

2013-10

# Preservación digital de ficheros máster de prensa actual: El caso del Consorcio de Digitalización de Prensa Vasca

Uriarte Gonzalo-Bilbao, Enrique

---

<http://hdl.handle.net/10016/17889>

---

*Descargado de e-Archivo, repositorio institucional de la Universidad Carlos III de Madrid*

**PRESERVACIÓN DIGITAL  
DE FICHEROS MÁSTER  
DE PRENSA ACTUAL**

El caso del Consorcio de Digitalización de  
Prensa Vasca

**Enrique Uriarte Gonzalo-Bilbao**

*Trabajo Fin de Máster*

MÁSTER UNIVERSITARIO EN BIBLIOTECAS Y SERVICIOS DE INFORMACIÓN DIGITAL  
Facultad de Humanidades, Comunicación y Documentación

**Universidad Carlos III de Madrid**

Madrid, Octubre 2013

Directora: **Dr. D<sup>a</sup>. Gema Bueno de la Fuente**

Enrique Uriarte Gonzalo-Bilbao, 2013

Obra bajo licencia *Creative Commons*

[Atribución-CompartirIgual 3.0 Unported](https://creativecommons.org/licenses/by-sa/3.0/) (CC BY-SA 3.0)



# RESUMEN / ABSTRACT / LABURPENA / RÉSUMÉ

---

## Resumen

Las bibliotecas del Consorcio de Digitalización de Prensa Vasca mantienen un proyecto de digitalización de los diarios contemporáneos relacionados con la cultura vasca. Se describen las características y contenidos del fondo digital generado. Se defiende la necesidad de acometer estrategias de preservación digital a la luz de las recomendaciones de diversos expertos y grupos internacionales. Se describen las opciones tecnológicas para establecer una infraestructura de almacenamiento capaz de afrontar los retos que la preservación digital plantea. Se concluye proponiendo las alternativas implementables por el Consorcio a menor coste y con mayor garantía de éxito.

## Palabras Clave

Digitalización; Prensa; Diarios; Hemerotecas; Preservación digital; Patrimonio digital; Patrimonio bibliográfico; Sistemas de almacenamiento

## Abstract

The libraries that make up the *Consortio de Digitalización de Prensa Vasca* keep a digitization project of contemporary newspapers related to Basque culture. Both characteristics and contents of the resulting digital collection are described. It is defended the need to undertake a project of digital preservation in the light of guidelines established by international experts and institutions. The paper also describes different technological choices in order to establish a storage infrastructure capable of facing the challenges posed by digital preservation. Finally, different choices are suggested to be implemented by the *Consortio* at a lower cost and with a higher guarantee of success.

## Keywords

Digitization; Press; Newspapers; Newspaper libraries; Digital preservation; Digital heritage; Bibliographical heritage; Digital storage systems

## Laburpena

Euskal Prentsa Digitalizatzeko Partzuergoa osatzen duten liburutegiek euskal kulturaren egunkari garaikideak digitalizazio proiektua mantentzen dute. Sortutako funts digitalaren ezaugarriak eta edukiak deskribatzen dira. Adituen eta nazioarteko ikerketa-talde desberdinen gomendioak jarraituz, babes digitaleko estrategiei ekiteko beharra argudiatzen da. Babes digitalak planteatzen dituen erronkei aurre egiteko biltegitratze azpiegitura ezartzeko aukera teknologikoak deskribatzen dira. Kostu txikiagora eta arrakastaren berme handiagoarekin ezar daitezkeen alternatibak proposatuz amaitzen du.

## Hitz Gakoak

Digitalizazioa; Prentsa; Egunkariak; Hemerotekak; Babes digitala; Ondare digitala; Ondare bibliografikoa; Biltegitratze sistemak

## Résumé

Les bibliothèques du Consortium de Numérisation de la Presse Basque soutiennent un projet de numérisation de journaux contemporains liés à la culture basque. On décrit les caractéristiques et les contenus du fond numérique généré. On défend la nécessité d'entreprendre des stratégies de préservation numérique, compte tenu des recommandations des experts et des groupes internationaux. On décrit les options technologiques pour l'établissement d'une infrastructure de stockage capable de relever les défis de la conservation numérique. On conclut en proposant des alternatives réalisables par le Consortium, à un moindre prix et avec une plus grande garantie de succès.

## Mots-clés

Numérisation ; Presse ; Journaux ; Hémérothèques ; Préservation numérique ; Patrimoine numérique ; Patrimoine bibliographique ; Systèmes de stockage numérique

# ÍNDICE DE FIGURAS

---

Figura 1.- Colección digitalizada del diario El País en el catálogo de la biblioteca de la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid.....	23
Figura 2.- Razones de la pérdida de datos (Fuente: Henriksen, Seuskens y Wijers, 2013, citando The insurance Agency Inc., 2001 y Ontrack Data International, inc., 2003) .....	27
Figura 3.- Factores de sostenibilidad en formatos para la preservación digital (Fuente: Formatos de difusión y formatos de preservación de contenidos digitales. Observatorio Vasco de la Cultura, 2011) .....	44
Figura 4.- Principales extensiones de archivos de imagen RAW (Fuente: Wikipedia).....	45
Figura 5.- Integrantes del Consorcio de Digitalización de Prensa Vasca, con fechas de incorporación y abandono del mismo.....	53
Figura 6.- Proceso de digitalización de diarios por el Consorcio de Digitalización de Prensa Vasca (Elaboración propia) .....	58
Figura 7.- Interfaz de búsqueda de Hemeroteca+ en la Biblioteca de la Universidad de Navarra .....	62
Figura 8.- Ciclo de costes del proyecto del Consorcio de Digitalización de Prensa Vasca (Elaboración propia) .....	70
Figura 9.- Discos duros externos almacenados en los depósitos controlados de la Fundación Sancho el Sabio. ....	74
Figura 10.- Metadatos técnicos EXIF incrustados en el fichero de imagen ráster TIFF de la primera página del diario Deia de uno de diciembre de 1996.....	84
Figura 11.- Digitalización de baja calidad: página no alineada, manchas negras, textos ilegibles, imagen no visible.....	85
Figura 12.- Diarios digitalizados en la Biblioteca Central del Gobierno Vasco, accesibles en sala mediante CD-ROM. ....	88
Figura 13.- Prensa actual difundida en acceso abierto en el repositorio digital “Hemeroteca Digitala”. .....	89
<i>Figura 14.- Mapa conceptual de medios de almacenamiento para la preservación digital (Elaboración propia) .....</i>	<i>92</i>
Figura 15.- Condiciones de almacenamiento para CD y DVD. (Fuente: Digital Media Group, 2007) .	102
Figura 16.- Mapa conceptual de las alternativas de almacenamiento para la preservación digital del fondo del Consorcio de Digitalización de Prensa Vasca (Elaboración propia).....	116
Figura 17.- Guía para el cuidado y manejo de discos ópticos (Fuente: Digital Media Group, 2007)..	135

## ÍNDICE DE TABLAS

---

Tabla 1.- Parámetros de digitalización contratados y realmente hallados en el muestreo (CD-ROM) y la verificación (HD) .....	60
Tabla 2.- Cabeceras digitalizadas y años cubiertos .....	64
Tabla 3.- Cabeceras en microfichas .....	64
Tabla 4.- Número de CD-ROM por cabecera, años y tamaño estimado de los datos .....	66
Tabla 5.- Relación de discos duros .....	67
Tabla 6.- Número de cabeceras digitalizadas por año en HD .....	68
Tabla 7.- Número de cabeceras digitalizadas por año en CD-ROM .....	68
Tabla 8.- Costes económicos .....	72
Tabla 9.- Marcas, modelos y características de los discos duros externos utilizados .....	76
Tabla 10.- Categorización de los problemas de la colección digital del Consorcio de Digitalización de Prensa Vasca.....	81
Tabla 11.- Precios estimados para diferentes tipos de medios de almacenamiento (Fuente: Henriksen et al., 2013).....	98
Tabla 12.- Condiciones ambientales de almacenamiento de soportes y años de duración. Tomado de la síntesis realizada por Lluca (2012a) a partir de Jones y Beagrie (2008) .....	102
Tabla 13.- Integrantes del Consorcio de Digitalización de Prensa Vasca, con fechas de incorporación y abandono del mismo.....	131

## LISTA DE ABREVIATURAS

---

<b>ALTO</b>	<i>Technical Metadata for Optical Character Recognition</i>
<b>ASCII</b>	<i>American Standard Code for Information Interchange</i>
<b>ATHA</b>	Archivo del Territorio Histórico Álava
<b>BFL</b>	<i>Bizkaiko Foru Liburutegia</i> - Biblioteca Foral de Bizkaia
<b>BNE</b>	Biblioteca Nacional de España
<b>CDPV</b>	Consortio de Digitalización de Prensa Vasca
<b>CEDALMAC</b>	Comisión Española sobre la Digitalización y la Accesibilidad en Línea del Material y de la Conservación Digital
<b>DC</b>	<i>Dublin Core</i>
<b>dSaaS</b>	<i>Data Storage as a Service</i>
<b>DPI</b>	<i>Dots Per Inch</i> (Puntos por pulgada, gotas por pulgada)
<b>DVD</b>	<i>Digital Versatile Disc</i>
<b>EITB</b>	<i>Euskal Irrati Telebista</i> /Radio Televisión Pública Vasca
<b>EPI</b>	El Profesional de la Información
<b>EXIF</b>	<i>Exchangeable Image File Format for Digital Still Camera</i>
<b>E-LIS</b>	<i>Eprints in Library and Information Science</i>
<b>FESABID</b>	Federación Española de Sociedades de Archivística, Biblioteconomía, Documentación y Museística
<b>FSS</b>	Fundación Sancho el Sabio
<b>HD</b>	<i>Hard Disc</i> , disco duro
<b>ICON</b>	<i>International Council on Archives</i> (ICA en francés: <i>Conseil International des Archives</i> )
<b>IECISA</b>	Informática El Corte Inglés, S.A.
<b>IFLA</b>	<i>International Federation of Library Associations and Institutions</i>
<b>ISO</b>	<i>International Organization for Standardization</i>
<b>KMK</b>	<i>Koldo Mitxelena Kulturunea</i>
<b>LC</b>	<i>Library of Congress</i>
<b>LISA</b>	<i>Library and Information Science Abstracts</i>
<b>LISTA</b>	<i>Library, Information Science &amp; Technology Abstracts</i>

<b>LOCKSS</b>	<i>Lots of Copies Keep Stuff Safe</i>
<b>LOT</b>	<i>Linear Tape-Open</i>
<b>LPI</b>	Ley de Propiedad Intelectual
<b>LPI</b>	<i>Lines Per Inch</i> (Líneas por pulgada)
<b>MARC</b>	<i>Machine Readable Cataloging</i>
<b>METS</b>	<i>Metadata Encoding and Transmission Standard</i>
<b>OAI</b>	<i>Open Archives Initiative</i>
<b>OAIS</b>	<i>Open Archival Information System</i>
<b>OCR</b>	( <i>Optical Character Recognition</i> , Reconocimiento Óptico de Caracteres)
<b>OMPI</b>	Organización Mundial de la Propiedad Intelectual
<b>PPP</b>	Píxel Por Pulgada
<b>PPI</b>	<i>Pixel Per Inch</i> (Píxel Por Pulgada)
<b>PREMIS</b>	<i>Preservation Metadata Implementation Strategies</i>
<b>RAE</b>	Real Academia Española
<b>RAID</b>	<i>Redundant Array of Independent Disks</i>
<b>RI</b>	Recuperación de la Información
<b>SMS</b>	<i>Short Message Service</i> (Servicio de mensajes cortos en telefonía)
<b>TXT</b>	<i>Text File</i> (formato de fichero de texto, independiente de plataforma)
<b>UNAV</b>	Universidad de Navarra
<b>UPS</b>	<i>Uninterrupted Power System</i> (Sistema de alimentación ininterrumpida)
<b>UPV/EHU</b>	<i>Euskal Herriko Unibertsitatea</i> /Universidad del País Vasco
<b>XML</b>	<i>Extensible Markup Language</i>



## **Agradecimiento**

*Etxeko hirurentzat, zinez eskertuz!*

*A las personas e instituciones que forman parte del Consorcio de Digitalización de Prensa Vasca –Pacto de Damas y Caballeros– por facilitar la realización de este estudio.*

# CONTENIDO

---

RESUMEN / ABSTRACT / LABURPENA / RÉSUMÉ .....	3
ÍNDICE DE FIGURAS .....	4
ÍNDICE DE TABLAS .....	5
LISTA DE ABREVIATURAS.....	6
CONTENIDO.....	9
1. INTRODUCCIÓN.....	11
1.1. JUSTIFICACIÓN.....	12
1.2. OBJETO .....	13
1.3. OBJETIVOS.....	13
1.4. PREGUNTAS DE INVESTIGACIÓN .....	14
1.5. METODOLOGÍA.....	14
1.6. FUENTES EMPLEADAS .....	16
1.7. ALCANCE Y LÍMITES .....	17
1.8. ESTRUCTURA DEL TEXTO .....	19
2. CONTEXTO Y MARCO TEÓRICO: DIGITALIZAR PARA PRESERVAR Y PRESERVAR LO DIGITALIZADO.....	20
2.1. PRESERVAR LA PRENSA DIARIA .....	21
2.2. UN TÉRMINO NO EXENTO DE PROBLEMÁTICA .....	24
2.3. LA PRESERVACIÓN DIGITAL A ESTUDIO EN DIFERENTES ÁMBITOS.....	29
2.4. ¿DIGITALIZAR PARA PRESERVAR? .....	30
2.5. PLANES DE PRESERVACIÓN DIGITAL .....	35
2.6. ESTRATEGIAS Y MODELOS TEÓRICOS.....	37
2.7. ESTÁNDARES PARA LA PRESERVACIÓN .....	42
3. CASO DE ESTUDIO: EL FONDO DIGITALIZADO DE PRENSA VASCA.....	50
3.1. EL CONSORCIO DE DIGITALIZACIÓN DE PRENSA VASCA .....	51
3.2. CARACTERIZACIÓN DE LA COLECCIÓN MAESTRA DEL CONSORCIO DE DIGITALIZACIÓN DE PRENSA VASCA.....	55
3.3. VALORACIÓN TÉCNICA DEL FONDO DIGITAL RESULTANTE.....	79
3.4. ASPECTOS LEGALES .....	86
4. ALTERNATIVAS DE ALMACENAMIENTO .....	90
4.1. MEDIOS Y SISTEMAS DE ALMACENAMIENTO .....	93

4.2.	<i>CONSIDERACIONES SOBRE ALMACENES Y DEPÓSITOS CONTROLADOS</i> .....	102
5.	PROPUESTA FUNDAMENTADA.....	104
5.1.	<i>ESTRATEGIAS DE PRESERVACIÓN VIABLES</i> .....	105
5.2.	<i>ESTRATEGIAS DE PRESERVACIÓN DESCARTABLES</i> .....	111
5.3.	<i>DECISIONES Y DEBATES A AFRONTAR</i> .....	113
5.4.	<i>PLAN DE PRESERVACIÓN DIGITAL</i> .....	114
6.	CONCLUSIONES .....	117
	BIBLIOGRAFÍA.....	119
	GLOSARIO .....	128
	ANEXOS .....	130
	Anexo I.....	130
	Anexo II.....	132
	Anexo III.....	133
	Anexo IV .....	135

# 1. INTRODUCCIÓN

---

Hace ya más de dos décadas que se están llevando a cabo proyectos de digitalización de documentos en todo el mundo. Entre estos, no son pocos los que se han dedicado a los periódicos. La especial fragilidad del papel en que los diarios están impresos ha supuesto un acicate para promover la digitalización con fines de conservación. Las principales iniciativas se han enfocado hacia la prensa histórica anterior a la Segunda Guerra Mundial, sin embargo también son abundantes los proyectos ejecutados con diarios actuales. Editores de prensa, archivos, bibliotecas y empresas de ingeniería informática están realizando hoy día digitalizaciones sistemáticas de diarios en curso de publicación.

Las bibliotecas, y los archivos en menor medida, se han ocupado siempre de ofertar diarios a los usuarios entre sus recursos de información. Esta práctica, además de posibilitar la lectura de las noticias de actualidad, ha permitido conformar colecciones hemerográficas para ofrecer la consulta de información retrospectiva. No cabe la más mínima duda de que “la prensa constituye una fuente fundamental para el conocimiento y el estudio del pasado y que, por ello, es relevante poder asegurar su acceso durante muchos años” (Guallar y Abadal, 2010, p. 140). En este sentido, las bibliotecas juegan un papel primordial.

Los programas de digitalización del patrimonio, junto a la expansión y popularización de los documentos nacidos digitales, han modificado la naturaleza y composición de las colecciones bibliográficas y documentales. Las bibliotecas, archivos y museos o “Instituciones de la Memoria”, como las denomina la UNESCO, gestionaban hasta hace pocos años fondos compuestos exclusivamente por patrimonio en soportes tradicionales. Hoy día, sin embargo, gestionan también amplias colecciones de patrimonio digital en gran variedad de soportes y formatos electrónicos.

En la medida en que muchos de estos *outputs* electrónicos, emanados de todas las esferas de la actividad social, conforman nuevos mecanismos de acceso a la información, la formación, la cultura y el conocimiento, forman ya parte del patrimonio bibliográfico y documental. De esta manera, los fondos patrimoniales digitales han irrumpido con fuerza en las colecciones de las instituciones encargadas de preservar la memoria de la humanidad. Nadie duda de que éstas han de asumir la gestión de la producción intelectual en formato digital.

Pero la gestión de estos soportes documentales no tradicionales implica inevitablemente un cambio en los procesos de trabajo y una renovación en el *modus operandi* por parte de quienes tradicionalmente han desempeñado el rol de conservadores. Nuevos medios requieren de nuevos conocimientos. El uso intensivo de las TIC demanda innovadoras soluciones de custodia y preservación de la documentación contemporánea.

En este nuevo escenario de la preservación de la memoria digital han aparecido también otros actores. Se trata de organizaciones eminentemente tecnológicas, que operan en el campo de la conservación de los registros electrónicos con gran conocimiento del medio digital y de las tecnologías asociadas al mismo. Vienen desempeñando un papel muy dinámico en la salvaguarda de los recursos digitales e, inevitablemente, deben ser tenidas en cuenta de cara a labrar alianzas estratégicas que permitan alcanzar los objetivos de la preservación digital a largo plazo.

La particular naturaleza de los medios digitales ha provocado un replanteamiento en el modo de abordar la preservación de este tipo de documentos. Se viene demostrando que los métodos tradicionales no son suficientes. Nacen, por tanto, nuevos retos en relación con la defensa del patrimonio digital.

Enmarcado en este contexto tecnológico, y de sensibilización por la salvaguarda del patrimonio bibliográfico, surge el proyecto cooperativo del Consorcio de Digitalización de Prensa Vasca (CDPV) con la finalidad de conservar la prensa diaria que se edita actualmente en Euskal Herria. Las bibliotecas integrantes del CDPV se enfrentan hoy, al igual que muchas otras instituciones de la memoria, al reto de encarar la preservación a largo plazo de los objetos digitales generados.

### 1.1. JUSTIFICACIÓN

La UNESCO viene alertando en la última década de la necesidad de abordar programas de preservación digital (UNESCO, 2003a, 2003b). En paralelo, diversas organizaciones de todo el mundo, preocupadas por el problema de la durabilidad de la información digital, desarrollan proyectos e iniciativas investigadoras de toda índole relacionadas con la preservación digital; Galloway (2005), Gallart y Keefer (2007) y Térmens (2009a) han realizado balances y recopilaciones de estas iniciativas multidisciplinares de forma exhaustiva.

En la conferencia internacional sobre el acceso permanente al patrimonio documental digital, (*The Memory of the World in the Digital Age: Digitization and Preservation*), organizada por la UNESCO en colaboración con la Universidad de British Columbia en septiembre de 2012, se aprobó la *Declaración de Vancouver* donde se alerta de que *“la conservación digital debe ser una prioridad de desarrollo y es indispensable invertir en infraestructura para garantizar la fiabilidad de los registros digitales así como su accesibilidad a largo plazo”* (UNESCO/UBC, 2012, p. 2).

Así las cosas, transcurridos cerca de veinte años desde que el CDPV comenzase las tareas sistemáticas y diarias de digitalización de las cabeceras de prensa, conviene analizar con cierto detalle la situación de la propia colección y el estado del arte sobre preservación digital, para valorar las mejores opciones de trabajo en relación con una política garantista de acceso a los datos en el futuro. Las bibliotecas vascas no pueden estar ajenas al devenir de los tiempos y al desarrollo de los acontecimientos en relación con la custodia del patrimonio digital. Por una parte, en este tiempo se han producido importantes cambios y avances tecnológicos relacionados con la digitalización y con la preservación digital. Y por otra, hay que considerar que la información convertida a formato digital durante el transcurso del proyecto del CDPV alcanza ya un gran volumen de datos.

El hecho de contar con este gran volumen de datos, junto con la ausencia de una política de preservación detallada e implementada, ha hecho pensar en repetidas ocasiones en la necesidad de establecer un marco operativo bien definido y un plan de preservación acorde a los fines que se persiguen. En diversas actas de reunión del Consorcio aparecen expuestas ciertas inquietudes en relación con la problemática de conservación de la colección de másteres y la necesidad de hacer frente a esta cuestión. Con el presente estudio se pretende ayudar a alcanzar ese objetivo.

## 1.2. OBJETO

El objeto de la investigación es la versión maestra de ficheros de imagen digital del fondo hemerográfico contemporáneo, de ámbito vasco, digitalizado por el Consorcio de Digitalización de Prensa Vasca. Se persigue elaborar una propuesta fundamentada de estrategias de acción y métodos de almacenamiento que ayude a garantizar su preservación a largo plazo. Todo ello a la luz del estado del arte en preservación digital.

Fundamentalmente, se trata de analizar las soluciones técnicas de almacenamiento con las que pueden contar las bibliotecas del CDPV para preservar sus recursos digitales, habida cuenta de que, en relación con la documentación digital, la mera conservación de los soportes no implica la adecuada accesibilidad, legibilidad o integridad de la información en el futuro. En la valoración de las distintas opciones posibles se tratará de medir los costes asociados a su implementación.

Las bibliotecas regionales medianas que han emprendido proyectos de digitalización hace años y que, como resultado de ello, custodian colecciones digitales de un tamaño considerable, en general no cuentan con experiencia previa en preservación digital, como se pueda dar en el caso de las bibliotecas universitarias en la gestión de recursos electrónicos o en el caso de las bibliotecas nacionales en la gestión de grandes proyectos de digitalización de patrimonio histórico. El problema que se plantea es que han de establecer ahora estrategias de acción, enmarcadas en planes de preservación digital, más allá del mero almacenamiento controlado, de la replicación periódica de los datos o del refresco de los soportes.

Las variables organizacionales, económicas y legales son vectores, que junto a los requerimientos técnicos, cobran notable importancia en todo programa de preservación digital y que lógicamente, afectan también a la gestión del fondo del Consorcio. Estos parámetros, aún sin ser el objeto central de esta investigación, recorren de forma transversal todo el planteamiento del problema.

## 1.3. OBJETIVOS

El objetivo general de este trabajo es proponer alternativas técnicas viables económicamente para la preservación a largo plazo de la colección de ficheros máster de los periódicos digitalizados por el Consorcio de Digitalización de Prensa Vasca.

A la vista de los condicionantes propios de la colección estudiada, y conforme a experiencias documentadas y soluciones teóricas extraídas de la literatura científica, se propondrá al CDPV un conjunto de recomendaciones prácticas y una propuesta fundamentada sobre las alternativas más adecuadas.

Como objetivos específicos, el detalle de tales alternativas pretende proporcionar información sobre:

- ~ las soluciones técnicas de almacenamiento y custodia adecuadas al fondo,
- ~ las estrategias de preservación más apropiadas,
- ~ y en lo posible, una estimación de costes asociados a la implementación de las mismas.

Todo ello, a fin de que el CDPV, si lo estima conveniente, pueda definir una política de preservación digital, plasmarla en un plan de acción y seleccionar justificadamente una opción de trabajo para los próximos años.

#### 1.4. PREGUNTAS DE INVESTIGACIÓN

Conforme al marco de trabajo expuesto, se plantean seguidamente varias cuestiones a las que se pretende dar respuesta con esta investigación:

- ~ El CDPV viene gestionando la preservación del fondo con una serie de acciones básicas. En relación con éstas, ¿es suficiente con dejar los ficheros en los soportes en que actualmente están almacenados? ¿Basta con realizar copias de seguridad, *backups* o copias de respaldo? ¿Es el refresco de los soportes una estrategia suficiente?
- ~ En lo relativo a la solución técnica de almacenamiento que requiere el fondo del CDPV, ¿es preferible el uso de discos ópticos, de discos duros externos, la transferencia de los datos a servidores propios, el uso de discos RAID, el establecimiento de una arquitectura cooperativa en *GRID*, la carga de los recursos en un repositorio digital o la contratación de un servicio de almacenamiento en la nube? ¿Es realmente el *cloud* de almacenamiento o *dSaaS*<sup>1</sup> una solución para este tipo de fondos?
- ~ En cuanto a la digitalización de fondos cuyos derechos de propiedad intelectual son de titularidad ajena a la institución, ¿pueden actualmente las bibliotecas asumir la responsabilidad de gestionar su preservación digital, sin contar con el permiso de los titulares de los derechos de las obras que se pretenden conservar? O, expresado de otra manera, ¿qué límites de actuación impone la carencia de tales permisos en cuanto a las estrategias y acciones de conservación que se pueden emprender?
- ~ Equilibrando los costes a los riesgos reales a que está expuesto el fondo ¿cuál es el cuadro general de costes económicos asociado a la preservación digital de colecciones como la del CDPV?
- ~ Finalmente, en relación con la descripción de los documentos electrónicos y los datos mínimos que deben contener para poder garantizarse su preservación ¿con qué tipo de metadatos se debería contar? ¿Qué esquemas de metadatos se deberían usar? ¿Qué granularidad hay que aplicar a la descripción del fondo?

#### 1.5. METODOLOGÍA

El trabajo fin de máster que presentamos aborda el estudio de la preservación digital desde la perspectiva de las Ciencias de la Documentación, concretamente desde una perspectiva básicamente biblioteconómica.

---

<sup>1</sup> Véase la lista de abreviaturas.

Se ha optado por realizar un estudio de caso considerando que, aunque tal elección impone límites que impiden abarcar toda la extensión de la materia, aporta la ventaja de centrar el objeto de estudio. Está planteado desde un enfoque investigador mixto, combinando el análisis cuantitativo de datos y el análisis cualitativo de la materia que se estudia.

La investigación consta de varias fases: 1) análisis de la bibliografía; 2) trabajo de campo; 3) análisis de la información obtenida; 4) recopilación de informaciones comerciales sobre soluciones técnicas; y 5) realización de una propuesta fundamentada.

La investigación se inicia con un acercamiento a la cuestión mediante la lectura de los materiales obtenidos tras una primera revisión de la bibliografía. Esta primera fase permite identificar el espacio que ocupa la preservación digital y su incidencia en las actuales organizaciones, así como comprender los diversos enfoques con que se está abordando su estudio desde las instituciones de la memoria. Conforme a las características del caso de estudio, la búsqueda se centra en identificar publicaciones que traten la temática desde la perspectiva de la “digitalización con fines de preservación”. En especial, se han tratado de identificar textos que estudien la problemática de las colecciones hemerográficas contemporáneas digitalizadas con fines de conservación. Se ha puesto el punto de mira en textos sobre digitalización como medio de archivo definitivo, sobre las implicaciones en la gestión de colecciones sujetas a derechos de explotación en manos de terceros y sobre soluciones de almacenamiento no condicionadas por la accesibilidad en línea de los datos.

La fase posterior se centra en el caso objeto de la investigación. Se realiza el trabajo de campo consistente en verificar y describir el estado de los ficheros de imagen de la versión maestra de la colección analizada. Para ello se ha tenido acceso a las copias máster de la colección de ficheros en CD-ROM y en discos duros externos, realizándose cargas completas de estos últimos, y accesos por muestreo a los ficheros en CD-ROM. Los discos duros han sido verificados en su totalidad, los CD-ROM mediante muestreo. Para el cálculo de la muestra sobre el universo total de CD-ROM se ha utilizado la herramienta informática *STATS 2.0*, incluida en el manual de 2010 de Hernández Sampieri *Metodología de la investigación*.

Durante el trabajo de campo y en relación con el software utilizado, para el chequeo de los ficheros de imagen digital ráster se han empleado los visores y editores de imágenes Microsoft Office Picture Manager, Galería Fotográfica de Windows Vista, Visor de imágenes de Windows, Nero PhotoSnap, GIMP y Adobe Photoshop CSS6. El uso de diferentes herramientas ha variado según la disponibilidad de los ordenadores empleados en distintos momentos y lugares. Para la visualización de la información de metadatos relativa a los ficheros de imagen se ha usado principalmente ExifTool 9.30, pero también los visores y editores de imágenes arriba mencionados.

Tras el trabajo de campo se realiza un análisis y estudio de la información obtenida, contrastando los resultados de la verificación de los ficheros de imagen. Se comparan los parámetros de digitalización contratados con los datos realmente hallados en los ficheros. Se evalúan y detallan defectos de digitalización. Y se analizan los medios de almacenamiento en que el fondo se encuentra soportado, las condiciones en que se encuentra almacenado, así como las acciones de preservación acometidas hasta el momento.



Con la descripción del fondo a la vista, a partir del volumen de datos conocido y de sus características, se busca establecer soluciones técnicas de almacenamiento adaptadas al mismo, valorando pros y contras en relación con el estado del arte en preservación digital. El hecho de tener presentes las tablas de datos que caracterizan el fondo, permitirá plantear las alternativas técnicas de custodia digital existentes en el mercado que mejor se adaptan al mismo.

Así se da paso a la siguiente fase, donde se estudian propuestas económicas y presupuestos comerciales para tratar de estimar el coste económico de implementación de cada una de las soluciones técnicas. Se analizan las soluciones técnicas existentes en el mercado que ayuden a conservar a largo plazo el patrimonio acumulado. Estas soluciones técnicas serán observadas siguiendo la literatura científica y el estado del arte en preservación digital, atendiendo a la variable del coste económico de su implementación real, y conforme a los condicionantes legales que atañen a las bibliotecas en relación con las obras sujetas a derechos de propiedad intelectual, como es el caso de los periódicos en curso de publicación.

La fase final es la consistente en la redacción de una propuesta fundamentada resultante de todos los análisis anteriores. Tal propuesta se estructura en torno a tres pilares: las estrategias de preservación digital que se pueden llevar a cabo; la necesaria toma de decisiones en relación con los retos de la preservación digital; y la articulación de un plan de preservación digital.

## 1.6. FUENTES EMPLEADAS

Para la búsqueda y obtención de información se ha recurrido a fuentes bibliográficas, a información comercial y a las actas de reunión del Consorcio de Digitalización de Prensa Vasca. Se ha tenido acceso a los ficheros de imagen de la colección maestra, y para trabajar con ellos y evaluarlos adecuadamente, se han utilizado varias herramientas de software.

Para la localización de bibliografía se han lanzado búsquedas generales en *Scopus* y en *Web of Science* a través del servicio de recursos electrónicos de la biblioteca de la UC3M, así como en *Google Scholar* o *Google académico*, localizando y seleccionando monografías, artículos en revistas científicas, actas de congresos y otros documentos en bases de datos y en repositorios digitales.

En concreto, se ha recurrido a las bases de datos especializadas *LISA: Library and Information Science Abstracts (ProQuest)*, y *LISTA: Library, Information Science & Technology Abstracts (EBSCO)*, y repositorios institucionales como *E-Archivo*, (UC3M), *GRECOS* (Universidad de Salamanca), *ADDI* (Universidad del País Vasco/Euskal Herriko Unibertsitatea), y a otros sistemas como *DIALNET*, *HEDATUZ* (repositorio digital de ciencia y cultura vascas), e *INGUMA*, la base de datos de la comunidad científica en euskera.

Para la identificación de tesis doctorales en lengua española se consultó la *Base de Datos de Tesis Doctorales Teseo*, las catalanas en *TDR Tesis Doctorales en Red* y las de lengua vasca en *Tesis Doctorales en Euskera*.

Se consultaron varios sitios web de revistas, como la revista española *El Profesional de la Información*, los artículos de *ThinkEpi* y la colección de libros EPI & UOC, la revista *Educación* y

*Bibliotecas, Boletín de la Anabad, Cuadernos de Documentación Multimedia, Bid textos universitaris de biblioteconomía i documentació* y la revista *Mi Biblioteca*. Se han buscado también artículos de revisión, a fin de conocer estudios generales y recopilatorios, en *Annual Review of Information Science and Technology (ARIST)*.

Se lanzaron búsquedas sobre repositorios de biblioteconomía como *Travesía, recursos digitales para la cooperación bibliotecaria*, del Ministerio de Educación, Cultura y Deporte, y *E-prints in Library and Information Science (E-LIS)*.

La estrategia de búsqueda inicial (luego repetida en otras bases de datos y recursos informativos) para la localización de bibliografía consistió en una búsqueda booleana planteada en los siguientes términos: “digital preservation” OR “digital conservation” OR “preservación digital” OR “conservación digital”. Se ha evitado el uso de “digital curation” pues hace referencia a una visión más amplia de la disciplina.

Esta estrategia de búsqueda se precisó posteriormente como “*digital preservation*” AND (“*digitization*” OR “*digitisation*”) AND “*newspapers*”, o bien “preservación digital” Y “digitalización” Y (“prensa” O “diarios” O “periódicos”). El análisis de la bibliografía se ha realizado atendiendo a un límite temporal de unos diez años.

Se han comprobado las web personales de los profesores Miquel Térmens o Ciro Llueca, así como la web *PRESERVA* del grupo de investigación en preservación digital de la Universidad de Barcelona al cual ambos pertenecen. Se han cotejado en detalle los documentos de los diez *WORKSHOP's* de REBIUN sobre proyectos digitales, en especial el VIII celebrado en Murcia en 2008 y dedicado a la preservación digital.

Por otra parte, se han seguido las informaciones sobre preservación digital aparecidas en Twitter, y se ha recibido periódicamente la suscripción del boletín *Digital Preservation Newsletter* de la *Library of Congress*. Asimismo, se han seguido los trabajos de los grupos sobre preservación digital en la versión web de *Mendeley*.

Se han estudiado las entradas del epígrafe preservación digital en las versiones inglesa y catalana de la Wikipedia. En esta última se ha desarrollado durante el curso 2012-2013 un wikiproyecto denominado *Viquiprojecte: Preservació de recursos d'informació digital* coordinado por la *Universitat Oberta de Catalunya*, que recoge información actualizada sobre el tema.

La información comercial sobre sistemas de almacenamiento se ha extraído de catálogos, páginas web especializadas y sitios web de las marcas comerciales. Se han realizado búsquedas precisas en Google para identificar proveedores e información sobre medios de almacenamiento. La información técnica ha sido contrastada y ampliada con el apoyo de diccionarios y manuales de informática.

## 1.7. ALCANCE Y LÍMITES

En primer lugar se ha señalado que las características propias de un Trabajo Fin de Máster imponen un límite de extensión, amplitud y cobertura a la investigación.

En relación con la cobertura, hay que destacar que no se presentan recopilaciones de recursos y proyectos sobre preservación digital, puesto que ello no encaja con el enfoque de este trabajo y excede los límites del mismo. Existen bastantes grupos de investigación muy activos, así como bibliotecas nacionales, bibliotecas universitarias y archivos nacionales que mantienen sitios web con los resultados actualizados de sus investigaciones y experiencias. Se entiende sin embargo que su consulta puede resultar de interés, por ello se recomienda la guía *Preservación digital: Guía de recursos* (Grupo de trabajo de la Línea estratégica 2 de REBIUN, 2009) que da cumplida información de las organizaciones más activas y dinámicas. Esta guía se puede actualizar y completar con los balances y recopilaciones comentados en el punto 1.1. Además, las actas de la reciente Conferencia Internacional de la UNESCO sobre el acceso permanente al patrimonio documental digital (*The Memory of the World in the Digital Age: Digitization and Preservation*) detalla muchas de las iniciativas emprendidas en distintas partes del mundo.

También rebasa los límites de esta investigación el uso y verificación de distintas herramientas y propuestas de evaluación emanadas de proyectos sobre preservación digital. Así, se ha renunciado a probar herramientas de software, aplicaciones de tipo *checklist* que ayudan en la toma de decisiones y tutoriales elaborados en el transcurso de alguno de estos proyectos, que guían en la evaluación de la preservación de una colección concreta o en la realización de planes de preservación digital.

Tampoco se exponen los modelos teóricos que detallan a nivel técnico las características y funcionalidades que un sistema de preservación digital debe contener, puesto que hay mucha bibliografía que explica pormenorizadamente estos modelos, principalmente el *Reference Model for an Open Archival Information System (OAIS)*.

Hay que resaltar que tampoco se aborda todo lo relacionado con la preservación de los documentos originales, colecciones de periódicos en papel en este caso, a pesar de que la digitalización como medio de preservación de estos materiales está en el origen mismo de los proyectos de digitalización desarrollados.

La preservación digital puede ser abordada desde diversos enfoques, como se detalla más adelante. No obstante, en este trabajo se analiza exclusivamente en relación con la necesidad de preservar los ficheros de imagen que son el resultado del proceso de escaneado acometido sobre periódicos como medio para su conservación preventiva.

No todos los fondos culturales patrimoniales son objeto de un tratamiento idéntico en el momento de la digitalización y como consecuencia de ello se generan productos digitales diferenciados, que presentan sus propias especificidades en relación con la preservación digital. En este estudio se analizan únicamente los condicionantes que afectan a la preservación digital de prensa moderna digitalizada.

Pese a que para el adecuado análisis del caso de estudio se ha recurrido a la presentación de información sobre decisiones, iniciativas, debates y diversas vicisitudes acaecidas en el seno del CDPV, en ningún momento se pretende aportar una visión pormenorizada del recorrido histórico del Consorcio. Es así que quienes han conocido de cerca su funcionamiento a lo largo de estos años echarán en falta multitud de detalles e informaciones que rebasan los límites de este estudio.

Finalmente, se ha de comentar que existen copias derivadas de la versión maestra de los ficheros de imagen originales en cada una de las bibliotecas del Consorcio. Hay, por tanto, varias colecciones en distintos depósitos, más o menos acondicionados, y en distintas ubicaciones. Pero estas copias derivadas no son objeto de este estudio, sino que se limita a la versión maestra de los ficheros de imagen que actualmente se conserva en las dependencias de la Fundación Sancho el Sabio (FSS) en Vitoria-Gasteiz.

## ***1.8. ESTRUCTURA DEL TEXTO***

El contenido de este trabajo se presenta dividido en cinco partes además de la presente introducción, tal y como se expone a continuación.

En la primera parte se detalla el marco teórico y el contexto por el cual discurre la investigación. A partir de la revisión bibliográfica llevada a cabo, se exponen cuáles son las principales teorías y estrategias existentes, esto es, cuál es el estado del arte en preservación digital. Se resalta la necesidad de acometer iniciativas de preservación digital mediante la ejecución de estrategias proactivas de trabajo. Y se repasan brevemente las principales recomendaciones y guías a nivel internacional, a la luz de las cuales se contextualiza la propuesta de trabajo aportada.

La segunda parte aborda el caso de estudio que nos ocupa: la copia máster digital del Consorcio de Digitalización de Prensa Vasca. Se da noticia del recorrido histórico del consorcio, sus objetivos, integrantes, cabeceras digitalizadas, medios y procesos de trabajo. Y se describe detalladamente la colección de ficheros máster de imágenes ráster digitales de prensa diaria que el CDPV ha ido generando a lo largo de las dos últimas décadas.

En la tercera parte se enumeran las alternativas técnicas de almacenamiento seguro para la documentación electrónica con vistas a hacer preservación a largo plazo. Y también se analizan en este apartado las implicaciones económicas asociadas a la implantación efectiva de cada una de las alternativas que se estudian.

En la cuarta parte se realiza una propuesta fundamentada, entendida como una serie de recomendaciones sobre las mejores alternativas disponibles para el consorcio desde los diversos puntos de vista analizados (técnico de almacenamiento, técnico descriptivo, legal y económico). Se detallan los esquemas mínimos de descripción mediante metadatos que se deberían emplear para garantizar la accesibilidad de los datos a largo plazo. Se plantean los requisitos legales mínimos que se han de cumplir. Finalmente, se presentan las principales conclusiones extraídas de la investigación realizada.

## 2. CONTEXTO Y MARCO TEÓRICO: DIGITALIZAR PARA PRESERVAR Y PRESERVAR LO DIGITALIZADO

---

Existe una producción bibliográfica muy cuantiosa sobre preservación digital, por lo cual su seguimiento y examen resulta complejo y laborioso. Esta literatura surge principalmente en torno a proyectos participados por bibliotecas universitarias, grandes bibliotecas y archivos nacionales, y grupos de investigación mixtos y multidisciplinares. La perspectiva con que los diferentes estudios publicados abordan la preservación digital gira alrededor de temas tan diversos y amplios como:

- ~ las revistas electrónicas,
- ~ la producción científico-técnica,
- ~ los grandes proyectos patrimoniales de colecciones de composición heterogénea,
- ~ la tramitación de expedientes administrativos electrónicos y administración electrónica en general,
- ~ los conjuntos de datos de investigación,
- ~ los repositorios digitales y los repositorios digitales de confianza,
- ~ los estándares y esquemas de metadatos,
- ~ el software y las herramientas para la preservación digital,
- ~ el archivo de la web,
- ~ etcétera.

Esta abundante bibliografía sobre preservación digital, procedente de muy diversas áreas de conocimiento, indica que el tema de la conservación de los recursos digitales es objeto de interés multidisciplinar, transversal y generalizado. En Europa, sin ir más lejos, la investigación en preservación digital ha sido abordada desde diversos proyectos de las convocatorias del quinto, sexto y séptimo Programa Marco de la Comisión Europea (Térmens, 2009a). Las universidades, centros de investigación, agencias estatales, empresas privadas y administraciones públicas europeas que han tomado parte en las distintas convocatorias proceden de sectores tan diversos como el aeronáutico y aeroespacial, la ingeniería informática, los grupos editoriales, los productores de bases de datos científicas (*datasets* científicos sobre física, biología, ingeniería, agricultura, química, medicina...), los archivos nacionales, las bibliotecas nacionales o las bibliotecas universitarias.

La literatura científica del campo de las Ciencias de la Documentación viene ocupándose de forma notable en las últimas décadas de la irrupción de “lo digital” en las instituciones de la memoria. Valga como ejemplo la recopilación bibliográfica publicada por Charles W. Bailey en 2010, y que actualiza anualmente.

La preservación digital es uno de los aspectos que sistemáticamente aborda la producción en nuestra área de conocimiento. Sin embargo, hay pocos trabajos publicados sobre experiencias

relacionadas con los retos de la preservación digital afrontados por bibliotecas con proyectos digitales pequeños o medianos relativos a patrimonio local, como es el caso que nos ocupa.

## **2.1. PRESERVAR LA PRENSA DIARIA**

Las instituciones de la memoria tienen como misión general promover y conservar el patrimonio cultural y los recursos de información de la comunidad en la que se asientan. La prensa en general, y los periódicos en particular, son un vehículo de transmisión de información de actualidad que refleja, a través de los acontecimientos principales, el devenir de la sociedad. Conservar los diarios para la posteridad entronca directamente con la misión y objetivos de estas instituciones.

Las bibliotecas, principalmente, se han ocupado de ofertar diarios entre sus recursos de información. Además de posibilitar la lectura de las noticias de actualidad, esta práctica ha permitido conformar colecciones hemerográficas para ofrecer la consulta de información retrospectiva. En la medida en que estas colecciones constituyen una fuente de primera mano para el conocimiento y el estudio del pasado, es primordial poder garantizar su accesibilidad futura.

Para garantizar la permanencia y el acceso a las colecciones, toda iniciativa de conservación de fondos hemerográficos contemporáneos ha de abordar inexorablemente varias cuestiones clave, como son: el problema del espacio de almacenamiento requerido para guardar los diarios; la fragilidad del soporte en que están impresos; y, en el caso de convertir o traspasar los periódicos a otro soporte o medio, la preocupación por garantizar el acceso a la información en su forma más íntegra y fiel posible.

Las colecciones de diarios en papel son, por su naturaleza, de crecimiento rápido y continuado: son de aparición diaria, se editan en gran formato y, normalmente, existen varias cabeceras de interés para una misma biblioteca. Ello genera, obviamente, un problema de gestión del espacio, por el rápido llenado de los depósitos. La custodia de las colecciones en su soporte original supone, por tanto, la reserva y el consumo de muchos metros lineales de estanterías, provocando la ocupación de un gran volumen de metros cúbicos en las instalaciones.

El segundo problema deriva de la fragilidad del papel utilizado para la edición de los diarios. La baja calidad del mismo provoca el rápido surgimiento de altos índices de acidez, con el consiguiente amarilleamiento del color, pérdida de flexibilidad y aparición de rigidez dándose, en consecuencia, mayor riesgo de roturas y rasgados. En casos graves de alta acidez, se puede llegar a la desintegración del soporte. Las tintas de los diarios, asimismo, suelen ser de calidades inferiores y pueden presentar problemas de desintegración o borrado y desaparición parcial o total y, en ocasiones, su composición química provoca la aceleración de la aparición de la acidez.

La tercera cuestión tiene que ver con el acceso a la información de la manera más fiel posible al original. Esto se viene resolviendo a través de la reproducción en soportes con mayor permanencia teórica. Así, desde los años 50 del siglo XX se apuesta por la microfilmación, y desde los años 90 del mismo siglo por la digitalización. Como se verá a lo largo de este informe y ha sido detallado en varios estudios, a día de hoy recurrir a la digitalización no parece resolver totalmente el acceso permanente y garantizado a la información a largo plazo.

En cualquier caso, el recurso a la digitalización de colecciones hemerográficas ha sido una constante en los últimos años. Se ha de distinguir, no obstante, entre iniciativas llevadas a cabo sobre fondos de prensa histórica e iniciativas acometidas sobre fondos de prensa actual. En principio, y en relación con la preservación de los objetos digitales generados tras la digitalización, ambos casos presentan idéntica problemática. Sin embargo, los proyectos realizados sobre fondos de prensa histórica aprovechan plenamente la gran versatilidad y capacidad de los medios electrónicos para su difusión telemática en abierto. Es debido a que no existen barreras legales que impidan la puesta a disposición para la comunidad de usuarios, de copias digitales en dominio público. Pero, dado que para la comunicación pública de la prensa actual necesariamente hay que contar con el acuerdo y permiso de los propietarios de los derechos de explotación, la gestión de su puesta a disposición se hace más compleja y, frecuentemente, impide la difusión abierta de los fondos digitalizados.

A pesar de esta barrera (infranqueable en muchos casos por los intereses comerciales legítimos de las editoras de los diarios), y dejando de lado los muchos casos de éxito de proyectos de prensa histórica, merece la pena destacar algunas iniciativas de nuestro entorno realizadas con prensa actual.

