerándolo un buen patrón de referencia para esta variable de calidad. Reforzando esta teoría, en la bibliografía consultada para la elaboración de la *Tabla 3* presentada en el epígrafe anterior, se relaciona idéntica profundidad de bits a cada modo de escaneo.

2.2.3.4. Patrón de Referencia de la Técnica y Nivel de Compresión basado en los atributos Tipo de Documento y Profundidad de Bits

La técnica de compresión utilizada y el nivel de compresión aplicado pueden afectar tanto la velocidad de entrega como la calidad de la imagen resultante.

La compresión con pérdida puede tener un marcado impacto sobre la calidad de la imagen, especialmente si el nivel de compresión es alto. En general, cuanto más enriquecido sea el archivo, tanto más eficiente y sustentable es la compresión. Por ejemplo, el escaneado bitonal de una página a 600 dpi es 4 veces más grande que una versión de 300 dpi, pero con frecuencia sólo dos veces más grande cuando se lo comprime. La compresión sin pérdida se utiliza con mayor frecuencia en el escaneado bitonal de material de texto. La compresión con pérdida típicamente se utiliza con imágenes tonales, y en particular imágenes de tono continuo en donde la simple abreviatura de información no tendrá como resultado un ahorro de archivo apreciable.

⁴⁸ Lara Pacheco, C.G., Castro Thompson, A., Ortiz Ancona, D., et al. Op. Cit.

⁴⁹ Grupo de Expertos de IFLA e ICA. Op. Cit. Apartado 2.1.2.2 Profundidad del bits. Conversión

Cuanto más compleja sea la imagen, tanto menor será el nivel de compresión que se puede alcanzar en un estado sin pérdida o sin pérdida visual. En el caso de las fotografías, la compresión sin pérdida por lo general proporciona un índice de tamaño de archivo de alrededor de 2:1; y en el caso de compresión con pérdida superior a 10 o 20:1, el efecto puede ser evidente.

A continuación se presenta una tabla con los atributos más destacables de las técnicas de compresión más comunes:

	ITU-T.6	JBIG ⁵⁰ /JBIG2 ⁵¹	JPEG	LZW ⁵²	Deflate	Wavelet	ImagePac
Estándar /Patentado	Estándar	Estándar	Estándar	Patentado	Estándar	Estándar o Patentado	Patentado
Sin pérdida /Con pérdida	Sin pérdida	Sin pérdida o Con pérdida	Con pérdida	Sin pérdida	Sin pérdida	Sin pérdida o Con pérdida	Con pérdida
Profundidad de bits	1 bit	Típicamente de 1 bit hasta 6 bits	8 bits o 24 bits	Típicamente de 1 bit a 8 bits	8, 16, and 24 bits		24 bits
Formatos de Archivo y	TIFF ⁵³ , PDF ⁵⁴ ,	TIFF, PDF, fax	JPEG/JFIF ⁵⁵ , TIFF,	ZIP ⁵⁸ , TIFF, GIF ⁵⁹ , PDF,	PNG ⁶¹ , Zip, PDF	JP2 ⁶² , LuraWave ⁶³ ,	

⁵⁰ *Joint Bilevel Image Group Bitmap Graphics*, escala de grises y color. Véase ITU-T T.82 en Glosario de Siglas y Acrónimos.

⁵¹ *Joint Bilevel Image Group Bitmap Graphics*, bitonal. Véase ITU-T T.88 en Glosario de Siglas y Acrónimos.

⁵² *Lempel-Ziv-Welch* es un algoritmo de compresión sin pérdida desarrollado por Terry Welch en 1984 como una versión mejorada del algoritmo LZ78 desarrollado por Abraham Lempel y Jacob Ziv. Ampliamente utilizada, por ejemplo, en archivos TIFF [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/LZW>].

⁵³ *Tagged Image File Format* es un formato de fichero para imágenes [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/TIFF>].

⁵⁴ *Portable Document Format*, es un formato de almacenamiento de documentos, desarrollado por la empresa Adobe Systems. Este formato es de tipo compuesto (imagen vectorial, mapa de bits y texto) [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/PDF>].

⁵⁵ *JPEG File Interchange Format*, es un archivo de mapa de bits comprimido con la tecnología JPEG [Consultado el 25-09-2011 en http://www.profesormolina.com.ar/tecnologia/informatica/glosario_pc.htm]

⁵⁸ *Zone Information Protocol*, es un formato de almacenamiento sin pérdida, muy utilizado para la compresión de datos como documentos, imágenes o programas [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/.zip>].

Aplicaciones	fax		FlashPix ⁵⁶ , SPIFF ⁵⁷ , PDF	Postscript ⁶⁰		MrSID ⁶⁴ , ERMapper ⁶⁵ , DjVu ⁶⁶	
---------------------	-----	--	--	--------------------------	--	---	--

⁵⁹ *Graphic Interchange Format*, es un formato gráfico utilizado ampliamente en la World Wide Web, tanto para imágenes como para animaciones [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/Gif>].

⁶¹ *Portable Network Graphics*, es un formato gráfico basado en un algoritmo de compresión sin pérdida para bitmaps no sujeto a patentes. Este formato fue desarrollado en buena parte para solventar las deficiencias del formato GIF y permite almacenar imágenes con una mayor profundidad de contraste [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/PNG>]

⁶² *JPEG 2000*, es un estándar de compresión y codificación digital de imágenes. Fue creado por el Joint Photographic Experts Group (JPEG), en el año 2000 con la intención de sustituir el formato original creado en 1992. El nuevo formato se basa en la transformada wavelet, en lugar de la transformada de coseno discreta establecida para el estándar original [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/JP2>]

⁶³ LuraWave: familia de software de compresión de imágenes que se basa en las técnicas de compresión wavelet [Consultado el 25-09-2011 en <http://www.luratech.com>].

⁵⁶ Formato de archivo donde se guarda la imagen en más de una resolución de forma que, cuando se envía una solicitud para mostrar la imagen a un navegador Web, sólo se manda la resolución necesaria para su representación en pantalla, lo que por que ahorra ancho de banda y tiempo de descarga [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/FlashPix>].

⁵⁷ *Still Picture Interchange File Format*, formato de archivo de intercambio de dibujos fijos, tiene la designación ISO 10918-3 y ofrece compresión, gestión de color y capacidad de metadatos más versátiles que JPEG/JFIF, pero tiene poco soporte [Consultado el 25-09-2011 en <http://www.library.cornell.edu/preservation/tutorial-spanish/presentation/table7-1.html>]

⁶⁰ Lenguaje de descripción de páginas utilizado en muchas impresoras y, de manera usual, como formato de transporte de archivos gráficos en talleres de impresión profesional [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/PostScript>].

⁶⁴ *Multi-resolution Seamless Image Database*, estándar abierto de compresión de imágenes raster desarrollado por Los Alamos National Laboratory y comercializado por la empresa LizardTech. Permite mostrar archivos digitales de gran tamaño con un tiempo de carga mínimo gracias a la tecnología wavelet [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/MrSID>].

⁶⁵ Software que emplea algoritmos matemáticos para el procesamiento digital de imágenes, teledetección y composición cartográfica, que permite la integración, realce, visualización e interpretación de datos geográficos [Consultado el 25-09-2011 en <http://www.erdas.com/products/ERDASERMapper/ERDASERMapper/Details.aspx>].

⁶⁶ Formato de archivo informático diseñado principalmente para almacenar imágenes escaneadas. Se caracteriza por incorporar avanzadas tecnologías tales como separación de capas de imágenes, carga progresiva, codificación aritmética y compresión sin pérdida para imágenes bitonales (dos

Tabla 8: Atributos de las técnicas de compresión más comunesFuente: Universidad de Cornell⁶⁷

2.2.3.5. Patrón de Referencia para la selección del Tipo de Escáner necesario basado en el atributo Formato

El proceso de digitalización, como ya se ha indicado, deberá llevarse a cabo empleando diferentes tipos de escáner en función de los atributos relativos al formato del soporte de los documentos a procesar:

Tipo Escáner	Atributos relativos al formato del soporte documental			
	Antiguos /Frágiles	Tipo Soporte	Formato	Sueltos /Encuadernados
Bandeja	No	Papel	Máximo DIN-A4	Sueltos
Plano	Sí	Papel	Máximo DIN-A4	Sueltos y Encuadernados
Cenital	Sí	Papel	Máximo DIN-A2	Sueltos y Encuadernados
Tambor	No	Papel Transparencias Negativos	Máximo Tabloid (432x279 mm.)	Sueltos
Grandes Formatos	Sí	Papel	Anchura igual ó < DIN-A0	Sueltos y Encuadernados
Ranura	No	Diapositivas Fotografías Negativos Clichés	-	Sueltos
Microfilm	No	Rollos de película Microfichas	-	Sueltos
Cámara Digital	Sí	Papel Fotografías	-	Sueltos y Encuadernados

Tabla 9: Patrón de referencia para la selección del tipo de escáner necesarioFuente: Elaboración propia basada en el artículo Digitalización de colecciones: texto e imagen⁶⁸.

colores), permitiendo que imágenes de alta calidad se almacenen en un mínimo de espacio [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/DjVu>].

⁶⁷ Departamento de Investigación de la Universidad de Cornell. Op. Cit. 3. Conversión.

⁶⁸ Lara Pacheco, C.G., Castro Thompson, A., Ortiz Ancona, D., et al. Op. Cit.

2.3. ANÁLISIS DE CONTENIDO

2.3.1. Definición y contextualización de la indización como proceso de apoyo a la recuperación de fondos digitalizados

Las organizaciones necesitan sistemas de información que capturen y gestionen información contextual que ayude al entendimiento, uso, acceso y gestión de sus documentos a lo largo del tiempo. Esta información es crítica para afirmar la autenticidad, fiabilidad, integridad, disponibilidad y valor probatorio de los documentos. En su conjunto, esta información se conoce como metadatos para la gestión de documentos.

Los modelos metadatos normalizados existentes difieren de las habituales reglas de descripción y catalogación en la disposición de categorías adicionales para soportar la navegación y gestión de ficheros de datos en un entorno digital.

Podemos establecer tres grandes categorías de metadatos según el tipo de información a la que hacen referencia. Hablamos, por tanto, de metadatos descriptivos, metadatos estructurales y metadatos técnicos. Aunque independientemente del tipo de metadatos podemos establecer una serie de funciones generales:

- ✓ Describir el objeto
- ✓ Permitir la búsqueda y recuperación
- ✓ Prevenir ciertos usos
- ✓ Proporcionar información para su interpretación
- ✓ Obtener información sobre las condiciones de uso y propiedad intelectual (propietario o autor)
- ✓ Aportar información acerca de su evolución
- ✓ Indicar relaciones con otros recursos
- ✓ Controlar su gestión

La siguiente tabla resume los objetivos y elementos que podrían atribuirse a cada una de las categorías o tipos de metadatos:

TIPO	DEFINICIÓN	EJEMPLOS
Metadatos descriptivos	Descripción, representación e identificación de recursos de información en la fase de organización de objetos digitales.	Identificadores únicos, descripciones físicas (medios, dimensiones), datos bibliográficos (título, autor/ creador, idioma, palabras clave), descripción o resumen, etc.
Metadatos estructurales	Facilitan la navegación y presentación de recursos electrónicos, es decir, proporcionan información sobre la estructura, la relación entre objetos y unen los archivos y los textos relacionados.	Páginas, tablas de contenido, capítulos, partes, índice, relación con otros objetos.
Metadatos técnicos	Facilitar la gestión y administración de los recursos. Incluyen datos técnicos sobre la creación, datos sobre la gestión de derechos, acceso y utilización, así como información sobre la conservación.	Datos técnicos tales como tipo y modelo del equipo de captura, formato y tamaño del archivo, resolución, profundidad, espacio de color, compresión, fuente de luz, copyright, datos sobre actualización, migración, etc.

Tabla 10: *Tipos de Metadatos*

Fuente: UNE: ISO 23081

Siguiendo los preceptos de la norma UNE: ISO 23081 sobre uso de metadatos para la gestión documental los metadatos descriptivos se englobarían dentro de los atributos fijados para la entidad “Documento” y sería esta agrupación, en tanto que digitalizamos documentos en cierto modo “descontextualizados” de la organización que los ha generado, sobre los que cabría determinar los pertinentes para su asignación en el entorno de un proyecto de digitalización⁶⁹.

El modelo de metadatos expuesto a continuación pretende servir de guía en el diseño de un esquema final de atributado de las unidades documentales mínimas que se hayan establecido:

- ✓ **Título:** nombre de la unidad documental mínima.

⁶⁹ Véase “III.2 Puntos de análisis de la documentación física para la extracción de requisitos relacionados con la indización” donde se introduce el modelo propuesto de metadatos descriptivos.

- ✓ **Clasificación:** información sobre la clasificación de la entidad de acuerdo con una fuente autorizada.
- ✓ **Resumen:** descripción textual no estructurada
- ✓ **Lugar:** información sobre la localización, el sitio o el espacio asociado al documento, como, por ejemplo dónde está ubicada, dónde se almacena la entidad o dónde se la puede encontrar. El lugar puede ser físico o virtual.
- ✓ **Identificadores externos:** cualesquiera identificadores únicos, actuales o históricos, asignados en un sistema externo al dominio de la gestión de documentos (por ejemplo, número del ISBN⁷⁰, número de Seguridad Social)

No siempre será necesario hacer uso de cada uno de los atributos apuntados en el modelo y, en cambio, en otros casos será necesario profundizar y obtener un mayor nivel de especificidad aumentando los metadatos a asignar a la unidad documental.

Por ejemplo, dentro del tipo “Clasificación” puede resultar provechoso indizar además de la materia el área geográfica a la que se circunscribe el documento o el contratante y/o contratista cuya relación contractual se tipifica en el mismo.

Sin embargo, dentro de esta variabilidad se considera que éste es un modelo factible respaldado por las directrices de la de la norma *UNE: ISO 23081*.

2.3.2. Puntos de análisis de la documentación física para la extracción de requisitos relacionados con la indización

Tal y cómo se ha introducido en el epígrafe anterior, la categoría de *metadatos descriptivos* sería más pertinente para la descripción de los documentos

⁷⁰ *International Standard Book Number*, identificador único para libros, previsto para uso comercial. Fue creado en el Reino Unido en 1966 por las librerías y papelerías británicas W. H. Smith y llamado originalmente *Standard Book Numbering*), abreviado SBN. Fue adoptado como estándar internacional ISO 2108 en 1970 [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/ISBN>].

digitalizados por lo que resulta posible ahondar, de forma más específica en el establecimiento de un modelo de metadatos que permita, a su vez, identificar los aspectos a analizar para la definición de estos requisitos.

Sin embargo, en analogía con las técnicas estadísticas, la clave no estaría únicamente en la correcta aplicación de las *fórmulas* que permitan la recuperación del resultado correcto (términos de indización o metadatos) sino también en la correcta determinación de la *muestra* (nivel o categoría de agrupación documental que pretendemos indizar).

2.3.2.1. Nivel de Asignación Requerido

De este modo, el documentalista deberá *reconocer* la agrupación mínima que el cliente *necesitará* para la correcta identificación y localización de los documentos que convertir a un formato digital. Dicho de otro modo, habrá de determinar la unidad documental sobre la que se efectuará la asignación de metadatos descriptivos:

- ✓ A nivel de **Grupo o agrupación de fondos** (archivo): Todos los documentos de una sociedad, jurisdicción o sector de actividad reunidos en un marco global.

Por ejemplo, documentos de varios servicios médicos o documentos de varias organizaciones no gubernamentales que contribuyen a la construcción de infraestructuras en los países en vías desarrollo.

- ✓ A nivel de **Fondo** (colección): Conjunto de documentos de una organización o individuo.

Por ejemplo, todos los documentos de un servicio médico o todos los documentos de una oficina regional de una empresa de seguros.

- ✓ A nivel de **Serie**: Agrupación de documentos, creados y mantenidos por un organismo o persona, que siguen un orden numérico, alfabético, cronológico

o cualquier otra secuencia identificable, o que son resultado de un mismo proceso de archivado y tienen una función, formato o contenido informativo similar.

Por ejemplo, las historias clínicas, o los expedientes de personal de los empleados de una empresa de seguros.

- ✓ A nivel de **Expediente compuesto o dossier fruto de varios procedimientos**: Secuencia de ítems enlazados física o virtualmente, que son testimonio de las actividades de la organización. Los distintos ítems individuales de un expediente están relacionados entre sí, por ejemplo una carta y su respuesta, y una respuesta a esa, etc., y estas relaciones se mantienen reuniéndolos en el expediente en el orden correcto que forma parte de la evidencia documental. Un expediente puede ser físico o electrónico.

Por ejemplo, una historia clínica.

- ✓ A nivel de **Expediente simple o fruto de un procedimiento**: Secuencia de ítems enlazados física o virtualmente, reflejo de una operación procedimentada de la que se deriva un resultado específico.

Por ejemplo, documentos resultantes de la prestación de un servicio determinado por un médico cuando atiende a un paciente en una consulta, o documentos resultantes de las tareas realizadas por un ayuntamiento para autorizar la apertura de un nuevo restaurante.

- ✓ A nivel de **Ítem o documento simple**: La unidad más pequeña de documentos gestionados como una entidad. Los ítems pueden contener componentes como por ejemplo un correo electrónico con ficheros adjuntos; no obstante, los componentes de un ítem son gestionados como una sola entidad dentro del sistema.

Por ejemplo, un correo electrónico que contiene una remisión de un paciente específico a un nuevo médico, o una propuesta de presupuesto para un nuevo proyecto.

Mediante entrevistas con los responsables y usuarios de la documentación, el examen físico de las unidades documentales y su ubicación, el documentalista encargado del proyecto deberá ser capaz de:

- ✓ Establecer la agrupación o unidad documental mínima
- ✓ Revisar sus criterios de clasificación y ordenación
- ✓ Definir el modelo de metadatos

Cuadro 5: *Recogida de Datos referentes al Contenido del Documento*

Fuente: *Elaboración propia*

2.3.2.2. Determinación de las Fuentes de Información

Igual de relevante que la definición de un esquema de metadatos adecuado para su correcta recuperación mediante la búsqueda documental, es la determinación de las *fuentes* de las que se han de extraer los valores de estos atributos.

Por ejemplo, el documentalista a cargo del análisis de requisitos de la colección de Proyectos de una empresa dedicada a la ejecución de obras hidráulicas considera oportuno, según los datos recabados de las entrevistas con el personal, establecer como metadato el número de proyecto que da sentido al expediente. Sin embargo, un análisis más profundo permite determinar que en la cubierta del expediente no se localiza este dato (o al menos no siempre) pero sí se encuentra ubicado en la cajetilla situada en la parte inferior derecha del tipo documental "Adjudicación Definitiva" que forma parte del expediente. Esto significa, a efectos prácticos, que será necesario localizar este tipo de documento dentro del expediente para poder rescatar este dato y que si éste, por cualquier motivo, no apareciese, no podría caracterizarse este atributo.

Entre las implicaciones que se derivan de este ejemplo, destaca la necesidad de analizar la calidad y detectabilidad de las fuentes de información que han de

respaldar el proceso de asignación de metadatos y las posibles repercusiones que pudiese tener en el proceso productivo y el resultado esperado por parte del cliente.

2.3.2.3. Establecimiento de Herramientas de Control de Vocabulario

El documentalista no sólo deberá establecer el nivel de indización, el conjunto de metadatos previsto para la descripción del documento y las fuentes de dónde extraerlos, sino que además deberá prever el establecimiento de normas relativas al Control del Vocabulario⁷¹.

A continuación se presenta un esquema en el que se pretenden resumir los recursos a tener en cuenta, en base al modelo de metadatos propuesto, para limitar el uso de sintaxis en lenguaje natural.

Metadato ó índice	Herramientas de control del vocabulario
Título	Siempre que sean títulos atribuidos, definición de normas de estilo en su redacción que limiten su variabilidad. Por ejemplo, uso de frases nominales (propias de los titulares periodísticos) para minimizar la incidencia de las desinencias verbales o el establecimiento de un orden predefinido en sus elementos.
Clasificación	Establecimiento o adopción de un cuadro de clasificación funcional, una tabla de clasificación por materias, una lista de encabezamientos autorizados o un tesoro.
Resumen	El resumen no es propiamente un índice o metadato ya que se expresa, por definición en lenguaje natural. No procede la aplicación de herramientas de control del vocabulario.
Lugar	Establecimiento o adopción de un sistema de codificación de la

⁷¹ Vocabulario Controlado: *Conjunto limitado de términos que deben utilizarse para representar las materias de los documentos. Este vocabulario puede ser una lista de encabezamiento de materias, un esquema de clasificación, un tesoro o simplemente una lista "autorizada" de frases o palabras clave (LANCASTER, FREDERICK W. El control del vocabulario en la recuperación de la información. Ed. Universitat de València, 1995. p. 19).*

Metadato ó índice	Herramientas de control del vocabulario
	signatura topográfica o directorio de ficheros.
Identificadores externos	Este metadato ha de ser, por definición, controlado (por ejemplo, número del ISBN, número de Seguridad Social). Sólo cabe establecer normas sobre el uso de signos de puntuación como guiones (-), puntos (.), etc.