La Biblioteca Nacional de España (BNE) tiene sus primeras experiencias de digitalización precisamente en la Hemeroteca Nacional donde “comenzó la digitalización de fondos modernos en 1999, pasando a CD-ROM los periódicos que recibía cada día” (Arquero Avilés y García-Ochoa Roldán, 2005, p. 85). Hasta el año 2003, los programas de digitalización en la BNE se realizaron exclusivamente sobre prensa actual, comenzando entonces el proceso de digitalización de los “Tesoros de la Biblioteca Nacional”. Anteriormente, solo había habido colaboraciones puntuales con la Universidad de Alicante, la Biblioteca de Valencia y Red.es (Salvador López, 2005). En el año 2000 la BNE digitalizaba 108 periódicos, las cabeceras más importantes de España, habiendo comenzado su programa de conservación mediante microfilmación en 1991. La BNE calcula su fondo de diarios actuales digitalizados en unos 200 títulos, no cotejados en control de calidad, por falta de medios de personal para hacerlo.

Esta prensa digitalizada no se difunde en acceso abierto a través de internet, sin embargo es consultable en la sala de prensa y revistas de la sede de la BNE en Madrid mediante petición anticipada. Según la información que se desprende de su propio catálogo, la BNE dispone de cuatro copias de cada cabecera:

- *Original en papel.* Ejemplar de conservación, localizado en la sede de Alcalá, no se sirve al público. Es el ejemplar reproducido mediante digitalización y microfotografía.
- *Reproducción microfotográfica en carrete de microfilme.* Ejemplar de conservación, localizado en la sede de Alcalá, no se sirve al público. Generado a partir del ejemplar en papel.
- *Reproducción digital en discos compactos.* Ejemplar de conservación, localizado en la sede de Alcalá, no se sirve al público. Generado a partir del ejemplar en papel.
- *Reproducción digital en discos compactos (segunda copia).* Ejemplar de consulta, localizado en la sede de Madrid, se sirve al público tras petición anticipada.

La situación descrita no es homogénea para todos los diarios. En algunos casos se dispone de otra copia de microfilm para la consulta pública, y en otros casos sólo aparecen en el catálogo tres copias (papel, microfilm y digital), no siendo ninguna de ellas consultable.

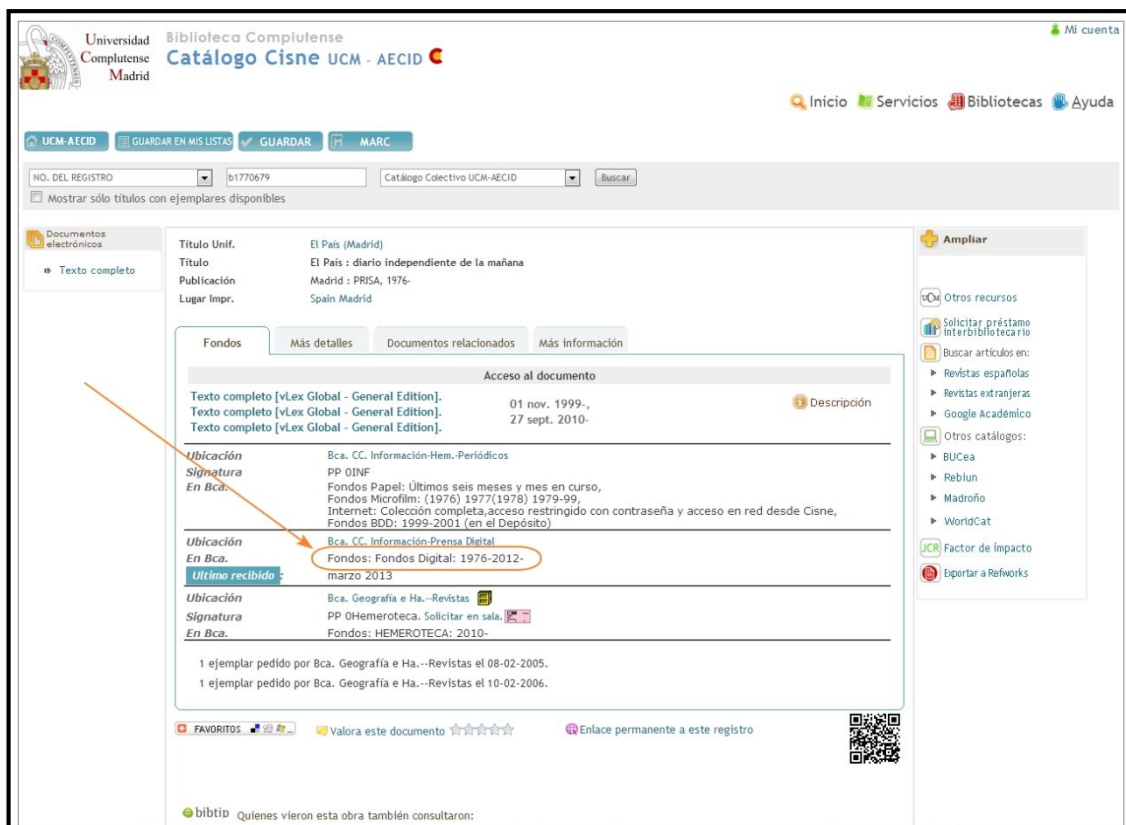


Figura 1.- Colección digitalizada del diario El País en el catálogo de la biblioteca de la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid.

La Hemeroteca Electrónica de la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid ha conformado una colección digital de prensa actual a partir de la digitalización de 62 cabeceras desde el año 2000, completada con fondos retrospectivos cuando han sido comercializados por las editoras de los diarios. Se trata de una colección en CD-ROM de acceso y consulta local (Arquero Avilés y García-Ochoa Roldán, 2005).

Aparte de los dos grandes proyectos mencionados hay otras iniciativas, como la de la *Hemeroteca del Ayuntamiento de Madrid*, que aunque se han dedicado principalmente a la prensa histórica no han descuidado la prensa actual de sus respectivas áreas de influencia.

Merece especial atención el proyecto *Jable: Archivo de Prensa Digital de la Universidad de Las Palmas de Gran Canaria*, que recoge toda la prensa viva canaria. Esta hemeroteca digital ha sido capaz de establecer un modelo mixto para ofertar la colección a sus usuarios, a través de consulta en sala, acceso restringido mediante intranet y acceso abierto en Internet, dependiendo de la gestión de derechos en cada caso (Macías Alemán y Torres Betancor, 2005).

A nivel internacional, cabe mencionar el proyecto *Europeana Newspapers* el cual recopila hemerotecas digitalizadas de bibliotecas públicas, nacionales y de investigación europeas. Varias de ellas han digitalizado prensa actual, encontrándose con las mencionadas barreras que la legislación de propiedad intelectual impone. En las conclusiones de una reciente



encuesta completada por 47 de estas bibliotecas, se afirma que proveer acceso a los contenidos de diarios del siglo XX sigue siendo problemático para las bibliotecas (Dunning, 2012).

En otro sentido, cabe mencionar que la gran mayoría de los diarios vascos guardan versiones de sus diarios impresos, algunas de las cuales están digitalizadas, otros diarios conservan la versión digital que sale a papel. *El Correo* está completamente digitalizado, con OCR y guardado en formato PDF indexado (Agirreazaldegui, Ronco y Camacho, 2009; Martín, 2009; Oleaga, 2009). *Berria* conserva las páginas del diario en formato PDF tal cual salieron a la luz mediante la conservación de la documentación nacida digital en el proceso productivo, evitando así tener que digitalizar a posteriori la edición papel (Eizagirre y Barandiaran, 2009). *Diario de Noticias*, *Diario de Noticias de Álava* y *Diario de Noticias de Gipuzkoa*, han sido digitalizados y pasados a formato PDF (Muñoz y Aldaz, 2009). *Deia* también se ha digitalizado, así como el Diario de Navarra y Gara (Agirreazaldegui et al., 2009).

Sin embargo, según un reciente estudio realizado con los medios de prensa local de la Comunidad de Madrid, alerta de que el 25% de ellos “han llegado a realizar expurgos masivos, eliminando de sus fondos contenidos publicados por falta de espacio o de presupuesto para su almacenamiento”. Otros medios conservan y almacenan el papel, y el 50% de ellos almacenan versiones digitalizadas en formato PDF (Rodríguez Mateos, Hernández Pérez, y Méndez Rodríguez, 2012, p. 20). Sin embargo, “en muchos casos, no han planteado al mismo tiempo estrategias mínimas de conservación de esos contenidos digitales” (Rodríguez Mateos et al., 2012, p. 24). En este escenario, es previsible que en pocos años sea difícil recuperar copias de estos diarios.

## 2.2. UN TÉRMINO NO EXENTO DE PROBLEMÁTICA

El término preservación digital necesita de una aclaración previa, dado que ha venido utilizándose indistintamente para hacer referencia a dos cuestiones bien diferentes. Por un lado, referido al proceso de digitalización de documentos en soporte analógico para asegurar su conservación en el tiempo evitando el deterioro del original por el uso y la manipulación; por otro lado, referido a la preservación de objetos digitales para asegurar su integridad y accesibilidad futura (Ferrerías, 2010).

La primera acepción surge cuando se multiplican los proyectos de digitalización de fondos bibliográficos y documentales, entre cuyos objetivos está la salvaguarda de los soportes analógicos, comenzándose entonces a hablar de preservación por medios digitales o preservación digital. En realidad, se trataba de hacer digitalización como medio de conservación preventiva. Posteriormente, surge en las organizaciones la necesidad de preservar tanto los *outputs* emanados de tales procesos de transformación como los documentos nacidos digitales (*born-digital*), y se utiliza nuevamente el término preservación digital.

Hoy día, no se recomienda el uso del término para la acepción primera de “digitalizar para conservar”, aunque existe todavía gran confusión y los usos referidos a ambas acepciones aparecen frecuentemente en la literatura. The National Archives (Reino Unido), una de las organizaciones más dinámicas en este ámbito, recomienda evitar su uso en este sentido,

sustituyéndolo simple y llanamente por *digitisation* (“Digital preservation FAQs”, 2013). En la misma línea, tanto Galloway (2005) como más recientemente Conway (2010) recomiendan utilizar el término preservación digital referido solamente a la preservación de los objetos digitales, y proponen para el otro uso los términos *digitization for preservation*, *digital reformatting*, *preservation reformatting* o, simplemente, *digitisation* (grafía británica) o *digitization* (grafía estadounidense).

A lo largo de este trabajo, se siguen los citados usos y se utilizan los términos “digitalización” y “preservación digital” en los sentidos arriba mencionados y hoy día ampliamente recomendados.

Con ánimo de acotar un poco más la terminología, cabe apuntar que otro concepto relacionado con la preservación digital es el de curación digital. Este término (del inglés *digital curation*, a veces también *data curation* o “preservación de datos de investigación”, en castellano) es más amplio que el de preservación digital, puesto que hace referencia al ciclo de vida completo del documento digital desde su generación (Pérez Cervera, 2012), atendiéndose a la selección, recopilación, archivo, conservación y mantenimiento de activos digitales en repositorios digitales. Se entiende por tanto, que la curación digital es un proceso más amplio que engloba entre otras tareas las propias de la preservación digital.

### 2.2.1. DEFINICIÓN DE PRESERVACIÓN

Pero, ¿qué es preservar? De acuerdo con la RAE, preservar significa “proteger, resguardar anticipadamente a una persona, animal o cosa, de algún daño o peligro”. Desde el punto de vista de la ética profesional, Conway (2010), citando a Barr (1946), afirma que preservar es un acto de custodia responsable. Ambas definiciones encajan con la misión de las instituciones de la memoria y con los principios éticos y deontológicos de la profesión, pero hace falta mayor precisión.

La conservación del patrimonio bibliográfico, documental y cultural en general, se ha entendido tradicionalmente como un conjunto de medidas y operaciones destinadas a prolongar la vida del bien a proteger. La investigación y la praxis en las disciplinas en torno al patrimonio cultural consideran las labores de custodia y salvaguarda de los bienes como un proceso de gestión integral. Así, y a pesar de cierta disparidad de definiciones (véanse por ejemplo, las aportadas por Ballart Hernández, 2007; Soler, 2008; o Zubiaur Carreño, 2004), la conservación engloba el examen del bien, su preservación o conservación preventiva y la restauración (Cruz Mundet, 2011). La preservación se engloba entre las actividades de conservación orientadas a proteger proactivamente los bienes, tratando de evitar las intervenciones restauradoras o métodos intrusivos de conservación reactiva.

A fin de aportar mayor precisión y al objeto de establecer un marco de trabajo claro, se siguen a continuación las definiciones aportadas por la Directora del Departamento de Preservación y Conservación de la BNE, quien a su vez sigue a Arsenio Sánchez (Salvador López, 2005, pp. 191–193).

- La *preservación* “abarca todas aquellas actividades encaminadas a mantener en óptimas condiciones las colecciones documentales y sus contenidos intelectuales, independientemente del soporte en el que se encuentren almacenadas, con el fin de

asegurar su permanencia tanto como se considere necesario y facilitar su acceso". Y entre tales actividades se enumeran "la gerencia administrativa y económica, el acondicionamiento de los locales de trabajo y de almacenamiento, la provisión de personal y la adopción de políticas, técnicas y métodos destinados a salvaguardar el contenido".

- La *conservación* es "una actividad encuadrada dentro del término más general de preservación". Engloba tanto las medidas preventivas como los procesos de reparación de materiales dañados. Entre las medidas preventivas cabe mencionar la conservación de la información mediante su reproducción por cualquier medio a fin de evitar el uso del original.
- La *restauración* son los "trabajos encaminados a la recuperación de las condiciones originales del libro o documento, una vez que se ha producido un daño", y entra dentro del campo de la conservación de los materiales físicos.
- La *conservación preventiva* "tiene como finalidad frenar el impacto negativo del medio ambiente en las colecciones". Se ocupa del control de los factores ambientales, le evaluación de edificios, planificación de siniestros, protección, limpieza y formación de trabajadores y usuarios.

Queda claro que la preservación es un término más genérico. También se observa una mayor frecuencia de aparición en la literatura científica. Pero en cualquier caso, a lo largo de este trabajo se usarán indistintamente los términos preservación y conservación a fin de amenizar la lectura y de evitar una redacción excesivamente reiterativa.

### 2.2.2. DEFINICIÓN DE PRESERVACIÓN DIGITAL

Los documentos digitales están expuestos a un alto riesgo de ilegibilidad, debido a que están sometidos a procesos imparables de caducidad derivados de la llamada obsolescencia tecnológica, la cual es fruto de la constante evolución de los medios tecnológicos y de su mercado y comercialización. Hay autores que distinguen entre obsolescencia lógica, o de los formatos de codificación de la información; y obsolescencia física de los medios o soportes de almacenamiento de los datos y de los dispositivos o unidades de lectura y escritura, en cuanto que están afectados por el deterioro y la degradación física inherente a todos los soportes materiales. Otros autores llegan a las mismas conclusiones pero plantean el asunto de otra forma y distinguen tres aspectos: la fragilidad de los medios de almacenamiento; la obsolescencia del hardware para la lectura; y la obsolescencia del software y de los formatos.

En cualquier caso, queda claro que en relación con la conservación del documento digital, se plantean los retos que el documento *Preservación de los documentos digitales: Guía para comenzar. ISO/TC 46/SC 11* (Elvira, 2011) distingue como:

- *Obsolescencia y degradación del formato físico (soporte)*. Los soportes son susceptibles de daño por almacenamiento o manipulación inadecuados; y asimismo, la rápida evolución de la tecnología comporta que los soportes de almacenaje queden obsoletos en un periodo corto de tiempo.

- *Obsolescencia del formato del fichero.* Los datos se representan mediante software, y estos sólo pueden leer o interpretar los datos que están en un formato específico de fichero. Las actualizaciones constantes para añadir mejoras y nuevas funcionalidades hacen que los formatos más antiguos corran el riesgo de no poder ser leídos con los nuevos sistemas de software.
- *Obsolescencia del software.* Los programas de software también se vuelven obsoletos, inutilizando así los ficheros codificados en el formato interpretable por tal software. Ello afecta tanto a los programas de aplicación como a los sistemas operativos.
- *Obsolescencia del hardware.* Los medios de lectura están también sujetos a daño físico y a obsolescencia tecnológica por el desarrollo de nuevos medios y el abandono del mantenimiento de anteriores sistemas.

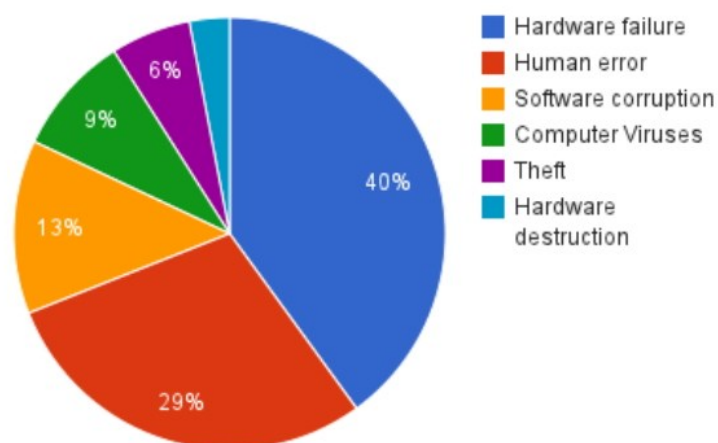


Figura 2.- Razones de la pérdida de datos (Fuente: Henriksen, Seuskens y Wijers, 2013, citando The insurance Agency Inc., 2001 y Ontrack Data International, inc., 2003)

Frente a estos problemas, la preservación digital puede ser entendida como un conjunto de procesos y actividades encaminados a asegurar el acceso continuado a la información existente en formato digital, tratando de anticiparse al momento en el que la obsolescencia haya afectado al registro electrónico hasta el punto de hacerlo inaccesible.

La gestión de los activos digitales de cara a su preservación, en comparación con la gestión de los activos analógicos, requiere de una mayor y constante atención. En relación con la documentación digital, la mera conservación de los soportes no implica la adecuada accesibilidad, legibilidad o integridad de la información en el futuro. Como más adelante se detallará se han de implementar unas estrategias de acción programadas tendentes a evitar la obsolescencia tecnológica. En este sentido, la gestión asociada a la preservación digital presenta mayor complejidad así como notables novedades en relación con la preservación documental en soportes tradicionales.

Según las *Directrices para la preservación del patrimonio digital* es el “conjunto de acciones destinadas a mantener la accesibilidad de los objetos digitales a largo plazo” (UNESCO, 2003a). La Digital Preservation Coalition, por su parte, define la preservación digital como “la gestión de las actividades necesarias para garantizar el acceso indefinido a los materiales digitales, tratando de evitar fallos en los soportes y afrontando los cambios tecnológicos”.

Gallart y Keefer (2007) la caracterizan como una actuación proactiva y continuada, y Lluca y Reoyo (2012; Lluca, 2012a) básicamente como una estrategia de acción cuyo fin último es asegurar el acceso a los bienes digitales patrimoniales. Estos autores llaman la atención sobre la naturaleza de la preservación digital fundamentada en una metodología de trabajo activa y transformadora, frente a la conservación tradicional donde el método pasivo suele acarrear buenos resultados. También la British Library entiende la preservación digital como “la serie de acciones e intervenciones necesarias para garantizar el acceso continuo y de confianza a objetos digitales auténticos durante el tiempo que se consideren de valor” (British Library, 2013, p. 17).

En este sentido, se han propuesto diversas estrategias de acción, habiendo gran consenso en las medidas básicas de actuación, y algunos modelos teóricos, aunque básicamente se viene aportando ciencia desde la praxis de la investigación aplicada. Tales estrategias están esencialmente orientadas a evitar daños, pérdidas o alteraciones, a facilitar el acceso permanente, y a actualizar formatos y soportes para asegurar la integridad y el acceso a los documentos (Lluca y Reoyo, 2012). También hay autores que proponen estrategias de acción más cercanas al uso que a la conservación del valor intrínseco del objeto, primando el acceso al contenido informativo del documento (Conway, 2010). Esta última visión entronca con la línea de los procesos de digitalización abordados desde una perspectiva reprográfica y no desde una perspectiva patrimonial, enfoque adoptado por algunos proyectos de digitalización de prensa actual.

Gran número de autores afirman que los actuales retos de la preservación digital no se encuentran en un nivel técnico, sino en los niveles organizativo o institucional (asunción de responsabilidades en preservación y compromiso de hacerlo), económico (cómo se sustenta económicamente la preservación a lo largo del tiempo), y legal (cómo reproducir, reformatear y poner a disposición si no se tiene el permiso del titular) (Keefer, 2007).

Debido a la existencia de tales retos no resueltos, hay quien afirma que la preservación digital está en un estado de beta continuo (LeFurgy, 2013), en el sentido de que no existe el sistema de preservación digital perfecto (Conway, 2010; Keefer y Gallart, 2007; Lluca y Reoyo, 2012) y que además es muy posible que no lo veamos en los próximos años.

También se ha afirmado que la preservación digital es una bomba de relojería para las bibliotecas digitales (Keefer y Gallart, 2003), expresión no carente de cierto alarmismo e incluso humor, pero que ilustra a la perfección la existencia de una necesidad perentoria de acometer el problema antes o después en el tiempo.

Finalmente, y justamente en relación con el tiempo y los plazos en los que se aporta garantía de acceso a los datos electrónicos, se habla de preservación digital a largo plazo, preservación digital a medio plazo y preservación digital a corto plazo.

Según Soler, la preservación a largo plazo “implica la provisión de acceso indefinido a los documentos” (2008, p. 28). Ortiz Ancona añade que es el acceso continuado a los documentos “o por lo menos a la información contenida en estos indefinidamente” (2012, p. 2). Ambos autores están de acuerdo en que el medio plazo es un periodo definido de tiempo, en el cual se pueden superar los cambios tecnológicos acaecidos en ese tiempo. Por último, la preservación digital a corto plazo “implica que el acceso a los documentos se mantiene [sólo]

hasta que los cambios tecnológicos los hacen menos accesibles o inaccesibles totalmente” (Soler, 2008, p. 28).

El proyecto *Digital Preservation Europe* cita diferentes fuentes y autores para no establecer claramente qué se entiende por largo plazo. Así, lo enmarcaba en plazos tan variables como “cinco años o más”, “no menor de diez años para todos los proyectos y mayor de veinte años para proyectos clínicos o de importancia social, medio-ambiental, o patrimonial”, o “preferiblemente permanente”. Lluca, citando a Waller y Sharpe (2006), fija el largo plazo en objetos digitales en veinte años (2012a).

### 2.3. LA PRESERVACIÓN DIGITAL A ESTUDIO EN DIFERENTES ÁMBITOS

El estudio de la preservación digital es abordado desde, al menos, cuatro puntos de vista, que tienen que ver con áreas de trabajo diferenciadas. De una parte, está la comunidad que se interesa por la preservación de la documentación electrónica producida en el ámbito de la administración pública y de las organizaciones. Se trata de documentos digitales no servidos en red pública, o distribuidos en entornos telemáticos cerrados de intranets o redes corporativas. Esta corriente de acercamiento a la cuestión se da en el ámbito de la archivística y el *records management*. Aborda aspectos como la fiabilidad, la autenticidad, la integridad, la identidad y la capacidad de uso y acceso a lo largo de tiempo (Soler, 2008). Su objeto de estudio es tanto la documentación nacida digital como la documentación digitalizada y validada electrónicamente para su uso administrativo, técnico y legal. En este contexto, cobra importancia el estudio de los flujos de información electrónica, la tramitación telemática, la firma digital, la marca de tiempo (*data stamp*), la encriptación de la información, los niveles de seguridad de acceso a los datos y, lógicamente, la propia preservación de la documentación digital. En el ámbito de las Ciencias de la Documentación varios autores vienen abordando el problema de la conservación de los documentos digitales desde una perspectiva fundamentalmente archivística como Serra, Soler, Alberch o Torres Freixinet y, en menor medida, Raventós, Navarro Bonilla y Térmens. Este último autor ha estudiado las diferencias y similitudes de enfoque entre archivos y bibliotecas con respecto a la preservación digital (Térmens, 2009b).

En otro contexto, se analiza la preservación de los documentos digitales en la web y el ciberespacio. En este ámbito, la documentación nace y permanece en un contexto digital de red, en entornos telemáticos abiertos. Los contenidos son publicados *online*. Piénsese en webs, blogs, colecciones y bases de datos abiertas, galerías de imágenes, radiodifusión y televisión por Internet, y, en general, en cualquier forma de difusión de contenidos a través de la red de redes. Es reseñable la preocupación existente por la salvaguarda de los contenidos publicados por los medios de difusión pública, la prensa digital, las webs institucionales que contienen grandes volúmenes documentales (*big data*), y, en general, por la permanencia de cualquier reflejo o rastro de la cultura ciudadana actual. En esta línea de trabajo, se vienen acometiendo acciones sistemáticas de preservación mediante el copiado de la información, principalmente de las páginas web, en repositorios digitales llamados habitualmente *archivos web* (Lluca y Cócera, 2012).

Por otra parte, en el entorno universitario e investigador destacan grandes esfuerzos por salvaguardar la documentación científico-técnica mediante el uso de repositorios digitales de

acceso abierto, conforme al modelo OAIS principalmente, y mediante infraestructuras tecnológicas y estrategias de gestión para el aseguramiento de grandes volúmenes de datos. Se está trabajando en preservación digital relacionada con documentación variada: *datasets* y conjuntos de datos de investigación; artículos y publicaciones de revistas científicas electrónicas de grupos editoriales, universidades y centros de investigación; y, en general, con toda publicación emanada del proceso investigador (*papers*; informes de proyectos; resultados de experimentos; ensayos; etcétera). Estos repositorios abiertos recogen documentación digital de cuatro grandes grupos: patrimonio institucional digitalizado, materiales docentes, producción investigadora y documentación de gestión y archivo de la organización. Las bibliotecas universitarias y de investigación están jugando un papel muy destacado, afrontando con dinamismo retos como la gestión de las colecciones digitales contratadas a terceras partes, y la integración del acceso a materiales muy diversos en sus catálogos.

La cuarta perspectiva de acercamiento al concepto de preservación digital se da en torno a la preocupación por la conservación de la documentación patrimonial no digital con valor cultural e histórico. El concepto *digitization for preservation* hace referencia a la digitalización como método preventivo para la conservación. En principio de modo generalizado, se recurrió a la digitalización como estrategia para minimizar el uso de la documentación en soportes tradicionales como papel, pergamino, papiro, imagen y sonido analógicos, etcétera (Ferrerías, 2010; Gómez, Martínez y Zubiaga, 2012; Gómez, 1995, 1997; Keefer y Gallart, 2003, 2007; Moreno, 2002; y muchos otros, dado que sobre este asunto hay abundantísima bibliografía). Pero pronto se observó la necesidad de interesarse también por la preservación de la propia copia digital generada (Keefer y Gallart, 2007). La conversión digital de documentos analógicos ha perseguido desde sus comienzos facilitar la difusión, proteger las copias originales evitando su uso y preservar el valor informativo de los mismos<sup>2</sup>.

En esta investigación, se aborda la preservación digital desde el último de los cuatro puntos de vista: exclusivamente en el contexto de los proyectos llevados a cabo con el fin de “digitalizar para preservar”, y más concretamente, en relación con la necesidad sobrevenida de preservar la copia digital generada.

#### 2.4. ¿DIGITALIZAR PARA PRESERVAR?

Digitalizar es convertir documentos en soporte no eléctrico a un formato digital procesable por un ordenador. La digitalización mediante imagen digital ráster (también llamada trama, mapa de bits o bitmap), que es una imagen compuesta por píxeles, permite conseguir imágenes muy cercanas a la apariencia de los documentos originales, lográndose una vista cuasi-facsimilar. Los documentos electrónicos presentan un nivel físico (soporte o medio de almacenamiento) y un nivel lógico (formato de datos) y para su interpretación es preciso una unidad o dispositivo de lectura y programas decodificadores o que interpreten los datos.

---

<sup>2</sup> También se han acometido proyectos de digitalización en los cuales se busca ganar espacio físico de almacenamiento, garantizando el acceso a la información original, pero evitando guardar los documentos digitalizados. En este caso se trata de organizaciones para las cuales no existe la preocupación por el patrimonio documental sino que están orientadas a la producción.

Si bien es cierto que muchas bibliotecas y archivos han aprovechado ampliamente las tecnologías para la reproducción digital, no es menos cierto que la selección de los documentos a digitalizar ha de obedecer rigurosamente a criterios claros si se lleva a cabo con fines de preservación. Esta selección documental para digitalizar fondos con fines de preservación debe priorizar:

- materiales frágiles, o con riesgo de deterioro por manipulación, y alta demanda,
- y, materiales libres de derechos de autor, en dominio público, o de los cuales se posea la cesión de los derechos de explotación y uso.

Las ventajas que aporta la digitalización, como son la evitación o reducción del uso del original y la potencial accesibilidad universal, desaparecen o se reducen drásticamente en caso de incumplirse alguna de estas premisas. Está claro que, en la selección de documentos a digitalizar, si nos encontramos frente a materiales sin riesgo de pérdida por manipulación o con baja demanda, disponer de una copia digital no presenta gran ventaja de cara a la conservación. Asimismo, si no se dispone de los permisos de uso para la comunicación pública, la difusión, la transformación, etc., nuevamente las ventajas que aporta el documento digital se ven notablemente mermadas. ¿De qué sirve disponer de una copia digital que apenas se demanda y que solo se puede poner a disposición de los usuarios de forma limitada?

Como contrapartida aparece además una nueva problemática asociada a la preservación, al tener que gestionar los artefactos electrónicos resultantes del proceso de digitalización. Esta gestión presenta una complejidad mayor y unos costes económicos añadidos.

Por tanto, la digitalización aporta su principal beneficio si se va a facilitar difusión y acceso abierto. En caso contrario, la conservación preventiva tradicional resulta más garantista y menos costosa.

Desde mediados los años noventa del siglo XX, las instituciones de la memoria de los países tecnológicamente desarrollados vienen acometiendo programas de digitalización de sus fondos patrimoniales de forma sistemática y masiva. Tras los éxitos alcanzados por las iniciativas pioneras y en la medida en que la tecnología progresa y los costes asociados a ella disminuyen, muchas bibliotecas y archivos comienzan a abordar en esas fechas proyectos de digitalización. La conversión digital de documentos analógicos persigue facilitar la difusión y proteger las copias originales evitando su uso. Como resultado de este proceso, surge la necesidad de custodiar y conservar la copia digital generada.

Es así que, la conversión masiva de documentos analógicos a formatos digitales, junto con la fuerte expansión de los documentos nacidos digitales, provoca que diversas organizaciones comiencen a estudiar la problemática asociada a la preservación de los recursos digitales desde mediados de la década de 1990 (Keefer y Gallart, 2003; Térmens, 2009a, 2010a).

Sin embargo, la conversión digital con fines de conservación continúa presentando varios problemas. Así, hoy día, para la salvaguarda del patrimonio bibliográfico y documental, bastantes autores siguen primando la microfilmación sobre la digitalización como medio de conservación preventiva, dado que su gestión presenta menor problemática y ofrece mayor estabilidad en el tiempo (Alberch, 2001; Guallar y Abadal, 2009, 2010; IFLA et al., 2002; Pujari y Pai, 2005; Salvador López, 2005).



La Jefa del Área de Preservación y Conservación de Fondos de la Biblioteca Nacional se mostraba sumamente cautelosa en Vitoria-Gasteiz en diciembre de 2003, en relación con el uso del soporte digital como medio de preservación a largo plazo, considerándolo “exclusivamente como vía de difusión y acceso” (Salvador López, 2005, p. 210).

Arquero y García-Ochoa consideran que “las tres ventajas fundamentales por las que el uso de microformas se aconseja para la prensa son: el ahorro de espacio, la duración y la seguridad” (2005, p. 76), y aunque afirman que la digitalización es “de gran utilidad no solo para conservar la información y almacenarla, sino también para difundirla” (2005, p. 78), posteriormente no se detienen a analizar tal uso para la conservación, mientras sí ponen en valor la utilidad para la difusión.

Abadal y Guallar afirman contundentemente: “aunque pueda parecer paradójico, las principales organizaciones relacionadas con la preservación (Library of Congress, British Library, National Library of Australia, ICON, etcétera) defienden en estos momentos la microfilmación como el formato fundamental para asegurar la preservación de la prensa impresa” (2010, p. 141).

Bien es cierto que este aspecto sigue generando disparidad de opiniones y polémica en la comunidad del patrimonio cultural. En sentido contrario al expuesto hasta ahora, cabe mencionar que recientemente las entidades integradas en *The Association for Research Libraries* han adoptado oficialmente la digitalización como una estrategia de preservación aceptable (Conway, 2010).

Sin embargo, en la década de 1990, cuando el proyecto del CDPV iniciaba su andadura, se recomendaba de forma generalizada recurrir a la digitalización como medio de conservación. Se trataba básicamente de “digitalización para la preservación”, con el objetivo principal de minimizar el uso de los documentos originales y favorecer su difusión. El estado del arte en aquellos momentos, y cierto optimismo tecnológico, no entraban a considerar, salvo excepciones, la problemática derivada de la conservación de los materiales digitales generados como resultado del proceso de digitalización. Habrá que esperar una década para que las publicaciones y recomendaciones de ámbito bibliotecario comiencen a estar atentas al problema de la preservación digital. Veamos algunos ejemplos.

Las *Normas y directrices para bibliotecas universitarias y científicas* recomendaban la digitalización para la preservación del patrimonio bibliográfico histórico, sin entrar a hacer ningún tipo de consideración en relación con la preservación digital (REBIUN, 1999).

Esta misma visión de la digitalización para la preservación, carente de valoraciones sobre la preservación digital de los *outputs* generados, está bastante extendida en los proyectos de la época. Habitualmente reparan únicamente en las ventajas del soporte digital, como son su versatilidad para la difusión o evitar el uso de los originales (véanse por ejemplo, en el ámbito bibliotecario vasco: Bilbao Ariño, 2005; Gómez, 1997; Valverde Tejedor, 2005).

Es alrededor del año 2000 cuando comienzan a aparecer publicaciones que resaltan la problemática de la conservación de los *outputs* digitales. Los *Principios para el cuidado y manejo de material de bibliotecas* (IFLA, 1998) recogen recomendaciones sobre las diversas razones por las que digitalizar:

- Preservar el contenido intelectual.

- Reducir el desgaste y destrucción de los originales.
- Ahorrar espacio, ya que el material quebradizo y seriamente dañado puede desecharse, siempre que su valor radique en el contenido intelectual y no en el objeto mismo.
- Mejorar el acceso. Las copias de microfilm y la información digitalizada pueden distribuirse a otros lugares, proporcionando acceso a más de un usuario a la vez.
- Duplicar ciertos registros por razones de seguridad, en caso de que los originales estén dañados, hayan sido robados o destruidos.

Y alertan ya sobre los problemas de almacenamiento de las copias electrónicas, no considerando el traspaso a formato digital un “verdadero archivo” y recomendando, finalmente, la combinación de la microfilmación para las copias maestras y la digitalización para brindar mayor acceso al contenido de los documentos.

La guía para proyectos de digitalización *Llevando la teoría a la práctica: tutorial de digitalización de imágenes*, que ha sido tan citada y utilizada en nuestro entorno gracias a contar con versión española desde su publicación en 2000, incluye ya un capítulo dedicado a la preservación digital (Cornell University Library, 2000).

El manual de la editorial complutense *Temas de biblioteconomía general y universitaria* coordinado por Magán Walls (2002), incluye un capítulo de Pilar Moreno donde se detallan los inconvenientes de la digitalización y se recogen las principales estrategias a seguir para la protección de los datos.

Ya en 2002, las *Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos* (IFLA et al., 2002) enumeran nuevas razones para decidir la implementación de un proyecto de digitalización:

- Mejorar los servicios.
- Ofrecer a la institución oportunidades de desarrollo tecnológico y formación técnica de su personal.
- Impulsar el desarrollo de recursos cooperativos.
- Buscar intereses comunes con otras instituciones para rentabilizar las ventajas económicas de un enfoque compartido.
- Aprovechar oportunidades financieras.

Y, en relación con la preservación, los expertos del Consejo Internacional de Archivos y de la IFLA que redactaron este informe para la UNESCO afirman que “digitalización no es lo mismo que preservación: la digitalización no es más barata, más segura, ni más fiable que la microfilmación” y se muestran tajantes al afirmar que “a diferencia de un fotograma de un microfilm de alta calidad, una imagen digital no es un máster de preservación”. Reducen la utilidad de la “digitalización para la preservación” a un cometido: “la única manera en la que la conversión digital contribuye positivamente a la preservación es cuando la copia digital reduce el desgaste y la rotura del original” (IFLA et al., 2002, p. 11).

Se ha de tener en cuenta además que con la digitalización se minimiza la manipulación del original, pero no se elimina totalmente, dado que siempre habrá un pequeño porcentaje de investigadores que necesiten acceder al original por una u otra causa relacionada con el objeto de su estudio.

Estas directrices dedicaron ya entonces, de modo bastante pionero, un capítulo a la preservación del contenido digital. Pensemos que en aquel mismo año de 2002 se produjo el desastre del *BBC Domesday Project*. Tan sólo siete años antes se había publicado el famoso y muy citado artículo de Jeff Rothenberg *Ensuring the longevity of digital information*, que aparece originalmente en la revista *Scientific American* en enero de 1995. Ese mismo año se publica en la revista *Investigación y ciencia*, edición española de la publicación americana, bajo el título *¿Son perdurables los documentos digitales?* (Rothenberg, 1995)<sup>3</sup>.

Un años después del informe de la IFLA y el ICON ven la luz las *Directrices para la preservación del patrimonio digital* (UNESCO, 2003a), que se convertirán pronto en guía de referencia en este campo.

Finalmente, cabe resaltar el informe de febrero de 2008 de la Comisión Española sobre la Digitalización y la Accesibilidad en Línea del Material y de la Conservación Digital (CEDALMAC), que incluía entre sus líneas estratégicas de trabajo la preservación digital (Vives et al., 2009).

En el ámbito bibliotecario vasco aparecen en 2005 las primeras voces que ponen el foco sobre los problemas asociados a la preservación digital (véanse Cuesta Ezeiza, 2007; Gómez, 2005), y en especial el informe encargado por el Gobierno Vasco a *Ibermática* y presentado en la *Conferencia Internacional sobre Políticas Culturales* (Cuesta, 2005).

El *Plan Vasco de la Cultura* (2004-2007), que trazaba las líneas de actuación para la digitalización de fondos de archivos, bibliotecas y museos, no hacía aún mención expresa a la cuestión de la preservación digital (Gobierno Vasco, 2004). Cabe destacar, sin embargo, que al amparo de dicho plan se elaboró el *Plan estratégico sobre la conservación y preservación del patrimonio digital en la Comunidad Autónoma de Euskadi* cuyos objetivos estratégicos eran “impulsar el desarrollo del patrimonio cultural digital; asegurar la preservación del patrimonio digital; y, facilitar la difusión del patrimonio cultural digital” (Cuesta, 2005, p. 4). Esta iniciativa se concretaría en la puesta en marcha del repositorio con fondos patrimoniales digitalizados *Memoria Digital Vasca*, y se darían los primeros pasos para la puesta en marcha de la política de preservación web.

El posterior *Contrato Ciudadano por las Culturas* de 2010 tampoco incluye referencias a la preservación digital, pero sí establece entre sus líneas estratégicas un proyecto transversal (L5.T3) llamado *Plan integral de digitalización de fondo histórico y fondo vivo* que pretende priorizar acciones ofreciendo un calendario con los fondos de interés a digitalizar (Gobierno Vasco, 2010). Su fruto más destacable ha sido la puesta en marcha del repositorio *Liburuklik – Biblioteca Digital Vasca* con fondos culturales patrimoniales digitalizados.

---

<sup>3</sup> Posteriormente, el mismo autor hace una versión ampliada (Rothenberg, 1999), que no se ha publicado en castellano, puesto que la versión aparecida después en la revista *Temas de investigación y ciencia* es nuevamente la traducción de la primera versión (Rothenberg, 2004).

En el informe *Mapa de digitalización* (Observatorio Vasco de la Cultura, 2011) se dedica por vez primera un esfuerzo notable a realizar un estudio detallado del asunto de la preservación digital. Aborda el estado de la cuestión describiendo el contexto y el marco conceptual, detallando los principales programas y proyectos internacionales, analizando experiencias y buenas prácticas en el ámbito de la Comunidad Autónoma de Euskadi, detallando aspectos clave en los proyectos de preservación digital, analizando requisitos y recomendaciones técnicas y exponiendo unas orientaciones finales. Supone el mayor esfuerzo de concreción sobre preservación digital que se ha realizado a nivel institucional en el País Vasco.

Por otra parte, durante la vigencia de la *Agenda Digital de Euskadi 2010* una de las líneas del plan de acción fueron las acciones enmarcadas bajo el epígrafe “Servicios para la cultura y el patrimonio cultural vascos”, detallándose acciones para la “Digitalización y preservación del patrimonio cultural” (Plan Euskadi en la Sociedad de la Información, 2008). Tales iniciativas se materializarían en la creación del archivo web *Ondarenet*.

No obstante, la *Agenda Digital de Euskadi 2015* no contempla ninguna acción encaminada a la preservación del patrimonio cultural digital o digitalizado (Plan Euskadi en la Sociedad de la Información, 2012). Se trata sin duda de una gran pérdida de oportunidad, puesto que los retos de la preservación digital suponen hoy un gran desafío y enfrentarlos representa un problema de gran vigencia.

## 2.5. PLANES DE PRESERVACIÓN DIGITAL

La preservación digital solo es alcanzable mediante el establecimiento de una política que detalle tanto las estrategias a seguir como las acciones a ejecutar sobre una colección digital. Para implementar dicha política se ha de articular un plan de preservación digital. El plan tiene la finalidad de programar y organizar la ejecución de las acciones conforme a las estrategias previstas y definidas en la política. El plan debe por tanto, atender a la política escrita, identificar los responsables y la colección sobre la cual se va a aplicar, programar las acciones, estimar los recursos disponibles y marcar un calendario.

Diversos proyectos internacionales resaltan la importancia de establecer un plan de preservación digital desde el inicio de todo proyecto digitalizador. Tales indicaciones pueden observarse en recomendaciones y guías para acometer proyectos de digitalización desde hace más de una década, como en *Llevando la teoría a la práctica: tutorial de digitalización de imágenes* (Cornell University Library, 2000), o en las *Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos* (IFLA et al., 2002). La premisa más extendida viene a alertar de que de nada vale digitalizar sino se atiende irremediablemente, y desde el primer momento, al problema de la preservación digital. Y la planificación de la preservación digital es la herramienta más adecuada para encarar los problemas que surgen en torno a la conservación de los activos digitales.

Llueca y Reoyo, después de enumerar las amenazas que afectan a la preservación de los objetos digitales y tras poner el acento en que ésta sólo es alcanzable mediante la continua transformación de los objetos originales para hacerlos accesibles (“preservación por alteración”, lo llaman), señalan que “probablemente el mayor reto al que se enfrentan los

profesionales de bibliotecas, archivos y museos es la planificación de la preservación digital” (2012, p. 13).

Muñoz de Solano, por su parte, remarca que “cada institución debe determinar sus objetivos particulares y a partir de éstos, respetando los parámetros normativos existentes, desarrollar su propio plan de preservación” (2010, p. 7). La autora defiende que solo de modo planificado, con un adecuado equilibrio entre coste y beneficio, se pueden alcanzar los objetivos de la preservación digital.

Carl Wilson, de la British Library, expuso en el *Digital Preservation Summit 2011* las lecciones aprendidas sobre planificación de la preservación digital en el marco de los proyectos desarrollados por la Open Planets Foundation (2011). Este autor define la planificación como el proceso organizacional de crear un plan, parándose a pensar en las actividades requeridas para alcanzar las metas deseadas. Establece razones de sentido común para la planificación, como que es bueno pensar sobre la cosas, que el hecho de planificar provee de objetivos claros sobre a dónde se quiere llegar, cuáles son las opciones existentes y cuáles son las mejores de estas opciones y porqué. También señala que planificar es bueno para afrontar los retos de la preservación digital porque ayuda a sentirse preparado. Alerta, sin embargo, de que no todo es planificable, pues hay aspectos incontrolables y los recursos siempre son limitados, y advierte de que un plan no puesto en acción es simplemente un plan.

Según este autor, un plan de preservación define una serie de acciones de preservación a llevar a cabo por la institución responsable para evitar los riesgos identificados sobre un set de objetos digitales dado (colección).

Wilson (2011) y Alberch (2012) enumeran los elementos a tener en cuenta para la formulación de un plan de preservación. Ambos se basan en los ofrecidos por el grupo de trabajo del Comité Técnico 46 de la International Organization for Standardization (ISO) en su documento *Preservación de los documentos digitales: Guía para comenzar. ISO/TC 46/SC 11. Versión española, octubre 2010* (Elvira, 2011) que mantiene que un plan de preservación es “un plan para preservar una colección concreta o una parte de una colección de objetos digitales, teniendo en cuenta:

- las políticas de preservación,
- las obligaciones legales,
- las limitaciones de la organización,
- las limitaciones técnicas,
- las necesidades de los usuarios,
- y los objetivos de la preservación.”

La planificación de la preservación se realiza mediante la programación de las estrategias que se van a seguir para un periodo dado (de corto, medio o largo plazo). Tal programación debe incluir detalle de las actividades y procesos a seguir, debe establecer una política acorde con la misión de la organización, y debe atender cuidadosamente el plan económico y de gasto. En este último sentido, todo plan “debe ser realista y aunque no sea una tarea fácil, desde el punto de vista de la planificación, es necesario estimar los gastos del programa a corto o

medio plazo (...) partiendo de las premisas anteriores hay que simplificar al máximo posible cada tarea y abarcar programas sencillos” (Muñoz de Solano, 2006, p. 577).

Finalmente, hemos de recordar que en el momento actual son pocas las instituciones de nuestro entorno que se han dotado de planes de preservación digital. Como ejemplo, baste decir que en el análisis DAFO realizado para el *Plan estratégico 2012-2014 de la Biblioteca Nacional de España* (2012) esta institución de referencia reconoce entre sus debilidades la “falta de política e infraestructuras para la preservación digital” y entre sus oportunidades el “buen momento para desarrollar políticas de digitalización y preservación digital”. Entre las estrategias a seguir la BNE se marca la “aplicación gradual del Plan de preservación y acceso al documento” y el “desarrollo del sistema de preservación digital”. Es, por tanto, interesante señalar que hasta las instituciones más grandes están aún comenzando a dar pasos en la planificación de la preservación digital.

## 2.6. ESTRATEGIAS Y MODELOS TEÓRICOS

Los requisitos a cumplir para que los recursos digitales perduren en condiciones de accesibilidad y legibilidad atañen tanto a cuestiones técnicas como a estrategias de acción; y desde luego, están sujetos a condicionantes legales, organizativos y económicos.