Tabla 11: Modelo de Metadatos propuesto

Fuente: Elaboración propia

Si el responsable de la documentación no dispusiese de instrumentos ya elaborados para el control a los metadatos del tipo “Clasificación”, el documentalista debería analizar, en clave de coste-beneficio, la idoneidad de construir una herramienta ad hoc o adaptar alguna de las existentes tras un proceso de selección.

En este sentido, el *Tesaurus de la UNESCO*⁷² sería una fuente adecuada para la selección de vocabularios controlados para la asignación de materias ya que permite la exportación de micro-tesauros clasificados por áreas temáticas.

⁷² <http://databases.unesco.org/thessp> [Consultado 05-04-2011].

UNESCO es la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura es un organismo especializado de las Naciones Unidas. Se fundó el 16 de noviembre de 1945 con el objetivo de contribuir a la paz y a la seguridad en el mundo mediante la educación, la ciencia, la cultura y las comunicaciones. La constitución firmada ese día entró en vigor el 4 de noviembre de 1946. Se dedica a orientar a los pueblos en una gestión más eficaz de su propio desarrollo, a través de los recursos naturales y los valores culturales, y con la finalidad de modernizar y hacer progresar a las naciones del mundo, sin que por ello se pierdan la identidad y la diversidad cultural. La Unesco tiene vocación pacifista, y entre varias cosas se orienta muy particularmente a apoyar la alfabetización. En la educación, este organismo asigna prioridad al logro de la educación elemental adaptada a las necesidades actuales. Colabora con la formación de docentes, planificadores familiares y vivienda, administradores educacionales y alienta la construcción de escuelas y la dotación de equipo necesario para su funcionamiento. Las actividades culturales buscan la salvaguarda del patrimonio cultural mediante el estímulo de la creación y la creatividad y la preservación de las entidades culturales y tradiciones orales, así como la promoción de los libros y de la lectura. En materia de información, la Unesco promueve la libre circulación de ideas por medios audiovisuales, fomenta la libertad de prensa y la independencia, el pluralismo y la diversidad de los

Complementariamente, la Web del *Instituto de Estudios Documentales sobre Ciencia y Tecnología (antes CINDOC)* del Consejo Superior de Investigaciones Científicas⁷³, ofrece la posibilidad de consultar en línea sus tesauros referentes a sus áreas de investigación, además de un tesoro de topónimos.

No obstante, cada fondo requerirá un nivel de especificidad por lo que puede resultar o no beneficioso para el proyecto ser exhaustivo en la elaboración y uso de las herramientas de control del vocabulario. El punto de equilibrio lo determinará, en cualquier caso, la intuición, experiencia y profesionalidad del documentalista.

2.3.3. Automatismos

Definidos los atributos necesarios para la correcta caracterización del documento escaneado (recuérdese que entendemos por documento tanto una unidad documental simple como la compuesta) conviene establecer una serie de mecanismos automáticos que permitan optimizar el proceso de asignación de los mismos. Es decir, hay que analizar la posibilidad de ahorrar costes a través de automatismos aplicados a la indización.

A continuación se presentan las cuestiones que se han considerado clave para el desarrollo de estos mecanismos y que, por tanto, el documentalista debería plantearse durante esta fase de definición del proyecto:

- ✓ ¿Existen metadatos comunes en un nivel superior de agrupación documental al designado para la indización?
- ✓ ¿El cliente dispone de algún fichero o listado manual que podamos aprovechar para la indización?

medios de información, vía el Programa Internacional para la Promoción de la Comunicación [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/Unesco>].

⁷³ http://thes.cindoc.csic.es/index_esp.php [Consultado 05-04-2011].

- ✓ ¿Existe la posibilidad de aplicar *eficazmente* técnicas de reconocimiento de caracteres?

- ✓ ¿Qué ayudas podríamos implementar en la herramienta informática diseñada para la indización dentro de la cadena de producción?

2.3.3.1. Procesamiento por Lotes

Entre las competencias básicas de un documentalista *debe* contarse la capacidad de detectar las clasificaciones supra-documentales y criterios de ordenación que dan sentido a una colección de Archivo. Precisamente, en el ámbito de la indización, esta habilidad puede y debe ponerse al servicio de la optimización del proceso de indización del modo que se expresa a continuación.

Dado un conjunto de documentos, se asume que los campos de indización o metadatos asociados a nivel del documento simple, como por ejemplo la “fecha del documento”, coexistirán con aquellos que “herede” de las agrupaciones documentales superiores (expediente, unidad de instalación, colección, etc.). En este caso sería recomendable agrupar en una misma unidad lógica todos los documentos que compartan los datos de indización de los niveles superiores al expediente para que sea suficiente con asignarlos colectivamente. Veamos un ejemplo,

En el archivo de una fundación la documentación se organiza en archivadores que contienen los expedientes de los aspirantes a becas de formación. Cada archivador recoge la totalidad de aspirantes en el ámbito de un proyecto concreto encontrándose la documentación de cada uno de ellos separada mediante cartulinas.

Los campos a indexar para cada aspirante serán los siguientes:

- ✓ *Número de Proyecto*
- ✓ *Nombre del Proyecto*
- ✓ *Convocatoria*
- ✓ *Solicitante*
- ✓ *Tipo de Documento (impreso de solicitud, acreditación de méritos, adjudicación, denegación, memoria, justificantes de gastos)*

Analizando la documentación, se detecta que la mayoría de atributos serán comunes a todos los aspirantes ya que se encuentran rotulados en la cubierta del archivador, siendo solamente el "Solicitante" asignable a nivel de expediente y "Tipo de Documento" a nivel de documento simple.

Cuadro 6: Ejemplo de asignación de metadatos en varios niveles de agrupación documental

Fuente: Elaboración propia

Intuitivamente podemos asumir que resultaría interesante operativamente poder asignar en un único gesto los datos comunes a todos los documentos sin tener que "picarlos" para cada página. Esto es posible mediante la colocación, de forma previa al escaneado, de **carátulas con códigos de barras con los datos comunes de indexación** precediendo los grupos documentales. Utilizando este automatismo, sólo sería necesario introducir los datos comunes una única vez por agrupación documental puesto que el software del escáner asociará los atributos indicados en la carátula a cada una de las páginas digitalizadas hasta la detección el código de barras siguiente.

La gran mayoría de escáneres de alta producción disponen de lector de código de barras incorporado y del software necesario para interpretarlo. Obviamente, se requiere parte de desarrollo informático para “generar” estas carátulas pero ha de ser el documentalista quien le indique los criterios de indización que determinarán estas agrupaciones documentales. El uso de carátulas redundará tanto en la optimización del tiempo de indización del total de páginas que integran cada agrupación como en la mejora de la eficiencia del escaneado, al posibilitar la formación de lotes de producción con documentación procedente de distintas unidades de instalación sin perder la trazabilidad⁷⁴.

2.3.3.2. Aprovechamiento de Ficheros Cliente

Pero aún podemos ir un paso más allá en la automatización del proceso de digitalización. En el caso de que la organización propietaria de la documentación dispusiera de “**registros informatizados**” de los fondos de archivo que contuviesen datos coincidentes con los atributos de indización (siguiendo el ejemplo, si la Fundación dispusiese de una relación de los solicitantes de cada proyecto en cada una de las convocatorias) éstos podrían tratarse informáticamente para automatizar la generación de las carátulas. De esta forma, para cada archivador se indicaría el código de proyecto, recuperando el resto de datos del registro del propietario de la documentación y generando con ellos una carátula que precediese cada expediente (puesto que el nombre del solicitante también nos vendría dado). Solamente restaría, según el ejemplo, indizar el tipo de documento en concreto para cada una de las páginas.

⁷⁴ Asociación Española de Documentación Digital. *Curso experto en gestión de documentos digitales en empresas e instituciones. Unidad 1. Gestión de un proyecto de digitalización de documentos para una entidad*. AEDOC Digital, 2007 (Pág. 18)

Además de la optimización de los tiempos, la disposición del fichero del propietario de la documentación permitirá realizar un **control de calidad pasivo**⁷⁵ en tanto que se minimizarán los errores cometidos al teclear los datos manualmente y, adicionalmente, resultará posible detectar, sin gran esfuerzo, la falta de expedientes previstos dentro de una unidad de instalación (siguiendo el ejemplo, veríamos que nos sobraría una la carátula del solicitante cuyo expediente no está en el archivador correspondiente) o el mal archivado de alguno de ellos.

Ponderando los beneficios que pueden obtenerse de la aplicación de estas utilidades, resulta obvio que el documentalista debe tener en cuenta estos aspectos en el análisis de requisitos.

2.3.3.3. Aplicación de Técnicas de Reconocimiento Óptico

En ocasiones, sobretodo en colecciones cuyo núcleo temático sea de índole técnico-científica (que presupone el uso de una terminología específica y precisa) no excesivamente disperso en cuanto a materias con documentos eminentemente textuales de una extensión corta-media, puede ser suficiente con la identificación de determinadas palabras clave en el global del texto.

Bajo estas condiciones, el documentalista podría sugerir la ejecución de una prueba de **Reconocimiento Óptico** (OCR⁷⁶ ó ICR⁷⁷) para verificar si resulta posible una

⁷⁵ UNE-ISO 15489. Op. Cit. "Se considera adecuado introducir en este punto el concepto de registro como medio para "formalizar la incorporación de un documento en un sistema de gestión de documentos"

⁷⁶ Reconocimiento Óptico de Caracteres, tecnología que proporciona a los sistemas de reproducción por escáner y sistemas de imágenes la habilidad de convertir imágenes de caracteres en letra de máquina, en caracteres capaces de ser interpretados o reconocidos por una computadora. Así, las imágenes de caracteres en letra de máquina son extraídas de un mapa de bits de la imagen reproducida por el escáner [Consultado el 25-09-2011 en <http://espanol.scantron.com/sistemas/comparacion.htm>].

conversión aceptable de la imagen en texto. De esta forma, la comparación del texto completo del documento con el patrón de palabras clave elaborado previamente podría automatizarse, optimizando los tiempos de indización.

⁷⁷ *Reconocimiento de Caracteres Inteligente*, proporciona a los sistemas de reproducción por escáner y sistemas de imágenes la habilidad de convertir caracteres en letra manuscrita (no cursiva) en caracteres capaces de ser interpretados o reconocidos por una computadora. Así, las imágenes de caracteres en letra manuscrita son extraídas de un mapa de bits de la imagen reproducida por el escáner [Consultado el 25-09-2011 en <http://espanol.scantron.com/sistemas/comparacion.htm>].

2.4. REQUISITOS RELACIONADOS CON EL PRODUCTO DIGITALIZADO ENTREGADO

2.4.1. Formato del Fichero de Imagen

El documentalista deberá establecer la necesidad de generar una o varias versiones de la imagen en función de los requisitos del propietario de la documentación. A continuación se presenta un cuadro que puede servir de guía tanto a nivel interno para detectar necesidades no expresadas por parte del cliente, como a nivel de traducir sus requisitos explícitos en información técnica para la configuración del proyecto:

Imagen <i>master</i>	Imagen de acceso	Imagen miniatura
Representa lo más fielmente posible la información contenida en la fuente original.		
Sin compresión.		
Sin edición.	Se utiliza en lugar de la imagen <i>master</i> para acceso general.	Imagen muy pequeña.
Uso como fuente de larga duración para hacer copias de ella.	Se ajusta al área de visualización de un monitor estándar.	Diseñada para desplegarse rápidamente en línea; permite al usuario decidir si desea o no desplegar o descargar la imagen (<i>imagen de acceso</i>)
Uso como sustituto del original.	Calidad aceptable para la visualización en pantalla (o incluso impresión).	Comúnmente se almacena en formato GIF o JPG.
Alta calidad.	Opcionalmente comprimida para agilizar el acceso.	No siempre es conveniente para las imágenes primordialmente textuales, partituras, etc.
Generalmente es de gran tamaño, pues conserva íntegra la información del escaneo.	Usualmente almacenada en formato estándar.	
Sirve para crear reproducciones impresas de alta calidad.		
Comúnmente es formato TIFF (texto bitonal), JPG (imágenes color) o PDF (texto o imagen/texto).		

Tabla 12: Formatos de los Archivos de Imagen según el modo de acceso previsto

Fuente: *Digital Imaging Best Practices*⁷⁸]

⁷⁸ Western States Digital Standards Group. *Western States Digital Imaging Best Practices vers. 1.0*. University of Denver and the Colorado Digitization Program, 2003. Citado en DLM-Forum on electronic records (Brussels, 18 to 20 December 1996). *Guidelines on best practices for using electronic information: How to deal with machine-readable data and electronic documents*. Office for Official Publications of the European Communities, 1997 (Pág. 57)

Adicionalmente se muestra un cuadro con el resumen de los mejores formatos a emplear dependiendo del caso:

Formato	Tipo de documento	Características	Extensión
JPG	Imagen (Incluye texto-imagen)	Ampliamente utilizado para fotografías e imágenes distribuidas por Internet Sin compresión produce imágenes de buena calidad y tamaño reducido Amplia capacidad de reproducción y despliegue de colores.	.jpg
PDF	Texto e imagen	Reproducción casi exacta de la fuente original Amplia difusión Estándar en Internet para distribución de textos.	.pdf
TIFF	Texto e imagen	Recomendado para texto simple Tamaño reducido Muy utilizado para archivos <i>master</i> Sin propietario	.tif
GIF	Imagen	Ampliamente utilizado en Internet Imágenes de mapa de bits Tamaño reducido.	.gif
BMP ⁷⁹	Imagen	Imágenes de mapa de bits Estándar para Windows y OS/2.	.bmp

Tabla 13: Características de los formatos de los Archivos de Imagen más comunes

Fuente: *Guidelines on best practices for using electronic information: How to deal with machine-readable data and electronic documents*⁸⁰

⁷⁹ *BitMaP*, es el formato propio del programa Microsoft Paint, que viene con el sistema operativo Windows. Puede guardar imágenes de 24 bits, 8 bits y menos. Puede darse a estos archivos una compresión sin pérdida de calidad [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/BMP>]

⁸⁰ DLM-Forum on electronic records. Op. Cit. (Pág. 38)

2.4.2. Formato de la Base de Datos con los Atributos de Indización

El documentalista debe conocer como mínimo los formatos de archivo de bases de datos manejados por su Departamento de Informática (en los casos en que la herramienta de visualización sea de desarrollo propio de la empresa digitalizadora) o determinar el que fuere necesario según los requisitos expresados por el propietario de la documentación (o de la herramienta sobre la que se vayan a digitalizar las imágenes).

En cualquier caso, no es cometido del documentalista ser “experto” en los distintos formatos y arquitecturas de bases de datos pero sí será necesario que haga constar la no recogida de este requisito antes de valorar un proyecto, para que, si es necesario, el personal informático de la empresa pueda contactar con su homólogo en la organización propietaria de la documentación y determinarlo.

CAPÍTULO 3. CONCLUSIONES

Como catarsis del trabajo presentado como Proyecto Final de Carrera puede extraerse la conclusión general de que ante cualquier proyecto de digitalización, resulta necesario analizar, de forma previa, las características diferenciales de la documentación en su formato original.

Asimismo, se concluye que, a falta de una metodología establecida a nivel profesional, resulta necesario consultar diversas fuentes a fin de recoger y establecer la forma de análisis de los aspectos documentales que deben ser observados.

Y, a la vista de la bibliografía consultada y la experiencia personal se puede afirmar que tanto el establecimiento como el análisis de estos requisitos documentales deben sistematizarse de forma que se apliquen siempre y del mismo modo ante cualquier proyecto de digitalización.

Pero la conclusión más importante de todas es aquella que dicta que la Documentación es una disciplina en continuo desarrollo que abarca dentro de su campo soluciones a problemas del mundo real como la acumulación de documentos en empresas e instituciones.

Por este motivo, el documentalista no puede permitirse una actitud pasiva ante los acontecimientos sino que, a través del continuo aprendizaje y reciclaje profesional puede ser parte activa en su resolución.

BIBLIOGRAFÍA Y NORMATIVA TÉCNICA

Normativa Técnica

- ✓ ISO 23081 Uso de metadatos para la gestión documental.
- ✓ UNE 15489 Información y Documentación. Gestión de Documentos.
- ✓ UNE 54110 de Requisitos de Almacenamiento de documentos para materiales de archivo y bibliotecas.

Bibliografía

- ✓ Abadal Falgueras, Ernest. *Gestión de proyectos en información y documentación*. Ediciones TREA, S.L., 2004.
- ✓ Asociación Española de Documentación Digital. *Curso experto en gestión de documentos digitales en empresas e instituciones. Unidad 1. Gestión de un proyecto de digitalización de documentos para una entidad*. AEDOC Digital, 2007.
- ✓ Biblioteca de la Universidad de Cornell. Departamento de Investigación. *Llevando la Teoría a la Práctica. Tutorial de Digitalización de Imágenes*. Disponible en [\[http://www.library.cornell.edu/preservation/tutorial-spanish/contents.html\]](http://www.library.cornell.edu/preservation/tutorial-spanish/contents.html)
- ✓ CNEDA. *Modelo Conceptual de Descripción Archivística y Requisitos de Datos Básicos de las Descripciones de Documentos de Archivo, Agentes y Funciones. Parte 1: Tipos de Entidad*. Borrador final de la CNEDA (5-12-2008). Disponible en [\[http://www.mcu.es/archivos/docs/NEDATiposEntidad_20081215.pdf\]](http://www.mcu.es/archivos/docs/NEDATiposEntidad_20081215.pdf)
- ✓ Cobacho Gómez, L. Reseña de Moralejo Álvarez, Remedios [et al.]; Coordinador, Francisco Alía Miranda. *Del texto al hipertexto: las bibliotecas universitarias ante el reto de la digitalización*. Anales de Documentación.

Servicio de Publicaciones, Universidad de Murcia. pp.291-294. Disponible en [\[http://www.um.es/ojs/index.php/analesdoc/article/view/3471/3371\]](http://www.um.es/ojs/index.php/analesdoc/article/view/3471/3371)

- ✓ Cubells y M^a José; París, Lidón. *Como implantar un Sistema de Gestión Documental y un sistema archivístico en una institución*. Universitat Jaume I, curso 2006/2007.
- ✓ DLM-Forum on electronic records (Brussels, 18 to 20 December 1996). *Guidelines on best practices for using electronic information: How to deal with machine-readable data and electronic documents*. Office for Official Publications of the European Communities, 1997
- ✓ García Caballero, R., & Méndez, E. *Nuevas Tecnologías y Servicios de información gráfica: reflexiones para el profesional de la información ante la digitalización de imágenes fijas*. IV Jornadas de Documentación Automatizada, Valencia (Spain). 19-21 October 1998. Fesabid⁸¹. Disponible en [\[http://hdl.handle.net/10760/12696\]](http://hdl.handle.net/10760/12696)
- ✓ Gimenez Chornet, V., & Campuzano Juan, F. *Recomendaciones técnicas para la captura y digitalización de fondos de archivos*, 2005. Disponible en [\[dglab.cult.gva.es/ArxiuRegne/documents/Recdigitalizacion.pdf\]](http://dglab.cult.gva.es/ArxiuRegne/documents/Recdigitalizacion.pdf)
- ✓ Gonzáles-Cam, C. *La Importancia de la Digitalización de Archivos para la Biblioteca*. Convención Nacional de Centros Binacionales, Trujillo (Perú), 11-12 Octubre 2007. Disponible en E-LIS [\[http://hdl.handle.net/10760/10647\]](http://hdl.handle.net/10760/10647)
- ✓ Grupo de Expertos de IFLA e ICA. *Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos ((Grupo de expertos de IFLA e ICA*

⁸¹ Federación Española de Sociedades de Archivística, Biblioteconomía, Documentación y Museística. Es una entidad de derecho privado, sin ánimo de lucro, fundada en 1988 con el fin de reunir a las asociaciones profesionales más relevantes del sector [Consultado el 25-09-2011 en <http://www.fesabid.org>].

(*International Council on Archives*) invitados a elaborarlas por la UNESCO).