Los autores y organismos que se han detenido a analizar la problemática asociada a la preservación digital señalan que más allá de los retos tecnológicos se detectan unos obstáculos más complejos de superar, como son los legales (contar con el permiso del titular para difundir o transformar un recurso digital), financieros (contar con los medios económicos necesarios para asegurar la continuidad de la iniciativa de conservación) e institucionales (contar con el compromiso e implicación a largo plazo de las instituciones embarcadas en el proyecto) (Ferreras, 2010; Keefer, 2007; Lluca, 2012b; Térmens, 2013).

En cualquier caso, en los últimos años se han experimentado y puesto en práctica diversas estrategias y acciones de índole técnico encaminadas a asegurar que los objetos digitales puedan ser recuperados, decodificados, usados y entendidos en el futuro. A pesar de los muchos informes, estudios y proyectos, no está disponible ni se ha desarrollado todavía ningún sistema informático que garantice la preservación digital a largo plazo (véanse Térmens, 2010a para conocer el avance en el desarrollo de software; y Térmens, 2013 para una relación de las herramientas que facilitan el trabajo de preservación digital mediante la comprobación de formatos, la comprobación de la integridad mediante sumas de comprobación –*checksum*– y la transferencia de ficheros). A día de hoy se han de combinar estrategias a medio plazo y actividades contrastadas y experimentadas que permitan prolongar en el tiempo la accesibilidad a los documentos digitales (Lluca, 2012a). Las acciones encaminadas al refresco de los soportes, a la migración de los formatos y a la emulación del entorno técnico originario son las que presentan mayor grado de maduración y consenso en la comunidad científica.

De cara a contextualizar debidamente el actual estado de las cosas en preservación digital seguiremos a Lluca, quien describe varias lecciones aprendidas en relación con la preservación digital (2012b):

- Los documentos digitales son frágiles y están expuestos a una rápida obsolescencia.

- No existe una solución técnica infalible para la preservación digital.
- Las dos técnicas más extendidas (migración de soportes y formatos, y emulación) deben ser empleadas lo mínimo posible, puesto que traerán complicaciones y no son infalibles.
- Hay que encontrar el punto de equilibrio entre la importancia relativa del documento y los factores de riesgo a que está sometido, puesto que los recursos son siempre limitados.
- A menudo hay que decidir entre preservar el contenido o mantener el recurso preservando el aspecto y el comportamiento original.

Partiendo de las anteriores premisas, el mismo autor señala varias ideas, de índole organizativo y de gestión, a contemplar en el diseño de una estrategia de preservación digital (2012a):

- Frente a la preservación tradicional orientada a mantener el documento en su estado original, la preservación digital es preservación por alteración.
- La preservación digital es más compleja que la preservación tradicional.
- La limitación de los recursos implica el tener que establecer una política, entendida ésta como una toma de decisiones que promueve unas acciones descartando otras.
- Se debe hallar el equilibrio entre la importancia relativa del documento y los factores de riesgo. Se ha de realizar un análisis de coste/beneficio.
- Hay que emprender actuaciones proactivas, puesto que las reactivas pueden ser incapaces de solucionar el daño.
- Las actuaciones deben ser continuadas, no esporádicas, afectando tanto al documento como a su descripción; ello requiere financiación permanente.
- La omisión no es una opción, el olvido es letal.
- Es preferible actuar desde el momento de la creación del documento.
- Se ha de decidir entre preservar el contenido o preservar el aspecto y comportamiento original.

A continuación se enumeran y describen las llamadas estrategias técnicas de preservación digital o simplemente estrategias para la preservación digital (se sigue a Alberch, 2012; Elvira, 2011; Ferreras, 2010; Keefer y Gallart, 2007; Lluca, 2012a; Ortiz Ancona, 2012; Rivera Donoso, 2009; Salvador Benítez y Ruiz Rodríguez, 2005; Térmens, 2013). Se trata de actividades de orden técnico que se realizan sobre la propia colección de ficheros digitales. En general, hay consenso en considerar las tres técnicas que se describen a continuación como las principales estrategias técnicas:

- *Actualización de soportes* (también denominada *refreshing*, refresco, renovación, rejuvenecimiento o migración de soportes). Se trata de transferir los datos de un soporte a otro. No actúa sobre los datos, pero es una opción repetible e imprescindible para garantizar el acceso a los mismos. Se ha de tener en cuenta la previsión de vida

del nuevo soporte, y la previsión de comercialización y vida de las unidades de lecto-escritura.

- *Migración de formatos* (también denominada conversión o reformateado). Se trata de transferir los datos de un formato a otro. Existe riesgo de pérdida de elementos del documento y de pérdida de datos. Es una operación que no requiere conocimientos técnicos especializados; aunque es difícil de programar en objetos digitales complejos, donde se multiplica el riesgo de alterar los documentos. La migración de formatos se ha de repetir a lo largo del tiempo y se ha de apostar por formatos estándar, abiertos o propietarios con amplia base de clientes.
- *Emulación*. Posibilita usar el software y los datos originales sin que el sistema original sobre el cual se ejecutaba exista. Ello se logra mediante la recreación del entorno técnico requerido para poder acceder a los documentos electrónicos, creando para ello un programa que simule el funcionamiento del hardware y del software original. Según algunos autores es la mejor solución para preservar los documentos digitales guardando su *look & feel* original.

En un segundo orden aparecen varias rutinas técnicas que van desde simples medidas de conservación preventiva a criterios y recomendaciones de orden más general. Convenimos con Keefer y Gallart, aunque sólo enumeran algunas de ellas, cuando dicen que son prácticas que no todos consideran estrategias (2007). En tal sentido hay que considerar acciones como las siguientes:

- *Utilización de medios persistentes*. Como CDs o DVDs de platino o de oro, o cualquier medio de almacenamiento de alta calidad.
- *Confianza en los estándares*. Procurando utilizar los formatos y modelos asentados puesto que ello es garantía de permanencia y soporte por parte de los desarrolladores. Se trata de primar el uso de formatos y medios estándar, validados tanto por un uso contrastado y por su penetración en el mercado (estándar de facto) como por haber pasado a ser norma técnica aprobada por una entidad normalizadora. Hay que decir que los formatos abiertos más difundidos están documentados en estándares internacionales.
- *Metadatos de preservación*. Se refiere al uso de la descripción y documentación de los ficheros. Posteriormente se ahondará en este asunto, que algunos consideran entre las estrategias o actividades básicas.
- *Encapsulación*. Se refiere al hecho de empaquetar en un único fichero los datos que se van a preservar junto a la metainformación que los describe.
- *Mantenimiento* (también denominado cuidado duradero). Actividad orientada a garantizar el buen estado de los recursos digitales, atendiendo al uso de medios de almacenamiento seguros y fiables, y cuidando que las condiciones de almacenamiento y manipulación sean óptimas, con la finalidad de optimizar las expectativa de vida de los soportes según las recomendaciones del mercado. En la medida en que no hace frente al problema de la obsolescencia es considerado sólo una estrategia para el corto plazo.



- *Mantenimiento físico de los equipos*. Se trata de tomar medidas tendentes a evitar la pérdida de los equipos por deterioro. Tales medidas de conservación preventiva van en la línea de controlar las condiciones de manipulación y el estado de los equipos.
- *Respaldo* (también denominado copia de seguridad, duplicado, *backup* o copia de respaldo). Es un duplicado exacto del objeto a preservar. En la medida en que no hace frente al problema de la obsolescencia es considerado sólo una estrategia para el corto plazo.
- *Respaldo analógico* (también denominado analogización). Consiste en revertir o pasar los documentos digitales a soportes analógicos que cuenten con solvencia contrastada para la preservación, como el papel o la microfotografía.
- *Replicación*. Es otra forma de realizar copias de seguridad o *backup*. Se trata de copiar la información a uno o más sistemas y su principal ventaja es que la información se almacena en más de un lugar. Programas como LOCKSS están basados en esta idea.
- *Preservación de la tecnología* (también denominado museo tecnológico o museo informático). Consiste en conservar en perfecto estado de funcionamiento tanto el hardware como los sistemas operativos, los programas y aplicaciones originales.
- *Canonicalización*. Se trata de crear un sistema que permita saber si un fichero ha sido transferido de un formato a otro; mediante caracterización, claves y algoritmos canónicos se puede saber si un fichero guarda su esencia tras la conversión a un nuevo formato.
- *Control de la autenticidad* (también denominado control de estabilidad, control de integridad o fijeza de la información). Aseguramiento de la integridad de la información y comprobación de no transformación mediante *checksum* o suma de verificación, marcas de agua digitales, códigos hash, encriptación y firmas digitales. En archivística se habla de preservar los valores esenciales del documento, como son la autenticidad, la fiabilidad, la integridad, la identidad y la manejabilidad o capacidad de uso a lo largo del tiempo (Soler, 2008).
- *Arqueología digital* (también denominada análisis forense digital). Puesta en marcha, reparación y rescate de datos, programas y hardware. Recuperación de información dañada o antigua.
- *Computador virtual universal (UVC)*. De momento se trata de una estrategia más teórica que práctica. “Es una representación virtual de una computadora simplificada que permite guardar información actual codificada y descifrarla en el futuro” (Rivera Donoso, 2009, p. 33).

Todas las estrategias técnicas de preservación digital están encaminadas a la consecución del objetivo último de la preservación digital, que tal y como se recoge en cualquiera de sus definiciones, se refiere al hecho de proveer acceso permanente a los recursos. Observando esta preocupación básica por el acceso y teniendo en cuenta el actual entorno social y tecnológico de la información, que como se sabe se trata un contexto distribuido y servido en red, no es de extrañar que, hoy por hoy, los repositorios digitales sean considerados la

herramienta por excelencia para garantizar el acceso abierto, permanente y estable a los recursos digitales.

En este campo, uno de los grandes consensos y el principal estándar, diseñado con orientación hacia la preservación digital es la solución propuesta por el modelo teórico *Reference Model for an Open Archival Information System (OAIS)*, que es norma ISO 14721:2003. Aunque ha habido voces que han reclamado mayor sencillez a este modelo y a pesar de que también se ha experimentado con otros modelos teóricos (como PAIMAS, CIDOC o DCC *Curation Lifecycle Model*), hoy día son muchos los repositorios digitales levantados conforme a los principios del modelo OAIS, que se ha convertido en un modelo de referencia.

OAIS plantea dos funciones básicas que debe tener un sistema de archivo seguro o repositorio fiable: proveer acceso a la información y preservarla. Para ello trata la información en paquetes, donde el objeto a preservar contiene además metadatos que documentan su naturaleza, origen y vida de modo que sea autosuficiente para poder ser interpretable y decodificable en momentos posteriores. No se van a detallar aquí las características del modelo y se recomienda acudir a la reciente monografía *Preservación digital* (Térmens, 2013) que en breves páginas da cumplida cuenta del funcionamiento del modelo propuesto.

OAIS no se debe confundir con la iniciativa *Open Archives Initiative (OAI)* que nació con la finalidad de desarrollar y promover estándares de interoperabilidad que facilitasen la difusión de contenidos en internet. Aunque OAI nace con la idea de proveer acceso abierto a los documentos digitales, los avances centrados en la interoperabilidad, la recuperación automática de los metadatos y el desarrollo de repositorios, han promovido fuertemente el desarrollo de protocolos y estándares para la comunicación y la recuperación, como *Protocol for Metadata Harvesting (PMH)*, *Dublin Core (DC)* o XML, que en la medida que normalizan y promueven la interoperabilidad facilitan asimismo la permanencia y preservación de los recursos digitales.

Del mismo modo que no todos los repositorios digitales cumplen con la norma ISO 14721:2003 (OAIS), tampoco todos los repositorios digitales son considerados repositorios de preservación. De hecho, la mayoría no lo son. Siguiendo a PREMIS, un repositorio de preservación es aquel que tiene como finalidad última o como parte de sus responsabilidades desempeñar la preservación a largo plazo de los objetos digitales que alberga (Keefer, 2007; Térmens, 2009a). Con la finalidad de certificar los repositorios que cumplen una serie de requisitos para ser considerados repositorios de confianza en relación con la preservación digital y con el cumplimiento de los estándares, surge la denominación *Trusted Digital Repositories*. Se trata de auditar y certificar conforme a una norma establecida para otorgar la denominación de “repositorio digital de confianza”, esto es, validado conforme a normativa para el cumplimiento de la preservación digital. El estándar tiene rango internacional y fue aprobado en 2012, se trata de la norma ISO 16363:2012 *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*.

Gallart y Keefer definen un depósito digital fiable como “una institución que pretende ofrecer acceso fiable a largo plazo a recursos digitales gestionados por él” (Keefer y Gallart, 2003) y se identifican tres niveles de fiabilidad:

- Que los usuarios se fíen de la institución.

- Que la institución se fie de los subcontratistas para la parte externalizada, normalmente técnica.
- Que el colectivo de usuarios se fie de los datos preservados.

Teniendo en cuenta que la preservación de grandes volúmenes de datos es cara y que el replicado de los datos es una estrategia preventiva básica, han surgido propuestas cooperativas consistentes en compartir recursos de computación para facilitar que los datos de diferentes organizaciones se conserven en servidores de otras organizaciones. Es muy conocida la iniciativa *Lots Of Copies Keep Stuff Safe* (LOCKSS) y su variante *Controlled Lots Of Copies Keep Stuff Safe* (CLOCKSS) promovida por un grupo de bibliotecas y que funciona “mediante una red cerrada *peer to peer* (que) permite la replicación de datos científicos y su preservación”(Térmens, 2009a, p. 616).

Otros proyectos como *Duracloud* que es un “sistema de almacenamiento y preservación con computación y replicación en la nube, fruto de la unión de Dspace y FedoraSpce de la mano del proyecto DuraSpace” (Térmens, 2009a, p. 618), *Portico* que es un servicio para la preservación de revistas electrónicas, o HathiTrust Digital Library promueven y facilitan también la externalización de los servicios de preservación digital. En España la Universidad Complutense de Madrid ha suscrito un acuerdo con esta última.

La Biblioteca de Catalunya ha puesto en marcha COFRE *Conservamos para el Futuro Recursos Electrónicos* (Pérez y Serra, 2010), que ha posibilitado que organizaciones como el Ateneu Barcelonès hayan externalizado la preservación mediante un acuerdo con la propia biblioteca (Montes et al., 2013).

Como se ha visto, entre todas las estrategias para la preservación las acciones encaminadas al refresco de los soportes, a la migración de los formatos y a la emulación del entorno técnico originario son las que presentan mayor grado de maduración y consenso en la comunidad científica. Además, el principal modelo teórico desarrollado, el modelo OAIS, es la alternativa más asentada en el contexto de los datos servidos en red.

## 2.7. ESTÁNDARES PARA LA PRESERVACIÓN

Como ya se ha comentado, el uso de estándares en formatos, en herramientas de software y en hardware, favoreciendo la utilización de los más asentados e interoperables, es un criterio básico que debe seguirse en aras a facilitar la preservación de los documentos digitales.

Uno de los hechos que denota la notable magnitud que la preocupación por la preservación digital ha levantado a nivel mundial, es la abundante producción normativa internacional que se ha generado en los últimos años. ISO ha publicado una notable colección de normas sobre modelos teóricos, repositorios, formatos, esquemas de metadatos, auditoría de repositorios de confianza, etcétera<sup>4</sup>.

---

<sup>4</sup> Se enumeran algunas de ellas, sin ánimo de exhaustividad: ISO 14721:2012, ISO/TR 15801:2009, ISO 15836:2009, ISO 16363:2012, ISO/TR 18492:2005, ISO 19005-1:2005, ISO 19005-2:2011, ISO 19005-3:2012, ISO 20652:2006, ISO 23081-1:2006, ISO 23081-2:2009, ISO/TR 23081-3:2011 o ISO 32000-1:2008.

Por otra parte, merece la pena destacar que tanto el *Real Decreto 3/2010, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica*, como 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*, mencionan expresamente la conservación de los documentos electrónicos; en el Esquema Nacional de Seguridad se cita además de forma explícita el asunto de la "preservación del patrimonio digital".

Se presentan a continuación unos breves criterios en relación con la elección más adecuada de formatos y esquemas de metadatos en función de su idoneidad para la preservación, y se enumeran aquellos sobre los que hay mayor consenso.

### 2.7.1. *FORMATOS DE IMAGEN RÁSTER PARA PRESERVACIÓN*

Un formato es una codificación de la información representable o legible mediante software. Es la estructura de un archivo que define la forma en que éste es almacenado y presentado. Puede ser simple, como en el caso de los archivos almacenados en forma de texto ASCII, o muy complejos e incluir diversos tipos de códigos e instrucciones de control a usar por programas y dispositivos.

En relación con la propiedad de los formatos existen: formatos propietarios, cuya codificación solo es conocida por la empresa que posee los derechos; y formatos no propietarios, que son libres y abiertos. La estructura de codificación de los formatos abiertos es pública, accesible y gratuita.

Todo formato es representable mediante software propietario, mediante software no propietario o mediante software de código abierto. Este último tiene las fuentes de codificación abiertas pero es propietario; por tanto, es abierto pero no libre, no gratuito.

De entre los tipos anteriores de formatos, bien sean propietarios o libres, hay algunos que se han establecido como formatos estándar, normalmente como resultado de un proceso natural de éxito o primacía en el mercado, pero también por consenso generalizado y conversión en norma sectorial, nacional o internacional. Entre ellos, hay algunos más estables que otros, en el sentido de que están soportados por empresas solventes y presentan mayor compatibilidad con otros formatos.

Existe bastante consenso en que los formatos libres, los de código abierto, los más estandarizados y los más estables, cumplen las condiciones para ser más perdurables, por ello se prefieren y se recomienda su uso para la preservación digital.

El manual *Preservation management of digital materials* hace las siguientes recomendaciones sobre formatos de fichero (Beagrie y Jones, 2008):

- Use formatos abiertos no propietarios y bien documentados, cuando sea posible.
- Alternativamente use formatos bien desarrollados, ampliamente adoptados y que sean estándares de facto en el mercado.
- Identifique formatos aceptables para cada finalidad: la transferencia, el almacenamiento y la distribución a los usuarios (pueden ser formatos distintos).

- Reduzca al mínimo, en la medida de lo posible, el número de formatos de archivo a gestionar.
- No use encriptación ni compresión en los ficheros de almacenamiento, si es posible.

Existen unos criterios claros para la selección del formato digital en relación con la finalidad de uso de los archivos de imagen. En aquellos cuya finalidad es la difusión se ha de atender a factores de calidad y funcionalidad (resolución, en relación con la velocidad de transferencia, y compatibilidad, en relación con el uso de diferente software y hardware en soportes de almacenamiento y redes). En aquellos formatos cuya finalidad última es la preservación se ha de atender a los llamados factores de sostenibilidad (véase la Figura 3), estos son: divulgación o transparencia, apertura, dependencia o interoperabilidad, estabilidad o compatibilidad, aceptación, estandarización y mecanismos de protección técnica (Observatorio Vasco de la Cultura, 2011). También Soler (2008) ha enumerado los criterios que deben tenerse en cuenta a la hora de elegir los formatos de fichero más adecuados de cara a la preservación digital.

<b>TABLA A-Relación de factores de sostenibilidad a tener en cuenta en la selección de los formatos establecidos por entidades internacionales de referencia</b>	
-	<b>Divulgación/Transparencia:</b> hace mención al grado en el que las especificaciones técnicas de los formatos son accesibles (es decir son de código abierto).
-	<b>Apertura:</b> se refiere a la dependencia de un formato determinado a una patente. En el caso de las entidades dedicadas a la preservación y/o difusión del patrimonio cultural, en general, se tiende a utilizar formatos abiertos, es decir formatos de archivo que no se hallan sujetos a patentes o derechos de autor. En el caso de las industrias culturales, existe un vínculo creciente entre las plataformas distribuidoras de los contenidos y los fabricantes de software y dispositivos (Apple, Amazon, etc.), que pueden imponer formatos propietarios (protegidos por patentes o derechos de autor) siguiendo criterios comerciales.
-	<b>Dependencia/Interoperabilidad:</b> se trata del grado de dependencia de un formato determinado respecto a un hardware, a un software o a un sistema operativo específico.
-	<b>Estabilidad/Compatibilidad:</b> grado en el que un formato mantiene su funcionalidad e integridad con versiones anteriores o posteriores.
-	<b>Aceptación:</b> hace referencia al grado de utilización de los formatos por parte de los creadores, distribuidores y usuarios de los recursos.
-	<b>Estandarización:</b> adecuación formal a los procesos o especificaciones establecidos por un organismo de normalización (ISO, NISO, W3C...) con el objetivo de garantizar la calidad de los archivos y su interoperabilidad.
-	<b>Mecanismos de protección técnica:</b> los mecanismos de protección técnica como por ejemplo el cifrado, utilizados habitualmente para proteger la propiedad intelectual, no deben dificultar la recuperación de datos, migración de los contenidos o su adaptación a nuevas necesidades derivadas de la evolución tecnológica.

Figura 3.- Factores de sostenibilidad en formatos para la preservación digital (Fuente: Formatos de difusión y formatos de preservación de contenidos digitales. Observatorio Vasco de la Cultura, 2011)

Hay nuevamente gran consenso en cuanto a los formatos de imagen considerados más adecuados para la preservación digital. Enumeramos a continuación solamente los formatos idóneos para conservar la versión maestra de los ficheros de imagen, no se mencionan los formatos de fichero para versiones derivadas dado que no es el objeto de este estudio.

**RAW** – Archivo de imagen en crudo o en bruto, sin modificaciones, que contiene todos los datos de la imagen tal y como ha sido captada por el sensor de la cámara de fotos digital o del escáner. Considerado el negativo digital. En realidad no se trata de un formato de imagen sino de muchos distintos pero del mismo tipo, dado que cada casa comercial de cámaras fotográficas o de escáneres suele tener su propia extensión de archivo RAW. Son, por tanto, formatos propietarios. En la Figura 4 se muestran las principales extensiones de archivos de imagen RAW. El hecho de contener toda la información bruta y sin transformar captada por la cámara o escáner los hace especialmente adecuados para la conservación de los ficheros maestros originales.

Archivo de imagen RAW	
Información general	
Extensión de archivo	.raf (Fuji) .crw .cr2 (Canon) .tif .k25 .kdc .dcs .dcr .drf (Kodak) .mrw (Minolta) .nef .nrw (Nikon) .orf (Olympus) .dng (Adobe) .ptx .pef (Pentax) .arw .srf .sr2 (Sony) .x3f (Sigma) .erf (Epson) .mef .mos (Mamiya) .nef (Nikon Electronic Format) .raw .rw2 (Panasonic) .cap .tif .iiq (Phase One) .r3d (Red) .fff (Imacon) .pxn (Logitech) .bay (Casio)

Figura 4.- Principales extensiones de archivos de imagen RAW (Fuente: Wikipedia)

Existe cierta prevención por parte de algunos autores a la hora de considerarlo un formato válido para la preservación, en relación con factores de sostenibilidad por el hecho de ser dependiente de fabricantes. El gran inconveniente que presenta este formato es precisamente la falta de estandarización. En la medida en que cada fabricante de cámaras usa su propia versión se pueden producir incompatibilidades, e incluso que un

fichero RAW concreto no se pueda usar en el futuro por obsolescencia del formato. Afortunadamente, la iniciativa *OPENRAW Digital Image Preservation Through Open Documentation* trabaja con el objetivo de que los fabricantes establezcan un formato RAW de código abierto y estándar.

No obstante este hecho, otros muchos autores entienden que conservar la imagen en crudo ha de ser una estrategia elemental de preservación. Es altamente recomendable guardar siempre la imagen en crudo generada por la cámara, independientemente de que se conserven alguno de los otros formatos específicos para la preservación digital de los ficheros maestros de imagen digital ráster.

**TIFF** (*Tagged Image File Format*) – Significa formato de archivo de imagen etiquetado, es un formato de archivo estándar comúnmente utilizado para escanear, almacenar e intercambiar imágenes gráficas. Puede almacenar imágenes en modo bitonal, escala de grises, paleta de colores o color real. Se han desarrollado distintas versiones, siendo la 5 y la 6 las más admitidas por la mayoría de aplicaciones. Admite distintas posibilidades para la compresión de datos, pero el uso del sistema de compresión sin pérdida LZW ha favorecido su uso para guardar ficheros maestros de preservación. Es el formato más utilizado para la preservación de versiones maestras de ficheros de imagen digital ráster.



JPEG2000 – Convertido en estándar internacional en 2001 es un formato que pretende sustituir al formato JPEG, usa un sistema de compresión más potente que éste y es capaz también de realizar compresión sin pérdida. Es adecuado para albergar copias máster usado con compresión sin pérdida. En compresión con pérdida logra ratios de compresión de tres a cinco veces mayores que JPEG con mejores calidades, dado que no genera los típicos problemas asociados a JPEG, como son los artefactos de compresión. Aunque su soporte por parte de la industria es limitado, está considerado un formato idóneo para la preservación digital y varios proyectos desarrollados por organizaciones de referencia en este campo se han decantado por él, como la Biblioteca del Congreso de los Estados Unidos.

### 2.7.2. ESQUEMAS DE METADATOS PARA PRESERVACIÓN Y PRENSA DIGITALIZADA

Son numerosos los autores que han remarcado la importancia de que los ficheros digitales vayan acompañados de su correspondiente descripción mediante metadatos. Es evidente que el proceso de documentar y describir los objetos digitales favorecerá su interpretación posterior tanto por máquinas como por humanos. Lluca (2012a) afirma que sin metadatos no se puede garantizar que los recursos sean recuperables y usables. Detalla que la inexistencia de datos descriptivos impedirá su localización, la inexistencia de datos técnicos impedirá su visualización o uso, la inexistencia de datos administrativos impedirá verificar su integridad y

autenticidad y, finalmente, la inexistencia de datos legales impedirá conocer los usos admisibles.

Los metadatos son datos incrustados o junto al documento, redactados conforme a un esquema normalizado que sirven para identificarlo, describirlo y facilitar su uso y gestión. NISO ofrece la siguiente definición de metadato: “Información estructurada que describe, explica, localiza, o, de cualquier otro modo, facilita la recuperación, uso y gestión de un recurso de información” (Keefer y Gallart, 2007). Un metadato incrustado es un documento secundario integrado en el documento primario. Los metadatos también pueden presentarse en ficheros independientes que acompañan a los ficheros de imagen (o de otro tipo) que describen.

Además de proveer información clave sobre los recursos digitales para el control de su ciclo vital y la gestión de la preservación, son necesarios para posibilitar su recuperación. Para proyectos de prensa histórica, por ejemplo, la Biblioteca Virtual de Prensa Histórica utiliza un amplio conjunto de esquemas de metadatos: Dublin Core (DC), MARC XML, METS, PREMIS, y recientemente METS/ALTO, relacionado con el reconocimiento óptico de caracteres.

Tanto DC como MARC XML permiten básicamente recuperabilidad dado que son fundamentalmente metadatos descriptivos. En las siguientes líneas se detallan los dos principales esquemas de metadatos de preservación digital: PREMIS y METS. Hay otros, pero se describen solamente aquellos que se han impuesto de un modo generalizado como estándares en este campo.

Por otra parte, hay dos grupos de metadatos relacionados con la fotografía: los EXIF y los IPTC. Los primeros se describen pormenorizadamente más abajo. Los metadatos IPTC (*International Press Telecommunications Council*) están relacionados con la parte social y de publicación de las fotografías, en ellos se describen los datos del fotógrafo, los derechos de copyright del autor así como los de las personas que aparecen en las fotografías (en caso de que aparezcan).

Keefer y Gallart, citando a Lavoie y Gartner (2005), aportan la siguiente definición de metadatos de preservación (2007, p. 140): “son metadatos descriptivos, estructurales y administrativos que apoyan [y documentan] la preservación a largo plazo de los recursos digitales”.

Muñoz de Solano (2010, p. 4), citando a la Biblioteca Nacional de Australia (2001), indica los metadatos de preservación “deben utilizarse para proporcionar:

- Información técnica de apoyo a los administradores del sistema de gestión para la ulterior toma de decisiones y acciones de conservación.
- Información relativa a las acciones de conservación adoptadas previamente como son las políticas de migración o emulación.
- Registrar los efectos y consecuencias de las estrategias de conservación.
- Garantizar la autenticidad de los recursos digitales.
- Información sobre la gestión de los derechos de autor y acceso a la información.”





PREMIS – *Preservation Metadata: Implementation Strategies*. Su origen está en el grupo de trabajo que en 2003 organizan OCLC y RLG, y que dará como fruto la versión 1.0 del *PREMIS Data Dictionary for Preservation Metadata*. La versión 2.0 fue lanzada en 2008. Actualmente es mantenido por la *Library of Congress*. Es una traslación del modelo OAIS y al igual que éste “no define ninguna estrategia o método de implementación, ni prescribe una tipología de acciones, ni el modo en el que deben recogerse, almacenarse o gestionarse los datos referidos a estas acciones” (Ferrerías, 2010, p. 38). Se trata de un “esquema XML que implementa un catálogo de metadatos de preservación. PREMIS se centra en documentar las propiedades técnicas, los agentes, los derechos y los eventos implicados en la preservación de un objeto digital a lo largo de su vida. Su objetivo es dejar constancia documentada, de una forma estandarizada y por tanto automatizable, de aquellos temas que pueden afectar a la conservación de un objeto digital.” (Térmens, 2013, p. 71).

Su objetivo es que pueda aplicarse a cualquier tipo de fondo digital que aplique estrategias de preservación sobre cualquier tipo de objeto digital. PREMIS suele utilizarse de forma conjunta con otros esquemas de metadatos, en particular con METS recogiendo la información de preservación junto a los metadatos descriptivos, de propiedad intelectual, técnicos y estructurales que aúna este otro esquema. El diccionario de datos de PREMIS es muy amplio por lo que muchas instituciones optan por adaptarlo e implementarlo conforme a sus necesidades.

## Metadata Encoding & Transmission Standard

METS – *Metadata Encoding and Transmission Standard*. Desarrollado y mantenido por la Library of Congress. Básicamente considerado un esquema de metadatos estructurales, es un estándar que combina tanto una parte descriptiva e identificativa del contenido del documento como una parte para la descripción de la estructura física y lógica de los documentos. Se compone de siete secciones que recogen datos descriptivos (autoría, título, editor...), administrativos (de derechos y técnicos) y estructurales (relaciones entre partes de un objeto digital, y relaciones del objeto con otros objetos digitales). Es muy versátil y extensible, permitiendo que se puedan integrar distintos tipos de metadatos codificados en XML, como DC, MARCXML, PREMIS, METSRights...

Además, se describen otros dos esquemas de metadatos que no son de preservación pero que a nuestro entender tienen especial incidencia en proyectos de prensa digitalizada. Se trata, por una parte, del estándar METS/ALTO usado en proyectos de reconocimiento óptico de caracteres (OCR). El reconocimiento óptico de caracteres cobra especial importancia en fondos de prensa, de cara a facilitar su recuperabilidad y explotación intensiva por parte de los investigadores y usuarios. Por otra parte, se referencia el estándar EXIF que es un esquema propio de herramientas de captura de imágenes digitales, como cámaras fotográficas y escáneres. Lógicamente, su trascendencia es muy notable en proyectos de documentación digitalizada, dado que tales digitalizaciones se hacen siempre con cámaras fotográficas y

escáneres que recogen una ingente cantidad de metadatos técnicos en el momento de la toma. Comprender EXIF y entender la riqueza de datos que aporta de cara a establecer acciones de preservación digital es fundamental.

## ALTO ...Technical Metadata for Optical Character Recognition...

ALTO – *Technical Metadata for Optical Character Recognition*. Es un estándar abierto y en XML que sirve para describir el texto tras el proceso de reconocimiento óptico de caracteres, y la información de diseño de los documentos impresos. Mantenido por la Biblioteca del Congreso. Es un estándar que permite establecer tanto la estructura de los ficheros como las coordenadas de los caracteres. Su uso es de especial interés para aplicarse en proyectos de digitalización que usen la técnica de OCR. Se utiliza a menudo conjuntamente con el estándar METS. Esta combinación se conoce como METS/ALTO y facilita la búsqueda y navegación por el texto mediante una representación visual del objeto original. Ello ha favorecido la expansión de su uso en programas de digitalización, es muy utilizado en prensa histórica.

## Exif Version 2.3

EXIF – *Exchangeable Image File Format for Digital Still Camera*. Orientado a la descripción técnica de los objetos digitales. Creado y mantenido por la industria japonesa de fabricantes de cámaras fotográficas. Este grupo de metadatos contiene toda la información relativa al equipo fotográfico y a los ajustes y características de la toma. En cuanto al equipo, la información indica el fabricante de la cámara y el objetivo, los modelos de ambos, sus números de serie y el propietario. Y en cuanto a la toma, se describen todos los ajustes realizados para capturar la imagen, así como las características de esta, incluyendo tamaño en píxeles, resolución, formato de imagen, fecha de creación, software utilizado, versiones, etc... Recopila, por tanto, todos los datos técnicos de la toma que han de ser clave para conocer el origen del objeto digital. Las transformaciones posteriores que se realizan sobre el mismo objeto digital quedarán reflejadas también a nivel técnico en el propio fichero Exif. Hay que remarcar que los ficheros de metadatos EXIF van siempre embebidos o incrustados en el propio objeto digital y lo acompañan a lo largo de toda su vida. Esta información se puede extraer e integrar en otros esquemas de metadatos, en especial toda aquella información relativa a la vida y características técnicas del objeto será de especial interés para documentar PREMIS de cara a establecer estrategias de preservación digital.

### 3. CASO DE ESTUDIO: EL FONDO DIGITALIZADO DE PRENSA VASCA

---

Los retos a los que se enfrentan todas las instituciones de la memoria en relación con la conservación de la prensa diaria, como son la fragilidad de los soportes, el gran consumo de espacio para su almacenamiento y la disyuntiva de decantarse por un medio de reproducción u otro, afectan también, como no podía ser de otra manera, a las bibliotecas, archivos y centros de documentación vascos. El Consorcio de Digitalización de Prensa Vasca surge como respuesta por parte de un grupo de instituciones a tales preocupaciones, y hubo de encarar desde sus inicios la problemática asociada tanto a la gestión de grandes colecciones hemerográficas como a la decisión de optar por la digitalización como medio de reproducción de tales colecciones.

El asunto que mayor preocupación suscita es el problema de la conservación de las colecciones en papel, así aparece mencionado reiteradamente a lo largo de los años en la documentación interna del Consorcio. El desafío planteado para la conservación de la documentación patrimonial no digital, junto a la aparición de los medios técnicos que posibilitan acudir a la digitalización como método preventivo para la conservación, concurren para decantar a las bibliotecas vascas hacia el abandono de la microfotografía y a volcar sus esfuerzos en un proyecto de digitalización conforme a la corriente de *digitization for preservation*.

El proyecto del CDPV nace con la finalidad de conservar cooperativamente la prensa diaria editada en Euskal Herria, valiéndose para ello, prácticamente desde el comienzo de su andadura, de la digitalización de los diarios. Hoy se ha de hacer frente a un nuevo reto: la preservación a largo plazo de los objetos digitales generados mediante el escaneado de la prensa.

En esta investigación, tal y como se detalla en la introducción, se aborda el concepto de preservación digital desde el punto de vista de los proyectos llevados a cabo con el fin de “digitalizar para preservar” y más concretamente, en relación con la necesidad sobrevenida de preservar la copia digital generada. Se trata de analizar la preservación de los recursos electrónicos fruto de la digitalización realizada para preservar un original no digital.

La colección hemerográfica que el CDPV viene digitalizando es la prensa diaria actual de ámbito cultural vasco. El consorcio y todas las instituciones que lo han integrado a lo largo del tiempo y en sus distintas etapas establecen el término “ámbito cultural vasco” en el sentido adoptado por la Sociedad de Estudios Vascos–Eusko Ikaskuntza (fundada en 1918 por las Diputaciones Forales de Álava, Bizkaia, Gipuzkoa y Navarra). Este hace referencia a los territorios donde tradicionalmente se ha hablado y habla el euskera, esto es, Euskal Herria como concepto y territorio cultural. La colección de prensa comprende, por tanto, todas las cabeceras publicadas diariamente en la Comunidad Foral de Navarra, en la Comunidad Autónoma de Euskadi y en los territorios de Lapurdi, Baja-Navarra y Zuberoa en el Departamento de Pirineos Atlánticos de la Región de Aquitania en el suroeste francés (*Iparralde*).

Esta colección está conformada únicamente por diarios. Como se sabe, los diarios son prensa escrita que se publica con periodicidad diaria. Pero la prensa escrita abarca también publicaciones de periodicidad semanal o superior, son las revistas (Arquero Avilés y García-Ochoa Roldán, 2005). A lo largo de este trabajo, centrado en diarios, aparecen los términos

prensa y colección hemerográfica haciendo referencia al conjunto de los mismos. Se usan indistintamente los términos diarios, prensa y colección hemerográfica porque entendemos, como Abadal y Guallar (2010), que prensa (término genérico) hace referencia al conjunto de diarios (término específico) y son, por tanto, términos compatibles que no comportan problemas de uso indistinto.

La iniciativa que se analiza tiene el valor de ser una de las pocas de nuestro entorno mantenidas en el tiempo a lo largo de tantos años, junto a los proyectos desarrollados por la Facultad de Ciencias de Información de la Universidad Complutense de Madrid, por la Biblioteca Nacional de España o por la Biblioteca de la Universidad de Las Palmas de Gran Canaria.

### **3.1. EL CONSORCIO DE DIGITALIZACIÓN DE PRENSA VASCA**

Alrededor de 1990 varias bibliotecas vascas comienzan a plantearse la puesta en marcha de un proyecto cooperativo de conservación de la prensa diaria que adquieren, sirven y almacenan. Se busca mitigar la gran cantidad de espacio requerido en los depósitos de sus instalaciones para la custodia de los periódicos en papel y, en paralelo, hacerse con una copia de respaldo que garantice el acceso a la información, de la manera más fiel posible, en caso de pérdida de los originales. El proyecto se inicia en 1991 (Urkia, 2008). En los primeros años, entre 1991 y 1995, se microfilman ocho cabeceras de prensa diaria editadas en el País Vasco y Navarra. A partir del uno de enero de 1996 se comienza la digitalización de las mismas (Gómez, 1997), abandonando la microfilmación. Paulatinamente otras instituciones y otros títulos se irán sumando al proyecto.

Como resultado de la experiencia que la Fundación Sancho el Sabio (FSS) había adquirido en procesos de digitalización, fruto de sus tempranas incursiones en la materia de la mano de la empresa Archivos y Bibliotecas del grupo Informática El Corte Inglés (IECISA), el Consorcio de Digitalización de Prensa Vasca decide apostar por los últimos avances tecnológicos en materia de reproducción digital para su programa de conservación de diarios. En aquellos años, IECISA era líder absoluto del mercado en conversión digital, en virtud del *know how* acumulado durante las digitalizaciones masivas del proyecto del Archivo de Indias de Sevilla.

Es importante señalar que el Consorcio carece de documentación formal que recoja de modo explícito, mediante estatutos, reglamento, normativa, política escrita, carta o cualquier otra forma, su misión y objetivos o sus políticas y procesos de trabajo. A falta de tal documentación, solamente las actas de reunión y otros documentos menores de gestión (como presupuestos, correspondencia, correos electrónicos, relaciones, hojas de cálculo, etcétera) permiten extraer información sobre la finalidad de la constitución del CDPV o sobre las actividades y evolución del mismo. Las menciones que se hacen del Consorcio en unas pocas publicaciones permiten obtener cierta información en este mismo sentido.

#### **3.1.1. FINES. INTEGRANTES. AUSENCIA DE PERSONALIDAD JURÍDICA**

El tradicional recurso a la microfilmación es una alternativa altamente interesante para afrontar el problema de la reserva de grandes cantidades de espacio en las instalaciones, sin embargo se trata de un proceso caro. Varias bibliotecas que se encuentran ante idéntica

situación, pronto descubren que están interesadas en la conservación de los mismos títulos y que es factible cooperar para asumir de modo compartido los costes derivados del proceso de transformación de los diarios. Por una parte se comparten los gastos económicos derivados de la transformación de las cabeceras y por otra parte, se asume la conservación cooperativa de los diarios en papel. La conversión de los documentos analógicos a formato digital, dejando de lado la microfotografía, también es un proceso económicamente costoso y que tampoco evita tener que conservar los originales escaneados, es por ello que se continuará cooperando para obtener las copias digitales.

En los años 2003 y 2004, momento en que se debate en el seno del Consorcio cierto cambio de orientación en relación con la operativa de trabajo, algunas actas recogen la preocupación de sus miembros por mantenerse fieles al objetivo que inicialmente se habían marcado. En ellas aparece claramente expresada la finalidad del proyecto<sup>5</sup>.

Hasta dieciséis instituciones han llegado a formar parte del Consorcio en diferentes momentos de su trayectoria: un archivo histórico, tres bibliotecas universitarias, dos centros de documentación, cuatro bibliotecas públicas o redes de bibliotecas públicas, dos bibliotecas especializadas, dos bibliotecas patrimoniales, una biblioteca nacional y un servicio central de lectura pública. En la Figura 5 se presenta una vista diacrónica de todas las entidades que han integrado el CDPV en los diversos momentos a lo largo de su historia.

Al comienzo de su andadura, a mediados de los años noventa del siglo veinte, fue cuando menos miembros participaban en el proyecto, siendo ocho las organizaciones que pusieron en marcha el Consorcio en aquellos primeros momentos. Entre enero y septiembre de 2003 hasta catorce organizaciones llegan a formar parte del CDPV. En los años siguientes cinco bibliotecas abandonan el proyecto cooperativo y tan sólo una se suma al mismo. Desde 2008 y hasta hoy en día, son diez las bibliotecas que están apostando por seguir adelante con la iniciativa.

A lo largo de estos años el proyecto ha llegado a tener carácter internacional con la participación de la Biblioteca de la Universidad de Nevada (con sede en Reno, Nevada, Estados Unidos) a través de su *Basque Studies Program*. También consiguió sumar la importante participación de la Biblioteca Nacional de España durante ocho años. Sin embargo, actualmente, tras el abandono del proyecto por parte de los centros de documentación y de los archivos, y tras la salida de la Biblioteca de la Universidad de Navarra, el proyecto se restringe hoy al ámbito de la Comunidad Autónoma de Euskadi y está conformado únicamente por bibliotecas. El Anexo I recoge información detallada, extraída de las actas de Consorcio, sobre la entrada y abandono de cada una de las instituciones.

Una de las dificultades que el Consorcio no ha podido superar en sus largos años de andadura, es la de constituirse como entidad con personalidad jurídica propia (Urkia, 2008). Varios servicios jurídicos de las organizaciones (públicas en su mayoría) de las cuales las bibliotecas, los archivos y los centros de documentación dependen, han realizado diferentes propuestas de

---

<sup>5</sup> “(...) no se debe perder el espíritu conservacionista que anima al Consorcio” (acta de 08/04/2003); “(...) el objetivo del Consorcio no es otro que la conservación de los fondos hemerográficos tal y como se editaron en soporte papel (...)” (acta de 14/01/2004); “(...) [se ha mantener] el espíritu fundacional del Consorcio en cuanto a conservar la integridad de la información” (acta de 04/02/2004).

estatutos, pero siempre algún otro servicio jurídico ha informado en contra o ha planteado dificultades insalvables para poder culminar la constitución de la nueva entidad.

Este hecho ha provocado notables inconvenientes, derivados de la incapacidad de obrar del Consorcio, no pudiéndose cumplir con eficacia tareas administrativas de gestión tales como emitir facturas, tener cuentas corrientes, firmar documentos con valor legal, etcétera. Consiguientemente, se han venido generando múltiples ralentizaciones en la ejecución de las tareas y se han tenido que dedicar no pocas horas a la búsqueda de alternativas viables para poder cumplir su cometido.

La falta de personalidad jurídica ha causado situaciones graves como la recogida en las actas 02/2007 (de 13/06/2007) y 03/2007 (de 14/11/2007), donde se debate nuevamente sobre la problemática derivada de la incapacidad de obrar, incapacidad de emitir facturas en este caso, provocando el abandono del Consorcio por parte de la Biblioteca Nacional de España.



Figura 5.- Integrantes del Consorcio de Digitalización de Prensa Vasca, con fechas de incorporación y abandono del mismo.

Todavía muy recientemente, en diversas reuniones, el tema reaparece continuamente como expresión de las dificultades existentes para poder trabajar con eficiencia: “FSS insiste en la necesidad de dar entidad jurídica al Consorcio” (acta 02/2008 de 04/12/2008); en acta 02/2011 de 28/09/2011 se menciona nuevamente la necesidad de dotarse de entidad jurídica y capacidad de obrar. En algún momento, también se ha planteado la posibilidad de trabajar bajo el modelo de convenio, pero este tipo de iniciativas, por las razones que sea, tampoco han cuajado.

A pesar de todo ello, hay que destacar el hecho de que el Consorcio sea un ejemplo de proyecto colaborativo entre administraciones de distintos niveles que gozan de plena autonomía organizativa y económica: local, foral, autonómica, universitaria, instituciones privadas, organismos autónomos...

En torno al año 2003 se llega a plantear que el Consorcio carece incluso de nombre, no ya oficial, sino hasta para uso interno en el seno del propio Consorcio y de las organizaciones participantes. En reuniones de ese mismo año (actas de 08/04/2003 y de 30/09/2003) se constata esta falta de denominación, no pudiéndose llegar a un consenso para establecer un nombre de uso interno. Algunos aducen la falta de personalidad jurídica como traba para dotar siquiera de nombre a la iniciativa, ya por entonces con más de dos lustros de trabajo a sus espaldas.

### 3.1.2. MARCO LEGAL. MANDATO. DEPÓSITO LEGAL

Todas las bibliotecas participantes en el Consorcio tienen como misión, en cuanto que instituciones de la memoria, conservar los fondos documentales propios para dar servicio a su comunidad de usuarios. En este sentido, todas consideran la prensa diaria y las colecciones hemerográficas retrospectivas como un recurso valioso. Además, todas las bibliotecas prestan siempre especial mimo al cuidado de la colección local. Se entiende, por tanto, como lógica y normal la decisión adoptada por estas instituciones de conservar los diarios locales. Ello entronca directamente con la misión y el mandato de las propias instituciones.

Sin embargo, otro asunto distinto es el mandato o imperativo legal. Esto es, qué institución o instituciones están obligadas por ley a preservar los diarios que se vienen conservando. Lo que parece claro es que el Consorcio de Digitalización de Prensa Vasca, que incluso carece de personalidad jurídica propia, no lo está. Otra cosa es si alguno de sus miembros tiene tal responsabilidad, aunque sea de modo parcial o solamente sobre algunas de las cabeceras.

Indiscutiblemente, para establecer responsabilidades y determinar a quién obliga la ley por mandato expreso se ha de acudir a la legislación sobre depósito legal. Pero el primer problema viene derivado del ámbito territorial de los diarios conservados. El sentido de “colección local” que se viene aplicando, como ya ha quedado dicho, rebasa el territorio de la comunidad autónoma vasca y abarca todo el ámbito cultural del euskera. Se trata, por consiguiente, de cabeceras sujetas tanto a la legislación francesa como a la legislación española sobre depósito legal. Incluso dentro del ámbito español hay cabeceras de dos comunidades autónomas diferentes, por lo cual es evidente que hay diferentes responsabilidades que recaen sobre administraciones públicas distintas en lo tocante a la conservación por depósito legal de esas cabeceras.