Disponible en [http://www.mcu.es/archivos/docs/pautas_digitalizacion.pdf]

- ✓ Hasan Kitapci, Barry W. Boehm. *Using a Hybrid Method for Formalizing Informal Stakeholder Requirements Inputs*. Presentado en Fourth International Workshop on Comparative Evaluation in Requirements Engineering, 2006 (CERE '06). Disponible en [http://www.di.unipi.it/CERE06/P5%20CERE%2006%20Kitapci_Hasan.pdf]
- ✓ Hernando-De-Larramendi, L., Domínguez-Muriel, J., Viedma-Peláez, A., et al. *Datos y metadatos: La normalización dinámica de los elementos y de los procesos constituyentes de una biblioteca virtual*. XI Jornadas Españolas de Documentación: 20, 21 y 22 de mayo de 2009, Auditorio Palacio de Congresos de Zaragoza, 2009, Zaragoza, 20, 21 y 22 de mayo de 2009. FESABID. pp.109-116. Disponible en [http://www.digibis.com/img_articulos/libro_actas_fesabid_2009_digibis.pdf]
- ✓ Lara Pacheco, C.G., Castro Thompson, A., Ortiz Ancona, D., et al. *Digitalización de colecciones: texto e imagen*. Universidad Nacional Autónoma de México, 2008. Disponible en [<http://www.digitalizacion.unam.mx/index.html>]
- ✓ Marín Fernández, Josefa. *Estadística aplicada a las Ciencias de la Documentación*. DM, 2008.
- ✓ Merchán Arribas, Montaña. *El documento, expediente y archivos electrónicos. Buenas Prácticas*. TECNIMAP⁸² Zaragoza 2010.

⁸² Jornadas sobre Tecnologías de la Información para la Modernización de las Administraciones Públicas organizadas por el Consejo Superior de Administración Electrónica, de acuerdo con lo previsto en el Real Decreto 589/2005, de 20 de mayo por el que se reestructuran los órganos colegiados responsables de la Administración electrónica [Consultado el 25-09-2011 en <http://www.tecnimap.es>].

- ✓ Mochón Bezares, Gonzalo. Sorlí Rojo, Ángela. *Tesaurus de Biblioteconomía y Documentación. Instituto de Estudios Documentales sobre Ciencia y Tecnología (antes CINDOC)*. Consejo Superior de Investigaciones Científicas (CSIC). Disponible en [http://thes.cindoc.csic.es/index_BIBLIO_esp.php]
- ✓ Moya Anegón, Félix, López Gijón, Javier, García Caro, Concepción. *Técnicas cuantitativas aplicadas a la Biblioteconomía y Documentación*. Síntesis, 1999.
- ✓ Navarro del Peso, Emilio. *Manual de Outsourcing Informático: Análisis y Contratación*. Ediciones Díaz de Santos, S.A. 2ª ed., 2003.
- ✓ Reilly, James M., Franziska Frey. *Recommendations for the evaluation of Digital Images produced from Photographic, Micrographic and Various Paper*. Rochester: Image Permanence Institute. Disponible en [<http://lcweb2.loc.gov/ammem/lpireprt>]
- ✓ Rodríguez, D. *Proyecto de digitalización de prensa en la Biblioteca Nacional*. Revista d'arxius i biblioteques: Compactus. Generalitat Valenciana. Conselleria de Cultura i Educació. Direcció General del Llibre, Arxius i Biblioteques. pp.35-38, 2003. Disponible en E-LIS [<http://hdl.handle.net/10760/6623>]
- ✓ Serra Serra, J. *El uso de la certificación electrónica en los archivos, 2010*. In *El documento electrónico en el archivo: conservación y certificación*. Presentado en Cáceres, 5 Noviembre 2010. Disponible en E-LIS [<http://hdl.handle.net/10760/15320>]
- ✓ Torralba Martínez, José María. *Introducción al Presupuesto de Proyectos Informáticos y Telemáticos: Costes para el proveedor*. Departamento de Organización de Empresas, Economía Financiera y Contabilidad. Facultad de Informática. Universidad Politécnica de Valencia, Editorial de la UPV, 2003.

- ✓ Torralba Martínez, José María (et. al.). El Proceso presupuestario en proyectos de construcción. Departamento de Organización de Empresas, Economía Financiera y Contabilidad. Escuela Superior de Ingeniería del Diseño. Universidad Politécnica de Valencia. Editorial de la UPV.

- ✓ Vives-Gràcia, J., Álvarez-García, J., & Portús-Vinyeta, M.D. *Del zoo victoriano al ecosistema electrónico: la digitalización del patrimonio en una sociedad digital*,. Revista de patrimonio cultural de España. Instituto del Patrimonio Cultural de España. pp.129-145, 2010. Disponible en [\[http://www.calameo.com/read/000075335d0a70e570294\]](http://www.calameo.com/read/000075335d0a70e570294)

ANEXO I. ANÁLISIS DE LOS REQUISITOS EXPRESADOS EN EL ESQUEMA NACIONAL DE INTEROPERABILIDAD

El Real Decreto 4/2010, de 8 de enero, por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica preveía la publicación de una serie de Normas Técnicas que desarrollasen los requerimientos técnicos planteados por la interoperabilidad electrónica entre las administraciones.

Recientemente se publicaron las resoluciones correspondientes a la siguiente normativa:

- ✓ Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Digitalización de Documentos.
- ✓ Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Documento Electrónico.
- ✓ Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Expediente Electrónico.
- ✓ Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Política de Firma Electrónica y de certificados de la Administración.
- ✓ Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Procedimientos de copiado auténtico y conversión entre documentos electrónicos.

- ✓ Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de requisitos de conexión a la red de comunicaciones de las Administraciones Públicas españolas.
- ✓ Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Modelo de Datos para el Intercambio de asientos entre las entidades registrales.

A continuación se analiza el contenido de las Normas Técnicas (NT) relacionadas con el objeto y propósito del presente Proyecto Final de Carrera:

La Norma Técnica de Copiado Auténtico y Conversión, en su punto 6[2], establece que los documentos generados resultantes de la digitalización deberán establecer el valor “Copia electrónica auténtica de documento papel” en el metadato obligatorio “**Estado de elaboración**”.

Esta NT, en su punto 5[6] establece la posibilidad de destrucción de los documentos originales tras la generación de la copia auténtica, que también tendrá consideración de original. Sin embargo, establece que para ser considerado como copia auténtica debe ser expedida por un Órgano Público según establece la Ley 30/1992 de Régimen Jurídico de las Administraciones Públicas, por lo que no podrían delegar el proceso completamente en personal externo.

En el punto 4[4] indica que la relación entre el documento copia auténtica y el documento origen vendrá establecido a través del metadato “**Identificador del documento origen**” que tomará el valor del documento origen. En la NT de Documento Electrónico se establece que este solo se rellenará en el caso de tener su origen en documento electrónico.

La NT de Digitalización es quizá el más avanzado en su redacción, ya que simplemente remite a otras Normativas Técnicas. El proceso de digitalización se establece formado por los siguientes pasos:

- ✓ Conversión digital
- ✓ Mejora de la imagen, si bien limita la mejora a procesos automáticos. Es comprensible que se incluyen las mejoras “manuales” como parte de la corrección de la conversión digital.
- ✓ Asignación de metadatos al documento (no a la página ni imagen).
- ✓ Si procede, firma digital. Evidentemente **siempre procede** por ser un requisito imprescindible para la consideración de **copia auténtica**.

La NT establece (punto 5[2]) la posibilidad de existencia de un **plan de calidad** que permita garantizar la calidad de las imágenes resultantes. Es decir, un mecanismo que obligue a realizar comprobaciones periódicas en las aplicaciones.

El resto de puntos vienen a repetir los requisitos de los documentos certificados digitalmente: resolución mínima de 200dpi, formatos abiertos, respeto a proporciones (en ningún caso ninguna norma obliga a tamaños mínimos determinados) y no adición de elementos inexistentes en el soporte original.

El primer punto a destacar de la NT de Documentos Electrónicos reincide en lo ya dicho anteriormente. Los documentos administrativos electrónicos y aquellos asociados a un expediente tendrán siempre al menos una firma electrónica; los inicialmente electrónicos por ser un requerido durante su generación, las copias auténticas por ser imprescindible para ser consideradas “auténticas”.

Sí que establece un conjunto **mínimo** de metadatos a la vez que permite ampliar aquellos por las necesidades específicas de gestión en cada caso:

✓ **Versión Norma Técnica de Interoperabilidad**

<Definición> Identificador normalizado de la versión de la Norma Técnica de Interoperabilidad de Documento electrónico conforme a la cual se estructura el documento electrónico.

<Cumplimentado> Automático, presumiblemente siempre el mismo valor durante todo el proceso.

<Valor a cumplimentar> URI⁸³ con el enlace al XSD⁸⁴ o DTD⁸⁵ con la definición oficial de los atributos.

✓ **Identificador**

<Definición> Identificador normalizado del documento.

<Cumplimentado> Automático.

<Valor a cumplimentar> Secuencia de 2+9+4+30 y 3 separadores (total de 48) caracteres alfanuméricos que identifica el país, órgano (puede ser tanto el de origen como el responsable del expediente), año de captura y un número identificador. Este

⁸³ *Uniform Resource Identifier*, es una cadena de caracteres corta que identifica inequívocamente un recurso (servicio, página, documento, dirección de correo electrónico, enciclopedia, etc.). Normalmente estos recursos son accesibles en una red o sistema. Los URI pueden ser localizadores uniformes de recursos, Uniform Resource Name, o ambos [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Uniform_Resource_Identifier].

⁸⁴ *XML Schema Definition*, es un lenguaje de esquema utilizado para describir la estructura y las restricciones de los contenidos de los documentos XML de una forma muy precisa, más allá de las normas sintácticas impuestas por el propio lenguaje XML. Se consigue así una percepción del tipo de documento con un nivel alto de abstracción. Fue desarrollado por el *World Wide Web Consortium* (W3C) y alcanzó el nivel de recomendación en mayo de 2001 [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/XSD>].

⁸⁵ *Document Type Definition*, es una descripción de estructura y sintaxis de un documento XML o SGML. Su función básica es la descripción del formato de datos, para usar un formato común y mantener la consistencia entre todos los documentos que utilicen la misma DTD. De esta forma, dichos documentos, pueden ser validados, conocen la estructura de los elementos y la descripción de los datos que trae consigo cada documento, y pueden además compartir la misma descripción y forma de validación dentro de un grupo de trabajo que usa el mismo tipo de información [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/DTD>].

número puede ser secuencial o aleatorio pero siempre único. En caso de procesarse documentación de forma simultánea habría que tener cuidado de no duplicarlo pero es lo bastante largo (cerca de 30 caracteres) como para que no surjan problemas. Nada obliga a que el número asignado coincida con cualquier posible identificador que tenga el original en papel, pero ese posible código, de existir, puede entenderse como un metadato adicional a indizar.