La actual norma española, *Ley 23/2011 de depósito legal*, en su artículo 4, apartado 3, párrafo c), establece que los diarios son publicaciones objeto de depósito legal. El artículo 9, en su apartado 2, marca como “centros de conservación” a la Biblioteca Nacional de España y a aquellos que determinen las Comunidades Autónomas en el ámbito de sus competencias. Esta misma ley en su artículo 10, apartado 4, párrafo c), obliga a la BNE a conservar al menos dos ejemplares de los diarios.

Actualmente, en la BNE se están conservando ejemplares de los diarios del País Vasco y Navarra impresos en papel, que además están siendo digitalizados y microfilmados con fines de preservación.

En la Comunidad Autónoma de Euskadi la competencia de depósito legal recae sobre el actual Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco. El Departamento, en la guarda de monografías y otros materiales diferentes de los diarios, comparte la custodia de los ejemplares con las bibliotecas de los Territorios Históricos de Álava, Bizkaia y Gipuzkoa en virtud de un acuerdo establecido a tal fin.

La conservación de los diarios, por otra parte, se está realizando por medio del proyecto del Consorcio de Digitalización de Prensa Vasca. En Euskadi no se ha desarrollado legislación relativa al depósito legal. En la medida en que tampoco existe convenio formal, contrato, ni ninguna forma de oficialización de las relaciones establecidas en torno al Consorcio, puede

decirse que las propias bibliotecas integrantes del mismo han asumido *motu proprio* el mandato de conservar los diarios locales.

En el caso navarro la competencia recae sobre la actual Consejería de Cultura, Turismo y Relaciones Institucionales del Gobierno de Navarra. Recientemente se ha aprobado el *Decreto Foral 33/2013, de 29 de mayo, por el que se regula el Depósito Legal en la Comunidad Foral de Navarra* que establece que el “centro de conservación” es la Biblioteca de Navarra. Los editores de prensa están obligados a entregar cuatro ejemplares de los diarios editados en navarra al centro depositario del depósito legal en Navarra, el cual circulará dos ejemplares a la BNE y los otros dos a la Biblioteca de Navarra.

El decreto foral incluye en su artículo 3 *Materiales en vascuence editados fuera de la Comunidad Foral de Navarra* la siguiente disposición: “Los editores y, en su defecto o en su caso, el productor, impresor, estampador o grabador que produzcan materiales en vascuence en cualquier territorio del Estado y no tengan su domicilio, residencia o establecimiento permanente en la Comunidad Foral de Navarra, deberán depositar un ejemplar en la Biblioteca de Navarra.” Tal requisito va a permitir que el diario *Berria*, único editado íntegramente en euskera, contará con un ejemplar de conservación en la Biblioteca de Navarra, a pesar de ser editado en Gipuzkoa.

Finalmente, para el caso de los diarios vasco-franceses se ha de tener en cuenta la legislación francesa, que convierte a la Bibliothèque Nationale de France en depositaria del depósito legal de aquel país. El depósito legal se rige en la actualidad por el Título III del Libro I del Código del Patrimonio (parte legislativa, codificada en 2004; parte normativa, codificada en 2011) y sus decretos de aplicación, publicados entre 1995 y 1997 y modificados en 2006. La legislación obliga a hacer entrega de dos ejemplares de los diarios. En el Departamento de Pirineos Atlánticos (Aquitania) la biblioteca encargada de la recogida del depósito legal es la Bibliothèque Municipale Classée (BMC) de Burdeos. Por tanto, los diarios vascos de *Iparralde* están siendo conservados en la citada biblioteca. Se desconoce si se están aplicando políticas de digitalización con fines de conservación en el caso de los diarios en Francia.

### 3.2. CARACTERIZACIÓN DE LA COLECCIÓN MAESTRA DEL CONSORCIO DE DIGITALIZACIÓN DE PRENSA VASCA

El Consorcio de Digitalización de Prensa Vasca posee un gran banco de imágenes digitales ráster, producto de la digitalización sistemática de diarios mediante procesos de escaneado. Como resultado de este esfuerzo, mantenido a lo largo de los años, se ha conformado una notable colección –por su volumen y contenido– de ficheros máster y una segunda colección de ficheros de imágenes derivadas. El volumen de datos recopilado otorga al fondo de imágenes una posición destacada por su tamaño y por la cobertura exhaustiva de la prensa diaria editada en Euskal Herria en los últimos veinte años, y en algunos casos más allá. Hoy, se puede hablar de una colección hemerográfica digital consolidada, cuyas dimensiones continúan incrementándose ininterrumpidamente desde 1996. Varias instituciones han ido descabalgándose del proyecto y varias cabeceras han dejado de publicarse, sin embargo toda la prensa diaria vasco-navarra de este período se conserva a día de hoy gracias a esta iniciativa.



En las siguientes líneas se analiza y describe el fondo mediante la obtención de datos actualizados sobre la composición exacta del mismo –al máximo detalle posible– en lo tocante a formatos digitales, soportes de almacenamiento y volumen de datos. Se estima necesario contar con una descripción precisa que detalle los materiales que componen el fondo y sus características. Para ello se ha trabajado in situ con las propias copias originales de los ficheros máster, accediendo al contenido de los discos duros y CD-ROM. Se ha tenido acceso a las actas del CDPV para obtener detalle de las casas comerciales que ejecutaron en cada momento las labores de escaneado, así como para conocer las especificaciones técnicas de digitalización contratadas, los presupuestos aprobados y los costes, reales o estimados, finales.

La parte de los fondos almacenada en discos duros externos ha sido chequeada completamente. Esto es, se han verificado uno a uno todos los discos duros, de una de las dos copias existentes. La verificación ha consistido en comprobar el correcto acceso a los datos (que los discos funcionen), acceder aleatoriamente a diversas carpetas de ficheros y visualizar varios de ellos con visores de imágenes. También se ha accedido a los metadatos técnicos fotográficos EXIF mediante la herramienta ExifTool. Se han documentado tipos de ficheros y volumen de datos por disco. Se han contabilizado número de carpetas y número de ficheros por disco duro.

La parte de los fondos almacenada en CD-ROM ha sido chequeada por muestreo. El muestreo se ha realizado sobre el 24,94% de los soportes, esto es, sobre 463 de los 1856 CD-ROM que componen la colección. Se ha realizado mediante muestra probabilística, con un tamaño de error estándar máximo aceptable de 3,947%. Para el cálculo de la muestra sobre el universo total de CD-ROM se ha utilizado la herramienta informática STATS 2.0, incluida en el manual de 2010 de Hernández Sampieri *Metodología de la investigación*. A continuación se presentan los datos del muestreo:

- Tamaño del universo: 1856
- Error de muestreo: 3,947%
- Nivel de porcentaje estimado: 50%
- Nivel de confianza: 95%
- Tamaño de la muestra: 463

Durante la manipulación de los ficheros se han mantenido ciertas medidas de protección y seguridad a fin de evitar daños en la colección. Para la carga y visualización de los ficheros que lo han requerido, se ha utilizado un ordenador con antivirus. Asimismo, se ha procurado manipular los soportes cuidadosamente, reubicándolos en sus cajas y fundas una vez utilizados, observando cuidadosamente no errar en la identificación y ubicación del material.

Durante el trabajo de campo y en relación con el software utilizado, para el chequeo de los ficheros de imagen digital ráster se han empleado los visores y editores de imágenes Microsoft Office Picture Manager, Galería Fotográfica de Windows Vista, Visor de imágenes de Windows, Nero PhotoSnap, GIMP y Adobe Photoshop CSS6. Para la visualización de la información de metadatos relativa a los ficheros de imagen se ha usado principalmente ExifTool 9.30, pero también los visores y editores de imágenes arriba mencionados.

Los datos que se han recabado del análisis de los soportes de almacenamiento de los ficheros de imagen durante el trabajo de campo han sido registrados en una base de datos de Access

diseñada a tal efecto. Posteriormente, para el cálculo del tamaño de la colección se han utilizado hojas de cálculo con el programa Excel. Aunque pudiera resultar interesante aportar los datos contenidos en ambas herramientas, los límites de extensión de este Trabajo Fin de Máster impiden adjuntar tablas e informes como anexos, dado que su gran tamaño incrementaría mucho el número de páginas del trabajo.

### 3.2.1. *METODOLOGÍA DE DIGITALIZACIÓN Y PROCESO DE TRABAJO*

La digitalización de los diarios siempre se ha realizado a través de empresas subcontratadas a tal efecto por parte del Consorcio de Digitalización de Prensa Vasca. Han sido seleccionadas anualmente mediante libre concurrencia, a través de la presentación de proyectos presupuestados. Las empresas que han realizado los trabajos han sido Archivos y Bibliotecas AIE (1996-1997), Informática El Cortes Inglés S.A. (1998-hasta primer semestre de 2002), Tratamientos Micrográficos TRAMIC S.L.L. (segundo semestre de 2002 y 2003), Scanbit S.L. (2004), Gudat S.L. (2005-2011) y Scanbit S.L. (2012-2013). La empresa Archivos y Bibliotecas AIE estaba formada por Informática El Cortes Inglés S.A. e IBM España S.A., por lo cual se puede decir que en toda la primera época, hasta el segundo semestre de 2002, los trabajos fueron realizados por Informática El Cortes Inglés S.A (IECISA).

Apenas existe documentación que explicita la metodología de trabajo seguida, tanto porque no se han elaborado pliegos técnicos detallados para la contratación de los servicios por parte del CDPV, como porque no se han recibido habitualmente informes de entrega por parte de las empresas sobre los procesos ejecutados. Los únicos documentos que exponen información sobre métodos de trabajo son siempre los proyectos técnicos que acompañan a los presupuestos. Una vez seleccionado un presupuesto, no se conoce si el detalle de la metodología propuesta en tales proyectos fue finalmente llevado a cabo en el desarrollo de las labores contratadas, puesto que no se entregan memorias finales. En cualquier caso, parece que el proceso de trabajo ha sido muy similar a lo largo de los años.

El Consorcio solicita a las empresas unos requisitos técnicos mínimos a cumplir, normalmente relativos únicamente a formatos de fichero (tanto para la copia maestra como para las derivadas), resolución óptica de captura, uso del color (bitonal, escala de grises, paleta de colores o color real), soportes de entrega de los datos y número de copias. Tales indicaciones se comunican escuetamente por carta o por correo electrónico, y posteriormente aparecen recogidas en las propuestas técnicas y presupuestos económicos de los proveedores.

Los miembros del Consorcio realizan desde hace años un notable esfuerzo en el control de calidad del producto final entregado, verificando una a una que todas las páginas de los diarios estén digitalizadas, y comprobando, en algunos casos, que la calidad de entrega de las imágenes sea adecuada.

Todos los procesos de escaneo se han realizado directamente desde los periódicos en papel. Diariamente, las empresas contratadas adquieren la prensa que va a ser objeto de digitalización mediante escáner. En algunos casos, y dependiendo del modelo de la máquina, se guillotinan a fin de favorecer la carga automática de las hojas. No están documentadas en ninguna parte las marcas y modelos de escáneres utilizados por ninguna de las empresas a lo largo de los años (como se ha dicho, existen relaciones de equipos y máquinas en las

propuestas comerciales, pero no se sabe si son los finalmente utilizados). También se desconocen los programas de software utilizados para generar y manipular las imágenes, así como los programas y elementos de hardware usados para grabar los CD-ROM y discos duros externos. Por otra parte, los ficheros digitales carecen por completo de metadatos técnicos EXIF de identificación de cámara. Se desconoce si en la fase de captura el primer fichero generado ha sido algún tipo de RAW, TIFF o JPEG.

Parece ser que, por lo menos en algunos casos, una vez generadas las imágenes se procede a la corrección de las imágenes mediante operaciones automáticas básicas como el enderezado, la detección de hojas faltantes o errores de escaneo graves. Posteriormente, se generan los archivos maestros (si es que han sido capturados en formato distinto al solicitado por el CDPV) y los archivos derivados o versiones de explotación.

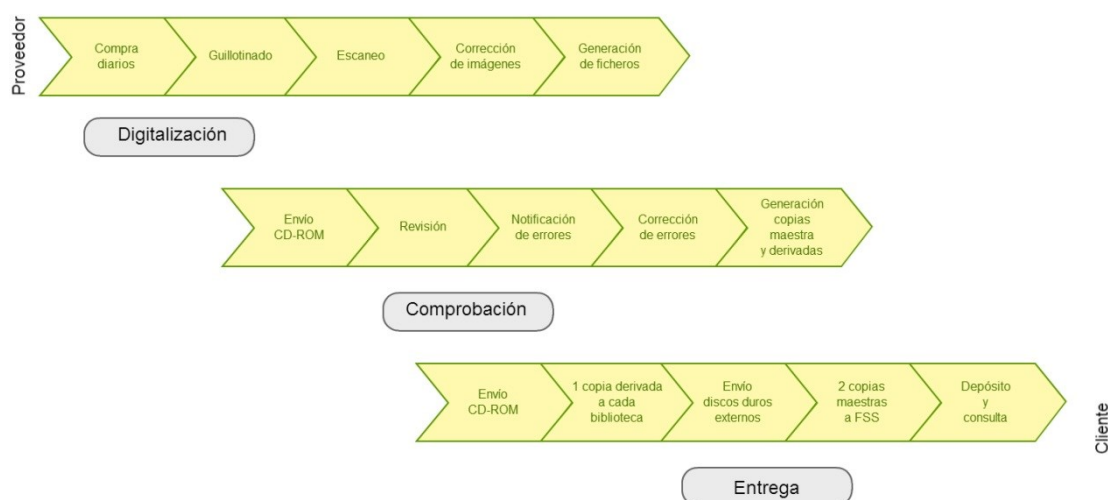


Figura 6.- Proceso de digitalización de diarios por el Consorcio de Digitalización de Prensa Vasca (Elaboración propia)

A partir de ese momento, la empresa realiza el envío de un CD-ROM por cabecera a cada una de las bibliotecas encargadas de su revisión, conteniendo la copia del mes natural completo. Tal copia, en los años más recientes, es en el formato requerido para la copia derivada (PDF). Normalmente, se revisa un diario por biblioteca, pero en algunos momentos, como actualmente, en que se digitalizan mayor número de diarios que bibliotecas integran el Consorcio, a alguna biblioteca le toca revisar más de un diario. La revisión consiste básicamente en comprobar que están digitalizadas todas las páginas del periódico. Se realiza visualizando en pantalla el diario frente al ejemplar en papel. Una vez comprobado el mes completo se notifican los errores a la empresa para que proceda a su corrección.

Una vez que la empresa recibe las notificaciones de errores y conformidades para todos los diarios del mes en curso, procede a corregir mediante nuevo escaneo o mediante la acción oportuna. Posteriormente, la empresa envía copias derivadas a cada una de las bibliotecas y entrega dos copias maestras a la FSS, que es quien custodia la versión máster. Los plazos de entrega tanto de las copias de uso de las bibliotecas como de las copias maestras han variado a lo largo de los años. En la última época viene siendo a año natural cerrado.

Los soportes de entrega han sido tradicionalmente CD-ROM de distintas calidades, como luego se detallará. Y desde 2005 discos duros externos de diversas marcas, modelos y capacidades.

### 3.2.1.1. PARÁMETROS DE DIGITALIZACIÓN EMPLEADOS

El fondo de imágenes ráster existente presenta características técnicas diversas como resultado de los distintos parámetros técnicos de digitalización aplicados a lo largo de estos años. Las especificaciones acordadas para la ejecución de las conversiones digitales han ido evolucionando hacia la aplicación de parámetros de mayor calidad, avanzando básicamente hacia el uso de mayores resoluciones de captura, uso del color, mejora en la edición de los archivos en formato PDF (uso de marcadores y multipágina) y evolución en el uso de los soportes de almacenamiento, pasando del uso de CD-ROM a discos duros externos. Así, los ficheros conservados presentan diversos parámetros en cuanto a resolución, color, formato y soportes de almacenamiento. La Tabla 1 presenta los parámetros de digitalización contratados en cada campaña anual. En las líneas blancas del cuadro están representados los valores contratados.

En el curso de la verificación mediante muestreo de los CD-ROM se han hallado valores diferentes de los contratados, tal y como puede comprobarse al consultar la Tabla 1. Las líneas azules de la tabla representan los valores hallados, destacándose en negrita aquellos casos donde hay diferencia entre lo contratado y lo hallado. Tal situación se da únicamente en la parte de la colección conservada en CD-ROM, en ningún caso ocurre en los archivos albergados en discos duros externos. Varios de los ficheros muestreados presentan cierta inconsistencia puesto que, en el mismo año y para el mismo diario, la resolución óptica de captura contratada de 200 ppp oscila realmente entre los valores de 96 ppp, 200 ppp, 220 ppp, 250 ppp, 270 ppp y 280 ppp. Se desconoce si tal situación se debe a la parametrización del escáner en el momento de la captura, a la fase posterior de generación de ficheros, a la fase de grabación de los soportes de almacenamiento o incluso a un posterior refresco de los soportes (realizado entre finales de 2006 y principios de 2007).

No se ha de pensar que estas resoluciones superiores a los 200 ppp contratados supongan una mejora en la calidad de las imágenes. Tal extremo sería cierto en el supuesto caso de que los valores mayores de 200 ppp se hubieran generado en la fase de captura, debido a una errónea parametrización al alza del escáner. Si el error se hubiese generado en alguno de los otros momentos (generación de ficheros, grabación de soportes o refresco) es probable que el aumento de resolución pueda haberse debido al uso de resolución interpolada.

La resolución interpolada disminuye la calidad final de las imágenes. Frente a la resolución óptica que contiene la información real que el dispositivo fotográfico ha capturado, convirtiendo la muestra de píxeles captados por el sensor en puntos de información, la resolución interpolada contiene puntos de información creados a partir de aplicar un proceso de interpolación que genera nuevos píxeles a partir de los píxeles ya existentes en la imagen. La resolución interpolada no aporta detalles de la imagen, no añade información real, sino que genera nuevos píxeles a partir de los valores de los píxeles adyacentes creando puntos de información que no están en el original.

Tabla 1.- Parámetros de digitalización contratados y realmente hallados en el muestreo (CD-ROM) y la verificación (HD)

	Empresa	Formato máster	Resolución máster	Color máster	Tamaño captura	Soporte	Formato derivado	Otras	Resolución derivado	Color derivado	Tamaño salida	Soporte
1996	Archivos y Bibliotecas	TIFF	200	Grises a 256	Original 1:1	CD-ROM	No					
		TIFF	200	Grises a 256	Original 1:1	CD-ROM	TIFF		200			CD-ROM dorado
1997	Archivos y Bibliotecas	TIFF	200	Grises a 256	Original 1:1	CD-ROM	No					
			No en muestra			CD-ROM	TIFF		200			CD-ROM dorado
1998	IECISA	TIFF	200	Grises a 256	Original 1:1	CD-ROM	No					
		TIFF	<b>200, 96</b>	Grises a 256	Original 1:1	CD-ROM	TIFF		200			CD-ROM Kodak dorado, Imation azul, HP dorado
1999	IECISA	TIFF	200	Grises a 256	Original 1:1	CD-ROM	No					
			No en muestra			CD-ROM	TIFF		200			CD-ROM HP verde, dorado
2000	IECISA	TIFF	200	Grises a 256	Original 1:1	CD-ROM	No					
		TIFF	<b>200, 96</b>	Grises a 256	Original 1:1	CD-ROM	TIFF		200			CD-ROM verde, dorado
2001	IECISA	TIFF	200	Grises a 256	Original 1:1	CD-ROM	No					
			No en muestra			CD-ROM	TIFF		200			CD-ROM verde
2002	IECISA/Tramic	TIFF	200	Grises a 256	Original 1:1	CD-ROM	No					
		TIFF	<b>200, 220, 250, 280</b>	Grises a 256	Original 1:1	CD-ROM	TIFF PDF 1.2	No etiquetado	200			CD-ROM verde, azul
2003	Tramic	TIFF	200	Grises a 256	Original 1:1	CD-ROM	No					
			No en muestra			CD-ROM	PDF 1.2	No etiquetado	200			CD-ROM azul
2004	Scanbit	TIFF	200	Grises a 256	Original 1:1	CD-ROM	?					
		TIFF	<b>200, 250, 270, 300</b>	Grises a 256	Original 1:1	CD-ROM	PDF 1.4	No etiquetado	200			CD-ROM azul
2005	Gudat	TIFF	200	Grises a 256	Original 1:1	HD	PDF		200	Grises = orig. o <		
		TIFF	200	Grises a 256	Original 1:1	HD	PDF 1.6	No etiquetado	200	Grises = orig. o <		CD-ROM azul
2006	Gudat	TIFF	200	Grises a 256	Original 1:1	HD	PDF		200	Grises = orig. o <		
		TIFF	200	Grises a 256	Original 1:1	HD	PDF 1.5		200	Grises = orig. o <		CD-ROM azul
2007	Gudat	TIFF	200	Grises a 256	Original 1:1	HD	PDF		200	Grises = orig. o <		
		TIFF	200	Grises a 256	Original 1:1	HD	PDF 1.6		200	Grises = orig. o <		CD-ROM azul
2008	Gudat	TIFF	200	Grises a 256	Original 1:1	HD	PDF	Marcadores	200	Grises = orig. o <		
		TIFF	200	Grises a 256	Original 1:1	HD	PDF 1.6	Marcadores	200	Grises = orig. o <		CD-ROM azul
2009	Gudat	TIFF	200	Grises a 256	Original 1:1	HD	PDF	Marcadores	200	Grises = orig. o <		
		TIFF	200	Grises a 256	Original 1:1	HD	PDF 1.6	Marcadores	200	Grises = orig. o <		CD-ROM azul
2010	Gudat	TIFF	200	Grises a 256	Original 1:1	HD	PDF	Marcadores	200	Grises = orig. o <		
		TIFF	200	Grises a 256	Original 1:1	HD	PDF 1.6	Marcadores	200	Grises = orig. o <		CD-ROM azul
2011	Gudat	TIFF	300	Color	Original 1:1	HD	PDF	Marcadores	300	Color = orig. o <		
		TIFF	300	Color	Original 1:1	HD	PDF 1.6	Marcadores	300	Color = orig. o <		CD-ROM azul
2012	Scanbit	TIFF	300	Color indexado	Original 1:1	HD	PDF	Multipágina	300	Color = orig. o <		
		TIFF	300	Color indexado	Original 1:1	HD	PDF 1.6	Multipágina	300	Color = orig. o <		CD-ROM azul
2013	Scanbit	TIFF	300	Color indexado	Original 1:1	HD	PDF	Multipágina	300	Color = orig. o <		
		TIFF	300	Color indexado	Original 1:1	HD	PDF 1.6	Multipágina	300	Color = orig. o <		CD-ROM azul

En relación con los formatos contratados y los realmente disponibles en la colección maestra, se han podido confirmar los siguientes extremos: se pide siempre TIFF a los proveedores, sin mayor concreción de versión, y se dispone realmente de TIFF versión T6/Group 4 Fax, y de JPEG en algunos casos; los derivados son solicitados en PDF, y están en tal formato, en las versiones PDF 1.2, PDF 1.4, PDF 1.5 y PDF 1.6.

### 3.2.1.2. OTRAS COLECCIONES

Durante los largos años que el CDPV ha estado digitalizando prensa, se han acercado hasta el mismo en diversos momentos empresas editoras de diarios y empresas informáticas ofertando colecciones de prensa digital y soluciones de gestión de contenidos digitales. En algunos casos, estas colecciones paralelas han servido para completar los fondos de algunas bibliotecas integrantes del Consorcio. Se llegó a pensar que la compra de la prensa digitalizada a los propios editores podría ser una buena solución en cuanto a calidad y precio como alternativa a las digitalizaciones del CDPV, que se planteó el abandono de las mismas. Sin embargo la carestía de precios, la escasa calidad de algunas publicaciones digitales y la baja permanencia e inestabilidad en el tiempo de las iniciativas comerciales diluyeron rápidamente tal idea.

En los últimos años del siglo XX y en los años iniciales del presente, fue habitual que los medios de prensa publicasen y comercializasen ediciones retrospectivas de sus cabeceras en soporte óptico. Abadal analizó con detalle en aquellos años esta cuestión (2001). El CDPV ha tenido la oportunidad de hacerse con otras colecciones (paralelas) de prensa vasca digitalizada.

Es el caso de la edición en CD-ROM que *El Correo Español-El Pueblo Vasco* comercializó, y que fue adquirida por varias bibliotecas mientras duró esta comercialización entre los años 2002 y 2005. En un único CD-ROM se vendían todas las ediciones del diario: Álava, Gipuzkoa, La Rioja, Miranda de Ebro y las diversas ediciones de Bizkaia (Costa, Duranguesado, Margen Derecha, Margen Izquierda y Nervión-Ibaizabal). El formato de los ficheros es PDF 1.4 en color, no etiquetado, y presenta la gran ventaja de estar con OCR.

El Diario de Navarra publicó con motivo de su centenario una edición histórica (1905-2004) de toda su colección en CD-ROM, que fue puesta a la venta en 2006 por 18.000 euros + IVA. En su momento fue ofertada a los miembros del Consorcio y un importante número de instituciones estuvieron interesadas en su adquisición, sin embargo lo elevado del precio disuadió de la compra a casi todas las bibliotecas.

Finalmente, hay que destacar también la colección realizada por la empresa Plazagunea S.L. por encargo del propio Consorcio. Esta empresa elabora y mantiene durante varios años una herramienta de software denominada *Hemeroteka+ Sistema Avanzado de Consulta de Prensa*, que contiene todas las cabeceras vascas digitalizadas, con OCR y servidas en una interfaz de interrogación muy avanzada, con búsquedas lematizadas especialmente adaptadas a las características morfológicas del euskera. La herramienta permite realizar dossieres (*clipping*) y distribuirlos por correo electrónico. Se trata de una herramienta de amplio uso y gran versatilidad que desgraciadamente dejó de tener soporte por su fabricante debido a la falta de explotación comercial. *Hemeroteka+* contiene la prensa digitalizada de los años 2004 a 2010. Arriazabalaga ha descrito brevemente las características y funcionalidades de esta herramienta (2009).

El hecho de que este software fuese realizado por encargo del CDPV y que sus miembros contratasen anualmente con Plazagunea S.L. la carga de datos en el sistema, ha propiciado que a pesar de la pérdida de mantenimiento de la aplicación los miembros del Consorcio cuentan a día de hoy con una copia en disco duro externo de otra colección de diarios digitalizados y con OCR, de todas las cabeceras vasco-navarras del periodo comprendido entre 2004 y 2010. El sistema, a pesar de no tener nuevas cargas de datos, sigue estando accesible por red mediante identificación para las bibliotecas que lo contrataron y continúan pagando una conexión anual.

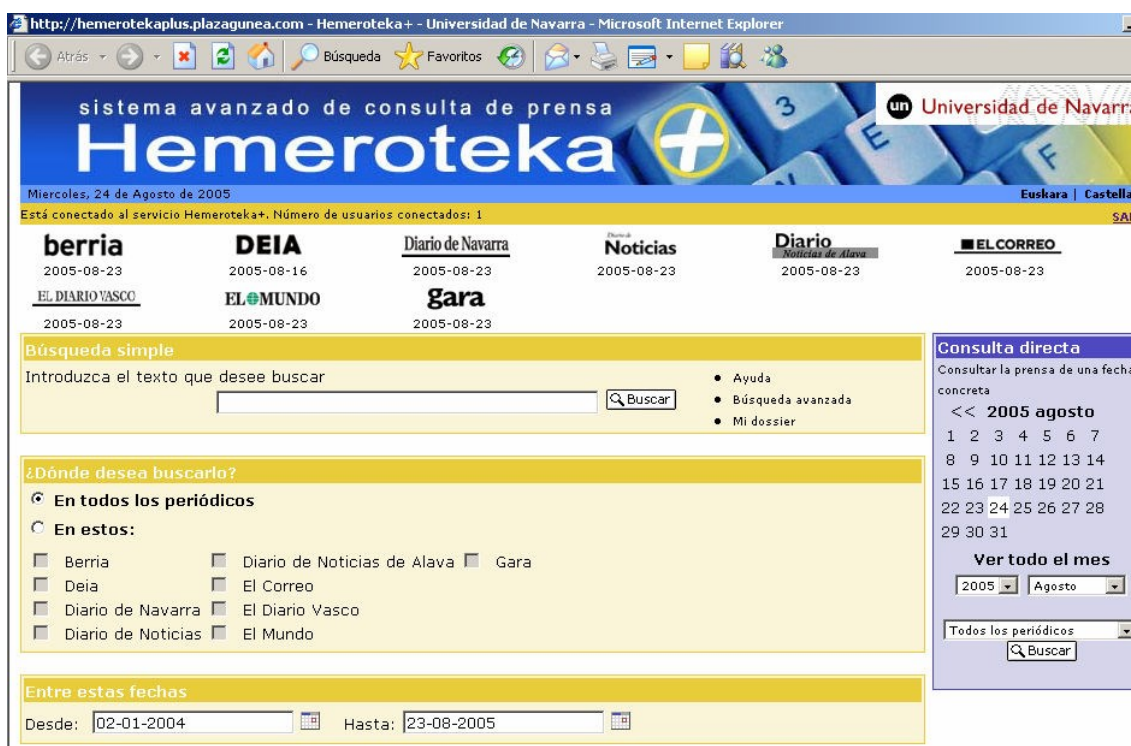


Figura 7.- Interfaz de búsqueda de Hemeroteca+ en la Biblioteca de la Universidad de Navarra

Finamente, cabe comentar que algunos editores también comercializaron en los últimos años del siglo XX sus diarios en microfichas, como por ejemplo *El País* entre 1980 y 2000, y algunas de las bibliotecas del Consorcio adquieren individualmente tales colecciones.

### 3.2.2. DIARIOS DIGITALIZADOS Y AÑOS CUBIERTOS

Cuando el primero de enero de 1996 se inicia el proceso de digitalización, comienzan a escanearse las ocho cabeceras que hasta ese momento se venían microfilmado. Se trata de los diarios: *Deia*, *Diario de Navarra*, *Diario de Noticias*, *Diario Vasco*, *Egin*, *Egunkaria*, *El Correo Español-El Pueblo Vasco* (edición Álava), y *El Mundo del País Vasco*.

Posteriormente, a medida que nuevos diarios van apareciendo en el mercado el CDPV emprende su digitalización. Son varias las cabeceras que se presentarán al público a partir de la fecha. *Gara* nace el 1 de febrero de 1999. *Euskadi Información* comienza a publicarse el 16 de julio de 1998. *Egunero* sale al mercado en febrero de 2003. *Berria* se publica desde junio de 2003. En mayo y en noviembre de 2004 aparecen *La Estafeta* y *Diario de Noticias de Álava*,

respectivamente. En la Tabla 2 se ofrece una vista diacrónica completa con todos los diarios digitalizados por el Consorcio de Digitalización de Prensa Vasca.

A menudo se decide comenzar a digitalizar cuando el periódico ya lleva unos meses publicándose, por lo cual en tales casos se decide hacer una digitalización retrospectiva desde el primer número. Tal es el caso de *El País, edición País Vasco* o de *El Periódico de Álava* que comienzan a digitalizarse en 1998, en ambos casos se hace desde el número uno aparecido en mayo de 1997 y diciembre de 1996, respectivamente. Del mismo modo se hace con el periódico *Zabalik* que comienza a digitalizarse en 1999 desde su primer número. Posteriormente y siguiendo el mismo criterio, en 2002 se decide digitalizar el periódico *Le Journal du Pays Basque* desde el número uno publicado en octubre de 2001. En 2006 se inicia la digitalización de *Diario de Noticias de Gipuzkoa*, haciendo una digitalización retrospectiva desde su aparición en noviembre de 2005.

Caso distinto es del periódico *El Correo Español, edición Bizkaia*, que se comienza a digitalizar el primero de enero de 1998, pues hasta ese momento sólo se hacía con la edición de Álava. La cabecera llevaba ya muchos años publicándose antes de comenzar con ella.

En sentido contrario, a medida que diferentes cabeceras dejan de publicarse o pasan de formato diario a revista se abandona su digitalización. *Egin* se cierra el 15 de julio de 1998. *Euskadi Información* deja de publicarse el 29 de enero de 1999. A partir de final del año 2002 se decide dejar de digitalizar el periódico *Zabalik* que ha pasado a ser revista de información general. *Egunkaria* se cierra en febrero de 2003. *Egunero* deja de publicarse en junio de 2003. *El Periódico de Álava* deja de publicarse en la primera quincena de julio de 2004. *La Estafeta* deja de publicarse en mayo de 2005.

Por otra parte, en el seno del Consorcio se plantea la posibilidad de abordar la digitalización retrospectiva completa de algunas cabeceras. Tal decisión se toma en virtud de la necesidad sentida de contar con colecciones completas, de la importancia de los diarios seleccionados en el devenir de la vida social vasca y en base al hecho de contar con una fecha de arranque para las digitalizaciones no muy lejana en el tiempo. Es por ello que se plantea la conversión de ciertos diarios aparecidos a partir de la ley sobre libertad de expresión de abril de 1977. Así, en acta de 16/04/1996 se decide abordar la digitalización retrospectiva del diario *Egin*, aparecido el 29 de septiembre de 1977. En otro momento dado, también se estima interesante acometer la digitalización retrospectiva (1977-1991) del diario *Deia*, aparecido el 8 de junio de 1977, pero finalmente tal digitalización retrospectiva no se lleva a cabo.

Sí se realizará la digitalización del diario *Egin* entre 1977 y 1991. Se ha de tener en cuenta que el Consorcio cuenta ya con microfichas de 1992 a 1995, y con digitalización de 1996 en adelante. Así que la colección de *Egin* quedará como sigue: 1977-1991 en CD-ROM; erratas de 1979 y 1980 en CD-ROM (Archivos y Bibliotecas AIE, empresa que hizo la digitalización, entregó posteriormente nuevos CD-ROM que completan la colección con la subsanación de las faltas halladas); 1992-1995 en microfichas; 1996-1998 en CD-ROM. Años después, en 2013, se generó una copia digital completa del diario, grabada esta vez en un disco duro externo. El Anexo II recoge la evolución de las cabeceras digitalizadas por el Consorcio de Digitalización de Prensa Vasca con mayor detalle.



Tabla 2.- Cabeceras digitalizadas y años cubiertos

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Retrospectiva
Berria								✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Desde nº1 junio 2003
Deia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Diario de Navarra	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Diario de Noticias	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Diario de Noticias de Álava									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Desde nº1 noviembre 2004
Diario Noticias Gipuzkoa										✓	✓	✓	✓	✓	✓	✓	✓	✓	Desde nº1 octubre 2005
Diario Vasco	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Egin	✓	✓	✓																Completa, 1977-1998
Egunero								✓											Desde nº1 febrero 2003
Egunkaria	✓	✓	✓	✓	✓	✓	✓	✓											
El Correo, edición Álava	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
El Correo, edición Bizkaia			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
El Mundo del País Vasco	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
El País, edición País Vasco		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Desde nº1 mayo 1997
El Periódico de Álava	✓	✓	✓	✓	✓	✓	✓	✓	✓										Desde nº1 diciembre 1996
Euskadi Información			✓	✓															Desde nº1 julio 2008
Gara				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Desde nº1 febrero 1999
La Estafeta									✓	✓									Desde nº1 mayo 2004
Le Journal du Pays Basque						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Desde nº1 octubre 2001
Zabalik				✓	✓	✓	✓												Desde nº1 noviembre 1999

Tabla 3.- Cabeceras en microfichas

	Deia	Diario de Navarra	Diario de Noticias	Diario Vasco	Egin	Egunkaria	El Mundo del País Vasco	Navarra Hoy
1992	✓	✓		✓	✓	✓	✓	✓
1993	✓	✓		✓	✓	✓	✓	✓
1994	✓	✓	✓	✓	✓	✓	✓	✓
1995	✓	✓	✓	✓	✓	✓	✓	

Finalmente, a fin de poder apreciar globalmente la colección de diarios completa que el CDPV posee, se adjunta también, a título informativo, la Tabla 3 que contiene la relación completa de la prensa reproducida en microfichas por el Consorcio de Digitalización de Prensa Vasca.

### 3.2.3. VOLUMEN DE DATOS Y TAMAÑO DE LA COLECCIÓN

La colección de ficheros máster del Consorcio de Digitalización de Prensa Vasca está compuesta a fecha de hoy por 83 discos duros y por 1.856 CD-ROM. En ambos casos existe una segunda copia de respaldo o *backup*. Dado que ambas copias están almacenadas juntas en la misma ubicación –en los mismos depósitos y en estanterías correlativas–, el número de discos en las dependencias de la Fundación Sancho el Sabio alcanza los 166 discos duros y los 3.712 CD-ROM. Tales soportes de almacenamiento conservan imágenes relativas a diarios entre 1996 y 2012.

En la Tabla 4 se presenta un estadillo del número de CD-ROM que hay por cada cabecera digitalizada y una estimación del tamaño que ocupa la información.

Las imágenes ráster contenidas en los CD-ROM alcanzan un tamaño de 831,17 GB o 0,812 TB<sup>6</sup>. La estimación del tamaño que ocupa la información en el total de la colección de CD-ROM se ha realizado mediante la extrapolación de los datos de los 463 CD-ROM que representan el tamaño de la muestra analizada. La extrapolación de datos se ha realizado título a título, dado que cada diario se edita con un número de páginas más o menos estable en similares periodos mensuales o semanales a lo largo del año, habiendo sin embargo diferencias notables entre títulos. Por ello, el cálculo del tamaño para cada cabecera será más aproximado partiendo de los datos reales de lo que ocupa la digitalización de la propia cabecera en etapas del año idénticas. Así por ejemplo, para el cálculo del tamaño de los 170 CD-ROM que ocupa el Diario Vasco entre enero de 1996 y diciembre de 2004 se han extrapolado los datos de los 42 discos verificados (muestra del 24,71%), tomados en los meses de septiembre a diciembre de 1996, 1998, 2000, 2002 y 2004. El valor obtenido mediante una simple regla de tres no ha sido corregido con ningún factor de ponderación.

Hay que considerar por otra parte, que la colección en CD-ROM contiene los diarios editados entre 1996 y 2004, periodo previo a la crisis actual de los medios de prensa en que se viene observando un descenso paulatino e importante del número de páginas publicadas, que está decayendo constantemente en los últimos años. Se estima que el periodo analizado es más o menos estable, aún no afectado por la crisis de los medios, y que los datos de unos años son extrapolables a los años colindantes.

Se presentan a continuación varios datos en relación con el tamaño de la colección, como son:

- Número total de discos duros externos.
- Tamaño total en discos duros externos.
- Tamaño total de la colección.

---

<sup>6</sup> En realidad, al tratarse de tamaños expresados en base binaria, deberían indicarse como 831,17 Gibibytes (GiB) o 0,812 Tebibytes (TiB), tal y como se explica en el apartado 4 del presente trabajo. Todas las cantidades expresadas en este apartado lo son en base binaria.

- Número de páginas digitalizadas por año.
- Número de carpetas por año.

Tabla 4.- Número de CD-ROM por cabecera, años y tamaño estimado de los datos

CABECERAS	Nº cedés	Periodo cubierto	Tamaño GB
<i>Beria</i>	33	Junio 2003-2004	9,74
<i>Deia</i>	154	1996-2004	76,17
<i>Diario de Navarra</i>	154	1996-2004	81,11
<i>Diario de Noticias de Álava</i>	4	Nov. 2004-Dic. 2004	1,61
<i>Diario de Noticias</i>	150	1996-2004	72,09
<i>Diario Vasco</i>	170	1996-2004	83,19
<i>Egin</i>	31	1996-Julio 1998	11,42
<i>Egunero</i>	5	Febrero 2003-Junio 2003	1,01
<i>Egunkaria</i>	106	1996-Febrero 2003	39,48
<i>El Correo Español-El Pueblo Vasco, ed. Álava</i>	206	1996-2004	101,20
<i>El Correo Español-El Pueblo Vasco, ed. Bizkaia</i>	175	1998-2004	81,00
<i>El Mundo del País Vasco</i>	177	1996-2004	83,38
<i>El País, edición País Vasco</i>	156	1997-2004	78,48
<i>El Periódico de Álava</i>	128	Diciembre 1996-Julio 2004	50,42
<i>Euskadi Información</i>	7	Julio 1998-Enero 1999	1,61
<i>Gara</i>	113	1999-2004	46,14
<i>Le Journal du Pays Basque</i>	44	Octubre 2001-2004	9,64
<i>Zabalik</i>	43	Nov. 1999-Dic. 2002	3,47
<b>TOTAL</b>	<b>1.856</b>	<b>1996-2004</b>	<b>831,17</b>

### 3.2.3.1. NÚMERO TOTAL DE DISCOS DUROS EXTERNOS

El número total de discos duros externos es de 83, que se reparten conforme a la relación que aparece en la Tabla 5. Cada año se realiza un disco duro externo por mes, por lo cual lo normal es que haya 12 discos duros por año. Conviene, no obstante, indicar que hay 7 discos duros del año 2010 porque el primero contiene el mes de enero, el siguiente disco febrero y el resto de discos contienen dos meses cada uno: marzo-abril, mayo-junio, julio-agosto, septiembre-octubre, noviembre-diciembre. Hay 11 discos duros del año 2011, uno por mes natural, excepto noviembre y diciembre que están en un único disco. Y finalmente, hay 3 discos duros externos del año 2012, cada uno de los cuales agrupa un cuatrimestre.

Aparte de estos, se debe mencionar que existe un disco duro que contiene el diario *La Estafeta* entre mayo y diciembre de 2004. También existe otro disco duro externo que contiene el diario

Egin completo, de septiembre de 1977 a julio de 1998, tal y como ya se ha indicado previamente en el apartado 3.2.2.

Tabla 5.- Relación de discos duros

Año	Nº discos duros	Cabeceras
2005	12	Todas
2006	12	Todas
2007	12	Todas
2008	12	Todas
2009	12	Todas
2010	7	Todas
2011	11	Todas
2012	3	Todas
Mayo a diciembre de 2004	1	La Estafeta
Septiembre 1977 - Julio 1998	1	Egin
<b>TOTAL</b>	<b>83</b>	

### 3.2.3.2. TAMAÑO TOTAL EN DISCOS DUROS EXTERNOS

Este valor se obtiene realizando el cálculo mediante dos métodos distintos. Como se observará, los valores obtenidos mediante ambos métodos no son idénticos, pero sí muy cercanos por lo cual se estima que la precisión de la cifra presentada es alta.

El primer método ha consistido en verificar el contenido de cada disco duro, cargando disco a disco, y obteniendo la información sobre el volumen ocupado del mismo mediante las propiedades de archivo a través de la herramienta "Mi PC" de Windows. Los valores obtenidos por el primer método son:

$$31.111.675,59 \text{ MB} = 30.382,50 \text{ GB} = \mathbf{29,67 \text{ TB}}$$

El segundo método ha consistido en tomar los datos de la pegatina adosada a cada disco duro<sup>7</sup>, donde están referidos los datos obtenidos por la empresa subcontratada en el momento de la grabación del disco duro. Se desconoce la herramienta utilizada para la grabación (no está documentada en ninguna parte) pero los datos aparecen claramente reflejados como *source bytes read* en el momento de verificar si el desarrollo de la grabación ha sido satisfactoria. Los valores obtenidos por el segundo método son:

$$25.413.419,93 \text{ MB} = 24.817,79 \text{ GB} = \mathbf{24,24 \text{ TB}}$$

---

<sup>7</sup> Años 205-2011. Los discos de 2012 carecen de pegatina. El tamaño por el primer método descontado 2012 es **24,68 TB**

Para una posible comparativa del volumen de datos contenido en los discos duros externos frente al volumen de los datos contenido en CD-ROM se han de tener en cuenta varios factores. En primer lugar se ha de cotejar el número de diarios digitalizados por año. También las variaciones anuales en número de páginas con que los diarios salen publicados van a generar disparidades en el número de las imágenes y por tanto en el tamaño final almacenado. Pero sobretodo, es la variable de la resolución de captura la que va a determinar el tamaño o peso de los archivos. Los datos han de ser entendidas conforme a estas variables. Los datos grabados en HD se refieren al número de diarios relacionados en la Tabla 6.

Tabla 6.- Número de cabeceras digitalizadas por año en HD

Año	2005	2006	2007	2008	2009	2010	2011	2012
Nº títulos	15	13	13	13	13	13	13	13

La colección de discos duros contiene además la digitalización completa del diario *Egin* entre 1977 y 1998. Sólo este diario ocupa 265 GB. Otra variable que se ha de contemplar para una posible comparativa es que los discos duros entre 2005 y 2011 contienen tanto las imágenes máster en TIFF como las imágenes derivadas en PDF. Los CD-ROM solamente contienen las imágenes máster TIFF.

Se presentan en la Tabla 7 los datos relativos al número de cabeceras cuyas imágenes digitalizadas se han grabado en CD-ROM.

Tabla 7.- Número de cabeceras digitalizadas por año en CD-ROM

Año	1996	1997	1998	1999	2000	2001	2002	2003	2004
Nº títulos	8	8	12	12	12	12	13	12	14

### 3.2.3.3. TAMAÑO TOTAL DE LA COLECCIÓN

El tamaño total estimado para la colección completa se obtiene mediante la suma de los datos obtenidos por el primer método del apartado anterior (referidos a los años 2005-2012 e HD) más los datos obtenidos mediante muestreo para los años 1996-2004 y almacenados en CD-ROM. Los valores obtenidos son:

$$30.382,50 \text{ GB} + 831,17 \text{ GB} = 31.195,66 \text{ GB}$$

$$29,67 \text{ TB} + 0,81 \text{ TB} = \mathbf{30,48 \text{ TB}}$$

Para poder valorar adecuadamente el volumen de datos alcanzado por el CDPV merece la pena poner estos datos en contexto, realizando una simple comparativa con el tamaño de otros proyectos. El archivo de Internet *Padicat*, realizado por la Biblioteca de Catalunya “contiene 39.587 capturas de 118.039 páginas web y está formado por 249 millones de ficheros, con un tamaño de 7,5 terabytes (datos a 15-05-2011)” (Lluca y Cócera, 2012). En datos más recientes, se contabilizan 249.609 URLs, 349 millones de ficheros y 13 terabytes espacio para este mismo

proyecto (Grupo de Trabajo de Patrimonio Digital, 2013). El proyecto de archivo web vasco *Ondarenet*, llevado adelante por el Gobierno Vasco, por su parte alcanza 16.943 URLs, 25.195.236 archivos y 824,04 GB espacio (Grupo de Trabajo de Patrimonio Digital, 2013; “Ondarenet, patrimonio digital vasco,” 2013). El proyecto de captura de la web llevado adelante por la BNE ha recolectado 1.780 millones archivos, alcanzando un tamaño de 85 TB (Grupo de Trabajo de Patrimonio Digital, 2013). Como puede apreciarse, los 30,48 TB que alcanza el fondo del CDPV lo convierten en un proyecto de tamaño mediano.