✓ **Órgano**

<Definición> Identificador normalizado de la organización responsable del documento o de la captura del mismo.

<Cumplimentado> Automático si lo entendemos siempre únicamente como “responsable de la captura”.

<Valor a cumplimentar> Código del órgano. Puede repetirse tantas veces como sea necesario, p.ej. una para el órgano de captura y otra para el órgano emisor.

✓ **Fecha de Captura**

<Definición> Fecha de alta en el sistema de gestión documental.

<Cumplimentado> Automático

<Valor a cumplimentar> Fecha según ISO 8601

✓ **Origen**

<Definición> Valor binario que diferencia entre “Administración” y “Ciudadano”

<Cumplimentado> *Manual*

<Valor a cumplimentar> Requerirá indizar cada documento, uno a uno, para definir si el origen del documento es el Ciudadano (Valor 0) o la Administración Pública (Valor 1)

✓ **Estado de elaboración**

<Definición> Indica la forma de generación del documento.

<Cumplimentado> Automático.

<Valor a cumplimentar> Siempre el valor textual “Copia auténtica de documento papel (Ley 11/2007 Art.30.2 y 30.3)”.

✓ **Nombre de formato**

<Definición> Formato electrónico que indica el tipo de fichero del documento.

<Cumplimentado>Automático

<Valor a cumplimentar>Valor según el Catálogo de Estándares (aún no publicado).

✓ **Tipo Documental**

<Definición> Tipo del documento

<Cumplimentado> Manual

<Valor a cumplimentar> Según la lista de códigos definida.

✓ **Tipo de firma**

<Definición> Formato de firma certificada

<Cumplimentado> Automático

<Valor a cumplimentar> No definido aún.

La NT de Expedientes Electrónicos establece a su vez un conjunto mínimo de metadatos que serían requeridos. Omitiendo los que son idénticos o muy similares a los del documento tenemos:

✓ **FechaAperturaExpediente**

<Definición> Momento de creación del expediente

<Cumplimentado> Manual/Automático

<Valor a cumplimentar> Fecha de apertura del expediente según ISO.

✓ **Estado**

<Definición> Indica el estado en que se encuentra el expediente en el momento de la captura

<Cumplimentado>Automático

<Valor a cumplimentar> “Cerrado” en caso de procesos de digitalización externalizados.

✓ **Interesado**

<Definición> Identificador de personas y órganos implicados

<Cumplimentado>Manual

<Valor a cumplimentar> DNI o similar de las personas o código del órgano según listado. Es posible repetir según se requiera.

También hay que generar un “Documento de índice” que referencia a todos los documentos que forman parte del expediente.

La NT de Firma Electrónica y de Certificados y la NT de su política se limita a englobar el marco teórico de funcionamiento de la firma electrónica y la política empleada para su generación y comparación. Es previsible que siguiendo el marco normativo español y europeo, asuma como buenas las firmas validadas por la Autoridad Pública de Certificación de cada país (FNMT en el caso de España) y cualquiera de los prestadores de servicio existentes: ACE⁸⁶, ACCV⁸⁷, IPSCA⁸⁸,

⁸⁶ *Agencia de Certificación Electrónica*, compañía del Grupo Telefónica que centra su actividad en proporcionar servicios de confianza que garanticen la seguridad de las transacciones electrónicas en el ámbito del comercio electrónico: autenticación (identificación) de las personas, dispositivos y organizaciones involucradas; integridad de la información intercambiada en la transacción; confidencialidad (secreto) de la información intercambiada; no repudio de la intervención de los participantes en una operación [Consultado el 25-09-2011 en <http://www.ace.es>].

⁸⁷ *Agencia de Tecnología y Certificación Electrónica*, proporciona a los ciudadanos, las empresas y las Administraciones Públicas los mecanismos de identificación telemática segura en los trámites administrativos a través de Internet: los certificados digitales y las tecnologías asociadas. Los certificados emitidos y el resto de servicios que la ACCV presta se ajustan a lo establecido en la Ley 59/2003, de 19 de diciembre, de firma electrónica, por lo que la ACCV goza de amplio reconocimiento en todas las Administraciones Públicas. Además, con los certificados digitales reconocidos expedidos por la ACCV se puede generar firma electrónica reconocida, que es equivalente a la manuscrita [Consultado el 25-09-2011 en <http://www.accv.es>].

⁸⁸ Certification Authority CA, compañía cuyo certificado raíz (ipsCA Global CA Root) está incorporado en los productos Microsoft desde 2009. Los certificados ofrecidos por IPSCA son los siguientes: certificados de servidor, certificados personales, certificados para firma de código y certificados para IPSEC-VPN [Consultado el 25-09-2011 en <http://web.ipsca.com/es>].

FESTE⁸⁹ o CAMEFIRMA⁹⁰ (Consortio formado por los principales bancos y cajas), entre muchas otras. Cualquiera de las entidades autorizadas por la FNMT⁹¹ tiene la misma validez legal en toda la Unión Europea. Es irrelevante la autoridad que facilite el certificado siempre que sea validada por la FNMT.

⁸⁹ Autoridad de certificación de la *Fundación para el Estudio de la Seguridad de las Telecomunicación*, entidad en la que participan, entre otros, los notarios y corredores de comercio electrónico [Consultado el 25-09-2011 en <http://www.invenia.es/feste>].

⁹⁰ Servicio de certificación digital de las Cámaras de Comercio, Industria y Navegación de España [Consultado el 25-09-2011 en www.camerfirma.com].

⁹¹ Fábrica Nacional de Moneda y Timbre, que lidera el denominado proyecto CERES (Certificación Española) consistente en el establecimiento de la misma como Entidad Pública de Certificación que permita autentificar y garantizar la confidencialidad de las comunicaciones entre ciudadanos, empresas u otras instituciones y administraciones públicas a través de las redes abiertas de comunicación [Consultado el 25-09-2011 en <http://www.cert.fnmt.es>].

ANEXO II. Glosario de Siglas y Acrónimos

ACCV: Agencia de Tecnología y Certificación Electrónica, proporciona a los ciudadanos, las empresas y las Administraciones Públicas los mecanismos de identificación telemática segura en los trámites administrativos a través de Internet: los certificados digitales y las tecnologías asociadas. Los certificados emitidos y el resto de servicios que la ACCV presta se ajustan a lo establecido en la Ley 59/2003, de 19 de diciembre, de firma electrónica, por lo que la ACCV goza de amplio reconocimiento en todas las Administraciones Públicas. Además, con los certificados digitales reconocidos expedidos por la ACCV se puede generar firma electrónica reconocida, que es equivalente a la manuscrita [Consultado el 25-09-2011 en <http://www.accv.es>].

ACE: Agencia de Certificación Electrónica, compañía del Grupo Telefónica que centra su actividad en proporcionar servicios de confianza que garanticen la seguridad de las transacciones electrónicas en el ámbito del comercio electrónico: autenticación (identificación) de las personas, dispositivos y organizaciones involucradas; integridad de la información intercambiada en la transacción; confidencialidad (secreto) de la información intercambiada; no repudio de la intervención de los participantes en una operación [Consultado el 25-09-2011 en <http://www.ace.es>].

AEDOC Digital: Asociación Española de Documentación Digital, es la asociación española de empresas especializada en gestión documental. Es una organización no lucrativa nacida con la idea de extender la incorporación de nuevas tecnologías aplicadas a la gestión de la información, de documentos y de contenidos en empresas españolas e instituciones como medio para reducir gastos, y aumentar la productividad económica. AedocDigital ayuda a empresas e instituciones a entrar en la sociedad de la información por promoviendo la desaparición gradual de papel para alcanzar la e-administración y el comercio electrónico [Consultado el 25-09-2011 en http://www.cdm-europe.info/index.php?option=com_content&view=article&id=55&Itemid=60].

AENOR: La Asociación Española de Normalización y Certificación es una institución española, privada, independiente, sin ánimo de lucro cuyas funciones son las de elaborar normas técnicas españolas (UNE) con la participación abierta a todas las partes interesadas y representar a España en los distintos organismos de normalización regionales e internacionales además de certificar productos, servicios y empresas. Fue designada para ello por la Orden del Ministerio de Industria y Energía, de 26 de febrero de 1986, de acuerdo con el Real Decreto 1614/1985 y reconocida como organismo de normalización y para actuar como entidad de certificación por el Real Decreto 2200/1995, en desarrollo de la Ley 21/1992, de Industria [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/AENOR>].

AIIM: Association for Information and Image Management, fundada en 1943 como la Asociación Nacional del Microfilm, es una organización internacional sin ánimo de lucro enfocada a estudiar y

difundir los avances relacionados con la gestión documental, de contenidos y procesos de negocio [Consultado el 18-07-2011 en www.aiim.org].

ANSI: American National Standards Institute es una organización sin ánimo de lucro que supervisa el desarrollo de estándares para productos, servicios, procesos y sistemas en los Estados Unidos. ANSI es miembro de la Organización Internacional para la Estandarización (ISO) y de la Comisión Electrotécnica Internacional (International Electrotechnical Commission, IEC) [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/Ansi>]

BMP: BitMaP, es el formato propio del programa Microsoft Paint, que viene con el sistema operativo Windows. Puede guardar imágenes de 24 bits, 8 bits y menos. Puede darse a estos archivos una compresión sin pérdida de calidad [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/BMP>]

C10: Comité de la AIMM (Association for Information and Image Management) responsable del establecimiento de procedimientos de control de calidad y estándares del documento origen y otras especificaciones en el ámbito de la conversión digital [consultado el 18-07-2011 en www.aiim.org].

CAMERFIRMA: Servicio de certificación digital de las Cámaras de Comercio, Industria y Navegación de España [Consultado el 25-09-2011 en www.camerfirma.com].

CCITT: Consultative Committee for International Telegraphy and Telephony. Véase ITU.

CCITT Grupo 3: Véase ITU-T y ITU-T T.4

CCITT Grupo 4: Véase ITU-T y ITU-T T.6

CERE: International Workshop on Comparative Evaluation in Requirements Engineering [Consultado el 25-09-2011 en http://www.di.unipi.it/CERE06/P5%20CERE%2006%20Kitapci_Hasan.pdf]

CINDOC: Centro de Información y Documentación Científica. Véase IEDCYT).