#### 3.2.3.4. NÚMERO DE PÁGINAS DIGITALIZADAS POR AÑO

Otro valor interesante para conocer el tamaño de la colección es el número de páginas digitalizadas por año. Los datos parciales con que se cuenta se han obtenido de dos fuentes. Por una parte, aparecen datos para tres años recogidos en un acta de reunión del Consorcio. Se desconoce el método de obtención, pero son datos reales para dos años y estimación de datos para el tercero. Por otra parte, están los datos de la pegatina adosada a cada disco duro, donde están referidos los datos obtenidos por la empresa subcontratada en el momento de la grabación del disco duro externo. Aparecen reflejados como *Check files* (archivos analizados) en el momento de verificar si el desarrollo de la grabación ha sido satisfactorio.

Los datos de los años 1998, 1999 y 2000 aparecen en el acta de 01/03/2001. Se trata de 330.459 páginas digitalizadas en 1998, 364.517 páginas digitalizadas en 1999 y una previsión de 393.678 páginas a digitalizar en 2000.

Los datos de los años 2005 a 2011 hacen referencia a la grabación en los HD de 3.397.656 archivos de imagen. Como ya se ha comentado se ha de tener en cuenta el número de cabeceras, el hecho de que aparece *Egin* completo y el hecho de que están grabados conjuntamente archivos máster y derivados.

#### 3.2.3.5. NÚMERO DE CARPETAS POR AÑO

Finalmente, también están disponibles los datos del número de carpetas (*folders*) por años para todas las cabeceras en los años 2005 a 2011. Las carpetas son los directorios donde se guardan los ficheros de imagen digital ráster en las unidades de almacenamiento (bien CD-ROM o bien discos duros externos, en el caso que nos ocupa). Los datos una vez más se han obtenido de la pegatina adosada a cada disco duro externo, donde están referidos los datos obtenidos por la empresa subcontratada en el momento de la grabación. Aparecen reflejados como *Check folders* (carpetas analizadas) en el momento de verificar si el desarrollo de la grabación ha sido satisfactorio.

Los datos de los años 2005 a 2011 hacen referencia la grabación en los HD de 123.395 carpetas conteniendo archivos de imagen. Como ya se ha comentado se ha de tener en cuenta el número de cabeceras, el hecho de que aparece *Egin* completo y el hecho de que están grabados conjuntamente archivos máster y derivados.

#### 3.2.4. ANÁLISIS DE COSTES ECONÓMICOS

La metodología de trabajo seguida por el CDPV para proteger y guardar el material hemerográfico contemporáneo, mediante la corriente de *digitization for preservation*, genera

en varias etapas del proceso técnico una serie de costes identificables. En algunos casos, estos costes también son cuantificables.

La esencia del proyecto del Consorcio es de naturaleza cooperativa pero el hecho de carecer de personalidad jurídica y de no tener, por tanto, capacidad de obrar en general ni de facturar en particular, provoca que los costes económicos reales se facturen siempre a cada una de las bibliotecas integrantes del mismo. El cargo y reparto de costes fue solidario y a partes iguales entre cada uno de los miembros del Consorcio en los primeros años de funcionamiento del mismo. Sin embargo, pronto cada organización asume las facturas derivadas de la conversión digital de los diarios por los cuales está interesada, no pagando por la digitalización de aquellos otros diarios que no le interesen, que son costeados por aquellas bibliotecas que los demandan.

En general, puede decirse que el mero principio de economía de escala ha favorecido la reducción de los costes a asumir por cada parte. El hecho de cooperar en la contratación de la digitalización ha provocado una minoración de costes.



*Figura 8.- Ciclo de costes del proyecto del Consorcio de Digitalización de Prensa Vasca (Elaboración propia)*

Los costes habidos hasta el momento pueden ser clasificados del siguiente modo:

- Adquisición de diarios.
- Conversión a formato digital.
- Almacenamiento y custodia de la prensa en papel.
- Almacenamiento y custodia de la prensa digitalizada.
- Costes de preservación digital.

Se excluyen de este análisis los costes asociados a la etapa de consulta de los fondos, pues su cálculo no presente interés en relación con el objeto de este estudio. Tampoco es analizable el gasto relacionado con la difusión, puesto que no existe publicación en red, ni abierta ni intranet corporativa. Sí deben destacarse, en cambio, los costes relacionados con la preservación digital, puesto que el almacenamiento pasivo no es una estrategia válida en la conservación de los recursos digitales y el mero cálculo de los costes de almacenamiento no permite establecer una visión completa de las cargas económicas que supone preservar los ficheros electrónicos.

#### *3.2.4.1. COSTE DE ADQUISICIÓN DE LOS DIARIOS*

Cada biblioteca realiza la compra diaria de la prensa que ha seleccionado para su colección. Normalmente, se realiza mediante suscripciones o mediante acuerdo con un distribuidor de prensa que sirve diariamente los periódicos. En la medida en que a las bibliotecas interesa hacerse con colecciones hemerográficas completas, se entiende que son costes recurrentes presentes en los presupuestos de todos los años y con continuidad en el tiempo mientras la cabecera es publicada. También ha de tenerse en cuenta que habitualmente las bibliotecas adquieren varios títulos para sus colecciones, multiplicándose así el coste por tantos diarios como se adquieran para la colección.

#### *3.2.4.2. COSTE DE CONVERSIÓN DE LOS DIARIOS A FORMATO DIGITAL*

Las bibliotecas contratan anualmente la realización de la conversión a formato digital de los diarios. Esto es, la digitalización se externaliza o subcontrata. El modo en que el proceso está articulado hace que haya costes que recaigan sobre la empresa contratada y costes que recaigan sobre cada una de las bibliotecas contratantes.

Se desconoce al detalle el proceso técnico exacto que los subcontratistas desarrollan para la digitalización de los periódicos (véase apartado 3.2.1), pero en general se puede establecer que sobre la empresa contratada recaen los siguientes costes:

- Adquisición de los diarios.
- Digitalización (guillotinado, alimentación de escáner, operador de escáner, correcciones, generación de ficheros).
- Grabación de los CDs de chequeo, uno por cabecera.
- Envío de los CDs de chequeo a las bibliotecas encargadas de verificar cada cabecera.
- Corrección de errores (nuevos escaneados, identificación de ficheros...)
- Grabación CDs definitivos con las copias derivadas para cada una de las bibliotecas.
- Envío CDs a cada una de las bibliotecas.
- Grabación de discos duros externos con la versión maestra.
- Envío de los discos duros a la Fundación Sancho el Sabio.

A pesar del desconocimiento pormenorizado de las tareas asociadas al, y que por tanto no se pueda establecer una tipología de operaciones generadoras de costes, sí cabe cuantificarse el



coste económico global, dado que la empresa recibe un pago que se computará como coste total de esa parte del proceso. En la Tabla 8 se presentan datos económicos que cuantifican estos costes.

En 1998 se estipula un precio de referencia por cabecera, el cual permite hacerse una idea del coste que había supuesto hasta ese momento la generación de copias de conservación. La cantidad establecida se cobrará a partir de ese momento a los nuevos miembros del Consorcio que deseen hacerse con copia de aquellas cabeceras reproducidas con anterioridad a su ingreso. El precio “por cada periódico y año” se valora en “la cantidad de 82.458 ptas. más IVA”, según tasación realizada por la Biblioteca Central del Gobierno Vasco cuantificando “el gasto efectuado por cada institución hasta 1996 y 1997, incluida microfilmación” (acta 22/10/1998).

Tabla 8.- Costes económicos

	Empresa	Coste**	IVA	Btc as.	€/btca**	Coste***	Coste por 10 Btcas
1992* <sup>8</sup>	CTD					3.605,54	36.055,40
1993*	CTD					4.984,67	49.846,70
1994*	CTD					4.359,31	43.593,10
1995*	CTD					4.327,28	43.272,80
1996	Archivos y Bibliotecas					3.471,35	34.713,50
1997	Archivos y Bibliotecas					3.186,03	31.860,30
1998	IECISA					6.160,54	61.605,40
1999	IECISA					6.160,54	61.605,40
2000	IECISA					6.160,54	61.605,40
2001	IECISA					6.160,54	61.605,40
2002	IECISA/Tramic					5.602,13	56.021,30
2003	Tramic					5.016,60	50.166,00
2004	Scanbit				2.620,80	5.016,60	50.166,00
2005	Gudat				2.589,23	4.579,21	45.792,10
2006	Gudat				2.685,03	2.685,21	26.852,10
2007	Gudat					5.953,56	59.535,60
2008	Gudat	67.870,60	Sí	10	6.787,06	6.787,08	67.870,80
2009	Gudat	78.838	Sí	10	7.883,80	7.871,76	78.717,60
2010	Gudat	79.396,20	Sí	10	7.939,62	7.962,24	79.622,40
2011	Gudat	2010+ipc				8.226,96	82.269,60
2012	Scanbit	52.700	Sí	10	5.270	5.792,68	57.926,80
2013	Scanbit	60.062	Sí	10	6.006,20	6.006,20	60.062,00
Egin						4.873,25	48.732,50
Refresco						569,13	7.232,60
Hemeroteca +						24.070,12	240.701,20
<b>Coste total</b>						<b>146.508,80 €</b>	<b>1.497.432 €</b>

En el acta 26/11/2003 se presentan unos cuadros de cálculo de precios de referencia (se supone que es lo que ha costado digitalizar cada cabecera a cada institución, para “vender” al mismo precio a otras instituciones).

Por otra parte, y como ya se ha mencionado “en 1999 se hizo una digitalización retrospectiva del diario Egin (1977-1991) (...) tuvo un coste de 6.296,63 euros”.

Existe otro coste de la conversión de los diarios a formato digital que recae sobre las bibliotecas, tiene que ver con la comprobación de los CDs de chequeo.

<sup>8</sup> \* 1992-1995 son microfichas. \*\* Coste presupuestado. \*\*\* Coste real soportado por una de las bibliotecas, por la compra de diez cabeceras.

- Costes de personal/hora en comprobación.

Las bibliotecas verifican que todos y cada uno de los periódicos estén plenamente digitalizados, sin errores de identificación, ni de paginación, sin hojas faltantes y sin errores graves de imágenes defectuosas. El proceso es manual, comprobando visualmente cada diario en papel contra la digitalización del mismo en pantalla. Hay aquí un coste de personal para cada una de las bibliotecas integrantes del Consorcio, cuantificable en horas de trabajo empleadas en comprobar todos los meses la cabecera asignada.

#### *3.2.4.3. COSTES DE ALMACENAMIENTO Y CUSTODIA DE LA PRENSA EN PAPEL*

El almacenamiento de grandes volúmenes de papel comporta el consumo de muchos metros lineales de estanterías. Esta ocupación de estanterías comporta lógicamente el llenado de los depósitos de las instalaciones, variable cuantificable en metros cúbicos.

Como se sabe, una de las finalidades principales del proyecto llevado adelante por el Consorcio ha consistido en minimizar el consumo de espacio en los depósitos de las bibliotecas a través de la generación de copias en formatos que consuman menos espacio para su almacenamiento. Sin embargo, es precisamente la conservación de las colecciones en papel la generadora del derecho a reproducción. Esto es, para las bibliotecas la posesión de los diarios en papel parece ser la clave que garantiza el origen de esos derechos (recogidos en la ley como límites a su favor), y existen notables dudas sobre la legalidad de la posesión de los diarios digitalizados careciendo de las copias en papel en caso de haber sido eliminadas tras su digitalización. Es por ello que en algunos casos se ha optado por la conservación cooperativa como fórmula garantista de derechos y que permite ahorrar espacio. El mantenimiento de colecciones en papel, aunque sea de modo cooperativo, comporta lógicamente un gran consumo de espacio.

En la medida en que se trata de colecciones vivas, puesto que los diarios siguen en curso de publicación y las bibliotecas continúan adquiriéndolos, el incremento del consumo de espacio va en aumento. Este incremento del consumo de espacio difícilmente se puede calcular conforme a un valor constante dado que la variación en el número de páginas y en el tamaño con que se están publicando los diarios es cambiante en los últimos años, debido a la situación de crisis de los medios impresos que anteriormente se ha comentado.

En cualquier caso, el coste del espacio ya ocupado en los depósitos de cualquiera de las bibliotecas es cuantificable atendiendo a la variable de los metros cúbicos ocupados en relación con el valor de amortización del edificio (3% del valor contable en la actual legislación fiscal de Álava), más los costes de mantenimiento de los depósitos.

#### *3.2.4.4. COSTES DE ALMACENAMIENTO Y CUSTODIA DE LA PRENSA DIGITAL*

El consumo de espacio en los depósitos para la custodia de la prensa digitalizada deriva de la ocupación del mismo por las unidades de almacenamiento informáticas (CD-ROM y discos duros externos, en este caso), y conlleva unos costes en el mismo sentido que el planteado en el apartado anterior.

Aunque el consumo de espacio es bastante menor en comparación con el almacenamiento del papel, lo cierto es que la ubicación de los soportes alcanza ya los 27,5 m lineales en las

instalaciones de la Fundación Sancho el Sabio (para las dos copias máster y contabilizados únicamente los discos duros externos hasta 2011). Hay ocupadas 21 baldas de 130 cm de largo y 40 cm de ancho. Cada balda, por tanto, ocupa 0,52 m<sup>2</sup>.

En metros cuadrados la ocupación es de una superficie de 0,52 m<sup>2</sup> X 21 baldas = 10,92 m<sup>2</sup>.

Tomando una altura estándar de dos metros y medio para los depósitos, el volumen ocupado ronda los 10,92 X 2,50 = 27,3 m<sup>3</sup>.



*Figura 9.- Discos duros externos almacenados en los depósitos controlados de la Fundación Sancho el Sabio.*

La ocupación de metros lineales aumenta anualmente en la medida en que va creciendo la colección de diarios digitalizados, puesto que se trata de una colección viva y el programa de digitalización sigue su curso. El número de discos en las dependencias de la Fundación Sancho el Sabio alcanza los 160 discos duros y los 3.712 CD-ROM (véase apartado 0).

Los costes de almacenamiento de las unidades informáticas en depósitos especiales son mayores a los costes de almacenamiento de papel en un depósito normal, debido a que la guarda de los soportes electrónicos se ha de realizar en condiciones de temperatura y humedad bajo controles más estrictos y dentro de unos valores sólo alcanzables mediante

refrigeración y deshumidificación por máquina. Estos sistemas de refrigeración y deshumidificación consumen energía y están necesitados de continuos controles humanos para mantener estables, y dentro de unos rangos aceptables, los valores idóneos para la conservación preventiva de las unidades informáticas. Por ello, el coste es mayor que el coste de consumo de espacio de almacenamiento en un depósito normal.

Finalmente y en cualquier caso, este coste, al igual que en el caso anterior, es cuantificable atendiendo a la variable de los metros cúbicos ocupados en relación con el valor de amortización del edificio (3% del valor contable), más los costes de mantenimiento de los depósitos (estos costes son mayores en almacenes con control medioambiental estricto).

Hay que destacar que no todas las bibliotecas poseen todas las colecciones digitales derivadas completas (y por tanto tampoco son propietarias de todas las colecciones de ficheros maestros), puesto que a partir de cierta etapa del proyecto se dio la oportunidad a las bibliotecas de elegir qué cabeceras eran de su interés. Quienes se interesan por un título concreto cooperan en la asunción de los gastos de digitalización de esa cabecera, reciben solamente las copias digitales derivadas de tal título y son propietarias de la copia maestra digital de ese diario.

#### *3.2.4.5. COSTES DE PRESERVACIÓN DIGITAL*

Hasta la fecha estos costes, más allá de los descritos en el apartado 3.2.4.4 y que tiene que ver con el control propio de un programa de conservación preventiva, se han limitado a la generación de una copia de refresco de los soportes realizada entre octubre de 2006 y enero de 2007.

El refresco realizado por Scanbit en 2006, consistente en paso a CD-ROM, costó 7.232,80 € (6.235,20 € + IVA). Cada una de las trece bibliotecas integrantes del consorcio en ese momento debió hacer frente a facturas por valor de 556,40 € IVA incluido. Durante la realización del trabajo se descubre que existe otros 145 CD-ROM de “errores” de El Corte Inglés que faltaban por presupuestar, así que finalmente el coste se eleva.

#### *3.2.5. ALMACENAMIENTO Y CUSTODIA DE LA COLECCIÓN*

La versión maestra de los ficheros de imagen se conserva actualmente en las dependencias de la Fundación Sancho el Sabio en Vitoria-Gasteiz. Desde el comienzo del proyecto ha estado localizada en las dependencias de la misma institución, primero en el Palacio Zulueta en el centro de Vitoria-Gasteiz, a partir de 2009 en la sede actual de la Fundación en el convento de las madres Carmelitas de Betoño. Se trata de un conjunto de edificios rehabilitados para uso de equipamientos culturales por el equipo de arquitectos de Roberto Ercilla y Miguel Ángel Campo, situado en Portal de Betoño nº 23. La Fundación se ubica en la construcción que albergaba el antiguo cementerio del convento, en unas instalaciones acondicionadas para uso bibliotecario.

La biblioteca de la FSS se dedicada a los estudios vascos y alberga un notable fondo patrimonial, que incluye tanto incunables como negativos fotográficos, folletos, carteles, libros, revistas, periódicos, fanzines, películas de cine y video..., en diversidad de soportes y formatos. Por lo cual, tanto los depósitos de la antigua sede como los nuevos están diseñados

y acondicionados para garantizar la preservación de los diversos materiales conforme a los requerimientos específicos de cada uno de ellos. Cuenta con depósitos diferenciados y acorde a distintas necesidades de conservación documental.

### 3.2.5.1. MEDIOS DE ALMACENAMIENTO

La colección maestra de imágenes ráster se encuentra grabada, según la época, en CD-ROM o en disco duro externo. La primera época, que abarca de 1996 a 2004 se encuentra soportada en CD-ROM, y la segunda época, de 2005 a 2011, lo está en HD. No existe ningún otro medio de almacenamiento, aunque en las actas y en varios presupuestos se menciona en diversas ocasiones el uso de DVD no existe información grabada en este medio. El consorcio posee además a prensa microfilmada entre 1992 y 1995.

Los CD-ROM que comprenden la información de los años 1996 a 2002 son CD-ROM de color dorado (véase apartado 4.1.1 para detalles sobre tipología de medios ópticos de almacenamiento). Fueron realizados por le empresa Scanbit entre finales de 2006 y principios de 2007 en virtud de un contrato para el refresco de los soportes. Los CD-ROM de 2002 son de color azul y los de 2003 son de color verde. Ambos fueron realizados por la empresa Tramic. Los de 2004 fueron realizados por Scanbit, en color azul.

Los discos duros utilizados para el almacenamiento de la información, entregados por los distintos proveedores entre los años 2005 a 2013, presentan gran diversidad. Se pueden contabilizar hasta 26 modelos diferentes de discos de diez casas comerciales distintas. Todos son accesibles mediante puerto USB 2.0, dos de ellos tienen conexión USB 3.0 (compatible con US 2.0, posibilita un acceso a los datos diez veces más rápido). Pueden observarse los detalles de marcas y modelos comerciales en la Tabla 9.

Tabla 9.- Marcas, modelos y características de los discos duros externos utilizados

MARCA	MODELO	CAPACIDAD	ACCESO
Conceptronic	Conceptronic 3'5" Hardisk Box Grab'n'Go	1 TB	USB 2.0
Eminent	Eminent 3.5" HDD Enclosure SATA+eSATA	1 TB	USB 2.0
Hitachi	Hitachi Simple Drive	1 TB	USB 2.0
	Hitachi Touro Desk	1 TB	USB 2.0
	Hitachi XL1000	1 TB	USB 2.0
iOmega	iOmega Prestige Desktop Hard Drive	1 TB	USB 2.0
	iOmega Prestige Desktop Hard Drive	500 GB	USB 2.0
	iOmega Select Portable Hard Drive	1 TB	USB 2.0
Maxtor	Maxtor Basics Personal Storage 3200	320 GB	USB 2.0
	Maxtor One Touch II External Hard Drive	300 GB	USB 2.0
	Maxtor One Touch III External Hard Drive	500 GB	USB 2.0
Philips	Philips External Har Disk	500 GB	USB 2.0

	Philips External Har Disk Dual Interface	500 GB	USB 2.0
Seagate	Seagate FreeAgent Desk External Drive	500 GB	USB 2.0
	Seagate Pushbutton Backup External Hard Drive	500 GB	USB 2.0
Toshiba	Toshiba External USB Hard Drive	500 GB	USB 2.0
	Toshiba STOR.E ALU 2	1 TB	USB 2.0
	Toshiba STOR.E ALU 2	2 TB	USB 2.0
	Toshiba STOR.E ALU 2S	1 TB	USB 3.0
Verbatim	Verbatim Desktop Hard Drive	1 TB	USB 2.0
Western Digital	My Passport	2 TB	USB 3.0
	WD Elements External Hard Drive	2 TB	USB 2.0
	WD Elements External Hard Drive	1 TB	USB 2.0
	WD My Book External Hard Drive Essential Edition	500 GB	USB 2.0
	WD My Book External Hard Drive Essential Edition	750 GB	USB 2.0
	WD My Book External Hard Drive Home Edition	500 GB	USB 2.0

### 3.2.5.2. DEPÓSITO

Las instalaciones de la Fundación Sancho el Sabio en que se encuentran los ficheros maestros cuentan con varios sistemas de seguridad y protección. El interior del edificio está equipado con cámaras de seguridad. Cuando se cierra al público se activa un sistema de alarma, conectado a una centralita, que incluye protección perimetral con infrarrojos y video-vigilancia para evitar vandalismo e intrusiones.

En relación con la protección frente a desastres, existe un sistema anti-inundaciones. Las principales dependencias del edificio, entre ellas los depósitos, se encuentran por debajo del nivel de calle, esto es, por debajo de la cota 0 del terreno. El edificio está construido en zona con riesgo de inundaciones, donde el nivel freático varía de media desde -1 m en invierno a -3 m en verano. El edificio, por tanto, fue equipado con un sistema para evitar inundaciones, consistente en dos cubos que lo protegen. El primero hace estanco el edificio y el segundo es un muro que rodea el perímetro del anterior. El agua que desde la capa freática va entrando en este hueco conformado por ambos muros estancos, es retirada con bombas eléctricas a dos pozos alejados del edificio.

En caso de fallo del sistema eléctrico, por interrupción del suministro, tres personas reciben notificación mediante SMS (*Short Message Service*, servicio que permite el envío de mensajes cortos entre teléfonos móviles), y se activa un generador de gasóleo con autonomía de 24 horas que permite mantener activas tanto las bombas del sistema anti-inundación como el sistema informático, en modo de UPS.

En relación con los depósitos de documentación estos cuentan con control de accesos con llave solo para personal autorizado. También están sometidos a controles anti-plagas, gestionados por una empresa subcontratista que periódicamente pasa informes del estado del

edificio y de los depósitos. Hay control de roedores, insectos y microorganismos con trampas de hormonas.

Existen también medios activos en la protección contra incendios. Hay sistema de detección de incendio, alerta y señalización y de extinción con agentes extintores conducidos por tuberías que rocían con gas HFC-23 (norma UNE 23573) o con agua pulverizada, dependiendo del tipo de depósito en relación con el tipo de materiales documentales albergados en el mismo.

El depósito en que se encuentran los CD-ROM y los discos duros cuenta con condiciones de temperatura y humedad estables y adecuadas a los soportes de almacenamiento óptico y magnético. La temperatura está establecida en unos valores constantes de entre 12°C y 13°C, y la humedad relativa es constante entre 30% y 40% de humedad relativa (HR).

Tanto los discos duros como los CD-ROM cuentan con protección física. Están debidamente protegidos guardados en sus fundas y en cajas. Los CD-ROM están en su correspondiente caja de metacrilato y organizados por título y año en cajas de preservación en armarios compactos. Los discos duros por su parte están en las cajas en las cuales se comercializan, normalmente compuestas por carcasas de protección de plástico rígido envueltas por cajas de cartón. Están ordenados por meses y año, dado que en un único disco duro mensual están almacenadas las imágenes de todos los diarios escaneados en ese periodo.

#### 3.2.5.3. BACKUP

La colección completa de discos duros tiene una segunda copia de respaldo. Sin embargo, no se trata de un *backup* al uso, realizado en un momento dado a partir de la copia original para contar con un duplicado de los datos, sino que se generó en el mismo momento que la copia principal a petición del Consorcio y con la idea de contar con un *backup* desde el origen, sin tener que realizarlo posteriormente. Los contratos obligan a los proveedores desde el año 2005, a entregar al CDPV dos copias con la versión maestra de los ficheros, y una copia con las versiones derivadas a cada una de las bibliotecas integrantes del Consorcio. La segunda copia máster es, por tanto, una réplica que las empresas que han realizado las labores de escaneo en las distintas etapas del proyecto han entregado junto con la primera, como producto final de su trabajo. Son dos copias idénticas en cuanto a datos, soporte de almacenamiento y fecha de creación.

No se puede saber cuál de las dos copias es la copia principal y cuál la secundaria. No existe ninguna anotación en las cajas o fundas que las identifique en este sentido. Están almacenadas en la misma ubicación y colocadas correlativamente en las estanterías, lo cual hace indistinguible una copia de la otra.

De la colección de másteres en CD-ROM existe una única copia.

#### 3.2.5.4. REFRESCO

Hasta la fecha, ha habido una única actuación en este sentido, consistente en la generación de una copia de refresco de los soportes realizada entre finales de 2006 y principio de 2007, tal y como se ha descrito en el apartado 3.2.4.5.

En acta de 30/09/2003 se menciona por primera vez la necesidad de realizar un “refresco” de las digitalizaciones realizadas por IECISA entre 1996 y 2002. Poco después, en acta de 05/11/2003, se detalla que se han recibido cinco ofertas de otras tantas empresas para realizar “el traspaso de la información contenida en soporte CD-ROM a soporte DVD”. E inmediatamente después, acta de 26/11/2003, se adjudica tal volcado a la empresa Gudat.

Sin embargo, finalmente no se lleva a cabo, y el asunto del refresco aparece recurrentemente en posteriores actas hasta adjudicar el trabajo, nuevamente y de forma ya definitiva, en 2006.

Este volcado de los TIFF de IECISA, se adjudicará a Scanbit por 7.232,6 (IVA incluido) (acta de 18/10/2006). Aparecen cuestiones técnicas de periodos volcados y nomenclatura de los CD-ROM en acta de 13/12/2006. Tal refresco consistió en la generación de CD-ROM, no DVD como en un principio se mencionaba (acta 01/2007 de 31/01/2007).

Parece ser que lo que en realidad ocurre es que en las entregas iniciales de la digitalizaciones de IECISA no se había generado “una copia de titularidad de todo el Consorcio”, o sea que no existía copia máster y que todas las copias entregadas a los miembros eran idénticas, no se disponía de las copias derivadas más una copia máster (acta de 14/01/2004); “(...) no existe copia TIFF de seguridad de los fondos digitalizados por El Corte Inglés” (acta de 10/05/2005). Además, existía cierta preocupación por la dependencia del programa de software propietario Archidoc para poder visualizar las imágenes iguales (acta de 14/01/2004).

Por tanto, lo que a lo largo de las actas del consorcio durante varios años aparece mencionado como un “refresco” se trata en realidad de una cosa distinta. Para empezar, se ha de tener en cuenta que las digitalizaciones de ese período incluyen únicamente la entrega de una copia en TIFF, no entregándose imágenes en ningún otro formato. Se desconoce si se trata de la imagen TIFF original de alta calidad en captura, por tanto copia maestra, o si se trata de imágenes TIFF de menor calidad generadas a partir de las imágenes de captura y por tanto derivadas. Es de suponer, que se trata de la primera y, en cualquier caso, como el Consorcio no dispone de copia de mayor calidad ha de ser considerada como copia maestra.

Afortunadamente, las entregas de IECISA fueron realizadas en CD-ROM de buena calidad y se puede considerar que, para las digitalizaciones de ese período, se dispone de tantas copias maestras como miembros había en el consorcio, más la nueva copia generada por Scanbit que es de peor calidad (véanse apartados 3.2.1.1y 3.3).

No se ha realizado ninguna otra actuación de refresco o actualización de soportes durante toda la vida del Consorcio de Digitalización de Prensa Vasca.

### **3.3. VALORACIÓN TÉCNICA DEL FONDO DIGITAL RESULTANTE**

Durante el estudio y análisis del fondo de imágenes ráster del CDPV se han detectado varios problemas dignos de mención, algunos de mayor importancia que otros. En ningún caso se pretende hacer una valoración pormenorizada del estado de la colección, sin embargo se aprovecha para describir ciertos errores encontrados y algunas cuestiones de calidad llamativas. Algunos de los problemas tienen fácil solución, otros tienen solución con coste económico asociado y por último, también existen problemas que no tienen solución.



Los principales problemas encontrados son: i) ficheros con baja resolución de cara a posible OCR; ii) ficheros sin copia derivada o de explotación; iii) inexistencia de archivos RAW; iv) inexistencia de información sobre las cámaras de captura de imágenes; v) inexistencia de gestión del color; vi) ficheros maestros en formato JPEG; vii) inexistencia de descripción con esquemas de metadatos y escasa información de metadatos EXIF; viii) baja calidad de imágenes digitales; ix) inexistencia de marca de agua; x) dos copias idénticas albergadas en la misma ubicación; o xi) ficheros de imagen sin copia de respaldo.

La Tabla 10 sistematiza y categoriza estos problemas detectados, y que se describen a continuación.

- i. El primer condicionante se deriva del hecho de que las imágenes de los primeros quince años, entre 1996 y 2010, presentan una resolución de captura de 200 ppp. Tales parámetros de digitalización son fruto, en los primeros años, del estado y desarrollo de la tecnología y en años siguientes, del interés por no encarecer los costes. Esta resolución de captura puede considerarse baja de cara al reconocimiento óptico de caracteres (OCR) que habría que aplicar al fondo si se pensase en una mejora de la recuperación de la información (RI) del mismo en un contexto de repositorio digital.

El alojamiento en repositorios digitales va a facilitar el acceso de los materiales, pero para ello es imprescindible que se puedan recuperar adecuadamente. En relación con la recuperabilidad de los recursos digitales en los repositorios hay que prestar atención a dos variables: los metadatos y el OCR. El caso del OCR en RI en prensa daría es crítico, puesto que la recuperación a través de los campos típicos en la descripción catalográfica tradicional (título, fechas, etc.) carecen de interés para recuperar contenidos en una colección amplia de diarios, pasando a primer plano las búsquedas a texto completo.

Abadal y Guallar señalan que el reconocimiento óptico de caracteres en el caso de los diarios presenta una serie de dificultades añadidas que no se dan en este proceso técnico para otro tipo de documentos. Entre los problemas más frecuentes mencionan “la existencia de originales deteriorados, el uso de fuentes tipográficas poco corrientes, la utilización de tipos un poco sucios, de letras superpuestas o muy juntas, columnas muy fragmentadas, etcétera” (2010, p. 139). Se ha de señalar además, que el OCR funciona mejor sobre ficheros ráster con buen nivel de contraste, por la cual las digitalizaciones en blanco y negro (profundidad de color bitonal) son las mejores para realizarlo, pero la resolución de captura, y aquí está el problema con que puede encontrarse el CDPV, debe ser siempre superior a 200 ppp (Robledano, 2012).

- ii. Otro inconveniente detectado en la colección del CDPV tiene que ver con que no se cuenta con ficheros derivados o versiones de explotación de los escaneos realizados por IECISA entre los años 1996 y 2002, tal y como ha quedado expuesto en el apartado 3.2.5.4. Toda la colección está en formato TIFF. Es un problema menor, dado que como se ha expuesto se puede considerar que se cuenta con varias copias maestras y la baja calidad del refresco realizado entre finales de 2006 y principios de 2007, invita a considerar la copia generada en tal refresco como copia de explotación. No conlleva un coste asociado. Se podría generar PDF con coste asociado

Tabla 10.- Categorización de los problemas de la colección digital del Consorcio de Digitalización de Prensa Vasca

Problema	Categoría	Solución	Coste
Ficheros con baja resolución	Grave	Re-escaneado	Muy Alto
Ficheros sin copia derivada	Leve	Generación	Medio
Inexistencia de archivos RAW	Grave	Próximas digitalizaciones	Bajo
Inexistencia de información de cámara	Grave	Próximas digitalizaciones	Muy Bajo
Inexistencia de gestión del color	Grave	Próximas digitalizaciones	Bajo
Ficheros maestros en formato JPEG	Muy Grave	Migración	Medio
Inexistencia de descripción con metadatos	Grave	Metadateado	Medio/Alto
Escasa información de metadatos EXIF	Muy Grave	Próximas digitalizaciones	Muy Bajo
Baja calidad de imágenes digitales	Muy Grave	Re-escaneado	Muy Alto
Inexistencia de marca de agua	Leve	Generación	Medio/Alto
Dos copias en la misma ubicación	Grave	Traslado	Muy Bajo
Ficheros máster sin <i>backup</i>	Grave	Generación	Medio/Alto

- iii. Por otra parte, no se existen en ningún caso ficheros en formato RAW, o de imagen bruta de captura de cámara. Tenerlos puede presentar la ventaja de poder mejorar la calidad del fondo con el paso de los años en función de la evolución de la tecnología, mediante la aplicación de nuevos procesos técnicos de generación de derivados y másteres, y mediante el procesado de la información de metadatos, conforme al estado del arte en tecnología de imagen digital de cada momento. Es un problema que no tiene solución, pero que podría y debería abordarse de cara a las nuevas campañas anuales de digitalización.

Dependiendo del equipo con que el proveedor esté generando las digitalizaciones, podrá ser capaz o no de entregar ficheros RAW. Hacerlo podría encarecer algo los costes, dado que la manipulación y control de mayor número de ficheros requiere personal y se duplicaría el tamaño del producto final entregable necesitando aumentar el número de discos. Finalmente, el consiguiente aumento del número de ficheros a custodiar por el CDPV y el aumento del tamaño de la colección lógicamente supone un aumento del coste en la preservación digital a largo plazo.

- iv. Tampoco se dispone de información de los escáneres o de las cámaras utilizadas por ninguna de las empresas a lo largo de estos años pasados. Se carece de datos sobre modelos y características; tan solo en alguno de los presupuestos aparecen

relacionadas las máquinas con que cuentan las compañías que se presentan a los concursos, sin conocerse si finalmente tales máquinas u otras fueron utilizadas.

- v. No se ha realizado gestión del color durante el proceso de digitalización. Si bien los diarios editados en los primeros años de actuación del Consorcio salían casi en su totalidad en blanco y negro, poco a poco se fue introduciendo el color, principalmente en las portadas y en las ediciones de fin de semana. Desde hace unos pocos años a esta parte, el color está presente diariamente en casi todas las páginas de los rotativos. En acta de 01/03/2001 se resalta la ausencia o “pérdida de color en los titulares. Es un problema técnico al digitalizarse sólo en blanco y negro, y que no tiene solución”, parece ser que la tecnología no estaba al alcance del Consorcio.

Justamente diez años después, en 2011, el CDPV comenzará con las digitalizaciones en color. Sin embargo, por lo que se ha podido comprobar, no se ha realizado en ningún caso gestión integral del color.

Los ficheros de imagen ráster del CDPV no contienen en ningún caso información en sus metadatos sobre el espacio de color utilizado. Bien es cierto que tratándose de fondos documentales donde la apariencia del color no presenta importancia alta, tal vez podría considerarse que el uso de del espacio de color virtual sRGB podría ser suficiente. Tal espacio de color ha sido utilizado únicamente con los ficheros derivados en formato PDF a partir de la inclusión del color en el año 2011.

Las imágenes TIFF no han sido escaneadas junto a cartas de perfilación (cartas de color estandarizadas, como Colorcheckr o IT8). Los proveedores no han entregado ficheros de perfil de color ICC. No se ha aplicado ningún software de Sistema de Gestión de Color (CMS). Se puede afirmar por tanto, que no ha habido gestión de color.

En este sentido, es clave considerar si la digitalización que se viene realizando está hecha desde una perspectiva reprográfica donde se prima el mero valor informativo de los diarios, o bien desde una perspectiva patrimonial donde se quiere mantener también el aspecto y valor icónico (color incluido) de los diarios a modo facsimilar.

- vi. En otro orden de cosas, existe un problema mayor relacionado con el hecho de que hay una serie de imágenes ráster de las cuales el Consorcio sólo dispone de ficheros maestros en formato JPEG.

Las imágenes en este formato presentan un problema de conservación relacionado con su degradación paulatina, que ha sido documentado y expuesto ampliamente, y que está relacionado con compresión con pérdida del formato. Cada vez que un fichero con formato JPEG se abre en un editor gráfico y se vuelve a guardar (no se copia, sino que se abre y se guarda), la calidad de la imagen se degrada (Carnicer Arribas et al., 2011; García Martínez, 2012). Se han realizado diversas pruebas que establecen que abriendo un fichero entre 600 y 1000 veces la degradación es casi absoluta. El problema deriva de la fuerte compresión que este formato aplica, aunque tiene varios modos o ratios de compresión que producen mayor o menor pérdida.

Cada vez que se abre y se guarda un fichero éste vuelve a comprimirse, de modo que al tratarse de compresión con pérdida se van sumando pérdidas de información, que se

acumulan hasta degradar fuertemente la imagen original y perder gran parte de la información de captura.

Existen en el mercado visores de imágenes, de uso muy común y popular, en los cuales es difícil saber si lo que se está haciendo al “abrir” una imagen es visualizarla o editarla. Tal problema no se da con las aplicaciones profesionales, como GIMP o Photoshop. Se requiere por tanto especial cuidado en la manipulación de imágenes con software no profesional.

Es el caso del diario *Egin*, chequeado con cierto detalle aunque no de modo sistemático. Se trata de un muestreo realizado en el disco duro externo elaborado en 2013, verificando el primer número mensual de cada año; por tanto, los datos no son totalmente fiables dado que no se trata de un muestreo realizado sobre una porción significativa del conjunto. No es muestreo sistemático ni muestreo aleatorio, sino tan solo muestreo de ejemplar o de testigo.

Se han detectado una serie de imágenes que están únicamente en formato JPEG, sobre el contenido completo del periódico entre el 29 de septiembre de 1977 y el 15 de julio de 1998:

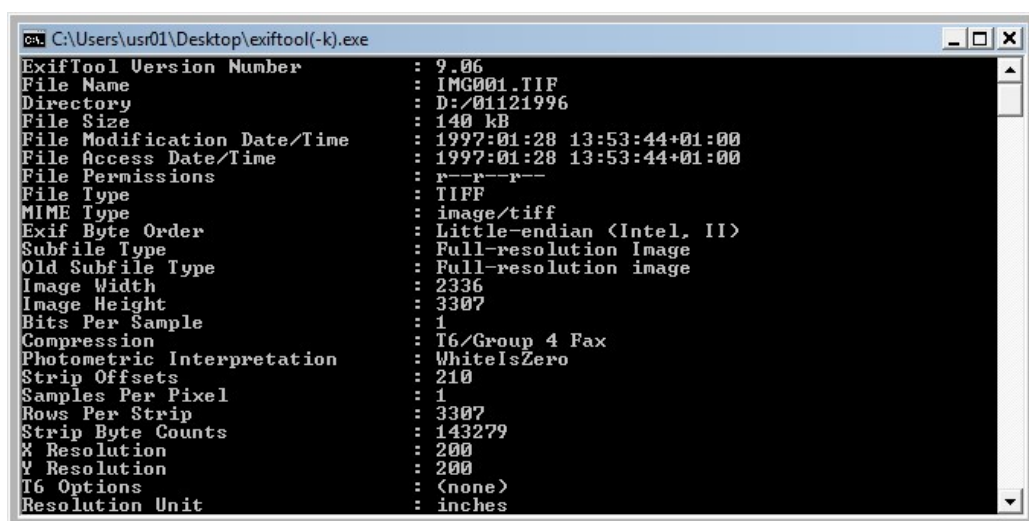
- ~ 1983, año completo en formato JPEG.
- ~ 2 de enero de 1991, en formato JPEG.
- ~ 1992, año completo en formato JPEG .
- ~ 1994, días sueltos en enero y todo de mayo a diciembre, en formato JPEG.
- ~ 1995, año completo en formato JPEG .

- vii. En relación con los metadatos se puede afirmar que en ningún caso existen ficheros independientes, con metainformación sobre las imágenes, almacenados junto a los ficheros de imagen ráster. La escasa información de metadatos de que se dispone es siempre información incrustada en los propios ficheros de imagen ráster, en formato EXIF.

Los visores de imágenes han permitido verificar la existencia, cantidad y calidad de datos técnicos sobre la imagen. Por ejemplo, el visor Galería de Fotografía de Windows Vista ofrece información muy completa y en una presentación muy cómoda para el usuario. Aporta datos de descripción (como título, asunto, clasificación, etiquetas y comentarios), datos de origen (como autor, fecha de captura, nombre del programa, fecha de adquisición y copyright), datos de imagen (como id de imagen, dimensiones, ancho, alto, resolución horizontal, resolución vertical, profundidad en bits, compresión, unidad de resolución, representación del color, bits comprimidos/píxel) datos de cámara (como fabricante, modelo, punto F, tiempo de exposición, velocidad ISO, compensación de exposición, distancia focal, apertura máxima, modo de medición, distancia al objeto, modo de flash, intensidad de flash y longitud focal de 35 mm), datos de fotografía avanzada (como creador de objetivo, modelo de objetivo, creador de flash, modelo de flash, número de serie de la cámara, contraste, brillo, fuente de luz, programación de exposición, saturación, nitidez, balance de blanco, interpretación fotométrica, zoom digital y versión EXIF) y datos de archivo (como nombre, tipo de elemento, ruta de acceso, fecha de creación, fecha de modificación, tamaño, atributos, dispositivo sin conexión, estado sin conexión, compartido con, propietario y equipo).

Sin embargo, se ha utilizado principalmente la herramienta ExifTool (versiones 9.06 y 9.30).

Existe muy escasa información integrada en los ficheros de imagen ráster, posiblemente fruto de una inadecuada parametrización de la máquina con la que se realizan las tomas, o bien fruto del procesamiento posterior de las imágenes en el cual se ha descuidado conservar la posible información contenida en la toma original. En la Figura 10 se puede observar la información de metadatos EXIF de que se dispone en el 100% de los casos. Compárese con las extensas posibilidades de descripción que este formato de descripción de metadatos presenta, en el Anexo III.



```
C:\Users\usr01\Desktop\exiftool(-k).exe
ExifTool Version Number      : 9.06
File Name                    : IMG001.TIF
Directory                   : D:/01121996
File Size                    : 140 kB
File Modification Date/Time  : 1997:01:28 13:53:44+01:00
File Access Date/Time       : 1997:01:28 13:53:44+01:00
File Permissions             : r--r--r--
File Type                    : TIFF
MIME Type                   : image/tiff
Exif Byte Order              : Little-endian (Intel, II)
Subfile Type                 : Full-resolution Image
Old Subfile Type             : Full-resolution image
Image Width                  : 2336
Image Height                 : 3307
Bits Per Sample              : 1
Compression                  : T6/Group 4 Fax
Photometric Interpretation   : WhiteIsZero
Strip Offsets                : 210
Samples Per Pixel            : 1
Rows Per Strip               : 3307
Strip Byte Counts            : 143279
X Resolution                  : 200
Y Resolution                  : 200
T6 Options                   : <none>
Resolution Unit              : inches
```

Figura 10.- Metadatos técnicos EXIF incrustados en el fichero de imagen ráster TIFF de la primera página del diario Deia de uno de diciembre de 1996.

Se ha de resaltar por último, que más allá de los de metadatos técnicos tampoco se dispone de documentación soportada en esquemas de metadatos descriptivos ni de preservación, careciéndose por tanto de datos descriptivos, estructurales, administrativos y legales relativos al fondo.

- viii. Otra cuestión a destacar es que se ha detectado que la calidad de las imágenes o de las digitalizaciones no es siempre todo lo buena que sería de esperar. Con una frecuencia bastante alta se han encontrado imágenes borrosas, satinadas en negro, textos e imágenes no alineados o torcidos, textos ilegibles, fotos no visibles y otra serie de errores menores. Apréciase a modo de ejemplo la digitalización de la Figura 11.
- ix. En otro orden de cosas, muchos proyectos han incluido marcas de agua en las imágenes que han digitalizado. Las imágenes del Consorcio de Digitalización de Prensa Vasca carecen por completo de identificación de propiedad, tanto mediante la inserción de descripción por metadatos, como se ha comentado, como mediante el sistema de marca de agua. En términos digitales, una marca de agua, visible o invisible, se crea mediante la inserción de un modelo de bits en una imagen digital para identificar que el propietario de la imagen es conocido. Tal vez fuese conveniente que el CDPV se plantearse la necesidad de incluir marcas de agua a fin de promover la visibilidad de la institución, proteger y concienciar sobre los derechos de propiedad intelectual y favorecer cierto retorno de la inversión.



Figura 11.- Digitalización de baja calidad: página no alineada, manchas negras, textos ilegibles, imagen no visible

- x. También se ha destacar como problema el hecho de que las dos copias de los discos duros, que contienen las imágenes ráster producto de las digitalizaciones realizadas entre los años 2005 y 2012 estén albergadas en la misma ubicación. Es evidente que en caso de desastre o daño accidental podría ser altamente probable el perder toda la colección, por no haber tomado la simple precaución de ubicar las copias en lugares distanciados geográficamente.
- xi. Finalmente, tampoco es despreciable el hecho de que toda la colección albergada en CD-ROM, la correspondiente a las imágenes tomadas entre los años 1996 y 2004, está en una única copia, sin que exista *backup* o copia de respaldo.

### 3.4. ASPECTOS LEGALES

Se han de considerar tres dimensiones, o vértices de un mismo escenario, en relación con las cuestiones legales que afectan al proyecto de digitalización emprendido por el Consorcio de Digitalización de Prensa Vasca: la cuestión ligada con el derecho de las propias instituciones de la memoria a acometer digitalizaciones; los aspectos legales relacionados con la preservación digital de los documentos electrónicos; y en tercer lugar, el asunto del uso y difusión de los recursos digitales bajo el amparo legal y en condiciones garantistas.

Guallar y Abadal identifican dos derechos de explotación de las obras que se ven afectados en los proyectos de digitalización: el de transformación, al pasar el original en papel a un formato digital, y el de comunicación pública, si es que se va a facilitar el acceso a los contenidos a través de Internet (2010). Otros autores como Fernández Molina (2010), a Térmens (2010b) y a Keefer (2007) han reparado en el otro aspecto, el que tiene que ver con el análisis de la problemática relacionada con la preservación digital.

Sobre la primera de las cuestiones hay poco debate, dado que el marco normativo es muy claro y hay unanimidad de opinión entre los especialistas que han abordado este aspecto. La legislación española permite a museos, bibliotecas, fonotecas, filmotecas, hemerotecas o archivos de titularidad pública o integradas en instituciones de carácter cultural o científico la reproducción de los documentos con fines de conservación según el artículo 37.1 de la Ley de Propiedad Intelectual (*Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia*). La posesión de los títulos en las colecciones propias genera los derechos de reproducción con fines de conservación, la posibilidad legal de digitalizar los diarios emana del mero hecho de poseerlos en la colección propia.