CNEDA: Comisión de Normas Españolas de Descripción Archivística, creada por Orden del Ministerio de Cultura de 25 de mayo de 2007 (Orden CUL/1524/2007) y adscrita al Ministerio de Cultura a través de la Dirección General del Libro, Archivos y Bibliotecas. Constituye un órgano colegiado con el carácter de grupo de trabajo de los previstos en el artículo 40.3 de la LOFAGE. El objetivo de la CNEDA es el asesoramiento al Ministerio de Cultura en el desarrollo y actualización de las NEDA, que quedan definidas como la herramienta normativa aplicable en la descripción de documentos de archivo, orientada a la mejora continua del acceso a los recursos archivísticos [Consultado el 25-09-2011 en <http://www.mcu.es/archivos/MC/CNEDA/Presentacion.html>].

CSIC: El Consejo Superior de Investigaciones Científicas es una agencia estatal adscrita al Ministerio de Ciencia e Innovación, a través de la Secretaría de Estado de Investigación cuyo objetivo fundamental es desarrollar y promover investigaciones en beneficio del progreso científico y tecnológico, para lo cual está abierta a la colaboración con entidades españolas y extranjeras. Tiene carácter multidisciplinar y realiza investigaciones avanzadas en todas las áreas científicas gracias a sus más de 130 centros distribuidos en todas las Comunidades Autónomas de España [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/CSIC>].

DjVu: formato de archivo informático diseñado principalmente para almacenar imágenes escaneadas. Se caracteriza por incorporar avanzadas tecnologías tales como separación de capas de imágenes, carga progresiva, codificación aritmética y compresión sin pérdida para imágenes bitonales (dos colores), permitiendo que imágenes de alta calidad se almacenen en un mínimo de espacio [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/DjVu>].

DIN: El Deutsches Institut für Normung es el organismo nacional de normalización de Alemania. Elabora, en cooperación con el comercio, la industria, la ciencia, los consumidores e instituciones públicas, estándares técnicos (normas) para la racionalización y el aseguramiento de la calidad. El DIN representa los intereses alemanes en las organizaciones internacionales de normalización. El acrónimo DIN también ha sido interpretado como Deutsche Industrie Norm y Das Ist Norm [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/DIN>].

DLF: Digital Libraries Federation es un consorcio de bibliotecas fundado en 1995 que promueve el empleo de tecnologías de información electrónica para ampliar las colecciones y servicios de las bibliotecas. A través de sus miembros, la DLF se dedica a la identificación de estándares para colecciones digitales y redes de acceso, la coordinación de actividades de investigación y desarrollo en el uso de tecnologías de la información, iniciar proyectos y servicios que las bibliotecas necesitan pero no pueden desarrollar individualmente [Consultado el 25-09-2011 en <http://www.clir.org/dlf.html>].

DLM-Forum: Originalmente "Máquina Données par Lisibles", en su conferencia de 2002 en Barcelona, se aprobó una resolución cambiando esto, así que ahora DLM significa Documento de gestión del ciclo vital. De acuerdo con su Constitución, uno de los objetivos del Foro DLM es: "para facilitar la transferencia de tecnología y conocimientos y servicios de información, directrices apropiadas práctica, los indicadores de referencia y de información, desarrollo educativo, habilidades y oportunidades de investigación. La constitución también define otros objetivos: Proporcionar los servicios de información (en línea portal de información, la lista de ser-, publicaciones, foro de discusión, noticias, archivo de documentos dinámicos), Disponer la transferencia de tecnología y conocimiento (vigilancia tecnológica, normas, legislación y supervisión de políticas, de consultoría, proveedores de tecnología de conexión con los usuarios, desarrollo de asociaciones); Proporcionar a la educación y el desarrollo de habilidades (conferencias, seminarios y exposiciones, talleres y

sesiones de capacitación, dirigido entornos de aprendizaje), Proporcionar a la investigación (análisis de las necesidades del usuario, modelos económicos, de desarrollo de ontologías, ingeniería del conocimiento, clasificación, recuperación de información) [Consultado el 25-09-2011 en http://en.wikipedia.org/wiki/DLM_Forum].

DPI: Dots Per Inch (o Puntos Por Pulgada -ppp-) es una unidad de medida para resoluciones de impresión, concretamente, el número de puntos individuales de tinta que una impresora o tóner puede producir en un espacio lineal de una pulgada [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Puntos_por_pulgada].

DTD: Document Type Definition, es una descripción de estructura y sintaxis de un documento XML o SGML. Su función básica es la descripción del formato de datos, para usar un formato común y mantener la consistencia entre todos los documentos que utilicen la misma DTD. De esta forma, dichos documentos, pueden ser validados, conocen la estructura de los elementos y la descripción de los datos que trae consigo cada documento, y pueden además compartir la misma descripción y forma de validación dentro de un grupo de trabajo que usa el mismo tipo de información [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/DTD>].

ER Mapper: software que emplea algoritmos matemáticos para el procesamiento digital de imágenes, teledetección y composición cartográfica, que permite la integración, realce, visualización e interpretación de datos geográficos [Consultado el 25-09-2011 en <http://www.erdas.com/products/ERDASERMapper/ERDASERMapper/Details.aspx>].

FESABID: Federación Española de Sociedades de Archivística, Biblioteconomía, Documentación y Museística. Es una entidad de derecho privado, sin ánimo de lucro, fundada en 1988 con el fin de reunir a las asociaciones profesionales más relevantes del sector [Consultado el 25-09-2011 en <http://www.fesabid.org>].

FESTE: Autoridad de certificación de la Fundación para el Estudio de la Seguridad de las Telecomunicación, entidad en la que participan, entre otros, los notarios y corredores de comercio electrónico [Consultado el 25-09-2011 en <http://www.invenia.es/feste>].

FlashPix: Formato de archivo donde se guarda la imagen en más de una resolución de forma que, cuando se envía una solicitud para mostrar la imagen a un navegador web, sólo se manda la resolución necesaria para su representación en pantalla, lo que por que ahorra ancho de banda y tiempo de descarga [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/FlashPix>].

FNMT: Fábrica Nacional de Moneda y Timbre, que lidera el denominado proyecto CERES (Certificación Española) consistente en el establecimiento de la misma como Entidad Pública de

Certificación que permita autentificar y garantizar la confidencialidad de las comunicaciones entre ciudadanos, empresas u otras instituciones y administraciones públicas a través de las redes abiertas de comunicación [Consultado el 25-09-2011 en <http://www.cert.fnmt.es>].

GIF: Graphic Interchange Format, es un formato gráfico utilizado ampliamente en la World Wide Web, tanto para imágenes como para animaciones [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/Gif>].

ICA: International Council of Archives es una organización profesional de la comunidad de archivos, dedicada a promover la conservación, desarrollo y utilización del patrimonio mundial de los archivos. Reúne a gestores de archivos nacionales, asociaciones profesionales de archiveros, archivos locales y regionales y archivos de otras organizaciones así como archiveros a título individual. CIA es una organización neutra no gubernamental, financiada por sus miembros [Consultado el 25-09-2011 en <http://www.ica.org>].

ICR: Reconocimiento de Caracteres Inteligente, proporciona a los sistemas de reproducción por escáner y sistemas de imágenes la habilidad de convertir caracteres en letra manuscrita (no cursiva) en caracteres capaces de ser interpretados o reconocidos por una computadora. Así, las imágenes de caracteres en letra manuscrita son extraídas de un mapa de bits de la imagen reproducida por el escáner [Consultado el 25-09-2011 en <http://espanol.scantron.com/sistemas/comparacion.htm>].

IEDCYT: Instituto de Estudios Documentales sobre Ciencia y Tecnología constituido a partir del antiguo Centro de Información y Documentación Científica (CINDOC). Es un organismo del Consejo Superior de Investigaciones Científicas (CSIC), cuya principal línea de actuación es analizar, recopilar, difundir y potenciar la información científica en todas las áreas del conocimiento [Consultado el 25-09-2011 en <http://www.cindoc.csic.es>].

IFLA: International Federation of Library Associations and Institutions es es una organización independiente, internacional, no gubernamental, sin fines de lucro fundada en Edimburgo, Escocia, en 1927. Sus objetivos principales son: promover estándares de provisión y prestación de servicios bibliotecarios y de información, favorecer una mayor comprensión del valor de los buenos servicios bibliotecarios y de información y representar los intereses de sus miembros en el mundo [Consultado el 25-09-2011 en <http://www.ifla.org>].

IPSCA: Certification Authority CA, compañía cuyo certificado raíz (ipsCA Global CA Root) está incorporado en los productos Microsoft desde 2009. Los certificados ofrecidos por IPSCA son los siguientes: certificados de servidor, certificados personales, certificados para firma de código y certificados para IPSEC-VPN [Consultado el 25-09-2011 en <http://web.ipsca.com/es>]

ISBN: International Standard Book Number, identificador único para libros, previsto para uso comercial. Fue creado en el Reino Unido en 1966 por las librerías y papelerías británicas W. H. Smith y llamado originalmente Standard Book Numbering), abreviado SBN. Fue adoptado como estándar internacional ISO 2108 en 1970 [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/ISBN>].

ISO: Organización Internacional de Normalización o ISO (del griego "isos", "igual"), nacida tras la Segunda Guerra Mundial (23 de febrero de 1947), es el organismo encargado de promover el desarrollo de normas internacionales de fabricación, comercio y comunicación para todas las ramas industriales a excepción de la eléctrica y la electrónica. Su función principal es la de buscar la estandarización de normas de productos y seguridad para las empresas u organizaciones a nivel internacional. La ISO es una red de los institutos de normas nacionales de 160 países, sobre la base de un miembro por país, con una Secretaría Central en Ginebra (Suiza) que coordina el sistema. La Organización Internacional de Normalización (ISO), con sede en Ginebra, está compuesta por delegaciones gubernamentales y no gubernamentales subdivididos en una serie de subcomités encargados de desarrollar las guías que contribuirán al mejoramiento ambiental. Las normas desarrolladas por ISO son voluntarias, comprendiendo que ISO es un organismo no gubernamental y no depende de ningún otro organismo internacional, por lo tanto, no tiene autoridad para imponer sus normas a ningún país. El contenido de los estándares está protegido por derechos de copyright y para acceder ellos el público corriente debe comprar cada documento, que se valoran en francos suizos (CHF). Está compuesta por representantes de los organismos de normalización nacionales, que produce normas internacionales industriales y comerciales. Dichas normas se conocen como normas ISO y su finalidad es la coordinación de las normas nacionales, en consonancia con el Acta Final de la Organización Mundial del Comercio, con el propósito de facilitar el comercio, el intercambio de información y contribuir con normas comunes al desarrollo y a la transferencia de tecnologías [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Organización_Internacional_para_la_Estandarización].

ITU: International Telecommunication Union (antes CCITT: Consultative Committee for International Telegraphy and Telephony) es el organismo especializado de la Organización de las Naciones Unidas encargado de regular las telecomunicaciones a nivel internacional entre las distintas administraciones y empresas operadoras. En general, la normativa generada por la ITU está contenida en un amplio conjunto de documentos denominados Recomendaciones, agrupados por Series. Cada serie está compuesta por las Recomendaciones correspondientes a un mismo tema. Aunque en las Recomendaciones nunca se "ordena", solo se "recomienda", su contenido, a nivel de relaciones internacionales, es considerado como mandatorio por las Administraciones y Empresas Operadoras [Consultado el 25-09-2011 en <http://www.itu.int>].

ITU-T: Sección de Normalización de las Telecomunicaciones de la ITU (International Telecommunication Union). Ha desarrollado una serie de protocolos de comunicaciones para la

transmisión de imágenes a través líneas telefónicas y redes de datos. Estos protocolos se conocen oficialmente como el ITU-T T.4 T.6, pero son más comúnmente conocido como Grupo 3 y Grupo 4 de compresión, respectivamente [Consultado el 25-09-2011 en <http://www.itu.int>].

ITU-T T.4: Conocido como CCITT Grupo 3, es un es un algoritmo unidimensional de compresión de imágenes desarrollado por la ITU utilizado por la mayoría de máquinas de fax [Consultado el 25-09-2011 en <http://www.itu.int>].

ITU-T T.6: Conocido como CCITT Grupo 4, es un algoritmo de compresión de imágenes de dos dimensiones desarrollado por la ITU. Obtiene resultados más eficientes que el T.4, por lo que es de común aplicación a las imágenes resultantes del escaneado [Consultado el 25-09-2011 en <http://www.itu.int>].

ITU-T T.82: Recomendación desarrollada por la ITU para la compresión de imágenes en múltiples capas. Se utiliza para codificar imágenes en escalas de grises y en color con número limitado de bits por pixel. Puede considerarse como una codificación de calidad tipo fax, similar a las codificaciones ITU-T T.4 e ITU-T T.6, con entre 20 y 80% de mejora en compresión sobre estos métodos [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Joint_Bi-level_Image_Experts_Group].

ITU-T T.88: Recomendación desarrollada por la ITU para la compresión de imágenes en blanco y negro [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Joint_Bi-level_Image_Experts_Group].

JBIG: Joint Bilevel Image Group Bitmap Graphics, escala de grises y color. Véase ITU-T T.82.

JBIG2: Joint Bilevel Image Group Bitmap Graphics, bitonal. Véase ITU-T T.88.

JFIF: JPEG File Interchange Format, es un archivo de mapa de bits comprimido con la tecnología JPEG [Consultado el 25-09-2011 en http://www.profesormolina.com.ar/tecnologia/informatica/glosario_pc.htm]

JP2: JPEG 2000, es un estándar de compresión y codificación digital de imágenes. Fue creado por el Joint Photographic Experts Group (JPEG), en el año 2000 con la intención de sustituir el formato original creado en 1992. El nuevo formato se basa en la transformada wavelet, en lugar de la transformada de coseno discreta establecida para el estándar original [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/JP2>]

JPEG: Joint Photographic Experts Group, es el nombre de un comité de expertos que creó un estándar de compresión y codificación de archivos de imágenes fijas. Además de ser un método de compresión, es a menudo considerado como un formato de archivo. JPEG/Exif es el formato de

imagen más común utilizado por las cámaras fotográficas digitales y otros dispositivos de captura de imagen, junto con JPG/JFIF, que también es otro formato para el almacenamiento y la transmisión de imágenes fotográficas en la World Wide Web. Estas variaciones de formatos a menudo no se distinguen, y se llaman JPEG [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/JPG>].

LuraWave: familia de software de compresión de imágenes que se basa en las técnicas de compresión wavelet [Consultado el 25-09-2011 en <http://www.luratech.com>]

LZW: Lempel-Ziv-Welch es un algoritmo de compresión sin pérdida desarrollado por Terry Welch en 1984 como una versión mejorada del algoritmo LZ78 desarrollado por Abraham Lempel y Jacob Ziv. Ampliamente utilizada, por ejemplo, en archivos TIFF [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/LZW>].

MrSID: Multi-resolution Seamless Image Database, estándar abierto de compresión de imágenes raster desarrollado por Los Alamos National Laboratory y comercializado por la empresa LizardTech. Permite mostrar archivos digitales de gran tamaño con un tiempo de carga mínimo gracias a la tecnología wavelet [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/MrSID>].

OCR: Reconocimiento Óptico de Caracteres, tecnología que proporciona a los sistemas de reproducción por escáner y sistemas de imágenes la habilidad de convertir imágenes de caracteres en letra de máquina, en caracteres capaces de ser interpretados o reconocidos por una computadora. Así, las imágenes de caracteres en letra de máquina son extraídas de un mapa de bits de la imagen reproducida por el escáner [Consultado el 25-09-2011 en <http://espanol.scantron.com/sistemas/comparacion.htm>].

PDF: Portable Document Format, es un formato de almacenamiento de documentos, desarrollado por la empresa Adobe Systems. Este formato es de tipo compuesto (imagen vectorial, mapa de bits y texto) [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/PDF>].

PMBOK: Project Management Body of Knowledge, es un estándar en la Administración de proyectos desarrollado por el Project Management Institute (PMI). La misma comprende dos grandes secciones, la primera sobre los procesos y contextos de un proyecto, la segunda sobre las áreas de conocimiento específico para la gestión de un proyecto [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/PMBOK>].

PNG: Portable Network Graphics, es un formato gráfico basado en un algoritmo de compresión sin pérdida para bitmaps no sujeto a patentes. Este formato fue desarrollado en buena parte para solventar las deficiencias del formato GIF y permite almacenar imágenes con una mayor profundidad de contraste [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/PNG>].

PostScript: lenguaje de descripción de páginas utilizado en muchas impresoras y, de manera usual, como formato de transporte de archivos gráficos en talleres de impresión profesional [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/PostScript>].

QI: Quality Index [consultado 21-03-2011 en www.library.cornell.edu/preservation/tutorial-spanish/contents.html]

SPIFF: Still Picture Interchange File Format, formato de archivo de intercambio de dibujos fijos, tiene la designación ISO 10918-3 y ofrece compresión, gestión de color y capacidad de metadatos más versátiles que JPEG/JFIF, pero tiene poco soporte [Consultado el 25-09-2011 en <http://www.library.cornell.edu/preservation/tutorial-spanish/presentation/table7-1.html>]

TECNIMAP: Jornadas sobre Tecnologías de la Información para la Modernización de las Administraciones Públicas organizadas por el Consejo Superior de Administración Electrónica, de acuerdo con lo previsto en el Real Decreto 589/2005, de 20 de mayo por el que se reestructuran los órganos colegiados responsables de la Administración electrónica [Consultado el 25-09-2011 en <http://www.tecnimap.es>].

TIFF: Tagged Image File Format es un formato de fichero para imágenes [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/TIFF>].

UNE: Una Norma Española, son un conjunto de normas tecnológicas creadas por los Comités Técnicos de Normalización (CTN), de los que forman parte todas las entidades y agentes implicados e interesados en los trabajos del comité. Por regla general estos comités suelen estar formados por AENOR, fabricantes, consumidores y usuarios, administración, laboratorios y centros de investigación [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Una_Norma_Española].

UNESCO: La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura es un organismo especializado de las Naciones Unidas. Se fundó el 16 de noviembre de 1945 con el objetivo de contribuir a la paz y a la seguridad en el mundo mediante la educación, la ciencia, la cultura y las comunicaciones. La constitución firmada ese día entró en vigor el 4 de noviembre de 1946. Se dedica a orientar a los pueblos en una gestión más eficaz de su propio desarrollo, a través de los recursos naturales y los valores culturales, y con la finalidad de modernizar y hacer progresar a las naciones del mundo, sin que por ello se pierdan la identidad y la diversidad cultural. La Unesco tiene vocación pacifista, y entre varias cosas se orienta muy particularmente a apoyar la alfabetización. En la educación, este organismo asigna prioridad al logro de la educación elemental adaptada a las necesidades actuales. Colabora con la formación de docentes, planificadores familiares y vivienda, administradores educacionales y alienta la construcción de escuelas y la dotación de equipo necesario para su funcionamiento. Las actividades culturales buscan la

salv guarda del patrimonio cultural mediante el estímulo de la creación y la creatividad y la preservación de las entidades culturales y tradiciones orales, así como la promoción de los libros y de la lectura. En materia de información, la Unesco promueve la libre circulación de ideas por medios audiovisuales, fomenta la libertad de prensa y la independencia, el pluralismo y la diversidad de los medios de información, vía el Programa Internacional para la Promoción de la Comunicación [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/Unesco>].

URI: Uniform Resource Identifier, es una cadena de caracteres corta que identifica inequívocamente un recurso (servicio, página, documento, dirección de correo electrónico, enciclopedia, etc.). Normalmente estos recursos son accesibles en una red o sistema. Los URI pueden ser localizadores uniformes de recursos, Uniform Resource Name, o ambos [Consultado el 25-09-2011 en http://es.wikipedia.org/wiki/Uniform_Resource_Identifier].

XSD: XML Schema Definition, es un lenguaje de esquema utilizado para describir la estructura y las restricciones de los contenidos de los documentos XML de una forma muy precisa, más allá de las normas sintácticas impuestas por el propio lenguaje XML. Se consigue así una percepción del tipo de documento con un nivel alto de abstracción. Fue desarrollado por el World Wide Web Consortium (W3C) y alcanzó el nivel de recomendación en mayo de 2001 [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/XSD>].

ZIP: Zone Information Protocol, es un formato de almacenamiento sin pérdida, muy utilizado para la compresión de datos como documentos, imágenes o programas [Consultado el 25-09-2011 en <http://es.wikipedia.org/wiki/.zip>]