Existen, por tanto, límites al derecho de autor establecidos en la LPI a favor de las bibliotecas que permiten a éstas realizar transformaciones con fines de conservación. Sin lugar a dudas, se puede considerar que la digitalización, en cuanto que transformación realizada bajo estas premisas, está siendo acometida en un marco legislativo garantista y acorde a derecho. *Digitization for preservation* es entendido en el ordenamiento jurídico español, y en todos los países del entorno, como uso legítimo o *fair use*.

En relación con la problemática legal relacionada con la preservación digital de los documentos electrónicos, Fernández Molina (2010), Térmens (2010b) y Keefer (2007) plantean que varias de las tareas clave en toda estrategia de preservación digital chocan actualmente con preceptos de la legislación de propiedad intelectual española. Los límites que juegan a favor de las bibliotecas permiten obrar con comodidad en el mundo analógico, pero en el contexto digital no. Las instituciones de la memoria que realicen labores de migración o refresco sobre sus copias digitales de preservación, realizando así una nueva transformación no contemplada originalmente entre los fines con los que el poseedor de los derechos de autor publicó el recurso informativo, está incurriendo en una infracción de la LPI.

Operaciones como el refresco, la migración o la emulación afectan al derecho de reproducción porque implican la realización de copias. La migración, en cuanto que modificación de contenidos afecta al derecho de transformación. La entrega de copias a otras instituciones para realizar custodia cooperativa de las réplicas de respaldo, supone afección del derecho de

distribución. La subcontratación del almacenamiento y custodia digital, mediante la modalidad *data storage as a service* de almacenamiento en la nube o *cloud computing*, también afecta al derecho de distribución a terceras partes. Y el acceso por parte de los usuarios a los documentos en red afecta al derecho de comunicación pública.

Es por ello que FESABID ha propuesto una serie de alegaciones a la legislación para poder realizar desde los archivos y las bibliotecas, entre otras cosas, preservación digital (2013). En este mismo sentido camina la nueva legislación de Depósito Legal, que en estos momentos se encuentra en fase de desarrollo reglamentario, que va suponer un gran avance en el campo de la preservación de los documentos electrónicos alojados en la red Internet, con facilidades para el archivo web y para la retención de publicaciones científicas y revistas electrónicas. Según la jefa de depósito legal de la Biblioteca Nacional de España (BNE), Montse Oliván los editores tendrán la obligación de dar las claves de acceso a redes con documentación patrimonial digital de acceso restringido, como periódicos de pago o libros electrónicos (Criado, 2013). Existe un conflicto entre la explotación comercial, la legislación de propiedad intelectual y el interés público. Términos se pregunta cuándo y cómo se hará el traspaso a la preservación pública (2012), parece ser que la nueva legislación sobre depósito legal camina en esta dirección.

Finalmente, en relación con aspectos que tiene que ver con el uso y difusión de los recursos digitales, cabe resaltar que las tecnologías permiten nuevas y amplias posibilidades de uso, reutilización, transformación y difusión, pero existen notables resistencias, amenazas y un claro vacío legal que impiden operar con garantías. El marco legislativo del uso legítimo del patrimonio bibliográfico y documental digital está por definir, y los diversos *lobbies* están tratando de amoldar la ley a sus intereses particulares. La Declaración de Vancouver de la UNESCO recomienda a las partes trabajar por una legislación que “garantice la conservación y accesibilidad del patrimonio cultural digitalizado, así como el acceso a ese patrimonio de un modo culturalmente apropiado” (UNESCO/UBC, 2012).

Las organizaciones del ámbito *libraries, archives and museums* (LAM) vienen apostando por una legislación que reproduzca en el ámbito digital los límites que concurren a su favor en la legislación hasta ahora conocida. Es el estado español FESABID está avanzando en esta línea (2013). A nivel mundial la Conferencia Internacional sobre Preservación y Digitalización de la UNESCO, que tuvo lugar en septiembre de 2012, trataba de identificar los marcos legales que faciliten la preservación digital a largo plazo y el fomento del acceso al patrimonio documental digitalizado. Por su parte, el Comité Permanente de Derecho de Autor y Derechos Conexos de la Organización Mundial de la Propiedad Intelectual está trabajando acerca de un instrumento jurídico internacional adecuado sobre excepciones y limitaciones para las bibliotecas y los archivos, y lógicamente está abordando la problemática relacionada con la preservación digital y el uso legítimo de los recursos digitales (OMPI, 2012).

Nuevos usos con la copia digital, tecnológicamente factibles a muy bajo coste, están prohibidos o restringidos por barreras de precio muy elevadas. Si se pretende, por ejemplo, hacer difusión telemática de un diario digitalizado por parte de una biblioteca pública, es considerado comunicación pública, y en la medida en que se trata de obras sujetas a derechos de explotación pertenecientes a sus editores conforme a la Ley de Propiedad Intelectual, aparecen condicionantes, hasta ahora inexistentes, restrictivos con el uso de las colecciones



bibliográficas. Lógicamente las bibliotecas que posean acuerdos de derechos pueden dar difusión a sus fondos digitalizados sin restricciones de ningún tipo.

El acceso a las hemerotecas que los propios diarios ofrecen en sus webs era, de modo generalizado, de pago hace unos años (Abadal, 2001). Sin embargo, actualmente las más importantes hemerotecas retrospectivas son de consulta y acceso gratuitos (véanse por ejemplo, los centenarios de ABC o La Vanguardia, o el caso de El País), ello no quiere decir sin embargo que tales hemerotecas digitales no representen un activo con interés comercial para sus editores (aunque sea en forma de atraer visitas a sus sedes web). Hay casos notables, como el del Diario de Navarra que siguen siendo de pago. Los diarios son editados por empresas privadas que tienen legítimos intereses económicos. Sus hemerotecas o archivos son o pueden ser fuentes de ingresos. De hecho, un estudio realizado por la Asociación para la Investigación de los Medios de Comunicación en el estado español señala que un alto porcentaje de quienes leen prensa en Internet (69,3 %) valoran de forma prioritaria precisamente el poder acceder a la hemeroteca de los diarios, siendo la segunda razón que se alega para acceder a la prensa digital (AIMC, 2011).

Las bibliotecas que quieran hacer comunicación pública han de averiguar quién posee los derechos de explotación (en el caso de los diarios parece claro que serán las empresas editoras de los mismos) y se ha establecer una negociación a fin de obtener los permisos necesarios para poder subir los contenidos a Internet. El acuerdo puede oscilar desde una autorización amplia para poner todos los contenidos en Internet en abierto, hasta una autorización para colocar los contenidos en un repositorio accesible sólo para los usuarios de la biblioteca bajo identificación mediante validación a través de usuario y clave.



Figura 12.- Diarios digitalizados en la Biblioteca Central del Gobierno Vasco, accesibles en sala mediante CD-ROM.

Recordemos que la actual legislación permite a las bibliotecas difundir los contenidos “mediante red cerrada e interna a través de terminales especializados instalados a tal efecto en los locales de los establecimientos” (artículo 37.3 de la LPI), por tanto todo lo que rebase estas restricciones debe contar con la autorización de los titulares de los derechos de explotación de las obras. La literalidad de la ley expone que ello se podrá realizar “siempre que tales obras figuren en las colecciones del propio establecimiento y no sean objeto de condiciones de adquisición o de licencia. Todo ello sin perjuicio del derecho del autor a percibir una

remuneración equitativa.” En el caso del CDPV los diarios que se digitalizan pertenecen a las colecciones de las bibliotecas integrantes del consorcio.

El Consorcio de Digitalización de Prensa Vasca ha optado por limitar la difusión de sus colecciones hemerográficas digitales, considerando que el fondo no es consultable por internet por problemas de propiedad intelectual (Gómez et al., 2012, p. 170). Este es también el caso de la hemeroteca digital de la Universidad Complutense de Madrid o de la BNE cuyos fondos digitalizados deben ser consultados en sala mediante CD-ROM.

Existen experiencias como la del archivo de Murcia, que dio acceso a la prensa a través de Internet y tuvo que retirar la conexión tras un proceso judicial (comentario de D<sup>a</sup> Dolores Rodríguez, de la BNE, en acta del CDPV de 10/05/2005).



Figura 13.- Prensa actual difundida en acceso abierto en el repositorio digital “Hemeroteca Digitala”.

Sin embargo, también son visibles en la red experiencias de varias hemerotecas digitales que han logrado acuerdos para colocar en abierto sus colecciones de prensa digitalizada. Es el caso de la Hemeroteca de Madrid a través de la web Memoria de Madrid, o el de la biblioteca de Koldo Mitxelena Kulturunea que ha puesto en acceso abierto en su hemeroteca digital la colección digitalizada de un periódico que se sigue publicando, se trata de El Diario Vasco, accesible desde 1936 hasta 1980 (Araiz, 2012). Por su parte, JABLE, el archivo de prensa digital de la Universidad de Las Palmas de Gran Canaria, mantiene ciertos títulos con acceso restringido sólo de su intranet.

## 4. ALTERNATIVAS DE ALMACENAMIENTO

---

El informe *Nothing last forever* del proyecto NESTOR hace una llamada de atención sobre el hecho de que, en la práctica, los medios de almacenamiento de datos habitualmente permanecen intactos mientras que los archivos guardados se han convertido en inaccesibles debido al hecho de que los programas requeridos para abrirlos no están disponibles. Con ello se quiere poner de relieve que la obsolescencia del software es el problema al que hay que estar más atento en preservación digital, poniendo en segundo plano la cuestión de los medios de almacenamiento aunque no por ello hay que dejar de prestar a este aspecto la atención debida (Rohde-Enslin y Allen, 2009).

Las colecciones digitales requieren de sistemas de almacenamiento masivo de datos. Hoy día están disponibles en el mercado varias soluciones que permiten articular infraestructuras y sistemas capaces de albergar grandes volúmenes de información binaria. Téngase en cuenta que los grandes proyectos de digitalización de fondos documentales suelen cuantificarse en petabytes (PB) y que los proyectos medianos y pequeños suelen medirse en terabytes (TB).

Las tecnologías de almacenamiento actuales son: tecnología magnética (discos magnéticos y cintas magnéticas), tecnología óptica (CD, DVD, Blu-ray) y tecnología de memoria flash (excepcionalmente rápida, pero de coste muy elevado). La tecnología magneto-óptica prácticamente se ha extinguido y ha desaparecido del mercado. Hay que señalar en cualquier caso, que la tecnología de almacenamiento evoluciona constantemente y se trata de un área muy técnica de la ingeniería informática.

Los datos pueden ser almacenados en diversos tipos de medios o soportes de almacenamiento. En función de las características del medio de almacenamiento se requerirá una unidad o dispositivo de lectura y escritura específico. La guía *Llevando la teoría a la práctica: tutorial de digitalización de imágenes*, establece los siguientes criterios para la evaluación de la tecnología de almacenamiento (Cornell University Library, 2000): velocidad (de lectura / escritura, y de transferencia de datos); capacidad; fiabilidad (estabilidad, redundancia); estandarización (hoy podríamos añadir, de código abierto); costo; y aptitud para la tarea.

Existen diversas clasificaciones de los sistemas de almacenamiento, según atiendan a unos u otros criterios.

### a) *Por su modo de acceso*

Es posible establecer una primera clasificación de los sistemas dependiendo del modo de acceso a los datos: RAM (primario), *online* (secundario), *nearline* (terciario) y *offline* (sin conexión).

La memoria RAM (*Random Access Memory*, memoria de acceso aleatorio) es el almacenamiento primario de los ordenadores, imprescindible para su funcionamiento, presenta una velocidad de acceso a los datos muy elevada y su contenido se borra y desaparece al apagar el ordenador.

El almacenamiento de datos secundario no es imprescindible para el funcionamiento del equipo, está conectado al mismo y forma parte de él. Es de acceso inmediato. Se distingue

entre almacenamiento de conexión directa y almacenamiento conectado a red. El primero incluye unidades dentro de la computadora o cableadas directamente a la misma, presenta límites de ampliación. El segundo se refiere a almacenamiento accesible para múltiples computadoras conectadas a un servidor de red (*online*), es ampliable o escalable a demanda.

El almacenamiento terciario, accesible tras un lapso de tiempo moderado de forma automática, sin intervención humana, mediante sistemas mecánicos robotizados (*nearline*).

Finalmente, está el almacenamiento en soportes no accesibles de forma inmediata, sin conexión directa a ningún equipo. Los soportes físicos que albergan los datos se han de recuperar y cargar manualmente, están *offline* o fuera de línea.

#### b) *Por la velocidad de acceso a los datos*

Otra clasificación posible está en relación con la velocidad de acceso a los datos. Se puede distinguir entre sistemas de almacenamiento de *acceso aleatorio* y sistemas de *acceso secuencial*. Los primeros permiten acceder a cualquier dato en un tiempo muy reducido, independientemente de la localización de los datos dentro del dispositivo. El acceso a los datos en los sistemas secuenciales (como cintas magnéticas) es más lento y dependerá de la localización física del dato en el dispositivo.

#### c) *Por la capacidad de almacenamiento*

Un aspecto determinante en proyectos de bibliotecas digitales que se debe tener en cuenta para la elección de un medio de almacenamiento es la capacidad del mismo. Los rápidos avances en las tecnologías de almacenamiento favorecen la continua ampliación de la capacidad y el abaratamiento de los precios de salida al mercado. A la hora de adquirir dispositivos de almacenamiento se ha de estar atento a la forma de expresar la capacidad.

En informática las medidas se hacen en modo binario: byte (B), kilobyte (kB), megabyte (MB), gigabyte (GB), terabyte (TB), petabyte (PB), exabyte (EB) zettabyte (ZB) y yottabyte (YB); siendo  $1 \text{ kB} = 2^{10} \text{ bytes}$ ,  $1 \text{ MB} = 2^{20} \text{ bytes}$ ,  $1 \text{ GB} = 2^{30} \text{ bytes}$ ,  $1 \text{ TB} = 2^{40} \text{ bytes}$ , etc. Sin embargo, los anuncios comerciales de los dispositivos de almacenamiento masivo emplean el sistema de medidas decimal, siendo así que un disco duro de 1 GB contiene  $10^9$  bytes (1.000.000.000) y no  $2^{30}$  bytes (1.073.741.824), perdiendo por tanto un 7,4% de la capacidad. En el caso del TB se pierde hasta un 10% de la capacidad de almacenamiento. Por tanto, cuando se compra un dispositivo de 750 GB en realidad se está adquiriendo un disco con capacidad para 698 GB y cuando se compra un disco de 2 TB en realidad se está adquiriendo uno con capacidad para 1,82 TB (Herrerías Rey, 2012). Es por ello, que se ha establecido la nomenclatura de las medidas en modo binario para uso cuando se quiere hacer referencia a la medida binaria exacta: Kibibyte (KiB), Mebibyte (MiB), Gibibyte (GiB), Tebibyte (TiB), Pebibyte (PiB), etc. Las casas comerciales siempre utilizan medidas de capacidad en base decimal.

Para seleccionar un sistema de almacenamiento se deben tener en cuenta aspectos como: el volumen de datos a almacenar, el crecimiento previsto, el modo de almacenamiento (*online*, *offline* o *nearline*), el tamaño de los ficheros, la frecuencia de acceso estimada, la velocidad

requerida de acceso a los datos, la infraestructura tecnológica con que se cuenta, la infraestructura de depósitos acondicionados de que se dispone, el espacio de almacenamiento de soportes del que se dispone en esos depósitos, el presupuesto con que se cuenta y la disponibilidad presupuestaria de ejercicios venideros para el mantenimiento y sostenibilidad del sistema establecido. Tampoco estará de más valorar si se cuenta con personal capacitado tecnológicamente para manejar el fondo digital, el sistema de almacenaje con que se cuenta y la infraestructura informática.

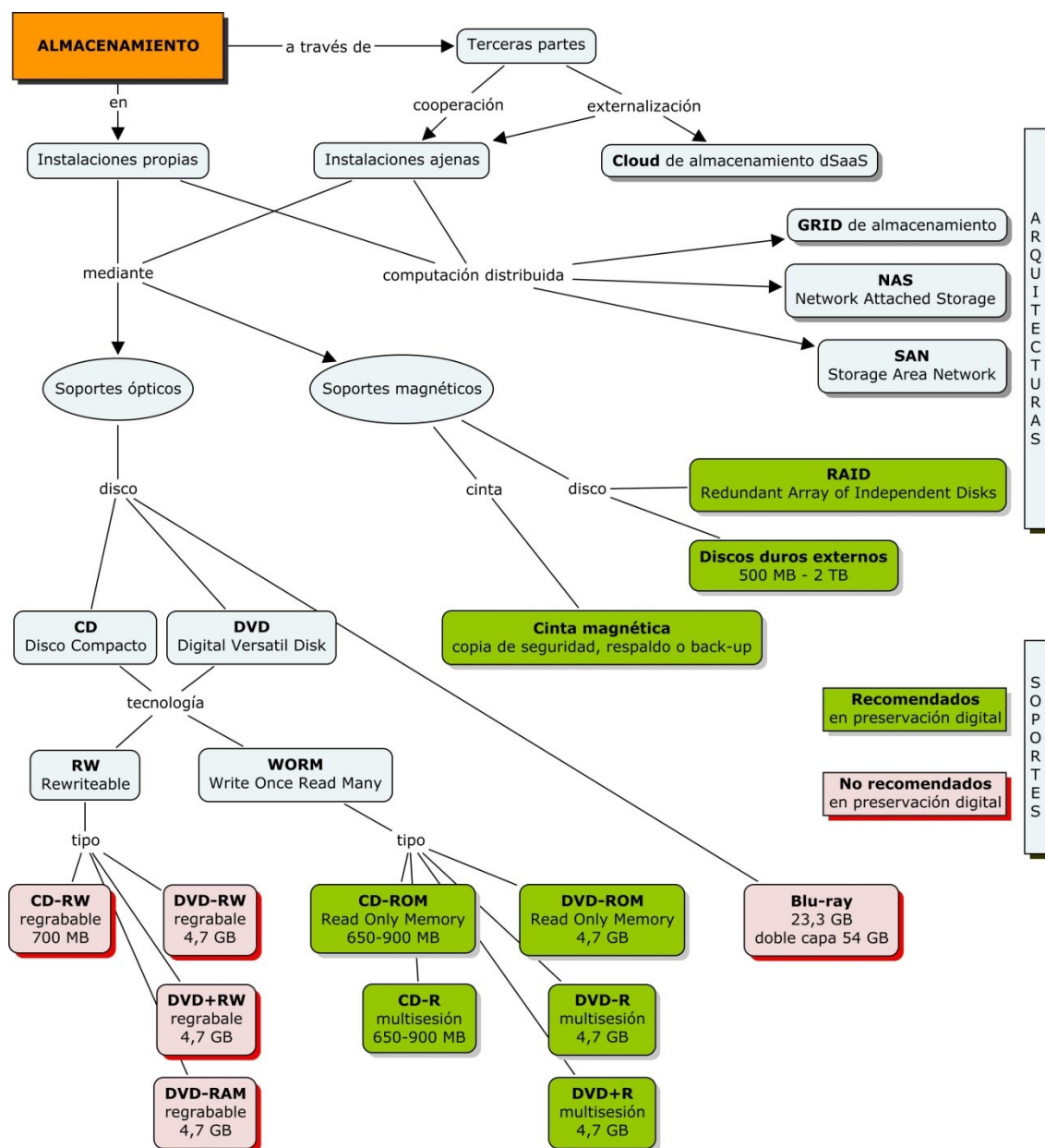


Figura 14.- Mapa conceptual de medios de almacenamiento para la preservación digital (Elaboración propia)

También hay que tener en cuenta a la hora de seleccionar un sistema u otro, que la velocidad media de un dispositivo y el tiempo de acceso a los datos dependerá siempre del tipo de dispositivo, del tipo de datos y de la modalidad de acceso a los datos. A nivel técnico, las

prestaciones de los sistemas de almacenamiento se miden conforme a dos parámetros: capacidad de almacenamiento (en bytes) y tiempo medio de acceso a los datos (en milisegundos). En los sistemas actuales ambas magnitudes son inversamente proporcionales, de modo que a mayor capacidad menor velocidad de acceso. El tiempo medio de acceso va desde los 0,04 milisegundos de un disco RAM, pasando por los 0,4 milisegundos de un disco flash, hasta los 12 milisegundos de un disco fijo. Las capacidades de almacenamiento son de 700 MB en un CD-ROM, 4,7 GB un DVD, entre 1 y 32 GB la memoria RAM, 50 GB un disco Blu-ray, hasta entre 4 y 8 TB un disco duro.

En el momento de la adquisición de los medios de almacenamiento se ha de tener en cuenta que sean de buena calidad, estandarizados y observar que no tengan fallos de fabricación. En cualquier caso, se recomienda como norma general almacenar los ficheros máster en soportes digitales de alta calidad. Así mismo, las unidades grabadoras y lectoras deben ser también de buena calidad y estar en correcto estado de funcionamiento.

En el momento de la grabación se ha de poner cuidado en su correcta manipulación, elegir una velocidad de grabación inferior a la máxima, grabar en sesión única (no usar la grabación multisesión), comprobar que la grabación ha sido satisfactoria e identificar los soportes mediante etiquetas en las cajas y no en el propio soporte. El proceso debe ser debidamente documentado.

Finalmente, se ha de tener en cuenta que en toda la tecnología de almacenamiento se ha de distinguir entre formatos lógicos y formatos físicos, haciendo referencia los primeros al modo de codificación de los datos más allá de las características físicas del soporte de almacenamiento. Es por ello que sistemas aparentemente iguales generan problemas de incompatibilidad por no reconocimiento de los formatos lógicos en que los datos están codificados.

Según se detalla en el tutorial de la Cornell University Library (2000, p. 81), las características de cada tecnología vienen definidas por:

- “el sistema de almacenamiento subyacente (magnético, óptico o magnetoóptico),
- el tipo de unidad (fija o removible),
- el material del medio (cinta, disco rígido, disco flexible), y
- la interfaz de hardware (ATA, ATAPI, SCSI, USB, Fireware / IEEE 1394, Canal de Fibra)”.

Este último elemento, la interfaz de hardware, que es el modo de conexión y acceso a los datos desde el exterior, es determinante en la velocidad con que se pueden recuperar los datos o velocidad de transferencia.

#### **4.1. MEDIOS Y SISTEMAS DE ALMACENAMIENTO**

En las siguientes líneas se detallan brevemente las características más reseñables de los principales sistemas y medios de almacenamiento de datos que actualmente se están utilizando en programas de archivo de grandes volúmenes de ficheros de imagen generados en proyectos de digitalización masivos.

#### 4.1.1. TECNOLOGÍAS ÓPTICAS (CD, DVD, BLU-RAY)

La tecnología óptica permite leer los datos codificados en un soporte mediante un haz óptico, normalmente un láser de una longitud de onda concreta, dependiendo del sistema. Los soportes ópticos, en general, son de acceso más lento a los datos y menos fiables que los sistemas de almacenamiento en disco duro. Se necesita siempre una unidad lectora/grabadora para poder acceder a los datos contenidos en los discos. Aunque el mercado evoluciona constantemente hacia la estandarización y compatibilidad entre productos y sistemas, es frecuente encontrarse con unidades de lectura y escritura que no son capaces de reconocer a otros sistemas ópticos.

Según Herrerías Rey (2012), las ventajas de la tecnología óptica para el almacenamiento de datos son:

- Los datos no se alteran por efecto de los campos magnéticos.
- Los soportes están hechos con materiales a los cuales apenas afecta la humedad y pueden aguantar golpes.
- Los datos apenas sufren deterioro por el paso del tiempo, son muy estables.

Las tecnologías ópticas presentan el inconveniente de no poder mantener la información en línea, pues se debe dotar al sistema con un *juke-box* que sólo permitirá (si la colección es mediana o grande) una configuración *nearline*. Son buena alternativa como copia de respaldo de una colección pequeña, dado que presentan fácil portabilidad, tienen capacidad de datos alta y se pueden almacenar de modo sencillo. Sin embargo, requieren refrescos periódicos en lapsos de cinco años, más o menos. El almacenaje en discos ópticos comporta cierto riesgo de pérdida de datos, el cual se reduce usando discos de buena calidad, aunque ello no evita tener que almacenarlos en depósitos debidamente acondicionados.

Sobre los métodos de escritura cabe señalar que pueden ser monosesión o multisesión. El primer método crea una única sesión en el disco y no permite que se añadan datos nuevos más adelante. A diferencia del método anterior, la grabación en multisesión permite que el soporte óptico se escriba varias veces, añadiendo nuevos datos en cada una de las sesiones hasta completar la capacidad del disco.

Los discos compactos (CD) son sistemas de almacenamiento óptico, que tienen un tamaño estándar de 12 cm y se componen de una o varias microláminas metálicas superpuestas sobre un corazón de policarbonato de plástico y recubiertas con una capa protectora de laca. Las capas metálicas son de aluminio, aunque también se utiliza oro y plata, dado que presentan mejor comportamiento de cara a su conservación y longevidad aunque encarecen el producto. Los datos se registran en microsurcos grabados en espiral, que suelen ser capaces de almacenar entre 650 y 900 MB de datos, siendo lo más habitual capacidades de 700 MB.

Existen varios tipos: discos WORM (*Write Once Read Many*) que pueden ser grabados por el usuario, la grabación no puede ser borrada ni reescrita, aunque sí se puede grabar en varias sesiones hasta completar la capacidad de almacenamiento de datos del disco; discos ROM (*Read Only Memory*) de sólo lectura, son una sub-tipología de los anteriores, pero son estampados en fábricas y comercializados, no pueden ser escritos por los usuarios; discos reescribibles, que pueden ser grabados, borrados y regrabados, por ello no son válidos con

finés de conservación. Estas tres tipologías se identifican en el mercado como CD-R, CD-ROM y CD-RW.

Las diferencias de calidades entre discos ópticos se debe al hecho de que se usan distintos materiales tanto para la capa reflectora (oro, plata y aluminio) como para la capa de soporte de datos (cianina y azon), dando lugar a distintas coloraciones (Rohde-Enslin y Allen, 2009). Existen discos de color dorado, verde y azul, siendo los primeros los que presentan mejor calidad y mayor esperanza de vida, seguido de los verdes y por último los azules. Comúnmente se acepta que los discos de oro y plata son los más estables y de mayor calidad.

En relación con los discos de alta calidad a la venta en el mercado fabricantes como DELKIN comercializan modelos como "*Archival Gold CD-R*" bajo la promesa de duración de "300 años", siendo el precio de una caja de 100 discos de 190 €. La casa MAM-A vende discos de oro de 24 Kt. "*Gold Archive CD-R*" también con promesas de 300 años de longevidad.

Los discos ópticos DVD son de similares características a los CD, pero el hecho de que los microsurcos para la grabación de datos (*pits*) sean más pequeños permite almacenar mayor cantidad de información, siendo el tamaño habitual 4,7 GB (que se corresponden con 4,38 GiB). Pueden alcanzar capacidades de hasta 17 GB en caso de que sean de doble cara (*dual side*) y de doble capa (*dual layer*).

Los modelos regrabables, no recomendados en ningún caso para la conservación, se identifican como DVD-RW, DVD+RW y DVD-RAM, dependiendo de la patente del fabricante. Por otra parte, están los de tecnología WORM, como son los DVD-ROM, y DVD-R y DVD+R que son multisesión y varía la tecnología en relación con la patente y tecnología de los fabricantes. Es por ello que hay que estar atento a la compatibilidad de los dispositivos con los distintos sistemas, puesto que aunque las últimas unidades lectoras que van saliendo al mercado son capaces de trabajar tanto con DVD-R como con DVD+R, no todas las unidades lo son.

En este caso, los discos de oro (también llamados dorados) o plata presentan mejor comportamiento que los discos de aluminio. En el mercado hay cierta variedad de casas comerciales; se presentan algunos modelos y precios, a título informativo y sin ánimo de exhaustividad:

- DELKIN *Archival Gold DVD-R* 100 años, caja de 100 unidades de 4.7GB, 260 €.
- VERBATIM *DVD-R Gold Archival DataLifePlus 4.7GB*, diseñado para un uso ilimitado y una gran longevidad ofrece la mayor vida útil de archivado posible. Archivado a largo plazo. Presentación: 25 Pack 49,00 € (IVA incluido).
- VERBATIM *Archival Grade Gold DVD-R* diseñado para alcanzar 100 años, 4.7GB.
- NANOARK *Archival Quality Gold DVD-R* con una expectativa de vida de 80 años.
- MITSUI *Gold Archival DVD-R* y *DVD+R* Comercializados para almacenamiento de datos de cara a la preservación digital a largo plazo, expectativa de vida superior a los 100 años. 100 discos en caja rígida a 95 €.
- FALCONMEDIA *Archival Gold DVD-R* almacenamiento seguro de la información por más de 100 años.



- MAM-A 24 Kt. *Gold Archive* DVD+/-R 116 años de longevidad.

El Sistema Blu-ray, también conocido como Blu-Ray Disc o BD, es el último miembro en salir al mercado de la familia de los discos ópticos. Se trata de un formato de disco óptico con capacidad de almacenamiento mayor al DVD, y del mismo modo que éste y que el CD mide 12 cm, y su aspecto externo es idéntico a éstos. Puede almacenar 23,3 GB, y en configuraciones de doble capa hasta 54 GB. Desarrollado por la Blu-ray Disc Association, un conjunto de compañías tecnológicas lideradas por Sony y Philips. Otras compañías y agrupaciones de compañías han desarrollado otros sistemas y tecnologías de almacenamiento óptico de gran capacidad, como HD DVD, HD-VMD o HVD, pero los discos BD se han impuesto.

Su funcionamiento se basa en un rayo láser de tonalidad azul (de ahí su nombre) cuya longitud de onda permite grabar en menor espacio, lo cual permite almacenar más cantidad de datos en la misma superficie de disco que la utilizada por la tecnología de los CD o DVD.

La penetración de esta tecnología en el mercado está siendo lenta y desigual debido a su elevado precio, puesto que las unidades grabadoras lectoras rondan todavía los 300 €, aunque es cierto que las unidades más modernas presentan compatibilidad hacia atrás con los diversos sistemas lógicos de CD y DVD. Veamos dos ejemplos de discos comercializados: TDK Blu-ray Disc BD-R *Gold-colored Disc*, 25 GB, usa una tasa de metales reducida lo cual permite evitar la corrosión y favorecer su larga permanencia; DELKIN *Blu-Ray Archival Gold* BD-R, 25 GB, 200 años de duración, la caja de 25 discos se comercializa por 215 €.

Sobre el uso de la tecnología Blu-ray para la preservación digital a largo plazo existen dudas sobre su penetración y su permanencia en el mercado, lo cual otorga a esta tecnología cierto riesgo de obsolescencia tecnológica (Henriksen, Seuskens, y Wijers, 2013; Rohde-Enslin y Allen, 2009). El informe *Nothing last forever* del proyecto NESTOR, expone literalmente que “en resumen, consideramos a los discos Blu-ray como un medio de archivo de datos inadecuado” (2009, p. 48).

#### 4.1.2. TECNOLOGÍAS MAGNÉTICAS (DISCOS DUROS, CINTAS MAGNÉTICAS)

La tecnología magnética se basa en la capacidad de almacenamiento de información aprovechando la capacidad de determinados materiales para almacenar de forma permanente un determinado estado magnético. Según Herrerías Rey (2012), las ventajas de la tecnología magnética para el almacenamiento de datos son:

- El almacenamiento de los datos es permanente (al contrario que la memoria RAM, no necesita de un flujo eléctrico continuado).
- Puede alterarse en cualquier momento para reutilizar el soporte (almacenamiento de lectura y escritura).
- Son de bajo precio, respecto a su gran capacidad (<0,08 € el GB en diciembre de 2011).
- Gran velocidad de acceso a los datos.
- Gran capacidad de almacenamiento de datos.

Y según el mismo autor las desventajas son:

- Los soportes son bastante delicados, puesto que pueden verse afectados por las altas y las bajas temperaturas, por la humedad, por los golpes y sobre todo por los campos magnéticos.

Herrerías Rey afirma que “la gran velocidad de acceso y la alta densidad de almacenamiento de datos de la tecnología magnética de almacenamiento resultan muy superiores a las alcanzables por la tecnología óptica (como la del CD o DVD)” (2012, p. 359).

Existen gran cantidad de dispositivos basados en esta tecnología, diferenciándose por el tipo de soporte: hilo de acero en los primeros diseños, varios tipos de cintas y discos recubiertos de material magnético como discos duros, disquetes, o discos Zip. Las unidades de disco duro, tradicionalmente y hasta hace escasas fechas, han estado siempre basadas en tecnología magnética. Recientemente ha aparecido en escena la tecnología flash, dando origen a los discos de estado sólido.

Las unidades de disco duro (HDD, *Hard Disk Drive*) se fabrican normalmente en formato de 3,5” para poder ser ensambladas en las carcasas de los ordenadores. Emplean interfaces (métodos de conexión al resto del PC) de serie SATA en ordenadores personales y SAS en sistemas profesionales, que determinarán a velocidad de transferencia de datos.

Por otro lado están los discos duros externos (*External Hard Disk*), que también son conocidos como discos duros portátiles dada la facilidad para su transporte sin necesidad de conexión eléctrica ni consumo. Ideados en principio pensando en su portabilidad, no de cara a la permanencia de los datos, los fabricantes nunca garantizan más allá de los cinco años de duración. Se trata de un soporte de almacenamiento magnético. Sus capacidades de almacenamiento más habituales son 500 GB y 1 TB, pero ya se están comercializando desde hace un tiempo con tamaños de 2 TB y 3 TB. Los discos SCSI y ATA requieren conexión eléctrica con transformador, los discos *Firewire* también requieren conexión eléctrica, pero sin transformador. Los discos con conexión USB no necesitan de una toma eléctrica. USB 3.0 multiplica por diez la velocidad de acceso a los datos en comparación con la tecnología USB 2.0.

Los discos de estado sólido (SSD, *Solid State Drives*) tienen la apariencia externa de un disco duro, normalmente están basados en memoria flash, son más rápidos, más silenciosos y de menor consumo, pero de un precio muy elevado, que ronda los 1,30 € el GB en diciembre de 2011, según Herrerías Rey (2012).

El uso de cinta magnética es una solución configurable *offline*, *nearline* u *online* incrementándose el precio según la configuración. En la actualidad la principal utilidad de las cintas magnéticas es para el grabado de copias de seguridad de la información contenida en discos. Existe gran diversidad de tipos y distintas tecnologías de unidades de cinta magnética (Prieto, Lloris, y Torres, 2010):

- Cintas clásicas o de carrete (cintas ½” de ancho)
- Cartuchos con cintas de ¼”, ½”, u 8 mm
- Cartuchos compactos (cartuchos pequeños)

- QUIC (*Quarter Inch Cartridge*)

- DAT (*Digital Audio Tapes*)
- EXABYTE
- SAIT (*Super Advanced Intelligent Tape*)
- DLT/SDLT (*Digital Linear Tape/Super DLT*)
- LTO (*Linear Tape Open*)

Tabla 11.- Precios estimados para diferentes tipos de medios de almacenamiento (Fuente: Henriksen et al., 2013)

	HDD	SDD	LTO Tape
<b>Media cost per TB</b>	€ 91	€ 2361	€ 36
<b>Max capacity of one unit</b>	3 TB	2 TB	3 - 6 TB
<b>Archive life according to manufacturers</b>	Three - five years	Five - seven years	Seven - ten years
<b>Total cost of ownership (TCO)<sup>27</sup></b>	High	Very high	Medium
<b>Maintenance per year for &lt;100TB</b>	Medium	-	High
<b>Maintenance per year for &gt;100TB</b>	High	-	Medium

La Junta de Castilla y León en las *Recomendaciones para la digitalización de documentos en los archivos*, publicadas en 2011, admite como sistema para la conservación a medio-largo plazo de ficheros maestros, en caso de no contar con repositorios o servidores, el formato LTO. Se han desarrollado seis generaciones de este formato, la última (LTO-6) tiene una capacidad de 3,2 TB (velocidad 270 MB/s) (Carnicer Arribas et al., 2011). Se trata una tecnología desarrollada como alternativa de estándares abiertos a los formatos de cinta magnética patentados.

Para establecer una solución *offline* o *nearline* las unidades grabadoras oscilan hoy día en el mercado entre los 1.800 y los 4.000 €. Por su parte, las cintas, que son de una sola escritura (WORM), son comercializadas por las principales marcas del mercado a precios que van de 70 a 130 €.

En relación con el asunto de los precios, y a pesar de su difícil cálculo sino es a través de estimaciones, el reciente estudio *Best practices for a digital storage infrastructure for the long-term preservation of digital files* ha presentado un acercamiento a esta cuestión, dando lugar a análisis comparativos como el que se presenta en la Tabla 11.

#### 4.1.3. ARQUITECTURAS EN RED

La configuración de red es independiente del tipo de sistema de almacenamiento, si se trata de un conjunto de discos duros (RAID), bibliotecas de cintas, *juke-box* de discos ópticos o conjuntos de discos SDD. Para acceder de forma remota los archivos almacenados existen diferentes configuraciones de acceso a los dispositivos conectados. Estas redes se conectan a los dispositivos de almacenamiento a través de un host. Las diferentes tipos de

configuraciones de red que actualmente se implementan son los sistemas NAS (*Network-Attached Storage*, almacenamiento conectado a red) y los sistemas SAN (*Storage Area Network*, red de área de almacenamiento). Ambos son transparentes para el usuario final y permiten tener los datos a mano como si estuviesen albergados en el propio ordenador (Henriksen et al., 2013; Herrerías Rey, 2012).

La instalación de toda infraestructura de equipos informáticos o de un centro de datos, que es la ubicación donde se concentran los recursos necesarios para el procesamiento de la información de una organización, ha de disponer de una ubicación acondicionada y especial. Se deben considerar varios factores a tener en cuenta: adecuada localización, diseño apropiado del espacio, ajustada refrigeración, seguridad, fuente de alimentación ajustada a las necesidades y organización de equipamiento.

Los sistemas RAID consisten en un conjunto de discos duros conectados con la finalidad de lograr grandes capacidades de almacenamiento y rapidez en el acceso a los datos. El acrónimo RAID (*Redundant Array of Independent Disks*) significa conjunto redundante de discos independientes. Es una alternativa que permite colocar grandes cantidades de datos en línea, por lo cual se está imponiendo en proyectos de fondos de imágenes digitalizadas. La redundancia de los datos permite alcanzar altas cotas de fiabilidad en estos sistemas. La fiabilidad de los sistemas RAID no exime de tener que realizar *backups*.

Es un sistema versátil en el sentido de que puede ir escalándose. Como consiste en una serie de discos conectados, a medida que se va necesitando mayor espacio de almacenamiento se pueden ir conectando más discos. El resultado es transparente para el usuario que percibe la arquitectura como si se tratase de un único disco independientemente de cuantos estén conectados en el conjunto. Se implementan mediante dispositivos y sistemas NAS o SAN.

Los sistemas compuestos por *clusters* de computadores están pensados para compartir ordenadores estableciendo una plataforma con potencia de procesamiento comparable a la de un supercomputador, pero no están pensados para lograr gran espacio de almacenamiento. Para aprovechar mejor los recursos de distintos centros están los sistemas GRID, que es un sistema distribuido de *clusters* y otros recursos o colección de recursos distribuidos conectados a una red, situados en diferentes dominios administrativos y que son accedidos por usuarios y aplicaciones buscando reducir sobrecostes e incrementar el rendimiento. La computación en Grid es especialmente interesante en la computación intensiva de datos, para computación colaborativa y para computación multimedia.

#### 4.1.4. REPOSITORIO DE ACCESO ABIERTO BAJO ESTÁNDAR OAIS

Según Ferreras Fernández un repositorio “es un conjunto de servicios prestados (...) para recopilar, administrar, difundir y preservar la producción documental digital generada en la institución, cualquiera que sea su tipología, a través de la creación de una colección digital organizada, abierta e interoperable a través del protocolo OAI-PMH (protocolo para la recolección de metadatos) para garantizar un aumento de la visibilidad e impacto de la misma” (2010, p. 7). Esta autora destaca sus principales características: interoperabilidad, recolección, difusión y accesibilidad. Lluca y Reoyo por su parte, dicen que “sirve para publicar en Internet, normalmente con acceso libre, una base de datos que incluye los documentos

digitales o digitalizados que forman los fondos y colecciones de una unidad de información.”(2012, p. 4).

En relación con el hardware y software que soportará el repositorio, Lluca y Reoyo plantean la posibilidad de elegir entre montar un servidor mediante un sistema de arquitectura informática local o externa, lo cual dependerá de varios factores, pero fundamentalmente de la capacidad económica de la organización. “Si la institución no tiene un equipo informático estable y personal técnico, será más rápido y barato empezar ubicando el repositorio en un sistema externo. Pero a largo plazo será más caro mantener esa estructura externa. Otros factores que determinarán el grado de complejidad de la arquitectura son el volumen de documentos a depositar (especialmente en el caso de documentos muy "pesados", como las imágenes), y el nivel de consultas de los usuarios potenciales.”(2012, p. 7).

Estos autores creen que un sistema básico de infraestructura técnica local para un repositorio se podría montar sobre la siguiente arquitectura “un nodo de la serie Proliant DL360 (HP, o su equivalente en Sun o Dell) con 2 procesadores de 6 núcleos más 24 GB de RAM y un disco duro interno de 300 GB” (2012, p. 8), estableciendo su coste inicial de unos 9.000 USD. Como ellos mismos remarcan “la tendencia mundial es implementar los repositorios con software libre.”

#### 4.1.5. CLOUD COMPUTING

Cabe definir el *cloud computing* como el uso de recursos de cómputo escalables ofertados como un servicio ajeno al entorno que los utiliza, habitualmente mediante pago por uso.

Recientemente el Instituto Nacional de Tecnologías de la Comunicación (INTECO), centro de ciberseguridad, ha publicado la *Guía para entidades locales: cómo ahorrar costes y mejorar la productividad con cloud computing* (Pérez San-José, De la Fuente Rodríguez, Álvarez Alonso, García Pérez, y Gutiérrez Borge, 2013), según esta guía el *cloud computing* “es un modelo tecnológico que permite el acceso ubicuo, adaptado y bajo demanda en red a un conjunto compartido de recursos de computación configurables (por ej. redes, servidores, almacenamiento, aplicaciones y servicios), que pueden ser rápidamente aprovisionados y liberados con un esfuerzo de gestión reducido o interacción mínima con el proveedor del servicio.”(2013, p. 4)

Ventajas del cloud de almacenamiento dSaaS (*Data Storage as a Service*):

- Sólo se paga por el espacio utilizado. Se usa lo que se necesita y se paga por lo usado. Tarificación por almacenamiento y transferencia.
- La información es accesible permanentemente. Se puede acceder a cualquiera de los recursos que están en la nube en cualquier momento y desde cualquier sitio en internet. Alta disponibilidad.
- Se evitan tareas de mantenimiento.

Y desventajas:

- Seguridad en el uso. Seguridad de los datos, confidencialidad de la información. Necesidad de garantías de privacidad por parte de los anfitriones

- Rendimiento, velocidad de acceso.
- Marco regulatorio y legal indefinido, incompleto, complejo y de alcance internacional.
- No viable en administraciones públicas por infracciones LOPD, necesidad de encriptamiento con lo cual el proceso se complica y se encarece.

El almacenamiento en cloud evitaría el refresco o la migración de soportes, así como la gestión física de la colección y el control de los depósitos de almacenamiento.

Existe hoy día en el mercado una amplia oferta de servicios de alojamiento de datos mediante *cloud computing*, estando la mayoría de ellos orientados al sector de público en general o de consumo de hogar. Suelen ofrecer *hosting* gratuito de entre 2 y 25 GB. Entre los servicios más conocidos se pueden citar: Dropbox 2 GB; SkyDrive 25 GB; SugarSync 5 GB; iCloud 5 GB; OpenDrive 5 GB; Mozy 2 GB; iDrive 5 GB; YouSendIt 5GB; Copy 5GB; o Google Drive 15 GB. Se trata de un sector muy cambiante y en estos momentos en plena efervescencia. Todos estos servicio tienen modalidades *premium* que ofertan mayores volúmenes de almacenamiento pero con coste asociado.

Por otra parte, están los servicios orientados a almacenamiento de grandes volúmenes de datos, pensados para empresas y con coste económico. Las política de precios suelen incluir varios epígrafes, estableciéndose las tarifas conforme a criterios como:

- capacidad de almacenamiento, con rangos de precio por volumen de datos a albergar
- transacciones de subida y bajada de datos
- acceso y consulta de datos
- redundancia geográfica
- eliminación de datos
- restauración de datos
- ...

Bajo esta clasificación se pueden enumerar servicios como Windows Azure, Amazon S3, Arsys Cloud Storage, Acens Cloud Storage, Google Cloud Plataform, LiveDrive Cloud Storage, UltiCloud, y muchos otros. Los precios actuales de mercado son muy elevados, no están basados en costes reales sino en valor de mercado y al contrario que en el caso de las tecnologías tradicionales de almacenamiento no están bajando, por ello se puede concluir que actualmente es una opción muy poco recomendable de cara a la preservación digital de grandes volúmenes de datos (Henriksen et al., 2013).

Finalmente, en el mundo bibliotecario han aparecido soluciones de almacenamiento en *cloud* establecidas como proyectos cooperativos y orientados a la preservación digital como LOCKSS, DuraCloud o HatiTrust. Se trata principalmente de iniciativas orientadas a garantizar el acceso permanente a colecciones de revistas actuales de las cuales se poseen los derechos de difusión y de fondo patrimonial en dominio público. Con relación a LOCKSS, por ejemplo, Castillo y Jorba afirman que “cuando el contenido a preservar no es necesariamente público, o no es accesible vía web, LOCKSS no es aplicable” (2008).

## 4.2. CONSIDERACIONES SOBRE ALMACENES Y DEPÓSITOS CONTROLADOS

Los depósitos para el almacenamiento de información en soporte digital han de tener la temperatura y humedad relativa conforme a unas constantes controladas, sin oscilaciones, sin contaminación por campos magnéticos, con controles de plagas y limpios. Han de estar cerrados y se deben controlar los accesos de personas.

**General recommendations for long-term storage conditions:**  
 For archiving recordable (R) discs, it is recommended to use discs that have a gold metal reflective layer.

*Archival Storage Facility—Recommendation for storing CDs and DVDs together*

Media	Temperature	Relative Humidity (RH)
CD, DVD	Less than 20°C (68°F) Greater than 4°C (39°F)	20% to 50% RH

A temperature of 18°C and 40% RH would be considered suitable for long-term storage.  
 A lower temperature and RH is recommended for extended-term storage.

Figura 15.- Condiciones de almacenamiento para CD y DVD. (Fuente: Digital Media Group, 2007)

En la Figura 15 se pueden ver las recomendaciones generales que hace *Digital Media Group* sobre condiciones de almacenamiento a largo plazo para soportes ópticos como CD y DVD.

Diversos autores han ofrecido datos sobre la esperanza de vida de los distintos soportes. Se trata habitualmente de arcos temporales que oscilan según varíen las condiciones ambientales de temperatura y humedad relativa en que estén almacenados. Lluca, citando a Jones y Beagrie (2008), aporta los datos que se pueden ver en la Tabla 12 (Lluca, 2012a).

Tabla 12.- Condiciones ambientales de almacenamiento de soportes y años de duración. Tomado de la síntesis realizada por Lluca (2012a) a partir de Jones y Beagrie (2008)

Soporte	25 h.r. /10 C	40 h.r. /20 C	50 h.r. /28 C
D3 cinta magnética	50 años	15 años	1 año
DLT cinta magnética	75 años	15 años	1 año
CD/DVD	75 años	20 años	2 años
CD-ROM	30 años	3 años	3 años

Por otra parte, existen una serie de recomendaciones básicas que deben seguirse con el objetivo de minimizar los riesgos de deterioro de los soportes de almacenamiento. Hacen referencia al momento previo a su uso (fabricación, adquisición), al momento de uso (grabación, lectura y manipulación) y al momento posterior, esto es, al almacenamiento. El manual *Preservation management of digital materials* relaciona las siguientes recomendaciones generales que atañen a diversos tipos y sistemas de almacenamiento (Beagrie y Jones, 2008):

- Mantenga las áreas de almacén y acceso a los depósitos libres de humo, polvo, suciedad y otros contaminantes.
- Almacene los soportes magnéticos lejos de campos magnéticos fuertes.

- Transporte los medios magnéticos en recintos con espacio libre de 50 mm.
- Almacene en un lugar fresco, seco y estable y seguro.
- Aclimate los soportes antes de su uso.
- Utilizar los medios y dispositivos de alta calidad.
- Mantenga los dispositivos de lectura bien cuidados y limpios.
- No coloque etiquetas en los discos ópticos ni marque con un bolígrafo o un lápiz.
- Siga las recomendaciones del fabricante para el etiquetado.
- Reduzca al mínimo el manejo y uso de los medios de archivo, el número de accesos y usos, y refresque los soportes.
- Grabe copias de archivo desde diferentes dispositivos y con diferente software.
- Haga copias de archivo de medios similares o idénticos comprados a diferentes proveedores.

Los equipos de lectura deben estar en buen estado de uso, pues de lo contrario los dispositivos de lectura pueden degradar los datos. Como ya se ha dicho, se ha de hacer una manipulación adecuada y cuidadosa de los soportes y medio, y en relación con la manipulación de soportes ópticos es muy interesante la breve guía *"CD and DVD Archiving: Quick Reference Guide for Care and Handling"* (Digital Media Group, 2007), que se adjunta en el Anexo IV.



## 5. PROPUESTA FUNDAMENTADA

---

Tras abordar a lo largo de las páginas precedentes de este estudio los aspectos relativos al estado del arte en preservación digital, al análisis del fondo del Consorcio de Digitalización de Prensa Vasca y al estudio de las alternativas sobre infraestructuras de almacenamiento digital, en el presente apartado se plantean, a la vista de los datos extraídos, las diversas posibilidades para configurar un sistema de almacenamiento para la preservación digital a largo plazo del fondo del Consorcio de Digitalización de Prensa Vasca.

Como señala Alice Keefer, la preservación digital es un reto al que, antes o después, habrá de enfrentarse toda institución y “más vale que empiecen con pasos modestos pero coherentes, siguiendo las mejores prácticas ya identificadas” (2007, p. 199). A lo largo de este estudio se ha prestado atención a los retos legales, financieros e institucionales que, más allá de los retos tecnológicos, suponen los verdaderos obstáculos para encarar con éxito la preservación digital.

Un condicionante que afecta a las bibliotecas es que no hay un límite temporal para la preservación de los objetos digitales. Al contrario de lo que ocurre en archivística donde hay un mandato legal que obliga a retener cierta documentación por un periodo dado, en las bibliotecas se ha de preservar el material “para la posteridad” sin un límite temporal marcado a priori (Castillo y Jorba, 2008). Por ello, es de especial importancia atender minuciosamente al proceso de selección de los materiales digitales que serán objeto de preservación, dado que cuanto mayor sea el volumen de información electrónica a conservar, mayor será el coste económico de la preservación digital (Wilson, 2011).

La guía del TC46 de la ISO afirma que “hay suficiente experiencia para sugerir que la adopción de procedimientos de preservación basados en estándares de sistemas y prácticas generalmente aceptadas dará como resultado documentos adecuadamente preservados” (Elvira, 2011, p. 474), siempre que se establezcan y ejecuten controles de forma rutinaria, y se sigan los procesos preestablecidos por parte del personal designado. Estos controles y procedimientos deben comprobar:

- Formatos y soportes en uso.
- Copias de seguridad y su estado.
- Análisis de riesgos.
- Control de calidad de los documentos y sistemas.
- Seguridad, acceso y legibilidad.

La solución o estrategia que se desee implementar debe tener en cuenta que cualquier gasto o inversión a realizar deberá ir acorde a los riesgos. Antes de decantarse por una alternativa concreta hay que medir los riesgos reales del fondo que se tiene entre manos, y hacer un balance equilibrado entre estos y el coste de la elección a implementar (Muñoz de Solano, 2006; Wilson, 2011). Es importante no incurrir en costosas operativas de trabajo en fondos con escaso riesgo de pérdida.

En relación con las actuaciones que se lleven a cabo con los ficheros (migración de formatos, descripción conforme a estándares de metadatos, etcétera), se ha de garantizar que estas vayan siempre encaminadas a asegurar la interoperabilidad para el intercambio y transferencia de información en el futuro.

En los próximos años habrá que estar atento a las herramientas de software de gestión de la preservación digital que se están desarrollando, pues la mayoría son de código abierto, basadas en XML y adaptables a cualquier entorno de trabajo (Térmens, 2009a), y facilitarán las tareas de gestión en las organizaciones.

## 5.1. ESTRATEGIAS DE PRESERVACIÓN VIABLES

Existen una serie de acciones, enmarcadas en estrategias ejecutables a corto y medio plazo, que el Consorcio puede llevar adelante sin incurrir en costes económicos de consideración. Se trata de acciones como gestionar de modo adecuado las copias de respaldo existentes o definir unas especificaciones técnicas detalladas para los proveedores. La segunda serie de acciones realizables, pero que muy posiblemente acarrearán costes económicos mayores, son tareas como la generación de *backups* actualizados, la migración de formatos en algunos pocos casos, la descripción con metadatos, el refresco. En un tercer estadio habrá que plantear que solución para el medio largo plazo es factible, si hay compromiso de las partes, y abordable, si hay presupuesto para ello.

### 5.1.1. ESTRATEGIA 1: REUBICACIÓN DE LA SEGUNDA COPIA DE RESPALDO

Una primera acción a realizar inmediatamente es trasladar la segunda copia maestra de ficheros ráster en HD a otra biblioteca del CDPV que tenga un depósito acondicionado para almacenar este tipo de soportes. De esta actividad se deriva un único coste, relacionado con la operativa de traslado entre dependencias de dos bibliotecas del propio CDPV.

### 5.1.2. ESTRATEGIA 2: ELABORACIÓN DE ESPECIFICACIONES TÉCNICAS PARA LA CONTRATACIÓN DE SERVICIOS DE DIGITALIZACIÓN

Sería altamente recomendable que el CDPV elaborase un pliego de requisitos técnicos detallados para la contratación de los servicios, a modo de *Request for Proposal* (RFP). Un modelo estandarizado podría servir para las diferentes campañas anuales. Jesús Robledano (2012) establece un cuadro de requisitos amplio que abarca la recogida de información relativa a aspectos como:

1. Objeto de los trabajos.
2. Las especificaciones técnicas muy detalladas exigidas en los parámetros de digitalización, proceso y almacenamiento digital.
3. Procedimientos de captura y procesamiento de imágenes postcaptura, incluyendo el sistema de calibración y caracterización de los equipamientos y el uso de los perfiles ICC.
4. El presupuesto por la realización de los trabajos.

5. Calendario de pagos.
6. Instalaciones donde se realizarán los trabajos.
7. Indicación y descripción de los trabajos y entregables incluidos y excluidos en el presupuesto.
8. Método de entrega de los materiales digitalizados.
9. Indicación de si los precios son firmes para toda la duración del proyecto.
10. Gasto promedio por cada imagen y el gasto total.
11. Plazos para la ejecución de las diferentes fases del proyecto.
12. Periodicidad en la entrega de los materiales digitalizados.
13. Supuestos para la cancelación del contrato. Responsabilidades de cada parte por cada supuesto.
14. Propiedad intelectual de los entregables.
15. Datos de la persona responsable del proyecto en la empresa y del responsable técnico de los trabajos.
16. Sistema que permita el seguimiento del proyecto por parte de la institución y su programación en el tiempo.
17. Garantías, responsabilidades y procedimientos de actuación por cada parte en el caso de deficiencias en los trabajos entregados.
18. Aplicación de un sistema propio de control de calidad de los trabajos y equipos de captura.
19. Empleo de una metodología de captura correcta y adecuada para los objetivos del proyecto.
20. Empleo de herramientas de captura, hardware y software, que permitan la obtención de los parámetros de calidad exigidos por las especificaciones técnicas del proyecto.
21. Soporte técnico apropiado para los equipamientos y software a emplear en la captura o almacenamiento digital. Se deberá garantizar la no interrupción por un período razonable de los trabajos debido a problemas técnicas con las herramientas de captura.
22. Experiencia demostrable previa en proyectos similares con resultados óptimos. Referencias de otros proyectos y personas y datos de contacto.
23. La empresa cuenta con personal suficientemente cualificado para la manipulación de documentos históricos con diversos estados de conservación y para el manejo de las herramientas de captura. Los operadores de escáner y, en su caso, otros profesionales de la empresa que tengan que manipular los documentos van, a recibir la formación adecuada en conservación que evite manipulaciones inadecuadas o una exposición a factores de riesgo para la conservación de los documentos.
24. La capacidad logística, económica y técnica para hacer frente a los requisitos y exigencia de los trabajos contratados durante el período de duración estimado.
25. La empresa ha realizado una prueba de digitalización exitosa de una muestra de los materiales de acuerdo con las especificaciones técnicas del proyecto.
26. El responsable de conservación y restauración de los documentos del archivo ha supervisado que tanto los equipos de captura que van a ser empleados por la empresa como las pautas de manipulación o de adaptación de los documentos a los equipos de captura no son perjudiciales para la conservación de los documentos.
27. Información sobre las medidas de prevención de los riesgos laborales que puedan implicar los trabajos de digitalización.

En el caso del CDPV, las especificaciones a los proveedores deberían incluir, concretamente: la parametrización adecuada de los escáneres o cámaras de captura de imágenes a fin de enriquecer al máximo posible los ficheros EXIF resultantes de las tomas; el detalle de cámaras o máquinas para realizar las digitalizaciones (modelos, marcas, software asociado, etc.); la inclusión de perfiles de color virtuales o de máquina tipo sRGB; indicación expresa de si ha habido procesamiento de las imágenes y con qué herramienta de software; y finalmente, la entrega de los archivos de imagen cruda RAW.

Estos tres aspectos permitirían atajar los problemas descritos en el apartado 3.3, epígrafes iii, iv, v y vii, que tal y como se ha explicitado en la Tabla 10.- Categorización de los problemas de la colección digital del Consorcio de Digitalización de Prensa Vasca” son problemas graves e incluso muy graves que pueden ser atajados en próximas campañas de digitalización a coste bajo o muy bajo.

Existe una herramienta interesante para ayudar a redactar un pliego de condiciones de especificaciones técnicas para la subcontratación de la digitalización (*Request for proposal* RFP) <http://www.updig.org/disg/generator.php>

También sería interesante el instar a las empresas a entregar por escrito un informe detallado de los procesos ejecutados en cada contratación anual.

Es importante resaltar que con relación a la grabación de datos en disco o cinta debe estarse atento al uso de sistema NTFS (admite tamaños de archivo de hasta unos 16 TB) o exFAT para discos USB (admite tamaño de archivo de 512 TB), y evitar el uso de FAT 32 (File Allocation Table, 1996) puesto que tiene un límite de tamaño de archivo de 4 GB.

### 5.1.3. ESTRATEGIA 3: REALIZACIÓN DE COPIAS DE RESPALDO DE LOS CD-ROM EXISTENTES

Una parte importante de la colección del CDPV, en concreto toda aquella contenida en CD-ROM, carece de copia de respaldo o *backup*, tal y como se ha señalado en el apartado 3.3, epígrafe xi. Parece evidente que su generación puede resultar costosa por el hecho de tener que realizarla desde una gran cantidad de CD-ROM que habría que manipular, 1.856 CD-ROM en concreto, con el consiguiente tiempo en horas hombre y máquina. Sin embargo vendría a resolver un problema grave.

La operativa podría ir realizándose por ciclos o períodos temporales. En las organizaciones los períodos de respaldo aparecen indicados habitualmente en el Plan de Contingencia de Seguridad Informática (formado por tres sub-planes: plan de respaldo, plan de emergencia y plan de recuperación).

### 5.1.4. ESTRATEGIA 4: MIGRACIÓN DE LAS IMÁGENES MASTER EN FORMATO JPEG

Existe un grave problema con los ficheros que únicamente están en formato JPEG. Se trata de las imágenes pertenecientes a la conversión retrospectiva del diario *Egin*, detalladas en el apartado 3.3, epígrafe xi. En principio, parece que no se trata de un gran número de ficheros y

aunque el problema es extremadamente grave, con un coste económico medio podría resolverse el asunto mediante una migración o cambio de formato.

En relación con el resto de la colección, tomemos el caso que expone Wilson (2011) para la colección “*JISC1 Newspapers*” de la *British Library* (BL), poniendo el foco en la copia maestra en formato TIFF que posee y que alcanza los 80 TB. Aunque se trata de una colección de diarios del siglo XIX, con OCR y alojada en un repositorio de acceso público, las indicaciones que vamos a traer a colación son extrapolables para el caso que nos ocupa.

Teniendo en cuenta que, en relación con el tamaño de los ficheros, más pequeño significa menores costes de preservación y, habida cuenta de que el formato TIFF es escasamente comprimible, el proyecto de la BL opta por el formato JPEG2000. Éste permite mayores ratios de compresión sin pérdida y ello redundaría en un ahorro económico en costes de almacenamiento. En la BL se rechaza la opción de “no hacer nada” dejando la colección en TIFF tal cual está, dado que tales ficheros son grandes y presentan ratios pobres de compresión. Se evita también la conversión a JPEG, puesto que aunque son ficheros de tamaño pequeño tienen compresión con pérdida (y problemas de conservación, como ya ha quedado explicado). Finalmente se rechaza la conversión tanto a PNG como a BMP, puesto que ambos formatos son considerados formatos de archivo pobres debido a su escasa capacidad de compresión y a su limitado espacio de color.

Partiendo del caso planteado, al CDPV se le plantean varias posibilidades. Por una parte, cabría mantener el fondo en formato TIFF, tal cual está, pese a saber que su elevada tasa de ocupación supondrá un coste elevado en gestión de la preservación digital y almacenamiento en los años venideros.

Otra opción es convertir todo el fondo a formato JPEG2000, para reducir costes de almacenamiento. Sin duda hay un coste en conversión de la colección que habrá que ponderar para saber si es mayor o menor que el coste de preservación de los TIFF. Como ya se ha dicho, el formato JPEG2000 presenta la desventaja de no contar con un gran desarrollo por parte de la industria, aunque las grandes organizaciones de la memoria están apostando por este formato lógico.

La tercera de las opciones consistiría en dejar la colección ya generada en TIFF y producir en JPEG2000 los nuevos *outputs*; aunque pudiera parecer que esta opción carece de todo fundamento de cara a establecer una política de preservación digital coherente y homogénea, lo cierto es que lo habitual es encontrarse frente a colecciones de composición mixta. El formato TIFF no presenta problemas de obsolescencia remarcables (es estándar, está soportado por la industria, es comúnmente utilizado...) y presenta ventajas como su amplia difusión o que permite compresión sin pérdida, por todo ello su permanencia está garantizada por muchos años. Pero si el fondo sigue creciendo como es previsible, no estará de más pensar en el volumen global y los costes de gestión derivados de ello.

El problema planteado apartado 3.3, epígrafe ii, relativo a los ficheros máster de los cuales el consorcio no dispone de copia derivada es solventable por dos vías. O bien se considera la actual copia máster, fruto del refresco detallado en el apartado 3.2.5.4, como copia de explotación, sin incurrir en costes mayores. O bien se genera una copia derivada en formato PDF, cuya realización conllevaría coste asociado definible como medio.

En cuanto a formatos, finalmente y en relación con los formatos de difusión, derivados o de explotación, estos han de ir más allá de los orientados al uso en PCs y ordenadores portátiles, como el formato PDF. Se debe intentar eliminar las barreras de accesibilidad que este formato presenta en dispositivos con pantallas de pequeño tamaño como móviles, *smartphones*, tabletas, *e-reader* (dispositivos de lectura electrónica). El problema fundamental es que se trata de un formato no repaginable. Los contenidos se han de difundir también en formatos abiertos como ePub, que permitan una lectura cómoda en este tipo de dispositivos que presentan gran implantación entre los lectores. Existe un formato adecuado para los derivados, está orientado a la preservación digital a largo plazo: PDF/A 1.7.

#### 5.1.5. *ESTRATEGIA 5: DESCRIPCIÓN CON METADATOS DE LAS IMÁGENES DE FUTURAS DIGITALIZACIONES*

Los ficheros de imagen ráster de las digitalizaciones venideras deberían contener en todo caso y siempre una adecuada integración de los metadatos EXIF de toma fotográfica como ha quedado establecido en el apartado 5.1.2.

Sería planteable la posibilidad de describir mediante diversos esquemas de metadatos las colecciones de diarios digitalizados, puesto que aunque no se volcasen a un repositorio de datos abiertos tales descripciones podrían servir para mejorar su identificación en un entorno de servidor.

Es cierto que existen ya unas descripciones mínimas a nivel de cabecera que están en formato MARC, MARC 21 y MARC XML en los distintos catálogos bibliográficos de las bibliotecas integrantes del CDPV. Pero las imágenes ráster están fuera de estos catálogos y su adecuada identificación y trazabilidad hace recomendable dotarlos de cuanta más metainformación mejor.

También es cierto que realizar descripciones METS solo tiene sentido en un entorno de servidor y de red, que hacerlo en DC sólo tiene sentido en un contexto de necesidad de recuperación mediante metadatos descriptivos, que hacerlo en ALTO sólo es necesario en relación con colecciones con OCR, y que hacerlo con PREMIS sólo tiene sentido en un contexto de repositorio de preservación. Aunque “la no existencia de un repositorio de preservación normalizado no exime de la necesidad de utilizar metadatos de preservación y en concreto PREMIS.” (Ferrerías, 2010, p. 38), pero esta autora habla en un contexto de repositorio abierto.

Si el Consorcio consigue que las imágenes de las campañas venideras contengan incrustados metadatos EXIF y vayan acompañados de ficheros METS con información estructural, de derechos y que integre MARC XML, los nuevos ficheros ráster de la colección se encontrarían en disposición de poder afrontar el futuro con garantías de perdurabilidad. Todo ello no tiene por qué suponer un coste elevado, puesto que los proveedores tendrían que mejorar algún proceso (captura) y añadir algún otro proceso más (generación de ficheros METS).

#### 5.1.6. ESTRATEGIA 6: PLANIFICACIÓN DE CICLOS DE REFRESCO DE LOS SOPORTES (CD-ROM Y HD) DE LA COLECCIÓN

El hecho de que la colección esté soportado, más o menos a partes iguales, en CD-ROM y en discos duros externos, y habida cuenta de las recomendaciones de los expertos y de los fabricantes sobre durabilidad de estos soportes, es recomendable acometer procesos de refresco. Hay una parte de la colección en CD-ROM de buena calidad. Además, el hecho de encontrarse albergados en una almacén adecuadamente controlado y estabilizado en lo relativo a temperatura y humedad, hace pensar que pueden durar aún varios años. Sin embargo, también existen CD-ROM de peor calidad. Y por otra parte, el CDPV tiene gran cantidad de discos duros externos, exactamente 83. Estos aparecen referenciados en toda la literatura y especificaciones de fabricantes con una esperanza de vida de alrededor de cinco años. En cualquier caso, parece que la vida útil que dan los fabricantes se refiere a dispositivos “en uso”, suponiendo que si están guardados y no conectados a toma eléctrica su expectativa de vida será mayor.

Lo recomendable, por tanto, es encarar el inevitable proceso de refresco por ciclos y tipos de soporte, de forma y manera que con diversas partida presupuestarias de años distintos se pueda ir traspasando la colección a nuevos medios de almacenamiento.

Siempre que se realicen copias se debe realizar un proceso de suma de verificación, suma de chequeo o *checksum* que garantice la posterior autenticidad e integridad de la copia realizada. Se trata de un proceso técnico que detecta cambios accidentales en la secuencia de datos, verificando que no haya discrepancia entre los valores iniciales y finales en un proceso de grabación.

Todas las acciones llevadas a cabo deben quedar reflejadas. Se deben documentar tanto por escrito, como dejar constancia de ello en aquellos esquemas de metadatos que permitan introducir información sobre tales acciones.

#### 5.1.7. ESTRATEGIA 7: DEFINICIÓN DE LA INFRAESTRUCTURA DE ALMACENAMIENTO A LARGO PLAZO

La estrategia de mayor calado es la que tiene que ver con la adopción de decisiones sobre qué tipo de medio de almacenamiento, sistema o infraestructura, escoger para encarar el reto de la preservación digital en los años venideros. Las alternativas que se han expuesto en el apartado 4.1 han de ser evaluadas conforme a lo que el informe *Best practices for a digital storage infrastructure for the long-term preservation of digital files* (Henriksen et al., 2013) define como factores clave:

- El tamaño de datos de la colección.
- La capacidad presupuestaria de la organización.
- La plantilla de personal con capacitación tecnológica.
- El uso que se vaya a hacer de los datos.

En relación con estas cuatro variables se apuesta por un medio u otro. Conforme al cuadrante visible en la Tabla 13 que informe plantea, el Consorcio de Digitalización de Prensa Vasca habría de decantarse por las siguientes soluciones:

- a) En relación con el tamaño de la colección, dado que el fondo supera ya los 30 TB, se recomiendan unidades de disco duro HDD, cintas LTO o discos de estado sólido SSD. Todas estas soluciones son planteables en sus diversas configuraciones posibles.
- b) En relación con la disponibilidad presupuestaria, entendiéndose que ésta es baja, se recomiendan unidades de disco duro HDD, cintas LTO o *cloud computing*.
- c) En relación con el personal disponible capacitado tecnológicamente, entendiéndose que este es cero: *cloud computing* o *backup* remoto realizado por un subcontratista en unidades de disco duro HDD en sus posibles configuraciones.
- d) En relación con el uso de los datos, como se trata de un archivo de preservación, se recomiendan cintas LTO o unidades de disco duro HDD, una vez más en sus diversas configuraciones posibles.

Tabla 13.- Medios de almacenamiento adecuados en función de cuatro variables clave (Fuente: Henriksen et al., 2013)

Size of data collection	MB	GB	TB
Recommended storage	HDD, LTO tape, cloud	HDD, LTO tape, SSD	HDD, LTO tape, SSD
Size of budget	Small	Medium	Large
Recommended storage	HDD, LTO tape, cloud	HDD, LTO tape	HDD, LTO tape, SSD
Internal IT staff	0	1-5	5<
Recommended storage	Cloud and remote backup on HDD by external vendor	LTO tape or HDD server systems	LTO tape, HDD and SSD server systems
Use of data	Daily access	Backup (refreshed)	Preservation archive
Recommended storage	HDD, SDD, cloud	LTO tape, HDD, SDD, cloud	LTO tape, HDD

A la vista de tales recomendaciones, no parece desaventurado el apostar por una configuración de datos accesibles en servidor *online* o *nearline* mediante discos duros o cintas magnéticas.

## 5.2. ESTRATEGIAS DE PRESERVACIÓN DESCARTABLES

Hay otra serie de acciones que por diversas razones, principalmente económicas y legales, no son estrategias ejecutables por el Consorcio en el corto y medio plazo. Se trata de acciones como volver a digitalizar parte de la colección debido a la baja resolución de captura para



poder pasar OCR, volver a digitalizar imágenes defectuosas, colocar en acceso abierto el fondo, levantar un repositorio fiable de preservación o contratar servicios de *cloud computing* para el almacenamiento de la colección

#### 5.2.1. *ESTRATEGIA 8: RE-ESCANEO DE DIARIOS CON DEFECTOS DE CAPTURA Y/O BAJA RESOLUCIÓN*

Habida cuenta de los costes que la digitalización conlleva, es totalmente inviable plantearse nuevas digitalizaciones de aquellas partes de la colección que presentan o bien baja resolución de captura, impidiendo así el procesado con OCR, o bien notables defectos de captura, derivando en aparición de manchas negras, textos ilegibles, imágenes no visibles, textos no alineados y otros defectos de calidad graves.

#### 5.2.2. *ESTRATEGIA 9: DISEÑO E IMPLEMENTACIÓN DE UN REPOSITORIO DIGITAL DE CONFIANZA*

Los repositorios digitales están inicialmente pensados para proveer acceso abierto a los recursos de información. Aquellos cuya arquitectura y diseño se ha levantado conforme al modelo teórico OAIS están pensados además para poder gestionar adecuadamente la preservación digital de los recursos que albergan. Pero solamente los repositorios digitales fiables están en condiciones de poder decir que son efectivamente capaces de hacerlo.

Varias bibliotecas participantes en el CDPV cuentan con repositorios digitales: Liburuklik – Biblioteca Digital Vasca, del Departamento de Cultura del Gobierno Vasco (Pulgar Vernalte, 2013); EMD – Memoria Digital Vasca, de la Fundación Sancho el Sabio (Gómez et al., 2012); ADDI, de la UPV/EHU; BLD – Bilboko Liburutegi Digitala, de las Bibliotecas Municipales de Bilbao; META, Repositorio Digital de la Diputación Foral de Gipuzkoa.

Pero en el ámbito vasco no hay ningún repositorio de confianza “*Trusted Digital Repository*” validado conforme a la norma ISO 16363:2012 *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*, para el cumplimiento de la preservación digital.

Por otra parte, se ha de tener en cuenta que los primeros quince años de la colección digital del CDPV corresponden a imágenes ráster con una resolución de captura de 200 ppp, por lo cual la opción de pasar un software de reconocimiento óptico de caracteres sobre estos ficheros de imagen muy posiblemente daría muy pobres resultados con la actual tecnología. Siendo la opción del OCR inviable para el 80% de la colección ¿qué sentido puede tener volcar unos fondos sobre los cuales no se pueden hacer búsquedas a un repositorio digital? La recuperación de información se limitaría a los campos de título y fechas, y tal vez geográficos.

Finalmente, cabe mencionar que el CDPV cuenta con un proyecto presupuestado para volcar todo su fondo al repositorio Liburuklik, ascendiendo el coste ofertado a 480.000 € + IVA, para una estimación de 7.000.000 de imágenes tras pasar OCR sin corrección y describir con metadatos el fondo.

### 5.2.3. *ESTRATEGIA 10: CONTRATACIÓN DE UN SERVICIO DE CLOUD DE ALMACENAMIENTO (dSaaS)*

El cloud es considerado como una “nebulosa jurídica” (Alberch, 2012, p. 12), que además supone afrontar un elevado precio (Henriksen et al., 2013).

La contratación de un servicio de externalización de almacenamiento del tipo cloud dSaaS no es una opción viable en el caso de la colección del CDPV por dos razones de peso:

- La carencia de autorización y permisos de los titulares de los derechos de los diarios “impide” poner la copia en manos de terceros, o aun habiendo cierto vacío legal al respecto, la prudencia aconseja evitar hacerlo.
- Su elevado precio, pues se trata de soluciones con un gran coste económico.

Además, el almacenamiento seguro de los ficheros no implica superar la barrera de la obsolescencia de los formatos. Aunque bien es cierto que se ha dejado sentado que los formatos TIFF (mayoritarios en el fondo del CDPV) no presentan problemas de obsolescencia a medio plazo.

## 5.3. *DECISIONES Y DEBATES A AFRONTAR*

El Consorcio tiene planteados ante sí una serie de retos que no podrá abordar sin debatir en su seno ciertos aspectos clave, como son la identificación clara de responsabilidades, la asunción de un mandato expreso, la financiación a medio y largo plazo, o la gestión de los derechos de propiedad intelectual.

Se ha de decidir si la digitalización que realiza el CDPV está hecha desde una perspectiva reprográfica o desde una perspectiva patrimonial, esto se deberá plantear abiertamente si prima el mero valor informativo de los diarios o se quiere mantener también su aspecto y valor icónico a modo facsimilar.

### 5.3.1. *PROMULGACIÓN DE MANDATOS Y DEFINICIÓN DE RESPONSABILIDADES*

Los actuales miembros del Consorcio están asumiendo la conservación de los diarios conforme a la misión y objetivos propios de cada una de sus organizaciones, como ya ha quedado detallado en el apartado 3.1.2, pero en la mayoría de los casos no existe un mandato legal que obligue a que esto se haga. Es por ello que las partes deben identificar a quién corresponde preservar la prensa diaria de Euskal Herria. En caso de no identificar a ningún organismo que tenga la encomienda legal y quieran dotarse de la responsabilidad de hacerlo, han de establecer un acuerdo expresamente mandatorio de cooperación entre las partes firmantes.

La falta de mandato expreso podría dar lugar a una situación delicada, puesto que en tiempos de baja disponibilidad económica, como actualmente, algunas bibliotecas podrían verse obligadas a abandonar el proyecto, haciéndolo inviable para el resto de las partes. Si fuesen varias las bibliotecas que abandonasen el CDPV, y el proyecto cooperativo se hiciese

económicamente inviable, la preservación de los diarios tendría que afrontarse en solitario por quien esté obligado a hacerlo por mandato legal.

Los miembros del Consorcio, por otra parte, deberán plantearse a la luz de las anunciadas reformas de la Ley de Bases del Régimen Local y de la Ley de Territorios Históricos si el modelo consorciado es válido para sus fines o si existen alternativas jurídicas al mismo. A todas luces parece arriesgado operar sin personalidad jurídica siendo poseedor de un notable patrimonio, habiendo incurrido en cuantiosas inversiones y gastos, y teniendo que afrontar retos de gestión patrimonial como el de los activos digitales.

### *5.3.2. ASEGURAR LA SOSTENIBILIDAD ECONÓMICA DEL PROYECTO*

Hoy día hay varios grupos de investigación que están estudiando el tema de los costes de la preservación digital mediante el diseño y evaluación distintos modelos de análisis de costes. Una idea parece clara, la preservación digital supone un coste y, por muy obvio que parezca, es evidente que las organizaciones que quieran preservar sus recursos digitales van a tener que comenzar a presupuestar partidas destinadas a tal fin.

En nuestro contexto una pregunta queda en el aire, ¿cómo se garantiza la financiación necesaria para hacer preservación digital a largo plazo en Euskadi?

### *5.3.3. ABORDAR LA PROBLEMÁTICA DE GESTIÓN DE DERECHOS DE LAS OBRAS DIGITALIZADAS*

En relación con la problemática legal derivada de la difusión de las colecciones de diarios se pueden tomar decisiones en un sentido o en otro. Se puede permanecer como hasta ahora, evitando difundir las colecciones en abierto, alternativa conservadora pero que impide a las bibliotecas explotar los activos digitales de los cuales son poseedoras y a los usuarios hacer uso de los mismos. En sentido contrario, se pueden buscar acuerdos con los titulares de los derechos de propiedad intelectual que permitan poner a disposición del público las colecciones digitalizadas por el Consorcio de Digitalización de Prensa Vasca. La búsqueda de una gestión activa de los derechos de propiedad intelectual es una opción. Habría que buscar acuerdos con los editores de los diarios para poder contrastar si son alcanzables o no.

## *5.4. PLAN DE PRESERVACIÓN DIGITAL*

Parece adecuado pensar en la necesidad de dotarse de un plan de preservación digital que a modo de programa continuado, conforme a una planificación temporal, establezca por escrito un mandato y recoja un marco de políticas claro. Tales políticas deben identificar responsabilidades administrativas, garantizar la viabilidad organizativa, y dotarse, conforme a tecnología adecuada, de un sistema seguro. Para todo ello deberá contarse con equipo humano cualificado.

El propio Consorcio de Digitalización de Prensa Vasca ha detectado en los últimos años la necesidad de emprender tal plan de acción. En el acta 02/2011, de 28/09/2011, se menciona por primera vez la necesidad de dotarse de un plan de preservación digital. Poco después

“Jesús Zubiaga comenta que los discos duros en el formato TIFF ubicados en Sancho el Sabio están empezando a ocupar un espacio excesivo. (...) tendremos que pensar en un protocolo de preservación” (acta 03/2011, de 30/11/2011).

Los *backups* son suficientes para las copias derivadas de cada una de las bibliotecas, pero no lo son para los recursos de alta calidad (versión maestra) a los cuales se les debe aplicar una política de preservación digital. Ortiz Ancona describe con claridad las diferencias entre preservación y copias de seguridad: “La preservación digital es diferente de las copias de seguridad. Lo que se guarda como copia de seguridad en una biblioteca digital son, básicamente, dos cosas: por un lado la información publicada en el servidor (recursos digitales más información de catálogo) y, por otro lado, los recursos digitales en proceso de edición. La preservación digital sin embargo, no se ocupa de respaldar ni los datos del servidor ni el material de trabajo diario, sino de salvaguardar los recursos digitales que necesitaremos en el futuro.” (2012, p. 2).

Tal y como ha quedado expuesto en el apartado 2.5 toda planificación de la preservación digital ha de contemplar las políticas de preservación, las obligaciones legales, las limitaciones de la organización, las limitaciones técnicas, las necesidades de los usuarios, y los objetivos de la preservación. Y la variable de costes deberá ir acorde a los riesgos reales a que el fondo está expuesto.

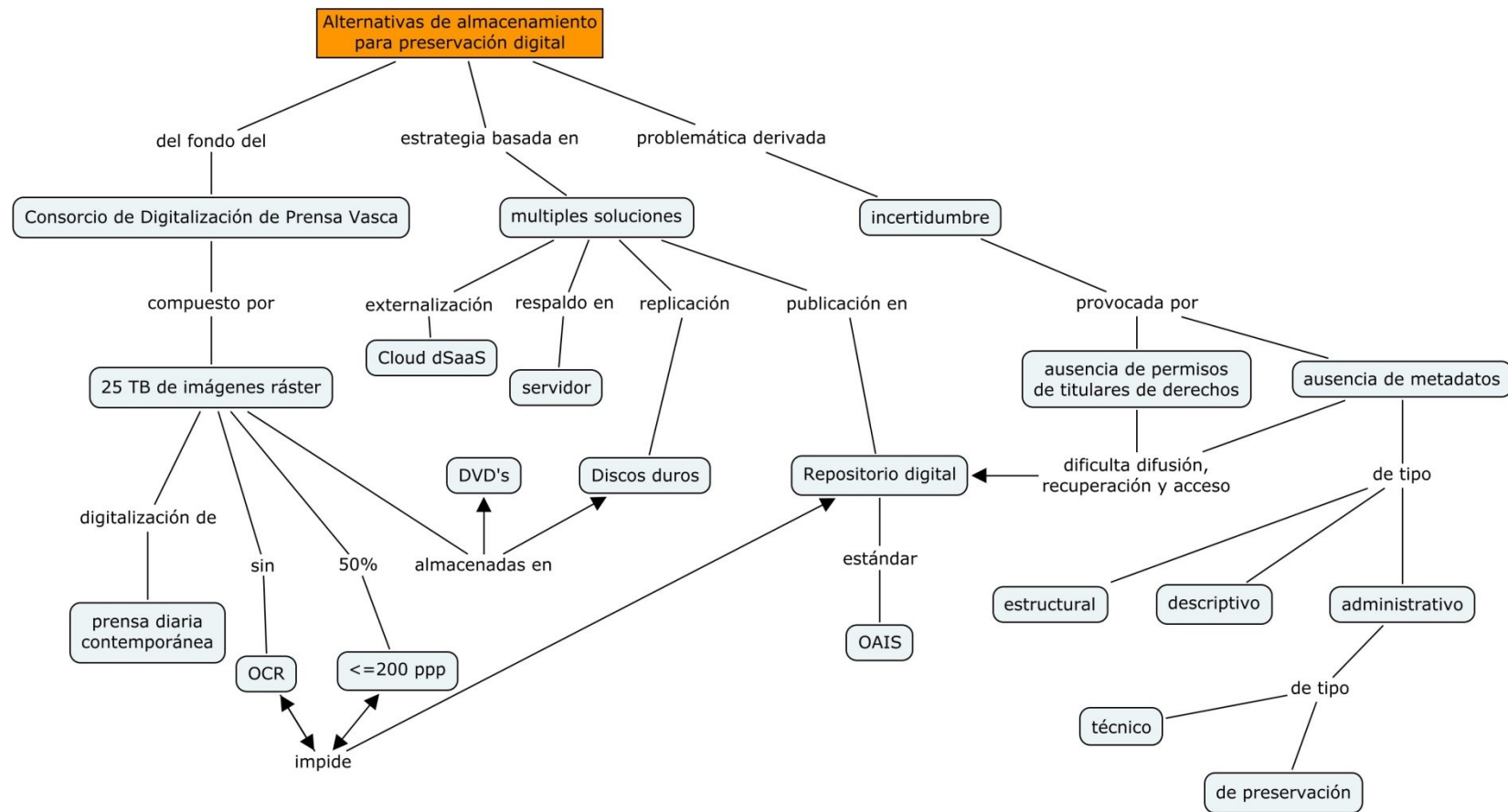


Figura 16.- Mapa conceptual de las alternativas de almacenamiento para la preservación digital del fondo del Consorcio de Digitalización de Prensa Vasca (Elaboración propia)

## 6. CONCLUSIONES

---

Los proyectos de digitalización medianos que se vienen desarrollando desde hace varios años, como el del CDPV, han de estar atentos a la cuestión de la preservación digital de sus activos digitales. El presente trabajo de investigación se ha abordado teniendo este reto presente, con la finalidad de proponer alternativas técnicas viables económicamente para la preservación a largo plazo de la colección de ficheros máster de los periódicos digitalizados por el Consorcio de Digitalización de Prensa Vasca.

Se ha podido analizar de una forma bastante detallada y pormenorizada el estado de la colección. Se ha podido asimismo realizar un acercamiento, por diversas vías, a la cuestión de los costes de digitalización (costes sobre lo presupuestado, costes reales y costes estimados) y a la cuestión de los costes de la preservación digital a largo plazo (mediante la presentación comparada de diferentes alternativas con indicación de costes de implementación aproximados).

Ha quedado establecido que no es una estrategia suficiente el contar únicamente con los datos almacenados en los soportes actuales, considerándose necesario tanto repartir copias por distintos lugares geográficos como realizar nuevos *backups*, y nuevos refrescos de los soportes.

Se estima que el *cloud computing* de almacenamiento dSaaS no es una opción madura, tanto por las incertidumbres derivadas de las condiciones legales de los contratos, como de la legislación española sobre cesión de datos e información sujeta a derechos de explotación conforme a la ley de propiedad intelectual a terceros. Las alternativas de mercado actuales son además económicamente muy gravosas para las organizaciones y exigirían presupuestar unas cantidades económicas importantes para los ejercicios venideros.

Ha habido dificultad en dibujar un cuadro general de costes asociados a la preservación digital, aunque, de la mano de diversos autores, se ha podido hacer un acercamiento inicial al tema. La idea principal es que no hay que perder de vista, tal y como plantean Muñoz de Solano y Wilson con gran sentido común, que se ha de hacer un balance ponderado entre los riesgos que el fondo realmente corre y los gastos en que incurrir para hacer preservación digital. De tal manera que no parece recomendable acudir a cuantiosas inversiones si el fondo en cuestión no está sujeto a grandes amenazas. Habrá por tanto, que realizar siempre un detallado estudio del fondo, como se ha hecho en el caso del Consorcio de Digitalización de Prensa Vasca, para poder plantear medidas acordes al riesgo real una vez que se dispone del diagnóstico preciso.

En relación con los esquemas de metadatos y la descripción de los ficheros de imagen ráster, es importante señalar que siempre y en todo caso los metadatos técnicos EXIF de toma fotográfica son claves y concluyentes, puesto que a través de ellos se tiene tanto información precisa sobre las características de la imagen digital, como información de trazabilidad o vida y ocurrencias de los ficheros. Esta última información de trazabilidad es muy importante de cara a la preservación digital y esquemas de metadatos como PREMIS hacen hincapié en recolectar toda la información técnica sobre la vida del objeto digital y todos los procesos técnicos a que ha estado sometido.

Es evidente también que los esquemas de metadatos descriptivos tienen mayor valor en contextos de datos e informaciones divulgados en red o en grandes servidores, donde la recuperación de la información es más complicada y por tanto donde mayor valor toman las buenas descripciones. La carencia de metadatos descriptivos en un fondo offline de diarios almacenados en soportes debidamente identificados es menos traumática, puesto que la recuperación de los datos es factible siempre por título y fecha, variables casi siempre suficientes en este tipo de documentos.

Se han dejado planteadas varias cuestiones no resueltas. Son decisiones no alcanzables entre las metas del presente estudio, pero cuya identificación posibilitará tenerlos presentes para abordarlos cuando corresponda. Son preguntas abiertas y sin respuesta clara de momento sobre aspectos relativos a responsabilidades de conservación del patrimonio bibliográfico en el ámbito vasco. Qué se hace, cómo y porqué. ¿Es el CDPV quien debe hacer lo que está haciendo en relación con la conservación del patrimonio bibliográfico? ¿Dónde están las responsabilidades de financiación? ¿Qué hay de la difusión y uso de documentos digitales sujetos a derechos de propiedad intelectual en el marco de las bibliotecas? ¿Es realmente inviable una estructura jurídica de consorcio con capacidad de obrar en el ámbito bibliotecario vasco?

Sería interesante que posteriores estudios reparasen en el análisis de estas cuestiones. Por nuestra parte, esperamos haber contribuido al proceso de reflexión y estudio sobre la preservación digital, con la esperanza de que los datos extraídos del caso del Consorcio de Digitalización de Prensa Vasca permitan a ésta y a otras organizaciones afrontar los retos existentes con mayor nivel de acierto y éxito.

## BIBLIOGRAFÍA

---

- Abadal, E. (2001). *Sistemas y servicios de información digital* (p. 152). Gijón: Trea.
- Agirreazaldegí, T., Ronco, M. M., & Camacho, I. (2009). Gida: euskal Herriko komunikabideetako dokumentazio zerbitzuak. In T. Agirreazaldegí (Ed.), *I Jornadas sobre Documentación y Medios de Comunicación, 2008* (pp. 273–326). Bilbao: Euskal Herriko Unibertsitatea = Universidad del País Vasco.
- AIMC. (2011). *La Prensa: Digital vs Papel - AIMC* (p. 30). Retrieved from <http://www.aimc.es/-La-Prensa-Digital-vs-Papel-.html>
- Alberch, R. (2001). Memoria, digitalización y mecenazgo. *El Profesional de la Información*, 10(9), 34–37.
- Alberch, R. (2012). Organizando el rompecabezas: la preservación digital como reto ineludible. In *XVIII Jornadas de la Conferencia de Archiveros de las Universidades Españolas (CAU/CRUE)*. CAU/CRUE. Retrieved from [http://cau.crue.org/export/sites/Cau/Quehacemos/jornadas/Cadiz12/Ramon\\_Alberch.pdf](http://cau.crue.org/export/sites/Cau/Quehacemos/jornadas/Cadiz12/Ramon_Alberch.pdf)
- Araiz, S. (2012). Koldo Mitxelena: liburutegiaren esparruak eta zerbitzuak zabaltzen. In T. Agirreazaldegí (Ed.), *Kazetaritza eta dokumentazioa, informazioaren sarbide irekiaren eta sare sozialen bidegurutzean = Periodismo y documentación en la encrucijada del acceso abierto a la información y las redes sociales* (pp. 76–83). Bilbao: Euskal Herriko Unibertsitatea = Universidad del País Vasco.
- Arquero Avilés, R., & García-Ochoa Roldán, M. L. (2005). *La hemeroteca de prensa* (p. 201). Gijón: Trea.
- Arrizabalaga, J. (2009). Hemeroteca Plus: publikazio idatzien ustiaketa digitala. In T. Agirreazaldegí (Ed.), *I Jornadas sobre Documentación y Medios de Comunicación, 2008* (pp. 267–269). Bilbao: Euskal Herriko Unibertsitatea = Universidad del País Vasco.
- Ballart Hernández, J. (2007). *Manual de museos* (p. 245). Madrid: Síntesis.
- Beagrie, N., & Jones, M. (2008). *Preservation management of digital materials: the handbook* (p. 160). Digital Preservation Coalition. Retrieved from <http://www.dpconline.org/publications/digital-preservation-handbook>
- Bilbao Ariño, C. (2005). Patrimonio bibliográfico de la Biblioteca Koldo Mitxelena, de la Diputación Foral de Gipuzkoa. In *I Seminario sobre Patrimonio Bibliográfico Vasco, 2003* (pp. 273–281). Vitoria-Gasteiz: Fundación Sanch el Sabio; Gobierno Vasco, Departamento de Cultura.



- BNE. (2012). Plan estratégico 2012-2014 de la Biblioteca Nacional de España. Madrid.  
Retrieved from <http://www.bne.es/es/LaBNE/PlanEstrategico/docs/PlanEstrategico12-14.pdf>
- British Library. (2013). *British Library Digital Preservation Strategy* (p. 19).
- Carnicer Arribas, D., Solana Hernández, C., Ruano Cerezo, J. J., San José Negro, I., Vázquez Mateo, C., López Morales, M., ... Rodríguez Alfageme, L. M. (2011). Recomendaciones para la digitalización de documentos en los archivos. Valladolid: Junta de Castilla y León.  
Retrieved from <http://www.jcyl.es/web/jcyl/CulturaPatrimonio/es/Plantilla100DetalleFeed/1246988230374/Noticia/1284166105811/Comunicacion>
- Castillo, J. M., & Jorba, F. (2008). Almacenamiento distribuido y preservación digital: una panorámica de alternativas. *BiD: textos universitaris de biblioteconomia i documentació, junio(20)*. Retrieved from <http://www.ub.edu/bid/20casti2.htm#similars>
- Conway, P. (2010). Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas. *The Library Quarterly, 80(1)*, 61–79. doi:10.1086/648463
- Cornell University Library. (2000). Llevando la teoría a la práctica: tutorial de digitalización de imágenes. Retrieved May 28, 2013, from <http://www.library.cornell.edu/preservation/tutorial-spanish/contents.html>
- Criado, M. A. (2013). La memoria digital de España no tiene quien la reglamente. *El Confidencial.com*. Retrieved June 04, 2013, from <http://www.elconfidencial.com/tecnologia/2013/05/13/la-memoria-digital-de-espana-no-tiene-quien-la-reglamente-4876/>
- Cruz Mundet, J. R. (2011). *Diccionario de archivística*. Madrid: Alianza editorial.
- Cuesta, A. (2005). Plan estratégico sobre la conservación y preservación del patrimonio digital en la Comunidad Autónoma de Euskadi. *I Conferencia Internacional sobre políticas culturales*. Bilbao. Retrieved from [http://www.kultura.ejgv.euskadi.net/r46-6614/es/contenidos/informacion/conferencia\\_politica\\_culturale/es\\_10312/adjuntos/ponencia\\_alejandro\\_cuesta\\_cast.pdf](http://www.kultura.ejgv.euskadi.net/r46-6614/es/contenidos/informacion/conferencia_politica_culturale/es_10312/adjuntos/ponencia_alejandro_cuesta_cast.pdf)
- Cuesta Ezeiza, A. (2007). Perspectivas del patrimonio digital. In *I Conferencia Internacional sobre Políticas Culturales, 2005* (pp. 310–312). Vitoria-Gasteiz: Gobierno Vasco, Departamento de Cultura.
- Digital Media Group. (2007). CD and DVD Archiving: Quick Reference Guide for Care and Handling. Retrieved July 15, 2013, from <http://www.itl.nist.gov/iad/894.05/docs/discare.html>

- Digital preservation FAQs. (2013). The National Archives, Kew, Richmond, Surrey, TW9 4DU. Retrieved May 07, 2013, from <http://www.nationalarchives.gov.uk/information-management/projects-and-work/digitisation.htm>
- Dunning, A. (2012). *European newspaper survey report* (pp. 1–20). Retrieved from <http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D4.1-Europeana-newspapers-survey-report.pdf>
- Eizagirre, M., & Barandiaran, J. (2009). Berria egunkariaren dokumentazio zerbitzua: bi datu-basekazetariaren zerbitzura. In T. Agirreazaldeggi (Ed.), *I Jornadas sobre Documentación y Medios de Comunicación, 2008* (pp. 211–228). Bilbao: Euskal Herriko Unibertsitatea = Universidad del País Vasco.
- Elvira, M. (2011). Preservación de los documentos digitales: Guía para comenzar . ISO / TC 46 / SC 11 . Version española , octubre 2010. *Revista española de documentación científica*, 34(3 julio-septiembre), 461–477.
- España. Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia (1996). Spain. Retrieved from [http://www.boe.es/diario\\_boe/txt.php?id=BOE-A-1996-8930](http://www.boe.es/diario_boe/txt.php?id=BOE-A-1996-8930)
- Fernández Molina, J. C. (2010). Preservación digital y derechos de autor: ¿un conflicto sin solución? In *V Congreso Nacional de Bibliotecas Públicas. Gijón, 2010*. Gijón.
- Ferreras, T. (2010). *Preservación digital en repositorios institucionales: GREDOS [Trabajo Fin de Máster]*. Universidad de Salamanca, Salamanca. Retrieved from <http://gredos.usal.es/jspui/handle/10366/74739>
- FESABID. (2013). *Temas para una reforma legislativa de la Ley de Propiedad Intelectual. Propuestas 2013 de FESABID*. Retrieved from <http://www.fesabid.org/repositorio/grupo-bpi/reforma-legislativa-de-la-ley-de-propiedad-intelectual-propuestas-2013>
- Galloway, P. (2005). Preservation of digital objects. *Annual Review of Information Science and Technology*, 38(1), 549–590. doi:10.1002/aris.1440380112
- García Martínez, D. S. (2012). *Digitalización de archivos fotográficos: filosofía y metodología [Trabajo Fin de Máster]*. Universidad Carlos III de Madrid, Madrid. Retrieved from <http://eldocumentalistaudiovisual.files.wordpress.com/2013/03/digitalizacic3b3n-de-archivos-fotogrc3a1ficos-filosofc3ada-y-metodologc3ada-diego-santos-garcc3ada-martc3adnez.pdf>
- Gobierno Vasco. (2004). Plan Vasco de la Cultura. Retrieved June 04, 2013, from [http://www.euskara.euskadi.net/r59-738/es/contenidos/informacion/argitalpenak/es\\_6092/adjuntos/plan\\_vasco\\_cultura\\_c.pdf](http://www.euskara.euskadi.net/r59-738/es/contenidos/informacion/argitalpenak/es_6092/adjuntos/plan_vasco_cultura_c.pdf)

- Gobierno Vasco. (2010). Contrato Ciudadano por las Culturas. Vitoria-Gasteiz. Retrieved June 04, 2013, from [http://plancultura-kulturaplana.wikispaces.com/Presentación+\(cas\)](http://plancultura-kulturaplana.wikispaces.com/Presentación+(cas))
- Gómez, C. (1995). Fundacion Sancho el Sabio. *RIEV. Revista Internacional de Estudios Vascos*, *T. XL, N°1*(Año 43), 182–194.
- Gómez, C. (1997). La digitalización en la Fundación Sancho el Sabio. *Educación y Biblioteca*, *80*, 42–47. Retrieved from <http://gredos.usal.es/jspui/handle/10366/113396>
- Gómez, C. (2005). El Patrimonio Bibliográfico Vasco en la Fundación Sancho el Sabio: aplicación de nuevas tecnologías (1993-2003). In *I Seminario sobre Patrimonio Bibliográfico Vasco, 2003* (pp. 249–258). Vitoria-Gasteiz: Fundación Sanch el Sabio; Gobierno Vasco, Departamento de Cultura.
- Gómez, C., Martínez, C., & Zubiaga, J. (2012). Gestión de contenidos: memoria digital de la Fundación Sancho el Sabio. In E. Roseras (Ed.), *Gestión de la innovación y nuevas estrategias de investigación y difusión del fondo documental artístico* (pp. 161–176). Gijón.
- Grupo de trabajo de la Línea estratégica 2 de REBIUN. (2009). Preservación digital: Guía de recursos. Retrieved from [http://www.rebiun.org/opencms/opencms/handle404?exporturi=/export/docReb/Guia\\_preservacion\\_rebiun.pdf&\]](http://www.rebiun.org/opencms/opencms/handle404?exporturi=/export/docReb/Guia_preservacion_rebiun.pdf&)
- Grupo de Trabajo de Patrimonio Digital. (2013). *Informe del Consejo de Cooperación Bibliotecaria*. Retrieved from [http://www.mcu.es/bibliotecas/docs/MC/ConsejoCb/GruposTrabajo/GT\\_Patrimonio\\_BP/patrimoniodiglinforme2012.pdf](http://www.mcu.es/bibliotecas/docs/MC/ConsejoCb/GruposTrabajo/GT_Patrimonio_BP/patrimoniodiglinforme2012.pdf)
- Guallar, J., & Abadal, E. (2009). Fuentes de información sobre prensa digital: una propuesta de clasificación. *I Congreso internacional de ciberperiodismo y web 2.0*. Bilbao. Retrieved from <http://eprints.rclis.org/13767/>
- Guallar, J., & Abadal, E. (2010). *Prensa digital y bibliotecas* (p. 176). Gijón: Trea. Retrieved from <http://www.trea.es/ficha.php?idLibro=00001022>
- Henriksen, S. L., Seuskens, W., & Wijers, G. (2013). *Best practices for a digital storage infrastructure for the long-term preservation of digital files* (pp. 1–51).
- Herrerías Rey, J. E. (2012). *El PC, hardware y componentes* (p. 735). Madrid: Anaya Multimedia.
- IFLA, ICA, & UNESCO. (2002). *Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos*. Retrieved from <http://travesia.mcu.es/portalnb/jspui/handle/10421/3342>

- Keefer, A. (2007). Preservación digital y depósitos institucionales. In *Anuario ThikEPI* (pp. 196–199).
- Keefer, A., & Gallart, N. (2003). La preservación digital y las universidades: el estado de la cuestión. In *8as Jornadas Españolas de Documentación*. Barcelona: Fesabid. Retrieved from <http://eprints.rclis.org/6780/>
- Keefer, A., & Gallart, N. (2007). *La preservación de recursos digitales: el reto para las bibliotecas del siglo XXI*. Barcelona: UOC.
- LeFurgy, B. (2013). Picturing perpetual beta for digital preservation. *The Signal: Digital Preservation [Blog, The Libray of Congress]*, (May 3). Retrieved from <http://blogs.loc.gov/digitalpreservation/2013/05/picturing-perpetual-beta-for-digital-preservation/>
- Llueca, C. (2012a). Preservación digital. Retrieved from <http://hdl.handle.net/2445/32212>
- Llueca, C. (2012b). Preservación digital en cinco pasos. In *OpenAccesWeek Perú*. Pontificia Universidad Católica del Perú (PUCP). Retrieved from <http://hdl.handle.net/10760/18922>
- Llueca, C., & Cócera, D. (2012). El archivo del patrimonio digital en Internet. In E. Roseras (Ed.), *Gestión de la innovación y nuevas estrategias de investigación y difusión del fondo documental artístico* (pp. 223–235). Gijón: Trea.
- Llueca, C., & Reoyo, S. (2012). Repositorios sostenibles; reflexiones a partir de la experiencia española. Lima. Retrieved from <http://hdl.handle.net/2445/32133>
- Macías Alemán, V. M., & Torres Betancor, J. (2005). Digitalización masiva y acceso en línea a prensa regional. Barcelona. Retrieved from [http://bibliotecna.upc.es/rebiun/workshop5/presentacions/23\\_bulpgc\\_macias\\_betancor.pdf](http://bibliotecna.upc.es/rebiun/workshop5/presentacions/23_bulpgc_macias_betancor.pdf)
- Martín, M. (2009). La documentación del grupo Vocento. In T. Agirreazaldegí (Ed.), *I Jornadas sobre Documentación y Medios de Comunicación, 2008* (pp. 203–209). Bilbao: Euskal Herriko Unibertsitatea = Universidad del País Vasco.
- Montes, S., Alcaraz-Martínez, R., Vicente-Hernández, L., Blanco, M., Comas, M., Nieto, L., ... Bilbeny, M. (2013). El archivo de la palabra del Ateneu Barcelonès: un repositorio sostenible para la curación de contenidos. In *XIII Jornadas Españolas de Documentación FESABID'13* (pp. 216–222). Toledo: Federación Española de Sociedades de Archivística, Biblioteconomía, Documentación y Museística. Retrieved from [http://eprints.rclis.org/19463/1/archivo\\_de\\_la\\_palabra.pdf](http://eprints.rclis.org/19463/1/archivo_de_la_palabra.pdf)
- Moreno, P. (2002). Las bibliotecas digitalizadas: la salvaguarda de los materiales libraríos. In J. A. Magán (Ed.), *Temas de biblioteconomía universitaria y general* (pp. 376–390). Madrid: Editorial Complutense.

- Muñoz de Solano, B. (2006). Aspectos económicos del proceso de preservación digital. *Revista española de documentación científica*, 29(4, octubre-diciembre), 572–587. Retrieved from <http://redc.revistas.csic.es/index.php/redc/article/download/309/384>
- Muñoz de Solano, B. (2010). La importancia de utilizar metadatos PREMIS. Primer paso para la preservación de objetos digitales. In *V Congreso Nacional de Bibliotecas Públicas. Gijón, 2010* (p. 11). Retrieved from <http://www.mcu.es/bibliotecas/docs/MC/2010/CongresoBP/2BarbaraMunoz.pdf>
- Muñoz, I., & Aldaz, I. (2009). El servicio de documentación del Grupo Noticias. In T. Agirreazaldegui (Ed.), *I Jornadas sobre Documentación y Medios de Comunicación, 2008* (pp. 229–234). Bilbao: Euskal Herriko Unibertsitatea = Universidad del País Vasco.
- Observatorio Vasco de la Cultura. (2011). Formatos de difusión y formatos de preservación de contenidos digitales. *Kultura 2.0 Cuadernos de formación*, 2.
- Oleaga, J. (2009). Sistema de trabajo del centro de documentación del periódico El Correo. In T. Agirreazaldegui (Ed.), *I Jornadas sobre Documentación y Medios de Comunicación, 2008* (pp. 184–202). Bilbao: Euskal Herriko Unibertsitatea = Universidad del País Vasco.
- OMPI. SCCR/23/8: Documento de trabajo en el que figuran los comentarios acerca de un Instrumento Jurídico Internacional Adecuado (independientemente de su forma) sobre Excepciones y Limitaciones para las Bibliotecas y los Archivos, y las Sugerencias de Contenido (2012). Ginebra: OMPI. Retrieved from [http://www.wipo.int/meetings/es/doc\\_details.jsp?doc\\_id=213607](http://www.wipo.int/meetings/es/doc_details.jsp?doc_id=213607)
- Ondarenet, patrimonio digital vasco. (2013). *Aurrera, Boletín Divulgativo de Innovación y Nuevas Tecnologías*, 45(septiembre 2013), 12.
- Ortiz Ancona, D. (2012, November 29). Preservación digital: problemáticas, estrategias, metadatos, infraestructura y políticas. Dirección General de Bibliotecas UNAM. Retrieved from <http://hdl.handle.net/10760/18590>
- Pérez Cervera, M. (2012, October 1). *Una aproximación al data curation y el rol del bibliotecario en su implantación: el caso de la Universitat Oberta de Catalunya [Trabajo Fin de Máster]*. Universidad Carlos III de Madrid. Retrieved from <http://e-archivo.uc3m.es/handle/10016/16251>
- Pérez, K., & Serra, E. (2010). *Repositori de preservació digital de la Biblioteca de Catalunya: informe descriptiu i de situació* (p. 24). Barcelona. Retrieved from <http://www.recercat.cat/handle/2072/97251>
- Pérez San-José, P., de la Fuente Rodríguez, S., Álvarez Alonso, E., García Pérez, L., & Gutiérrez Borge, C. (2013). *Guía para entidades locales: cómo ahorrar costes y mejorar la productividad con cloud computing* (Enero 2013., p. 52). Madrid: INTECO.

- Plan Euskadi en la Sociedad de la Información. (2008). *Agenda Digital de Euskadi 2010*. Vitoria-Gasteiz. Retrieved from [http://www.euskadi.net/r33-2287/es/contenidos/enlace/psi2010/es\\_psi2010/adjuntos/PESI2010\\_c.pdf](http://www.euskadi.net/r33-2287/es/contenidos/enlace/psi2010/es_psi2010/adjuntos/PESI2010_c.pdf)
- Plan Euskadi en la Sociedad de la Información. (2012). *Agenda Digital de Euskadi 2015*. Vitoria-Gasteiz. Retrieved from [http://www.innova.euskadi.net/contenidos/informacion/innovacion\\_euskadi/es\\_innova/AGENDA\\_DIGITAL\\_DE\\_EUSKADI\\_2015.pdf](http://www.innova.euskadi.net/contenidos/informacion/innovacion_euskadi/es_innova/AGENDA_DIGITAL_DE_EUSKADI_2015.pdf)
- Prieto, A., Lloris, A., & Torres, J. C. (2010). *Introducción a la informática* (4ª ed., p. 808). Madrid: McGraw-Hill.
- Pujari, A., & Pai, P. (2005). Newspaper archives : a knowledge asset. In *International Conference on Information Management in a Knowledge Society : ICIM*. Mumbai, India. Retrieved from <http://eprints.rclis.org/8276/>
- Pulgar Vernalte, F. (2013). Liburuklik: patrimonio digital vasco en acceso abierto. In *XIII Jornadas Españolas de Documentación. FESABI'13* (pp. 204–210). Toledo: Federación Española de Sociedades de Archivística, Biblioteconomía, Documentación y Museística.
- REBIUN. (1999). *Normas y directrices para bibliotecas universitarias y científicas* (2ª ed. aum., p. 99). Madrid: Ministerio de Educación y Cultura, Dirección General del Libro, Archivos y Bibliotecas. Retrieved from [http://www.rebiun.org/opencms/opencms/handle404?exporturi=/export/docReb/Bibliotecas\\_Universitarias\\_y\\_Cientificas.pdf&\]](http://www.rebiun.org/opencms/opencms/handle404?exporturi=/export/docReb/Bibliotecas_Universitarias_y_Cientificas.pdf&)
- Rivera Donoso, M. A. (2009). Directrices para la creación de un programa de preservación digital. *Serie Bibliotecología y Gestión de Información*, 43, 1–63. Retrieved from <http://hdl.handle.net/10760/12989>
- Robledano, J. (2012). Tema 1: Tecnología de la imagen digital ráster. Apuntes de la asignatura “Digitalización y Preservación”. Máster en Bibliotecas y Servicios de Información Digital. UC3M. Curso 2010/2011.
- Rodríguez Mateos, D., Hernández Pérez, T., & Méndez Rodríguez, E. (2012). Archivos y centros de documentación en la prensa local de la Comunidad de Madrid. *Documentación de las Ciencias de la Información*, 35, 11–30. doi:10.5209/rev\_DCIN.2012.v35.40444
- Rohde-Enslin, S., & Allen, K. R. (2009). *Nothing last forever* (p. 66). Berlin. Retrieved from <http://nbn-resolving.de/urn:nbn:de:0008-2010031529>
- Rothenberg, J. (1995). ¿Son perdurables los documentos digitales? *Investigación y ciencia*, 222, 8–13.
- Rothenberg, J. (1999). *Ensuring the Longevity of Digital Information*. Santa Monica, CA.

- Rothenberg, J. (2004). ¿Son perdurables los documentos digitales? *Temas de investigación y ciencia*, 36, 52–57.
- Salvador Benítez, A., & Ruiz Rodríguez, A. Á. (2005). Metadatos para la preservación de colecciones digitales. *Cuadernos de documentación multimedia*, 16, 48–60. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=2048508>
- Salvador López, F. (2005). Políticas y planes de preservación en bibliotecas: actitudes, contenido y objetivos. In *I Seminario sobre Patrimonio Bibliográfico Vasco, 2003* (pp. 189–210). Vitoria-Gasteiz: Fundación Sanch el Sabio; Gobierno Vasco, Departamento de Cultura.
- Soler, J. (2008). *La preservación de los documentos electrónicos*. Barcelona: UOC.
- Térmens, M. (2009a). Investigación y desarrollo en preservación digital: un balance internacional. *El Profesional de la Información*, 18(6), 613–624. doi:10.3145/epi.2009.nov.04
- Térmens, M. (2009b). Los archivos y las bibliotecas ante la preservación digital ¿sólo un enfoque? In SEDIC (Ed.), *Actas XI Jornadas de Gestión de la Información* (pp. 115–122). Madrid.
- Térmens, M. (2010a). Preservación digital en 2009. In *Anuario ThikEPI* (vol. 4., pp. 224–230).
- Térmens, M. (2010b). Gestión, uso y preservación de contenidos digitales: las necesidades técnicas frente a los derechos de propiedad intelectual. In *V Congreso Nacional de Bibliotecas Públicas. Gijón, 2010* (pp. 37–42). Gijón. Retrieved from <http://bd.ub.edu/pub/termens/docs/VCongresoBibsPublicas.pdf>
- Térmens, M. (2012). Digitalizació del patrimoni hemerogràfic i la seva preservació. In *Taula de Nova Recerca*. Barcelona. Retrieved from [http://www.upf.edu/tnr/\\_pdf/Miquel-Termens-TNR-Digitalitzacio.pdf](http://www.upf.edu/tnr/_pdf/Miquel-Termens-TNR-Digitalitzacio.pdf)
- Térmens, M. (2013). *Preservación digital* (p. 109). Barcelona: UOC.
- UNESCO. (2003a). *Directrices para la preservación del patrimonio digital*. Retrieved from <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>
- UNESCO. (2003b). *Carta sobre preservación del patrimonio digital*. Retrieved from <http://unesdoc.unesco.org/images/0013/001331/133171s.pdf#page=85>
- UNESCO/UBC. (2012). *Declaración de Vancouver. La Memoria del Mundo en la era digital: digitalización y preservación*. Vancouver, Canada. Retrieved from [http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/unesco\\_abc\\_vancouver\\_declaration\\_es.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/unesco_abc_vancouver_declaration_es.pdf)

- Urkia, A. (2008). La Red de bibliotecas municipales de Donostia San Sebastián año 2008 . Su modelo de gestión , sus proyectos cooperativos. A Coruña. Retrieved from <http://hdl.handle.net/10421/534>
- Valverde Tejedor, C. (2005). Un proyecto de digitalización en el Parlamento Vasco: entre la preservación y el acceso. In *I Seminario sobre Patrimonio Bibliográfico Vasco, 2003* (pp. 259–271). Vitoria-Gasteiz: Fundación Sanch el Sabio; Gobierno Vasco, Departamento de Cultura.
- Vives, J., Alberch, R., Álvarez, J., Cuevas, A., Labastida, I., Munilla, G., ... Solanilla, L. (2009). *La digitalización del patrimonio: archivos, bibliotecas y museos en la red.* (J. Vives, Ed.). Editorial UOC. Retrieved from <http://dialnet.unirioja.es/servlet/libro?codigo=439431>
- Wilson, C. (2011). Introduction to Preservation Planning. In *Digital Preservation Summit 2011.* Hamburg.
- Zubiaur Carreño, F. J. (2004). *Curso de museología.* Gijón: Trea.



# GLOSARIO

---

Esta lista únicamente abarca los términos realmente utilizados en el texto. Las definiciones de los términos en **negrita** han sido extraídas literalmente del glosario de las **Directrices para proyectos de digitalización** (IFLA, ICA, y UNESCO, 2002).

**ASCII** – (*American Standard Code for Information Interchange*) La norma internacional de facto para los números codificados utilizados por los ordenadores para representar las letras mayúsculas y minúsculas del alfabeto latino, los números, puntuación, etc.

**bit** – (*Binary DigIT*) La unidad de medida de información equivalente a la elección entre dos posibilidades igualmente probables.

**byte** – Unidad de información digital. 1 byte equivale a 8 bits. Un megabyte equivale a 1 millón de bytes.

**Captura de la imagen** – Proceso por el que se obtiene una representación digital de un original constituida por un conjunto de elementos pictóricos o píxeles mediante el escaneado o fotografía digital.

CD-ROM – (*Compact Disc Read Only Memory*) Soporte óptico de almacenamiento de sólo lectura de 650/700 MB de capacidad.

**Compresión** – La reducción del tamaño del fichero de la imagen para su proceso, almacenamiento y transmisión. La compresión puede ser con pérdida o sin pérdida.

**Compresión con pérdida** – Proceso de compresión que reduce el espacio de almacenamiento necesario para el fichero de una imagen mediante la eliminación de datos de la imagen. Al descomprimir una imagen que ha experimentado una compresión con pérdida siempre será distinta de la imagen antes de que se comprimiera, incluso aunque la diferencia sea difícil de detectar para el ojo humano.

**Compresión sin pérdida** – Proceso de compresión que reduce el espacio de almacenamiento necesario para el fichero de una imagen sin pérdida de datos. Si una imagen ha experimentado una compresión sin pérdida, será idéntica a la imagen antes de que se comprimiera.

Digitalizar – Conversión de documentos en soporte no eléctrico a un formato digital procesable por un ordenador.

*Digitisation* o *Digitization* – Digitalizar.

Disco duro externo – (*External Hard Disc*, HD) Disco duro independiente portátil, con su propia carcasa y fuente de alimentación que se conecta al ordenador a través de cable de datos o USB.

Dominio público – Obras libres de derechos de explotación conforme a la legislación de propiedad intelectual.

Dublin Core – Modelo de metadatos elaborado y mantenido por la *Dublin Core Metadata Initiative* (DCMI), de amplio uso. “DC simple” es un conjunto de 15 elementos de metadatos utilizado para describir recursos electrónicos con vistas a facilitar la recuperación. “DC cualificado” extiende el uso a más elementos. El *Dublin Core Metadata Element Set* es norma NISO Z39.85-2001 e ISO 15836:2009.

DVD-R – (*Digital Versatile Disc*) Disco óptico grabable multisesión, puede almacenar 4,7 GB de datos (monocapa), 8,5 GB (bicapa) o 17 GB (doble cara)

DVD+R – (*Digital Versatile Disc*) Disco óptico grabable multisesión, puede almacenar 4,7 GB de datos (monocapa), 8,5 GB (bicapa) o 17 GB (doble cara).

**Escala de grises** – El número de grises en la escala entre el blanco y el negro. Una imagen codificada con 8 bits puede representar una escala de 254 grises entre el negro y el blanco.

**Escáner** – Dispositivo de digitalización que entrega los datos de las imágenes escaneadas a un ordenador.

**Ficheros de miniaturas** – Ficheros digitales que representan las imágenes en versiones pequeñas, de baja resolución.

Imagen ráster – También llamada trama, mapa de bits o *bitmap*. Imagen compuesta por píxeles. Permite conseguir imágenes muy cercanas a la apariencia de los documentos originales, lográndose una vista cuasi-facsimilar.

JPEG – (*Joint Photographic Experts Group*) Norma ideada para conseguir ratios de compresión muy altos en imágenes fotográficas. Más que un formato es una serie de algoritmos de compresión. Realiza la compresión con pérdida siendo la

tasa de compresión escalable, de modo que a mayor compresión más evidente es la pérdida de información. Provoca artefactos de compresión visibles al ojo humano como halos alrededor de zonas de detalle y bloques de color sólidos.

**LAM** – (*Libraries, Archives and Museums*) Bibliotecas, archivos y museos, la UNESCO las denomina *Instituciones de la Memoria*.

**LZW** – (*Lempel-Ziv Welch*) Tipo de sistema de compresión de un fichero digital ampliamente utilizado con los formatos TIFF y GIF para reducir el tamaño del fichero sin pérdida de datos.

**MARC** – (*Machine Readable Cataloguing*) Los formatos MARC son normas para la representación y comunicación de información bibliográfica y relacionada de forma legible por ordenador.

**Marca de agua** – En términos digitales, una marca de agua visible o invisible que se crea mediante la inserción de un modelo de bits en una imagen digital para identificar que el propietario de la imagen es conocido.

**Metadatos** – Datos acerca de los datos, o la información que se conoce acerca de la imagen para proporcionar acceso a dicha imagen. Por lo general, incluye información sobre el contenido intelectual de la imagen, sobre la representación de los datos digitales e información sobre la gestión de derechos o de la seguridad.

**Migración** – Preservación de la integridad de las imágenes digitales mediante la transferencia de éstas entre las configuraciones de hardware y de software y entre generaciones sucesivas de tecnología informática.

**OAIS** – (*Open Archival Information System*)

**OCR** – (*Optical Character Recognition*) Reconocimiento Óptico de Caracteres. Capacidad de un dispositivo de entrada de un ordenador u otra máquina de leer ópticamente caracteres de texto individuales de una página y convertir la información a un fichero de texto almacenado electrónicamente.

**Píxel** – Abreviatura de *picture element*. Una fotografía electrónica escaneada a partir de un documento original se compone de un conjunto de píxeles. A cada píxel se le asigna un valor tonal (blanco, negro, un tono de gris o un color) y se representa digitalmente mediante un código binario (ceros y unos).

**PPP** – Píxel Por Pulgada.

**PPI** – (*Pixel Per Inch*) Píxel Por Pulgada.

**Profundidad del bit** – La profundidad del bit de una imagen se refiere al número de bits utilizado para describir el color de cada píxel.

**Resolución** – El número de píxeles (tanto en altura como a lo ancho) que constituye una imagen, expresado normalmente como el número de píxeles por unidad lineal, p. ej. 300ppi (píxeles por pulgada) o a veces como dpi (puntos por pulgada). Cuanto mayor es el número de píxeles de una imagen, más alta es su resolución, y cuanto mayor es la resolución de una imagen, mayores son su claridad y su definición (y mayor es el tamaño del fichero). Para la resolución del color, véase Profundidad del bit.

**TIC** – Tecnologías de la Información y la Comunicación.

**TIFF** – (*Tagged Image File Format*) Formato de fichero para el almacenamiento de imágenes mapeadas bit-0. Puede presentar cualquier resolución, blanco y negro, escala de grises o color.

**UPS** – (*Uninterruptible Power Supply*) Sistema de Alimentación Ininterrumpida. Dispositivo eléctrico diseñado para intervenir inmediatamente y suministrar energía durante un corto período de tiempo a dispositivos electrónicos tales como ordenadores cuando fallan los medios normales de suministro.

**XML** – (*Extensible Mark-up Language*) Lenguaje de Marcado Extensible. Metalenguaje que permite crear lenguajes de marcado para definir la estructura y contenido de un tipo concreto de documentos. Es una versión abreviada de SGML. Sistema ampliamente utilizado para definir formatos de datos, documentos especialmente complejos y estructuras de datos. Apto para cualquier plataforma porque los archivos XML son archivos de texto llano.

# ANEXOS

## ANEXO I

---

### **Evolución de los integrantes del Consorcio de Digitalización de Prensa Vasca.**

En la primera acta de reunión del Consorcio de Digitalización de Prensa Vasca que se conserva, fechada el 25/04/1994, se enumeran las ocho instituciones participantes en aquellos primeros momentos. Se trata de las siete instituciones fundadoras (1991), a las cuales se había sumado en ese mismo de 1994 el Centro de Documentación del Departamento de Interior del Gobierno Vasco:

- Archivo del Territorio Histórico de Álava (Diputación Foral de Álava).
- Biblioteca Pública de Vitoria-Gasteiz (Diputación Foral de Álava).
- Biblioteca de la Facultad de Filología e Historia de la Universidad del País Vasco.
- Biblioteca del Parlamento Vasco.
- Biblioteca General del Gobierno Vasco.
- Centro de Documentación del Departamento de Interior del Gobierno Vasco.
- Fundación Sancho el Sabio.
- Koldo Mitxelena Kulturunea (Diputación Foral de Gipuzkoa).

Paulatinamente otras instituciones se van sumando al proyecto. La Universidad de Nevada (Reno, Nevada, EE.UU) solicita participar en el proyecto de digitalización de prensa en 1996 (acta del 16/04/1996). Tanto el *Basque Studies Program* de *The University Library* de la citada universidad, como la Red de Bibliotecas Municipales de Bilbao se incorporan al proyecto con fecha 1 de enero de 1997. En la misma fecha cursa baja el Archivo del Territorio Histórico de Álava (acta 19/03/1997).

En acta de 22/01/1998 se aprueba la entrada del Ayuntamiento de San Sebastián, a través de su Red Municipal de Bibliotecas. El acta de 22/10/1998 recoge la aprobación de la incorporación de la Biblioteca de la Universidad de Navarra. En la misma se deniega a la Hemeroteca de Caixa Tarragona la incorporación y se acuerda “que no entrarán en el Consorcio instituciones que no sean de Euskal Herria”. En acta de 16/12/1999 se aprueba la entrada del Departamento de Documentación y Archivo de *Euskal Telebista* (ETB, Televisión Pública Vasca). Esta misma acta contiene una relación de las doce instituciones participantes en ese momento:

- Biblioteca de la Universidad de Navarra.
- Biblioteca de la Universidad del País Vasco.
- Biblioteca del Parlamento Vasco.
- Biblioteca General del Ayuntamiento de San Sebastián.
- Biblioteca General del Gobierno Vasco.
- Departamento de Documentación y Archivo de Euskal Telebista.
- Departamento de Interior del Gobierno Vasco.
- Diputación Foral de Álava, Casa de Cultura de Vitoria-Gasteiz.
- Diputación Foral de Gipuzkoa, Koldo Mitxelena Kulturunea.
- Fundación Sancho el Sabio.
- Red de Bibliotecas Municipales de Bilbao.
- Universidad de Nevada, Reno.

En acta de 04/05/2000 se aprueba la incorporación de la Biblioteca Nacional de España (BNE). El acta de 16/01/2003 recoge la aceptación de la incorporación del Departamento de Cultura del Gobierno Vasco. En acta de 30/09/2003 se acuerda comunicar a la Universidad de Nevada que “(...) al no haber recibido noticias de la institución [desde 1999] se dio de baja a la misma”. Con fecha 12 de diciembre de 2006 el Departamento de Interior del Gobierno Vasco abandona el Consorcio (acta de 13/12/2006). En la primavera del año 2007 tanto BNE como EITB comunican que en 2008 abandonarán el Consorcio (acta de 13/06/2007). El 18 de junio de 2007 la Biblioteca Foral de Bizkaia solicita su admisión en el Consorcio. Se acepta su entrada a partir del 1 de enero de 2008 (acta de 14/11/2007). En carta de 8 de mayo de 2008 la Biblioteca de la Universidad de Navarra comunica su abandono del Consorcio (acta de 20/11/2008).

Tabla 14.- Integrantes del Consorcio de Digitalización de Prensa Vasca, con fechas de incorporación y abandono del mismo

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	
ATHA	✕					✕																		
Biblioteca Pública Vitoria	✕																							⇒
UPV/EHU	✕																							⇒
Parlamento Vasco	✕																							⇒
Biblioteca Central Gob. Vasco	✕																							⇒
Dpto. Interior Gob. Vasco	✕																✕							
Fundación Sancho el Sabio	✕																							⇒
Koldo Mitxelena	✕																							⇒
Universidad de Nevada							✕						✕											
Red Municipal Bilbao							✕																	⇒
Red Municipal Donostia								✕																⇒
Universidad de Navarra								✕											✕					
EITB									✕									✕						
BNE										✕								✕						
Dpto. Cultura Gob. Vasco													✕											⇒
Biblioteca Foral Bizkaia																			✕					⇒

## ANEXO II

### Evolución de las cabeceras digitalizadas por el Consorcio de Digitalización de Prensa Vasca.

Al arrancar el proceso de digitalización el 1 de enero de 1996 se comienzan a escanear las ocho cabeceras que hasta ese momento se venían microfilmando, se trata de: Deia, Diario de Navarra, Diario de Noticias, Diario Vasco, Egin, Egunkaria, El Correo Español-El Pueblo Vasco (ed. Álava) y El Mundo del País Vasco. En acta de 16/04/1996 se decide abordar la digitalización retrospectiva del diario Egin. El primero de enero de 1998 se comienza a digitalizar El Correo Español edición Bizkaia. En 1998 se comienza también a digitalizar El País, edición País Vasco, y El Periódico de Álava retrospectivamente desde el primer número en ambos casos. En 1999 se decide (acta de 16/12/1999) digitalizar el periódico Zabalik desde el nº 1.

Relación de periódicos que se están digitalizando y sus ediciones en 1999, según acta de 16/12/1999:

1. El Correo Español-El Pueblo Vasco, edición Álava. Desde 1996
2. El Correo Español-El Pueblo Vasco, edición Bizkaia. Desde 1998
3. El Mundo del País Vasco. Desde 1996
4. Gara. Desde 1998
5. Diario de Noticias. Desde 1996
6. Diario de Navarra. Desde 1996
7. Egunkaria. Desde 1996
8. Diario Vasco. Desde 1996
9. Deia. Desde 1996
10. El País, edición País Vasco. Desde mayo de 1998
11. El Periódico de Álava. Desde 1998

Relación de periódicos que se están digitalizando en 2002 y sus ediciones, según acta de 14/02/2002:

1. El Correo Español-El Pueblo Vasco, edición Álava. Desde 1996
2. El Correo Español-El Pueblo Vasco, edición Bizkaia. Desde 1998
3. El Mundo del País Vasco. Desde 1996
4. Gara. Desde 1998
5. Diario de Noticias. Desde 1996
6. Diario de Navarra. Desde 1996
7. Egunkaria. Desde 1996
8. Diario Vasco. Desde 1996
9. Deia. Desde 1996
10. El País, edición País Vasco. Desde mayo de 1998
11. El Periódico de Álava. Desde 1998
12. Zabalik. Desde noviembre de 1999, nº 1

En 2002 se decide (acta de 18/04/2002) digitalizar el periódico Le Journal du Pays Basque, desde su primer número publicado en octubre de 2001. Se decide dejar de digitalizar el periódico Zabalik "que ha pasado a ser revista de información general" a partir de final del año 2002 (acta de 27/06/2002). En 2004 se decide (acta de 14/12/2004) digitalizar el periódico Diario de Noticias de Álava, que acaba de aparecer. También se aprueba la digitalización de La Estafeta, aparecido asimismo en el transcurso del año 2004. En 2006 se decide (acta de 01/05/2006) digitalizar el periódico Diario de Noticias de Gipuzkoa, aparecido en noviembre de 2005.

Relación de periódicos que se están digitalizando entre los años 2008 y 2013 y sus ediciones, según presupuestos presentados por las empresas:

- |   |                                 |
|---|---------------------------------|
| 1. Berria   | 10. El Mundo del País Vasco     |
| 2. Deia   | 11. El País, edición País Vasco |
| 3. Diario de Navarra                                  | 12. Gara                        |
| 4. Diario de Noticias                                 | 13. Le Journal du Pays Basque   |
| 5. Diario de Noticias de Álava                        |                                 |
| 6. Diario de Noticias de Gipuzkoa                     |                                 |
| 7. Diario Vasco                                       |                                 |
| 8. El Correo Español-El Pueblo Vasco, edición Álava   |                                 |
| 9. El Correo Español-El Pueblo Vasco, edición Bizkaia |                                 |

## ANEXO III

### Información completa de una toma fotográfica en formato de metadatos EXIF.

Descripción de metadatos técnicos EXIF de toma fotográfica, extraída mediante la herramienta ExifTool 9.30. Obsérvese la gran riqueza de metadatos y la amplia posibilidad para autodocumentar las imágenes en el momento mismo de la toma mediante una parametrización de cámara adecuada.

```
---- ExifTool ----
ExifTool Version Number: 9.30

--- System ---
File Name: IMG_0950.JPG
Directory: C:/UC3M/TFM
File Size: 3.6 MB
File Modific. Date/Time: 2013:05:03
07:49:51+02:00
File Access Date/Time: 2013:10:09
06:34:56+02:00
File Creation Date/Time: 2013:10:09
06:34:56+02:00
File Permissions: rw-rw-rw-

---- File ----
File Type: JPEG
MIME Type: image/jpeg
Exif Byte Order: Little-endian (Intel, II)
Image Width: 4320
Image Height: 3240
Encoding Process: Baseline DCT, Huffman
coding
Bits Per Sample: 8
Color Components: 3
Y Cb Cr Sub Sampling: YCbCr4:2:2 (2 1)

---- IFD0 ----
Image Description:
Make: Canon
Camera Model Name: Power Shot SX210 IS
Orientation: Horizontal (normal)
X Resolution: 180
Y Resolution: 180
Resolution Unit: inches
Modify Date: 2013:04:01 12:39:34
Y Cb Cr Positioning: Co-sited

---- ExifIFD ----
Exposure Time: 1/100
F Number: 5.0
ISO: 640
Exif Version: 0221
Date/Time Original: 2013:04:01 12:39:34
Create Date: 2013:04:01 12:39:34
Components Configuration: Y, Cb, Cr, -
Compressed Bits Per Pixel: 3
Shutter Speed Value: 1/101
Aperture Value: 5.0

Exposure Compensation: 0
Max Aperture Value: 5.0
Metering Mode: Multi-segment
Flash: Off, Did not fire
Focal Length: 24.7 mm
User Comment:
Flashpix Version: 0100
Color Space: sRGB
Exif Image Width: 4320
Exif Image Height: 3240
Focal Plane X Resolution: 17777.77778
Focal Plane Y Resolution: 17802.1978
Focal Plane Resolution Unit: inches
Sensing Method: One-chip color area
File Source: Digital Camera
Custom Rendered: Normal
Exposure Mode: Auto
White Balance: Auto
Digital Zoom Ratio: 1
Scene Capture Type: Standard

---- Canon ----
Macro Mode: Normal
Self Timer: Off
Quality: Fine
Canon Flash Mode: Off
Continuous Drive: Single
Focus Mode: Single
Record Mode: JPEG
Canon Image Size: Large
Easy Mode: Full auto
Digital Zoom: None
Contrast: Normal
Saturation: Normal
Sharpness: 0
Camera ISO: Auto
Metering Mode: Evaluative
Focus Range: Auto
AF Point: Face Detect
Canon Exposure Mode: Easy
Lens Type: Unknown (-1)
Max Focal Length: 70 mm
Min Focal Length: 5 mm
Focal Units: 100/mm
Max Aperture: 5
Min Aperture: 11
Flash Bits: (none)
Focus Continuous: Continuous
AE Setting: Normal AE
```

**Image Stabilization:** On  
**Zoom Source Width:** 4320  
**Zoom Target Width:** 4320  
**Spot Metering Mode:** Center  
**Manual Flash Output:** n/a  
**Focal Type:** Zoom  
**Focal Length:** 24.71 mm  
**Focal Plane X Size:** 6.32 mm  
**Focal Plane Y Size:** 4.75 mm  
**Canon Flash Info:** 0 0 0 0  
**Auto ISO:** 658  
**Base ISO:** 100  
**Measured EV:** 9.22  
**Target Aperture:** 5  
**Target Exposure Time:** 1/101  
**Exposure Compensation:** 0  
**White Balance:** Auto  
**Slow Shutter:** Off  
**Shot Number In Continuous Burst:** 0  
**Optical Zoom Code:** 80  
**Flash Guide Number:** 0  
**Flash Exposure Compensation:** 0  
**Auto Exposure Bracketing:** Off  
**AEB Bracket Value:** 0  
**Control Mode:** Camera Local Control  
**Focus Distance Upper:** 10.89 m  
**Focus Distance Lower:** 0 m  
**F Number:** 5  
**Exposure Time:** 1/97  
**Bulb Duration:** 0  
**Camera Type:** Compact  
**Auto Rotate:** None  
**ND Filter:** Off  
**Self Timer 2:** 0  
**Flash Output:** 0  
**Canon 0x0000:** 0 0 0 0 0 0  
**Canon Image Type:** PowerShot SX210 IS JPEG  
**Canon Firmware Version:** Firmware Version 1.00  
**File Number:** 120-0950  
**Owner Name:**  
**ISO:** 668  
**F Number:** 5  
**Exposure Time:** 1/97  
**Rotation:** 0  
**Camera Temperature:** 31 C  
**Canon Model ID:** PowerShot SX210 IS  
**AF Info Size:** 98  
**AF Area Mode:** Multi-point AF or AI AF  
**Num AF Points:** 9  
**Valid AF Points:** 9  
**Canon Image Width:** 4320  
**Canon Image Height:** 3240  
**AF Image Width:** 100  
**AF Image Height:** 100  
**AF Area Widths:** 18 18 18 18 18 18 18 18  
**AF Area Heights:** 18 18 18 18 18 18 18 18  
**AF Area X Positions:** -18 0 18 -18 0 18 -18 0 18  
**AF Area Y Positions:** -18 -18 -18 0 0 0 18 18 18  
**AF Points In Focus:** 4

**Canon AF Info 2 0x000d:** 0 0  
**Primary AF Point:** 4  
**Thumbnail Image Valid Area:** 0 0 0 0  
**Canon 0x0018:** 0 0 0 0 0 0 0 0 0 0 [...]  
**Canon 0x0019:** 1  
**Date Stamp Mode:** Off  
**My Color Mode:** Off  
**Firmware Revision:** 1.00 rev 3.00  
**Canon 0x001f:** 138 1 0 4 8 0 0 0 0 0 0 [...]  
**Canon 0x0022:** 416 0 0 16 8 1 1 640 4 [...]  
**Categories:** People  
**Intelligent Contrast:** Off  
**Image Unique ID:**  
 4215b98606e26e90337dc9a9591c2f51  
**VRD Offset:** 0  
**Canon 0x002d:** 0  
**Canon 0x002e:** 8 32767 32767 65535  
**Faces Detected:** 2

---- InteropIFD ----

**Interoperability Index:** R98 - DCF basic file (sRGB)  
**Interoperability Version:** 0100  
**Related Image Width:** 4320  
**Related Image Height:** 3240

---- IFD1 ----

**Compression:** JPEG (old-style)  
**X Resolution:** 180  
**Y Resolution:** 180  
**Resolution Unit:** inches  
**Thumbnail Offset:** 5120  
**Thumbnail Length:** 8192

---- XMP-xmp ----

**Rating:** 0

---- XMP-microsoft ----

**Date Acquired:** 2013:04:21 10:20:50

---- Composite ----

**Aperture:** 5.0  
**Drive Mode:** Single-frame Shooting  
**ISO:** 658  
**Image Size:** 4320x3240  
**Lens:** 5.0 - 70.0 mm  
**Lens ID:** Unknown 5-70mm  
**Scale Factor To 35 mm Equivalent:** 5.6  
**Shooting Mode:** Full auto  
**Shutter Speed:** 1/100  
**Thumbnail Image:** (Binary data 8192 bytes, use -b option to extract)  
**Circle Of Confusion:** 0.005 mm  
**Field Of View:** 14.8 deg  
**Focal Length:** 24.7 mm (35 mm equivalent: 138.7 mm)  
**Hyperfocal Distance:** 22.81 m  
**Lens:** 5.0 - 70.0 mm (35 mm equivalent: 28.1 - 392.7 mm)  
**Light Value:** 8.6

**Guía rápida para el cuidado y manejo de CD y DVD.**

Se adjunta la guía “CD and DVD Archiving: Quick Reference Guide for Care and Handling” (Digital Media Group, 2007).

**Quick Reference Guide for Care and Handling**

**Do:**

1. Handle discs by the outer edge or the center hole.
2. Use a non-solvent-based felt-tip permanent marker to mark the label side of the disc.
3. Keep dirt or other foreign matter from the disc.
4. Store discs upright (book style) in plastic cases specified for CDs and DVDs.
5. Return discs to storage cases immediately after use.
6. Leave discs in their packaging (or cases) to minimize the effects of environmental changes.
7. Open a recordable disc package only when you are ready to record data on that disc.
8. Store discs in a cool, dry, dark environment in which the air is clean.
9. Remove dirt, foreign material, fingerprints, smudges, and liquids by wiping with a clean cotton fabric in a straight line from the center of the disc toward the outer edge.
10. Use CD/DVD-cleaning detergent, isopropyl alcohol, or methanol to remove stubborn dirt or material.
11. Check the disc surface before recording.

**Do not:**

1. Touch the surface of the disc.
2. Bend the disc.
3. Use adhesive labels.
4. Store discs horizontally for a long time (years).
5. Open a recordable optical disc package if you are not ready to record.
6. Expose discs to extreme heat or high humidity.
7. Expose discs to extremely rapid temperature or humidity changes.
8. Expose recordable discs to prolonged sunlight or other sources of ultraviolet light.
9. Write or mark in the data area of the disc (the area the laser “reads”).
10. Clean by wiping in a direction going around the disc.

**For CDs especially do not:**

1. Scratch the label side of the disc.
2. Use a pen, pencil, or fine-tip marker to write on the disc.
3. Write on the disc with markers that contain solvents.
4. Try to peel off or reposition a label.

**General recommendations for long-term storage conditions:**  
 For archiving recordable (R) discs, it is recommended to use discs that have a gold metal reflective layer.

<i>Archival Storage Facility—Recommendation for storing CDs and DVDs together</i>		
Media	Temperature	Relative Humidity (RH)
CD, DVD	Less than 20°C (68°F) Greater than 4°C (39°F)	20% to 50% RH

A temperature of 18°C and 40% RH would be considered suitable for long-term storage.  
 A lower temperature and RH is recommended for extended-term storage.

Figura 17.- Guía para el cuidado y manejo de discos ópticos (Fuente: Digital Media Group, 2007)